Concentration Bounds from Parallel Repetition Theorems

by

Taylor Hornby

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2018

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis contributes to the study of parallel repitition theorems and concentration bounds for nonlocal games and quantum interactive proofs. We make the following contributions:

- A lemma that is useful for converting parallel repetition theorems (bounds on the probability of winning *all* instances of a nonlocal game which is being repeated in parallel) into concentration bounds (bounds on winning a certain *fraction* of the instances).
- Exponentially-decaying concentration bounds for two-player games on the uniform distribution and k-player free games, against quantum strategies.
- A proof that given a quantum interactive proof system with parameters α (the probability with which the verifier can be convinced to accept when they should accept) and β (the soundness error), as long as $\alpha > \beta$, both the soundness error and completeness error can be reduced exponentially by repeating the protocol in parallel and requiring an $(\alpha + \beta)/2$ fraction of the repetitions to be won. Our result requires a log-factor more repetitions than are necessary in the classical case.

Acknowledgements

This thesis documents the work I did while I was a master's student at the University of Waterloo under the supervision of John Watrous. I would like to thank the University of Waterloo and the Institute for Quantum Computing for providing an excellent environment for research and learning.

I would like to thank David R. Cheriton for funding the David R. Cheriton Graduate Scholarship, as well as the Government of Ontario for providing the Ontario Graduate Scholarship. These scholarships were the main source of funding for my work over the past two years, including the work being presented here.

I would like to thank John Watrous for supervising me, for helping me learn the theory of quantum information, and for helping me find my way to the results you will read in this thesis. I would also like to thank him for providing me with feedback on draft versions of this thesis. Thanks also to my readers Eric Blais and Richard Cleve for suggesting improvements and catching some mistakes.

Lastly, I would like to thank my friends and family for the motivation and encouragement they so generously provide.

Table of Contents

Chapter 1

Introduction

In this thesis, we study parallel repetition of nonlocal games and interactive proofs.

Nonlocal games are played between a referee and two or more players. The referee sends questions to the players and, without being allowed to communicate, the players must send answers back. The referee decides whether or not the players win the game based on the questions they sent and the answers they receive. Nonlocal games are important for understanding entanglement. For example, the CHSH game—where two players are sent random bits x_1 and x_2 and are required to return answers a_1 and a_2 such that $a_1 \oplus a_2 = x_1 \cdot x_2$—can be won with probability approximately 85% by quantum-entangled players, but only 75% by classical players. By setting up an instance of the game in real life and observing a success rate higher than 75%, we can empirically verify that whatever the players are doing can't be described in a classical model.

Interactive proofs are protocols between a verifier and a prover. In the usual setup, some string x and language L are known to both the verifier and the prover, and the prover's goal is to convince the verifier that $x \in L$. Typically, the verifier is bounded (say, to polynomial time) and cannot compute membership in the language on their own. The interactive proof protocol allows them to use the help of a cooperative-but-untrusted prover to decide whether $x \in L$. The probability that the verifier can be fooled into thinking $x \in L$ when $x \notin L$ is called the soundness error, and the probability that a good prover fails to convince the verifier that $x \in L$ when it is really true that $x \in L$ is called the completeness error. *Quantum* interactive proofs allow the verifier and prover to be quantum-mechanical systems and exchange quantum messages.

Nonlocal games and quantum interactive proofs are related, and one aspect common to their study is that of reducing error through repetition. An interactive proof system's

verifier would like to be able to reduce the chance they are fooled by a lucky prover by repeating the protocol. Similarly, in a physical experiment based on the CHSH game, many repetitions will be needed to statistically see the difference between a 75% and 85% success rate. It is well-known that error is reduced when the interactive proof protocols and nonlocal games are repeated sequentially in time. When they are repeated sequentially, the prover (or players) have no knowledge of what questions they will be asked in the future, so they cannot do anything to correlate the outcomes of the repetitions. But reducing error this way increases the number of back-and-forth trips between the prover and verifier (or players and referee), which is not always desirable. For example, some complexity classes like $\mathsf{QIP}[2]$ are based on constant-round interactive proofs. Sequential repetition cannot be used to reduce error for these classes, since it increases the number of rounds.

Another way to potentially reduce error is to repeat the protocols *in parallel*. That is, instead of repeating the protocol sequentially in time, we repeat it in space by sending multiple questions at once and receiving multiple answers back. Of course, now there is a possibility that the prover (or players) can somehow correlate their answers and do better than playing each repetition independently.

When discussing results about parallel repetition, it is useful to make the distinction between "parallel repetition theorems" and "concentration bounds." On the one hand, parallel repetition theorems say something akin to, "The probability of winning *all* of the parallel repetitions is less than..." and on the other hand, concentration bounds[1] say "The probability of winning at least this *fraction* of the parallel repetitions is less than..." The main difference is that parallel repetition theorems are only useful for reducing error when the original protocol's error is one-sided. For example, imagine modifying a proof system so that it is repeated in parallel and we require the prover to win 100% of the repetitions. In the case where the verifier should reject and the prover cannot convince the verifier to accept an individual reptition with 100% certainty, a parallel repetition theorem would imply that the chance that the verifier gets fooled decreases exponentially. But unless the prover can win with 100% certainty in the case where the verifier *should* accept, the probability that the verifier accepts when they should accept also decreases exponentially. For protocols with two-sided error, we need concentration bounds to show that error can be reduced. The idea is to modify the protocol so that it is repeated n times in parallel and requires that at least an $(\alpha + \beta)/2$ fraction of the repetitions accept, where, when we

[1] It should be noted that our use of the term "concentration bound" differs from the meaning it usually takes on in probability theory, i.e. an inequality bounding how concentrated a distribution is around some value. The two ideas are related to each other, but for the remainder of this document we will take "concentration bound" to mean results about parallel repetition taking the form described above, not the general concept from probability theory.

consider a single iteration of the protocol, β is the maximum probability that the verifier can be fooled, and α is the probability that a good prover can make the verifier accept when they should accept. The concentration bound is then used to show that, in the case where the verifier should reject, the parallel-repetition-with-threshold-check makes the error decrease as the number of repetitions grows larger.

In Chapter 2, we establish a lemma that relates parallel repetition theorems to concentration bounds. Informally, what this lemma says is that when we consider a sequence $X_1, \ldots X_n$ of probabilistic outcomes (0s or 1s), winning a certain threshold of the outcomes is at most as likely as winning *multiple* (slighly lower) thresholds in equal-sized random partitions of the outcomes. We use this lemma to essentially reduce the problem of winning *all* parallel repetitions to the problem of winning *a fraction* of the parallel repetitions, thus making it possible to turn a parallel repetition theorem into a concentration bound. However, the problem we are reducing is not the problem of winning all parallel repetitions of the original protocol, it is the problem of winning all parallel repetitions of *the protocol that requires you to win a certain fraction of some parallel repetitions of the original protocol.* So, for our technique to work we need to know a parallel repetition theorem that applies to the threshold-checking protocol. We will see some cases where this prevents our technique from working.

We combine this technique with known parallel repetition theorems in Chapter 3 to prove new concentration bounds for some classes of nonlocal games. Then, in Chapter 4, we prove that simple parallel-repetition-and-threshold-checking as we described above is sufficient to reduce error for quantum interactive proof protocols. This improves on the results of Molina [MP12] who had previously shown that parallel-repetition-and-threshold-checking works for α and β which are separated by a specific function. We close the gap between α and β and show that it works for an arbitrarily-small (but constant) difference.

The results from Chapters 2, 3 and 4 are expected to appear in a joint paper with John Watrous, which is currently in preparation.

Throughout this thesis, we will assume the reader is familiar with probability theory (random variables, linearity of expectation, etc.), linear algebra (finite-dimensional Hilbert spaces, operators, maps), and core concepts from quantum information theory (quantum registers, states, density operators, measurements, and channels). We recommend the texts [NC11] and [Wat18] as references.

We will begin by discussing probability theory in Chapter 2.

Chapter 2

Probability Theory

This chapter introduces some facts from probability theory and proves a lemma which will be central to our results in Chapters 3 and 4.

One probability distribution that is important to the topic of parallel repetition and concentration bounds is the binomial distribution. With parameters $n \in \mathbb{N}$ and $p \in [0, 1]$, the binomial distribution is the distribution over the number of successful outcomes of n experiments which each have independent probability of success p. The binomial distribution is relevant to concentration bounds because if the prover plays each of n parallel repetitions independently, then the probability that they win any given repetition is some probability p, and the amount of the repetitions that are won follows a binomial distribution.

A fact about the binomial distribution that we will use in later chapters is that the probability that there are fewer than k successful outcomes can be upper-bounded using the Chernoff bound,

$$\Pr[X \leq k] \leq \exp\left(-\frac{1}{2p}\frac{(np-k)^2}{n}\right), \tag{2.1}$$

where X is a random variable representing the number of successful outcomes.

Another distribution that will be important for proving our first lemma is the hypergeometric distribution. Whereas the binomial distribution can be understood as describing the number of green balls obtained by making n draws *with replacement* from a sac of N balls, pN of which are green and $(1-p)N$ of which are red, the hypergeometric distribution describes the expected number of green balls when drawing n balls from the same sac

without replacement. The bound on the hypergeometric distribution corresponding to the one we just gave for the binomial distribution is,

$$\Pr[X \leq (p-t)n] \leq \exp\bigl(-2t^2 n\bigr), \tag{2.2}$$

where X is the number of green balls drawn, n is the number of balls drawn, and p is the fraction of balls that are green [Hoe63].

We will use this bound on the hypergeometric distribution in the proof of the following lemma. The essense of the idea is as follows. Imagine that a quantum interactive proof prover (or a set of nonlocal game players) has won an $\alpha + \epsilon$ fraction of mn parallel repetitions. We show that if the parallel repetition outcomes are randomly assigned into n groups of m, it's likely for all of the groups to contain a fraction at least α of winning outcomes. The hypergeometric distribution comes in because each m-sized outcome group can be understood as being drawn without-replacement from sac of all mn outcomes.

In this document, the notation $[n]$ means the set $\{1, 2, \ldots, n\}$.

Lemma 2.0.1. *Let $X_1, \ldots, X_{mn}$ be binary-valued random variables. Let $\sigma : [m] \times [n] \to [mn]$ be a function chosen uniformly at random from the set of all bijections from $[m] \times [n]$ to $[mn]$. Then for any $\alpha \geq 0$ and any $\epsilon > 0$,*

$$(1 - ne^{-2\epsilon^2 m}) \Pr\left[\frac{1}{mn} \sum_{j \in [mn]} X_j \geq \alpha + \epsilon\right] \leq \Pr\left[\frac{1}{m} \sum_{j \in [m]} X_{\sigma(j,\ell)} \geq \alpha \text{ for all } \ell \in [n]\right]. \tag{2.3}$$

Proof. If $\alpha + \epsilon > 1$, the probability on the left-hand side is zero and the lemma is trivially true, so we can assume $\alpha + \epsilon \leq 1$ which implies $\alpha < 1$ and $\epsilon \leq 1$. Call the event that

$$\frac{1}{mn} \sum_{j \in [mn]} X_j \geq \alpha + \epsilon \tag{2.4}$$

occurs M. For an $\ell \in [n]$, call the event that

$$\frac{1}{m} \sum_{j \in [m]} X_{\sigma(j,\ell)} \geq \alpha \tag{2.5}$$

occurs C_ℓ. Then,

$$\begin{aligned}
&\Pr\left[\frac{1}{m}\sum_{j\in[m]} X_{\sigma(j,\ell)} \geq \alpha \text{ for all } \ell \in [n]\right] \\
&\qquad = \Pr[C_\ell \text{ for all } \ell] && (2.6)\\
&\qquad = \Pr[M]\Pr[C_\ell \text{ for all } \ell|M] + \Pr\left[\bar{M}\right]\Pr\left[C_\ell \text{ for all } \ell|\bar{M}\right] && (2.7)\\
&\qquad \geq \Pr[M]\Pr[C_\ell \text{ for all } \ell|M] && (2.8)\\
&\qquad = \Pr[M](1 - \Pr\left[\bar{C}_\ell \text{ for some } \ell|M\right]) && (2.9)\\
&\qquad \geq \Pr[M](1 - \sum_{\ell\in[n]} \Pr\left[\bar{C}_\ell|M\right]). && (2.10)
\end{aligned}$$

The last line follows from the union bound.

For any ℓ, we can interpret the event $\bar{C}_\ell|M$ as follows. That M occured means that at least $(\alpha+\epsilon)mn$ of the values of $X_1,\ldots,X_{mn}$ are 1. Then, because σ is a random bijection, we can think about C_ℓ as follows. We draw m bits at random from the multiset $\{X_1,\ldots X_{mn}\}$ without replacement, and C_ℓ corresponds to drawing at least αm 1s. So, $\bar{C}_\ell$ corresponds to drawing fewer than αm 1s.

Using the bound for the hypergometric distribution we gave above, replacing n with m, p with $\alpha+\epsilon$, and t with ϵ, we get,

$$\Pr\left[\bar{C}_\ell|M\right] \leq e^{-2\epsilon^2 m}. \tag{2.11}$$

So, using this fact to continue the sequence of inequalities above, we have that,

$$\Pr\left[\frac{1}{m}\sum_{j\in[m]} X_{\sigma(j,\ell)} \geq \alpha \text{ for all } \ell \in [n]\right] \geq \Pr[M](1 - ne^{-2\epsilon^2 m}), \tag{2.12}$$

which completes the proof. □

We use this lemma to prove concentration bounds for nonlocal games in Chapter 3 and an error-reduction theorem for quantum interactive proofs in Chapter 4. It is worth noting that Impagliazzo and Kabanets have also also connected parallel repetition theorems (Direct Product Theorems, in their language) with concentration bounds (Threshold Direct Product Theorems, in their language), using a different technique [IK10]. They show that when $X_1,\ldots,X_n$ are binary-valued random variables, if there is some $0 \leq \delta \leq 1$ such that for all subsets $S \subseteq [n]$,

$$\Pr[\wedge_{i\in S} X_i = 1] \leq \delta^{|S|}, \tag{2.13}$$

then the sum $X_1 + \cdots + X_n$ is likely to be near its expected value. The way this gets used is that a good enough parallel repetition theorem guarantees condition (2.13) is satisfied by all subsets S, and then the fact that the sum likely falls near its expected value implies a concentration bound. We did not investigate if using Impagliazzo's and Kabanets's technique instead of Lemma 2.0.1 would give better concentration bounds than the ones we are about to prove. Lemma 2.0.1 has the advantage that it can be used with parallel repetition theorems that are too weak to show that all subsets of the random variables satisfy condition (2.13). A disadvantage, as we will see in the next chapter, is that using Lemma 2.0.1 to prove a concentration bound for a special class of games will require a parallel repetition theorem for a game that is potentially outside of that special class.

In the following chapters, we will also make use of Markov's inequality, which says that for a random variable X,

$$\Pr[X \geq a] \leq \frac{\mathrm{E}[X]}{a}, \tag{2.14}$$

where $\mathrm{E}[X]$ is the expectation value of X.

Chapter 3

Concentration Bounds for Nonlocal Games

3.1 Introduction

A nonlocal game is a game played between k players and a referee. According to some probability distribution, the referee samples k questions, one for each player, and sends them to the players. Each player responds back to the referee with an answer. The players win the game if the questions and answers satisfy some predicate (which is known to all of the players).

Of course, if the players are allowed to communicate with each other, they can look at each others' questions and, since they know the predicate that decides whether or not they win, they can all agree upon a winning set of responses and win with certainty (as long as there *are* winning answers for the questions they were given). So, we are interested in what happens when we restrict the players so that they cannot communicate.

There are several different ways to model the idea that the players cannot communicate. The most basic way is to model the players as classical probabilistic algorithms, so that each player's answer is an independent random function of the question they are given. This model is quite easy to analyze. If $Q_1, \ldots Q_k$ are the sets of questions each of the k players could be asked, $A_1, \ldots, A_k$ are the sets of possible answers each player can give, $\mu(x_1, x_2, \ldots, x_k)$ for $(x_1, x_2, \ldots, x_k) \in Q_1 \times Q_2 \times \cdots \times Q_k$ is the probability that the referee selects the questions $x_1, \ldots, x_k$, and $V : Q_1 \times \cdots Q_k \times A_1 \times \cdots A_k \to \{0, 1\}$ is the predicate that decides whether or not the players win, then the probability that the players win the

game is exactly

$$\sum_{x_1,\ldots,x_k \in Q_1 \times \cdots \times Q_k} \mu(x_1, \ldots, x_k) \sum_{\substack{a_1,\ldots,a_k \in A_1 \times \cdots \times A_k \\ V(x_1,\ldots,x_k,a_1,\ldots,a_k)=1}} \prod_{j=1}^{k} p_j(x_j, a_j), \tag{3.1}$$

where $p_j(x, a)$ is the probability player j outputs a on input x.

A more interesting model allows the players to communicate before the game begins, so that they can agree upon a random string beforehand and use shared randomness to gain an advantage at winning the game. Another model allows the players to exchange quantum messages before the games begin, allowing them to take advantage of entanglement. These two different models are important for understanding the properties of entanglement. For example, in the well-known CHSH game, two players are given random bits x_1 and x_2 and they win if they answer a_1 and a_2 such that $a_1 \oplus a_2 = x_1 \cdot x_2$. CHSH can be won with at most 75% probability in the shared-randomness model, and $\cos^2(\pi/8) \approx 85\%$ in the shared-entanglement model. This is useful because if the referee repeats the game many times and observes the players consistently winning more than 75% of the time, they can conclude that the players are either entangled or doing something more than just using shared randomness (e.g. communicating). A central research topic in the study of nonlocal games is understanding which kinds of games can be won more easily by players who share entanglement.

Nonlocal games are also related to multi-prover interactive proof systems. In a multi-prover interactive proof system, there is a string x known to a verifier and multiple non-communicating provers. The goal of the proof system is for the verifier to be convinced that x is a member of some language L, when the verifier is not powerful enough to discover that fact for themselves. If we consider 1-round proof protocols, where the verifier sends messages to all of the provers and then bases their decision on the responses they receive back, then for a fixed string x and a fixed verifier, the situation can be seen as a nonlocal game where the provers are the players and the verifier is the referee.

Just like how with the CHSH game the referee would like to repeat the game many times to gain more confidence that the players are entangled, a multi-prover proof verifier would like to repeat the protocol multiple times to gain more confidence that they are not being fooled by lucky provers. If a nonlocal game is repeated sequentially, then each run of the game is independent and so, as long as the game cannot be won with certainty, the probability that the players win *all* of the repetitions decreases exponentially. But repeating the game sequentially increases the number of round trips between the players and the referee. An interesting question is whether or not it is safe to repeat the game in

parallel, by sending n questions from n repetitions of the game to the players all at once. Against two players with shared randomness, Raz's parallel repetition theorem implies that the probability of winning all the repeated games decreases exponentially in the number of parallel repetitions [Raz98]. However, for two players who share entanglement, it is only known that the probability of winning all of the repeated games decreases inverse-polynomially in the number of parallel repetitions [Yue16], except for some special kinds of games like XOR games where exponentially-decaying bounds are known [CSUU08].

These "parallel repetition theorems" upper-bound the probability of winning *all* of the parallel-repeated instances of the game. It is also useful to have an upper bound on the probability of winning a certain fraction of the games. We call upper bounds on the probability of winning fractions of parallel-repeated games "concentration bounds." For example, when running the CHSH game in parallel, in order to know that the players are doing something more powerful than a shared randomness strategy, we would like to check that they are winning some fraction above 75%, and know that as the number of parallel repetitions increases, the probability that they will win at least that fraction decreases exponentially. From the work of Rao, we know that such a concentration bound holds for shared-randomness players [Rao11]. No exponentially-decaying concentration bound is known for entanglement-sharing players, since by setting the fraction to 1, it would imply an exponentially decaying parallel repetition theorem (which would be better than the best known parallel repetition theorem for entangled players [Yue16]). Against non-signalling players, who are only constrained so that their behaviour does not imply communication, exponentially-decaying parallel repetition theorems and concentration bounds are known [LW15].

In this chapter, we develop a technique for converting parallel repetition theorems into concentration bounds. The technique works against both shared-randomness and entanglement-sharing players. For some kinds of games with known exponentially-decaying parallel repetition theorems, we use our technique to prove exponentially-decaying concentration bounds.

Of course, concentration bounds for some kinds of games have already been proven. For example, an exponentially-decaying concentration bound for XOR games has been established using methods similar to the alternative to Lemma 2.0.1 (due to Impagliazzo and Kabanets) that we discussed in Chapter 2 [Ung09]. To the best of our knowledge, all of the concentration bounds we prove in this chapter are new.

3.2 Definitions for Nonlocal Games

We begin by precisely defining what a k-player nonlocal game is, and what we mean by a (classical or quantum) strategy for playing the game.

Definition 3.2.1 (k-Player Nonlocal Games)**.** For integers $k \geq 2$, a k-player nonlocal game G is a collection of question sets $Q_1^G, \dots Q_k^G$ (all finite and nonempty), a collection of answer sets $A_1^G, \dots, A_k^G$ (all finite and nonempty), a probability distribution μ^G over $Q_1^G \times \cdots \times Q_k^G$, and a verification function $V^G : Q_1^G \times \cdots \times Q_k^G \times A_1^G \times \cdots \times A_k^G \to \{0, 1\}$.

Operationally, a k-player nonlocal game G is played when a referee samples questions for each player according to μ^G, sends each question to the corresponding player, then receives back answers from each player. The players win if the verification function outputs 1 and they lose if not. If the players are allowed to communicate with each other, then they can always win as long as there is at least one set of winning answers for the questions they were given. To make these games "nonlocal", we restrict the players to strategies where they are not allowed to communicate, including the special cases of the players being classical with shared randomness, or quantum with shared entanglement.

An (unconstrained) strategy for a nonlocal game defines a probability distribution over the player's answers for all possible combinations of questions they can be given.

Definition 3.2.2 (Strategy for a k-Player Nonlocal Game)**.** A strategy S for a k-player nonlocal game G is a function which on input $x \in Q_1^G \times \cdots \times Q_k^G$ outputs a probability distribution over $A_1^G \times \cdots \times A_k^G$.

A classical strategy is one that can be implemented by classical players using shared randomness.

Definition 3.2.3 (Classical Strategy for a k-Player Nonlocal Game)**.** A strategy S for a k-player nonlocal game G is a classical strategy if there exist finite and nonempty sets $X_1, \dots X_k$, a probability distribution χ over $X_1 \times \cdots \times X_k$ and functions $p_1, \dots, p_k$ where

$$p_j : Q_j^G \times X_j \to A_j^G, \tag{3.2}$$

such that for all $(x_1, \dots, x_k) \in Q_1^G \times \cdots \times Q_k^G$ and all $(a_1, \dots a_k) \in A_1^G \times \cdots \times A_k^G$,

$$S_{x_1,\dots,x_k}(a_1, \dots, a_k) = \Pr[p_j(x_j, s_j) = a_j \text{ for all integers } 1 \leq j \leq k], \tag{3.3}$$

where $(s_1, \dots, s_k)$ is sampled according to χ.

In other words, a strategy S is classical when the distributions over the answers it defines can be implemented by players who each start in a random state (which may be correlated with the other players' starting states) and whose answers are determined by their starting state and the question they receive.

A quantum strategy is one where the players are allowed to share entanglement before the game begins.

Definition 3.2.4 (Quantum Strategy for a k-Player Nonlocal Game)**.** Let G be a k-player nonlocal game. For all integers $1 \leq j \leq k$ and all finite-dimensional Hilbert spaces $\mathcal{X}$, let

$$M_j^G(\mathcal{X}) = \{\nu : A_j^G \to \text{Pos}(\mathcal{X}) \mid \sum_{x \in A_j^G} \nu(x) = \mathbb{1}\}, \tag{3.4}$$

where $\text{Pos}(\mathcal{X})$ is the set of all positive semidefinite operators over $\mathcal{X}$. In other words, $M_j^G(\mathcal{X})$ is the set of all quantum measurements on states in the space $\mathcal{X}$ with possible outcomes A_j^G.

A strategy S for G is a quantum strategy if there exist finite and nonempty sets $X_1, \ldots, X_k$, a quantum state $|\psi\rangle \in \mathbb{C}^{|X_1|} \otimes \cdots \otimes \mathbb{C}^{|X_k|}$, and functions $p_1, \ldots, p_k$, where

$$p_j : Q_j^G \to M_j^G(\mathbb{C}^{|X_j|}) \tag{3.5}$$

such that for all $(x_1, \ldots, x_k) \in Q_1^G \times \cdots \times Q_k^G$ and all $(a_1, \ldots, a_k) \in A_1^G \times \cdots \times A_k^G$,

$$S_{x_1,\ldots,x_k}(a_1, \ldots, a_k) = \langle\psi| \left(\bigotimes_{j=1}^{k} \nu_j^{x_j}(a_j) \right) |\psi\rangle , \tag{3.6}$$

where $\nu_j^{x_j} = p_j(x_j)$.

In other words, S is a quantum strategy when it can be implemented by players who start in some possibly-entangled state $|\psi\rangle$ and each arrive at their answer by performing a question-dependent measurement on their part of $|\psi\rangle$.

Quantum strategies are at least as powerful as classical strategies, since if χ is a classical strategy's distribution over the players' possible starting states $X_1 \times \cdots \times X_k$, then a quantum strategy can simulate this by using the state

$$|\psi\rangle = \sum_{(s_1,\ldots,s_k) \in X_1 \times \cdots \times X_k} \sqrt{\chi(s_1, \ldots, s_k)}\, |s_1, \ldots, s_k\rangle , \tag{3.7}$$

and then using measurements which are equivalent to performing a measurement in the computational basis and then determining the answer from the question and the measurement result the same way as the classical strategy does.

Given a strategy S for a game G, its "value" for G is the probability that it wins G.

Definition 3.2.5. Let S be a strategy for a game k-player nonlocal game G. Then,

$$\mathrm{val}(S,G) = \sum_{(x_1,\ldots,x_k)\in Q_1^G\times\cdots\times Q_k^G} \mu^G(x_1,\ldots,x_k) \sum_{\substack{(a_1,\ldots,a_k)\in A_1^G\times\cdots A_k^G \\ V^G(x_1,\ldots,x_k,a_1\ldots,a_k)=1}} S_{x_1,\ldots x_k}(a_1,\ldots,a_k), \quad (3.8)$$

or, equivalently,

$$\mathrm{val}(S,G) = \Pr\big[V^G(X,A)=1\big], \quad (3.9)$$

where X is sampled according to μ^G and then A is sampled according to S_X.

Given some nonlocal game G, we are interested in the maximum probability it can be won, called the "value" of G. When we maximize over classical strategies, we call the maximum probability of success the "classical value" and when we maximize over quantum strategies we call the maximum probability of success the "quantum value."

Definition 3.2.6 (The Classical Value of a k-Player Nonlocal Game)**.** The classical value $\mathrm{val_c}(G)$ of G is,

$$\mathrm{val_c}(G) = \max_S\{\mathrm{val}(S,G)\}, \quad (3.10)$$

where the maximum is taken over all possible classical strategies S for G.

It is safe to take the maximum here, rather than the supremum, because it is known that the optimum can always be achieved by a deterministic strategy, of which there are finitely many.

Definition 3.2.7 (The Quantum Value of a k-Player Nonlocal Game)**.** The quantum value $\mathrm{val_q}(G)$ of G is,

$$\mathrm{val_q}(G) = \sup_S\{\mathrm{val}(S,G)\}, \quad (3.11)$$

where the supremum is taken over all possible quantum strategies S for G.

It is known that the supremum is necessary here [Slo17].

Given a nonlocal game G, we are interested in discussing games which are constructed by repeating G in parallel. We start by defining what it means to play two games G_1 and G_2 in parallel.

Definition 3.2.8 (Parallel Play)**.** Given two k-player nonlocal games G_1 and G_2, a new game $G_1 \otimes G_2$ can be created by playing G_1 and G_2 in parallel. This game is defined by, for all $1 \leq i \leq k$,

$$Q_i^{G_1 \otimes G_2} = Q_i^{G_1} \times Q_i^{G_2} \tag{3.12}$$

$$A_i^{G_1 \otimes G_2} = A_i^{G_1} \times A_i^{G_2}, \tag{3.13}$$

and for all $((x_1, y_1), \ldots, (x_k, y_k)) \in Q_i^{G_1 \otimes G_2}$ and all $((a_1, b_1), \ldots, (a_k, b_k)) \in A_i^{G_1 \otimes G_2}$,

$$\mu^{G_1 \otimes G_2}((x_1, y_1), \ldots, (x_k, y_k)) = \mu^{G_1}(x_1, \ldots, x_k)\mu^{G_2}(y_1, \ldots, y_k), \tag{3.14}$$

which says that the referee chooses questions for G_1 and G_2 independently and,

$$\begin{aligned} V^{G_1 \otimes G_2}((x_1, y_1), \ldots, (x_k, y_k), (a_1, b_1), \ldots, (a_k, b_k)) \\ = V^{G_1}(x_1, \ldots, x_k, a_1, \ldots, a_k) \\ \wedge V^{G_2}(y_1, \ldots, y_k, b_1, \ldots, b_k) \end{aligned} \tag{3.15}$$

which says that the players win $G_1 \otimes G_2$ if and only if they win both G_1 and G_2.

In other words, each of the players get sent questions from G_1 and G_2, respond with answers for G_1 and G_2, and they win if their answers satisfy both G_1's and G_2's verification function.

Parallel repetition theorems upper-bound the probability of winning all of the parallel repetitions of a game G. In other words, parallel repetition theorems upper-bound the probability of winning the parallel repetition game based on G, which is defined as follows.

Definition 3.2.9 (Parallel Repetition Games)**.** Given a k-player nonlocal game G we can construct the parallel-repetition game $G^{\otimes n}$ for some $n \geq 1$ by repeating G in parallel n times. This is defined by $G^{\otimes n} = G \otimes G^{\otimes n-1}$ and $G^{\otimes 1} = G$.

We are interested in concentration bounds, which upper-bound the probability of winning a certain fraction of the parallel repetitions of some game G. In other words, a concentration bound upper-bounds the probability of winning the threshold game based on G, which is defined as follows.

Definition 3.2.10 (Threshold Games)**.** Given a k-player nonlocal game G, we can construct the threshold game $\mathcal{T}(G, \alpha, n)$ for some $n \geq 1$ and $\alpha \geq 0$ by repeating G in parallel n times and requiring that at least αn of the repeated games be won. This is defined the same as $G^{\otimes n}$, except with a different verification function,

$$V^{\mathcal{T}(G,n,\alpha)}(x_1, \ldots, x_k, a_1, \ldots, a_k) = \begin{cases} 1 & \text{if } \sum_{j=1}^{n} V^G(x_1^{(j)}, \ldots, x_k^{(j)}, a_1^{(j)}, \ldots, a_k^{(j)}) \geq \alpha n \\ 0 & \text{otherwise,} \end{cases} \tag{3.16}$$

where $x_i = (x_i^{(1)}, \ldots, x_i^{(n)})$ and $a_i = (a_i^{(1)}, \ldots, a_i^{(n)})$.

3.3 Deriving Concentration Bounds from Parallel Repetition Theorems

In this section, we establish lemmas which will be useful for deriving concentration bounds from parallel repetition theorems. The technique we use is based on a technique Jain et al. have used to reduce the soundness error of a two-message quantum interactive proof [JUW09].

Using Lemma 2.0.1 we can obtain the following result, which upper-bounds the value of a threshold game based on a game G.

Lemma 3.3.1. *Let G be any k-player game, and let $m, n \in \mathbb{N}$, and $\epsilon > 0$. Then,*

$$(1 - ne^{-2\epsilon^2 m/4}) \operatorname{val}_{\mathrm{q}}(\mathcal{T}(G, \operatorname{val}_{\mathrm{q}}(G) + \epsilon, mn)) \leq \operatorname{val}_{\mathrm{q}}(C^{\otimes n}), \tag{3.17}$$

where $C = \mathcal{T}(G, \operatorname{val}_{\mathrm{q}}(G) + \epsilon/2, m)$.

Proof. Let $v = \operatorname{val}_{\mathrm{q}}(G)$ and let S be any quantum strategy for $B = \mathcal{T}(G, v + \epsilon, mn)$. When the game B is played, the referee samples kmn questions according to μ^B. For all integers $1 \leq i \leq k$ and $1 \leq j \leq mn$, let $q_j^{(i)}$ be a random variable taking on the value of the question sent to player i for the j-th parallel repetition of G. After the questions are sampled according to μ^B, the players' answers are sampled according to S_Q where Q are the questions that were sampled. For all integers $1 \leq i \leq k$ and $1 \leq j \leq mn$, let $a_j^{(i)}$ be a random variable taking on the value of the answer player j chooses for the i-th parallel repetition of G. Now, for all integers $1 \leq j \leq mn$, define

$$X_j = V^G(q_j^{(1)}, \ldots, q_j^{(k)}, a_j^{(1)}, \ldots, a_j^{(k)}). \tag{3.18}$$

The probability that S wins B is equal to,

$$\Pr\left[\frac{1}{mn}\sum_{j\in[mn]} X_j \geq v+\epsilon\right]. \tag{3.19}$$

Now consider the game $C^{\otimes n}$. By the definition of $C^{\otimes n}$, this game's questions are sampled identically to B's, except that they are organized into n groups of m. For any bijection $\sigma : [m]\times[n] \to [mn]$, let S^σ be the strategy where the players follow strategy S, treating the question that comes in on index (j,ℓ) for $1 \leq j \leq m$ and $1 \leq \ell \leq n$ as if it came in on index $\sigma(j,\ell)$ while playing B. The players win the (j,ℓ)-th parallel repetition of G if and only if strategy S wins the $\sigma(j,\ell)$-th parallel repetition. So, the probability that S^σ wins $C^{\otimes n}$ is,

$$\Pr\left[\frac{1}{m}\sum_{j\in[m]} X_{\sigma(j,\ell)} \geq v+\frac{\epsilon}{2} \text{ for all } l\in[n]\right]. \tag{3.20}$$

Let S^R be the strategy where the players agree upon a bijection $\sigma : [m]\times[n] \to [mn]$ chosen uniformly at random from the set of all such bijections and then follow strategy S^σ. S^R is still a quantum strategy, since all of the S^σ strategies are quantum and the players can use entanglement to agree upon a random σ. The probability that S^R wins $C^{\otimes n}$ is the same expression,

$$\Pr\left[\frac{1}{m}\sum_{j\in[m]} X_{\sigma(i,\ell)} \geq v+\frac{\epsilon}{2} \text{ for all } l\in[n]\right], \tag{3.21}$$

except now σ is random. By replacing ϵ with $\epsilon/2$ in the statement of Lemma 2.0.1 and choosing $\alpha = v+\epsilon/2$, we obtain,

$$\begin{aligned}(1-ne^{-2\epsilon^2 m/4})\Pr\left[\frac{1}{mn}\sum_{j\in[mn]} X_j \geq v+\frac{\epsilon}{2}+\frac{\epsilon}{2}\right] \\ \leq \Pr\left[\frac{1}{m}\sum_{j\in[m]} X_{\sigma(j,\ell)} \geq v+\frac{\epsilon}{2} \text{ for all } \ell\in[n]\right]\end{aligned} \tag{3.22}$$

The second factor on the left-hand side is the probability that S wins the game B, and the right-hand side is the probability that S^R wins $C^{\otimes n}$. So, since S was arbitrary, and the probability that S^R wins $C^{\otimes n}$ is at most $\mathrm{val_q}(C^{\otimes n})$,

$$(1-ne^{-2\epsilon^2 m/4})\,\mathrm{val_q}(B) \leq \mathrm{val_q}(C^{\otimes n}), \tag{3.23}$$

which completes the proof. □

The only place we use entanglement in the proof is for sampling the random bijection σ. Obviously classical players can sample σ using shared randomness, so the lemma holds true when the quantum value is replaced with the classical value.

Lemma 3.3.2. *Let G be any k-player game, $m, n \in \mathbb{N}$, and $\epsilon > 0$. Then,*

$$(1 - ne^{-2\epsilon^2 m/4}) \operatorname{val}_{\mathrm{c}}(\mathcal{T}(G, \operatorname{val}_{\mathrm{c}}(G) + \epsilon, mn)) \leq \operatorname{val}_{\mathrm{c}}(C^{\otimes n}), \tag{3.24}$$

where $C = \mathcal{T}(G, \operatorname{val}_{\mathrm{c}}(G) + \epsilon/2, m)$.

An additional lemma will be helpful when applying Lemma 3.3.1.

Lemma 3.3.3. *Let G be any k-player game, $n \in \mathbb{N}$, and $\epsilon > 0$, and $\operatorname{val} \in \{\operatorname{val}_{\mathrm{c}}, \operatorname{val}_{\mathrm{q}}\}$. Then,*

$$\operatorname{val}(\mathcal{T}(G, \operatorname{val}(G) + \epsilon, n)) \leq \frac{\operatorname{val}(G)}{\operatorname{val}(G) + \epsilon}. \tag{3.25}$$

Proof. When $\operatorname{val}(G) + \epsilon > 1$, the left-hand side is zero and the inequality is trivially true, so we can assume $\operatorname{val}(G) + \epsilon \leq 1$. Let S be any strategy (classical if $\operatorname{val} = \operatorname{val}_{\mathrm{c}}$ or quantum if $\operatorname{val} = \operatorname{val}_{\mathrm{q}}$) for $\mathcal{T}(G, \operatorname{val}(G) + \epsilon, n)$, and let $X_1, \ldots, X_n$ be random variables such that X_j takes on the value 1 if S wins the game in index j and 0 otherwise. For all j, $\mathrm{E}[X_j] = \Pr[X_j = 1] \leq \operatorname{val}(G)$ since if this was false for some index j, then G could be won with probability higher than $\operatorname{val}(G)$ by providing the questions to the strategy S in index j and using shared randomness to sample questions for all of the other indices according to μ^G. So, using Markov's inequality,

$$\begin{aligned}
\operatorname{val}(\mathcal{T}(G, \operatorname{val}(G) + \epsilon, n)) &= \Pr\left[\frac{1}{n}\sum_{j=1}^{n} X_j \geq \operatorname{val}(G) + \epsilon\right] && (3.26)\\
&\leq \frac{\mathrm{E}[\frac{1}{n}\sum_{j=1}^{n} X_j]}{\operatorname{val}(G) + \epsilon} && (3.27)\\
&= \frac{\frac{1}{n}\sum_{j=1}^{n} \mathrm{E}[X_j]}{\operatorname{val}(G) + \epsilon} && (3.28)\\
&\leq \frac{\frac{n}{n}\operatorname{val}(G)}{\operatorname{val}(G) + \epsilon}, && (3.29)
\end{aligned}$$

which completes the proof. □

In the next section, we use these lemmas to prove concentration bounds for certain classes of nonlocal games that have known parallel repetition theorems.

3.4 Special Cases of Nonlocal Games

3.4.1 Games on the Uniform Distribution and Games with Complete Support

Chailloux and Scarpa have proven a parallel repetition theorem for two-player games on the uniform distribution [CS14].

Definition 3.4.1 (Games on the Uniform Distribution)**.** A two-player game G is on the uniform distribution if $\mu^G(x,y) = \frac{1}{|Q_1^G||Q_2^G|}$ for all $x \in Q_1^G$ and $y \in Q_2^G$.

Their parallel repetition theorem is:

Theorem 3.4.1 (Chailloux and Scarpa [CS14])**.** *For any game G on the uniform distribution such that $\mathrm{val}_{\mathrm{q}}(G) \leq 1-\epsilon$ for $\epsilon > 0$ and $Q_1^G = Q_2^G = I$ and $A_1^G = A_2^G = O$, we have that*

$$\mathrm{val}_{\mathrm{q}}(G^{\otimes n}) \leq (1-\epsilon^2)^{\Omega(\frac{n}{\log(|I||O|)} - |\log(\epsilon)|)}. \tag{3.30}$$

Given this parallel repetition theorem, we can derive the following concentration bound.

Theorem 3.4.2. *Let G be any game on the uniform distribution with, $Q_1^G = Q_2^G$, $A_1^G = A_2^G$ and $\mathrm{val}_{\mathrm{q}}(G) < 1$. Then for any $m, n \in \mathbb{N}$ and $\epsilon > 0$,*

$$(1 - ne^{-2\epsilon^2 m/4})\,\mathrm{val}_{\mathrm{q}}(\mathcal{T}(G, \mathrm{val}_{\mathrm{q}}(G)+\epsilon, mn)) \leq (1-\delta^2)^{\Omega\left(\frac{n}{m\log\left(|Q_1^G||A_1^G|\right)} - |\log(\delta)|\right)}, \tag{3.31}$$

where $\delta = 1 - \frac{\mathrm{val}_{\mathrm{q}}(G)}{\mathrm{val}_{\mathrm{q}}(G)+\epsilon/2}$.

Proof. Let G be any game on the uniform distribution with $Q_1^G = Q_2^G$ and $A_1^G = A_2^G$ and $\mathrm{val}_{\mathrm{q}}(G) < 1$. Let $\epsilon > 0$, $m, n \in \mathbb{N}$ and consider the game $C = \mathcal{T}(G, \mathrm{val}_{\mathrm{q}}(G) + \epsilon/2, m)$. By Lemma 3.3.3,

$$\begin{aligned}
\mathrm{val}_{\mathrm{q}}(C) &\leq \frac{\mathrm{val}_{\mathrm{q}}(G)}{\mathrm{val}_{\mathrm{q}}(G)+\epsilon/2} && (3.32)\\
&= 1 - \left(1 - \frac{\mathrm{val}_{\mathrm{q}}(G)}{\mathrm{val}_{\mathrm{q}}(G)+\epsilon/2}\right) && (3.33)\\
&= 1 - \delta. && (3.34)
\end{aligned}$$

By the definition of $\mathcal{T}$, the game C is also a game on the uniform distribution, with $Q_1^C = Q_2^C$, $A_1^C = A_2^C$, and $\delta > 0$ since $\epsilon > 0$, so by Theorem 3.4.1,

$$\mathrm{val}_{\mathrm{q}}(C^{\otimes n}) \leq (1-\delta^2)^{\Omega\left(\frac{n}{\log(|I||O|)} - |log(\delta)|\right)}, \tag{3.35}$$

where $|I| = |Q_1^C| = |Q_1^G|^m$, $|O| = |A_1^C| = |A_1^G|^m$.

Finally, by Lemma 3.3.1,

$$\begin{aligned}
(1 - ne^{-2\epsilon^2 m/4})\,\mathrm{val}_{\mathrm{q}}(\mathcal{T}(G, \mathrm{val}_{\mathrm{q}}(G)+\epsilon, mn)) &\leq \mathrm{val}_{\mathrm{q}}(\mathcal{T}(G, \mathrm{val}_{\mathrm{q}}(G)+\epsilon/2, m)^{\otimes n}) && (3.36)\\
&= \mathrm{val}_{\mathrm{q}}(C^{\otimes n}) && (3.37)\\
&\leq (1-\delta^2)^{\Omega\left(\frac{n}{\log\left(|Q_1^G|^m |A_1^G|^m\right)} - |\log(\delta)|\right)} && (3.38)\\
&= (1-\delta^2)^{\Omega\left(\frac{n}{m\log\left(|Q_1^G||A_1^G|\right)} - |\log(\delta)|\right)}. && (3.39)
\end{aligned}$$

□

For example, we can set $n = m^2$ to get an exponentially-decreasing concentration bound for games on the uniform distribution.

Corollary 3.4.3. *Let G be any game on the uniform distribution with, $Q_1^G = Q_2^G$, $A_1^G = A_2^G$ and $\mathrm{val}_{\mathrm{q}}(G) < 1$. Then for any $m \in \mathbb{N}$ and $\epsilon > 0$,*

$$(1 - m^2 e^{-2\epsilon^2 m/4})\,\mathrm{val}_{\mathrm{q}}(\mathcal{T}(G, \mathrm{val}_{\mathrm{q}}(G)+\epsilon, m^3)) \leq (1-\delta^2)^{\Omega\left(\frac{m}{\log\left(|Q_1^G||A_1^G|\right)} - |\log(\delta)|\right)}, \tag{3.40}$$

where $\delta = 1 - \frac{\mathrm{val}_{\mathrm{q}}(G)}{\mathrm{val}_{\mathrm{q}}(G)+\epsilon/2}$.

We can further simplify the bound by noting that $\lim_{m\to+\infty}(1 - m^2 e^{-2\epsilon^2 m/4}) = 0$ and

so for large enough m, $(1 - m^2 e^{-2\epsilon^2 m/4}) \geq \frac{1}{2}$. Also, if $\epsilon \leq 1$ then,

$$\delta = 1 - \frac{\mathrm{val_q}(G)}{\mathrm{val_q}(G) + \epsilon/2} \tag{3.41}$$

$$= \frac{\mathrm{val_q}(G) + \epsilon/2}{\mathrm{val_q}(G) + \epsilon/2} - \frac{\mathrm{val_q}(G)}{\mathrm{val_q}(G) + \epsilon/2} \tag{3.42}$$

$$= \frac{\mathrm{val_q}(G) + \epsilon/2 - \mathrm{val_q}(G)}{\mathrm{val_q}(G) + \epsilon/2} \tag{3.43}$$

$$= \frac{\epsilon/2}{\mathrm{val_q}(G) + \epsilon/2} \tag{3.44}$$

$$= \frac{\epsilon}{2\,\mathrm{val_q}(G) + \epsilon} \tag{3.45}$$

$$\geq \frac{\epsilon}{3}, \tag{3.46}$$

and so the following simplified corollary is true.

Corollary 3.4.4. *Let G be any game on the uniform distribution with, $Q_1^G = Q_2^G$, $A_1^G = A_2^G$ and $\mathrm{val_q}(G) < 1$. Then for large enough $m \in \mathbb{N}$ and any $0 < \epsilon \leq 1$,*

$$\mathrm{val_q}(\mathcal{T}(G, \mathrm{val_q}(G) + \epsilon, m^3)) \leq 2(1 - (\epsilon/3)^2)^{\Omega\left(\frac{m}{\log\left(|Q_1^G||A_1^G|\right)} - |\log(\epsilon/3)|\right)}. \tag{3.47}$$

The same simplifications can be made to all corollaries in this chapter, so we will make them implicitly from now on.

Chailloux and Scarpa have also proven parallel repetition theorems for games that have complete support [CS14].

Definition 3.4.2 (Games with Complete Support)**.** A two-player game G has complete support if $\mu^G(x, y) > 0$ for all $x \in Q_1^G$ and $y \in Q_2^G$.

Their parallel repetition theorem is as follows.

Theorem 3.4.5 (Chailloux and Scarpa [CS14])**.** *For any game G with complete support such that $\mathrm{val_q}(G) \leq 1 - \epsilon$ for $\epsilon > 0$ and $Q_1^G = Q_2^G = I$ and $A_1^G = A_2^G = O$, we have that*

$$\mathrm{val_q}(G^{\otimes n}) \leq (1 - \epsilon^2)^{\Omega\left(\frac{n}{Q\log(|I||O|)} - \frac{|\log(\epsilon)|}{Q}\right)}, \tag{3.48}$$

where $Q = \frac{|I|^2 \max_{x,y}(\mu^G(x,y))^2}{\min_{x,y}\mu^G(x,y)}$.

Unfortunately, applying the same technique using Theorem 3.4.5 does not seem to provide a useful concentration bound. We include the following proof as an example of our technique not working.

Theorem 3.4.6. *Let G be any game with complete support, $Q_1^G = Q_2^G$, $A_1^G = A_2^G$ and $\mathrm{val_q}(G) < 1$. Then for any $m, n \in \mathbb{N}$ and $\epsilon > 0$,*

$$(1 - ne^{-2\epsilon^2 m/4})\,\mathrm{val_q}(\mathcal{T}(G, \mathrm{val_q}(G) + \epsilon, mn)) \leq (1-\delta^2)^{\Omega\left(\frac{n}{Rm\log\left(|Q_1^G||A_1^G|\right)} - \frac{|\log(\delta)|}{R}\right)}, \tag{3.49}$$

where $\delta = 1 - \frac{\mathrm{val_q}(G)}{\mathrm{val_q}(G)+\epsilon/2}$ and,

$$R = \frac{|Q_1^G|^{2m}(\max_{x,y}(\mu^G(x,y))^2)^m}{(\min_{x,y}\mu^C(x,y))^m}. \tag{3.50}$$

$$\tag{3.51}$$

Proof. Let G be any game with complete support with $Q_1^G = Q_2^G$ and $A_1^G = A_2^G$ and $\mathrm{val_q}(G) < 1$. Let $\epsilon > 0$, $m, n \in \mathbb{N}$ and consider the game $C = \mathcal{T}(G, \mathrm{val_q}(G) + \epsilon/2, m)$. By Lemma 3.3.3,

$$\begin{aligned}
\mathrm{val_q}(C) &\leq \frac{\mathrm{val_q}(G)}{\mathrm{val_q}(G) + \epsilon/2} && (3.52)\\
&= 1 - (1 - \frac{\mathrm{val_q}(G)}{\mathrm{val_q}(G) + \epsilon/2}) && (3.53)\\
&= 1 - \delta. && (3.54)
\end{aligned}$$

By the definition of $\mathcal{T}$, the game C is also a game with complete support, $Q_1^C = Q_2^C$, $A_1^C = A_2^C$, and $\delta > 0$, so by Theorem 3.4.5,

$$\mathrm{val_q}(C^{\otimes n}) \leq (1-\delta^2)^{\Omega\left(\frac{n}{R\log(|I||O|)} - \frac{|log(\delta)|}{R}\right)}, \tag{3.55}$$

where $|I| = |Q_1^C| = |Q_1^G|^m$, $|O| = |A_1^C| = |A_1^G|^m$, and,

$$\begin{aligned}
R &= \frac{|I|^2 \max_{x,y}(\mu^C(x,y))^2}{\min_{x,y}\mu^C(x,y)} && (3.56)\\
&= \frac{|I|^2(\max_{x,y}(\mu^G(x,y))^2)^m}{(\min_{x,y}\mu^G(x,y))^m} && (3.57)\\
&= \frac{|Q_1^G|^{2m}(\max_{x,y}(\mu^G(x,y))^2)^m}{(\min_{x,y}\mu^G(x,y))^m}. && (3.58)
\end{aligned}$$

Finally, by Lemma 3.3.1,

$$\begin{aligned}
(1 - ne^{-2\epsilon^2 m/4})\, &\mathrm{val_q}(\mathcal{T}(G, \mathrm{val_q}(G) + \epsilon, mn)) \\
&\leq \mathrm{val_q}(\mathcal{T}(G, \mathrm{val_q}(G) + \epsilon/2, m)^{\otimes n}) && (3.59)\\
&= \mathrm{val_q}(C^{\otimes n}) && (3.60)\\
&\leq (1-\delta^2)^{\Omega\left(\frac{n}{R\log\left(|Q_1^G|^m |A_1^G|^m\right)} - \frac{|\log(\delta)|}{R}\right)} && (3.61)\\
&= (1-\delta^2)^{\Omega\left(\frac{n}{Rm\log\left(|Q_1^G||A_1^G|\right)} - \frac{|\log(\delta)|}{R}\right)}. && (3.62)
\end{aligned}$$

□

We won't bother giving a simplified corollary since this bound is of little use. The problem is that the size of the question set of the game $C^{\otimes m}$ grows exponentially in m. Because Q in Theorem 3.4.5 includes a factor of the question set size $|I|$, m needs to be logarithmic in n for the exponent to stay positive. When m is logarithmic in n, the factor $(1 - ne^{2\epsilon^2 m/4})$ does not exponentially approach 1.

3.4.2 k-Player Free Games

A free game is one where the questions sent to each player are chosen independently. Any free game can be converted into an equivalent game with complete support without loss of generality by removing the questions that are asked with probability 0, so they can be seen as a subset of the games with complete support.

Definition 3.4.3 (k-player Free Game)**.** A k-player nonlocal game G is a free game if for all $x_1 \in Q_1^G, \ldots, x_k \in Q_k^G$,

$$\mu^G(x_1, \ldots, x_k) = \mu_1^G(x_1) \cdots \mu_k^G(x_k), \tag{3.63}$$

where for each j, μ_j^G is the distribution on player j's questions induced by μ^G.

Chung et al. have proven parallel repetition theorems for k-player free games in both the quantum and classical settings. We begin with the classical setting.

Theorem 3.4.7 (Chung et al. [CWY15])**.** *Let G be a k-player free game with classical value $\mathrm{val_c}(G) \leq 1 - \epsilon$ for $\epsilon \geq 0$. Then,*

$$\mathrm{val_c}(G^{\otimes n}) \leq (1-\epsilon^2)^{\Omega\left(\frac{n}{sk}\right)}, \tag{3.64}$$

where $s = \max_{1\leq j\leq k} \log_2(|A_j^G|)$.

From this parallel repetition theorem we obtain the following concentration bound.

Theorem 3.4.8. *Let G be any k-player free game with $\mathrm{val_c}(G) < 1$. Then for any $m, n \in \mathbb{N}$ and $\epsilon > 0$,*

$$(1 - ne^{-2\epsilon^2 m/4}) \,\mathrm{val_c}(\mathcal{T}(G, \mathrm{val_c}(G) + \epsilon, mn)) \leq (1 - \delta^2)^{\Omega\left(\frac{n}{msk}\right)}, \tag{3.65}$$

where $s = \max_{1 \leq j \leq k} \log_2(|A_j^G|)$ and $\delta = 1 - \frac{\mathrm{val_c}(G)}{\mathrm{val_c}(G) + \epsilon/2}$.

Proof. Let $\epsilon > 0$, $m, n \in \mathbb{N}$ and consider the game $C = \mathcal{T}(G, \mathrm{val_c}(G) + \epsilon/2, m)$. By Lemma 3.3.3,

$$\begin{aligned} \mathrm{val_q}(C) &\leq \frac{\mathrm{val_c}(G)}{\mathrm{val_c}(G) + \epsilon/2} && (3.66)\\ &= 1 - (1 - \frac{\mathrm{val_c}(G)}{\mathrm{val_c}(G) + \epsilon/2}) && (3.67)\\ &= 1 - \delta. && (3.68) \end{aligned}$$

By the definition of $\mathcal{T}$, the game C is also a free game, and $\delta > 0$ since $\epsilon > 0$, so by Theorem 3.4.7,

$$\mathrm{val_c}(C^{\otimes n}) \leq (1 - \delta^2)^{\Omega\left(\frac{n}{tk}\right)}, \tag{3.69}$$

where $t = \max_{1 \leq j \leq k} \log_2(|A_j^C|) = m \max_{1 \leq j \leq k} \log_2(|A_j^G|)$ by definition of $\mathcal{T}$.

Finally, by Lemma 3.3.2,

$$\begin{aligned} &(1 - ne^{-2\epsilon^2 m/4}) \,\mathrm{val_c}(\mathcal{T}(G, \mathrm{val_c}(G) + \epsilon, mn)) && (3.70)\\ &\qquad \leq \mathrm{val_c}(\mathcal{T}(G, \mathrm{val_c}(G) + \epsilon/2, m)^{\otimes n})\\ &\qquad = \mathrm{val_c}(C^{\otimes n}) && (3.71)\\ &\qquad \leq (1 - \delta^2)^{\Omega\left(\frac{n}{tk}\right)} && (3.72)\\ &\qquad = (1 - \delta^2)^{\Omega\left(\frac{n}{msk}\right)}. && (3.73) \end{aligned}$$

□

By setting $n = m^2$ and applying the same simplifications we discussed in the previous section, we get an exponentially-decaying concentration bound for k-player free games.

Corollary 3.4.9. *Let G be any k-player free game with $\mathrm{val_c}(G) < 1$. Then for large enough $m \in \mathbb{N}$ and any $0 < \epsilon \leq 1$,*

$$\mathrm{val_c}(\mathcal{T}(G, \mathrm{val_c}(G) + \epsilon, m^3)) \leq 2(1 - (\epsilon/3)^2)^{\Omega\left(\frac{m}{sk}\right)}, \tag{3.74}$$

where $s = \max_{1 \leq j \leq k} \log_2(|A_j^G|)$.

Chung et al.'s parallel repetition theorem in the quantum setting is as follows.

Theorem 3.4.10 (Chung et al. [CWY15])**.** *Let G be a k-player free game with $\mathrm{val_q}(G) \leq 1 - \epsilon$ for $\epsilon \geq 0$. Then, for $n > sk^4 \log(k/\epsilon)/\epsilon^{3/2}$),*

$$\mathrm{val_q}(G^{\otimes n}) \leq (1 - \epsilon^{3/2})^{\Omega(n/k^4 s)} \tag{3.75}$$

where $s = \max_{1 \leq j \leq k} \log_2(|A_j^G|)$.

From this we derive the following concentration bound for the quantum setting.

Theorem 3.4.11. *Let G be any k-player free game with $\mathrm{val_q}(G) < 1$. Then for any $m, n \in \mathbb{N}$ and $\epsilon > 0$ where $n > msk^4 \log(k/\delta)/\delta^{3/2}$,*

$$(1 - ne^{-2\epsilon^2 m/4}) \, \mathrm{val_q}(\mathcal{T}(G, \mathrm{val_q}(G) + \epsilon, mn)) \leq (1 - \delta^{3/2})^{\Omega\left(\frac{n}{k^4 ms}\right)}, \tag{3.76}$$

where $s = \max_{1 \leq j \leq k} \log_2(|A_j^G|)$ and $\delta = 1 - \frac{\mathrm{val_q}(G)}{\mathrm{val_q}(G) + \epsilon/2}$.

Proof. Let G be any k-player free game with $\mathrm{val_q}(G) < 1$. Let $\epsilon > 0$, $m, n \in \mathbb{N}$ and consider the game $C = \mathcal{T}(G, \mathrm{val_q}(G) + \epsilon/2, m)$. By Lemma 3.3.3,

$$\mathrm{val_q}(C) \leq \frac{\mathrm{val_q}(G)}{\mathrm{val_q}(G) + \epsilon/2} \tag{3.77}$$

$$= 1 - (1 - \frac{\mathrm{val_q}(G)}{\mathrm{val_q}(G) + \epsilon/2}) \tag{3.78}$$

$$= 1 - \delta. \tag{3.79}$$

By the definition of $\mathcal{T}$, the game C is also a free game, and $\delta > 0$ since $\epsilon > 0$, so by Theorem 3.4.7, as long as $n > sk^4 \log(k/\delta)/\delta^{3/2}$,

$$\mathrm{val_q}(C^{\otimes n}) \leq (1 - \delta^{3/2})^{\Omega\left(\frac{n}{k^4 s}\right)}, \tag{3.80}$$

where $t = \max_{1\le j\le k} \log_2(|A_j^C|) = m \max_{1\le j\le k} \log_2(|A_j^G|)$ by definition of $\mathcal{T}$.

Finally, by Lemma 3.3.2,

$$(1 - ne^{-2\epsilon^2 m/4}) \operatorname{val}_{\mathrm{q}}(\mathcal{T}(G, \operatorname{val}_{\mathrm{q}}(G) + \epsilon, mn)) \le \operatorname{val}_{\mathrm{q}}(\mathcal{T}(G, \operatorname{val}_{\mathrm{q}}(G) + \epsilon/2, m)^{\otimes n}) \tag{3.81}$$

$$= \operatorname{val}_{\mathrm{q}}(C^{\otimes n}) \tag{3.82}$$

$$\le (1 - \delta^{3/2})^{\Omega\left(\frac{n}{k^4 t}\right)} \tag{3.83}$$

$$= (1 - \delta^{3/2})^{\Omega\left(\frac{n}{k^4 ms}\right)}. \tag{3.84}$$

□

We can simplify by setting $n = m^2$, applying the same simplifications as before and noting that with k and ϵ constant, $m^2 > msk^4(\log(k/\delta)/\delta^{3/2}$ is true for large enough m.

Corollary 3.4.12. *Let G be any k-player free game with $\operatorname{val}_{\mathrm{q}}(G) < 1$. Then for any $0 < \epsilon \le 1$ and large enough $m \in \mathbb{N}$,*

$$\operatorname{val}_{\mathrm{q}}(\mathcal{T}(G, \operatorname{val}_{\mathrm{q}}(G) + \epsilon, mn)) \le 2(1 - (\epsilon/3)^{3/2})^{\Omega\left(\frac{m}{k^4 s}\right)}, \tag{3.85}$$

where $s = \max_{1\le j\le k} \log_2(|A_j^G|)$.

In the same article, Chung et al. also study k-player games where the players are allowed to output quantum states, called classical-quantum (CQ) games. They prove a parallel repetition theorem for CQ games. We expect that a variant of the technique we are using here, based on Lemma 2.0.1, would provide concentration bounds for CQ games.

3.5 Future Work

There are several other types of games which have known parallel repetition theorems against quantum players, but we cannot derive concentration bounds for them using Lemma 3.3.1. A goal we leave for future work is to prove concentration bounds for these kinds of games, either by inventing a technique that is more generally applicable than Lemma 3.3.1, or by taking ad-hoc approach for each type of game. These game types are:

- *Projection Games and Unique Games.* A projection game is a two-player game where for every pair of questions and every answer from the first player, there is at most one

possible answer from the second player that would win the game. Unique games are projection games where there is always exactly one winning answer from the second player. Lemma 3.3.1 does not work for projection games or unique games, since the game C in the statement of the lemma is not any kind of game for which a parallel repetition theorem is known.

- *Anchored Games.* A two-player single-round game is said to be α-anchored if there is some subset $X_\perp$ of Alice's questions and some subset $Y_\perp$ of Bob's questions, such that both subsets have probability at least α, and conditioned on the questions coming from those sets, it looks like the referee is choosing the questions independently [BVY15]. Lemma 3.3.1 does not work for these games since although the game C will be anchored, if the original game is α-anchored, we can't assume that the C game will be more than α^{2m}-anchored. Since the C game's anchoring value is exponentially smaller, the bound provided by parallel repetition theorem for anchored games won't be strong enough for Lemma 3.3.1 to give a useful bound.

- *Other Types of Games.* We have not tried to apply Lemma 3.3.1 to all known parallel repetition theorems, for example the free games studied by Chung et al. where the players are allowed to output quantum states instead of classical messages [CWY15], or fortified games [Mos14]. Proving concentration bounds for game types not mentioned in this paper is left as future work.

3.6 Conclusion

The contributions of this chapter are:

- Lemma 3.3.1. If one is given a game G and knows a parallel repetition theorem that applies to threshold games constructed from G, then through Lemma 3.3.1 one can obtain a concentration bound for parallel repetitions of G.

- Concentration bounds for two-player games on the uniform distribution and k-player free games (against entangled players), obtained by combining known parallel repetition theorems with Lemma 3.3.1.

Chapter 4

Simple Parallel Repetition Reduces Error For Quantum Interactive Proofs

4.1 Introduction

In this chapter we develop a concentration bound for quantum interactive proofs and show that an interactive proof protocol's error rate can be reduced through parallel repetition, using techniques similar to the ones we used in the last chapter for nonlocal games.

In complexity theory, an interactive proof is a protocol between two parties: a verifier and a prover. Typically, there is some string x known to both parties, and the goal of the protocol is to convince the verifier that $x \in L$ where $L \subseteq \{0,1\}^*$ is a language. The verifier is computationally bounded, say to polynomial time in the size of x. The prover is unbounded, and thus, by answering the verifier's questions, the prover may be able to convince the verifier that $x \in L$ when the verifier cannot discover that fact on their own. During an execution of an interactive proof protocol,

1. The verifier sends a polynomial-in-$|x|$-sized message to the prover.
2. The prover sends a polynomial-in-$|x|$-sized response back to the verifier.
3. Steps (1) and (2) repeat some polynomial-in-$|x|$ number of times until finally the verifier outputs an "accept" or "reject" answer.

In order for the protocol to be useful for deciding a language L, it must satisfy two properties:

1. **Completeness.** There is an $\alpha > 0$ such that if $x \in L$, then there is at least one prover that can cause the verifier to accept with probability at least α.

2. **Soundness.** If $x \notin L$, then *no* unbounded prover can cause the verifier to output "yes" with probability greater than β, where $\beta < \alpha$.

The probability β that the verifier is fooled when $x \notin L$ is called the "soundness error" and the probability $1 - \alpha$ that the verifier is not convinced when $x \in L$ is called the "completeness error."

We say that there is an interactive proof system for a language L if there is a verifier for which the above properties hold. If we add the additional constraint that $\alpha \geq 2/3$ and $\beta \leq 1/3$, then in the case where the verifier is classical (i.e. is described by a classical algorithm with access to random bits) and runs in polynomial time, this defines the complexity class IP. When the messages are quantum states and the verifier is a polynomial-time quantum algorithm, it defines the complexity class QIP.

Interactive proof systems are powerful for expressing languages. For example, the class IP of languages with classical interactive proof systems is the same as the class PSPACE of languages where membership can be decided by polynomial-*space* algorithms [Sha92]. It is also true that $\mathsf{QIP} = \mathsf{PSPACE}$, but unlike in the classical case, we know that $\mathsf{QIP}[3] = \mathsf{QIP} = \mathsf{PSPACE}$, where $\mathsf{QIP}[3]$ is like QIP but restricted to interactive proof protocols of only 3 messages (beginning with a message from the prover) [JJUW10]. This motivates the study of classes like $\mathsf{QIP}[2]$ (quantum interactive proofs restricted to one message from the verifier and one response from the prover). When studying constant-round classes like $\mathsf{QIP}[2]$, it is useful to be able to reduce the soundness and completeness error without increasing the number of rounds.

One way to reduce the soundness and completness error of a quantum interactive proof system, which *does* increase the number of messages, is to repeat the protocol n times sequentially and accept if and only if that at least $\lfloor n(\alpha+\beta)/2 \rfloor$ of the repetitions accept. If the original protocol uses k messages, then the error-reduced protocol will use nk messages. Since the repetitions of the protocol come after each other, the verifier can be sure that the prover is playing each repetition independently, and so if the original protocol has soundness error at most β, then the probability the prover can win m of the n repetitions

is at most,

$$\sum_{j=m}^{n} \binom{n}{j} \beta^j (1-\beta)^{n-j}, \tag{4.1}$$

which decreases exponentially in n when $\lfloor m = n(\alpha+\beta)/2 \rfloor$.

Repeating the protocol in sequence n times like this has the disadvantage that the error-reduced protocol requires a factor of n more rounds than the original protocol. We would like to be able to decrease the soundness and completeness error of an interactive proof system by repeating it in parallel, rather than sequentially.

For classical interactive proofs, it is known that the error reduction from repeating an interactive proof protocol in parallel is the same as the error reduction from repeating it in sequence. However, the situation is different for quantum interactive proofs. For example, Molina found protocols for which provers can win some k out of n parallel repetitions with a higher likelihood than they could win k out of n sequential repetitions (for some values of k and n) [MP12]. From the same work of Molina's, we know that if the maximum probability that any prover can win a single instance of an interactive proof protocol is p, then the probability any prover can win at least k out of n parallel repetitions of the protocol is at most,

$$p^k \binom{n}{k}. \tag{4.2}$$

Using this upper bound, Molina shows that completeness and soundness error can be exponentially reduced by repeating the protocol in parallel, as long as α and β are separated by a certain function:

Theorem 4.1.1 (Molina [MP12])**.** *Let the parameters α and β for a quantum interactive proof system be constant real numbers, with $0 \leq \beta < 2^{-H(\alpha)} < \alpha \leq 1$, where $H(\alpha) = -\alpha \log_2(\alpha) - (1-\alpha)\log_2(1-\alpha)$. Then, a strategy based on parallel repetition followed by a threshold value computation will bring the soundness and completeness errors below $\epsilon > 0$ in $O(\log(\frac{1}{\epsilon}))$ repetitions.*

For valid values of α and β, Theorem 4.1.1 removes the need to use error-reduction techniques that are more complicated than simple parallel repetition, e.g. as are used in [JUW09] and [KW00].

In this chapter, we will remove the constraint that α and β have to be separated by $2^{-H(\alpha)}$, although we will require a log-factor more parallel repetitions to achieve the same reduction in error. This chapter's main theorem is:

Theorem 4.4.1. *Let $0 \leq 1 - \alpha < 1$ be the completeness error of a quantum interactive proof verifier, and let $0 \leq \beta < \alpha$ be the soundness error. Write $\alpha = \beta + \epsilon$ for $\epsilon > 0$. Then, by repeating the interactive proof system in parallel and checking for a threshold of at least $\beta + \epsilon/2$, both the completeness and soundness error can be reduced to at most $\delta > 0$ in $O(\log(1/\delta)\log(\log(1/\delta)))$ parallel repetitions.*

We will begin in the next section with formal definitions for interactive proof systems.

4.2 Definitions

We begin with the definition of a k-round interactive proof verifier. For all of the following definitions, let $\mathcal{Q} = \mathbb{C}^2$ be the state space of a single-qubit register. For any complex euclidean space $\mathcal{X}$, $\mathrm{D}(\mathcal{X})$ is the set of density operators on $\mathcal{X}$.

Definition 4.2.1. A k-round interactive proof verifier is a pair (V, p), where p is some positive integer-valued function and V is a function which on input $x \in \{0,1\}^*$ outputs $k+1$ quantum channels $V_1^x, \ldots, V_{k+1}^x$, where for all $x \in \{0,1\}^*$ and integers $1 \leq j \leq k+1$,

$$V_j^x : \mathrm{D}(\mathcal{Q}^{\otimes p(|x|)} \otimes \mathcal{Q}^{\otimes p(|x|)}) \to \mathrm{D}(\mathcal{Q}^{\otimes p(|x|)} \otimes \mathcal{Q}^{\otimes p(|x|)}). \tag{4.3}$$

In this definition, V is a function representing the actions of a verifier during an execution of the protocol. The channel V_1^x is used to prepare the verifier's initial state and the first message to the prover. The channels $V_2^x, \ldots, V_k^x$ are used to process the prover's response and prepare the next message to the prover, and the channel V_{k+1}^x is used to process the prover's last message and decide whether to accept or reject. The verifier's memory is of size $p(|x|)$, and the messages sent between the verifier and prover will all be of size $p(|x|)$. A more general definition would allow the size of the messages and the verifier's memory to all be different, but in terms of defining complexity classes like QIP, where we require p to be bounded by a polynomial, we haven't lost any generality.

Next, we define a prover who can interact with the verifier.

Definition 4.2.2. A prover compatible with a k-round interactive proof verifier (V, p) is a pair (P, f) where f is some positive integer-valued function and P is a function which on input $x \in \{0,1\}$ outputs k quantum channels $P_1^x, \ldots P_k^x$. For all integers $1 \leq j \leq k$ and $x \in \{0,1\}^*$,

$$P_j^x : \mathrm{D}(\mathcal{Q}^{\otimes p(|x|)} \otimes \mathcal{Q}^{\otimes f(|x|)}) \to \mathrm{D}(\mathcal{Q}^{\otimes p(|x|)} \otimes \mathcal{Q}^{\otimes f(|x|)}). \tag{4.4}$$

Given some verifier $\mathcal{V}$, prover $\mathcal{P}$, and string x, the interaction between $\mathcal{V}$ and $\mathcal{P}$ on x is a probabilistic process which results in either acceptance or rejection, according to the following definition.

Definition 4.2.3. Let $\mathcal{P} = (P, f)$ be a prover compatible with a k-round interactive proof verifier $\mathcal{V} = (V, p)$, and let $x \in \{0,1\}^*$. The result of $\mathcal{V}$ interacting with $\mathcal{P}$ on x is the result of the following process.

1. Prepare two registers X and Y of $p(|x|)$ qubits each in the state $|0\rangle^{\otimes p(|x|)}$, and apply V_x^1 to the register (X, Y).
2. Initialize an $f(|x|)$-qubit register Z to the state $|0\rangle^{\otimes f(|x|)}$.
3. For $j \in \{1, \ldots, k\}$ in sequence, apply P_j^x to the register (Y, Z), and then apply V_{j+1}^x to the register (X, Y).
4. Measure the first qubit of register X. If the outcome is 1 we say that the verifier accepts, otherwise we say that the verifier rejects.

In this definition, the register X holds the verifier's memory and Z holds the prover's memory. The register Y is sent back and forth between the verifier and the prover, containing either the message from the verifier to the prover or vice-versa.

Now, we can use this definition of an interaction between a verifier and a prover to say how good the verifier is at recognizing a language L in terms of completeness error and soundness error.

Definition 4.2.4. Let $\mathcal{V}$ be a verifier and let L be a language. We say that $\mathcal{V}$ has soundness error at most β for L if for all $x \notin L$ and all compatible provers $\mathcal{P}$, the probability that $\mathcal{V}$ accepts after an interaction with $\mathcal{P}$ on the string x is at most β. We say that $\mathcal{V}$ has completeness error at most $1 - \alpha$ for L if there exists a compatible prover $\mathcal{P}$ such that for all $x \in L$, the probability that $\mathcal{V}$ accepts after an interaction with P on the string x is at least α.

With the aim of reducing the completeness and soundness error, given a verifier $\mathcal{V}$, we can construct a new verifier which runs multiple instances of $\mathcal{V}$ in parallel and checks that the number of acceptances reaches a certain threshold.

Definition 4.2.5. Let $\mathcal{V} = (V, p)$ be a k-round verifier. Then for $n \in \mathbb{N}$ and $\gamma \geq 0$, $\mathcal{V}_\gamma^{\otimes n} = (C, f)$ is a verifier where $f(x) = 2np(x) + 1$, and C is a function based on V which, on input $x \in \{0,1\}^*$, outputs the channels $C_1^x, \ldots, C_{k+1}^x$, where for all $1 \leq j \leq k$,

$$C_j^x = \mathbb{1} \otimes (V_j^x)^{\otimes n}, \tag{4.5}$$

and,

$$C_{k+1}^x = T(\mathbb{1} \otimes (V_{k+1}^x)^{\otimes n}), \tag{4.6}$$

where the action of the channel T is as follows. Index the qubits of a $2np(|x|) + 1$ qubit register from from 0 up to $2np(x)$. Then T is the channel that measures the qubits in indices $\{2jp(|x|) + 1 \mid j \in \{0, \ldots n-1\}\}$ and sets the state of the first qubit to 1 if at least γn of the measurement results are 1, and to 0 if not.

In other words, the verifier $\mathcal{V}_\gamma^{\otimes n}$ runs $\mathcal{V}$ in parallel n times and will accept if and only if at least γn of the $\mathcal{V}$'s accept.

Molina's upper bound on the probability of winning k out of n parallel repetitions that we mentioned in the introduction can be restated in the language of our definitions by taking $k = \lceil \gamma n \rceil$.

Corollary 4.2.1 (of a result due to Molina [MP12])**.** *Let $\mathcal{V}$ be a verifier with soundness error at most β for L. Then $\mathcal{V}_\gamma^n$ has soundness error at most,*

$$\binom{n}{\lceil \gamma n \rceil} \beta^{\lceil \gamma n \rceil}, \tag{4.7}$$

for L.

In the next section, we prove a different upper bound, which we will use in the proof of our main theorem.

4.3 Bounding Soundness Error

The technique we will use to upper-bound the soundness error is very similar to the one we used to derive concentration bounds for two-player games in the last chapter. It is based on the technique for reducing error in two-message quantum interactive proofs [JUW09]. The following lemma can be seen as the quantum interactive proof analogue of Lemma 3.3.3.

Lemma 4.3.1. *Let $\mathcal{V} = (V, p)$ be a verifier with soundness error at most β for a language L. Let $n \in \mathbb{N}$ and $\epsilon > 0$. Then $\mathcal{V}_{\beta+\epsilon}^{\otimes n}$ has soundness error at most*

$$\frac{\beta}{\beta + \epsilon} \tag{4.8}$$

for L.

Proof. Let $\mathcal{P} = (P, f)$ be any prover compatible with $\mathcal{V}_{\beta+\epsilon}^{\otimes n} = (C, p)$ and let $x \notin L$. Let X, Y, and Z be quantum registers as defined in Definition 4.2.5 for an interaction between $\mathcal{P}$ and $\mathcal{V}_{\beta+\epsilon}^{\otimes n}$. Let $X_1, \ldots, X_n$ be random variables over $\{0, 1\}$ whose outcomes are defined as follows. Carry out the interaction between $\mathcal{P}$ and $\mathcal{V}$ on x as defined in Definition 4.2.5, and now index the qubits of register X from 0 up to $2p(|x|)$. For all integers $1 \leq j \leq n$, the random variable X_j takes on the value of the qubit in index $2jp(|x|) + 1$, which we know will be in a classical state from measurements in the definition of $\mathcal{V}_{\beta+\epsilon}^{\otimes n}$.

For all $1 \leq j \leq n$, $\mathrm{E}[X_j] = \Pr[X_j = 1] \leq \beta$, since if this was false then let j be an index for which it is false. We will show a contradiction. Let $\mathcal{Q} = (Q, g)$ be a prover compatible with V that works as follows on a string x. The prover's memory is of size $f(|x|) = f(|x|) + 2(n-1)p(|x|)$. The first $f(|x|)$ qubits are used to store the memory for a simulation of $\mathcal{P}$, and the remaining memory is divided into $(n-1)$ sections of $2p(|x|)$ qubits, which are used to store the memory of $(n-1)$ simulations of $\mathcal{V}$ and their output messages. In round i of the protocol, $\mathcal{Q}$ applies V_i^x to the sections of memory for each of the simulations of $\mathcal{V}$, and then applies P_i^x to (1) the memory set aside for the simulation of $\mathcal{P}$, (2) the message from the real $\mathcal{V}$, and (3) the output messages of the $(n-1)$ simulations of $\mathcal{V}$. The message from the real $\mathcal{V}$ is provided to the jth input index of P_i^x, and the messages from the $(n-1)$ simulations of $\mathcal{V}$ are provided to the other input indices by some arbitrary but fixed way. The jth index of the output of P_i^x is sent back to the real $\mathcal{V}$, and the other indexes of the output are sent back to their corresponding simulated $\mathcal{V}$'s. If $\Pr[X_j = 1] \leq \beta$ is false, then $\mathcal{Q}$ causes $\mathcal{V}$ to accept with probability greater than β on an input $x \notin L$, contradicting the fact that V has soundness error at most β for L.

So, using Markov's inequality, the probability that $\mathcal{V}_{\beta+\epsilon}^{\otimes n}$ accepts on input $x \notin L$ when

interacting with $\mathcal{P}$ is,

$$\Pr\left[\frac{1}{n}\sum_{j=1}^{n} X_j \geq \beta + \epsilon\right] \leq \frac{\mathrm{E}[\frac{1}{n}\sum_{j=1}^{n} X_j]}{\beta + \epsilon} \tag{4.9}$$

$$= \frac{\frac{1}{n}\sum_{j=1}^{n} \mathrm{E}[X_j]}{\beta + \epsilon} \tag{4.10}$$

$$\leq \frac{\frac{n}{n}\beta}{\beta + \epsilon}. \tag{4.11}$$

Since P and $x \notin L$ were arbitrary, $V_{\beta+\epsilon}^{\otimes n}$ has soundness error at most $\frac{\beta}{\beta+\epsilon}$ for L. □

The next theorem can be seen as the analogue of Lemma 3.3.1 for quantum interactive proofs. It establishes that repeating a verifier V in parallel and then checking for a threshold slightly greater than V's soundness error always reduces the soundness error exponentially in the number of repetitions.

Theorem 4.3.2. *Let $\mathcal{V}$ be a verifier with soundness error at most β for a language L. Let $m, n \in \mathbb{N}$ and $\epsilon > 0$. Then $V_{\beta+\epsilon}^{\otimes mn}$ has soundness error at most*

$$\frac{\left(\frac{\beta}{\beta+\epsilon/2}\right)^n}{1 - ne^{-2\epsilon^2 m/4}} \tag{4.12}$$

for L, as long as the denominator is positive.

Proof. Let $\mathcal{P}$ be any prover compatible with $\mathcal{V}_{\beta+\epsilon}^{\otimes mn}$, and let $x \notin L$. Define the binary-random variables $X_1, \ldots, X_{mn}$ as follows. Carry out the interaction between $\mathcal{P}$ and $\mathcal{V}_{\beta+\epsilon}^{\otimes mn}$ on x, and then for all integers $1 \leq j \leq n$, the random variable X_j takes on the value 1 if the jth parallel repetition of $\mathcal{V}$ accepts, and takes on the value 0 otherwise. The probability that $\mathcal{V}_{\beta+\epsilon}^{\otimes mn}$ accepts is,

$$\Pr\left[\frac{1}{n}\sum_{j=1}^{n} X_j \geq \beta + \epsilon\right]. \tag{4.13}$$

Now consider the verifiers $\mathcal{C} = \mathcal{V}_{\beta+\epsilon/2}^{m}$ and $\mathcal{E} = C_1^n$. By Lemma 4.3.1, $\mathcal{C}$ has soundness error at most $\frac{\beta}{\beta+\epsilon/2}$ for L. By Corollary 4.2.1 (setting $\gamma = 1$), $\mathcal{E}$ has soundness error at most $\left(\frac{\beta}{\beta+\epsilon/2}\right)^n$ for L. Let $\mathcal{Q}$ be the prover compatible with $\mathcal{E}$ defined as follows. At the

start of the interaction with $\mathcal{E}$, $\mathcal{Q}$ samples a bijection $\sigma : [m] \times [n] \to [mn]$ uniformly at random from the set of all such bijections. $\mathcal{Q}$ saves σ to its memory and also initializes memory for a simulated instance of $\mathcal{P}$ in its memory. Upon receipt of a message from $\mathcal{E}$, it treats it as mn registers indexed by pairs (i, j) for integers $1 \leq i \leq m$ and $1 \leq j \leq n$, where register (i, j) contains the message from the ith parallel repetition of $\mathcal{V}$ within the jth parallel repetition of $\mathcal{C}$. We can think of $\mathcal{P}$ as accepting mn messages from mn parallel repetitions of $\mathcal{V}$ indexed from 1 up to mn. $\mathcal{Q}$ provides its input messages to $\mathcal{P}$, so that the (i, j)th register of $\mathcal{Q}$'s input is given to $\mathcal{P}$ in index $\sigma(i, j)$. By the definition of $\mathcal{E}$ and the way we have constructed $\mathcal{Q}$, we can write the probability that $\mathcal{Q}$ accepts as,

$$\Pr\left[\frac{1}{m}\sum_{j\in[m]} X_{\sigma(j,\ell)} \geq \beta + \frac{\epsilon}{2} \text{ for all } \ell \in [n]\right]. \tag{4.14}$$

By replacing ϵ with $\epsilon/2$ in Lemma 2.0.1, and choosing $\alpha = \beta + \epsilon/2$, we obtain,

$$(1 - ne^{-2\epsilon^2 m/4}) \Pr\left[\frac{1}{mn}\sum_{j\in[mn]} X_j \geq \beta + \frac{\epsilon}{2} + \frac{\epsilon}{2}\right] \leq \Pr\left[\frac{1}{m}\sum_{j\in[m]} X_{\sigma(j,\ell)} \geq \beta + \frac{\epsilon}{2} \text{ for all } \ell \in [n]\right] \tag{4.15}$$

The second factor in the left-hand side is the probability $\mathcal{V}^{mn}_{\beta+\epsilon}$ accepts after interacting with $\mathcal{P}$ on x. The right-hand side is the probability that $\mathcal{E}$ accepts when interacting with $\mathcal{Q}$ on x. And so, using the fact that $x \notin L$ and that $\mathcal{E}$ has soundness error at most $\left(\frac{\beta}{\beta+\epsilon/2}\right)^n$, the probability that $\mathcal{V}^{mn}_{\beta+\epsilon}$ accepts after interacting with $\mathcal{P}$ on x is at most

$$\frac{\left(\frac{\beta}{\beta+\epsilon/2}\right)^n}{1 - ne^{-2\epsilon^2 m/4}}, \tag{4.16}$$

as long as the denominator is positive. Since $\mathcal{P}$ and $x \notin L$ were arbitrary (and $\mathcal{Q}$ is independent of x), $\mathcal{V}^{mn}_{\beta+\epsilon}$ has soundness error at most

$$\frac{\left(\frac{\beta}{\beta+\epsilon/2}\right)^n}{1 - ne^{-2\epsilon^2 m/4}} \tag{4.17}$$

for L, as long as the denominator is positive. □

4.4 Parallel Repetition Can Always Reduce Error

We can now prove our main theorem.

Theorem 4.4.1. *Let $0 \leq 1 - \alpha < 1$ be the completeness error of a quantum interactive proof verifier, and let $0 \leq \beta < \alpha$ be the soundness error. Write $\alpha = \beta + \epsilon$ for $\epsilon > 0$. Then, by repeating the interactive proof system in parallel and checking for a threshold of at least $\beta + \epsilon/2$, both the completeness and soundness error can be reduced to at most $\delta > 0$ in $O(\log(1/\delta)\log(\log(1/\delta)))$ parallel repetitions.*

Proof. Let $\mathcal{V}$ be a verifier with completeness error at most $0 \leq 1 - \alpha < 1$ and soundness error at most $0 \leq \beta < \alpha$ for a language L. Write $\alpha = \beta + \epsilon$. Let $n \in \mathbb{N}$ where $n \geq 2$ and let $m = 16\log(n)/\epsilon^2$. We will consider the completeness error and soundness error for the verifier $\mathcal{V}^{\otimes mn}_{\beta+\epsilon/2}$ separately, beginning with the soundness error.

By Theorem 4.3.2, the verifier $\mathcal{V}^{\otimes mn}_{\beta+\epsilon/2}$ has soundness error at most,

$$\frac{\left(\frac{\beta}{\beta+\epsilon/4}\right)^n}{1 - ne^{-2\epsilon^2 m/16}} = \frac{\left(\frac{\beta}{\beta+\epsilon/4}\right)^n}{1 - ne^{-2\log(n)}} \tag{4.18}$$

$$= \frac{\left(\frac{\beta}{\beta+\epsilon/4}\right)^n}{1 - 1/n} \tag{4.19}$$

$$\leq 2\left(\frac{\beta}{\beta+\epsilon/4}\right)^n, \tag{4.20}$$

since $1/n \leq 0.5$ for all $n \geq 2$. We want,

$$2\left(\frac{\beta}{\beta+\epsilon/4}\right)^n \leq \delta, \tag{4.21}$$

which is trivially true if $\beta = 0$. Otherwise, when $\beta > 0$, we can let $b = \left(\frac{\beta}{\beta+\epsilon/4}\right)^{-1}$ and since $\epsilon > 0$ we know that $b > 1$, so we can take the log-base-b of both sides to get an equivalent inequality,

$$-n + \log_b(2) \leq \log_b(\delta), \tag{4.22}$$

which is equivalent to,

$$n \geq -\log_b(\delta) + \log_b(2) = \log_b(1/\delta) + \log_b(2). \tag{4.23}$$

So if we let $n = \lceil \log_b(1/\delta) + \log_b(2) \rceil$, then by repeating $\mathcal{V}$ in parallel $mn = 16n\log(n)/\epsilon^2 \in O(\log(1/\delta)\log(\log(1/\delta)))$ times and checking for a threshold of $\beta + \epsilon/2$, we have reduced the soundness error to at most δ.

For the completeness error, let $\mathcal{P}$ be a prover compatible with $\mathcal{V}$ such that for all $x \in L$, $\mathcal{V}$ accepts when interacting with $\mathcal{P}$ on x with probability at least $\alpha > 0$. Consider the prover $\mathcal{Q}$ compatible with $\mathcal{V}^{\otimes mn}_{\beta+\epsilon/2}$ which simulates mn independent copies of $\mathcal{P}$. Then $\mathcal{Q}$ has independent probability at least α of winning each parallel repetition of $\mathcal{V}$ when interacting with $V^{\otimes mn}_{\beta+\epsilon/2}$ on an $x \in L$. Each parallel repetition can be thought of as an independent experiment that yields success with probability $p \geq \alpha$, and so the number of successful outcomes is described by the binomial distribution with $N = mn$ experiments each with independent probability of success p. The probability that fewer than $k = (\beta+\epsilon/2)N < pN$ experiments yield success (in other words, the probability that P' fails to convince $V^{mn}_{\beta+\epsilon/2}$), can be upper-bounded by the Chernoff bound we discussed in Chapter 2. Where X is the number of successful outcomes,

$$\begin{aligned}
\Pr[X \leq k] &\leq \exp\left(-\frac{1}{2p}\frac{(Np-k)^2}{N}\right) && (4.24)\\
&= \exp\left(-\frac{1}{2p}\frac{(N(p-(\beta+\epsilon/2)))^2}{N}\right) && (4.25)\\
&= \exp\left(-\frac{1}{2p}N(p-(\beta+\epsilon/2))^2\right) && (4.26)\\
&\leq \exp\left(-\frac{1}{2}N(p-(\beta+\epsilon/2))^2\right) && (4.27)\\
&\leq \exp\left(-\frac{1}{2}N(\alpha-(\beta+\epsilon/2))^2\right) && (4.28)\\
&= \exp\left(-\frac{1}{2}N(\epsilon/2)^2\right) && (4.29)\\
&= \exp\left(-\frac{1}{2}mn\epsilon^2/4\right) && (4.30)\\
&= \exp\left(-\frac{1}{2}16n\log(n)/4\right) && (4.31)\\
&= \exp(-2n\log(n)). && (4.32)\\
& && (4.33)
\end{aligned}$$

If we choose $n \geq \log(1/\delta)$ then by repeating $\mathcal{V}$ in parallel mn times, we have reduced

the completeness error to at most

$$\exp(-2\log(1/\delta)\log(\log(1/\delta))) \leq \exp(-\log(1/\delta)) = \exp(\log(\delta)) = \delta. \tag{4.34}$$

Combining our analysis of the soundness error and completeness error, if we choose

$$n = 2 + \begin{cases} \lceil \log(1/\delta) \rceil & \text{if } \beta = 0 \\ \lceil \log(1/\delta) \rceil + \lceil \log_b(1/\delta) + \log_b(2) \rceil & \text{otherwise,} \end{cases} \tag{4.35}$$

then $\mathcal{V}^{mn}_{\beta+\epsilon/2}$ has completeness and soundness error at most δ. Furthermore, $mn = 16n\log(n)/\epsilon^2 \in O(\log(1/\delta)\log(\log(1/\delta)))$. □

(Note that the completeness error part of our proof is pretty much a copy of the completeness error part of Molina's proof of 4.1.1 with some extra detail added).

4.5 Conclusion

The contributions of this chapter are:

- A proof that soundness error can be reduced exponentially by repeating a quantum interactive proof system in parallel and checking that the fraction of repetitions that accept is slightly larger than the original proof system's soundness error.
- A proof that given a quantum interactive proof system with parameters α (the probability with which the verifier can be convinced to accept when they should accept) and β (the soundness error), as long as $\alpha > \beta$, both the soundness error and completeness error can be reduced exponentially by repeating the protocol in parallel and requiring an $(\alpha+\beta)/2$ fraction of the repetitions to be won. Our result requires a log-factor more repetitions than are necessary in the classical case.

Chapter 5

Conclusion

We began in Chapter 2 by introducing some facts from probability theory and then proved Lemma 2.0.1, a fact about binary-valued random variables. We went on to use Lemma 2.0.1 in our study of parallel repetition and concentration bounds in Chapters 3 and 4.

In Chapter 3, we developed a technique for converting parallel repetition theorems into concentration bounds for nonlocal games. We proved new concentration bounds for certain kinds of games using the parallel repetition theorems that are currently available.

In Chapter 4, we proved that the soundness and completeness errors of a quantum interactive proof system can be reduced through simple parallel repetition, eliminating the need to rely on more complicated error-reduction strategies.

Chapters 3 and 4 are are examples of how Lemma 2.0.1 can be used to reduce the problem of winning n threshold games repeated in parallel to the problem of winning one threshold game (which has a slightly higher threshold). We expect this technique to be applicable to other problems that we did not discuss, e.g. reducing the error in multi-prover quantum interactive proof protocols.

References

[BVY15] Mohammad Bavarian, Thomas Vidick, and Henry Yuen. Anchoring games for parallel repetition. *arXiv preprint arXiv:1509.07466*, 2015.

[CS14] André Chailloux and Giannicola Scarpa. Parallel repetition of entangled games with exponential decay via the superposed information cost. In *International Colloquium on Automata, Languages, and Programming*, pages 296–307. Springer, 2014.

[CSUU08] Richard Cleve, William Slofstra, Falk Unger, and Sarvagya Upadhyay. Perfect parallel repetition theorem for quantum XOR proof systems. *Computational Complexity*, 17(2):282–299, 2008.

[CWY15] Kai-Min Chung, Xiaodi Wu, and Henry Yuen. Parallel repetition for entangled k-player games via fast quantum search. In *Proceedings of the 30th Conference on Computational Complexity*, pages 512–536. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[IK10] Russell Impagliazzo and Valentine Kabanets. Constructive proofs of concentration bounds. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 617–631. Springer, 2010.

[JJUW10] Rahul Jain, Zhengfeng Ji, Sarvagya Upadhyay, and John Watrous. Qip= pspace. *Communications of the ACM*, 53(12):102–109, 2010.

[JUW09] Rahul Jain, Sarvagya Upadhyay, and John Watrous. Two-message quantum interactive proofs are in PSPACE. *arXiv preprint arXiv:0905.1300*, 2009.

[KW00] Alexei Kitaev and John Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof systems. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 608–617. ACM, 2000.

[LW15] Cécilia Lancien and Andreas Winter. Parallel repetition and concentration for (sub-) no-signalling games via a flexible constrained de finetti reduction. *arXiv preprint arXiv:1506.07002*, 2015.

[Mos14] Dana Moshkovitz. Parallel repetition from fortification. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 414–423. IEEE, 2014.

[MP12] Abel Molina Prieto. Parallel repetition of prover-verifier quantum interactions. Master's thesis, University of Waterloo, 2012.

[NC11] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, New York, NY, USA, 10th edition, 2011.

[Rao11] Anup Rao. Parallel repetition in projection games and a concentration bound. *SIAM Journal on Computing*, 40(6):1871–1891, 2011.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.

[Sha92] Adi Shamir. IP=PSPACE. *Journal of the ACM (JACM)*, 39(4):869–877, 1992.

[Slo17] William Slofstra. The set of quantum correlations is not closed. *arXiv preprint arXiv:1703.08618*, 2017.

[Ung09] Falk Unger. A probabilistic inequality with applications to threshold direct-product theorems. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 221–229. IEEE, 2009.

[Wat18] John Watrous. *The theory of quantum information*. Cambridge University Press, 2018.

[Yue16] Henry Yuen. A parallel repetition theorem for all entangled games. *arXiv preprint arXiv:1604.04340*, 2016.