Spring 2022

The Existence and Quantum Approximation of Optimal Pure State Ensembles

Ryan Thomas McGaha

The Existence and Quantum Approximation of Optimal Pure State Ensembles

by

Ryan Thomas McGaha

Bachelor of Arts
University of South Carolina - Upstate 2017

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Mathematics

College of Arts and Sciences

University of South Carolina

2022

Accepted by:

George Androulakis, Major Professor

Changhui Tan, Committee Member

Wolfgang Dahmen, Committee Member

Stephen Fenner, Committee Member

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

Dedication

To my wife Desiree, and to my parents Marion and Jacqueline. Your support through these four years made all of this possible.

Acknowledgments

I want to thank Dr. George Androulakis for his mentorship and guidance through this journey. He always encouraged my creative freedom while providing the right amount of guidance or gentle criticism needed to keep me on track. He has also been incredibly supportive and empathetic during the personal struggles I have faced during this time.

I'm also thankful for the unconditional love and support of my wife Desiree and my parents. All of them helped make this journey just a little bit easier and more manageable. I couldn't have done this without them.

Next I'd like to thank the various professors who have taught me in and out of class here at UofSC. Their wisdom, knowledge, and feedback have helped to mold me into a professional mathematician and I'm truly grateful.

Lastly I'd like to thank Dr. Brett Barwick who helped me to begin my journey in mathematics, and Charles Alexander and Craig Shiver for encouraging my curiosity in mathematics early on.

Abstract

In this manuscript we study entanglement measures defined via the convex roof construction. In the first chapter we build the notion of an entanglement measure from the ground up and discuss various issues that arise when trying to measure the amount of entanglement present in an arbitrary density operator. Through this introduction we will motivate the use of the convex roof construction. In the second chapter we will show that the infimum in the convex roof construction is achieved for a specific set of entanglement measures and provide canonical examples of such measures. We also describe LOCC operations via a tree structure and show this tree structure's utility in proving LOCC monotonicity for candidate entanglement measures. In the final chapter we supply a variational quantum algorithm which allows for the approximation of the convex roof construction and show that the algorithm experiences the problem of exponentially vanishing gradients for a functional we call an entanglement detector. We showcase some numerical experiments of the algorithm that illustrate convergence for a small number of qubits. The work presented herein is the merging of two works coauthored with Dr. George Androulakis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction to Entanglement

1.1 Overview

The quantification and classification of quantum entanglement play a fundamental role in quantum information theory. Bipartite entanglement, which is entnaglement between only two quantum systems, is well understood and is straightforward to detect for pure quantum states. However, for ensembles of pure states, the quantification of entanglement becomes more computationally intensive. Moreover there exist different notions of entanglement for ensembles, distillable and nondistillable. These two types of entanglement are differentiated by their ability to be used as a resource for quantum teleportation, a vital phenomenon in the transmission of quantum information. Distillable states allow for teleportation while nondistillable do not. This distinction has inspired different notions of what it means to measure the amount of entanglement present in a system. Some types of measures can separate distallable states from nondistillable and nonentangled states, while another can separate entangled states (whether they are distillable or not) from nonentangled states. In this text, we will focus on the second notion of measuring entanglement. We seek to only determine whether a state is entangled, and do not care if the state is distillable.

Entanglement can be detected via the use of a function called an *entanglement measure*. Usually these measures are interpreted as detecting the amount of information present when half of a bipartite system is discarded. If there is information still present, then state was entangled. If no information is present after discarding a system, then the state was nonentangled. These measures of information are usually constructed by an extension of generalized entropies to quantum states. For pure states, these entropies can be calculated precisely via the functional calculus. But for ensembles of states, these entropies must be extended via the *convex roof construction*, which is often difficult to compute directly. In what follows, we will show that convex roof construction indeed exists for all ensembles of quantum states as well as how to approximate the convex roof for a certain class of entanglement measures.

1.2 Entanglement

In this section we discuss the basic notions of entanglement. Broad overviews of the subject can be found in [31, 52, 26].

Denote by $\mathrm{D}(\mathcal{H})$ the set of density operators on a Hilbert space $\mathcal{H}$, and by $\mathrm{S}(\mathcal{H})$ the unit sphere of a Hilbert space $\mathcal{H}$. Throughout this work we will consider quantum states on a particular type of Hilbert space. Namely, we want to study Hilbert spaces of the form $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ where $\mathcal{A}$ and $\mathcal{B}$ are both finite dimensional Hilbert spaces over $\mathbb{C}$. The main questions that we will investigate are 1) whether or not a state $|\psi\rangle \in \mathcal{H}$ can be "factored" as $|\psi\rangle = |\psi_{\mathcal{A}}\rangle \otimes |\psi_{\mathcal{B}}\rangle$ for some $|\psi_{\mathcal{A}}\rangle \in \mathcal{A}$ and $|\psi_{\mathcal{B}}\rangle \in \mathcal{B}$ and 2) whether a density matrix in $\mathrm{D}(\mathcal{H})$ can be expressed as a convex sum of factored states.

We'll start by considering the first question: How do we know if a state can be factored, and what does this mean physically? Let $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ be a tensor product of finite dimensional Hilbert spaces over $\mathbb{C}$ and let $|\psi\rangle \in \mathcal{H}$. If $|\psi\rangle$ can be written as $|\psi\rangle = |\psi_{\mathcal{A}}\rangle \otimes |\psi_{\mathcal{B}}\rangle$ for some $|\psi_{\mathcal{A}}\rangle \in \mathcal{A}$ and $|\psi_{\mathcal{B}}\rangle \in \mathcal{B}$, then $|\psi\rangle$ is called a *product state*, otherwise $|\psi\rangle$ is said to be *entangled*. As an example, let $\mathcal{H} = \mathbb{C}^{2\otimes 2}$ and $|\psi\rangle = |01\rangle$. Then $|\psi\rangle = |0\rangle \otimes |1\rangle$ by definition, and so $|\psi\rangle$ is a product state. Physically, ff the $|0\rangle$ were discarded on the left side of the tensor product, then the state on the right side would be unaffected. The $|1\rangle$ would remain a $|1\rangle$. The "factor states" on $\mathcal{A}$ and $\mathcal{B}$ do not affect each other when $|\psi\rangle$ is a product state. This is in contrast to the case in which a state is entangled, which allows operations on $\mathcal{A}$ to alter the state of $\mathcal{B}$ and vice versa as we'll see in the next example. Now let $|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and let's observe what happens when we measure one of the qubits. If we apply the projection operator $|0\rangle\langle 0|$ to the first qubit of $|\psi\rangle$ we see that

$$\Big(|0\rangle\langle 0| \otimes I_{\mathcal{B}}\Big)|\psi\rangle = \frac{1}{\sqrt{2}}|00\rangle\,. \tag{1.1}$$

The measurement forced both states on $\mathcal{A}$ and $\mathcal{B}$ into the $|0\rangle$ state even though only space $\mathcal{A}$ was acted upon, thus Einstein's nickname for this phenomenon "spooky action at a distance". This use of entanglement will be useful later when we construct a quantum algorithm to measure the "amount of entanglement" present.

So how does one determine if a state is entangled? For bipartite pure states there a couple of options. One possibility is to use the purity theorem presented below along with its usual proof.

Theorem 1.2.1 (Purity Theorem)**.** *Let $|\psi\rangle$ on $\mathcal{H}$ where $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ for some finite dimensional Hilbert spaces $\mathcal{A}$ and $\mathcal{B}$. Then $|\psi\rangle$ is factorable if and only if $Tr_{\mathcal{A}}|\psi\rangle\langle\psi|$ is a pure state.*

Proof. Denote the purity,γ, of an density operator ρ by $\gamma(\rho) = \mathrm{Tr}\,(\rho^2)$. If ρ is a pure state, then it is a rank 1 projection and $\rho^2 = \rho$, and so it will have purity 1. Otherwise, ρ will have eigenvalues that are strictly between 0 and 1, and therefore the $\gamma(\rho)$ will be strictly less than 1. Using the purity function, we can rephrase the the above theorem as "$|\psi\rangle$ is factorable if and only if $\gamma(\mathrm{Tr}_{\mathcal{A}}|\psi\rangle\langle\psi|) = 1$.

$\implies\Big)$ Suppose that $|\psi\rangle$ can be written as $|\psi\rangle = |\alpha\rangle \otimes |\beta\rangle$ for some $\alpha \in \mathcal{A}$ and some $\beta \in \mathcal{B}$. Then $\mathrm{Tr}_{\mathcal{A}}|\psi\rangle\langle\psi| = |\beta\rangle\langle\beta|$ which is a pure state. Thus $\gamma\Big(\mathrm{Tr}_{\mathcal{A}}|\psi\rangle\langle\psi|\Big) = 1$.

$\impliedby\Big)$ Now suppose that $\gamma\Big(\mathrm{Tr}_{\mathcal{A}}|\psi\rangle\langle\psi|\Big) = 1$ for some pure state $|\psi\rangle \in \mathcal{H}$. Let

$\operatorname{Tr}_{\mathcal{A}} |\psi\rangle\langle\psi|$ have spectral decomposition $\operatorname{Tr}_{\mathcal{A}} |\psi\rangle\langle\psi| = \sum_j \sigma_j \Pi_j$ where each $\sigma_j \in [0, 1]$ with $\sum_j \sigma_j = 1$ and the Π_j are mutually orthonormal projections. Then

$$\gamma\Big(\operatorname{Tr}_{\mathcal{A}} |\psi\rangle\langle\psi|\Big) = \sum_j \sigma_j^2 = 1. \tag{1.2}$$

And since $\operatorname{Tr}_{\mathcal{A}} |\psi\rangle\langle\psi|$ is still a state,

$$\operatorname{Tr}\Big(\operatorname{Tr}_{\mathcal{A}} |\psi\rangle\langle\psi|\Big) = \sum_j \sigma_j = 1. \tag{1.3}$$

Now suppose towards a contradiction that all of the σ_j are strictly less than 1. Then $\sigma_j^2 < \sigma_j$ for each nonzero σ_j and therefore

$$\sum_j \sigma_j^2 < \sum_j \sigma_j = 1, \tag{1.4}$$

which is a contradiction. Thus there must be a unique eigenvalue σ_k with $\sigma_k = 1$ and all other eigenvalues 0. $\operatorname{Tr}_{\mathcal{A}}$ must then be a rank 1 projection, and thus a pure state. □

Another possibility is to use the *Schmidt-Decomposition* of a state, which can offer slightly more information about a state than the purity theorem. The Schmidt decomposition takes advantage of the conjugate linear one to one correspondence between $\mathcal{A} \otimes \mathcal{B}$ and $\operatorname{Hom}_{\mathbb{C}}\big(\mathcal{A}, \mathcal{B}\big)$ in which a state $|\psi\rangle$ on $\mathcal{A} \otimes \mathcal{B}$ is transformed into a linear operator by mapping basis elements $|i\rangle \otimes |j\rangle \mapsto |i\rangle\langle j|$, and then extending by linearity. Thus a state $|\psi\rangle = \sum_{i,j} c_{i,j} |i, j\rangle$ is mapped to the operator $E = \sum_{i,j} c_{i,j} |i\rangle\langle j|$. One can then use the singular value decomposition of E to write E uniquely as $E = \sum_{k=1}^{d} \lambda_k |x_k\rangle\langle y_k|$, where λ_k are the singular values, and $\{|x_k\rangle : 1 \leq k \leq d\}$ and $\{|y_k\rangle : 1 \leq k \leq d\}$ are orthonormal sets of norm 1 vectors. Lastly we map this expression linearly back to $\mathcal{A} \otimes \mathcal{B}$ in order to express ψ as

$$|\psi\rangle = \sum_{k=1}^{d} \lambda_k |x_k\rangle \otimes |y_k\rangle\,. \tag{1.5}$$

The singular values of E are known as the *Schmidt Coefficients* of $|\psi\rangle$. If the length of the singular value decomposition, d, is equal to 1, then $|\psi\rangle$ can be written as

$|\psi\rangle = \lambda_1 |x_1\rangle \otimes |y_1\rangle$ and thus $|\psi\rangle$ is a product state. If however d is greater than 1, then $|\psi\rangle$ is entangled. The length of this singular value decomposition is often called the *Schmidt Rank* of a state $|\psi\rangle$.

On top of telling us whether a state is entangled, the Schmidt decomposition allows us to extract even more information about the state $|\psi\rangle$. Suppose that $|\psi\rangle$ has full Schmidt rank with Schmidt coefficients $\{\lambda_1, \ldots, \lambda_d\}$. If all of the Schmidt coefficients are equal, i.e. $\lambda_i = \frac{1}{\sqrt{d}}$ for each $i \in \{1, \ldots, d\}$, then $|\psi\rangle$ is said to be maximally entangled."Most" quantum states are going to be entangled, but may not have full Schmidt rank, or if they do, the Schmidt coefficients may not all be equal, so the natural question of what these numbers mean physically arises. One approach is to first notice that the set $\{\lambda_1^2, \ldots, \lambda_k^2\}$ is a discrete probability distribution since $\langle\psi|\psi\rangle = 1$. Once we interpret the squares of these coefficients as probabilities, we can then measure the amount of "uncertainty" present in this probability distribution using classical entropies. Perhaps the most famous such entropy is the Shannon entropy [52] which is defined by

$$\mathrm{S}(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{1.6}$$

where $\{p_1, \ldots, p_n\}$ is any finite probability distribution. Using the above definition of entropy, we then define the entanglement entropy [26] of a state $|\psi\rangle$ as

$$\mathrm{S}(\psi) = -\sum_{i=1}^{d} \lambda_i^2 \log_2(\lambda_i^2), \tag{1.7}$$

where $\{\lambda_1, \ldots, \lambda_d\}$ are the Schmidt coefficients of $|\psi\rangle$. The entanglement entropy is one of the fundamental and most commonly used *entanglement measures* [26, 51, 41, 50].

An interesting property of the entanglement entropy and entanglement measures in general is that they can be computed in a more natural way, although the computational complexity is the same [34]. Suppose that $|\psi\rangle$ has Schmidt decomposition $|\psi\rangle = |\psi\rangle = \sum_{k=1}^{d} \lambda_k |x_k\rangle \otimes |y_k\rangle$. We can then apply the von Neumman entropy [52], which is defined by

$$\mathrm{N}(\rho) := -\mathrm{Tr}\left(\rho \log_2(\rho)\right) \tag{1.8}$$

for all density operators ρ, to the reduced density matrix of $|\psi\rangle\langle\psi|$ in order to compute the same quantity. To see this, let's first compute $\mathrm{Tr}_{\mathcal{A}} |\psi\rangle\langle\psi|$.

$$\begin{aligned}
\mathrm{Tr}_{\mathcal{A}} |\psi\rangle\langle\psi| &= \Big(\mathrm{Tr} \otimes I_{\mathcal{B}}\Big) \sum_{i,j=1}^{d} \lambda_i \lambda_j |x_i\rangle\langle x_j| \otimes |y_i\rangle\langle y_j| \\
&= \sum_{i,j=1}^{d} \lambda_i \lambda_j \langle x_j | x_i\rangle |y_i\rangle\langle y_j| \\
&= \sum_{i,j=1}^{d} \lambda_i \lambda_j \delta_{i,j} |y_i\rangle\langle y_j| \qquad \text{(orthonormality of } |x_i\rangle) \\
&= \sum_{i=1}^{d} \lambda_i^2 |y_i\rangle\langle y_i| .
\end{aligned} \tag{1.9}$$

Using the functional calculus with the function $-x\log_2(x)$ we can compute the von Neumann entropy of $\mathrm{Tr}_{\mathcal{A}} |\psi\rangle\langle\psi|$.

$$\begin{aligned}
\mathrm{N}\Big(\mathrm{Tr}_{\mathcal{A}} |\psi\rangle\langle\psi|\Big) &= -\mathrm{Tr}\left(\sum_{i=1}^{d} \lambda_i^2 \log_2(\lambda_i^2) |y_i\rangle\langle y_i|\right) \\
&= -\sum_{i=1}^{d} \lambda_i^2 \log_2(\lambda_i^2) \langle y_i | y_i\rangle \\
&= -\sum_{i=1}^{d} \lambda_i^2 \log_2(\lambda_i^2)
\end{aligned} \tag{1.10}$$

Note that the computation would've yielded the same result had we instead computed the von Neumann entropy of $\mathrm{Tr}_{\mathcal{B}} |\psi\rangle\langle\psi|$ instead due to the orthonormality of the singular vectors. It turns that all entanglement measures on pure states, not just the von Neumann entropy, can be computed in the same way as above [51], as we'll see in more detail later on. This functional approach to quantifying entanglement allows to “easily" determine if a pure state is entangled by simply computing the von Neumann entropy of the reduced density matrix of the state. If the entropy yields 0, then the state would have to have been a product state as the classical Shannon entropy can only achieve 0 when the probability distribution is of the form $p_i = 1$ for some $i \in \{1, \ldots, n\}$ and $p_j = 0$ else. Otherwise, the state is entangled.

The question of how to measure entanglement for density matrices then arises, but before we can measure entanglement we must first discuss what it means for a density operator to be entangled as the definition of entanglement differs slightly from the pure state definition at first glance. We say that a density operator $\rho \in \mathrm{D}(\mathcal{A} \otimes \mathcal{B})$ is *separable* [52] if and only if there exist convex coefficients $\{p_1, \ldots, p_n\}$ and pure states $\{|\psi_1\rangle_{\mathcal{A}}, \ldots, |\psi_n\rangle_{\mathcal{A}}\} \subset \mathcal{A}$ and $\{|\psi_1\rangle_{\mathcal{B}}, \ldots, |\psi_n\rangle_{\mathcal{B}}\} \subset \mathcal{B}$ such that

$$\rho = \sum_{i=1}^{n} p_i |\psi_i\rangle \langle\psi_i|_{\mathcal{A}} \otimes |\psi_i\rangle \langle\psi_i|_{\mathcal{B}}, \tag{1.11}$$

or in other words, ρ can be expressed as a convex sum of product states. If ρ is not separable, then ρ is said to be entangled. But how does one find such a decomposition? Since we used an algebraic invariant of pure states to measure entanglement, a natural guess might be to use a natural invariant of a density matrix such as the spectral decomposition. However, as we'll see in the next example, this won't work.

Example 1.2.2. *Let $\mathcal{H} = \mathbb{C}^{2\otimes 2}$ equipped with the computational basis and define $\rho \in D(\mathcal{H})$ by $\rho = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11|$. If an observer were measuring this state in the computational basis, they would observe $|00\rangle$ half the time, and $|11\rangle$ half the time. No matter which state is observed, they would observe a product state. These measurements would lead the observer to the conclusion that ρ is separable, and they would be correct.*

What if instead, the observer measured ρ used the Bell Basis *[31], which consists of the states*

$$\begin{aligned} \left|\Phi^+\right\rangle &:= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \\ \left|\Phi^-\right\rangle &:= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle), \\ \left|\Psi^+\right\rangle &:= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \\ \left|\Psi^-\right\rangle &:= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \end{aligned} \tag{1.12}$$

which are known as the Bell States*, all of which can be seen to be maximally entangled. Since ρ can also be written as $\rho = \frac{1}{2} |\Phi^+\rangle\langle\Phi^+| + \frac{1}{2} |\Phi^-\rangle\langle\Phi^-|$, the observer would observe $|\Phi^+\rangle$ half the time, and $|\Phi^-\rangle$ half of the time. These measurements would lead our observer to erroneously believe that ρ is entangled. Since both of these decompositions of ρ are valid spectral decompositions, we can see that spectral decomposition cannot tell us whether or not a state is entangled. So if a natural algebraic invariant of ρ doesn't work, what will ?*

One possibility is to use the geometry of the set of density operators to extend the domain of entanglement measures, such as the entanglement entropy, to general density operators instead of just pure states. It can be shown that the set of density operators is compact and convex. Moreover the extreme points of $\mathrm{D}(\mathcal{H})$ are precisely the set of projections onto the pure states of $\mathcal{H}$. And since we saw in the previous example that a density operator can have more than decomposition into pure states, a viable solution is then to find the decomposition of $\{(p_i)_{i\in I}, (\psi_i)_{i\in I}\}$ of ρ such that $\sum_{i\in I} p_i \mathrm{S}(\psi_i)$ is a minimum. The problem here is that it's not clear if such a minimum exists, and if it does, what conditions are necessary to ensure its existence. It's also not clear how one would even begin to approach this minimization process. In order to address these concerns we need to first study entanglement measures in a more axiomatic fashion.

Chapter 2

Entanglement Measures

2.1 Basic Definitions

Let $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ for some finite dimensional Hilbert spaces $\mathcal{A}$ and $\mathcal{B}$, then a function $\mu : \mathrm{D}(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ is an entanglement measure if and only if the following three properties hold for all $\rho \in \mathrm{D}(\mathcal{H})$:

1. $\mu(\rho) = 0$ if and only if ρ is separable, (i.e. μ is *faithful*).

2. $\mu(\Lambda(\rho)) \leq \mu(\rho)$ where Λ is an LOCC channel, (i.e. μ is an *LOCC monotone*). LOCC stands for Local Operations and Classical Communication. See Section 2.5 for a precise description of these channels.

3. $\mu((U_1 \otimes U_2)\rho(U_1^* \otimes U_2^*)) = \mu(\rho)$ where U_1 and U_2 are unitaries, (i.e. μ is *invariant under local unitaries*).

A function μ on the states of a Hilbert space is called an *LOCC monotone* [51] if property (2) is satisfied. LOCC channels are hard to represent, [7]. In Section 2.5 we provide a tree representation of LOCC channels which contributes to a better understanding of these channels. Our result, Theorem 2.5.1 which is a slight strengthening of a result of Vidal [51] with a simplified proof, shows that many common entanglement measures and the recently proposed “entanglement number" are indeed LOCC monotones.

The goal of this section is to formalize the notion of what it means to find an ensemble whose average entanglement is a minimum. First we define the set of convex decompositions of a density matrix $\rho \in \mathcal{H}$, $\mathcal{CD}(\rho)$ by

$$\mathcal{CD}(\rho) = \left\{ \{(\lambda_i)_{i=0}^{\infty}, \{\psi_i\}_{i=0}^{\infty}\} : \lambda_i \geq 0, \sum_{i=1}^{\infty} \lambda_i = 1, \psi_i \in \mathcal{H}, \text{ and } \sum_{i=0}^{\infty} \lambda_i \left|\psi_i\right\rangle \left\langle\psi_i\right| = \rho \right\}. \tag{2.1}$$

Now let μ be a defined on the pure states of $\mathcal{H}$ taking values in $\mathbb{R}_{\geq 0}$. One can then extend μ to the set of density operators $\mathrm{D}(\mathcal{H})$ of a Hilbert space $\mathcal{H}$ by the *convex roof construction*:

$$\mu(\rho) = \inf \left\{ \sum_{i=0}^{\infty} \lambda_i \mu(\psi_i) : \{(\lambda_i)_{i=0}^{\infty}, \{\psi_i\}_{i=0}^{\infty}\} \in \mathcal{CD}(\rho) \right\}. \tag{2.2}$$

As mentioned before in the previous chapter, the only problem with this construction is that it's not immediately clear that the infimum should even exist. In the next section however, we prove existence in the affirmative.

2.2 Existence of OPSE

In this section we prove that the infimum in the definition of convex roof construction which extends entanglement measures from pure states to mixed states, is always achieved for a certain class of entanglement measures as summarized below.

Theorem 2.2.1. *If μ is a norm-continuous function on the pure states of a finite dimensional Hilbert space $\mathcal{H}$, then the infimum in Equation (2.2) is attained for all $\rho \in D(\mathcal{H})$.*

Acknowledgement 2.2.2. *We would like to thank Shirokov who made us aware of his paper [47] where an extension of our Theorem 2.1 has been proved that is applicable even for infinite dimensional Hilbert spaces. We were not aware of publication [47], and the method used here in the proof of Theorem 2.1 is different that the one used in this publication. We decided to keep our proof of Theorem 2.1 for the completeness of this work.*

Proof. An important role in the proof of this theorem will be played by a compact metrizable topological space (Π, τ) that we define now. Throughout the proof, for any Hilbert space $\mathcal{H}$, we will denote by $S(\mathcal{H})$ the *unit sphere* of $\mathcal{H}$, (i.e. the vectors of $\mathcal{H}$ of norm equal to 1). As usual, the set $S(\mathcal{H})$ is identified with the set of pure states of $\mathcal{H}$. Define

$$P_\infty = \{(\lambda_i)_{i=0}^{\infty} : \lambda_i \geq 0 \text{ for each i }, \sum_{i=0}^{\infty} \lambda_i = 1\} \tag{2.3}$$

i.e., P_∞ is the intersection of $S(\ell^1)$ with positive cone of ℓ^1. Next we define

$$\Pi = P_\infty \times S(\mathcal{H})^{\mathbb{N}} \tag{2.4}$$

which intuitively can be thought of as the collection of all possible pairings of "convex coefficients" with sequences of pure states. Equip Π with the product topology τ of the weak* topology on ℓ_1 with the product topology of (any) norm topologies of the unit sphere $S(\mathcal{H})$ of the Hilbert space $\mathcal{H}$.

Notice that P_∞ is compact since it is a closed subspace of a weak* compact space. And as long as $\dim(\mathcal{H}) < \infty$, $S(\mathcal{H})$ is norm compact. This implies that space of all sequences of pure states, $S(\mathcal{H})^{\mathbb{N}}$ is norm compact by Tychonoff's theorem. Thus Π is the product of two compact spaces and must therefore be compact. Note that we had to restrict ourselves to finite dimensional Hilbert spaces since in general the set of pure states is not weak* compact, (see for example [17, Theorem 2.8]).

Also notice that (Π, τ) is metrizable. Indeed the weak* topology on bounded subsets of ℓ_1 is metrizable since the predual of ℓ_1, (which is usually denoted by c_0), is separable. Also $S(\mathcal{H})^{\mathbb{N}}$ is the countable product of metric spaces, hence its product topology is metrizable. Thus (Π, τ) must be metrizable.

We split the rest of the proof into two Claims in order to make it more readable.

Claim 1: If $\mathcal{H}$ is a finite dimensional Hilbert space, then for every $\rho \in \mathrm{D}(\mathcal{H})$, the set $\mathcal{CD}(\rho)$ is τ-compact.

In order to prove Claim 1, assume that $\mathcal{H}$ is a finite dimensional Hilbert space and define $f : \Pi \to \mathrm{D}(\mathcal{H})$ by

$$((\lambda_i)_{i=0}^{\infty}, \{\psi_i\}_{i=0}^{\infty}) \mapsto \sum_{i=0}^{\infty} \lambda_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \tag{2.5}$$

Since the λ_i's are summable and the target space of f, (i.e. $\mathcal{D}(\mathcal{H})$), is complete, this series will always converge. Moreover notice that $f^{-1}(\{\rho\})$ is the set of all possible convex decompositions of ρ into pure states, i.e. $f^{-1}(\{\rho\}) = \mathcal{CD}(\rho)$. Since $f^{-1}(\{\rho\})$ is a subset of the compact space Π, in order to show that $f^{-1}(\{\rho\})$ is compact, it is enough to show that f is continuous on Π. Since both spaces, Π and $\mathcal{D}(\mathcal{H})$ are metrizable, we will use sequences to check the continuity of f, i.e. we will prove that $\pi_n \to \pi$ in Π implies that $f(\pi_n) \to f(\pi)$ in $\mathrm{D}(\mathcal{H})$.

So let $\pi_n = ((\lambda_i^n)_{i=0}^\infty, \{\psi_i^n\}_{i=0}^\infty)$ be a sequence in Π converging to $\pi = ((\lambda_i)_{i=0}^\infty, \{\psi_i\}_{i=0}^\infty)$ as n goes to infinity, and let $\epsilon > 0$. Then there exists some $I \in \mathbb{N}$ such that $\sum_{i=I+1}^{\infty} \lambda_i < \epsilon$, and so $\sum_{i=0}^{I} \lambda_i \geq 1-\epsilon$. Moreover, since $(\lambda_i^n)_{i=0}^\infty \to (\lambda_i)_{i=0}^\infty$ in ℓ^1, there exists some $N_0 \in \mathbb{N}$ such that $||(\lambda_i^n)_{i=0}^I - (\lambda_i)_{i=0}^I||_{\ell^1} < \epsilon$ for all $n \geq N_0$. And so $\sum_{i=0}^{I} \lambda_i^n \geq 1 - 2\epsilon$ which yields that $\sum_{i=I+1}^{\infty} \lambda_i^n < 2\epsilon$. Thus we have the following inequalities:

$$
\begin{aligned}
||f(\pi_n) - f(\pi)||_1 &\leq \left\| \sum_{i=0}^{I} \lambda_i^n \left|\psi_i^n\right\rangle \left\langle\psi_i^n\right| - \sum_{i=0}^{I} \lambda_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \right\|_1 + \left\| \sum_{i=I+1}^{\infty} \lambda_i^n \left|\psi_i^n\right\rangle \left\langle\psi_i^n\right| - \sum_{i=I+1}^{\infty} \lambda_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \right\|_1 \\
&\leq \left\| \sum_{i=0}^{I} \lambda_i^n \left|\psi_i^n\right\rangle \left\langle\psi_i^n\right| - \sum_{i=0}^{I} \lambda_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \right\|_1 \\
&\quad + \sum_{i=I+1}^{\infty} \lambda_i^n \left\| \left|\psi_i^n\right\rangle \left\langle\psi_i^n\right| \right\|_1 + \sum_{i=I+1}^{\infty} \lambda_i \left\| \left|\psi_i\right\rangle \left\langle\psi_i\right| \right\|_1 \\
&= \left\| \sum_{i=0}^{I} \lambda_i^n \left|\psi_i^n\right\rangle \left\langle\psi_i^n\right| - \sum_{i=0}^{I} \lambda_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \right\|_1 + \sum_{i=I+1}^{\infty} \lambda_i^n + \sum_{i=I+1}^{\infty} \lambda_i \qquad (2.6) \\
&< \left\| \sum_{i=0}^{I} \lambda_i^n \left|\psi_i^n\right\rangle \left\langle\psi_i^n\right| - \sum_{i=0}^{I} \lambda_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \right\|_1 + 3\epsilon \\
&= \left\| \sum_{i=0}^{I} (\lambda_i^n - \lambda_i + \lambda_i) \left|\psi_i^n\right\rangle \left\langle\psi_i^n\right| - \sum_{i=0}^{I} \lambda_i \left|\psi_i\right\rangle \left\langle\psi_i\right| \right\|_1 + 3\epsilon \\
&\leq \sum_{i=0}^{I} |\lambda_i^n - \lambda_i| \cdot \left\| \left|\psi_i^n\right\rangle \left\langle\psi_i^n\right| \right\|_1 + \sum_{i=0}^{I} \lambda_i \left\| \left|\psi_i^n\right\rangle \left\langle\psi_i^n\right| - \left|\psi_i\right\rangle \left\langle\psi_i\right| \right\|_1 + 3\epsilon \\
&\leq \sum_{i=0}^{I} \left(|\lambda_i^n - \lambda_i| + \left\| \left|\psi_i^n\right\rangle \left\langle\psi_i^n\right| - \left|\psi_i\right\rangle \left\langle\psi_i\right| \right\|_1 \right) + 3\epsilon
\end{aligned}
$$

Now since weak-* convergence implies coordinate-wise convergence in ℓ^1, we have that for all n large enough, $|\lambda_i^n - \lambda_i| < \frac{\epsilon}{I+1}$ for each $i \in \{0, 1, \ldots, I\}$. We can also make sure that for all such n and i, $\| |\psi_i^n\rangle \langle\psi_i^n| - |\psi_i\rangle \langle\psi_i| \|_1 < \frac{\epsilon}{I+1}$, since ψ_i^n converges to ψ_i. Thus for large enough n, we get that $\|f(\pi_n) - f(\pi)\|_1 < 5\epsilon$, which implies that f is continuous. This finishes the proof of Claim 1.

Next we define $g : \Pi \to \mathbb{R}_{\geq 0}$ by

$$((\lambda_i)_{i=0}^{\infty}, \{\psi_i\}_{i=0}^{\infty}) \mapsto \sum_{i=0}^{\infty} \lambda_i \mu(\psi_i) \tag{2.7}$$

Thus the statement that the infimum in Equation (2.2) is always achieved, is equivalent to the fact that g always attains its infimum on the set $\mathcal{CD}(\rho)$, for any density matrix ρ. Since by Claim 1, $\mathcal{CD}(\rho)$ is always τ-compact, the proof of Theorem 2.2.1 will be complete once we prove the following:

Claim 2: If μ is a norm-continuous function on the set of pure states of a finite dimensional Hilbert space, then g is continuous on (Π, τ).

The argument for the continuity of g will be almost exactly the same as that for f, that is given in the Claim 1. Since μ is norm-continuous on the unit sphere of a finite dimensional Hilbert space which is norm-compact, we have that μ is bounded, there exists some finite number M such that $\mu(\psi) \leq M$ for all pure states ψ. Again, let $\pi_n = ((\lambda_i^n)_{i=0}^{\infty}, \{\psi_i^n\}_{i=0}^{\infty})$ be a sequence in Π converging to $\pi = ((\lambda_i)_{i=0}^{\infty}, \{\psi_i\}_{i=0}^{\infty})$ as n goes to infinity, and let $\epsilon > 0$. Then using the same choice of I as in the proof of the continuity of f, we have the following inequalities:

$$
\begin{aligned}
|g(\pi) - g(\pi_n)| &\le \left|\sum_{i=0}^{I} \lambda_i \mu(\psi_i) - \sum_{i=0}^{I} \lambda_i^n \mu(\psi_i^n)\right| + \left|\sum_{i=I+1}^{\infty} \lambda_i \mu(\psi_i)\right| + \left|\sum_{i=I+1}^{\infty} \lambda_i^n \mu(\psi_i^n)\right| \\
&\le \left|\sum_{i=0}^{I} \lambda_i \mu(\psi_i) - \sum_{i=0}^{I} \lambda_i^n \mu(\psi_i^n)\right| + \sum_{i=I+1}^{\infty} \lambda_i M + \sum_{i=I+1}^{\infty} \lambda_i^n M \\
&\le \sum_{i=0}^{I} |(\lambda_i^n - \lambda_i + \lambda_i)\mu(\psi_i^n) - \lambda_i \mu(\psi_i)| + 3\epsilon M \\
&\le \sum_{i=0}^{I} |\lambda_i^n - \lambda_i| \mu(\psi_i^n) + \sum_{i=0}^{I} \lambda_i |\mu(\psi_i^n) - \mu(\psi_i)| + 3\epsilon M \\
&\le \sum_{i=0}^{I} |\lambda_i^n - \lambda_i| M + \sum_{i=0}^{I} |\mu(\psi_i^n) - \mu(\psi_i)| + 3\epsilon M
\end{aligned}
\tag{2.8}
$$

Now taking n large enough so that $|\lambda_i^n - \lambda_i| < \epsilon$ and $|\mu(\psi_i^n) - \mu(\psi_i)| < \epsilon M$ for each $i \in \{0, 1, \ldots, I\}$, we get that $|g(\pi) - g(\pi_n)| < 5\epsilon M$. Thus g is also continuous of Π This finishes the proof of Claim 2 and thus the proof of Theorem 2.2.1. $\square$

Terminology: The decomposition of a density operator ρ with respect to a convex roof extension of an entanglement measure μ which is initially defined on the pure states of a multipartite Hilbert space is usually called an *Optimal Pure State Ensemble*, which is often abbreviated as OPSE [51, 49].

Using the above theorem, we are also able to prove that an entanglement measure that is faithful on pure states is then faithful for general density operators.

Corollary 2.2.3. *Let $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ for some finite dimensional Hilbert spaces $\mathcal{A}$ and $\mathcal{B}$ and let μ be a norm-continuous function on the pure states of $\mathcal{H}$ with values in $\mathbb{R}_{\geq 0}$ which is faithful, in the sense that it only vanishes on the factorable pure states. Extend μ on the set of all states of $\mathcal{H}$ via the convex roof construction. Then μ is faithful, i.e. for any density matrix ρ, we have $\mu(\rho) = 0$ if and only if ρ is separable.*

Proof. $\Longrightarrow$) Suppose that $\rho \in \mathrm{D}(\mathcal{H})$ with $\mu(\rho) = 0$. Then by Theorem 2.2.1, there exist some $\{(\lambda)_i, (\psi_i)\} \in \mathcal{CD}(\rho)$ such that

$$\sum_i \lambda_i \mu(\psi_i) = \mu(\rho). \tag{2.9}$$

Since μ only takes non-negative values, we obtain that $\mu(\psi_i) = 0$ for each i. Since μ is assumed to be faithful on the pure states, this implies that each ψ_i is factorable, and so ρ is the convex combination of some factorable pure states and is therefore separable.

$\Longleftarrow$) Conversely, suppose that ρ is separable. Then ρ is the convex combination of some factorable pure states in $\mathcal{H}$. Let $\rho = \sum_i \lambda_i \psi_i$ where λ_i's are convex coefficients and ψ_i's are factorable pure states of $\mathcal{H}$. Since μ is faithful on the pure states of $\mathcal{H}$, we have that $\mu(\psi_i) = 0$ for every i. Then

$$\sum_i \lambda_i \mu(\psi_i) = 0 \tag{2.10}$$

which, by the definition of the convex roof extension, implies that $\mu(\rho) = 0$ since each μ takes on nonnegative values. for each i. □

2.3 Applications of Theorem (2.2.1)

Now that we've shown the existence of OPSE for general norm-continuous functions on pure states of finite dimensional Hilbert spaces we can present some common applications. In literature the existence of OPSE has many times been taken for granted, but other times it has been questioned [19], and other times claimed in a particular special case [49].

In [19], Gudder in [19] introduced a general theory of entanglement which applies even to classical probability measures with discrete support. In the same article, he also introduced a quantity that he called the *entanglement number* which quantifies entanglement for classical probability measures as well as for density operators. Indeed the formula $e(|\psi\rangle) = \sqrt{1 - \sum_i \lambda_i^2}$ (where $(\sqrt{\lambda_i})_i$ are the Schmidt coefficients of the pure state ψ, with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$) defines the entanglement number $e(|\psi\rangle)$ of a pure bipartite state $|\psi\rangle$ and can also be naturally extended to classical probability measures. The advantage of the entanglement number besides the fact that a similar formula makes sense even in classical probability theory, is that it can be computed easily for pure bipartite states. Indeed, in [20, Theorem 4.2] a closed form formula is given for computing the entanglement number of a pure bipartite state $|\psi\rangle$. Indeed, if $\mathcal{X}$, $\mathcal{Y}$ are two finite dimensional Hilbert spaces with orthonormal bases $(|x_i\rangle)$ and $(|y_j\rangle)$ respectively, and $|\psi\rangle \in \mathcal{X} \otimes \mathcal{Y}$ is a pure state written as $|\psi\rangle = \sum_{i,j} c_{i,j} |x_i\rangle |y_j\rangle$, then the entanglement number $e(|\psi\rangle)$ can be computed by

$$e(|\psi\rangle) = \sqrt{1 - \operatorname{Tr}(|C|^4)} \tag{2.11}$$

where C is the matrix $(c_{i,j})$. Moreover, it is shown that if (c_j)'s are the columns of C then

$$\operatorname{Tr}(|C|^4) = \sum_{r,s} |\langle c_r, c_s\rangle|^2. \tag{2.12}$$

The extension of the entanglement number from pure states to mixed states is done via the *convex roof* construction:

$$e(\rho) = \inf \left\{ \sum_{i=0}^{\infty} \lambda_i e(\psi_i) : \lambda_i \geq 0, \sum_i \lambda_i = 1, \psi_i \text{ are pure states, and } \sum_{i=0}^{\infty} \lambda_i |\psi_i\rangle \langle \psi_i| = \rho \right\} \tag{2.13}$$

Gudder states in [19] the open question of whether the infimum in the above convex roof construction is always attained. Indeed some of his results, (such as [19, Theorem 3.3]), depend on that assumption. An application of our Theorem (2.2.1) answers his open question in the following corollary.

Corollary 2.3.1. *Let $\mathcal{H} = \mathcal{X} \otimes \mathcal{Y}$ be a finite dimensional bipartite Hilbert space and consider the entanglement number defined on pure states of $\mathcal{H}$ via Equations (2.11) and (2.12), and extended to mixed states via Equation (2.13). Then every mixed state $\rho \in D(\mathcal{H})$ admits an OPSE.*

Proof. It is obvious that Equations (2.11) and (2.12) define a norm-continuous function e on the pure states of $\mathcal{H}$. Hence the result follows from Theorem 2.2.1. □

Another common entanglement measure to which we can apply our result is the Entanglement entropy. Recall that for Hilbert spaces $\mathcal{A}$ and $\mathcal{B}$ the *entanglement entropy* S of a pure state $|\psi\rangle \in \mathcal{A} \otimes \mathcal{B}$ is defined by

$$\mathrm{S}(\psi) = -\mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{A}}(|\psi\rangle\langle\psi|)\log(\mathrm{Tr}_{\mathcal{A}}|\psi\rangle\langle\psi|)\right) = -\mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{B}}(|\psi\rangle\langle\psi|)\log(\mathrm{Tr}_{\mathcal{B}}|\psi\rangle\langle\psi|)\right) \tag{2.14}$$

The entanglement entropy is then extended by the convex roof construction as in Equation (2.2) by defining

$$\mathrm{S}(\rho) = \inf\left\{\sum_{i=0}^{\infty}\lambda_i\mathrm{S}(\psi_i) : \{(\lambda_i)_{i=0}^{\infty}, \{\psi_i\}_{i=0}^{\infty}\} \in \mathcal{CD}(\rho)\right\} \tag{2.15}$$

for all density operators $\rho \in \mathrm{D}\,(\mathcal{A}\otimes\mathcal{B})$. The convex roof extension of the entanglement entropy is also often called the *entanglement of formation*. It's been shown that the von Neumann entropy, $\mathrm{H}(\rho) = -\rho\log\rho$, is a continuous mapping on density operators [12]. Moreover it is well known that the partial trace is a bounded linear operator from the space of trace class operators on a bipartite Hilbert space to the space of trace class operators of one of its parts, and therefore continuous. Thus the entanglement entropy is continuous on pure states since it is the composition of two continuous maps, and therefore, by Theorem 2.2.1, it must always exhibit OPSE.

Next we apply our result to the convex roof extended *negativity* [32, 15] of a state. Recall that the negativity [15] of a state $\rho \in \mathrm{D}\,(\mathcal{A}\otimes\mathcal{B})$ is defined by

$$\mathcal{N}(\rho) = \frac{1}{d-1}(\|\rho^{T_{\mathcal{B}}}\| - 1) \tag{2.16}$$

where $d = \min\{\dim\mathcal{A}, \dim\mathcal{B}\}$, $\rho^{T_{\mathcal{B}}}$ is the partial transpose of ρ on the space $\mathrm{D}\,(\mathcal{B})$, and the norm is the Hilbert-Schmidt norm on operators on $\mathcal{A}\otimes\mathcal{B}$. While $\mathcal{N}$ is indeed defined for all density operators, it cannot distinguish bound entangled states and separable states. One solution [32] to this problem is to first define the negativity for pure states $|\psi\rangle \in \mathcal{A} \otimes \mathcal{B}$ by

$$\mathcal{N}(\psi) = \frac{1}{d-1}(\| \, |\psi\rangle \langle\psi|^{T_{\mathcal{B}}} \| - 1) \tag{2.17}$$

then extend the domain of $\mathcal{N}$ by the convex roof construction. This definition of $\mathcal{N}$ can then distinguish bound entangled and separable states and is an entanglement monotone [32]. Now since the norm of a Hilbert space is obviously a norm-continuous function, and partial transpose is also continuous, it follows that $\mathcal{N}$ is continuous on pure states. Therefore, by Theorem 2.2.1, the convex roof extended negativity of a state must admit OPSE.

We can also apply our result to the family of concurrence monotones C_k [18] of a bipartite pure state $\psi \in \mathcal{A} \otimes \mathcal{B}$ which are defined by

$$C_k(\psi) = \Big(\frac{\mathrm{S}_k(\lambda_1, \ldots, \lambda_d)}{S_k(\frac{1}{d}, \ldots, \frac{1}{d})}\Big)^{\frac{1}{k}} \tag{2.18}$$

where S_k is the k-th elementary symmetric polynomial [11], the λ_j are the Schmidt coefficients of ψ , $d = \min\{\dim \mathcal{A}, \dim \mathcal{B}\}$, and k ranges from 1 to d. The C_k are then extended to the set of all density operators on $\mathcal{A}\otimes\mathcal{B}$ by the convex roof construction as well. Recall that the k-th elementary symmetric polynomial of d variables is defined by $\mathrm{S}_k(x_1, \ldots, x_d) = \sum_{1\leq i_1<i_2<\ldots<i_k\leq d} x_{i_1}x_{i_2}\ldots x_{i_k}$. Since the C_k are just the $1/k$-th powers of the normalized elementary symmetric polynomials in the Schmidt coefficients of a state, they are continuous on $\mathcal{A} \otimes \mathcal{B}$ for each k. Thus, by Theorem 2.2.1, the generalized concurrence monotones also admit OPSE.

A particularly interesting example is the *geometric measure of entanglement* [2]. Let $\mathcal{H} = \mathcal{X}_1 \otimes \cdots \otimes \mathcal{H}_n$ be the tensor product of some finite dimensional Hilbert spaces H_i. Then for pure states $|\psi\rangle \in \mathcal{H}$, the *geometric measure of entanglement* is a family of entanglement measures defined by

$$\mathrm{E}_{k_1,\ldots,k_n}(\psi) = \sup_{P_1,\ldots,P_n} ||P_1 \otimes \cdots \otimes P_n \, |\psi\rangle \, ||^2 \tag{2.19}$$

where each P_j is a rank k_j orthogonal projector on the space H_j for each index j. The supremum in this definition is achieved if $\mathcal{H}$ is a finite dimensional Hilbert space, since the set of projections of a finite dimensional Hilbert space is compact with the norm topology. This measure is an example of an *increasing LOCC monotone*; i.e., E cannot *decrease* under LOCC channels. While at first glance such monotones seem to do the opposite of what we discuss in Section 2.5 there is a one to one correspondence between the set increasing LOCC monotones and the set of (decreasing) LOCC monotones (the type discussed in Section 2.5). To see this relationship, consider the map $F(\Lambda) = \sup_{\psi} \Lambda(\psi) - \Lambda$ for all increasing LOCC channels Λ. We leave it to the reader to verify that this map is indeed a bijection between the two types of non-negative and bounded monotones. Another difference between the two types of monotones is that increasing monotones are extended to general density operators by a concave roof construction:

$$\mu(\rho) = \sup \left\{ \sum_{i=1}^{\infty} \lambda_i \mu(\psi_i) : \{(\lambda_i)_{i=1}^{\infty}, (\psi_i)_{i=1}^{\infty}\} \in \mathcal{CD}(\rho) \right\}. \tag{2.20}$$

But because of the correspondence between the two types of monotones, we've also shown the existence of OPSE for concave roof constructions as well. To apply our result to the geometric measure of entanglement, we must first show that it is continuous on the set of pure states of a general multipartite space $\mathcal{H}$.

Proposition 2.3.2. *Let $\mathcal{H} = H_1 \otimes \cdots \otimes \mathcal{H}_n$ be a tensor product of finite dimensional Hilbert space, then $E_{k_1 \ldots k_n}$ is continuous for all choices of $k_j \leq dim(\mathcal{H}_j)$ for each j.*

Proof. Let $(\psi_i)_{i=1}^{\infty}$ be a sequence of pure states in $\mathcal{H}$ converging to ψ for some pure state ψ. Then

$$\begin{aligned} |\mathrm{E}_{k_1 \ldots k_n}(\psi_i) - \mathrm{E}_{k_1 \ldots k_n}(\psi)| &= |\sup_{P_1 \ldots P_n} ||P_1 \otimes \cdots \otimes P_n |\psi_i\rangle ||^2 - ||Q_1 \otimes \cdots \otimes Q_n |\psi\rangle ||^2| \\ &\leq \sup_{P_1 \ldots P_n} | \, ||P_1 \otimes \cdots \otimes P_n |\psi_i\rangle ||^2 - ||P_1 \otimes \cdots \otimes P_n |\psi\rangle ||^2| \end{aligned} \tag{2.21}$$

Now using the reverse triangle inequality and the fact that orthogonal projectors have operator norm 1, the right hand side of the last inequality is less than or equal to

$$
\begin{aligned}
&\sup_{P_1 \dots P_n} |\; ||P_1 \otimes \dots \otimes P_n \left|\psi_i\right\rangle|| - ||P_1 \otimes \dots \otimes P_n \left|\psi\right\rangle||\;| \cdot |\; ||P_1 \otimes \dots \otimes P_n \left|\psi_i\right\rangle|| + ||P_1 \otimes \dots \otimes P_n \left|\psi\right\rangle||\;| \\
&\qquad \leq 2 \sup_{P_1 \dots P_n} |\; ||P_1 \otimes \dots \otimes P_n \left|\psi_i\right\rangle|| - ||P_1 \otimes \dots \otimes P_n \left|\psi\right\rangle||\;| \\
&\qquad \leq 2 \sup_{P_1 \dots P_n} ||P_1 \otimes \dots \otimes P_n (\left|\psi_i\right\rangle - \left|\psi\right\rangle)|| \\
&\qquad \leq 2 \sup_{P_1 \dots P_n} ||P_1 \otimes \dots \otimes P_n|| \cdot ||\psi_i - \psi|| \\
&\qquad \leq 2||\psi_i - \psi||
\end{aligned} \tag{2.22}
$$

Now letting i tend to infinity, it follows that $|\mathrm{E}_{k_1 \dots k_n}(\psi_i) - \mathrm{E}_{k_1 \dots k_n}(\psi)| \to 0$, implying that $\mathrm{E}_{k_1 \dots k_n}$ is continuous for all possible combinations of ranks $k_1 \dots k_n$ of local projections on $\mathcal{H}$. □

Thus Theorem 2.2.1 applies to the geometric measure of entanglement, and therefore this entanglement measure also exhibits OPSE.

While convex roof constructions often appear in the discussion of entanglement measures, they also arise in other contexts in quantum information theory. For instance, the entropy of a channel Φ with respect to a state ρ [39] is defined by

$$
\mathrm{H}_\rho(\Phi) = \mathrm{H}\Big(\Phi(\rho)\Big) - \inf\left\{\sum_{i=1}^{\infty} \lambda_i \mathrm{H}\Big(\Phi(\psi_i)\Big) : \{(\lambda_i)_{i=1}^{\infty}, (\psi_i)_{i=1}^{\infty}\} \in \mathcal{CD}(\rho)\right\} \tag{2.23}
$$

where $\mathrm{H}(\sigma)$ is the von Neumann entropy of a state σ. The existence of OPSE for this case was proven by Uhlmann [49] but our Theorem 2.2.1 is applicable to this problem as well. Since the von Neumann entropy is continuous [12] and since all quantum channels are continuous [52], $\mathrm{H}\Big(\Phi(\psi)\Big)$ is a continuous function on pure states ψ of a Hilbert space. Thus $\mathrm{H}_\rho(\Phi)$ exhibits OPSE for all states ρ and fixed channels Φ, verifying Uhlmann's result.

2.4 F,D-Extensions of Entanglement Measures

Despite knowing that OPSE exist for a large class of functions on quantum states, it turns out that approximating these OPSE will be difficult. While we derive an algorithm in the next chapter to approximate OPSE, we won't be able to analyze the convergence of the algorithm because of the highly non-polynomial nature of many entanglement measures. We do however manage to analyze the convergence for a certain transformation of the Tsallis entanglement entropy. In this section we develop the theory of these transformations, which we call $f, d-$extensions, so that we can refer to their properties later on. This extension method is defined using a method similar to the convex roof extension. It depends on a fixed function $f : [0, 1] \to [0, \infty)$ which vanishes only at zero, and it yields a family of extensions indexed by the integers $d \in \mathbb{N}$. The domains of these extensions increase with d, and they become equal to the set of density operators on the Hilbert space when d is large enough, (more precisely, when d strictly exceeds the dimension of the manifold of the density operators on the given Hilbert space). These extensions decrease with respect to $d \in \mathbb{N}$ at any fixed density operator, and their infimum for all d is equal to zero. More precisely, we have the following definition.

Definition 2.4.1. *Let $\mathcal{H}$ be a bipartite Hilbert space, μ be an entanglement measure on the set of pure states of $\mathcal{H}$, and $f : [0, 1] \to [0, \infty)$ be a function that vanishes only at 0. For every $d \in \mathbb{N}$ define the set $D(\mathcal{H})_d$ of density operators of $\mathcal{H}$ that can be written as a convex combination of at most d many pure states, i.e.*

$$D(\mathcal{H})_d = \Big\{ \rho \in D(\mathcal{H}) : \text{ there exists a family of pure states } (|\psi_i\rangle \langle\psi_i|)_{i=1}^d \text{ and } (p_i)_{i=1}^d \subseteq [0, 1] \text{ with } \sum_{i=1}^d p_i = 1 \text{ and } \rho = \sum_{i=1}^d p_i |\psi_i\rangle \langle\psi_i| \Big\}.$$

Note that

(i) $D(\mathcal{H})_1$ is equal to the set of pure states of $\mathcal{H}$.

(ii) $D(\mathcal{H})_d \subseteq D(\mathcal{H})_{d+1}$ for every $d \in \mathbb{N}$.

(iii) There exists $d \in \mathbb{N}$ such that $D(\mathcal{H})_d = D(\mathcal{H})$, (indeed, by Caratheodory's theorem in convex analysis, this happens when $d \geq dim(D(\mathcal{H})) + 1$).

Define a function $\mu_{f,d} : D(\mathcal{H})_d \to [0,\infty)$ *by*

$$\mu_{f,d}(\rho) = \inf\left\{\sum_{i=1}^{d} f(p_i)\mu(|\psi_i\rangle\langle\psi_i|) : \rho = \sum_{i=1}^{d} p_i |\psi_i\rangle\langle\psi_i| \in D(\mathcal{H})_d\right\}.$$

We call the function $\mu_{f,d}$ *the "*f*-*d *extension of* μ*". We call the sequence* $(\mu_{f,d})_{d\in\mathbb{N}}$ *the "sequence of* f*-extensions of* μ*".*

Remark 2.4.2. *Let* $\mathcal{H}$ *be a finite dimensional bipartite Hilbert space,* μ *be an entanglement measure on the pure states of* $\mathcal{H}$*, and* $f : [0,1] \to [0,\infty)$ *be a continuous function which vanishes only at* 0*. Then for every* $d \in \mathbb{N}$ *the infimum in the definition of* $\mu_{f,d}$ *is attained, (i.e. it is a minimum).*

Indeed, since $\mathcal{H}$ is finite dimensional, the set of pure states on $\mathcal{H}$ is a compact set. Let $\mathcal{P}$ denote the set of pure states on $\mathcal{H}$. Also, the set

$$\mathcal{C} = \{(p_i)_{i=1}^{d} : p_i \geq 0 \text{ for all } i = 1,\dots,d \text{ and } \sum_{i=1}^{d} p_i = 1\}$$

is compact as well. Now the function defined on $\mathcal{P}^d \times \mathcal{C}$ by

$$\mathcal{P}^d \times \mathcal{C} \ni (|\psi_i\rangle\langle\psi_i|)_{i=1}^{d} \times (p_i)_{i=1}^{d} \mapsto \sum_{i=1}^{d} f(p_i)\mu(|\psi_i\rangle\langle\psi_i|),$$

is a continuous function, since both f and μ are assumed to be continuous. Hence it achieves its minimum on its compact domain, which finishes the proof of Remark 2.4.2.

Remark 2.4.2 justifies the next terminology.

Terminology 2.4.3. *Let* $\mathcal{H}$ *be a finite dimensional bipartite Hilbert space,* μ *be an entanglement measure on the set of pure states of* $\mathcal{H}$*, and* $f : [0,1] \to [0,\infty)$ *be a continuous function which vanishes only at* 0*. Let* $d \in \mathbb{N}$*, and* $\rho \in D(\mathcal{H})_d$*. The tuple* $((p_i)_{i=1}^{d}, (|\psi_i\rangle\langle\psi_i|)_{i=1}^{d})$ *with* $p_i \geq 0$ *for all* $i = 1,\dots,d$*,* $\sum_{i=1}^{d} p_i = 1$*,* $|\psi_i\rangle\langle\psi_i|$ *being pure states on* $\mathcal{H}$*, and* $\mu_{f,d}(\rho) = \sum_{i=1}^{d} f(p_i)\mu(|\psi_i\rangle\langle\psi_i|)$*, is called "Optimal Pure State Ensemble (OPSE) for* μ*,* f *and* d*".*

Recall that the Tsallis entanglement entropy , T_2, is defined by

$$T_2(|\psi\rangle\langle\psi|) = 1 - \mathrm{Tr}\left((\mathrm{Tr}_{\mathcal{A}}(|\psi\rangle\langle\psi|))^2\right) \tag{2.24}$$

for all pure states $|\psi\rangle \in \mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ and is then extended to general mixed states ρ of $\mathcal{H}$ via the convex roof construction given in Equation (2.2), namely

$$T_2(\rho) = \inf \left\{ \sum_{i=0}^{\infty} p_i T_2(|\psi_i\rangle\langle\psi_i|) : \{(p_i)_{i=0}^{\infty}, \{\psi_i\}_{i=0}^{\infty}\} \in \mathcal{CD}(\rho) \right\}. \tag{2.25}$$

In Chapter 3 we study the sequence of f-extensions of Tsalis entropy T_2, where $f : [0,1] \to [0,\infty)$ is defined by $f(x) = x^2$. We will denote by $(T_{f,d})_{d\in\mathbb{N}}$ the sequence of f-extensions of Tsalis entropy T_2 defined on the set of pure states of a bipartite Hilbert space $\mathcal{H} = \mathcal{H} \otimes \mathcal{B}$. More precisely, by combining Equation (2.24) and Definition 2.4.1 we obtain

$$T_{f,d}(\rho) = \inf \left\{ \sum_{i=1}^{d} p_i^2 (1 - \mathrm{Tr}\left((\mathrm{Tr}_{\mathcal{A}}(|\psi_i\rangle \langle\psi_i|))^2\right)) : \rho = \sum_{i=1}^{d} p_i \,|\psi_i\rangle \langle\psi_i| \in \mathrm{D}\left(\mathcal{H}\right)_d \right\} \tag{2.26}$$

Remark 2.4.4. *The reason that we bound the length of the convex combinations in Definition 2.4.1 by a finite number d, is because otherwise the infimum in Equation* (2.26) *would be equal to zero for every ρ.*

Indeed, in order to verify the statement of Remark 2.4.4, notice that if $\rho = \sum_{i=1}^{d} p_i \,|\psi_i\rangle \langle\psi_i| \in \mathrm{D}\left(\mathcal{H}\right)_d$ for some $(p_i)_{i=1}^{d} \subset [0,1]$ with $\sum_{i=1}^{d} p_i = 1$ and a sequence of pure states $(|\psi_i\rangle \langle\psi_i|)_{i=1}^{d}$, then for every $n \in \mathbb{N}$ we can write $\rho = \sum_{i=1}^{d}\sum_{j=1}^{n} \frac{p_i}{n} \,|\psi_i\rangle \langle\psi_i|$ and $\sum_{i=1}^{d}\sum_{j=1}^{n} (\frac{p_i}{n})^2 \mu(|\psi_i\rangle \langle\psi_i|) = n \sum_{i=1}^{d} \frac{p_i^2}{n^2} \mu(|\psi_i\rangle \langle\psi_i|) = \frac{1}{n}\sum_{i=1}^{d} p_i^2 \mu(|\psi_i\rangle \langle\psi_i|)$, which tends to zero as n tends to infinity.

Given an entanglement measure μ on the set of pure states of a finite dimensional bipartite Hilbert space, and a continuous function $f : [0,1] \to [0,\infty)$ which vanishes only at 0, the sequence of the f-extensions of μ can be used in order to detect entanglement in the following sense:

Definition 2.4.5. *Let $\mathcal{H}$ be a bipartite Hilbert space, and for every $d \in \mathbb{N}$ let $D(\mathcal{H})_d$ be a subset of the set $D(\mathcal{H})$ of density matrices of $\mathcal{H}$, satisfying properties (i), (ii) and (iii) of Definition 2.4.1. For every $d \in \mathbb{N}$ let a function $\mu_d : D(\mathcal{H})_d \to [0, \infty)$. We say that the family $(\mu_d)_{d\in\mathbb{N}}$ detects entanglement, if for every density operator ρ on $\mathcal{H}$, ρ is separable if and only if there exists $d \in \mathbb{N}$ such that rho $\in D(\mathcal{H})_d$ and $\mu_d(\rho) = 0$.*

Proposition 2.4.6. *Let μ be any entanglement measure on the pure states of a finite dimensional bipartite Hilbert space $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$. Consider the f-d extensions $(\mu_{f,d})_{d\in\mathbb{N}}$ of μ to the density operators of $\mathcal{H}$, where $f : [0, 1] \to [0, \infty)$ is a function that vanishes only at zero. Let ρ be a mixed state on $\mathcal{H}$. Then ρ is separable if and only if there exists $d \in \mathbb{N}$ such that $\mu_{f,d}(\rho) = 0$.*

Proof. $\Longleftarrow \big)$ Let $\mu_{f,d}(\rho) = 0$ where μ is an entanglement measure, f is as in the above proposition, ad d is some positive integer. Then there exists some pure state ensemble $\left\{(p_i)_{i=1}^d, (|\psi_i\rangle)_{i=1}^d\right\}$ with $\sum_{i=1}^d p_i |\psi_i\rangle\langle\psi_i| = \rho$ and

$$\mu_{f,d}(\rho) = \sum_{i=1}^{d} f(p_i)\mu(\psi_i) = 0. \tag{2.27}$$

If we assume without loss of generality that each $p_i > 0$, then it must follow that $\mu(\psi_i) = 0$ for each $i \in \{1, \ldots, d\}$. Since μ is an entanglement measure, it must be true that ψ_i is factorable for each $i \in \{1, \ldots, d\}$. Thus ρ can be written as a convex sum of factorable pure states and is therefore separable.

$\Longrightarrow$) Conversely, suppose that μ is separable. Then there exists some ensemble $\left\{(p_i)_{i=1}^d, (|\psi_i\rangle)_{i=1}^d\right\}$ of ρ such that each $|\psi\rangle_i$ is factorable. Therefore $\mu(\psi_i) = 0$ for each $i \in \{1, \ldots, d\}$ and so

$$\mu_{f,d}(\rho) = \sum_{i=1}^{d} f(p_i)\mu(\psi_i) = 0. \tag{2.28}$$

□

In the next chapter, we show that for $f(x) = x^2$, the sequence of f-extensions that is obtained when the Tsalis entanglement measure T_2 is extended from the set of pure states of a bipartite Hilbert space to the set of all density matrices, experiences the problem of exponentially vanishing gradients.

2.5 LOCC Channels and LOCC Monotones

In this section we show that Gudder's entanglement number is an LOCC monotone using a slight extension of a criterion due to Vidal [51]. Moreover we simplify Vidal's proof of this criterion. Furthermore during the course of its proof, we provide a representation of LOCC operations using trees, which we believe gives a better understanding to the complicated notion of LOCC channels.

Vidal [51] shows the following result: Assume that $\mathcal{X}$ is a finite dimensional Hilbert space, and $f : \mathrm{D}(\mathcal{X}) \to \mathbb{R}_{\geq 0}$ is a function which is invariant under unitaries (i.e. $f(U\rho U^*) = f(\rho)$ for every $\rho \in \mathrm{D}(\mathcal{X})$ and every unitary operator U on $\mathcal{X}$), and concave (i.e. $f(\lambda\sigma_1 + (1-\lambda)\sigma_2) \geq \lambda f(\sigma_1) + (1-\lambda) f(\sigma_2)$ for all $\lambda \in [0,1]$ and all $\sigma_1, \sigma_2 \in \mathrm{D}(\mathcal{X})$). Then define a function μ on pure states of $\mathcal{X}_1 \otimes \mathcal{X}_2$ where $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$, by

$$\mu(\psi) = f(\mathrm{Tr}_{\mathcal{X}_1} |\psi\rangle \langle\psi|) = f(\mathrm{Tr}_{\mathcal{X}_2} |\psi\rangle \langle\psi|) \tag{2.29}$$

Extend the function μ from the pure states of $\mathcal{X}_1 \otimes \mathcal{X}_2$ to all states of $\mathcal{X}_1 \otimes \mathcal{X}_2$ via the convex roof construction. If the infimum in the definition of the convex roof is always attained, (i.e. if OPSE's exist for every mixed state), then the extension of μ to mixed states of $\mathcal{X}_1 \otimes \mathcal{X}_2$ via the convex roof construction is an LOCC monotone.

Recall that a function μ defined on density operators of a multipartite Hilbert space $\mathcal{H}$ and taking non-negative real values is called an LOCC monotone if $\mu(\Lambda(\rho)) \leq \mu(\rho)$ for all density operators $\rho \in \mathrm{D}(\mathcal{H})$ and all LOCC channels Λ. The image of an LOCC channel may not be the density operators on the tensor product of two identical Hilbert spaces, and indeed it can be easily verified that Vidal's proof extends to that case very easily. Moreover, it is well known that for any two finite dimensional

Hilbert spaces $\mathcal{X}_1$ and $\mathcal{X}_2$, and for every normalized vector ψ of $\mathcal{X}_1 \otimes \mathcal{X}_2$, the set of non-zero eigenvalues of $\operatorname{Tr}_{\mathcal{X}_1} |\psi\rangle\langle\psi|$ is equal to the set of non-zero eigenvalues of $\operatorname{Tr}_{\mathcal{X}_2} |\psi\rangle\langle\psi|$. Thus if the Hilbert spaces $\mathcal{X}_1$ and $\mathcal{X}_2$ are equal, (or at least have equal dimension), then the matrices $\mathrm{T}_{\mathcal{X}_1} |\psi\rangle\langle\psi|$ and $\operatorname{Tr}_{\mathcal{X}_2} |\psi\rangle\langle\psi|$ are unitarily equivalent. In the proof that we present below we do not assume that the Hilbert spaces $\mathcal{X}_1$ and $\mathcal{X}_2$ are equal, or have equal dimension, and we simply assume that f is concave function that depends only on the nonzero eigenvalues of the densities matrices. Moreover, our proof is simpler than Vidal's proof, and along the proof we give a pictorial tree representation of LOCC channels that helps to understand this notion. More precisely, our main result of this section is the following:

Theorem 2.5.1. *Assume that a function f is defined on the density operators of all finite dimensional Hilbert spaces and takes values in non-negative real numbers. Assume that f is concave, and it depends only the non-zero eigenvalues of its argument. Let μ be defined on pure states of any bipartite Hilbert space via Equation (2.29), and extended to all mixed states of the bipartite Hilbert space via Equation (2.2), and assume that the infimum in (2.2) is always achieved. Then μ is an LOCC monotone.*

Before providing the proof of this result, we will provide a discussion on the structure of LOCC channels using the notion of the tree.

Let $\mathcal{A}$ be a Hilbert space whose states can be manipulated by only one party Alice, and $\mathcal{B}$ be a a Hilbert space whose states can be manipulated by only one party Bob (*local operations*). Then Alice may perform quantum channels on her space which will be of the form $\Phi_{\mathcal{A}}(X) = \sum_i A_i X A_i^*$ and Bob may perform channels on his space which may be of the form $\Phi_{\mathcal{B}}(Y) = \sum_j B_j X B_j^*$ where $\sum_i A_i^* A = I_{\mathcal{A}}$ and $\sum_j B_j^* B_j = I_{\mathcal{B}}$. Note that for simplicity, we keep the domains and ranges of each channel vague and simply write $\mathcal{A}$ to be the current space of Alice and $\mathcal{B}$ to be the current space of Bob, even though these spaces $\mathcal{A}$ and $\mathcal{B}$ keep changing during the application of every channel of the LOCC communication. Now suppose that Alice and Bob share a state

$\rho \in \mathrm{D}(\mathcal{A} \otimes \mathcal{B})$, then Alice could perform $\Phi_{\mathcal{A}}$ and Bob could perform $\Phi_{\mathcal{B}}$ and the post operation state would be $(\Phi_{\mathcal{A}} \otimes \Phi_{\mathcal{B}})(\rho)$. But operations of this form do not explicitly allow for communication between the parties and therefore do not adequately describe LOCC channels. In order to incorporate classical communications, we must have that each party's channels depend on the other's in some way.

Consider the following operation on a state $\rho \in \mathrm{D}(\mathcal{A} \otimes \mathcal{B})$: Alice performs the channel $\Phi_{\mathcal{A}}$ then measures the outcomes of her operation and sends the result to Bob via some method of classical communication. To an outside observer ,the operation on ρ would appear as

$$\Phi(\rho) = \sum_i \mathcal{E}_i(\rho) \text{ where } \mathcal{E}_i(\rho) = (A_i \otimes I_{\mathcal{B}})\rho(A_i^* \otimes I_{\mathcal{B}}) \tag{2.30}$$

The outcome states of such an operation would be given by

$$\rho_i = \frac{\mathcal{E}_i(\rho)}{\mathrm{Tr}\left(\mathcal{E}_i(\rho)\right)} = \frac{(A_i \otimes I_{\mathcal{A}})\rho(A_i^* \otimes I_{\mathcal{A}})}{\mathrm{Tr}\left((A_i^* A_i \otimes I_{\mathcal{B}}\rho)\right)} \text{ with probability } \mathrm{Tr}\left((A_i^* A_i \otimes I_{\mathcal{B}}\rho)\right) \text{ for each } i. \tag{2.31}$$

Alice then measures the state on her system and sends the result to Bob. Then Bob acts accordingly and applies the channel

$$\Phi_i(\sigma) = \sum_j \mathcal{E}_{i,j}(\sigma) \text{ where } \mathcal{E}_{i,j}(\sigma) = (I_{\mathcal{A}} \otimes B_{ij})\sigma(I_{\mathcal{A}} \otimes B_{ij}^*) \tag{2.32}$$

when he learns that the measured state was ρ_i. Since Φ_i is a channel for every fixed i, we obtain that,

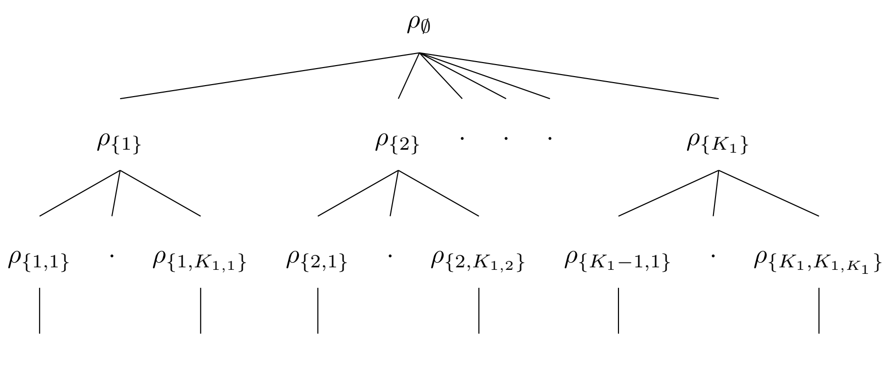
$$\sum_j B_{ij}^* B_{ij} = I_{\mathcal{B}} \tag{2.33}$$

and the channel that describes the final outcome of the scenario above will be of the form

$$\Lambda(\rho) = \sum_{i,j} (A_i \otimes B_{ij})\rho(A_i \otimes B_{ij}^*) = \sum_i \sum_j \mathcal{E}_{i,j} \circ \mathcal{E}_i(\rho), \tag{2.34}$$

But not all LOCC channels will be this simple. This scenario was only one round of operations and the communication only went in one direction. To extend the definition to more rounds of communication going in various directions, we can iteratively extend our scenario using the following tree structure.

Let $\mathcal{T} \subset \{\emptyset\} \cup \bigcup_{n=1}^{\infty} \mathbb{N}^n$ be a finite collection of finite subsets of $\mathbb{N}$ including the empty set. Endow $\mathcal{T}$ with a partial order $\prec$ by defining $\{x_1, \ldots, x_k\} \prec \{y_1, \ldots, y_l\}$ if $k < l$ and $x_i = y_i$ for all $i \in \{1, \ldots, k\}$. Moreover we define $\emptyset \prec x$ for all $x \in \mathcal{T} \backslash \{\emptyset\}$, and we denote the length $l(x) = k$ if $x = \{x_1, \ldots, x_k\}$ for some $x_i \in \mathbb{N}$ for each i. Also define $l(\emptyset) = 0$. Lastly we denote the *immediate successors* of an element x by $I(x) = \{y \in \mathcal{T} \big| x \prec y \text{ and } l(y) = l(x)+1\}$ and we denote $\mathcal{F}(\mathcal{T})$ to be the collection of *final nodes* of the tree $\mathcal{T}$ where the final nodes are the nodes which have no immediate successors. Using this notion of a tree, we can index all possible states that occur in an LOCC process by a tree $\mathcal{T}$. We will also use the convention $\rho_\emptyset$ to represent ρ before any operations have been applied. Thus an LOCC process can be visualized using the following tree



where ρ_x represents the *unnormalized* state that occurs after the x-th Krauss operator has been applied to previous *normalized* state. Moreover, each row of the tree represents a single party applying conditionally operations while the other party does nothing. Hence, each node in the tree ρ_y is of the form

$$\rho_y = (A_y \otimes I_\mathcal{B}) \frac{\rho_x}{\operatorname{Tr} \rho_x} (A_y^* \otimes I_\mathcal{B}) \text{ or } \rho_y = (I_\mathcal{A} \otimes B_y) \frac{\rho_x}{\operatorname{Tr} \rho_x} (I_\mathcal{A} \otimes B_y^*) \tag{2.35}$$

where $y \in I(x)$, A_x is an operation on Alice's space, and B_x is an operation on Bob's space. The physical significance of this construction being that $\dfrac{\rho_y}{\operatorname{Tr} \rho_y}$ is the state observed with probability $\operatorname{Tr} \rho_y$ after the previous party's operations. In order to refer to the Krauss operators at each node, we can define the maps $\mathcal{E}_y$ by

$$\mathcal{E}_y(X) = (A_y \otimes I_{\mathcal{B}})X(A_y^* \otimes I_{\mathcal{B}}) \text{ or } \mathcal{E}_y(X) = (I_{\mathcal{A}} \otimes B_y)X(I_{\mathcal{A}} \otimes B_y^*) \tag{2.36}$$

depending on which party is applying operations at node y. Thus using this notation, we express an LOCC channel Λ as

$$\Lambda(\rho) = \sum_{y_1 \in I(\emptyset)} \cdots \sum_{y_n \in I(y_{n-1})} \mathcal{E}_{y_n} \circ \cdots \circ \mathcal{E}_{y_1}(\rho_\emptyset) \tag{2.37}$$

where the y_n are the final nodes of the tree $\mathcal{T}$. Then to prove LOCC monotonicity, we need to ensure

$$\mu\Big(\sum_{y_1 \in I(\emptyset)} \cdots \sum_{y_n \in I(y_{n-1})} \mathcal{E}_{y_n} \circ \cdots \circ \mathcal{E}_{y_1}(\rho_\emptyset)\Big) \leq \mu(\rho_\emptyset). \tag{2.38}$$

But before we prove this we'll need the following lemma to ensure the self containment of the proof of our main result:

Lemma 2.5.2. *Let $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ be a finite dimensional Hilbert space, then the partial traces $Tr_{\mathcal{A}}$ and $Tr_{\mathcal{B}}$ are cyclic on the spaces $\mathcal{A}$ and $\mathcal{B}$ respectively.*

Proof. Let $A_1 \in \mathrm{L}(\mathcal{A}_1, \mathcal{A}_2)$, $A_2 \in \mathrm{L}(\mathcal{A}_2, \mathcal{A}_1)$, $\mathcal{B}_1 \in \mathrm{L}(\mathcal{B}_1, \mathcal{B}_2)$, and $\mathcal{B}_2 \in \mathrm{L}(\mathcal{B}_2, \mathcal{B}_1)$, then using the cyclicity of the trace operator on $\mathcal{A}$ we have that

$$\begin{aligned} \mathrm{Tr}_{\mathcal{A}_2}\Big(A_1A_2 \otimes B_1B_2\Big) &= (\mathrm{Tr} \otimes I_{\mathcal{B}_2})\Big(A_1A_2 \otimes B_1B_2\Big) = \mathrm{Tr}\Big(A_1A_2\Big)B_1B_2 \\ = \mathrm{Tr}\Big(A_2A_1\Big)B_1B_2 &= (\mathrm{Tr} \otimes I_{\mathcal{B}_2})\Big(A_2A_1 \otimes B_1B_2\Big) = \mathrm{Tr}_{\mathcal{A}_1}\Big(A_2A_1 \otimes B_1B_2\Big) \end{aligned} \tag{2.39}$$

which completes the proof. □

Henceforth, instead of writing the precise space that is being traced away, we will use the convention that $\mathrm{Tr}_{\mathcal{A}}$ traces out the space on the left, no matter what Hilbert space $\mathcal{A}$ is, and $\mathrm{Tr}_{\mathcal{B}}$ traces out the Hilbert space on the right. With this convention, the previous lemma can be written as

$$\mathrm{Tr}_{\mathcal{A}}\Big(A_1A_2 \otimes B_1B_2\Big) = \mathrm{Tr}_{\mathcal{A}}\Big(A_2A_1 \otimes B_1B_2\Big) \tag{2.40}$$

Using this lemma we can prove another useful property of the partial trace.

Corollary 2.5.3. *Let $A \in L(\mathcal{A}_1, \mathcal{A}_2)$, and $\rho \in D(\mathcal{A}_1 \otimes \mathcal{B})$, then*

$$Tr_{\mathcal{A}}\Big((A \otimes I_{\mathcal{B}})\rho(A^* \otimes I_{\mathcal{B}})\Big) = Tr_{\mathcal{A}}\Big((A^*A \otimes I_{\mathcal{B}})\rho\Big) \tag{2.41}$$

Proof. Let $\{|i\rangle\}_{i\in I}$ and $\{|j\rangle\}_{j\in J}$ be orthonormal bases for the Hilbert spaces $\mathcal{A}_1$ and $\mathcal{B}$ respectively. Then every $\rho \in \mathrm{D}(\mathcal{A}_1 \otimes \mathcal{B})$ can be written as $\rho = \sum_{i,j,k,l} \rho_{i,j,k,l} |i\rangle \langle k| \otimes |j\rangle \langle l|$. Thus

$$\begin{aligned}
\mathrm{Tr}_{\mathcal{A}}\Big((A \otimes I_{\mathcal{B}})\rho(A^* \otimes I_{\mathcal{B}})\Big) &= \sum_{i,j,k,l} \rho_{i,j,k,l} \mathrm{Tr}_{\mathcal{A}}\Big((A \otimes I_{\mathcal{B}})(|i\rangle \langle k| \otimes |j\rangle \langle l|)(A^* \otimes I_{\mathcal{B}})\Big) \\
&= \sum_{i,j,k,l} \rho_{i,j,k,l} \mathrm{Tr}_{\mathcal{A}}\Big(A |i\rangle \langle k| A^* \otimes |j\rangle \langle l|\Big) \\
&= \sum_{i,j,k,l} \rho_{i,j,k,l} \mathrm{Tr}_{\mathcal{A}}\Big(A^*A |i\rangle \langle k| \otimes |j\rangle \langle l|\Big) \\
&= \mathrm{Tr}_{\mathcal{A}}\Big((A^*A \otimes I_{\mathcal{B}})\rho\Big)
\end{aligned} \tag{2.42}$$

which is the desired result. □

Finally we are ready to give the

Proof of Theorem 2.5.1. Let $\rho \in \mathrm{D}(\mathcal{A} \otimes \mathcal{B})$ and let Λ be an LOCC channel with its corresponding tree $\mathcal{T}$. In order to show that μ decreases under Λ, we will first show that

$$\mu(\frac{\rho_x}{\mathrm{Tr}\,\rho_x}) \geq \sum_{y \in I(x)} \mathrm{Tr}\,\rho_y \mu(\frac{\rho_y}{\mathrm{Tr}\,\rho_y}) \tag{2.43}$$

where $x \in \mathcal{T} \setminus \mathcal{F}(\mathcal{T})$ and the ρ_x and ρ_y are the unnormalized states at nodes x and y respectively. Thus we need to show that μ decreases on average at each node in the tree. So let $x \in \mathcal{T}$ be a non-final node and let ρ_x be the unnormalized state at node x. Then without loss of generality we can write

$$\rho_y = (A_y \otimes I_{\mathcal{B}}) \frac{\rho_x}{\mathrm{Tr}\,\rho_x} (A_y^* \otimes I_{\mathcal{B}}) \tag{2.44}$$

for each $y \in I(x)$, where

$$\sum_y A_y^* A_y = I_{\mathcal{A}}. \tag{2.45}$$

Thus the normalized state at node y will be of the form $\frac{\rho_y}{\mathrm{Tr}\,(\rho_y)}$. Now let $\{(\lambda_i)_i, (\psi_i)_i\}$ be an OPSE for $\dfrac{\rho_x}{\mathrm{Tr}\,\rho_x}$. Then

$$\begin{aligned}
\mu(\frac{\rho_x}{\mathrm{Tr}\,\rho_x}) &= \sum_i \lambda_i \mu(\psi_i) \\
&= \sum_i \lambda_i f\Big(\mathrm{Tr}_{\mathcal{A}} \left|\psi_i\right\rangle \left\langle\psi_i\right|\Big) \\
&= \sum_i \lambda_i f\Big(\mathrm{Tr}_{\mathcal{A}} \sum_{y \in I(x)} (A_y \otimes I_{\mathcal{B}}) \left|\psi_i\right\rangle \left\langle\psi_i\right| (A_y^* \otimes I_{\mathcal{B}})\Big)
\end{aligned} \tag{2.46}$$

where the last equality follows from Corollary 2.5.3 and Equation (2.45).

Notice that for every $y \in I(x)$ we have

$$\sum_i \lambda_i (A_y \otimes I_{\mathcal{B}}) \left|\psi_i\right\rangle \left\langle\psi_i\right| (A_y^* \otimes I_{\mathcal{B}}) = (A_y \otimes I_{\mathcal{B}}) \frac{\rho_x}{\mathrm{Tr}\,\rho_x} (A_y^* \otimes I_{\mathcal{B}}) = \rho_y. \tag{2.47}$$

Next define $p_{i,y} = \left\langle\psi_i\right| A_y^* A_y \otimes I_{\mathcal{B}} \left|\psi_i\right\rangle$ and ψ_{iy} by $\left|\psi_{iy}\right\rangle = \dfrac{A_y \otimes I_{\mathcal{B}} \left|\psi_i\right\rangle}{\sqrt{p_{iy}}}$ for each i and each $y \in I(x)$ so that

$$\sum_i \lambda_i \frac{p_{iy}}{\mathrm{Tr}\,\rho_y} \left|\psi_{iy}\right\rangle \left\langle\psi_{iy}\right| = \frac{\rho_y}{\mathrm{Tr}\,\rho_y}. \tag{2.48}$$

Now taking the trace of both sides of the above equation we get $\sum_i \dfrac{p_{iy}\lambda_i}{\mathrm{Tr}\,\rho_y} = 1$. This implies that $\left\{\dfrac{\lambda_i p_{iy}}{\mathrm{Tr}\,\rho_y}, \left|\psi_{iy}\right\rangle\right\}_{i \in I}$ is a pure state ensemble (*not necessarily optimal*) of $\dfrac{\rho_y}{\mathrm{Tr}\,\rho_y}$ for each $y \in I(x)$. Thus

$$\begin{aligned}
\sum_i \lambda_i f\Big(\mathrm{Tr}_{\mathcal{A}} \sum_{y \in I(x)} (A_y \otimes I_{\mathcal{B}}) \left|\psi_i\right\rangle \left\langle\psi_i\right| (A_y^* \otimes I_{\mathcal{B}})\Big) &= \sum_i \lambda_i f\Big(\sum_y p_{iy} \mathrm{Tr}_{\mathcal{A}} \left|\psi_{iy}\right\rangle \left\langle\psi_{iy}\right|\Big) \\
&\geq \sum_{iy} \lambda_i p_{iy} f\Big(\mathrm{Tr}_{\mathcal{A}} \left|\psi_{iy}\right\rangle \left\langle\psi_{iy}\right|\Big) \\
&= \sum_{iy} \lambda_i p_{iy} \mu\Big(\psi_{iy}\Big)
\end{aligned} \tag{2.49}$$

Now since $\left\{\frac{\lambda_i p_{iy}}{\mathrm{Tr}\,\rho_y}, |\psi_{iy}\rangle\right\}_{i\in I}$ is not necessarily an optimal decomposition of $\frac{\rho_y}{\mathrm{Tr}\,\rho_y}$, we arrive at the inequality

$$\sum_i \frac{\lambda_i p_{iy}}{\mathrm{Tr}\,\rho_y}\mu\Big(\psi_{iy}\Big) \geq \mu\Big(\frac{\rho_y}{\mathrm{Tr}\,\rho_y}\Big) \tag{2.50}$$

for each $y \in I(x)$. It then follows that

$$\begin{aligned}
\sum_{iy} \lambda_i p_{iy}\mu\Big(\psi_{iy}\Big) &= \sum_{iy} \lambda_i p_{iy}\frac{\mathrm{Tr}\,\rho_y}{\mathrm{Tr}\,\rho_y}\mu\Big(\psi_{iy}\Big) \\
&= \sum_y \mathrm{Tr}\,\rho_y \sum_i \lambda_i \frac{p_{iy}}{\mathrm{Tr}\,\rho_y}\mu\Big(\psi_{iy}\Big) \\
&\geq \sum_{y\in I(x)} \mathrm{Tr}\,\rho_y \mu\Big(\frac{\rho_y}{\mathrm{Tr}\,\rho_y}\Big)
\end{aligned} \tag{2.51}$$

And therefore

$$\mu(\frac{\rho_x}{\mathrm{Tr}\,\rho_x}) \geq \sum_{y\in I(x)} \mathrm{Tr}\,\rho_y \mu(\frac{\rho_y}{\mathrm{Tr}\,\rho_y}) \tag{2.52}$$

Thus μ monotonically decreases at each node in $\mathcal{T}$. Next to show that $\mu(\rho) \geq \mu(\Lambda(\rho))$, we simply iterate the argument over all nodes in the tree to obtain

$$\begin{aligned}
\mu(\rho_\emptyset) &\geq \sum_{y_1\in I(\emptyset)} \mathrm{Tr}\,\rho_{y_1}\mu\Big(\frac{\rho_{y_1}}{\mathrm{Tr}\,\rho_{y_1}}\Big) \\
&\geq \sum_{y_1\in I(\emptyset)} \mathrm{Tr}\,\rho_{y_1} \sum_{y_2\in I(y_1)} \mathrm{Tr}\,\rho_{y_2}\mu\Big(\frac{\rho_{y_2}}{\mathrm{Tr}\,\rho_{y_2}}\Big) \\
&\geq \cdots \\
&\geq \sum_{y_1\in I(\emptyset)} \mathrm{Tr}\,\rho_{y_1} \cdots \sum_{y_n\in I(y_{n-1})} \mathrm{Tr}\,\rho_{y_n}\mu\Big(\frac{\rho_{y_n}}{\mathrm{Tr}\,\rho_{y_n}}\Big)
\end{aligned} \tag{2.53}$$

where each ρ_{y_j} is the unnormalized state at node $y_j \in \mathcal{T}$ and y_n are the final nodes of $\mathcal{T}$. Many authors [26, 7, 10, 50, 51] consider the above inequality the defining quality of an LOCC monotone and say that μ decreases *on average* under local operations and classical communication. But for a more functional result, we will go a step further and show that $\mu(\rho_\emptyset) \geq \mu(\Lambda(\rho_\emptyset))$. In this last step we will repeatedly use the fact that μ is convex which is an immediate consequence of the definition of convex roof extension. This repeated use of the convexity of μ corresponds to loss of information to the communicating parties, (Alice and Bob), and can be thought as black boxes in Alice's and Bob's labs dismissing states produced by the LOCC channel without informing Alice or Bob, (see [51, Citation [19]). Continuing the previous argument, we use convexity in the nodes y_n to obtain

$$\begin{aligned}&\sum_{y_1\in I(\emptyset)} \operatorname{Tr}\rho_{y_1}\cdots\sum_{y_n\in I(y_{n-1})}\operatorname{Tr}\rho_{y_n}\mu\Big(\frac{\rho_{y_n}}{\operatorname{Tr}\rho_{y_n}}\Big)\\ &\geq \sum_{y_1\in I(\emptyset)}\operatorname{Tr}\rho_{y_1}\cdots\sum_{y_{n-1}\in I(y_{n-2})}\operatorname{Tr}\rho_{y_{n-1}}\mu\Big(\sum_{y_n\in I(y_{n-1})}\rho_{y_n}\Big)\end{aligned} \tag{2.54}$$

Now can assume without loss of generality that $\rho_{y_n} = \mathcal{E}_{y_n}(\frac{\rho_{y_{n-1}}}{\operatorname{Tr}\rho_{y_{n-1}}})$ for each $y_n \in I(y_{n-1})$. Using this expression for ρ_{y_n}, it follows that

$$\sum_{y_n\in I(y_{n-1})}\rho_{y_n} = \sum_{y_n\in I(y_{n-1})}\mathcal{E}_{y_n}(\frac{\rho_{y_{n-1}}}{\operatorname{Tr}\rho_{y_{n-1}}}) \tag{2.55}$$

is a *normalized* state, which allows us to use convexity in the index y_{n-1} so that

$$\sum_{y_{n-1}\in I(y_{n-2})}\operatorname{Tr}\rho_{y_{n-1}}\mu\Big(\sum_{y_n\in I(y_{n-1})}\rho_{y_n}\Big) \geq \mu\big(\sum_{y_{n-1}\in I(y_{n-2})}\sum_{y_n\in I(y_{n-1})}\mathcal{E}_{y_n}(\rho_{y_{n-1}})\big) \tag{2.56}$$

for each $y_{n-1} \in I(y_{n-2})$. We then iterate the above argument to arrive at the inequality

$$\mu(\rho_\emptyset) \geq \mu\Big(\sum_{y_1\in I(\emptyset)}\cdots\sum_{y_n\in I(y_{n-1})}\mathcal{E}_{y_n}\circ\cdots\circ\mathcal{E}_{y_1}(\rho_\emptyset)\Big) = \mu(\Lambda(\rho_\emptyset)) \tag{2.57}$$

□

Corollary 2.5.4. *The entanglement number is an entanglement measure.*

Proof. Instead of using the definitions of the entanglement number given in Equations (2.11) and (2.12), we will use an alternative definition ([19, 20]) so that we can invoke our proof for establishing Theorem 2.6.3 in the next section.

First define a function f on density operators of finite dimensional Hilbert spaces by

$$f(\rho) = \sqrt{1 - \|\rho\|_2^2}, \tag{2.58}$$

where $\|\cdot\|_2$ denotes the Hilbert-Schmidt norm. Then if $\mathcal{A} \otimes \mathcal{B}$ is a bipartite Hilbert space, and $|\psi\rangle$ is a pure state of $\mathcal{A} \otimes \mathcal{B}$, we can define the entanglement number by

$$e(\psi) = f(\operatorname{Tr}_{\mathcal{A}} |\psi\rangle \langle\psi|) = f(\operatorname{Tr}_{\mathcal{B}} |\psi\rangle \langle\psi|). \tag{2.59}$$

Finally extend the definition of the entanglement number e to all mixed states of $\mathcal{A} \otimes \mathcal{B}$ using the convex roof construction.

As Gudder proved, the entanglement number is faithful when restricted to pure states [19]. Moreover since the function f is norm-continuous, the entanglement number is norm continuous on pure states by Equation (2.59). Therefore by Corollary 2.2.3 the entanglement number is faithful, i.e. it vanishes only on separable states.

In order to apply Theorem 2.5.1 to the entanglement number, (and deduce that the entanglement number is an LOCC monotone), notice that the function f defined in Equation (2.58) is concave since the Hilbert-Schmidt norm is convex, the square function is increasing and convex on the positive numbers, and the square root function is increasing and concave on the positive numbers. Moreover $f(\rho)$ depends only on the singular numbers (and hence on the non-zero eigenvalues) of the density matrix ρ for every density matrix ρ. Furthermore, the extension of the entanglement number to the mixed states of $\mathcal{A} \otimes \mathcal{B}$ via the convex roof function guarantees the existence of OPSE for all mixed states by Corollary 2.3.1. Thus by Theorem 2.5.1, the entanglement number is an LOCC monotone.

Finally notice that the entanglement number is invariant under local unitary transformations. Indeed for any pure state $|\psi\rangle \langle\psi|$ of a bipartite Hilbert space $\mathcal{A} \otimes \mathcal{B}$, local unitary transformations will not effect the non-zero eigenvalues of $\operatorname{Tr}_{\mathcal{A}} |\psi\rangle \langle\psi|$, or of $\operatorname{Tr}_{\mathcal{B}} |\psi\rangle \langle\psi|$, and therefore will not effect $e(\psi)$ according to Equation (2.59). This will remain valid under the convex roof extension of the entanglement number to mixed states. □

Motivated by Equations (2.58) and (2.59) we now define a family of entanglement measures. *We assume that all Hilbert spaces mentioned in this section are finite dimensional.* Let $\mathcal{Z}$ be a (finite dimensional) Hilbert space and $1 < p < \infty$ then define $f_p : \mathrm{D}(\mathcal{Z}) \to \mathbb{R}_{\geq 0}$ by

$$f_p(\rho) = \left(1 - ||\rho||_p^p\right)^{\frac{1}{p}} \tag{2.60}$$

where $||\cdot||_p$ is the Schatten p-norm on $\mathrm{L}(\mathcal{Z})$.

Remark 2.6.1. *For $1 < p < \infty$ the function f_p has the following properties:*

1. *f_p depends only on the non-zero eigenvalues of its argument.*
2. *f_p is concave.*
3. *f_p is norm-continuous.*

The proof of the Remark is similar to the corresponding properties of the function f defined in Equation (2.58).

Then we define a function μ_p on pure states of a bipartite Hilbert space $\mathcal{A} \otimes \mathcal{B}$ by

$$\mu_p(\psi) = f_p(\mathrm{Tr}_{\mathcal{A}} |\psi\rangle \langle\psi|) = f_p(\mathrm{Tr}_{\mathcal{B}} |\psi\rangle \langle\psi|) \tag{2.61}$$

for all pure states $|\psi\rangle \in \mathcal{A} \otimes \mathcal{B}$, and extending μ_p by the convex roof construction to all mixed states as in Equation (2.2). We call μ_p *the p-number of a state.* Now notice that for $p = 2$, the p-number of a state coincides with the entanglement number of the state.

Remark 2.6.2. *For all $p \in (1, \infty)$, every mixed state on a bipartite Hilbert space $\mathcal{A} \otimes \mathcal{B}$ admits an OPSE for the convex roof construction defining μ_p.*

Indeed since f_p is norm-continuous we obtain that μ_p is norm-continuous and the Remark follows from Theorem 2.2.1.

Theorem 2.6.3. *For all $p \in (1, \infty)$, μ_p is an entanglement measure.*

Proof. The proof of Corollary 2.5.4 repeats verbatim here, with the only addition that needs to be made is to verify that μ_p is faithful when restricted to pure states, (a statement that for the entanglement number e was proved by Gudder [19]).

For pure states, notice that we can compute μ_p using only the Schmidt decomposition of a state as follows. Let $|\psi\rangle \in \mathcal{A} \otimes \mathcal{B}$ be a pure state with Schmidt decomposition $\sum_k \sqrt{\lambda_k} \, |\alpha_k\rangle_{\mathcal{A}} \otimes |\beta_k\rangle_{\mathcal{B}}$. Then

$$\operatorname{Tr}_{\mathcal{A}} |\psi\rangle \langle\psi| = \operatorname{Tr}_{\mathcal{A}} \Big(\sum_{k,l} \sqrt{\lambda_k \lambda_l} \, |\alpha_k\rangle \langle\alpha_l| \otimes |\beta_k\rangle \langle\beta_l| \Big) = \sum_k \lambda_k \, |\beta_k\rangle \langle\beta_k| \tag{2.62}$$

Notice that a similar result will follow if $\mathcal{B}$ is traced out instead of $\mathcal{A}$ because of the orthonormality of the $(|\alpha_k\rangle)_k$ as well as the $(|\beta_k\rangle)_k$. Thus

$$\mu_p(\psi) = \Big(1 - \sum_k \lambda_k^p\Big)^{\frac{1}{p}}. \tag{2.63}$$

Finally we verify that for a pure state ψ in a bipartite Hilbert space $\mathcal{A} \otimes \mathcal{B}$, $\mu_p(\psi) = 0$ if and only if ψ is factorable.

$\Longrightarrow$) Suppose that $\mu_p(\psi) = 0$ and let $\sum_k \sqrt{\lambda_k} \, |\alpha\rangle_{\mathcal{A}} \otimes |\beta_k\rangle_{\mathcal{B}}$ be the Schmidt decomposition of $|\psi\rangle$. Then the Schmidt coefficients satisfy the following system of equations.

$$\sum_k \lambda_k^p = 1 \text{ and } \sum_k \lambda_k = 1. \tag{2.64}$$

Since each $\lambda_k \in [0, 1]$, it must follow that $\lambda_l = 1$ for some l and that $\lambda_j = 0$ for all $j \neq l$. Thus $|\psi\rangle = |\alpha_l\rangle \otimes |\beta_l\rangle$ as desired.

$\Longleftarrow$) Conversely, suppose that $|\psi\rangle = |u\rangle \otimes |v\rangle$ for some states $u \in \mathcal{A}$ and $v \in \mathcal{B}$. Then the length of the Schmidt decomposition is 1 with Schmidt coefficient 1. Thus $\mu_p(\psi) = 0$. $\square$

We also leave as an exercise to the reader to show that for any pure state $|\psi\rangle$ in a bipartite Hilbert space,

$$\lim_{p \to 1} \frac{d}{dp} \log \mu_p(\psi) = \mathrm{S}(\psi) \tag{2.65}$$

The p-numbers of a state also obey the following monotonicity property:

Remark 2.6.4. *Let $1 < p < q < \infty$, then $0 < \mu_p(\rho) < \mu_q(\rho) < 1$ for all states $\rho \in D(\mathcal{A} \otimes \mathcal{B})$.*

This can be easily seen first for pure states using Equation (2.63) and the fact that the sum of the p-th powers of the eigenvevalues of any density operator is in the interval $(0, 1]$ for any $p \in (1, \infty)$, and then observing that the monotonicity passes to the mixed states via the convex roof extension.

Remark 2.6.5. *A measure of entanglement similar to the p-number was proposed by Cirone [8]. Cirone defined his measure ν_p only for pure bipartite states by $\nu_p(\psi) = 1 - \sum_{i=1}^{n} \lambda_i^p$ where the $\sqrt{\lambda_i}$ are the Schmidt coefficients of the pure state ψ. Majorization techniques were used to discuss conversion of pure states to pure states via LOCC operations and LOCC monotonicity was restricted to pure states only. It was left as an open problem to extend ν_p to general density matrices. The solution for states on bipartite Hilbert spaces is implied by Gudder's [19] result on the entanglement number for the case of $p = 2$, and our results in Section 2.6 for all other p's. Namely μ_p^p is the extension of ν_p to all bipartite states.*

2.7 Measurement Inequalities

In this section we provide a list of measurement and lipschitz inequalities related to general quantum entropies. The first inequality is a generalization of the inequality by von Neumann [38] stating that application of a measurement channel to a state cannot decrease the entanglement entropy. We show that this inequality is true for a large class of functions on a Hilbert space which may not even be considered a proper "entropy".

Theorem 2.7.1 (Generalized Measurement Increasing Inequality)**.** *Let $\mathcal{H}$ be a finite dimensional Hilbert space and $f : D(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ be concave, continuous, and unitarily* *invariant. Define* $\Phi_{\mathcal{B}}(\rho) = \sum_{|e\rangle \in \mathcal{B}} \langle e| \rho |e\rangle |e\rangle \langle e|$ *for any orthonormal basis* $\mathcal{B}$ *and operator* ρ*. Then* $f(\rho) \leq f\Big(\Phi_{\mathcal{B}}(\rho)\Big)$ *for every orthonormal basis* $\mathcal{B}$*, and density operator* ρ*.*

Proof. Since f is unitarily invariant, it is a function of the spectrum of a given density ρ. And so f can be thought of as a concave function on the probability vectors of length N where $N = \dim(\mathcal{H})$. Suppose ρ has spectrum $\{\sigma_1, \ldots, \sigma_N\}$ and notice that the spectrum of $\Phi_{\mathcal{B}}$ is $\{\langle e_1| \rho |e_1\rangle, \ldots, \langle e_N| \rho |e_N\rangle\}$ for any orthonormal basis $\mathcal{B}$. Thus to prove our result, it suffices to show that $f(\sigma_1, \ldots, \sigma_N) \leq f\Big(\langle e_1| \rho |e_1\rangle, \ldots, \langle e_N| \rho |e_N\rangle\Big)$. Now suppose that ρ has spectral decomposition $\rho = \sum_i \sigma_i |\chi_i\rangle \langle\chi_i|$, then for any j, $\langle e_j| \rho |e_j\rangle = \sum_i \sigma_i \Big|\langle\chi_i|e_j\rangle\Big|^2$. This observation motivates us to define $X_{i,j} := \Big|\langle\chi_i|e_j\rangle\Big|^2$, so that

$$f\Big(\langle e_1| \rho |e_1\rangle, \ldots, \langle e_N| \rho |e_N\rangle\Big) = f\Big(\sum_i \sigma_i X_{i,1}, \ldots, \sum_i \sigma_i X_{i,N}\Big) \tag{2.66}$$

Expressing f in this way allows us to prove the desired inequality by minimizing over all relevant X. But before we can minimize, we must first determine the domain over which the minimization is to take place. Notice that for each j, $\sum_i X_{ij} = \mathrm{Tr}\, |e_j\rangle \langle e_j| = 1$ and for each i, $\sum_j X_{ij} = \mathrm{Tr}\, |\chi_i\rangle \langle\chi_i| = 1$. These constraints tell us that we're actually minimizing over the set of doubly stochastic matrices. The Birkhoff-von Neumann theorem states that the set of doubly stochastic matrcies is the convex hull of the set of permuation matrices. Thus the domain over which the minimization takes place is a polytope whose vertices are the permutation matrices. And since f can easily be shown to be concave in X, the minimum must occur at one of the vertices of the polytope. This implies that for each i, j, $\Big|\langle\chi_i|e_j\rangle\Big| \in \{0, 1\}$, therefore the basis $\mathcal{B}$ differs from the eigenbasis of ρ only by local phase shift, and will thus produce the same measurements. The result follows. $\square$

Recall that the Tsallis entropy of a quantum state ρ is defined by

$$T_\alpha(\rho) = \frac{1}{1-\alpha}\Big(\mathrm{Tr}\,(\rho^\alpha) - 1\Big) \tag{2.67}$$

for all positive α not equal to 1. An interesting consequence of the above lemma is that we are able to reproduce some Lipschitz continuity bounds in a similar but slightly different way as other others [22, 43] for the Tsallis α-entropy for $\alpha > 1$.

Proposition 2.7.2. *For any two density operators $\rho, \sigma \in D(\mathcal{H})$ for some finite dimensional Hilbert space $\mathcal{H}$, it follows that*

$$|T_\alpha(\rho) - T_\alpha(\sigma)| \leq \frac{\alpha}{\alpha - 1}\|\rho - \sigma\|_1 \tag{2.68}$$

Proof. Our proof will mimic the proof of Lemma III.2 in [25] where the Tsallis entropy takes the place of the von Neumann Entropy. Let $\mathcal{H}$ be finite dimensional with $\rho, \sigma \in \mathrm{D}(\mathcal{H})$. Suppose without loss of generality that $T_\alpha(\rho) \geq T_\alpha(\sigma)$ and let $\{|1\rangle, \ldots, |n\rangle\}$ be the eigenbasis of σ. Lastly let Λ be the measurement channel over the eigenbasis of σ. We then have that

$$\begin{aligned} |T_\alpha(\rho) - T_\alpha(\sigma)| &= T_\alpha(\rho) - T_\alpha(\sigma) \\ &\leq T_\alpha\Big(\Lambda(\rho)\Big) - T_\alpha\Big(\Lambda(\sigma)\Big) \end{aligned} \tag{2.69}$$

by the measurement increasing inequality. Next we use the functional calculus to write

$$T_\alpha\Big(\Lambda(\rho)\Big) = \frac{1}{1-\alpha} f\Big(\, \langle 1|\,\rho\,|1\rangle, \ldots, \langle n|\,\rho\,|n\rangle\,\Big) \tag{2.70}$$

where $f(x_1, \ldots, x_n) = 1 - \sum_{i=1}^n x^\alpha$ for $(x_1, \ldots, x_d)$ on the $n-$th probability simplex. Notice that $\|\nabla f\|_\infty = \alpha$ where the norm is taken over the probability simplex.

Putting everything together, we can then write

$$\begin{aligned}
&T_\alpha\Big(\Lambda(\rho)\Big) - T_\alpha\Big(\Lambda(\sigma)\Big) \\
&= \frac{1}{\alpha - 1}\Big(f\Big(\langle 1|\rho|1\rangle, \ldots, \langle n|\rho|n\rangle\Big) - f\Big(\langle 1|\sigma|1\rangle, \ldots, \langle n|\sigma|n\rangle\Big)\Big) \\
&\leq \frac{\|\nabla f\|_\infty}{\alpha - 1}\Big(\sum_{i=1}^{n}\Big(\langle i|\rho|i\rangle - \langle i|\sigma|i\rangle\Big)^2\Big)^{\frac{1}{2}} \\
&= \frac{\alpha}{\alpha - 1}\Big(\sum_{i=1}^{n}\Big(\langle i|\rho - \sigma|i\rangle\Big)^2\Big)^{\frac{1}{2}}.
\end{aligned} \tag{2.71}$$

Next we can use Jensen's operator inequality to arrive at

$$\sum_{i=1}^{n}\Big(\langle i|\rho - \sigma|i\rangle\Big)^2 \leq \sum_{i=1}^{n}\langle i|\Big(\rho - \sigma\Big)^2|i\rangle. \tag{2.72}$$

Lastly, using the definition of the Hilbert Schmidt norm and the monotonicity of the Schatten p-norms, it follows that

$$\begin{aligned}
\frac{\alpha}{\alpha - 1}\Big(\sum_{i=1}^{n}\Big(\langle i|\rho - \sigma|i\rangle\Big)^2\Big)^{\frac{1}{2}} &\leq \frac{\alpha}{\alpha - 1}\Big(\sum_{i=1}^{n}\langle i|\Big(\rho - \sigma\Big)^2|i\rangle\Big)^{\frac{1}{2}} \\
&= \frac{\alpha}{\alpha - 1}\|\rho - \sigma\|_{HS} \\
&\leq \frac{\alpha}{\alpha - 1}\|\rho - \sigma\|_1
\end{aligned} \tag{2.73}$$

thus proving the claim. □

The next lemmas outline methods of extending Lipschitz functions on the set of density operators to the set of positive Hermitian operators with trace less than or equal to 1. These types of operators can be thought of as *incomplete measurements* of a quantum state. Thus the following lemmas can be used to construct Lipschitz functions on the measurements of a quantum state.

Lemma 2.7.3. *Let $f : D(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ be continuous, bounded by a constant M, and Lipschitz with constant η. Define the extension $\tilde{f} : \Big(\mathcal{P}(\mathcal{H}) \cap Herm(\mathcal{H})\Big)_{Tr \leq 1} \to \mathbb{R}_{\geq 0}$ by $\tilde{f}(P) = Tr(P)f\Big(\frac{P}{Tr(P)}\Big)$. The extension $\tilde{f}$ is then Lipschitz with constant at most $M + 2\eta$.*

Proof. Suppose that $p, q \in \mathcal{P}(\mathcal{H}) \cap Herm(\mathcal{H})$ with $\operatorname{Tr} p, \operatorname{Tr} q \leq 1$ and suppose without loss of generality that $\operatorname{Tr} q \leq \operatorname{Tr} p$.

$$\begin{aligned}
|f\tilde{(p)} - \tilde{f}(q)| &= |\operatorname{Tr} p f(\frac{p}{\operatorname{Tr} p}) - \operatorname{Tr} q f(\frac{q}{\operatorname{Tr} q})| \\
&= |\operatorname{Tr} p f(\frac{p}{\operatorname{Tr} p}) - \operatorname{Tr} q f(\frac{p}{\operatorname{Tr} p}) + \operatorname{Tr} q f(\frac{p}{\operatorname{Tr} p}) - \operatorname{Tr} q f(\frac{q}{\operatorname{Tr} q})| \\
&\leq |\operatorname{Tr} p - \operatorname{Tr} q| f(\frac{p}{\operatorname{Tr} p}) + \operatorname{Tr} q |f(\frac{p}{\operatorname{Tr} p}) - f(\frac{q}{\operatorname{Tr} q})| \\
&\leq M\|p - q\| + \eta \operatorname{Tr} q \|\frac{p}{\operatorname{Tr} p} - \frac{q}{\operatorname{Tr} q}\| \\
&= M\|p - q\| + \eta \|\frac{\operatorname{Tr} q}{\operatorname{Tr} p} p - q\|
\end{aligned} \tag{2.74}$$

Now we need to bound the term $\|\frac{\operatorname{Tr} q}{\operatorname{Tr} p} p - q\|$ in the right hand side of the above ineqaulity.

$$\begin{aligned}
\|\frac{\operatorname{Tr} q}{\operatorname{Tr} p} p - q\| &= \|\frac{\operatorname{Tr} q}{\operatorname{Tr} p} p - \frac{\operatorname{Tr} q}{\operatorname{Tr} p} q + \frac{\operatorname{Tr} q}{\operatorname{Tr} p} - q\| \\
&\leq \frac{\operatorname{Tr} q}{\operatorname{Tr} p} \|p - q\| + |\frac{\operatorname{Tr} q}{\operatorname{Tr} p} - 1| \cdot \|q\| \\
&\leq \frac{\operatorname{Tr} q}{\operatorname{Tr} p} \|p - q\| + |\frac{\operatorname{Tr} q}{\operatorname{Tr} p} - 1| \cdot \|p\| \\
&= \|p - q\| + \frac{1}{\operatorname{Tr} p} |\operatorname{Tr} q - \operatorname{Tr} p| \cdot \|p\| \\
&= \|p - q\| + |\operatorname{Tr} q - \operatorname{Tr} p| \\
&\leq 2\|p - q\|
\end{aligned} \tag{2.75}$$

Putting everything together, we can then see that

$$|\tilde{f}(p) - \tilde{f}(q)| \leq (M + 2\eta)\|p - q\|. \tag{2.76}$$

□

Using the above lemma, we can also extend functions that are Lipschitz on the set of pure states to the set $\Big(\mathcal{P}(\mathcal{H}) \cap Herm(\mathcal{H})\Big)_{\operatorname{Tr} \leq 1}$ in a similar manner.

Lemma 2.7.4. *Let $\mathcal{H} = \mathcal{A} \otimes \mathcal{B} \otimes \mathcal{W}$ for some Hilbert spaces $\mathcal{A}, \mathcal{B}$, and $\mathcal{W}$ where $\mathcal{W}$ has basis $\{|i\rangle : 1 \leq i \leq d\}$ and define $P_i = |i\rangle\langle i|$ for each i. Lastly, let f :*

$\Pi\Big(\mathcal{A}\otimes\mathcal{B}\Big)\to\mathbb{R}_{\geq 0}$ be Lipschitz and bounded on the pure states of $D(\mathcal{A}\otimes\mathcal{B})$, the the function $f_M:\Pi\Big(\mathcal{A}\otimes\mathcal{B}\otimes\mathcal{W}\Big)\to\mathbb{R}_{\geq 0}$ defined by

$$f_M(X)=\sum_i Tr\Big(P_iX\Big)f\left(\frac{Tr_{\mathcal{W}}\Big(P_iXP_i\Big)}{Tr\Big(P_iX\Big)}\right) \tag{2.77}$$

is Lipschitz in the trace norm with constant $M+2\eta$, where η is the Lipschitz constant of f and M is the supremum of f on $\mathcal{A}\otimes\mathcal{B}$.

Proof. Let X,Y be pure states on $\mathcal{A}\otimes\mathcal{B}\otimes\mathcal{W}$ and define $X_i=P_iXP_i$ and $Y_i=P_iY_i$. Then using both the previous lemma and the fact that the partial trace is a contraction map, it follows that

$$\begin{aligned}|f_M(X)-f_M(Y)| &\leq \sum_i\left|\mathrm{Tr}\,(X_i)f\left(\frac{\mathrm{Tr}_{\mathcal{W}}X_i}{\mathrm{Tr}\,X_i}\right)-\mathrm{Tr}\,(Y_i)f\left(\frac{\mathrm{Tr}_{\mathcal{W}}Y_i}{\mathrm{Tr}\,Y_i}\right)\right|\\ &\leq (M+2\eta)\sum_i\|\mathrm{Tr}_{\mathcal{W}}\Big(X_i-Y_i\Big)\|_1 \\ &\leq (M+2\eta)\sum_i\|X_i-Y_i\|_1.\end{aligned} \tag{2.78}$$

Lastly, since X_i-Y_i is diagonal and $(X_i-Y_i)(X_j-Y_j)=0$ when $i\neq j$,

$$\sum_i\|X_i-Y_i\|_1=\|X-Y\|_1 \tag{2.79}$$

which implies that $|f_M(X)-f_M(Y)|\leq(M+2\eta)\|X-Y\|_1$ □

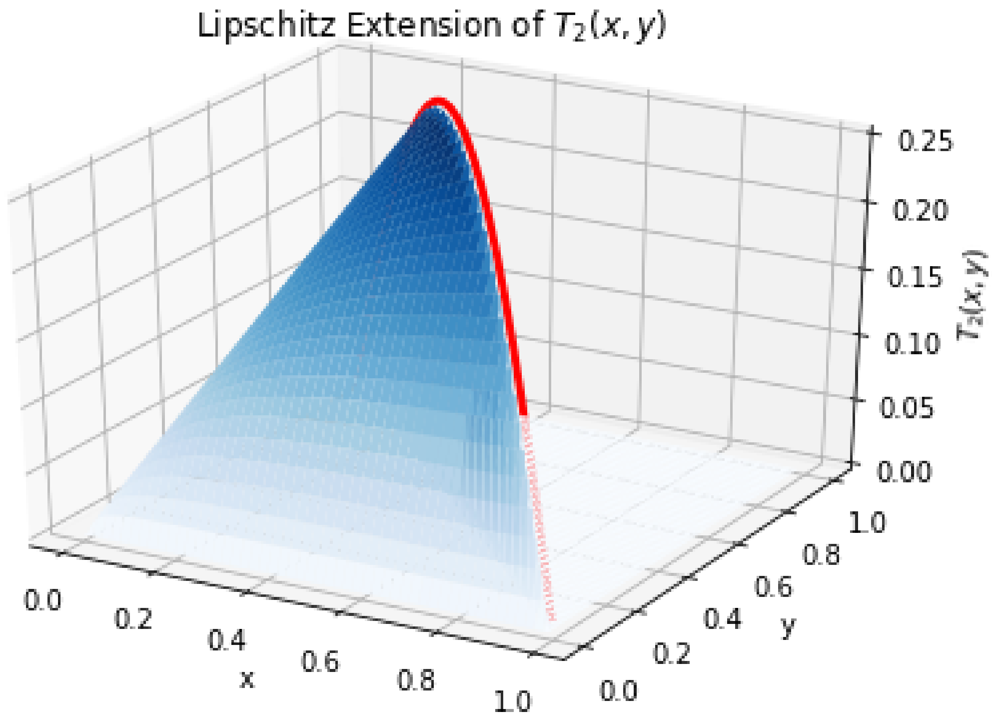


Figure 2.1: Lipschitz extension of T_2 on the probability simplex in two variables.

Chapter 3

A Quantum Algorithm for Approximating OPSE

3.1 Approximating Entanglement Measures

Even after discussing the existence of optimal pure state ensembles, we still don't much about how to actually find them, as the proofs for their existence were not constructive. However, methods for approximating them on a quantum computer are heavily implied by the following theorem [46, 27, 21, 37, 16].

Theorem 3.1.1 (Purification Theorem)**.** *Let $\mathcal{H}$ be a finite dimensional Hilbert space and suppose that $\rho \in D(\mathcal{H})$ can be expressed as both of the following pure state ensembles:*

$$\sum_{i=1}^{d} p_i \left|\psi_i\right\rangle \left\langle\psi_i\right| = \rho = \sum_{i=1}^{d} q_i \left|\phi_i\right\rangle \left\langle\phi_i\right|$$

and define

$$\left|\psi\right\rangle = \sum_{i=1}^{d} \sqrt{p_i} \left|\psi_i\right\rangle \otimes \left|i\right\rangle \tag{3.1}$$

where $\{\left|i\right\rangle\}_{i=1}^{d}$ is an orthonormal basis for $\mathcal{W} := \mathbb{C}^d$. Then there exists a unitary $U \in \mathbb{U}(d)$ such that $(I_{\mathcal{H}} \otimes U)\left|\psi\right\rangle = \sum_{i=1}^{d} \sqrt{q_i} \left|\phi_i\right\rangle \otimes \left|i\right\rangle$.

This theorem tells us that any two pure state ensembles of the same the length are related by a unitary map. So if we already know one pure state ensemble of a given density, we can find all others simply by applying unitaries onto the ancilla space of a purification $\left|\psi\right\rangle$ of ρ. Hence we can minimize the quantity $\sum_{i=1}^{d} q_i \mu(\phi_i)$ over the unitary group, where the q_i and ϕ_i are as in the previous theorem, and μ is an arbitrary entanglement measure. Moreover, the theorem hints at using the following quantum circuit to achieve our goal.

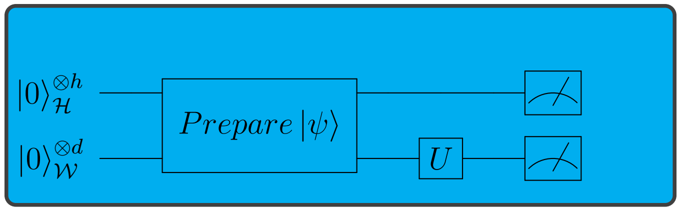


Figure 3.1: The quantum circuit implied by the purification theorem.

Here $h = \dim(\mathcal{H})$ where $\rho \in \mathcal{H}$ and $d = \dim(\mathcal{W})$ where $\mathcal{W}$ is the ancilla space. In particular, d is also the length of the pure state decompositions output from the circuit. Usually d is taken to be greater than h.

Examining the circuit, notice that after preparing the system in the state $|\psi\rangle$ and applying the unitary U, the system is put into the superposition $\sum_{i=1}^{d} \sqrt{q_i}\, |\phi_i\rangle \otimes |i\rangle$. When measuring the system, if the state $|j\rangle$ is observed on the ancilla space, then it must be the case that the top qubits are in the state $|\phi_j\rangle$ with certainty. Running this circuit a number of times allows the observer to experimentally reconstruct both $|\phi_j\rangle$ and q_j, and therefore compute the quantity $\sum_{i=1}^{d} q_i \mu(\phi_i)$. But in order to find the best ensemble we need expressions for $|\varphi_i\rangle$ and q_i in terms of U. So examining step by step, the circuit first prepares the purification $|\psi\rangle$. The unitary U is then applied to the ancilla space resulting in the state $(I_{\mathcal{H}} \otimes U)\, |\psi\rangle \langle\psi|\, (I_{\mathcal{H}} \otimes U^*)$. Lastly, measuring the ancilla yields the states

$$\frac{1}{q_i}\Big(I_{\mathcal{H}} \otimes |i\rangle \langle i|\, U\Big)\, |\psi\rangle \langle\psi| \Big(I_{\mathcal{H}} \otimes U^* \,|i\rangle \langle i|\Big) \tag{3.2}$$

where

$$\begin{aligned} q_i &= \mathrm{Tr}\left(\Big(I_{\mathcal{H}} \otimes |i\rangle \langle i|\, U\Big)\, |\psi\rangle \langle\psi| \Big(I_{\mathcal{H}} \otimes U^* \,|i\rangle \langle i|\Big)\right) \\ &= \langle\psi| \Big(I_{\mathcal{H}} \otimes U^* \,|i\rangle \langle i|\, U\Big)\, |\psi\rangle\,. \end{aligned} \tag{3.3}$$

And to access the state on the space $\mathcal{H}$, we partial trace the ancilla and use the partial cyclicity of the partial trace to arrive at the following expression for the states in the new ensemble,$|\varphi_i\rangle\langle\varphi_i|$:

$$|\varphi_i\rangle\langle\varphi_i| = \frac{1}{q_i}\mathrm{Tr}_{\mathcal{W}}\Bigg(\Big(I_{\mathcal{H}} \otimes U^* |i\rangle\langle i| U\Big)|\psi\rangle\langle\psi|\Bigg) \tag{3.4}$$

for each i in $\{1, \ldots, d\}$.

The natural question that arises is how to search the unitary group for unitary minimizing this quantity. There are several ways to accomplish this task, the most common being the utilization of a parametrization of the unitary group [48], or the use of random parametrized quantum circuits [30, 45, 35]. We choose the latter, as this choice will greatly simplify calculations in the proof of our main result. Specifically, we will use parametrized unitaries of the following form:

$$U(\theta_1, \ldots, \theta_L) = \prod_{l=1}^{L} exp\Big(i\theta_l V_l\Big)E_l \tag{3.5}$$

where V_l is a Hermitian operator and $\theta_l \in [0, 2\pi)$ for each $l \in \{1, \ldots, L\}$, and the E_l is an entangling gate independent of the parameters $\theta_1, \ldots, \theta_L$. Common choices for E_l are a CX or CZ ladder. In applications, the V_l are often taken to be a randomly chosen string of Pauli operators, which is the convention we use. Since we have introduced the parameters $\theta_1, \ldots, \theta_L$ into our unitaries, the ansatz we wish to train can be written as $\sum_{i=1}^{d} q_i(\theta_1, \ldots, \theta_L)\mu\Big(\varphi_i(\theta_1, \ldots, \theta_L)\Big)$, thus allowing us to use common optimization techniques to approximate the minimum. Thus a quantum algorithm to approximate optimal pure state ensembles can be summarized as below:

Algorithm 1: Variational Approximation of OPSE

Input: An ensemble $\{(p_i)_{i=1}^d, (\psi_i)_{i=1}^d\}$ of a density matrix $\rho \in \mathrm{D}(\mathbb{C}^h)$ and an optimization routine [42].

Parameters: Initial parameters $\theta_1^0, \ldots, \theta_L^0$.

while *Optimization routine has not converged* **do**

Prepare the purification $|\psi\rangle = \sum_{i=1}^d \sqrt{p_i}\, |\psi_i\rangle \otimes |i\rangle$ on $\mathcal{H} \otimes \mathcal{W}$.

Construct $U(\theta_1, \ldots, \theta_L)$ as above and apply it to the ancilla, $\mathcal{W}$.

Measure.

Use measurements to compute $\sum_{i=1}^d q_i(\theta_1, \ldots, \theta_L)\mu\Big(\varphi_i(\theta_1, \ldots, \theta_L)\Big)$ (or related quantities such as gradients as required by the chosen optimization technique).

Update $\theta_1, \ldots, \theta_L$ according to optimization routine.

return $\{\Big(q_i(\theta_1, \ldots, \theta_L)\Big)_{i=1}^d, \Big(\varphi_i(\theta_1, \ldots, \theta_L)\Big)_{i=1}^d\}$.

An algorithm of this type is called a *variational quantum algorithm*(VQA) [3]. VQA has shown great success in various settings such as Hamiltonian Simulation and Quantum Machine Learning [30, 3, 6, 29] where most cost functions are physically implementable on a quantum computer. In the algorithm described above, we instead train a nonlinear cost function on the measurements of a physically implementable ansatz. While VQA is believed (not yet proven) to offer a possible quantum advantage over classical algorithms for certain problems [3] Chapter 1 - section C, there are still drawbacks, namely the issue of *Barren Plateaus* [36, 5, 1, 40]. These are regions of the training landscape on which the partial derivatives with respect to the parameters decrease exponentially as the number of qubits in the experiment increase. More precisely, we give the following definition.

Consider the ansatz in Equation (3.5) on d qubits, with objective function

$$C(\theta_1, \ldots, \theta_L) = \sum_{i=1}^d q_i(\theta_1, \ldots, \theta_L)\mu\Big(\varphi_i(\theta_1, \ldots, \theta_L)\Big). \tag{3.6}$$

Then the training landscape will contain barren plateaus if and only if for each

$j \in \{1, \ldots, L\}$,

$$\mathrm{Var}_{\mathbb{U}(2^k)}\Big[\partial_{\theta_j} C(\theta_1, \ldots, \theta_L)\Big] \sim \mathcal{O}(a^k) \tag{3.7}$$

where the variance is taken with respect to the Haar measure on the Unitary group, k is the number of qubits, and is a real number such that $a \in (0,1)$ [1]. When this variance is exponentially small, the ansatz can become untrainable in practice because an exponentially large amount of precision will be required on the classical hardware to evaluate the derivatives. For instance, if it were shown that the variance is on the order of $\frac{1}{2^k}$, where k is the number of qubits, then systems with only 100 qubits can start creating issues since the number of classical bits required to differentiate the values of the gradient from 0 will be in the order of 2^{-100}.

We attempted to determine the existence of barren plateaus for both the von Neumann entanglement entropy, and the Tsallis entanglement entropy, but were unsuccessful due to the non-polynomial nature of $\sum_{i=1}^{d} q_i(\theta_1, \ldots, \theta_L)\mu\Big(\varphi_i(\theta_1, \ldots, \theta_L)\Big)$ on the entries of the unitary $U(\theta_1, \ldots, \theta_L)$. However, we were successful in showing that the x^2, d-extension of the Tsallis entropy does indeed exhibit barren plateaus. Notice that $T_{f,d}$ will retain the faithfullness and local unitary invariance of the Tsallis entanglement entropy but it will not be LOCC monotonic. $T_{f,d}$ is given by

$$T_{f,d}(\rho) = \inf\left\{\sum_{i=1}^{d} \lambda_i^2 T_2(\psi_i) : \{(\lambda_i)_{i=1}^d, \{\psi_i\}_{i=0}^d\} \in \mathcal{CD}(\rho)\right\}. \tag{3.8}$$

While the OPSE of $T_{f,d}$ will yield $T_{f,d}(\rho) = 0$ if we allow infinitely long decompositions as considered in remark (2.4.4), $T_{f,d}$ can still be useful in trying to determine whether a state is entangled as long as we consider decompositions of finite length. For if we let $T_{f,d}$ take the place of the entanglement measures in the variational algorithm discussed above, then we are not only able to determine the existence of entanglement, we can also prove the existence of barren plateaus. Instead of writing $T_{f,d}$, we will suppress the length d and assume that all ensembles have some fixed length d for the rest of this work. Continuing in this direction define the new cost function for the algorithm to be $T_f(\rho; \theta_1, \ldots, \theta_L)$ which is given by

$$T_f(\rho;\theta_1,\ldots,\theta_L) = \sum_i q_i^2(\theta_1,\ldots,\theta_L) T_2(\varphi_i(\theta_1,\ldots,\theta_L)). \tag{3.9}$$

To simplify notation somewhat, we define Φ_i by

$$\begin{aligned}\Phi_i(\theta_1,\ldots,\theta_L) &= q_i(\theta_1,\ldots,\theta_L) \left|\varphi_i(\theta_1,\ldots,\varphi_L)\right\rangle \left\langle\varphi_i(\theta_1,\ldots,\varphi_L)\right| \\ &= \operatorname{Tr}_{\mathcal{W}}\left(\left(I_{\mathcal{H}} \otimes U^* \left|i\right\rangle \left\langle i\right| U\right) \left|\psi\right\rangle \left\langle\psi\right|\right)\end{aligned} \tag{3.10}$$

For readability, we will usually suppress the parameters $(\theta_1,\ldots,\theta_L)$ and simply write Φ_i. This notation allows us to write T_f succinctly as

$$\sum_i \left(\left(\operatorname{Tr}\operatorname{Tr}_{\mathcal{A}}\Phi_i\right)^2 - \operatorname{Tr}\left(\operatorname{Tr}_{\mathcal{A}}\Phi_i^2\right)\right) = \sum_i \left(q_i^2 - \operatorname{Tr}\left(\operatorname{Tr}_{\mathcal{A}}\Phi_i^2\right)\right). \tag{3.11}$$

To prove the non-negativity of the above quantity, we only need to show that $\left(\operatorname{Tr}(P)\right)^2 \geq \operatorname{Tr}(P^2)$ for all positive operators P. Let $P = \sum_i \lambda_i \Pi_i$ be the spectral decomposition of P. Then all eigenvalues λ_i are non-negative and the Π_i are mutually orthogonal projections with $\operatorname{Tr}\Pi_i = 1$ for each i. Thus

$$\left(\operatorname{Tr}(P)\right)^2 = \sum_{i,j}\lambda_i\lambda_j = \sum_i \lambda_i^2 + \sum_{i\neq j}\lambda_i\lambda_j \geq \sum_i \lambda_i^2 = \operatorname{Tr}(P^2). \tag{3.12}$$

Faithfulness can then be seen via proposition (2.4.6). It turns out that the training landscape of $T_f(\rho;\theta_1,\ldots,\theta_L)$ will indeed exhibit barren plateaus, which is interesting because this demonstrates that cost functions that are non linear in the unitaries U and U^* can have barren plateaus despite being defined by a sum containing exponentially many terms.

3.2 The Existence of Barren Plateaus

Theorem 3.2.1. *For each $j \in \{1,\ldots,L\}$, Var$[\frac{\partial}{\partial_{\theta_j}}T_f] \sim \mathcal{O}(\frac{1}{2^k})$ where the variance is taken with respect to the Haar measure on the unitary group and k is the number of qubits.*

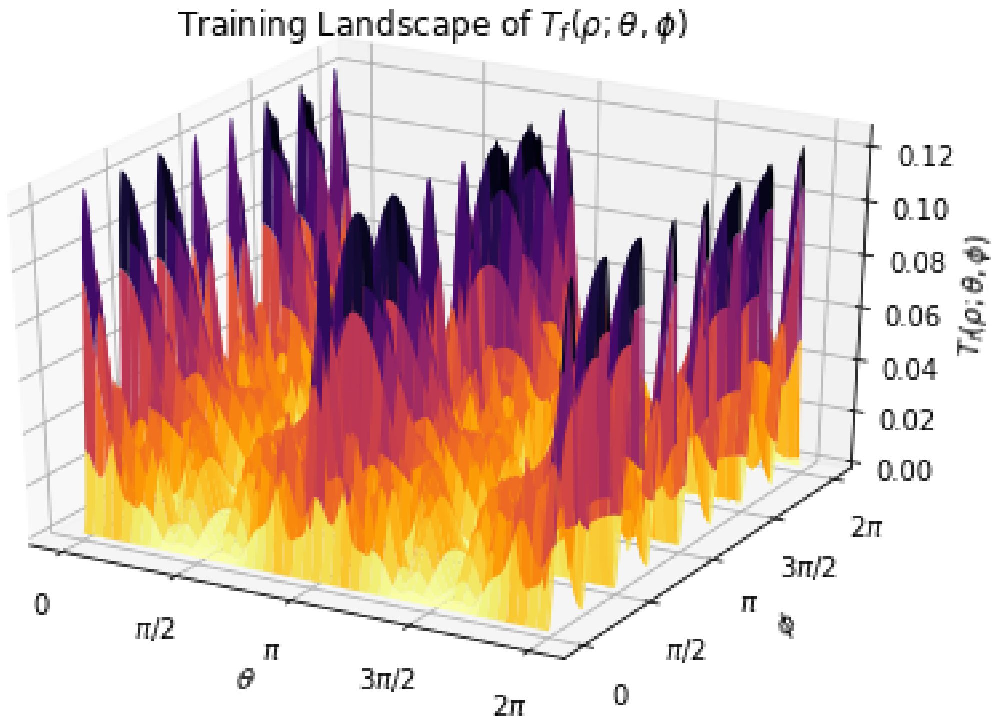


Figure 3.2: Training landscape of $T_f(\rho;\theta,\phi)$ where ρ is the maximally mixed state and $\theta,\phi\in[0,2\pi]$.

In order to prove this result, we'll need the following results on integrals over the unitary group [53, 9]:

Lemma 3.2.2. *For any strings of indices $i_1,\dots,i_p,j_1,\dots,j_p$ and $i'_1,\dots,i'_q,j'_1,\dots,j'_q$ whose values range from 1 to d*

$$\int_{\mathcal{U}(d)} U_{i_1,j_1}\cdots U_{i_p,j_p}\overline{U_{i'_1,j'_1}}\cdots\overline{U_{i'_q,j'_q}}dU=\delta_{p,q}\cdot\sum_{\sigma,\tau\in S_p}Wg(\sigma\tau^{-1})\delta_{i_1,i'_{\sigma(1)}}\delta_{j_1,j'_{\tau(1)}}\cdots\delta_{i_p,i'_{\sigma(p)}}\delta_{j_p,j'_{\tau(p)}} \tag{3.13}$$

where dU is the Haar-Measure on the Unitary group and Wg is the Weingarten function [33, 13].

While notation can make this lemma difficult to parse at first glance, notice that the integral is 0 when the number of complex conjugate entries of the matrix is different the number of non-conjugated entries. Unfortunately, the integrals used in this work aren't lucky enough to fall into this category and we must instead use the expression on the right hand side of the equation to evaluate our integrals. Since this expression can be difficult to understand at first, we supply a simple example of computing such an integral via this lemma.

Example 3.2.3. *Let U, V be unitary matrices on $\mathbb{C}^d$. Then the average gate fidelity between U and V, $F_{avg}(U, V)$, is defined by*

$$F_{avg}(U,V) := \int_{S(\mathcal{H})} |\langle\psi| U^*V |\psi\rangle|^2 d\psi \tag{3.14}$$

where $S(\mathcal{H})$ is unit ball of $\mathcal{H}$.

Such an integral over pure states is defined by

$$\int_{S(\mathcal{H})} |\langle\psi| U^*V |\psi\rangle|^2 d\psi := \int_{\mathcal{U}(d)} |\langle\psi_0| W^*U^*VW |\psi_0\rangle|^2 dW \tag{3.15}$$

where each ψ is replaced by $W|\psi_0\rangle$ for some arbitrary state $|\psi_0\rangle \in \mathcal{H}$. Now let $\{|i\rangle : 1 \leq i \leq d\}$ be a basis for $\mathbb{C}^d$ and without loss of generality, set $|\psi_0\rangle := |1\rangle$. Next write $H = U^*V$ and suppose that H has spectral decomposition

$$H = \sum_i \sigma_i P^* |i\rangle \langle i| P \tag{3.16}$$

where $\sigma_i \in \mathbb{C}$ with $|\sigma_i| = 1$ for each $1 \leq i \leq d$ and P is some unitary on $\mathcal{H}$. Using these substitutions, the integral in Equation(3.15) transforms as

$$\begin{aligned} &\int_{\mathcal{U}(d)} |\langle\psi_0| W^*U^*VW |\psi_0\rangle|^2 dW \\ &= \int_{\mathcal{U}(d)} \langle 1| W^*HW |1\rangle \overline{\langle 1| W^*HW |1\rangle} dW \\ &= \sum_{i,j} \sigma_i \overline{\sigma}_j \int_{\mathcal{U}(d)} \langle 1| W^*P^* |i\rangle \langle i| PW |1\rangle \overline{\langle 1| W^*P^* |j\rangle \langle j| PW |1\rangle} dW \end{aligned} \tag{3.17}$$

Now using the translation invariance of the Haar measure on $\mathcal{U}(d)$, we use a change of variable to simplify the above integral. Set $Q = PW$ so that $dQ = d(PW) = dW$.

Therefore

$$\begin{aligned}
&\sum_{i,j} \sigma_i \overline{\sigma_j} \int_{\mathcal{U}(d)} \langle 1| W^* P^* |i\rangle \langle i| PW |1\rangle \overline{\langle 1| W^* P^* |j\rangle \langle j| PW |1\rangle} dW \\
&= \sum_{i,j} \sigma_i \overline{\sigma_j} \int_{\mathcal{U}(d)} \langle 1| Q^* |i\rangle \langle i| Q |1\rangle \overline{\langle 1| Q^* |j\rangle \langle j| Q |1\rangle} dQ \\
&= \sum_{i,j} \sigma_i \overline{\sigma_j} \int_{\mathcal{U}(d)} Q_{i,1} Q_{j,1} \overline{Q_{i,1}}\ \overline{Q_{j,1}} dQ
\end{aligned} \tag{3.18}$$

We can now use the integration formula given in (3.2.2) to evaluate the integral in the last expression of the previous equation and we arrive at

$$\begin{aligned}
&\sum_{i,j} \sigma_i \overline{\sigma_j} \int_{\mathcal{U}(d)} Q_{i,1} Q_{j,1} \overline{Q_{i,1}}\ \overline{Q_{j,1}} dQ \\
&= \sum_{i,j} \sigma_i \overline{\sigma_j} \Big(\mathrm{Wg}((1))(1 + \delta_{i,j}) + \mathrm{Wg}((1,2))(1 + \delta_{i,j}) \Big) \\
&= \frac{1}{d(d+1)} \sum_{i,j} \sigma_i \overline{\sigma_j} (1 + \delta_{i,j}).
\end{aligned} \tag{3.19}$$

where $\mathrm{Wg}\big((1)\big) = \frac{1}{d^2-1}$ and $\mathrm{Wg}\big((1,2)\big) = \frac{-1}{d(d^2-1)}$. Since H is a unitary, each σ_i has absolute value 1, and so

$$\sum_{i,j} \sigma_i \overline{\sigma_j} \delta_{i,j} = \sum_i |\sigma_i|^2 = d. \tag{3.20}$$

To compute the other sum, we use the fact that the sum of the eignvalues of a matrix is its trace, so that

$$\sum_{i,j} \sigma_i \overline{\sigma_j} = \Big(\sum_i \sigma_i \Big) \Big(\sum_j \overline{\sigma_j} \Big) = |\mathrm{Tr}\,(H)|^2. \tag{3.21}$$

Recalling that $H = U^* V$, we can see that $\mathrm{Tr}\,(H) = \langle U, V \rangle_{HS}$ where the inner product is taken to be the Hilbert-Schmidt inner product. Thus it follows that

$$F_{avg}(U, V) = \frac{|\langle U, V \rangle|^2 + d}{d(d+1)} \tag{3.22}$$

We can see that in order to compute these integrals, we must be familiar with the Weingarten function on the unitary group. In general, the Weingarten function can be difficult to compute, and we therefore include a table of values for the Weingarten function for the first four permutation groups as we'll need these values in the computation of the variance of concern. In the tables, we list the cycle structure of a given permutation σ in the left hand column and the the value of $\mathrm{Wg}(\sigma)$ in the right hand column. These values were computed using the tool [14] written by M. Fukuda et alii.

Table 3.1: Table of Values for the Weingarten Function

Cycle type of σ	$\mathrm{Wg}(\sigma)$
1	$\frac{1}{d}$
1,1	$\frac{1}{d^2-1}$
2	$\frac{-1}{d(d^2-1)}$
1,1,1	$\frac{d^2-2}{d(d^4-5d^2+4)}$
2,1	$\frac{-1}{d^4-5d^2+4}$
3	$\frac{2}{d(d^4-5d^2+4)}$
1,1,1,1	$\frac{d^4-8d^2+6}{d^2(d^6-14d^4+49d^2-36)}$
2,1,1	$\frac{-1}{d(d^4-10d^2+9)}$
2,2	$\frac{d^2+6}{d^2(d^6-14d^4+49d^2-36)}$
3,1	$\frac{2d^2-3}{d^2(d^6-14d^4+49d^2-36)}$
4	$\frac{-5}{d(d^6-14d^4+49d^2-36)}$

Notice that the value of the Weingarten function of a given permutation is dependent only on the cycle structure of a permutation. For example, let $\sigma = (12)(34) \in \mathrm{S}_4$. This permutation has cycle structure 2,2 and so $\mathrm{Wg}(\sigma) = \frac{d^2+6}{d^2(d^6-14d^4+49d^2-36)}$. Now before computing the mean and variance of the derivatives of the cost function $T_f(\rho; \theta_1, \ldots, \theta_L)$, we must first discuss the probability distribution of unitaries produced by the circuits defined in Equation(3.5).

In order to make the notation more readable, suppose that $\theta = \theta_j$ for some $j \in \{1, \ldots, L\}$ and that $V = V_j$. Then the derivative of U with respect to θ can be written as

$$\frac{\partial}{\partial\theta}U(\theta) = i\Big(\prod_{l_1=1}^{j} \exp(i\theta_{l_1}V_{l_1})E_{l_1}\Big)V\Big(\prod_{l_2=j+1}^{L} \exp(i\theta_{l_2}V_{l_2})E_{l_2}\Big) \tag{3.23}$$

The derivative splits the circuit into two “pieces", the left side and the right side which we define accordingly

$$L = \prod_{l_1=1}^{j} \exp(i\theta_{l_1} V_{l_1}) E_{l_1} \text{ and } R = \prod_{l_2=j+1}^{L} \exp(i\theta_{l_2} V_{l_2}) E_{l_2} \tag{3.24}$$

which lets us write $\partial_\theta U(\theta) = iLVR$. Moreover, using this expression for the derivative of U lets us express $\partial_\theta U^* |i\rangle \langle i| U$ as

$$\frac{\partial}{\partial\theta} U^* |i\rangle \langle i| U = \sqrt{-1} R^* \Big[V, L^* |i\rangle \langle i| L \Big] R \tag{3.25}$$

We will often write $K_i = \sqrt{-1}\Big[V, L^* |i\rangle \langle i| L\Big]$ for readability so that

$$\frac{\partial}{\partial\theta} U^* |i\rangle \langle i| U = R^* K_i R \tag{3.26}$$

This splitting forces us to consider the distributions of the L and R separately. Therefore, just as in the paper [35] by McClean et alii, we write the probability distribution generated by the entire circuit as

$$\rho(U)dU = \delta(U - LR)\rho_1(L)\rho_2(R)dLdR \tag{3.27}$$

where dU, dL, and dR are the Haar measure on $\mathcal{U}(d)$, ρ_1 and ρ_2 are densities on $\mathcal{U}(d)$, and δ is the dirac measure centered at the $0-$matrix. This distribution forces any unitary U to be split as the product of two other unitaries L and R with their own distributions ρ_1 and ρ_2 respectively. To quantify these distribution we'll need to use the notion of *unitary k-designs*. A distribution P is defined to be a k-design if P matches the Haar measure on the unitary group up to and including the k-th moment [24, 4, 53, 9]. In other words,

$$\int_{\mathcal{U}(d)} U^{\otimes l} \otimes (U^*)^{\otimes l} P(U) d\mu(U) = \int_{\mathcal{U}(d)} U^{\otimes l} \otimes (U^*)^{\otimes l} d\mu(U) \tag{3.28}$$

for all $0 \leq l \leq k$. In practice, exact $k-$designs are costly to produce. And so we'll also need the notion of an $\varepsilon-$approximate $k-$design [4]. We say that a a measure ν

is an $\varepsilon-$approximate $k-$design if and only if

$$(1-\varepsilon)\int_{\mathcal{U}(d)} U^{\otimes k} X (U^*)^{\otimes k} dU \leq \int_{\mathcal{U}(d)} U^{\otimes k} X (U^*)^{\otimes k} d\nu(U) \leq (1+\varepsilon)\int_{\mathcal{U}(d)} U^{\otimes k} X (U^*)^{\otimes k} dU \tag{3.29}$$

for any $X \in \mathrm{L}\left(\mathbb{C}^{d\otimes k}\right)$, where $\varepsilon \in (0,1)$ and dU is the Haar measure on $\mathcal{U}(d)$. Such measures ν are attainable in practice using only polynomially many gates [4].

With the distribution ρ from Equation (3.27) in mind, we will split the computation of the mean and variance of $\partial_\theta T_f(\rho;\theta)$ into two cases; One in which we assume that ρ_1 is a 4-design, and the other in which ρ_2 is a 4-design. Note that this assumption is a practical one since it has been proven [24] that polynomial depth quantum circuits such as those given in Equation (3.5) are $\epsilon-$approximate k-designs, if they are of sufficient depth.

Next let's expand the various components of $\frac{\partial}{\partial\theta}T_f(\rho;\theta)$ using the definitions of q_i and Φ_i, given in Equations (3.3) and (3.10) respectively, for each $i \in \{1,\ldots,d\}$.

$$\begin{aligned} q_i &= \langle\psi|\left(I_{\mathcal{AB}} \otimes U^* |i\rangle\langle i| U\right)|\psi\rangle \\ &= \left(\sum_{j_1'} \sqrt{p_{j_1'}}\left\langle\psi_{j_1'}\right| \otimes \langle j_1'|\right)\left(I_{\mathcal{AB}} \otimes U^* |i\rangle\langle i| U\right)\left(\sum_{j_1}\sqrt{p_{j_1}}\,|\psi_{j_1}\rangle \otimes |j_1\rangle\right) \\ &= \sum_{j_1,j_1'} \sqrt{p_{j_1}p_{j_1'}}\left\langle\psi_{j_1'}\middle|\psi_{j_1}\right\rangle \langle j_1'| U^* |i\rangle\langle i| U |j_1\rangle \\ &= \sum_{j_1,j_1'} \sqrt{p_{j_1}p_{j_1'}}\left\langle\psi_{j_1'}\middle|\psi_{j_1}\right\rangle \langle i| U |j_1\rangle \overline{\langle i| U |j_1'\rangle} \\ &= \sum_{j_1,j_1'} \sqrt{p_{j_1}p_{j_1'}}\left\langle\psi_{j_1'}\middle|\psi_{j_1}\right\rangle \langle i| LR |j_1\rangle \overline{\langle i| LR |j_1'\rangle} \end{aligned} \tag{3.30}$$

It can be shown in a similar fashion that

$$\mathrm{Tr}_{\mathcal{A}}\Phi_i = \sum_{j_1,j_1'} \sqrt{p_{j_1}p_{j_1'}}\ \mathrm{Tr}_{\mathcal{A}}\,|\psi_{j_1}\rangle\left\langle\psi_{j_1'}\right| \langle i| LR |j_1\rangle \overline{\langle i| LR |j_1'\rangle}. \tag{3.31}$$

Using the commutator expression given in Equation (3.25) for the derivative of q_i

as follows:

$$\begin{aligned}\frac{\partial q_i}{\partial \theta} &= \langle\psi| \left(I_{\mathcal{AB}} \otimes \frac{\partial}{\partial\theta} U^* |i\rangle \langle i| U\right) |\psi\rangle \\ &= \langle\psi| \left(I_{\mathcal{AB}} \otimes R^* K_i R\right) |\psi\rangle \\ &= \sum_{j_1, j_1'} \sqrt{p_{j_1} p_{j_1'}} \left\langle \psi_{j_1'} \middle| \psi_{j_1} \right\rangle \langle j_1'| R^* K_i R |j_1\rangle\end{aligned} \tag{3.32}$$

In the same way, a similar expression is found for $\frac{\partial}{\partial\theta}\mathrm{Tr}_{\mathcal{A}}\Phi_i$:

$$\frac{\partial \mathrm{Tr}_{\mathcal{A}}\Phi_i}{\partial\theta} = \sum_{j_1, j_1'} \sqrt{p_{j_1} p_{j_1'}}\ \mathrm{Tr}_{\mathcal{A}} |\psi_{j_1}\rangle \left\langle \psi_{j_1'} \right| \langle j_1'| R^* K_i R |j_1\rangle \tag{3.33}$$

Putting all of this together we see that

$$\begin{aligned}\frac{\partial}{\partial\theta} q_i^2 &= 2q_i \frac{\partial q_i}{\partial\theta} \\ &= \sum_{j_1, j_2, j_1', j_2'} \sqrt{p_{j_1} p_{j_1'} p_{j_2} p_{j_2'}} \left\langle \psi_{j_1'} \middle| \psi_{j_1} \right\rangle \left\langle \psi_{j_2'} \middle| \psi_{j_2} \right\rangle \langle i| LR |j_1\rangle \overline{\langle i| LR |j_1'\rangle} \langle j_2'| R^* K_i R |j_2\rangle\end{aligned} \tag{3.34}$$

And again, $\frac{\partial}{\partial\theta}\mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{A}}\Phi_i^2\right)$ has a similar expression:

$$\frac{\partial}{\partial\theta}\mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{A}}\Phi_i^2\right) = 2\mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{A}}\Phi_i \cdot \frac{\partial \mathrm{Tr}_{\mathcal{A}}\Phi_i}{\partial\theta}\right) \tag{3.35}$$

which can be seen to be equal to

$$2 \sum_{j_1, j_2, j_1', j_2'} \sqrt{p_{j_1} p_{j_1'} p_{j_2} p_{j_2'}}\ \mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{A}} |\psi_{j_1}\rangle \langle \psi_{j_1'}| \mathrm{Tr}_{\mathcal{A}} |\psi_{j_2}\rangle \langle \psi_{j_2'}| \right) \langle i| LR |j_1\rangle \overline{\langle i| LR |j_1'\rangle} \langle j_2'| R^* K_i R |j_2\rangle \tag{3.36}$$

Putting all of this together, we can expand $\frac{\partial}{\partial\theta}T_f(\rho;\theta)$ as

$$\frac{\partial}{\partial\theta}T_f(\rho;\theta) = 2\sum_i \left(q_i \frac{\partial q_i}{\partial\theta} - \mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{A}}\Phi_i \cdot \frac{\partial \mathrm{Tr}_{\mathcal{A}}\Phi_i}{\partial\theta}\right)\right), \tag{3.37}$$

and then integrate this expression term by term using the expressions above. In what follows, we'll fix an i and integrate the terms $q_i \frac{\partial q_i}{\partial\theta}$ and $\mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{A}}\Phi_i \cdot \frac{\partial \mathrm{Tr}_{\mathcal{A}}\Phi_i}{\partial\theta}\right)$ separately for clarity. After the computation of the integrals, we'll then sum the results to arrive at the appropriate expected value.

Now notice that the expressions given in equations (3.34) and (3.36) are of the form

$$2 \sum_{j_1,j_2,j_1',j_2'} \sqrt{p_{j_1}p_{j_1'}p_{j_2}p_{j_2'}}c_{j_1,j_1',j_2,j_2'} \langle i| LR |j_1\rangle \overline{\langle i| LR |j_1'\rangle} \langle j_2'| R^*K_iR |j_2\rangle \tag{3.38}$$

where $|c_{j_1,j_1',j_2,j_2'}| \leq 1$ for any choice of indices. Because we only need an upper estimate for our variance, we'll soon see that we don't need the exact values of the $c_{j_1,j_1',j_2,j_2'}$ and that we'll only use the fact that they are bounded in absolute value by 1. Next we need to integrate the above expression with respect to the distribution $\rho_1(L)\rho_2(R)dLdR$. Since only the expression $\langle i| LR |j_1\rangle \overline{\langle i| LR |j_1'\rangle} \langle j_2'| R^*K_iR |j_2\rangle$ is dependent on L and R, we can integrate this expression and substitute its value back into Equation (3.38).

In the integration formula given in Lemma 3.2.2, the expressions that were integrated were given in terms of the coordinates of the unitaries, but the expression we currently have is not in this form. We must therefore force the expression into this form and expand the commutator R^*K_iR. It's easy to see that

$$\langle i| LR |j_1\rangle \overline{\langle i| LR |j_1'\rangle} \langle j_2'| R^*K_iR |j_2\rangle \tag{3.39}$$

is, up to a multiple of $\sqrt{-1}$, equal to

$$\langle i| LR |j_1\rangle \overline{\langle i| LR |j_1'\rangle} \cdot \Big(\langle j_2'| R^*VL^* |i\rangle \langle i| LR |j_2\rangle - \langle j_2'| R^*L^* |i\rangle \langle i| LVR |j_2\rangle \Big). \tag{3.40}$$

For our computations, we'll just use the first term, $\langle j_2'| R^*VL^* |i\rangle \langle i| LR |j_2\rangle$, of the commutator since the computations with the second term are nearly identical. Suppose for now that $\rho_1(L)dL$ is at least a 2-design, then $\rho_1(L)dL$ is equal to the Haar-distribution dL up to the second moment. Thus

$$\begin{aligned} &\int\int \langle i| LR |j_1\rangle \overline{\langle i| LR |j_1'\rangle} \langle j_2'| R^*VL^* |i\rangle \langle i| LR |j_2\rangle \rho_1(L)\rho_2(R)dLdR \\ &= \int\int \langle i| LR |j_1\rangle \langle i| LR |j_2\rangle \overline{\langle i| LR |j_1'\rangle}\ \overline{\langle i| LVR |j_2'\rangle}\rho_2(R)dLdR \end{aligned} \tag{3.41}$$

Using the translation invariance of the Haar-integral on the Unitary group, we define the change of variables $W = LR$ with $dW = dL$. The integral then becomes

$$\begin{aligned}
&\int\int \langle i|\, LR\, |j_1\rangle \langle i|\, LR\, |j_2\rangle \overline{\langle i|\, LR\, |j_1'\rangle}\ \overline{\langle i|\, LVR\, |j_2'\rangle}\rho_2(R)dLdR \\
&= \int\int \langle i|\, W\, |j_1\rangle \langle i|\, W\, |j_2\rangle \overline{\langle i|\, W\, |j_1'\rangle}\ \overline{\langle i|\, WR^*VR\, |j_2'\rangle} dW\rho_2(R)dR \\
&= \int\int W_{i,j_1}W_{i,j_2}\overline{W_{i,j_1'}}\ \overline{\langle i|\, WR^*VR\, |j_2'\rangle} dW\rho_2(R)dR
\end{aligned} \tag{3.42}$$

The last thing that we need to do before using the integration formula is to express $\overline{\langle i|\, WR^*VR\, |j_2'\rangle}$ in terms of the coordinates of W. To do this, first define $Q = R^*VR$ so that

$$\begin{aligned}
R^*VR\, |j_2'\rangle &= Q\, |j_2'\rangle \\
&= \sum_\alpha Q_{\alpha,j_2'}\, |\alpha\rangle\, ,
\end{aligned} \tag{3.43}$$

and thus

$$\begin{aligned}
\overline{\langle i|\, WR^*VR\, |j_2'\rangle} &= \overline{\langle i|\, WQ\, |j_2'\rangle} \\
&= \sum_\alpha \overline{Q_{\alpha,j_2'}}\ \overline{\langle i|\, W\, |\alpha\rangle}.
\end{aligned} \tag{3.44}$$

Substituting this expression back into the integral we see that

$$\begin{aligned}
&\int\int W_{i,j_1}W_{i,j_2}\overline{W_{i,j_1'}}\ \overline{\langle i|\, WR^*VR\, |j_2'\rangle} dW\rho_2(R)dR \\
&= \sum_\alpha \int \overline{Q_{\alpha,j_2'}} \int W_{i,j_1}W_{i,j_2}\overline{W_{i,j_1'}}\ \overline{\langle i|\, W\, |\alpha\rangle} dW\rho_2(R)dR \\
&= \sum_\alpha \int \overline{Q_{\alpha,j_2'}} \int W_{i,j_1}W_{i,j_2}\overline{W_{i,j_1'}}\ \overline{W_{i,\alpha}} dW\rho_2(R)dR.
\end{aligned} \tag{3.45}$$

Notice that all of the row coordinates of the W's in the integral are equal to i, so we can suppress the $\delta_{i_k,i'_{\sigma(k)}}$ terms in the integration formula from Lemma (3.2.2) since they are always equal to 1. The index α also took the place of j_2' in the column coordinate of one of the Ws in the last expression of the above equation.Therefore the α must be treated as a j_k' when using the integration formula from Lemma (3.2.2).

For clarity, we isolate the inner integral and compute it as follows:

$$\begin{aligned}
&\int W_{i,j_1} W_{i,j_2} \overline{W_{i,j_1'}}\ \overline{W_{i,\alpha}} dW \\
&= \mathrm{Wg}\Big((1)_2\Big)\delta_{j_1,j_1'}\delta_{j_2,\alpha} \qquad\qquad (\sigma = \tau = (1)_2) \\
&+ \mathrm{Wg}\Big((1,2)_2\Big)\delta_{j_1,j_1'}\delta_{j_2,\alpha} \qquad\qquad (\sigma = (1,2)_2, \tau = (1)_2) \\
&+ \mathrm{Wg}\Big((1,2)_2\Big)\delta_{j_1,\alpha}\delta_{j_2,j_1'} \qquad\qquad (\sigma = (1)_2, \tau = (1,2)_2) \\
&+ \mathrm{Wg}\Big((1)_2\Big)\delta_{j_1,\alpha}\delta_{j_2,j_1'} \qquad\qquad (\sigma = \tau = (1,2)_2) \\
&= \Big(\mathrm{Wg}(1,1) + \mathrm{Wg}(2)\Big)\Big(\delta_{j_1,j_1'}\delta_{j_2,\alpha} + \delta_{j_1,\alpha}\delta_{j_2,j_1'}\Big).
\end{aligned} \tag{3.46}$$

Defining $C := \Big(\mathrm{Wg}(1,1) + \mathrm{Wg}(2)\Big) = \frac{1}{d(d+1)}$ and substituting the last expression of the above equation back into Equation (3.45), we see that the value of the integral is equal to

$$\begin{aligned}
&C \sum_{\alpha} \int \overline{Q_{\alpha,j_2'}} \Big(\delta_{j_1,j_1'}\delta_{j_2,\alpha} + \delta_{j_1,\alpha}\delta_{j_2,j_1'}\Big)\rho_2(R)dR \\
&= C\int \overline{Q_{j_2 j_2'}}\delta_{j_1,j_1'}\rho_2(R)dR + C\int \overline{Q_{j_1,j_2'}}\delta_{j_2,j_1'}\rho_2(R)dR.
\end{aligned} \tag{3.47}$$

Ignoring the integral and outside constants for the moment, let's plug the expressions $\overline{Q_{j_2 j_2'}}\delta_{j_1,j_1'}$ and $\overline{Q_{j_1,j_2'}}\delta_{j_2,j_1'}$ into Equation (3.38) and try to bound the sum in absolute value. Substituting $\overline{Q_{j_2,j_2'}}\delta_{j_1,j_1'}$, we see that

$$\sum_{j_1,j_2,j_1',j_2'} \sqrt{p_{j_1}p_{j_1'}p_{j_2}p_{j_2'}} c_{j_1,j_1',j_2,j_2'} \overline{Q_{j_2,j_2'}}\delta_{j_1,j_1'} = \sum_{j_1,j_2,j_2'} p_{j_1}\sqrt{p_{j_2}p_{j_2'}} c_{j_1,j_1',j_2,j_2'} \overline{Q_{j_2,j_2'}} \tag{3.48}$$

Now taking absolute values,

$$\begin{aligned}
\Big|\sum_{j_1,j_2,j_2'} p_{j_1}\sqrt{p_{j_2}p_{j_2'}} c_{j_1,j_1',j_2,j_2'} \overline{Q_{j_2,j_2'}}\Big| &\le \sum_{j_1,j_2,j_2'} p_{j_1}\sqrt{p_{j_2}p_{j_2'}}\Big|c_{j_1,j_1',j_2,j_2'}\Big| \cdot \Big|Q_{j_2,j_2'}\Big| \\
&\le \sum_{j_1,j_2,j_2'} p_{j_1}\sqrt{p_{j_2}p_{j_2'}}\Big|Q_{j_2,j_2'}\Big| \\
&= \left(\sum_{j_1} p_{j_1}\right)\cdot\left(\sum_{j_2,j_2'}\sqrt{p_{j_2}p_{j_2'}}\Big|Q_{j_2,j_2'}\Big|\right) \\
&= \sum_{j_2,j_2'}\sqrt{p_{j_2}p_{j_2'}}\Big|Q_{j_2,j_2'}\Big|.
\end{aligned} \tag{3.49}$$

Lastly using the Cauchy-Schwartz inequality and the fact that the Hilbert-Schmidt norm of a $d \times d$ unitary matrix is equal to $\sqrt{d}$, we can see that

$$\begin{aligned}\sum_{j_2,j_2'} \sqrt{p_{j_2}p_{j_2'}}\left|Q_{j_2,j_2'}\right| &\leq \left(\sum_{j_2,j_2'} p_{j_2}p_{j_2'}\right)^{\frac{1}{2}} \left(\sum_{j_2,j_2'} \left|Q_{j_2,j_2'}\right|^2\right)^{\frac{1}{2}} \\ &= \|Q\|_{HS} \\ &= \sqrt{d}.\end{aligned} \tag{3.50}$$

The computation when substituting the term $\overline{Q_{j_1,j_2'}}\delta_{j_2,j_1'}$ is almost exactly the same so we omit it for brevity. We have thus showed that when $\rho_1(L)dL$ is at least a 2 design and $\rho_2(R)dR$ is an arbitrary distribution, the integral of Equation (3.38) is bounded in absolute value as

$$\begin{aligned}&\left|\int\int 2\sum_{j_1,j_2,j_1',j_2'} \sqrt{p_{j_1}p_{j_1'}p_{j_2}p_{j_2'}}c_{j_1,j_1',j_2,j_2'} \langle i| LR |j_1\rangle \overline{\langle i| LR |j_1'\rangle} \langle j_2'| R^* K_i R |j_2\rangle \rho_1(L)dL\rho_2(R)dR\right| \\ &\leq \int 2C\sqrt{d}\rho_2(R)dR \\ &= \frac{2\sqrt{d}}{d(d+1)}\end{aligned} \tag{3.51}$$

And since both $\mathbb{E}\left[2q_i\frac{\partial q_i}{\partial\theta}\right]$ and $\mathbb{E}\left[\frac{\partial}{\partial\theta}\text{Tr}\left(\text{Tr}_{\mathcal{A}}\Phi_i^2\right)\right]$ are both bounded by the above integral, it follows that

$$\mathbb{E}\left[2q_i\frac{\partial q_i}{\partial\theta}\right] \sim \mathcal{O}\left(\frac{1}{d^{\frac{3}{2}}}\right) \text{ and } \mathbb{E}\left[\frac{\partial}{\partial\theta}\text{Tr}\left(\text{Tr}_{\mathcal{A}}\Phi_i^2\right)\right] \sim \mathcal{O}\left(\frac{1}{d^{\frac{3}{2}}}\right). \tag{3.52}$$

Therefore

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}T_f(\rho;\theta)\right] = \sum_i \left(\mathbb{E}\left[2q_i\frac{\partial q_i}{\partial\theta}\right] - \mathbb{E}\left[\frac{\partial}{\partial\theta}\text{Tr}\left(\text{Tr}_{\mathcal{A}}\Phi_i^2\right)\right]\right) \sim \mathcal{O}\left(\frac{1}{\sqrt{d}}\right). \tag{3.53}$$

This shows that the expected gradients converge exponentially to 0 in the number of qubits when L is at least a 2-design and the distribution of R is arbitrary.

Suppose now that the distribution of R is at least a 2-design and the the distribution of L is arbitrary. We will make almost the same change of variables as before with $W = LR$ and $dW = dR$. However we now define $Q = LVL^*$ so that

$$\begin{aligned}\overline{\langle i|\,LVR\,|j_2'\rangle} &= \overline{\langle i|\,QW\,|j_2'\rangle}\\ &= \sum_\alpha \overline{Q_{i,\alpha}}\ \overline{W_{\alpha,j_2'}}\end{aligned} \tag{3.54}$$

Now using the same steps as in the first case in Equations (3.43) and (3.44), we arrive at the integral

$$\begin{aligned}&\sum_\alpha \int \overline{Q_{i,\alpha}} \int W_{i,j_1}W_{i,j_2}\overline{W_{i,j_1'}}\ \overline{W_{\alpha,j_2'}}dW\rho_1(L)dL\\ &= \sum_\alpha \int \overline{Q_{i,\alpha}}C\delta_{i,\alpha}\Big(\delta_{j_1,j_1'}\delta_{j_2,j_2'}+\delta_{j_1,j_2'}\delta_{j_2,j_1'}\Big)\rho_1(L)dL\\ &= C\int \overline{Q_{i,i}}\Big(\delta_{j_1,j_1'}\delta_{j_2,j_2'}+\delta_{j_1,j_2'}\delta_{j_2,j_1'}\Big)\rho_1(L)dL\end{aligned} \tag{3.55}$$

And again we now substitute one of the expressions containing deltas into Equation (3.38) and bound the expression.

$$\sum_{j_1,j_2,j_1',j_2'} \sqrt{p_{j_1}p_{j_1'}p_{j_2}p_{j_2'}}c_{j_1,j_1',j_2,j_2'}\overline{Q_{i,i}}\delta_{j_1,j_1'}\delta_{j_2,j_2'} = \sum_{j_1,j_2} p_{j_1}p_{j_2}c_{j_1,j_1,j_2,j_2}\overline{Q_{i,i}} \tag{3.56}$$

Since both the c_{j_1,j_1,j_2,j_2} and $\overline{Q_{i,i}}$ are bounded in absolute value by 1, we see that

$$\begin{aligned}&\left|\int\int 2\sum_{j_1,j_2,j_1',j_2'} \sqrt{p_{j_1}p_{j_1'}p_{j_2}p_{j_2'}}c_{j_1,j_1',j_2,j_2'}\langle i|\,LR\,|j_1\rangle\,\overline{\langle i|\,LR\,|j_1'\rangle}\,\langle j_2'|\,R^*K_iR\,|j_2\rangle\,\rho_1(L)dL\rho_2(R)dR\right|\\ &= \left|\sum_i 2C\sum_{j_1,j_2} p_{j_1}p_{j_2}c_{j_1,j_1,j_2,j_2}\overline{Q_{i,i}}\right|\\ &\le 2C\sum_i\sum_{j_1,j_2} p_{j_1}p_{j_2}\big|c_{j_1,j_1,j_2,j_2}\big|\cdot\big|Q_{i,i}\big|\\ &\le 2C\sum_i\sum_{j_1,j_2} p_{j_1}p_{j_2}\\ &= 2C\sum_i 1\\ &= \frac{2}{d(d+1)}\cdot d\\ &= \frac{2}{d+1}.\end{aligned} \tag{3.57}$$

Now we the same steps as in Equations (3.52) and (3.53) to compute the expected value again. Thus in the case when $\rho_2(R)dR$ is at least a 2-design and $\rho_1(L)dL$ is given by an arbitrary distribution, we see that

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}T_f(\rho;\theta)\right] \sim \mathcal{O}\left(\frac{1}{d}\right) \tag{3.58}$$

We now need to compute the second moment of the $\frac{\partial}{\partial\theta}T_f(\rho;\theta)$. We will accomplish this by squaring the expression in Equation (3.37) and integrating term by term just as we did with mean. In order to compute these integrals, we need stronger,but still modest, assumptions on the distributions of ρ_1 amd ρ_2. Namely we must split into cases when ρ_1 is a 4-design and ρ_2 is not, and vice versa. Note that such a circuit is indeed possible in practice and only needs to be of polynomial depth [4].

First let's square the sum in Equation (3.37) to understand what we're working with.

$$\begin{aligned}\left(\frac{\partial}{\partial\theta}T_f(\rho;\theta)\right)^2 &= 4\left(\sum_i\left(q_i\frac{\partial q_i}{\partial\theta} - \mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{A}}\Phi_i\cdot\frac{\partial \mathrm{Tr}_{\mathcal{A}}\Phi_i}{\partial\theta}\right)\right)\right)^2\\ &= 4\sum_{i,j}\left(q_i\frac{\partial q_i}{\partial\theta} - \mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{A}}\Phi_i\cdot\frac{\partial \mathrm{Tr}_{\mathcal{A}}\Phi_i}{\partial\theta}\right)\right)\left(q_j\frac{\partial q_j}{\partial\theta} - \mathrm{Tr}\left(\mathrm{Tr}_{\mathcal{A}}\Phi_j\cdot\frac{\partial \mathrm{Tr}_{\mathcal{A}}\Phi_j}{\partial\theta}\right)\right)\end{aligned} \tag{3.59}$$

When multiplying term by term in the last line of the equation above, making substitutions using Equations (3.30) through (3.33) and using the same change of variables $W = LR$, we encounter expressions of the form

$$\sum_{\substack{j_1,j_2,j_1',j_2'\\ j_3,j_4,j_3',j_4'}} \sqrt{p_{j_1}p_{j_1'}p_{j_2}p_{j_2'}p_{j_3}p_{j_3'}p_{j_4}p_{j_4'}}\,c_{j_1,j_1',j_2,j_2'}c_{j_3,j_3',j_4,j_4'}W_{i,j_1}\overline{W_{i,j_1'}}W_{i,j_3}\overline{W_{i,j_3'}}\langle j_2'|R^*K_iR|j_2\rangle\langle j_4'|R^*K_jR|j_4\rangle \tag{3.60}$$

for each each fixed i and j. Just as we did in the computation of the mean, we will fix i and j, compute the appropriate integrals for each term, and sum the results at the end.

Again, we only need to worry about the terms of the above equation that involve only entries of W, L, or R when integrating, and can substitute the values of these integrals back into the Equation. Because there are now two commutators, K_i and

K_j, in the expression, we'll need to break our proof into cases based on which terms of the commutator are multiplied together. To understand the cases better, let's expand fully expand $\langle j_2'| R^*K_iR|j_2\rangle \langle j_4'| R^*K_jR|j_4\rangle$. Using the definitions of K_i and k_j respectively, we can see that

$$\begin{aligned}
&\langle j_2'| R^*K_iR|j_2\rangle \langle j_4'| R^*K_jR|j_4\rangle \\
&= -\langle j_2'| R^*\Big[V, L^*|i\rangle\langle i| L\Big]R|j_2\rangle \langle j_4'| R^*\Big[V, L^*|j\rangle\langle j| L\Big]R|j_4\rangle \\
&= -\langle j_2'| R^*VL^*|i\rangle\langle i| LR|j_2\rangle \langle j_4'| R^*VL^*|j\rangle\langle j| LR|j_4\rangle \\
&+ \langle j_2'| R^*VL^*|i\rangle\langle i| LR|j_2\rangle \langle j_4'| R^*L^*|j\rangle\langle j| LVR|j_4\rangle \\
&+ \langle j_2'| R^*L^*|i\rangle\langle i| LVR|j_2\rangle \langle j_4'| R^*VL^*|j\rangle\langle j| LR|j_4\rangle \\
&- \langle j_2'| R^*L^*|i\rangle\langle i| LVR|j_2\rangle \langle j_4'| R^*L^*|j\rangle\langle j| LVR|j_4\rangle .
\end{aligned} \tag{3.61}$$

This first case we need to consider, *case A*, is when either the first term,

$$-\langle j_2'| R^*VL^*|i\rangle\langle i| LR|j_2\rangle \langle j_4'| R^*VL^*|j\rangle\langle j| LR|j_4\rangle, \tag{3.62}$$

or the last term

$$-\langle j_2'| R^*L^*|i\rangle\langle i| LVR|j_2\rangle \langle j_4'| R^*L^*|j\rangle\langle j| LVR|j_4\rangle, \tag{3.63}$$

are used in the computation of the integral of Equation (3.60). The computations involved in bounded the integral of those expression are almost idential to each other. The second case, *case B*, is when either of the middle terms in the above expansion of the commutator, are used in the computation of the integral of Equation (3.60). Let's look at *case A* first. The expression that we want to integrate is given by

$$W_{i,j_2}W_{j,j_4}\langle j_2'| R^*VL^*|i\rangle \langle j_4'| R^*VL^*|j\rangle \tag{3.64}$$

Thus we need to integrate expressions of the form:

$$W_{i,j_1}\overline{W_{i,j_1'}}W_{i,j_3}\overline{W_{i,j_3'}}W_{i,j_2}W_{j,j_4}\langle j_2'| R^*VL^*|i\rangle \langle j_4'| R^*VL^*|j\rangle \tag{3.65}$$

Since the computations that follow are very similar for when either $\rho_1(L)dL$ or $\rho_2(R)dR$ is a 4-design respectively, we only provide the details for the first case. So suppose that $\rho_1(L)dL$ is at least a 4-design , and again define Q to be R^*VR so that the integral of the above expression is equal to

$$\begin{aligned}
&\int\int W_{i,j_1}\overline{W_{i,j_1'}}W_{i,j_3}\overline{W_{i,j_3'}}W_{i,j_2}W_{j,j_4}\langle j_2'|\,R^*VL^*\,|i\rangle\,\langle j_4'|\,R^*VL^*\,|j\rangle\,\rho_1(L)dL\rho_2(R)dR\\
&=\int\int W_{i,j_1}\overline{W_{i,j_1'}}W_{i,j_3}\overline{W_{i,j_3'}}W_{i,j_2}W_{j,j_4}\langle j_2'|\,QW^*\,|i\rangle\,\langle j_4'|\,QW^*\,|j\rangle\,\rho_1(L)dL\rho_2(R)dR\\
&=\int\int W_{i,j_1}\overline{W_{i,j_1'}}W_{i,j_3}\overline{W_{i,j_3'}}W_{i,j_2}W_{j,j_4}\overline{\langle i|\,WQ^*\,|j_2'\rangle\langle j|\,WQ^*\,|j_4'\rangle}\rho_1(L)dL\rho_2(R)dR\\
&=\int\int W_{i,j_1}\overline{W_{i,j_1'}}W_{i,j_3}\overline{W_{i,j_3'}}W_{i,j_2}W_{j,j_4}\overline{\langle i|\,WQ\,|j_2'\rangle\langle j|\,WQ\,|j_4'\rangle}\rho_1(L)dL\rho_2(R)dR\\
&=\sum_{\alpha,\beta}\overline{Q_{\alpha,j_2'}}\;\overline{Q_{\beta,j_4'}}\int\int W_{i,j_1}W_{i,j_2}W_{j,j_3}W_{j,j_4}\overline{W_{i,j_1'}}\;\overline{W_{j,j_3'}}\;\overline{W_{i,\alpha}}\;\overline{W_{j,\beta}}\rho_1(L)dL\rho_2(R)dR
\end{aligned}\tag{3.66}$$

where the two lines are derived using the fact that Q is hermitian, and by expanding Q in terms of its coordinates. Since computing the exact value of this integral involves summing over $S_4\times S_4$, we'll instead show the value of one of terms in the sum using $\sigma=(1234)$ and $\tau=(13)(24)$. It'll be easy to see that each term has value on the order $\frac{1}{d^4}$ and that the choice of permutation doesn't play a significant role in the computation.

These choices of permutations lead us to the following expression:

$$\mathrm{Wg}\Big(\sigma\tau^{-1}\Big)\sum_{\alpha\beta}\overline{Q_{\alpha,j_2'}}\;\overline{Q_{\beta,j_4'}}\delta_{i,j}\delta_{j_1,\alpha}\delta_{j_2,\beta}\delta_{j_3,j_1'}\delta_{j_4,j_3'}=\mathrm{Wg}\Big(\sigma\tau^{-1}\Big)\overline{Q_{j_1,j_2'}}\;\overline{Q_{j_2,j_4'}}\delta_{i,j}\delta_{j_3,j_1'}\delta_{j_4,j_3'}\tag{3.67}$$

Ignoring the constant $\mathrm{Wg}\Big(\sigma\tau^{-1}\Big)$ and substituting the right hand side of the equation into Equation (3.60) we obtain

$$\begin{aligned}
&\sum_{\substack{j_1,j_2,j_1',j_2'\\ j_3,j_4,j_3',j_4'}}\sqrt{p_{j_1}p_{j_1'}p_{j_2}p_{j_2'}p_{j_3}p_{j_3'}p_{j_4}p_{j_4'}}c_{j_1,j_1',j_2,j_2'}c_{j_3,j_3',j_4,j_4'}\overline{Q_{j_1,j_2'}}\;\overline{Q_{j_2,j_4'}}\delta_{j_3,j_1'}\delta_{j_4,j_3'}\delta_{i,j}\\
&=\sum_{\substack{j_1,j_2,j_2'\\ j_3,j_4,j_4'}}p_{j_3}p_{j_4}\sqrt{p_{j_1}p_{j_2}p_{j_2'}p_{j_4'}}c_{j_1,j_3,j_2,j_2'}c_{j_3,j_4,j_4,j_4'}\overline{Q_{j_1,j_2'}}\;\overline{Q_{j_2,j_4'}}\delta_{i,j}
\end{aligned}\tag{3.68}$$

Then ignoring the $\delta_{i,j}$ term for the moment, taking absolute values, and using the Cuachy-Schwartz inequality again, the above sum is less than or equal to

$$\begin{aligned}
&\sum_{\substack{j_1,j_2,j_2'\\ j_3,j_4,j_4'}} p_{j_3}p_{j_4}\sqrt{p_{j_1}p_{j_2}p_{j_2'}p_{j_4'}}|Q_{j_1,j_2'}|\cdot|Q_{j_2,j_4'}| \\
&=\sum_{j_1,j_2,j_2',j_4'} \sqrt{p_{j_1}p_{j_2}p_{j_2'}p_{j_4'}}|Q_{j_1,j_2'}|\cdot|Q_{j_2,j_4'}| \\
&=\Big(\sum_{j_1,j_2'}\sqrt{p_{j_1}p_{j_2'}}|Q_{j_1,j_2'}|\Big)\cdot\Big(\sum_{j_2,j_4'}\sqrt{p_{j_2}p_{j_4'}}|Q_{j_2,j_4'}|\Big) \\
&\leq\Big(\sum_{j_1,j_2'}p_{j_1}p_{j_2'}\Big)^{\frac{1}{2}}\cdot\Big(\sum_{j_1,j_2'}|Q_{j_1,j_2'}|^2\Big)^{\frac{1}{2}}\cdot\Big(\sum_{j_2,j_4'}p_{j_2}p_{j_4'}\Big)^{\frac{1}{2}}\cdot\Big(\sum_{j_2,j_4'}|Q_{j_2,j_4'}|^2\Big)^{\frac{1}{2}} \\
&=\|Q\|_{HS}^2 \\
&=d
\end{aligned} \tag{3.69}$$

Choosing any other pair of permutation σ and τ will yield a result of the same magnitude. Since $\mathrm{Wg}\big(\sigma\tau^{-1}\big)$ is of order at most $\frac{1}{d^4}$ for any $\sigma,\tau\in \mathrm{S}_4$, the integral of these terms must be of order at most $\frac{1}{d^3}$. The computation for when R is a 4-design will be similar to the previous case and will result in a value whose magnitude is of the order $\frac{1}{d^4}$.

Now we need to consider *case B*, in which we substitute the middle terms from the commutator in Equation (3.61) into Equation (3.60) and integrate. Namely, we want to integrate expressions of the form

$$W_{i,j_1}\overline{W_{i,j_1'}}W_{i,j_3}\overline{W_{i,j_3'}}\,\langle j_2'|\,R^*VL^*\,|i\rangle\,\langle i|\,LR\,|j_2\rangle\,\langle j_4'|\,R^*L^*\,|j\rangle\,\langle j|\,LVR\,|j_4\rangle\,. \tag{3.70}$$

Now suppose that L is at least a 4-design and define $Q=R^*VR$ again. Then the above expression reduces to

$$\begin{aligned}
&W_{i,j_1}W_{i,j_2}W_{i,j_3}\overline{W_{i,j_1'}}\ \overline{W_{i,j_3'}}\ \overline{W_{j,j_4'}}\,\langle j_2'|\,QW^*\,|i\rangle\,\langle j|\,WQ\,|j_4\rangle \\
&=\sum_{\alpha,\beta}\overline{Q_{j_2',\alpha}}Q_{\beta,j_4}W_{i,j_1}W_{i,j_2}W_{i,j_3}W_{j,\beta}\overline{W_{i,j_1'}}\ \overline{W_{i,j_3'}}\ \overline{W_{j,j_4'}}\ \overline{W_{i,\alpha}}
\end{aligned} \tag{3.71}$$

This integral is a little trickier than the rest. Up to now the indices ,α and β, that were added when applying Q to a ket or bra, have been paired with a j,j', or an i

when integrating. However, there is now the possibility that α and β can be paired with each other this time. When looking at the string of Kronecker deltas produced from the Haar integral, we must simplify the sum in Equation (3.60) based on two distinct cases. The first, when α and β are not paired with each other, and the second when α and β are paired with each other. Computationally, the first case is nearly identical to *case A*. We will only focus on the case in which α and β are paired with each other in the integration formula in Lemma (3.2.2). Once again, we only focus on example permutations of this form, and leave it to the reader to see that the computations are unchanged by a different permutation as long as α and β are still paired with each other. For simplicity, take $\sigma = \tau = (1)_4$. Then integrating the above expression with respect to W yields

$$\begin{aligned}&\int \sum_{\alpha,\beta} \overline{Q_{j_2,'\alpha}} Q_{\beta,j_4} W_{i,j_1} W_{i,j_2} W_{i,j_3} W_{j,\beta} \overline{W_{i,j_1'}}\ \overline{W_{i,j_3'}}\ \overline{W_{j,j_4'}}\ \overline{W_{i,\alpha}} dW \\ &= \mathrm{Wg}(1,1,1,1) \sum_{\alpha,\beta} \overline{Q_{j_2',\alpha}} Q_{\beta,j_4} \delta_{j_1,j_1'} \delta_{j_2,j_3'} \delta_{j_3,j_4'} \delta_{\alpha,\beta} \\ &= \mathrm{Wg}(1,1,1,1) \sum_{\alpha} \overline{Q_{j_2',\alpha}} Q_{\alpha,j_4} \delta_{j_1,j_1'} \delta_{j_2,j_3'} \delta_{j_3,j_4'}\end{aligned} \tag{3.72}$$

Just like before, we ignore the leading constant for readability and substitute the above expression into 3.60 and bound it in absolute value.

$$\begin{aligned}&\sum_{\alpha} \sum_{\substack{j_1,j_2,j_1',j_2' \\ j_3,j_4,j_3',j_4'}} \sqrt{p_{j_1} p_{j_1'} p_{j_2} p_{j_2'} p_{j_3} p_{j_3'} p_{j_4} p_{j_4'}}\, c_{j_1,j_1',j_2,j_2'} c_{j_3,j_3',j_4,j_4'} \overline{Q_{j_2',\alpha}} Q_{\alpha,j_4} \delta_{j_1,j_1'} \delta_{j_2,j_3'} \delta_{j_3,j_4'} \\ &= \sum_{\alpha} \sum_{j_1,j_2,j_3,j_4,j_2'} p_{j_1} p_{j_2} p_{j_3} \sqrt{p_{j_2'} p_{j_4}}\, c_{j_1,j_1,j_2,j_2'} c_{j_3,j_2,j_4,j_3} \overline{Q_{j_2',\alpha}} Q_{\alpha,j_4}\end{aligned} \tag{3.73}$$

Taking absolute values using Cauchy Schwartz again, we see that the magnitude of this sum is bounded by

$$\begin{aligned}
&\sum_{\alpha}\sum_{j_4 j_2'} \sqrt{p_{j_2'} p_{j_4}} |Q_{j_2',\alpha}| \cdot |Q_{\alpha,j_4}| \\
&= \sum_{\alpha}\Big(\sum_{j_4} \sqrt{p_{j_4}} |Q_{\alpha,j_4}|\Big) \cdot \Big(\sum_{j_2'} \sqrt{p_{j_2'}} |Q_{j_2',\alpha}|\Big) \\
&\leq \sum_{\alpha}\Big(\sum_{j_4} p_{j_4}\Big)^{\frac{1}{2}} \cdot \Big(\sum_{j_4} |Q_{\alpha,j_4}|^2\Big)^{\frac{1}{2}} \cdot \Big(\sum_{j_2'} p_{j_2'}\Big)^{\frac{1}{2}} \cdot \Big(\sum_{j_2'} |Q_{\alpha,j_2'}|^2\Big)^{\frac{1}{2}} \\
&\leq \sum_{\alpha} 1 \\
&= d
\end{aligned} \tag{3.74}$$

Thus when multiplying by the appropriate evaluation of the Weingarten function, this expression is of order at most $\frac{1}{d^3}$. And since there are d^2 terms in the integrand of the second moment given in Equation (3.59), the second moment must be of order $\frac{1}{d}$. And therefore

$$\mathrm{Var}\Big[\frac{\partial}{\partial\theta} T_f(\rho;\theta)\Big] \sim \mathcal{O}\Big(\frac{1}{d}\Big) \tag{3.75}$$

Lastly, since $d = 2^k$ where k is the number of qubits, we see that the variance decreases exponentially in the number of qubits, implying the existence of barren plateaus in the training landscape of $T_f(\rho;\theta)$.

3.3 Numerical Simulations

Recall that for a two qubit state $|\psi\rangle = \sum_{i,j=0}^{1} c_{i,j} |ij\rangle$, the *concurrence* [26, 44] of a state $\mathrm{C}(\psi)$ is defined by

$$\mathrm{C}(\psi) = 2|c_{0,0}c_{1,1} - c_{0,1}c_{1,0}|. \tag{3.76}$$

The concurrence is related to the von Neumann entanglement entropy via the formula

$$S(\psi) = h\Big(\frac{1+\sqrt{1-C(\psi)^2}}{2}\Big) \tag{3.77}$$

where h is the standard binary entropy defined by $h(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ [26]. Interestingly enough, it is possible to directly measure the concurrence of a

2 qubit quantum system if there is access to two decoupled copies of the same state [44]. One can do this using the following quantum circuit:

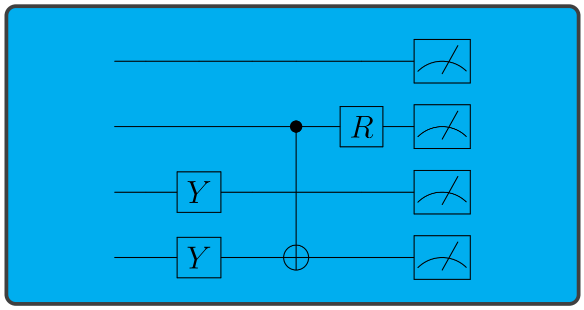


Figure 3.3: Quantum circuit for directly measuring concurrence in 2 qubit states.

where the first two wires and the second two wires are in the state $|\psi\rangle$ respectively, and R is the unitary

$$R = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}. \tag{3.78}$$

After applying these operations to the state $|\psi\rangle \otimes |\psi\rangle$, the probability amplitude of the state $|00\rangle$ will be $\frac{C(\psi)^2}{8}$.

Using the above ideas, we can append the circuit for measuring the concurrence of a state to the pure state ensemble circuit, to the measure the concurrence of formation of a 2 qubit ensemble of states directly, and therefore efficiently compute the entanglement of formation for 2 qubit states. The entire circuit for this process is given below when for a an arbitrary density matrix ρ with a purification $|\psi\rangle$ only requiring two qubits.

Using the above ideas, we can append the circuit for measuring the concurrence of a state to the pure state ensemble circuit, so that we can measure the concurrence of a 2 qubit ensemble of states directly, and therefore efficiently compute the entanglement of formation for 2 qubit states. The entire circuit for this process is given in Figure 3.4 for an arbitrary density matrix ρ with a purification $|\psi\rangle$ requiring two qubits.

Let's go through the circuit in Figure 3.4 step by step since there's a lot going on. We first prepare the purification $|\psi\rangle$ of a two qubit density operator on the first four and last four quantum registers so that we can access two decoupled copies of the pure state ensemble. Next we apply $U(\hat{\theta})$ to both of the ancilla spaces and measure the ancillae. This will put the top two registers into the state $|\varphi_i\rangle$, (as in Equation (3.4)), the third and fourth registers into the state $|i\rangle$, the 5th and 6th registers into the state $|\varphi_j\rangle$, (again, as in Equation (3.4)), and the last two registers into the state $|j\rangle$, where $i, j \in \{0, 1, 2, 3\}$. Thus the state of circuit at this point is then $|\varphi_i i \varphi_j j\rangle$. If $i = j$, then we can measure the concurrence by appending the circuit in Figure 3.3 to wires 1,2,5, and 6, and taking note of the frequency of the state $|00i00j\rangle$ where the first two registers are in state $|00\rangle$, the third and fourth are in state $|i\rangle$, the fifth and sixth are in state $|00\rangle$, and lastly the 7th and 8th are in the state $|j\rangle$. We then run this experiment a desired number of times to also make note of the probabilities q_i as defined as in Equation (3.3) and the concurrence $C(\varphi_i)$ for each $i \in \{0, 1, 2, 3\}$. Lastly we use the measurements to approximate the entanglement entropy in the equation below via

$$\sum_{i=0}^{3} q_i h\Big(\frac{1 + \sqrt{1 - \mathrm{C}^2(\varphi_i)}}{2}\Big) \approx \mathrm{S}(\rho). \tag{3.79}$$

Using the maximally mixed state and its canonical purification with 4 ancillae, we simulated the above variational algorithm and showcase our results in Figures 3.5 and 3.6. Notice that in the circuit above, the first two quantum registers will be in the state $|\phi_i\rangle$ for some $i \in \{1, \ldots, 4\}$ and the fourth and fifth quantum registers will be in the state $|\phi_j\rangle$ for some $j \in \{1, \ldots, 4\}$, each appearing with probability q_i and q_j respectively, after the ancillae have been measured. If the two sets of ancillae are not in the same state, then we can make note of the probability $p(i|j)$ which is defined as the probability that the second ancilla is observed to be in state $|j\rangle$ given that the first ancilla was observed to be in state $|i\rangle$, and then we can discard the

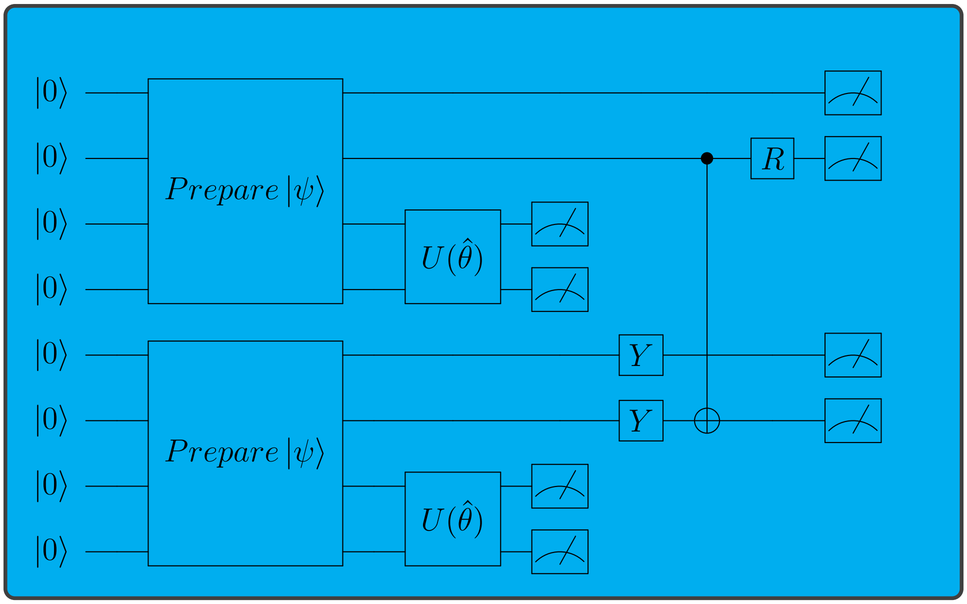


Figure 3.4: Quantum circuit for variationally approximating the entanglement of formation by directly measuring the concurrence of pure state ensembles.

state of the system. However, if they are in the same state, then amplitude of the $'00i00i'$ quantum register will be $\mathrm{C}(\phi_i)$ for each $i \in \{1, \ldots, 4\}$. Running this circuit a number of times allows us to approximate p_i by computing the marginal distributions of $p(x|y)$ while also taking note of $\mathrm{C}(\phi_i)$ for each $i \in \{1, \ldots, 4\}$. We can then compute

$$\sum_i p_i h\Big(\frac{1 + \sqrt{1 - \mathrm{C}^2(\phi_i)}}{2}\Big) \approx \mathrm{S}(\rho). \tag{3.80}$$

Using the maximally mixed state and its canonical purification with 4 ancillae, we simulated the above variational algorithm and showcase our results in figures 3.5 and 3.6.

The python code corresponding to the simulations and plots can be found at `https://github.com/TheMathDoctor/code_for_vqa_paper`. These simulations made extensive use of both the numpy[23] and matplotlib [28] python libraries.

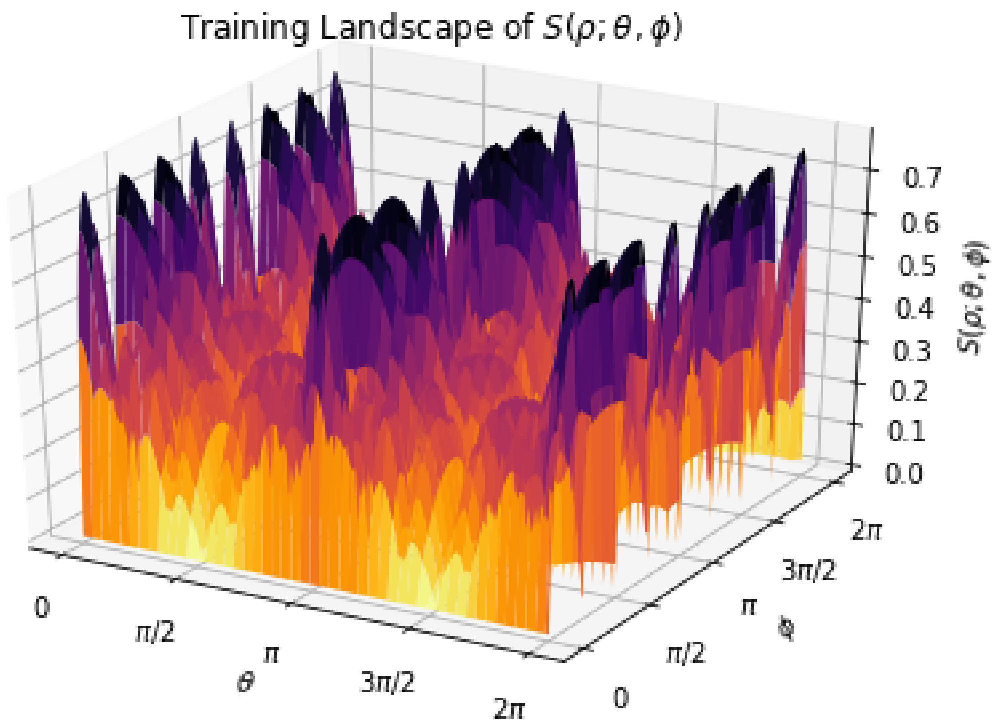


Figure 3.5: The training landscape of $S(\rho;\theta,\phi)$ where ρ is the maximally mixed state and $\theta,\phi \in [0,2\pi]$.

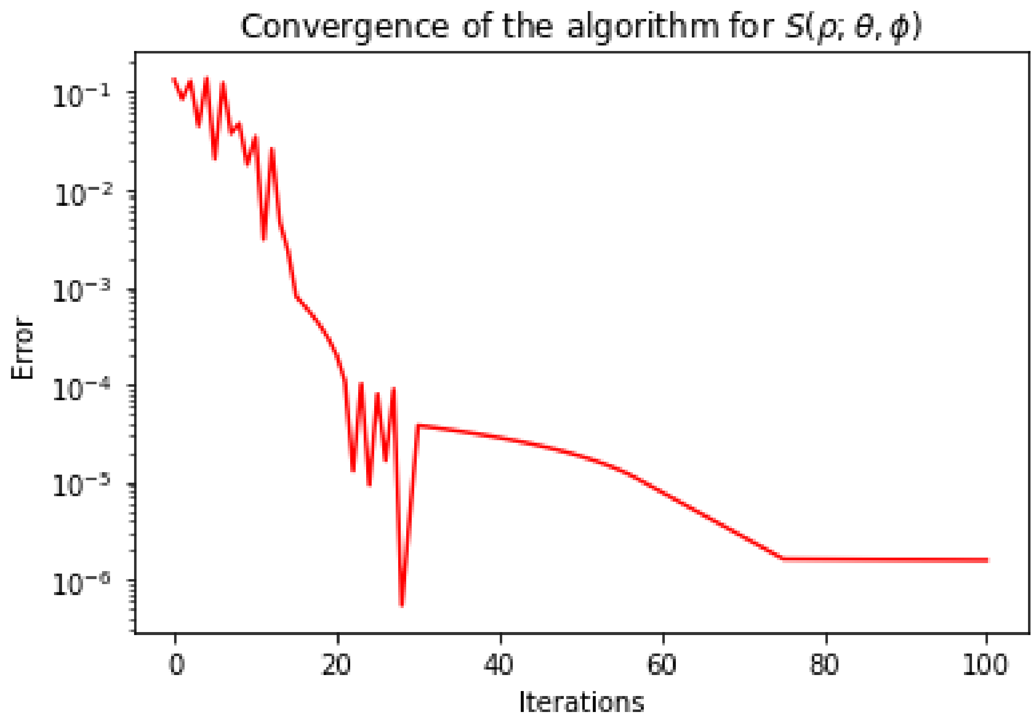


Figure 3.6: Convergence of the algorithm for the entanglement of formation of the maximally mixed state.

Bibliography

[1] A. Arrasmith et al. "Equivalence of quantum barren plateaus to cost concentration and narrow gorges". In: (2021). eprint: https://arxiv.org/pdf/2104.05868.pdf.

[2] H. Barnum and N. Linden. "Monotones and invariants for multi-particle quantum states". In: *J. Phys. A* 34 (2001), pp. 6787–6805. DOI: 10.1088/0305-4470/34/35/305. eprint: arXiv:quant-ph/0103155.

[3] K. Bharti et al. "Noisy intermediate-scale quantum (NISQ) algorithms". In: (2021). eprint: https://arxiv.org/abs/2101.08448.

[4] F. Brandão, A. Harrow, and M. Horodecki. "Local Random Quantum Circuits are Approximate Polynomial-Designs". In: *Communications in Mathematical Physics* 346 (2016), pp. 397–434. DOI: https://doi.org/10.1007/s00220-016-2706-8. eprint: https://arxiv.org/pdf/1208.0692.pdf.

[5] M. Cerezo et al. "Cost function dependent barren plateaus in shallow parametrized quantum circuits". In: *Nature Communications* 12 (2021). DOI: https://doi.org/10.1038/s41467-021-21728-w.

[6] M. Cerezo et al. "Variational Quantum Algorithms". In: (2020). eprint: https://arxiv.org/abs/2012.09265.

[7] E. Chitambar et al. "Everything You Always Wanted to Know About LOCC (But Were Afraid to Ask)". In: *Comm. Math. Phys.* 328 (2014), pp. 303–326. DOI: 10.1007/s00220-014-1953-9. eprint: arXiv:1210.4583.

[8] M. Cirone. "Quantifying entanglement with probabilities". In: (2001). eprint: arXiv:quant-ph/0110139.

[9] B. Collins. "Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability". In: *International Mathematics Research Notices* 17 (2003), pp. 953–982. eprint: https://arxiv.org/abs/math-ph/0205010.

[10] M. Donald, M. Horodecki, and O. Rudolph. “The Uniqueness Theorem for Entanglement Measures”. In: *J. Math. Phys.* 43 (2002), pp. 4252–4272. DOI: 10.1063/1.1495917. eprint: arXiv:quant-ph/0105017.

[11] D. Eisenbud. *Commutative Algebra With A View Towards Algebraic Geometry.* New York: Spring-Verlag, 1995.

[12] M. Fannes. “A Continuity Property of The Entropy Density for Spin Lattice Systems”. In: *Comm. Math. Phys.* 31 (1973), pp. 291–294. DOI: https://doi.org/10.1007/bf01646490.

[13] M. Fukuda, R. König, and I. Nechita. “RTNI—A symbolic integrator for Haar-random tensor networks”. In: *Journal of Physics A: Mathematical and Theoretical* 52 (), pp. 425–303. DOI: https://doi.org/10.1088/1751-8121/ab434b. eprint: https://arxiv.org/abs/1902.08539.

[14] M. Fukuda, R. König, and I. Nechita. “RTNI_light_web”. In: (2019). eprint: https://motohisafukuda.pythonanywhere.com/.

[15] R.F. Werner G. Vidal. “Computable Measure of Entanglement”. In: *Phys. Rev. A* 65 (2001). DOI: https://doi.org/10.1103/physreva.65.032314. eprint: arXiv:quant-ph/0102117.

[16] Nicolas Gisin. “Stochastic quantum dynamics and relativity”. In: *Helvetica Physica Acta* 62 (1989), pp. 363–371. DOI: http://doi.org/10.5169/seals-116034.

[17] J. Glimm. “On a certain class of operator algebras”. In: *Trans. Amer. Math. Soc.* 95 (1960), pp. 318–340. DOI: https://doi.org/10.1090/s0002-9947-1960-0112057-5.

[18] G. Gour. “Family of Concurrence Monotones and its Applications”. In: *Phys. Rev. A* 71 (2005). DOI: 10.1103/PhysRevA.71.012318. eprint: arXiv:quant-ph/0410148.

[19] S. Gudder. “A Theory of Entanglement”. In: *Quanta* 9 (2020), pp. 7–15. DOI: https://doi.org/10.12743/quanta.v9i1.115. eprint: arXiv:1904.06589.

[20] S. Gudder. “Spooky Action at a Distance”. In: *Quanta* 9 (2020), pp. 1–6. DOI: https://doi.org/10.12743/quanta.v9i1.113. eprint: arXiv:2005.11870.

[21] N. Hadjisavvas. “Properties of mixtures on non-orthogonal states”. In: *Letters in Mathematical Physics* 5 (1981), pp. 327–332. DOI: https://doi.org/10.1007\%2FBF00401481.

[22] E. Hanson and N. Datta. "Universal proofs of entropic continuity bounds via majorization flow". In: (2021). eprint: https://arxiv.org/abs/1909.06981v3.

[23] Charles R. Harris et al. "Array programming with NumPy". In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: https://doi.org/10.1038/s41586-020-2649-2.

[24] A. Harrow and R. Low. "Random Quantum Circuits are Approximate 2-designs". In: *Communications in Mathematical Physics* 291 (2009), pp. 257–302. DOI: https://doi.org/10.1007/s00220-009-0873-6. eprint: https://arxiv.org/abs/0802.1919.

[25] P. Hayden, D. Leung, and A. Winter. "Aspects of Generic Entanglement". In: *Communications in Mathematical Physics* 265 (1 2006), pp. 95–117. DOI: https://doi.org/10.1007/s00220-006-1535-6. eprint: https://arxiv.org/abs/quant-ph/0407049.

[26] R. Horodecki et al. "Quantum entanglement". In: *Rev.Mod.Phys.* 81 (2009), pp. 865–942. DOI: 10.1103/RevModPhys.81.865. eprint: arXiv:quant-ph/0702225.

[27] L. Houghston, R. Josza, and W. Wooters. "A complete classification of quantum ensembles having a given density matrix". In: *Physical Letters A* 483 (1993), pp. 14–18. DOI: https://doi.org/10.1016\%2F0375-9601\%2893\%2990880-9.

[28] J. D. Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[29] J.Lau et al. "Quantum assisted simulation of time dependent Hamiltonians". In: (2021). eprint: https://arxiv.org/abs/2101.07677.

[30] A. Kandala et al. "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets". In: *Nature* 548 (2017), pp. 242–246. DOI: https://doi.org/10.1038/nature23879.

[31] P. Kaye, R. Laflamme, and M. Mosca. *An Introduction too Quantum Computing*. Oxford, New York: Oxford University Press, 2006.

[32] S. Lee et al. "Convex-roof extended negativity as an entanglement measure for bipartite quantum systems". In: *Phys. Rev. A* 68 (2003). DOI: 10.1103/PhysRevA.68.062304. eprint: arXiv:quant-ph/0310027.

[33] S. Matsumoto. “Weingarten calculus for matrix ensembles associated with compact symmetric spaces”. In: *World Scientific* 2 (2013). DOI: https://doi.org/10.1142/S2010326313500019. eprint: https://arxiv.org/abs/1301.5401.

[34] J. Maziero. “Computing partial traces and reduced density matrices”. In: *International Journal of Modern Physics C* 28 (2017). DOI: https://doi.org/10.1142/s012918311750005x. eprint: https://arxiv.org/abs/1601.07458.

[35] J. McClean et al. “Barren plateaus in quantum neural network training landscapes”. In: *Nature Communications* 9 (2018). DOI: https://doi.org/10.1038/s41467-018-07090-4.

[36] J. McClean et al. “Barren plateaus in quantum neural network training landscapes”. In: *Nature Communications* 9 (2018). DOI: https://doi.org/10.1038/s41467-018-07090-4.

[37] N. Merman. “What Do These Correlations Know about Reality? Nonlocality and the Absurd”. In: *Foundations of Physics* 29 (1999), pp. 571–587. DOI: https://doi.org/10.1023\%2FA\%3A1018864225930.

[38] J. von Neumann. “Thermodynamik quantenmechanischer Gesamtheiten”. ger. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1927 (1927), pp. 273–291. URL: http://eudml.org/doc/59231.

[39] M. Ohya and D. Petz. *Quantum Entropy and its Use*. Berlin, Germany: Texts and Monographs in Physics, 1993.

[40] A. Pesah et al. “Absence of Barren Plateaus in Quantum Convolutional Neural Networks”. In: (2021). eprint: https://arxiv.org/abs/2011.02966.

[41] M. B. Plenio and S. Virmani. “An introduction to entanglement measures”. In: *Quant. Inf. Comput.* 7 (2007), pp. 1–51. eprint: arXiv:quant-ph/0504163.

[42] W. Press et al. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 2007.

[43] G. A. Raggio. “Properties of q-entropies”. In: *Journal of Mathematical Physics* 36 (9 1995), pp. 4785–4791. DOI: https://doi.org/10.1063/1.530920.

[44] G. Romero et al. “Direct measurement of concurrence for atomic two-qubit pure states”. In: *Physical Review A* 75 (2007). DOI: https://doi.org/10.1103/physreva.75.032303. eprint: https://arxiv.org/abs/quant-ph/0611016.

[45] J. Romero et al. “Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz”. In: *Quantum Science and Technology* 4 (2018). DOI: https://doi.org/10.1088/2058-9565/aad3e4.

[46] E. Schr/"odinger. “Probability Relations Between Separated Systems”. In: *Proceedings of the Cambridge Philosophical Society* 32 (1936), pp. 446–452. DOI: https://doi.org/10.1017\%2FS0305004100019137.

[47] M. Shirokov. “On properties of the space of quantum states and their application to the construction of entanglement monotones”. In: *Izvestiya: Math.* 2 (2010), pp. 849–882. eprint: https://arxiv.org/abs/0804.1515.

[48] C. Spengler, M. Huber, and B. Hiesmayr. “A composite parameterization of unitary groups, density matrices and subspaces”. In: *Journal of Physics A: Mathematical and Theoretical* 43 (2010). DOI: https://doi.org/10.1088/1751-8113/43/38/385306.

[49] A. Uhlmann. “Optimizing entropy relative to a channel or a subalgebra”. In: *Open Sys. and Inf. Dyn.* 5 (1998), pp. 209–227. eprint: arXiv:quant-ph/9701014.

[50] V. Vedral et al. “Quantifying Entanglement”. In: *Phys.Rev.Lett.* 78 (1997), pp. 2275–2279. DOI: 10.1103/PhysRevLett.78.2275. eprint: arXiv:quant-ph/9702027.

[51] G. Vidal. “Entanglement monotones”. In: *J.Mod.Opt.* 47 (2000), pp. 355–376. DOI: 10.1080/09500340008244048. eprint: arXiv:quant-ph/9807077.

[52] J Watrous. *The theory of quantum information*. Cambridge, United Kingdom: Cambridge University Press, 2018.

[53] L. Zhang. “Matrix integrals over unitary groups: An application of Schur-Weyl duality”. In: (2015). eprint: https://arxiv.org/abs/1408.3782.

Appendix A

Code For Plots

A.1 Lipschitz Extension Plots

```
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from typing import Callable
from  matplotlib import rcParams
rcParams.update ({'figure.autolayout': True })

class lipschitz_extension(object):
  """
  A class for extending a lipschitz
  function on the unit circle
  to the unit disc using the lipschitz
  extension theorem
  """

  def __init__(self, f: Callable[[float, float], float])
  -> None:
    """
    :type f: Callable[[float, float], float]
```

```
    :rtype : None

    f should be a function that takes in two float parameters
    and outputs 1 float parameter.

    The domain of f should include the unit circle in two
    variables.

    """
    self.f = f

def extension(self,x: float,y: float) -> float:
    """
    type: x: float
    type: y: float
    rtype: float

    The extension of self.f to the unit disc.
    """

    norm = np.sqrt(x*x + y*y)
    if norm > 0:
      X, Y = np.abs(x)/norm, np.abs(y)/norm

    #return 0 for x + y > 1 so that plot
    doesn't extend past unit ball
    return norm*self.f(X,Y) if norm > 0 and x + y < 1 else 0
```

```
def plot_extension(self) -> None:
    """
    Plots self.extension on the domain [0,1]x[0,1].
    """
    a = np.arange(0,1,.01)
    b = np.arange(0,1,.01)
    X,Y = np.meshgrid(a,b)

    L = len(a)
    Z = np.zeros((L,L))
    for i in range(L):
        for j in range(L):
            Z[i,j]=self.extension(a[i],b[j])

    fig = plt.figure()
    ax = fig.add_subplot(111, projection = '3d')
    ax.plot_surface(X, Y, Z,cmap = 'Blues')
    ax.set_xlabel('x')
    ax.set_ylabel('y')
    ax.set_zlabel(r'$T_2(x,y)$')
    ax.set_zlim((0,.25))

    x = np.arange(0,1,.01)
    y = 1-abs(x)
    n = len(x)
    z = [extended_tsallis(x[i],y[i]) for i in range(n)]
```

```
    ax.plot3D(x,y,z,'red',linewidth=3)
    ax.set_title(r'Lipschitz Extension of $T_2(x,y)$')
    #ax.view_init(0,0)


#tsallis entropy
def tsallis(x,y):
  return x-x**2 + y - y**2

tsallis_extension = lipschitz_extension(tsallis)
tsallis_extension.plot_extension()
```

A.2 Code For Training Landscapes and Convergence Charts

The following code is taken from a jupypter notebook and will supply the plots for training landscapes as well the convergence chart.

```
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import rcParams
rcParams.update({'figure.autolayout': True})

def adjoint(M):
  return np.transpose(np.conj(M))

def binary_entropy(x):
  return -x*np.log2(x)-(1-x)*np.log2(1-x)
  if x > 0 and x <1 else 0
```

```
nomralization = 1/np.sqrt(2)

I2 = np.identity(2)
I4 = np.identity(4)
I8 = np.identity(8)
I16 = np.identity(16)
I64 = np.identity(64)
proj0 = [[1,0],[0,0]]
proj1 = [[0,0],[0,1]]
X = [[0,1],[1,0]]
Y = [[0,complex(0,-1)],[complex(0,1),0]]

def rx(theta):
    a = complex(np.cos(theta),0)
    b = complex(0,np.sin(theta))
    M = np.array([[a,-b],[b,a]])
    return M

def ry(theta):
    a = complex(np.cos(theta),0)
    b = complex(np.sin(theta),0)
    M = np.array([[a,b],[-b,a]])
    return M
```

```
cx_next = np.array([[1,0,0,0],
                    [0,1,0,0],
                    [0,0,0,1],
                    [0,0,1,0]])

#the following two lines are written this way so
#that the code fits into the margin requirements
#of my dissertation.
cx_15 = np.kron(I2,np.kron(proj0,I64))
cx_15 += + np.kron(I2,np.kron(proj1,
np.kron(I8,np.kron(X,I4))))

R = np.array([[nomralization,nomralization],
              [-nomralization,nomralization]])



initial = [.5,0,0,0,
           0,.5,0,0,
           0,0,.5,0,
           0,0,0,.5]

def pqc(theta,phi):
  #U = np.kron(rx(theta),ry(phi)) # first training landscape
  U = np.kron(ry(theta),ry(phi))
  U = np.matmul(U,cx_next)
  #return U
```

```
    #return np.matmul(U,U)
    U = np.matmul(U,U)
    U = np.matmul(U,U)
    return U

def measure(i,j):
    bit0 = proj0 if i == 0 else proj1
    bit1 = proj0 if j == 0 else proj1
    M = np.kron(bit0,bit1)
    partial_measurement = np.kron(I4,M)
    return partial_measurement

def concurrence(theta,phi,i,j):
    U = pqc(theta,phi)
    one_side = np.kron(I4,U)
    first_operation = np.kron(one_side,one_side)
    state = np.kron(initial,initial)
    state = np.matmul(first_operation,state)
    M = measure(i,j)
    M2 = np.kron(M,M)
    state = np.matmul(M2,state)
    ygates = np.kron(I16,np.kron(Y,np.kron(Y,I4)))
    state = np.matmul(ygates,state)
    state = np.matmul(cx_15,state)
    r_gate = np.kron(I2,np.kron(R,I64))
    state = np.matmul(r_gate,state)
    #conc = np.abs(state[0])**2
```

```
  index = i*(2**5+2) + j*(2**4+1)
  #concurrence squared divided by 8
  conc = np.abs(state[index])
  return conc

def probability(theta,phi,i,j):
  state = initial
  state = np.matmul(np.kron(I4,pqc(theta,phi)),state)
  M = measure(i,j)
  state = np.matmul(M,state)
  probability = np.dot(np.conj(state),state)
  return probability

def entanglement(theta,phi):
  entropy = 0
  for i in range(2):
    for j in range(2):
      prob = probability(theta,phi,i,j)
      conc = concurrence(theta,phi,i,j)*8
      val = (1+np.sqrt(1-conc))/2
      entropy += prob*binary_entropy(val)

  return np.abs(entropy)

def grad_entanglement(theta,phi):
  delta = .0001
  first = entanglement(theta+delta,phi)
```

```
    first -= -entanglement(theta-delta,phi)
    first /= 2*delta
    second = entanglement(theta,phi+delta)
    second -= -entanglement(theta,phi-delta)
    second /= 2*delta
    return np.array([first,second])

entanglement(0,0)
probability(.1,np.pi,1,1)


a = np.arange(0,2*np.pi,.5)
b = np.arange(0,2*np.pi,.5)
thetas,phis = np.meshgrid(a,b)
L = len(a)
Z = np.zeros((L,L))
for i in range(L):
    for j in range(L):
        x = a[i]
        y = b[j]
        Z[i,j]=entanglement(x,y)



fig = plt.figure()
ax = fig.add_subplot(111, projection = '3d')
ax.plot_surface(thetas, phis, Z,cmap='inferno_r')
ax.set_xticks(np.arange(0,2*np.pi+.001,np.pi/2))
ax.set_yticks(np.arange(0,2*np.pi+.001,np.pi/2))
```

```
labels = ['$0$', r'$\pi/2$', r'$\pi$'
, r'$3\pi/2$', r'$2\pi$']
ax.set_xticklabels(labels)
ax.set_yticklabels(labels)

ax.set_xlabel('x')
ax.set_ylabel('y')
ax.set_zlabel('z')

ax.view_init(0,0)
fig

epochs = 100
eta = .01
theta = 2*np.random.rand()
phi = 2*np.random.rand()
thetas = [theta]
phis = [phi]

for epoch in range(epochs):
  if epoch in range(15):
    eta = .1
  elif epoch in range(15,30):
    eta = .01
  elif epoch in range(30,75):
    eta = .0001
  else:
```

```
      eta = .000001
    grad = grad_entanglement(theta,phi)
    theta -= eta*grad[0]
    phi -= eta*grad[1]
    thetas.append(theta)
    phis.append(phi)

entanglements = [entanglement(thetas[i],phis[i])
for i in range(epochs+1)]

plt.plot(range(epochs+1),entanglements,color='red')
plt.yscale('log')
plt.title(r'Convergence of the
algorithm for $S(\rho;\theta,\phi)$')
plt.xlabel('Iterations')
plt.ylabel('Error')

#two ry
a = np.arange(0,2*np.pi,.01)
b = np.arange(0,2*np.pi,.01)
thetas,phis = np.meshgrid(a,b)
L = len(a)
Z = np.zeros((L,L))
for i in range(L):
    for j in range(L):
        x = a[i]
        y = b[j]
```

```
        Z[i,j]=entanglement(x,y)

fig = plt.figure()
ax = fig.add_subplot(111, projection = '3d')
ax.plot_surface(thetas, phis, Z,cmap='inferno_r')
ax.set_xlabel(r'$\theta$')
ax.set_ylabel(r'$\phi$')
ax.set_xticks(np.arange(0,2*np.pi+.001,np.pi/2))
ax.set_yticks(np.arange(0,2*np.pi+.001,np.pi/2))
labels = ['$0$', r'$\pi/2$',
r'$\pi$', r'$3\pi/2$', r'$2\pi$']
ax.set_xticklabels(labels)
ax.set_yticklabels(labels)
ax.set_zlabel(r'$S(\rho;\theta,\phi)$')
ax.set_title(r'Training Landscape of $S(\rho;\theta,\phi)$')

ax.view_init(0,0)
fig
ax.view_init(45,45)
fig

#t_f training landscape and convergence
def probs_and_phis(theta,phi):
    probabilities = []
    big_phis = []

    state = initial
```

```
  state = np.matmul(np.kron(I4,pqc(theta,phi)),state)
  for i in range(2):
    for j in range(2):
      M = measure(i,j)
      measurement = np.matmul(M,state)
      probability = np.dot(np.conj(measurement),measurement)
      probabilities.append(probability)

      outer_product = np.outer(measurement,measurement)
      #below is written so that the code fits into
      #the margin requirements
      #for my dissertation
      A =[[np.trace(outer_product[0:8,0:8]),
      np.trace(outer_product[0:8,8:16])],
        [np.trace(outer_product[8:16,0:8]),
        np.trace(outer_product[8:16,8:16])]]
      partial_trace = np.array(A)

      big_phis.append(partial_trace)
  return probabilities,big_phis

def t_f(theta,phi):
  probabilities_and_big_phis = probs_and_phis(theta,phi)
  t = 0
  for i in range(4):
    p = probabilities_and_big_phis[0][i]
    state = probabilities_and_big_phis[1][i]
```

```
        t += p**2 - np.trace(np.matmul(state,state))

    return np.abs(t)


stuff = probs_and_phis(np.pi,np.pi)
stuff[0]


x,y= np.random.rand(2,1)
t_f(x,y)


a = np.arange(0,2*np.pi,.01)
b = np.arange(0,2*np.pi,.01)
thetas,phis = np.meshgrid(a,b)
L = len(a)
Z = np.zeros((L,L))
for i in range(L):
    for j in range(L):
        x = a[i]
        y = b[j]
        Z[i,j]=t_f(x,y)


fig = plt.figure()
ax = fig.add_subplot(111, projection = '3d')
ax.plot_surface(thetas, phis, Z,cmap='inferno_r')
ax.set_xlabel(r'$\theta$')
ax.set_ylabel(r'$\phi$')
ax.set_xticks(np.arange(0,2*np.pi+.001,np.pi/2))
```

```
ax.set_yticks(np.arange(0,2*np.pi+.001,np.pi/2))
labels = ['$0$', r'$\pi/2$',
r'$\pi$', r'$3\pi/2$', r'$2\pi$']
ax.set_xticklabels(labels)
ax.set_yticklabels(labels)
ax.set_zlabel(r'$T_f(\rho;\theta,\phi)$')
ax.set_title(
r'Training Landscape of $T_f(\rho;\theta,\phi)$')

ax.view_init(45,45)
fig

ax.view_init(0,0)
fig
```