No-Regret Learning and Equilibrium Computation in Quantum Games

Wayne Lin, Georgios Piliouras, Ryann Sim, and Antonios Varvitsiotis

Singapore University of Technology and Design, Singapore

As quantum processors advance, the emergence of large-scale decentralized systems involving interacting quantum-enabled agents is on the horizon. Recent research efforts have explored quantum versions of Nash and correlated equilibria as solution concepts of strategic quantum interactions, but these approaches did not directly connect to decentralized adaptive setups where agents possess limited information. This paper delves into the dynamics of quantum-enabled agents within decentralized systems that employ no-regret algorithms to update their behaviors over time. Specifically, we investigate two-player quantum zero-sum games and polymatrix quantum zero-sum games, showing that no-regret algorithms converge to separable quantum Nash equilibria in time-average. In the case of general multi-player quantum games, our work leads to a novel solution concept, that of the separable quantum coarse correlated equilibria (QCCE), as the convergent outcome of the time-averaged behavior no-regret algorithms, offering a natural solution concept for decentralized quantum systems. Finally, we show that computing QCCEs can be formulated as a semidefinite program and establish the existence of entangled (i.e., non-separable) QCCEs, which are unlearnable via the current paradigm of no-regret learning.

1 Introduction

As quantum computing reaches maturity and quantum computing processors become more affordable and scalable, large-scale systems with interacting quantum-enabled agents will become more commonplace. Quantum games offer a powerful framework to predict the behavior and guide the design of such systems [1, 2, 3, 4, 5]. In a quantum game, agents can process and exchange quantum information, and their utilities are determined by performing a measurement on a quantum state that is shared among all agents.

A significant portion of quantum game literature studies well-known games such as the Prisoner's Dilemma [6] and Battle of the Sexes [7], aiming to uncover potential advantages of using quantum strategies when compared to classical ones. Another significant research avenue centers on identifying suitable solution concepts for quantum games, which correspond to system states that exhibit stability against unilateral player deviations and are collectively referred to as quantum equilibria. In particular, two notions of quantum equilibria have been studied: quantum Nash equilibria and quantum correlated equilibria [2, 3, 8, 9]. Nevertheless, computing quantum Nash equilibria is intractable [2], casting doubt on their suitability as a viable solution concept. Indeed, in view of the hardness of computing quantum equilibria, how are agents expected to reach such states? To make matters worse, the hardness result holds even in settings where agent's utilities are known, an unrealistic assumption for large-scale decentralized systems.

A more pragmatic setup is to consider that agents interact with each other over a series of rounds

Wayne Lin: wayne_lin@sutd.edu.sg

Georgios Piliouras: georgios@sutd.edu.sg

Ryann Sim: ryann_sim@sutd.edu.sg

Antonios Varvitsiotis: antonios@sutd.edu.sg

and they have the opportunity to improve their strategies over time based on the outcomes of previous interactions. One established method to enable this dynamic learning process is *no-regret learning*, where agents update their strategies using an algorithm that meets a natural benchmark; it performs as well as the best fixed strategy in hindsight. This leads to the following key question which we seek to answer in this paper:

For which classes of quantum games can agents reach equilibria using no-regret learning? What types of equilibria do they arrive at?

In the realm of classical normal-form games, where strategies are probability simplex vectors that capture classical randomness over a finite set of actions, it is well understood that *no-regret learning converges to equilibrium states* [10, 11]. However, the type of equilibrium and notion of convergence depends on the specific setting and underlying applications. Comparatively, the study of no-regret learning in quantum games is in its infancy [8, 12, 13, 14]. Our main goal in this work is to develop distributed algorithms for learning in quantum games and explore what types of equilibrium solutions emerge across different game classes.

Model, approach, and contributions. In this paper we focus on a model of quantum games which is a natural extension of prior models, while still being amenable to no-regret learning. Formally, we focus on non-interactive quantum games, where each player i controls a quantum register $\mathcal{H}_i$ and has as their strategy a density matrix $\rho_i \in \mathrm{D}(\mathcal{H}_i)$. The joint strategy is given by a state $\rho \in \mathrm{D}(\bigotimes_i \mathcal{H}_i)$, and the payoff of the i-th player is given by the expected value of an observable R_i on the joint state, i.e.,

$$u_i(\rho) = \mathrm{Tr}(\rho R_i). \tag{QG}$$

A QG is called zero-sum if players' payoffs add up to zero. More generally, we also consider polymatrix QGs, where there are k players and each player is situated at a node within an undirected graph. Every player engages in two-player QGs with each of their neighboring agents, employing a single state $\rho_i \in \mathrm{D}(\mathcal{H}_i)$ to participate in all games with their neighbors, and their individual payoff is the cumulative payoff earned across all these games.

To investigate learning in quantum games, we draw inspiration from insights derived from no-regret learning in classical games. Within this line of research, we single out two important results. Firstly, no-external-regret learning gives rise to decentralized algorithms that converge in the time-average sense to *coarse correlated equilibria* [11] in arbitrary normal-form games. Secondly, no-external-regret learners converge in the time-average sense to *Nash equilibria* in both two-player zero-sum games [10, 15] and also polymatrix (globally) zero-sum games [16, 17].

All these results can be unified within the $\mathbf{\Phi}$-regret framework [18]. The benchmark of no-$\mathbf{\Phi}$-regret arises by allowing agents to deviate from an action x to $\phi(x)$, where ϕ is an admissible deviation mapping in a family $\mathbf{\Phi}$ of linear deviation maps. A unifying result of noteworthy relevance to our work is that players using a no-$\mathbf{\Phi}$-regret algorithm converge to a corresponding notion of $\mathbf{\Phi}$-equilibria in classical normal-form games [18].

Our results in this work show that all aforementioned results for no-regret learning in classical normal-form games carry over to the quantum regime. Specifically, in this work:

- We introduce the notion of quantum $\mathbf{\Phi}$-equilibria (Q$\mathbf{\Phi}$E) for any admissible family of quantum deviation maps $\mathbf{\Phi}$. Notably, the well-explored concepts of quantum Nash (QNE) and quantum correlated equilibria (QCE) both emerge as specific instances within this broader framework.
- For any QG, we show that no-$\mathbf{\Phi}$-regret learning converges to Q$\mathbf{\Phi}$E. Moreover, we show that the set of *separable* quantum coarse correlated equilibria (QCCE) coincides with the limit points of the time-averaged achieved through no-external-regret algorithms. On the other hand, *entangled* QCCEs cannot be reached through the current paradigm of learning in games, and we demonstrate that this class of equilibria is not vacuous by constructing examples of entangled QCCEs in Appendix A.

- For two-player quantum zero-sum games and polymatrix quantum zero-sum games, we show that the limit points of no-external-regret algorithms are separable QNEs.

Related work. Research on no-regret learning in quantum games is relatively limited. In a pioneering work, [8] focused on Matrix Multiplicative Weights Update (MMWU), a matrix extension of the widely-used Multiplicative Weights Update algorithm [19]. Their research focused on two-player quantum zero-sum games, demonstrating convergence in a time-average sense to the set of QNEs. Our results provide an alternative, simpler proof of that result, which furthermore holds for any no-external-regret algorithm.

More recently, [13] and later on [12] studied continuous-time variants of MMWU and variants thereof. They demonstrated a cyclic behavior known as Poincaré recurrence in the dynamics, a phenomenon reminiscent of classical results indicating that regret minimization alone is insufficient for last-iterate equilibrium convergence, e.g. see [20, 21, 22]. Beyond the zero-sum setting, [14] studies the continuous and discrete variants of a linear version of MMWU in quantum potential games, showing that players' utilities strictly increase when using these algorithms.

The particular context of learning within QGs investigated in this study can be regarded as a specific instance of the broader framework of learning in convex games, as outlined in works like [23] and [24]. Consequently, it becomes essential to elucidate the applicability of these general findings to our specific setting.

In particular, [23] studies $\mathbf{\Phi}$-regret minimization in general convex games and provides a template for designing no-$\mathbf{\Phi}$-regret algorithms, which entails that:

1. The set of transformations $\mathbf{\Phi}$ is a *reproducing kernel Hilbert space* (RKHS).
2. We have access to an algorithm $\mathcal{A}'$ which computes approximate fixed points of any $\phi \in \mathbf{\Phi}$.
3. We have access to an algorithm $\mathcal{A}''$ for no-regret learning in the setting where actions correspond to choosing a deviation $\phi \in \mathbf{\Phi}$.

In terms of using this framework for no-regret learning in QGs, the most restrictive assumption is the third one. An an example, in the case where $\mathbf{\Phi}$ is the set of all completely positive trace-preserving maps (i.e., linear maps that transform valid quantum states to valid quantum states), the third assumption necessitates the existence of a no-regret algorithm for learning over the domain of completely positive, trace-preserving (CPTP) maps. To the best of our knowledge, such an algorithm is not available, and obtaining one is the focus of ongoing work.

Finally, [24] studies no-internal-regret convergence in convex games, however, their algorithm is not practically applicable as the time and space requirements grow exponentially with the number of timesteps. In contrast, our approach is efficiently computable and in Section 6 we complement our theoretical results with related experiments.

2 Quantum Games, Equilibria and Online Optimization

In this section, we introduce a broad formulation of quantum games and study non-commutative analogues of classical equilibrium concepts in such games, before turning our attention to the equilibria we can learn and how to learn them in the subsequent sections.

Quantum preliminaries. A d-dimensional quantum register is mathematically described as the set of unit vectors in a d-dimensional Hilbert space $\mathcal{H}$. The *state* of a qudit quantum register $\mathcal{H}$ is represented by a *density matrix*, i.e., a $d \times d$ Hermitian positive semidefinite matrix with trace equal to 1. A qudit is the unit of quantum information described by a superposition of d states. If the number of states d is equal to two then it is referred to as a qubit. The state space of a quantum register $\mathcal{H}$ is denoted by $\mathrm{D}(\mathcal{H})$.

One mathematical formalism of the process of measuring a quantum system is the positive-operator-valued measure (POVM), defined as a set of positive semidefinite operators $\{P_i\}_{i=1}^{m}$ such that $\sum_{i=1}^{m} P_i = \mathbb{1}_{\mathcal{H}}$, where $\mathbb{1}_{\mathcal{H}}$ is the identity matrix on $\mathcal{H}$. If the quantum register $\mathcal{H}$ is in a state described by density matrix $\rho \in \mathrm{D}(\mathcal{H})$, upon performing the measurement $\{P_i\}_{i=1}^m$ we get the outcome i with probability $\langle P_i, \rho\rangle$, where $\langle A, B\rangle = \mathrm{Tr}\big(A^\dagger B\big)$ is the *Hilbert-Schmidt inner product* defined on the linear space of Hermitian matrices. A POVM can also be seen as a collection of observables, each corresponding to a Hermitian operator. In this paper we focus on the POVM formalism for quantum measurement, but there are other formalisms in the literature which we defer to Appendix B for completeness.

Given a finite-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^n$, we denote by $\mathrm{L}(\mathcal{H})$ the set of linear operators acting on $\mathcal{H}$, i.e., the set of all $n \times n$ complex matrices over $\mathcal{H}$. When two quantum registers with associated spaces $\mathcal{A}$ and $\mathcal{B}$ of dimension n and m respectively are considered as a joint quantum register, the associated state space is given by the density operators on the tensor product space, i.e., $D(\mathcal{A}\otimes\mathcal{B})$. A linear operator that maps matrices to matrices, i.e., a mapping $\Theta : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$, is called a *super-operator*. The set of admissible super operators is equivalently the set of completely positive and trace preserving (CPTP) maps. The adjoint super-operator $\Theta^\dagger : \mathrm{L}(\mathcal{A}) \to \mathrm{L}(\mathcal{B})$ is uniquely determined by the equation $\langle A, \Theta(B)\rangle = \langle \Theta^\dagger(A), B\rangle$. A super-operator $\Theta : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$ is *positive* if it maps PSD matrices to PSD matrices. There exists a linear bijection between matrices $R \in \mathrm{L}(\mathcal{A}\otimes\mathcal{B})$ and super-operators $\Theta : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$ known as the *Choi-Jamiołkowski isomorphism.* Specifically, for a super-operator Θ its *Choi matrix* is:

$$C_\Theta = \sum_{1\le i,j\le m} \Theta(E_{i,j}) \otimes E_{i,j} \in \mathrm{L}(\mathcal{A}\otimes\mathcal{B}), \tag{1}$$

where $\{E_{i,j}\}_{i,j=1}^m$ is the standard orthonormal basis of $\mathrm{L}(\mathcal{B}) = \mathbb{C}^{m\times m}$. Conversely, given an operator $R = \sum_{1\le i,j\le m} A_{i,j} \otimes E_{i,j} \in \mathrm{L}(\mathcal{A}\otimes\mathcal{B})$, we can define $\Theta_R : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$ by setting $\Theta_R(E_{i,j}) = A_{i,j}$ from which it easily follows that $C_{\Theta_R} = R$. Explicitly, we have

$$\Theta_R(B) = \mathrm{Tr}_{\mathcal{B}}(R(\mathbb{1}_{\mathcal{A}} \otimes B^\top)), \tag{2}$$

where the partial trace $\mathrm{Tr}_{\mathcal{B}} : \mathrm{L}(\mathcal{A}\otimes\mathcal{B}) \to \mathrm{L}(\mathcal{A})$ is the *unique* function that satisfies:

$$\mathrm{Tr}_{\mathcal{B}}(A\otimes B) = A\,\mathrm{Tr}(B),\ \forall A, B.$$

Moreover, the adjoint map is $\mathrm{Tr}_{\mathcal{B}}^\dagger(A) = A \otimes \mathbb{1}_{\mathcal{B}}$. Lastly, a superoperator Θ is completely positive (i.e., $\mathbb{1}_m \otimes \Theta$ is positive for all $m \in \mathbb{N}$) if and only if the Choi matrix of Θ is positive semidefinite. In particular, if the Choi matrix of the super-operator Θ is PSD, it follows that Θ is positive.

Finally, if a state $\rho \in \mathrm{D}(\bigotimes_i \mathcal{H}_i)$ can be written as a convex combination of product states, i.e.,

$$\rho = \sum_j \lambda_j \bigotimes_i \rho_{i,j} \tag{3}$$

where $\lambda_j \ge 0\ \forall j$, $\sum_j \lambda_j = 1$, and $\rho_{i,j} \in \mathcal{H}_i\ \forall j$, then it is called *separable.* A separable state can be interpreted as a classical probability distribution over the product states $\rho_j := \bigotimes_i \rho_{i,j}$. A state that is not separable is said to be *entangled.*

2.1 Quantum games

In a QG, there are k players and each player i has register $\mathcal{H}_i$ and selects a density matrix $\rho_i \in \mathrm{D}(\mathcal{H}_i)$. A joint strategy is given by a joint state $\rho \in \mathrm{D}(\bigotimes_i \mathcal{H}_i)$. Each player has an observable $R_i = \sum_{m_i} m_i P_{m_i}$ and thus their utility function is the (multilinear) expected value of the observable R_i on the joint state, i.e.,

$$u_i(\rho) = \mathrm{Tr}(\rho R_i) \tag{QG}$$

for some Hermitian $R_i \in \mathrm{L}(\otimes_i \mathcal{H}_i)$. We henceforth refer to R_i as player i's utility tensor.

It is useful to note that if $\rho = \rho_i \otimes \rho_{-i}$ for some $i \in [k]$, $\rho_{-i} \in \mathrm{D}(\bigotimes_{i'\neq i} \mathcal{H}_{i'})$, then we can also write the utility (QG) in the alternative form

$$u_i(\rho) = \mathrm{Tr}(\rho R_i) = \big\langle \rho_i, \Theta_i(\rho_{-i}^\top)\big\rangle \tag{4}$$

where $\Theta_i : \mathrm{L}(\bigotimes_{i' \neq i} \mathcal{H}_{i'}) \to \mathrm{L}(\mathcal{H}_i)$ and $R_i \in \mathrm{L}(\bigotimes_{i'} \mathcal{H}_{i'}) = \mathrm{L}(\mathcal{H}_i \otimes (\bigotimes_{i' \neq i} \mathcal{H}_{i'}))$ are related via the Choi-Jamiołkowski isomorphism. This is because, by (2),

$$\langle \rho_i, \Theta_i(\rho_{-i}^\top) \rangle = \langle \rho_i, \mathrm{Tr}_{\mathcal{B}}(R(\mathbb{1}_{\mathcal{A}} \otimes \rho_{-i})) \rangle = \langle \rho_i \otimes \mathbb{1}_{\mathcal{B}}, R(\mathbb{1}_{\mathcal{A}} \otimes \rho_{-i}) \rangle = \mathrm{Tr}(R(\rho_i \otimes \rho_{-i})).$$

An important special case of the QG framework was introduced by Zhang [3] in an attempt to explore whether quantum resources provide advantages in classical games. In Zhang's model, the strategies of each player are encoded by a Hilbert space $\mathcal{H}_i$ with an orthonormal basis $|s_i\rangle, s_i \in S_i$ and the probability of strategy profile $s \in \times_i S_i$ is given by $\mathrm{Tr}(|s\rangle\langle s| \rho)$, where the shared state for all players is $\rho \in D(\otimes_i \mathcal{H}_i)$. Consequently the expected utility of the i-th player is given by

$$u_i(\rho) = \sum_{s \in S} \mathrm{Tr}(|s\rangle\langle s| \rho) u_i(s) = \mathrm{Tr}\left(\sum_{s \in S} u_i(s) |s\rangle\langle s| \rho \right),$$

which is the form of QG when restricted to diagonal utility tensors, i.e., $R_i = \sum_{s \in S} u_i(s) |s\rangle\langle s|$.

Moreover, the QG framework also captures the work of [8, 25], who consider a related model of non-interactive quantum games wherein players each control a quantum register. They each independently prepare a quantum state in the register they hold, which are subsequently sent to a referee who performs a joint measurement to determine a real payoff to each player. Crucially, the QG framework can be seen as the first stage of the more general quantum game framework found in [1, 2]. In this broad framework, quantum strategies are defined over n rounds of interaction. Each round is characterized by an input and output space, as well as admissible mappings between rounds which can capture quantum memory. Finally, at the last memory state, a joint measurement is made to determine the payoffs. We focus on a specialization of this framework, where players select a strategy without prior communication or entanglement, and a measurement is performed using the joint state of the players. This allows us to study a closer quantum analogy to the framework of classical, simultaneous *non-cooperative* games.

In our setting, the extended utility function of QG, $u_i(\rho) = \mathrm{Tr}(\rho R_i)$, can be interpreted as the expected utility if the product state (i.e., strategy profile) to be played is separable. On the other hand, an entangled state ρ can only be played by a central agent who plays on behalf of all the players. We explore this problem in more detail in Appendix A.

2.2 Various notions of quantum equilibria

In prior works in quantum game theory, classical notions of equilibria have been studied in the quantum context, and several classical results have been shown to have non-commutative analogues. In this section, we recall some of these notions and introduce the quantum coarse correlated equilibrium (QCCE), which we first write in terms of deviation mappings and later reformulate in terms of partial traces.

A seminal equilibrium concept in classical game theory is the Nash equilibrium. In our formulation (QG), we have the notion of an ϵ-quantum Nash equilibrium (ϵ-QNE), from which players can only make utility gains of $\leq \epsilon$ by deviating. This formulation of the quantum Nash equilibrium has already been studied in several works [2, 8, 25] and we repeat it using our notation for completeness.

Definition 2.1. *A product state $\rho = \bigotimes_i \rho_i \in \mathrm{D}(\otimes_i \mathcal{H}_i)$ is a quantum Nash equilibrium (QNE) if*

$$u_i(\rho) \geq u_i(\rho_i' \otimes \rho_{-i}) \quad \forall\, i \in [k],\ \rho_i' \in \mathrm{D}(\mathcal{H}_i), \tag{QNE}$$

where $\rho_{-i} := \bigotimes_{j \neq i} \rho_j$. Moreover ρ is called an ϵ-approximate quantum Nash equilibrium (ϵ-QNE) if the inequality is satisfied up to an additive error of ϵ.

The *exploitability* (see e.g. [26]) of player i at state ρ is the maximum utility they can gain by deviating, and is given by $\lambda_{\max}(\Theta_i(\rho_{-i})) - u_i(\rho) \geq 0$. This leads to an alternative characterization of QNEs as the product states that have zero total exploitability, i.e.,

$$\rho \text{ is a QNE} \iff \sum_{i=1}^{k} \Big(\lambda_{\max}(\Theta_i(\rho_{-i})) - u_i(\rho) \Big) = 0. \tag{5}$$

Clearly, QNEs are states which are stable under unilateral player deviations. However, when we consider non-product states, then we need to formulate a better way to capture families of unilateral deviations permitted within the quantum mechanics framework. These deviations are captured by quantum channels and mathematically formalized by completely positive, trace-preserving (CPTP) maps, which are the family of linear mappings from density matrices to density matrices. This leads to the following definition:

Definition 2.2. *Consider a family of CPTP maps $\mathbf{\Phi} = \{\Phi_i\}_{i=1}^k$. A state $\rho \in \mathrm{D}(\otimes_i \mathcal{H}_i)$ is called a quantum $\mathbf{\Phi}$-equilibrium if*

$$u_i(\rho) \geq u_i((\phi_i \otimes \mathbb{I}_{-i})(\rho)) \quad \forall\ i,\ \phi_i \in \Phi_i. \tag{Q$\mathbf{\Phi}$E}$$

Moreover ρ is an ϵ-quantum $\mathbf{\Phi}$-equilibrium *(ϵ-Q$\mathbf{\Phi}$E) if the inequality is satisfied up to an additive error of ϵ.*

Specializing this formulation and allowing for all possible CPTP maps, we arrive at the notion of quantum correlated equilibria (QCE), first defined in [3] and analyzed further in [9]. Concretely, we have that:

Definition 2.3. *A state $\rho \in \mathrm{D}(\otimes_i \mathcal{H}_i)$ is called a quantum correlated equilibrium if*

$$u_i(\rho) \geq u_i((\phi_i \otimes \mathbb{I}_{-i})(\rho)) \tag{QCE}$$

where $\phi_i : \mathrm{L}(\mathcal{H}_i) \to \mathrm{L}(\mathcal{H}_i)$ is a CPTP map on player i's subsystem. Moreover ρ is an ϵ-quantum correlated equilibrium *(ϵ-QCE) if the inequality is satisfied up to an additive error of ϵ.*

As a second instantiation of $\mathbf{\Phi}$-equilibria, we consider the set of "constant maps", which in our setting corresponds to replacement channels. This leads to a new notion of quantum equilibria, described below:

Definition 2.4. *A state $\rho \in \mathrm{D}(\otimes_i \mathcal{H}_i)$ is called a quantum coarse correlated equilibrium if*

$$u_i(\rho) \geq u_i((\phi_i \otimes \mathbb{I}_{-i})(\rho)) \tag{QCCE}$$

for all players $i \in [k]$ and all replacement channels

$$\phi_i : \mathrm{L}(\mathcal{H}_i) \to \mathrm{L}(\mathcal{H}_i),\ X \mapsto \mathrm{Tr}(X)\rho_i', \quad \text{where } \rho_i' \in \mathrm{D}(\mathcal{H}_i). \tag{6}$$

Equivalently, we can express the QCCE definition in terms of partial traces. In particular, ρ is a QCCE if

$$u_i(\rho) \geq u_i(\rho_i' \otimes \mathrm{Tr}_i\, \rho) \tag{7}$$

for all players $i \in [k]$ and $\rho_i' \in \mathrm{D}(\mathcal{H}_i)$, where $\mathrm{Tr}_i : \mathrm{L}(\bigotimes_{i'} \mathcal{H}_{i'}) \to \mathrm{L}(\bigotimes_{i' \neq i} \mathcal{H}_{i'})$ is the partial trace with respect to player i's subsystem. Moreover, ρ is an ϵ-approximate quantum coarse correlated equilibrium (ϵ-QCCE) if the inequality is satisfied up to an additive error of ϵ. Finally, if a QCCE ρ is a separable state (i.e., it can be expressed as a convex combination of product states), we call it a separable QCCE. A proof of equivalence between QCCE and (7) can be found in Lemma C.1 of the Appendix.

By definition, a product Q$\mathbf{\Phi}$E is a QNE. Thus, the space of states considered matters greatly in our equilibrium definitions. For QGs it turns out that, unlike the classical case, there is a space of states of interest between that the narrow class of product states and the broad class of (possibly entangled) general states, which is the space of separable states (3). Separable equilibria are an interesting class of equilibria to consider not only because of the significance of separable states in quantum theory, but also because separable states are the set of states reachable by the current paradigm of no-regret learning, in which a product state is played in each round of play and equilibria are obtained by taking a time average—i.e., a convex combination.

Nevertheless, entangled (i.e., non-separable) equilibria can exist. In particular, in we provide explicit constructions of maximally-entangled QCCEs in Appendix A.

Spectrahedral characterization of QCCEs. Analogous to the classical setting where the set of CCEs can be described as the feasible space of a linear program [27], the set of QCCEs of a game can be described as the feasible space of a semidefinite program (SDP). Suppose that, for each $i \in [k]$, $\Theta_i : \mathrm{L}(\bigotimes_{i' \neq i} \mathcal{H}_{i'}) \to \mathrm{L}(\mathcal{H}_i)$ and $R_i \in \mathrm{L}(\bigotimes_{i'} \mathcal{H}_{i'}) = \mathrm{L}(\mathcal{H}_i \otimes (\bigotimes_{i' \neq i} \mathcal{H}_{i'}))$ are related via the Choi-Jamiołkowski isomorphism. Then a density matrix ρ^* is a QCCE if and only if

$$\begin{aligned}
& u_i(\rho^*) \geq u_i(\rho_i' \otimes \operatorname{Tr}_i \rho^*) && \forall i \in [k], \rho_i' \in \mathcal{H}_i \\
\Leftrightarrow\ & \operatorname{Tr}(R_i \rho^*) \geq \operatorname{Tr}(R_i(\rho_i' \otimes \operatorname{Tr}_i \rho^*)) && \forall i \in [k], \rho_i' \in \mathcal{H}_i \\
\Leftrightarrow\ & \operatorname{Tr}(R_i \rho^*) \geq \max_{\rho_i' \in \mathcal{H}_i} \left\langle \rho_i', \Theta_i((\operatorname{Tr}_i \rho^*)^\top) \right\rangle && \forall i \in [k] \\
\Leftrightarrow\ & \operatorname{Tr}(R_i \rho^*) \geq \lambda_{\max}(\Theta_i((\operatorname{Tr}_i \rho^*)^\top)) && \forall i \in [k] \\
\Leftrightarrow\ & \operatorname{Tr}(R_i \rho^*) I_i - \Theta_i((\operatorname{Tr}_i \rho^*)^\top) \succeq 0 && \forall i \in [k],
\end{aligned}$$

so

$$\text{QCCEs} = \{\rho^* \in \mathcal{H} : \operatorname{Tr}(R_i \rho^*) I - \Theta_i((\operatorname{Tr}_i \rho^*)^\top) \succeq 0 \quad \forall i \in [k],\ \operatorname{Tr} \rho^* = 1,\ \rho^* \succeq 0\}.$$

These conic inequality constraints can be combined into a single, block-diagonal linear matrix inequality in terms of the entries of ρ^*, giving us an SDP characterization of the set of QCCEs of a given game. Hence, the set of QCCEs is a spectrahedron, mirroring the classical result that the set of CCEs is a polyhedron.

$\mathbf{\Phi}$-equilibria in classical games. To give context for quantum $\mathbf{\Phi}$-equilibria to readers who may be unfamiliar with with the concept, we shall express well-known classical equilibria in the $\mathbf{\Phi}$-equilibria framework. A normal-form game consists of a set of players $\mathcal{N} = \{1, \ldots, k\}$ where player i may select from a finite set of actions or pure strategies S_i. Additionally, each player has a payoff function $u_i : S \equiv \prod_i S_i \to \mathbb{R}$ assigned over pure strategy profiles $s = (s_1, \ldots, s_k)$. A joint strategy $p \in \Delta(\times_i S_i)$ is a probability distribution over the space $\times_i S_i$ of pure strategy profiles, where $p(s_1, \ldots, s_k)$ is the probability the system is in the pure state $(s_1, \ldots, s_k)$. The expected payoff of the i-th player given that the joint strategy p is played is given by $u_i(p) = \sum_{s_i \in S_i : i \in \mathcal{N}} p(s_1, \ldots, s_k) u_i(s_1, \ldots, s_k)$,

The analysis of normal-form games typically boils down to equilibrium analysis: studying what the players of the game eventually fall to as the best strategy to employ. The most famous form of equilibrium is the Nash equilibrium [28], which is however intractable to compute. Several alternative equilibrium concepts have been proposed, namely correlated equilibria (CE) [29] and coarse correlated equilibria (CCE) [30]. All three of these equilibrium types can be effectively unified and examined through the lens of $\mathbf{\Phi}$-equilibria [18].

Formally, let Φ_i be a family of deviations for agent i, i.e., for any $\phi_i \in \Phi_i$ we have that $\phi_i(\Delta(S_i)) \subseteq \Delta(S_i)$, so for each $s_i' \in S_i$ the vector $\phi(s_i')$ is a distribution. Then, for any joint strategy $p \in \Delta(\times_i S_i)$ and $\phi_i \in \Phi_i$, we define a new joint strategy $(\phi_i \otimes \mathbb{I}_{-i})(p)$ that assigns to the strategy profile (s_i, s_{-i}) the probability $\sum_{s_i'} p(s_i', s_{-i}) \mathrm{Prob}(s_i' \xrightarrow{\phi_i} s_i)$. Setting $\mathbf{\Phi} = \{\Phi_i\}_{i=1}^k$, the joint strategy p is called a $\mathbf{\Phi}$-equilibrium if

$$u_i(p) \geq u_i((\phi_i \otimes \mathbb{I}_{-i})(p)) \quad \forall\ i,\ \phi_i \in \Phi_i, \tag{$\mathbf{\Phi}$-equilibrium}$$

or explicitly:

$$u_i(p) \geq \sum_{s_i, s_{-i}} u_i(s_i, s_{-i}) \sum_{s_i'} p(s_i', s_{-i}) \mathrm{Prob}(s_i' \xrightarrow{\phi_i} s_i). \tag{8}$$

Nash, correlated, and coarse correlated equilibria can all be seen as specific instances of $\mathbf{\Phi}$-equilibria through suitable choice of deviation mappings. Concretely, correlated equilibria correspond to the case where permissible deviations are linear maps ϕ_i that map distributions to distributions, i.e., $\phi_i(\Delta(S_i)) \subseteq \Delta(S_i)$. By linearity, any such map can be written as $\phi_i(x) = A_i x$, and since ϕ_i preserves the simplex A_i is entrywise nonnegative and column stochastic. On the other hand, coarse correlated equilibria correspond to the case where the allowable deviations are the constant maps, i.e., the map $\Delta(S_i)$ to a single point $x_i \in \Delta(S_i)$. Finally, a joint strategy p is a Nash equilibrium iff it is a CCE and a product distribution.

2.3 No-$\mathbf{\Phi}$-regret learning in quantum games

The concept of regret serves as a well-established measure in assessing the performance of online algorithms [10, 31]. In the ensuing discussion, we introduce the notion of $\mathbf{\Phi}$-regret within the context of online linear optimization over the set of density matrices.

Let's consider an algorithm $\mathbf{A}$ that generates a sequence of iterates $\rho^t \in D(\mathcal{H})$. The $\mathbf{\Phi}$-regret benchmark compares the cumulative utility achieved by the trajectory $\{\rho^t\}_{t=0}^T$ with the best attainable utility when deviating from ρ^t to $\phi(\rho^t)$ at each step, with $\phi \in \mathbf{\Phi}$ being a fixed deviation map, i.e.,

$$\mathrm{regret}^{T,\mathbf{\Phi}}(\mathbf{A}) = \max_{\phi \in \mathbf{\Phi}} \sum_{t=1}^{T} \left\langle \phi(\rho^t), R \right\rangle - \sum_{t=1}^{T} \left\langle \rho^t, R \right\rangle, \tag{9}$$

An online algorithm $\mathbf{A}$ is called "no-$\mathbf{\Phi}$-regret" if the normalized $\mathbf{\Phi}$-regret $\frac{1}{T}\mathrm{regret}^{T,\mathbf{\Phi}}(\mathbf{A})$ tends towards zero as T grows. In line with conventional terminology in the literature, an algorithm is deemed "no-external-regret" (or simply "no-regret," where context permits) when all admissible constant maps are considered as deviations. In this case, these deviations correspond to all replacement channels, as defined in (6).

The first example of a no-external-regret algorithm for online linear optimization over the set of density matrices is the Matrix Multiplicative Weights Update (MMWU) method, e.g. see [32, Theorem 10] and Algorithm 1. MMWU is a widely applicable no-external-regret algorithm which has found applications for online optimization over the set of density matrices [19, 32, 33]. Some specific applications include solving semidefinite programs (SDPs) [34], proving QIP=PSPACE [35] and spectral sparsification [36].

Furthermore, in [36] it is shown how MMWU arises naturally as an instance of the Follow-the-Regularized-Leader framework where the regularizer is chosen to be the entropy function. Based on this, they introduce the novel $\mathrm{FTRL}_{\mathrm{exp}}$ framework for a different class of regularizers and provide corresponding regret bounds. It turns out that this class of algorithms is also no-external-regret in our setting.

Algorithm 1: Matrix Multiplicative Weights Update (MMWU)

```
Initialize weight matrix $W_i^1 = \mathbb{1}_d$ and stepsize $\eta \leq \frac{1}{2}$.
for $t = 1 \dots T$ do
    Play using $\rho_i^t = \frac{W_i^t}{\mathrm{Tr}(W_i^t)}$
    Update weight matrix $W_i^{t+1} = \exp\Big(\eta \sum_{\tau=1}^{t} \Theta\left((\rho_{-i}^{\tau})^\top\right)\Big)$
end for
```

Beyond online optimization, in this work we consider the setup where players in a game interact with each other over a series of rounds and improve their strategies using a no-regret algorithm $\mathbf{A}$. Let Φ_i be a set of deviation mappings for each agent and let $\mathbf{\Phi} = \{\Phi_i\}_{i=1}^k$. Recalling, that in the QG setup the payoffs are multilinear (4), each player i's regret for using an online algorithm $\mathbf{A}$ with deviations Φ_i is

$$\mathrm{regret}_i^{T,\Phi_i}(\mathbf{A}, \mathbf{\Phi}) = \max_{\phi_i \in \Phi_i} \sum_{t=1}^{T} \left\langle \phi_i(\rho_i^t), \Theta\left((\rho_{-i}^t)^\top\right) \right\rangle - \sum_{t=1}^{T} \left\langle \rho_i^t, \Theta\left((\rho_{-i}^t)^\top\right) \right\rangle.$$

Moreover, $\mathbf{A}$ has the no-$\mathbf{\Phi}$-regret property if $\frac{1}{T}\mathrm{regret}^T(\mathbf{A}, \mathbf{\Phi}) \to 0$. In Theorem 3.1(a) we show that for any set of deviation mappings $\mathbf{\Phi} = \{\Phi_i\}_{i=1}^k$, the limit points of the time-averaged joint history of play $\left\{\frac{1}{T}\sum_{t=1}^T \bigotimes_i \rho_i^t\right\}_T$ generated by no-$\mathbf{\Phi}$-regret algorithms are separable QΦE.

No $\mathbf{\Phi}$-regret in classical games. In a classical normal-form game, joint strategies are distributions over pure action profiles and agents use deviation mappings $\mathbf{\Phi}$ that are linear maps that map distributions to distributions. Moreover, as payoffs are multilinear in a normal-form game, each agent updates their strategy using an algorithm for online optimization over their corresponding probability simplex.

A pivotal result highlighted in [18] demonstrates the existence of no-$\mathbf{\Phi}$-regret algorithms for any family $\mathbf{\Phi}$ of linear deviation maps. Crucially, there exists an interesting connection between no-$\mathbf{\Phi}$-regret and game theoretic equilibria. Specifically, the time-average behavior of players using a no-$\mathbf{\Phi}$-regret algorithm converges to the corresponding notion of $\mathbf{\Phi}$-equilibria in general normal-form games.

Furthermore, the significance of no-external-regret algorithms is underscored by the folk theorem, which posits that if all players employ external regret-minimizing algorithms to select their strategies, the players' time-average behavior converges to the set of coarse correlated equilibria [27, 37]. In addition, in the setting of two-player and polymatrix zero-sum games, the product of the players' individual time-averaged strategies converges to the set of Nash equilibria [10, 15, 16, 17].

3 No-Regret Learning in General Quantum Games

In this section we study QGs from the perspective of no-$\mathbf{\Phi}$-regret learning and provide an analogue to the classical CCE convergence result for no-external-regret learning.

Theorem 3.1 (Main Theorem)**.** *For any quantum game we have the following:*

(a) For any deviation mappings $\mathbf{\Phi} = \{\Phi_i\}_{i=1}^{k}$, the limit points of the time-averaged joint history of play $\left\{\frac{1}{T}\sum_{t=1}^{T}\bigotimes_i \rho_i^t\right\}_T$ generated by no-$\mathbf{\Phi}$-regret algorithms are separable Q$\mathbf{\Phi}$E. In particular, if all players update their strategies with a no-$\mathbf{\Phi}$-regret algorithm that guarantees a time-averaged regret of $\leq \epsilon$ after T timesteps, then the time-averaged joint history of play after T timesteps is a separable ϵ-Q$\mathbf{\Phi}$E.

(b) For any separable QCCE ρ^*, there exist no-external-regret algorithms for each player so that their time-averaged joint history of play $\left\{\frac{1}{T}\sum_{t=1}^{T}\bigotimes_i \rho_i^t\right\}_T$ converges to ρ^*.

In particular, Theorem 3.1 implies that for any quantum game the set of separable QCCEs is equal to the limit points of the time-averaged history $\overline{\rho}(T) := \frac{1}{T}\sum_{t=1}^{T}\rho^t$ of joint play of players using no-external-regret algorithms. We note that this is the best statement that can be written for learning QCCEs since taking the time-averaged history of joint play can only ever yield separable states (due to the fact that at each round a product state is played). On the other hand, because entangled QCCEs can exist (see Appendix A), this means that there exist QCCEs that are unlearnable via the current paradigm of no-regret learning.

We dedicate the rest of this subsection to proving Theorem 3.1, as well deriving an explicit convergence rate to separable ϵ-QCCEs when MMWU (Algorithm 1) is used.

Proof of Theorem 3.1(a). Suppose that after T iterations of running no-$\mathbf{\Phi}$-regret algorithms, every player has $\mathbf{\Phi}$-regret $\leq \epsilon = \epsilon(T)$. Let $\rho^t = \bigotimes_{i=1}^{k}\rho_i^t$ denote the strategy profile (product distribution) at time t, and let $\overline{\rho} = \overline{\rho}(T) := \frac{1}{T}\sum_{t=1}^{T}\rho^t$ be the time-averaged history of these strategy profiles. (It is thus a classical probability distribution over product distributions.)

That player i has $\mathbf{\Phi}$-regret $\leq \epsilon$ means that the player i's realized time-averaged utility is at least the time-averaged utility obtained by applying the channel $\phi_i \otimes \mathbb{I}_{-i}$ for any $\phi_i \in \Phi_i$, i.e.,

$$\frac{1}{T}\sum_t \operatorname{Tr}\left(R_i\Big(\bigotimes_i \rho_i^t\Big)\right) \geq \frac{1}{T}\sum_t \operatorname{Tr}\left(R_i(\phi_i \otimes \mathbb{I}_{-i})\Big(\bigotimes_i \rho_i^t\Big)\right) - \epsilon \qquad \forall \phi_i \in \Phi_i. \tag{10}$$

But player i's realized time-averaged utility is simply the (expected) utility he would get in one round if all the players play according to the time-averaged joint history ρ^* since

$$\frac{1}{T}\sum_t \operatorname{Tr}\left(R_i\Big(\bigotimes_i \rho_i^t\Big)\right) = \operatorname{Tr}\left(R_i\left(\frac{1}{T}\sum_t\bigotimes_i \rho_i^t\right)\right) = \operatorname{Tr}(R_i\overline{\rho}) = u_i(\overline{\rho}),$$

while on the right-hand side of (10) we have that

$$\frac{1}{T}\sum_t \operatorname{Tr}\left(R_i(\phi_i\otimes\mathbb{I}_{-i})\Big(\bigotimes_i \rho_i^t\Big)\right) = \operatorname{Tr}\left(R_i(\phi_i\otimes\mathbb{I}_{-i})\left(\frac{1}{T}\sum_t\bigotimes_i \rho_i^t\right)\right) = u_i((\phi_i\otimes\mathbb{I}_{-i})(\overline{\rho})).$$

Thus we have from the regret bound (10) written for each player i that

$$u_i(\overline{\rho}(T)) \geq u_i((\phi_i\otimes\mathbb{I}_{-i})(\overline{\rho}(T))) - \epsilon(T) \quad \forall i\in[k],\ \phi_i\in\Phi_i,$$

and taking the limit of these equations as $T\to\infty$ we get that

$$\lim_{T\to\infty} u_i(\overline{\rho}(T)) \geq \lim_{T\to\infty} u_i((\phi_i\otimes\mathbb{I}_{-i})(\overline{\rho}(T))) \quad \forall i\in[k],\ \phi_i\in\Phi_i.$$

Finally, where $\rho^* := \lim_{m\to\infty}\overline{\rho}(T_m)$ is any limit point of the no-regret play (here $(T_m)_{m=1}^\infty$ is a subsequence of $\mathbb{N}$ for which the subsequence $(\overline{\rho}(T_m))_{m=1}^\infty$ converges), we have from the continuity of the payoff functions u_i and the quantum channels $\phi_i\otimes\mathbb{I}_{-i}$ for all $i\in[k]$, $\phi_i\in\Phi_i$ that

$$u_i(\rho^*) = \lim_{m\to\infty} u_i(\overline{\rho}(T_m)) \geq \lim_{m\to\infty} u_i\left((\phi_i\otimes\mathbb{I}_{-i})(\overline{\rho}(T_m))\right) = u_i((\phi_i\otimes\mathbb{I}_{-i})(\rho^*)) \quad \forall i\in[k],\ \phi_i\in\Phi_i,$$

i.e. ρ^* is a Q**Φ**E. Since the set of separable states is compact and each $\overline{\rho}(T)$ is separable, so the limit point ρ^* is itself separable, and hence a separable Q**Φ**E. □

Proof of Theorem 3.1(b). Let $\rho^* = \sum_{j=1}^m \lambda_j \bigotimes_i \rho_{ij}$ be a separable QCCE. We can create a sequence of play $\{\bigotimes_i \rho_i^t\}_t$ that converges to ρ^* in terms of its time-averaged joint history as $t\to\infty$, i.e.

$$\lim_{T\to\infty}\overline{\rho}(T) = \rho^* \qquad \text{where} \qquad \overline{\rho}(T) := \frac{1}{T}\sum_{t=1}^T\bigotimes_i \rho_i^t,$$

by simply creating a sequence whose terms are in $[m]$ and whose frequencies converge to the distribution $(\lambda_j)_j$, then playing the product distribution $\bigotimes_i \rho_{ij}$ in time t if the t-th element of the sequence is i.

Now define the function $f : \bigotimes_i \mathcal{H}_i \to \mathbb{R}$ such that

$$f(\rho) := \max_i \sup_{\rho_i'\in\mathcal{H}_i}\big[u_i(\rho_i'\otimes\operatorname{Tr}_i\rho) - u_i(\rho)\big].$$

The function f is continuous by Lemma C.2 since the functions $h_i : (\bigotimes_{i'}\mathcal{H}_{i'})\times\mathcal{H}_i$, $h_i(\rho,\rho_i') := u_i(\rho_i'\otimes\operatorname{Tr}_i\rho)$ are continuous $\forall i$, and so we have that $\lim_{T\to\infty} f(\overline{\rho}(T))$ exists and

$$\lim_{T\to\infty} f(\overline{\rho}(T)) = f\left(\lim_{T\to\infty}\overline{\rho}(T)\right) = f(\rho^*) \leq 0,$$

with the last inequality due to the fact that ρ^* is a QCCE.

But the value of the function $f(\overline{\rho}(T))$ is equal to the maximal time-averaged regret that any player obtains up till time T, since

$$\begin{aligned}\text{regret}_i^T &= \sup_{\rho_i'\in\mathcal{H}_i}\frac{1}{T}\sum_{t=1}^T\left[u_i\left(\rho_i'\otimes\left(\bigotimes_{i'\neq i}\rho_{i'}^t\right)\right) - u_i\left(\bigotimes_{i'}\rho_{i'}^t\right)\right]\\ &= \sup_{\rho_i'\in\mathcal{H}_i}\left[u_i\left(\rho_i'\otimes\left(\frac{1}{T}\sum_{t=1}^T\bigotimes_{i'\neq i}\rho_{i'}^t\right)\right) - u_i\left(\frac{1}{T}\sum_{t=1}^T\bigotimes_{i'}\rho_{i'}^t\right)\right]\\ &= \sup_{\rho_i'\in\mathcal{H}_i}\big[u_i(\rho_i'\otimes\operatorname{Tr}_i\overline{\rho}(T)) - u_i(\overline{\rho}(T))\big].\end{aligned}$$

Thus the maximal regret that any player obtains up till time T converges to a value ≤ 0 as $T \to \infty$, i.e. the sequence of play is no-regret.

Finally, the no-regret algorithms that converge in time-averaged joint history to ρ^* can be defined as follows: for player i, for time $t = 1$ play ρ_i^1 (from the above sequence of play), and for time $t \geq 2$ check if all other players i' have played according to the sequence $(\rho_{i'}^\tau)_\tau$ for all $\tau < t$. If YES, continue playing ρ_i^t; if NO, default to a guaranteed no-regret algorithm (e.g., MMWU) for all future time. □

Finally, since we have explicit no-external regret algorithms for learning in quantum games, we can give an explicit convergence rate to QCCEs obtained by all players using one such algorithm (MMWU):

Remark 3.1. *For a quantum game, if all utilities are in $[-1, 1]$ and all players use MMWU with stepsize $\eta = \frac{\epsilon}{2}$ to update their strategies, then for any $\epsilon \leq 2$ their time-averaged joint history of play $\frac{1}{T}\sum_{t=1}^T \bigotimes_i \rho_i^t$ after $T = \frac{4 \ln n}{\epsilon^2}$ steps is a separable ϵ-QCCE.*

To see why this is the case, recall from Section 2.1 that the utility of player i at time t can be written as $u_i(\rho^t) = \operatorname{Tr}(\rho^t R_i) = \left\langle \rho_i^t, \Theta_i({\rho_{-i}^t}^\top)\right\rangle$ where Θ_i and R_i are related via the Choi-Jamiołkowski isomorphism. The assumption that the utilities are in $[-1, 1]$ implies that, at each time t, the eigenvalues of player i's gain matrix $\Theta_i({\rho_{-i}^t}^\top)$ are in $[-1, 1]$ for each player i. Then, if player i were to run MMWU with fixed stepsize $\eta \leq 1$ for T timesteps, she would accumulate average regret $\leq \eta + \frac{\ln n}{\eta T}$ (see, e.g. [34]). Thus, after $T = \frac{4 \ln n}{\epsilon^2}$ timesteps of running MMWU with fixed stepsize $\eta = \frac{\epsilon}{2}$, she is guaranteed to have average regret $\leq \epsilon$. Finally, by Theorem 3.1(a), the time-averaged joint history of play is a separable ϵ-QCCE.

4 No-Regret Learning in Two-Player Quantum Zero-Sum Games

Next, we consider the special case of quantum zero-sum games, where the utility is defined such that the sum of all players' payoffs is always zero. In this section, we restrict ourselves to the two-player case in order to present an analogue of a standard classical result - that no-external-regret algorithms go to the set of Nash equilibria in two-player zero-sum games. The main result in this section shows that no-external-regret algorithms converge to the set of QNE in two-player quantum zero-sum games.

In a two-player quantum zero-sum game, Alice and Bob play density matrices $\rho \in \mathrm{D}(\mathcal{A})$ and $\sigma \in \mathrm{D}(\mathcal{B})$ respectively. For notational simplicity, we depart from previous convention and say that Alice's payoff is $u_A(\rho, \sigma) = \langle \rho, \Theta(\sigma)\rangle = \operatorname{Tr}\big(R(\rho \otimes \sigma^\top)\big)$, where $\Theta : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$ and $R \in \mathrm{L}(\mathcal{A} \otimes \mathcal{B})$ are related by the Choi-Jamiołkowski isomorphism. By the definition of zero-sum, Bob receives payoff $u_B(\rho, \sigma) = -\langle \rho, \Theta(\sigma)\rangle$. We begin by showing that QNEs attain the value of the game in a two-player quantum zero-sum game.

Theorem 4.1 (Quantum Minimax Theorem)**.** *Every two-player quantum zero-sum game has a well-defined value, i.e., all QNEs attain the same utility*

$$v = \max_{\rho \in \mathrm{D}(\mathcal{A})} \min_{\sigma \in \mathrm{D}(\mathcal{B})} \langle \rho, \Theta(\sigma)\rangle = \min_{\sigma \in \mathrm{D}(\mathcal{B})} \max_{\rho \in \mathrm{D}(\mathcal{A})} \langle \Theta^\dagger(\rho), \sigma\rangle. \tag{11}$$

Moreover, the set of QNEs is the product of two spectrahedra, i.e.,

$$\mathit{QNEs} = \{\rho \in \mathrm{D}(\mathcal{A}) : \Theta^\dagger(\rho) \succeq vI_{\mathcal{B}}\} \times \{\sigma \in \mathrm{D}(\mathcal{B}) : \Theta(\sigma) \preceq vI_{\mathcal{A}}\}. \tag{12}$$

Proof. The equivalence of max-min and min-max in Equation (11) comes as a direct consequence of Von Neumann's minimax theorem which holds for compact convex sets [38, 39]. Equation (12) was proven in [25], but for completeness we provide a simplified proof of it here that also proves along the way that all QNEs attain the max-min value. First, introducing an auxiliary variable t, the max-min term in (11) can be rewritten as

$$\begin{aligned}\max_{\rho,t} \quad & t \\ \text{s.t.} \quad & \langle \rho, \Theta(\sigma)\rangle \geq t \quad \forall \sigma \in \mathrm{D}(\mathcal{B}) \\ & \rho \in \mathrm{D}(\mathcal{A}),\end{aligned} \tag{13}$$

which can in turn be rewritten as

$$\max_{\rho,t} \left\{ t : \Theta^\dagger(\rho) \succeq tI_{\mathcal{B}},\ \rho \in \mathrm{D}(\mathcal{A}) \right\}. \tag{14}$$

The dual of this semidefinite program is given by

$$\min_{\sigma,t'} \left\{ t' : \Theta(\sigma) \preceq t' I_{\mathcal{A}},\ \sigma \in \mathrm{D}(\mathcal{B}) \right\}. \tag{15}$$

For proving one direction of Equation (12), suppose that (ρ^*, σ^*) is a QNE, i.e.,

$$\lambda_{\max}(\Theta(\sigma^*)) = \langle \rho^*, \Theta(\sigma^*)\rangle = \lambda_{\min}(\Theta^\dagger(\rho^*)).$$

This implies that $\Theta(\sigma^*) \preceq \langle \rho^*, \Theta(\sigma^*)\rangle I_{\mathcal{A}}$ and $\langle \rho^*, \Theta(\sigma^*)\rangle I_{\mathcal{B}} \preceq \Theta^\dagger(\rho^*)$, which respectively imply that $(\sigma^*, \langle \rho^*, \Theta(\sigma^*)\rangle)$ is feasible for (15) and $(\rho^*, \langle \rho^*, \Theta(\sigma^*)\rangle)$ is feasible for (14). But since these programs are a primal-dual pair and the primal-dual feasible solutions attain the same objective value, we have that the dual-feasible solution $(\sigma^*, \langle \rho^*, \Theta(\sigma^*)\rangle)$ and the primal-feasible solution $(\rho^*, \langle \rho^*, \Theta(\sigma^*)\rangle)$ are in fact optimal for the dual (15) and the primal (14) respectively. This means that the utility $\langle \rho^*, \Theta(\sigma^*)\rangle$ is equal to the max-min value v, and that (ρ^*, σ^*) satisfies $\Theta(\sigma^*) \preceq vI_{\mathcal{A}}$, $\Theta^\dagger(\rho^*) \succeq vI_{\mathcal{B}}$.

To prove the other set inclusion in Equation (12), suppose that (ρ^*, σ^*) satisfies

$$\Theta^\dagger(\rho^*) \succeq vI_{\mathcal{B}}, \qquad \Theta(\sigma^*) \preceq vI_{\mathcal{A}}.$$

Taking inner product of the first inequality with σ^* and the second inequality with ρ^* gives us that $\langle \rho^*, \Theta(\sigma^*)\rangle \geq v$ and $\langle \rho^*, \Theta(\sigma^*)\rangle \leq v$ respectively, which together imply that $\langle \rho^*, \Theta(\sigma^*)\rangle = v$. Substituting this fact back into the two inequalities gives

$$\lambda_{\min}(\Theta^\dagger(\rho^*)) \geq \langle \rho^*, \Theta(\sigma^*)\rangle, \qquad \lambda_{\max}(\Theta(\sigma^*)) \leq \langle \rho^*, \Theta(\sigma^*)\rangle,$$

i.e., that (ρ^*, σ^*) is a QNE. □

We can use Theorem 4.1 to show the main result of this section, that no-external-regret dynamics converge to QNE in two-player quantum zero-sum games.

Theorem 4.2. *For any two-player quantum zero-sum game, the limit points of the product of time-averaged individual histories of play* $\left\{ \frac{1}{T}\sum_{t=1}^T \rho_t, \frac{1}{T}\sum_{t=1}^T \sigma_t \right\}_T$ *generated by no-external-regret algorithms are separable QNE. In particular, if all players update their strategies with a no-external-regret algorithm that guarantees a time-averaged regret of* $\leq \epsilon$ *after* T *timesteps, then the time-averaged individual sequences of play after* T *timesteps is a separable* 2ϵ*-QNE.*

Proof. For any no-external-regret algorithm, we can select parameters such that, at time T, each player's time-averaged regret is at most ϵ. Consider the regret for the sequences of play of Alice (denoted by ρ) and Bob (denoted by σ) respectively. Them we have that:

$$\min_{\sigma} \frac{1}{T}\sum_{t=1}^T \langle \rho_t, \Theta(\sigma)\rangle + \epsilon \geq \frac{1}{T}\sum_{t=1}^T \langle \rho_t, \Theta(\sigma_t)\rangle \geq \max_{\rho} \frac{1}{T}\sum_{t=1}^T \langle \rho, \Theta(\sigma_t)\rangle - \epsilon. \tag{16}$$

Next, let $\overline{\rho} = \frac{1}{T}\sum_{t=1}^{T}\rho_t$ and $\overline{\sigma} = \frac{1}{T}\sum_{t=1}^{T}\sigma_t$ so that Equation 16 can be written as

$$\min_{\sigma}\langle\overline{\rho}, \Theta(\sigma)\rangle + \epsilon \geq \max_{\rho}\langle\Theta^{\dagger}(\rho), \overline{\sigma}\rangle - \epsilon \tag{17}$$

By taking the maximum over ρ for Equation 17, we obtain the following:

$$\begin{aligned}\max_{\rho}\min_{\sigma}\langle\rho, \Theta(\sigma)\rangle &\geq \min_{\sigma}\langle\overline{\rho}, \Theta(\sigma)\rangle \\ &\geq \max_{\rho}\langle\Theta^{\dagger}(\rho), \overline{\sigma}\rangle - 2\epsilon \\ &\geq \min_{\sigma}\max_{\rho}\langle\Theta^{\dagger}(\rho), \sigma\rangle - 2\epsilon\end{aligned}$$

By Theorem 4.1, the left-hand side of the inequality is v, the value of the game. Now let us consider Nash strategies, which are strategies for each player that achieve the minimax value regardless of the other player's strategy. Thus, since the time-average value of σ, $\overline{\sigma}$, satisfies the maximin inequalities above up to a 2ϵ error, it is a 2ϵ-approximate Nash strategy for Bob by Theorem 4.1. A similar argument holds for the case of Alice with $\overline{\rho}$, and thus we have that the time average values $(\overline{\rho}, \overline{\sigma})$ are a 2ϵ-QNE strategy of the zero-sum game, and $\langle\overline{\rho}, \Theta(\overline{\sigma})\rangle$ is the 2ϵ-equilibrium value of the game.

Finally, Theorem 3.1(b), we have that for any separable QCCE there exists no-external-regret algorithms that converge to that QCCE. Thereafter, we can take the marginals over the players' joint history of play to obtain a QNE of the game. □

We can, with similar reasoning to Remark 3.1, give an explicit convergence rate for players using MMWU to update their strategies:

Remark 4.1. *For any two-player quantum zero-sum game, if utilities are in $[-1, 1]$ and all players use MMWU with fixed stepsize $\eta = \frac{\epsilon}{4}$ to update their strategies, for any $\epsilon \leq 4$, the product of their time-averaged individual sequences of play $\left(\frac{1}{T}\sum_{t=1}^{T}\rho_t, \frac{1}{T}\sum_{t=1}^{T}\sigma_t\right)$ after $T = \frac{16\ln n}{\epsilon^2}$ steps is an ϵ-QNE.*

5 No-Regret Learning in Polymatrix Quantum Zero-Sum Games

A key question in the quantum game regime is whether there exist classes of multiplayer games where quantum Nash equilibria are tractable and can be converged to via no-regret learning. As it turns out, in classical polymatrix zero-sum games, [16, 17] show that no-external-regret learning converges in time-average to Nash equilibria. In this section we show an analogous result: in polymatrix quantum zero-sum games, no-external-regret learning converges in time-average to approximate QNE. In order to show this, we first need to define the notion of a polymatrix quantum zero-sum game.

Definition 5.1. *A polymatrix quantum game $\mathcal{G}$ is a game defined on an undirected graph (V, E) such that the following holds:*

- *The vertices (or nodes) $V = \{1, \dots, k\}$ represent players, and edges E represent two-player quantum games (QG) between a pair of players (i, j), where $i \neq j$.*
- *Each player $i \in V$ has register $\mathcal{H}_i$.*
- *For each edge $(i, j) \in E$, we associate a two-player quantum game (QG) between players i and j where player i has register $\mathcal{H}_i$ and utility tensor R_{ij}, while player j has register $\mathcal{H}_j$ and utility tensor R_{ji}. Where $\rho_{ij} := \mathrm{Tr}_{-ij}\,\rho = \mathrm{Tr}_i(\mathrm{Tr}_{-j}\,\rho)) \in \mathrm{D}(\mathcal{H}_i \otimes \mathcal{H}_j)$ is the joint state of the two players' registers, player i's utility from this two-player game is then $u_{ij}(\rho_{ij}) = \mathrm{Tr}(\rho_{ij}R_{ij})$, while player j's utility from this two-player game is $u_{ji}(\rho_{ij}) = \mathrm{Tr}(\rho_{ij}R_{ji})$.*
- *For each joint state $\rho \in \mathrm{D}(\bigotimes_i \mathcal{H}_i)$, the total utility attained by player $i \in V$ under ρ is $u_i(\rho) = \sum_{(i,j)\in E} u_{ij}(\rho_{ij}) = \sum_{(i,j)\in E}\mathrm{Tr}(\rho_{ij}R_{ij})$.*

- *Finally, the game $\mathcal{G}$ is called zero-sum if for all joint states $\rho \in \mathrm{D}(\bigotimes_i \mathcal{H}_i)$, we have that $\sum_{i\in V} u_i(\rho) = 0$.*

Note that this definition refers to the zero-sum property of the game in a global sense, as opposed to the stronger notion of *pairwise* zero-sum polymatrix quantum games, the definition of which includes the additional constraint that every two-player edge game is a quantum zero-sum game.

We next establish a lemma stating that a polymatrix quantum game is also a quantum game in the sense of our earlier definition.

Lemma 5.1. *A polymatrix quantum game is also a quantum game in the sense of* (QG)*.*

Proof. The utility of player i can be expressed as

$$u_i(\rho) = \sum_{j:(i,j)\in E} u_{ij}(\rho_{ij}) = \sum_{j:(i,j)\in E} \mathrm{Tr}(\rho_{ij} R_{ij}).$$

Subsequently,

$$\sum_{j:(i,j)\in E} \mathrm{Tr}(\rho_{ij} R_{ij}) = \sum_{j:(i,j)\in E} \mathrm{Tr}((\mathrm{Tr}_{-ij}\,\rho) R_{ij}) = \sum_{j:(i,j)\in E} \langle \mathrm{Tr}_{-ij}\,\rho, R_{ij}\rangle\,,$$

and finally

$$\sum_{j:(i,j)\in E} \langle \mathrm{Tr}_{-ij}\,\rho, R_{ij}\rangle = \sum_{j:(i,j)\in E} \langle \rho, R_{ij}\otimes I_{-ij}\rangle = \mathrm{Tr}\left(\left(\sum_{j:(i,j)\in E} R_{ij}\otimes I_{-ij}\right)\rho\right),$$

so setting $R_i := \sum_{j:(i,j)\in E} R_{ij}\otimes I_{-ij}$ gives $u_i(\rho) = \mathrm{Tr}(R_i\rho)$ $\forall i$, which fits into the QG formulation. □

Next, we prove a property connecting QCCEs and QNEs in the class of polymatrix quantum zero-sum games.

Lemma 5.2. *Let $\mathcal{G}$ be a polymatrix quantum zero-sum game. For any joint state ρ that is a QCCE, its* marginalized state *$\hat{\rho}$ defined by*

$$\hat{\rho} = \bigotimes_{i\in[n]} \hat{\rho}_i, \qquad \hat{\rho}_i = \mathrm{Tr}_{-i}\,\rho$$

is a QNE of $\mathcal{G}$.

Proof. First note that $\forall i \in [k]$, $\forall \rho_i' \in \mathrm{D}(\mathcal{H}_i)$ we have that

$$u_i(\rho_i' \otimes \hat{\rho}_{-i}) = u_i(\rho_i' \otimes \mathrm{Tr}_i\,\rho). \tag{18}$$

This is because if the joint state on all registers is $\rho_i' \otimes \hat{\rho}_{-i}$, then for each $j : (i,j) \in E$ the joint state on player i and j's register is $\mathrm{Tr}_{-ij}(\rho_i' \otimes \hat{\rho}_{-i}) = \rho_i' \otimes \mathrm{Tr}_{-j}(\hat{\rho}_{-i}) = \rho_i' \otimes \hat{\rho}_j$, and thus on the left-hand side of the equation we have that player i's expected utility given this joint state is $u_i(\rho_i' \otimes \hat{\rho}_{-i}) = \sum_{j:(i,j)\in E} u_{ij}(\rho_i' \otimes \hat{\rho}_j)$. On the other hand, if the joint state on all registers is $\rho_i' \otimes \mathrm{Tr}_i\,\rho$, then for each $j : (i,j) \in E$ we have that the joint state on player i and j's register is also $\mathrm{Tr}_{-ij}(\rho_i' \otimes \mathrm{Tr}_i\,\rho) = \rho_i' \otimes \mathrm{Tr}_{-j}(\mathrm{Tr}_i\,\rho) = \rho_i' \otimes \mathrm{Tr}_{-j}\,\rho = \rho_i' \otimes \hat{\rho}_j$, so on the right hand side of the equation we have that player i's expected utility given this joint state is also $u_i(\rho_i' \otimes \mathrm{Tr}_i\,\rho) = \sum_{j:(i,j)\in E} u_{ij}(\rho_i' \otimes \hat{\rho}_j)$.

Now fix an $i \in [k]$ and a $\rho_i' \in \mathrm{D}(\mathcal{H}_i)$. By (18) and the fact that ρ is a QCCE, we have that $u_i(\rho) \geq u_i(\rho_i' \otimes \mathrm{Tr}_i\,\rho) = u_i(\rho_i' \otimes \hat{\rho}_{-i})$, and also that $u_j(\rho) \geq u_j(\hat{\rho}_j \otimes \mathrm{Tr}_j\,\rho) = u_j(\hat{\rho})$ $\forall j$. Then, summing up the utilities attained by each player on the joint state ρ and using the fact the $\mathcal{G}$ is zero-sum, we have that

$$0 = \sum_{j \in [n]} u_j(\rho) = \sum_{j \neq i} u_j(\rho) + u_i(\rho) \geq \sum_{j \neq i} u_j(\hat{\rho}) + u_i(\rho_i' \otimes \hat{\rho}_{-i}) = -u_i(\hat{\rho}) + u_i(\rho_i' \otimes \hat{\rho}_{-i}),$$

i.e., that $u_i(\hat{\rho}) \geq u_i(\rho_i' \otimes \hat{\rho}_{-i})$. Since this holds for any given $i \in [n]$ and $\rho_i' \in \mathrm{D}(\mathcal{H}_i)$, $\hat{\rho}$ is a QNE. □

Finally, we use the previously proved result about convergence to QCCEs in general quantum games, in conjunction with the lemmas presented above, to prove convergence of no-external-regret algorithms to QNEs in polymatrix quantum zero-sum games. This generalizes the analogous result of [16] from classical polymatrix zero-sum games to the quantum setting.

Theorem 5.1. *If all players in a polymatrix quantum zero-sum game (Definition 5.1) use no-external-regret algorithms, then the product of their time-averaged individual histories of play converges to the set of QNE. In particular, if all players update their strategies with a no-external-regret algorithm that guarantees time-averaged regret of $\leq \epsilon$ after T timesteps, then the time-averaged joint history of play after T timesteps is a separable $k\epsilon$-QNE.*

Proof. Suppose that after T iterations of running no-external-regret algorithms, every player has time-average regret $\leq \epsilon = \epsilon(T)$. Let $\rho^t = \bigotimes_{i=1}^k \rho_i^t$ denote the strategy profile (product distribution) at time t, and let $\overline{\rho} = \overline{\rho}(T) := \frac{1}{T}\sum_{t=1}^T \rho^t$ be the time-averaged history of these strategy profiles. Moreover, note that if players i and j play with strategies ρ_i and ρ_j respectively, we can write the utility for player i in the form $u_{ij}(\rho_i \otimes \rho_j)$. For any $\rho_i' \in \mathrm{D}(\mathcal{H}_i)$ we can write:

$$\frac{1}{T}\sum_{t=1}^T \sum_{(i,j)\in E} u_{ij}(\rho_i' \otimes \rho_j^t) = \sum_{(i,j)\in E} u_{ij}(\rho_i' \otimes \overline{\rho}_j)$$

Let z_i be the best response of i if all other players use $\overline{\rho}_j$. Then for all i and any $\rho_i' \in \mathrm{D}(\mathcal{H}_i)$,

$$\sum_{(i,j)\in E} u_{ij}(z_i \otimes \overline{\rho}_j) \geq \sum_{(i,j)\in E} u_{ij}(\rho_i' \otimes \overline{\rho}_j).$$

Next, by the no-external-regret property, we have that

$$\frac{1}{T}\sum_{t=1}^T \sum_{(i,j)\in E} u_{ij}(\rho_i^t \otimes \rho_j^t) \geq \frac{1}{T}\sum_{t=1}^T \left(\sum_{(i,j)\in E} u_{ij}(z_i \otimes \rho_j^t) \right) - \epsilon = \sum_{(i,j)\in E} u_{ij}(z_i \otimes \overline{\rho}_j) - \epsilon$$

Summing both sides of the above over all $i \in V$, we have from the LHS that

$$\sum_{i\in V} \left(\frac{1}{T}\sum_{t=1}^T \sum_{(i,j)\in E} u_{ij}(\rho_i^t \otimes \rho_j^t) \right) = \frac{1}{T}\sum_{t=1}^T \left(\sum_{i\in V} \sum_{(i,j)\in E} u_{ij}(\rho_i^t \otimes \rho_j^t) \right) = 0,$$

which is due to the global zero-sum property of the quantum polymatrix game. Moreover, the sum on the RHS is given by

$$\sum_{i\in V} \sum_{(i,j)\in E} u_{ij}(z_i \otimes \overline{\rho}_j) - k\epsilon$$

since there are k players. Combining the two, and using the fact that the LHS is at least as large as the RHS,

$$0 \geq \sum_{i\in V} \sum_{(i,j)\in E} u_{ij}(z_i \otimes \overline{\rho}_j) - k\epsilon \implies k\epsilon \geq \sum_{i\in V} \sum_{(i,j)\in E} u_{ij}(z_i \otimes \overline{\rho}_j).$$

We now show that each player i playing $\overline{\rho}_i$ is a $k\epsilon$- approximate QNE. Note that if each player i plays $\overline{\rho}_i$, the sum of all players' payoffs is 0, i.e.

$$\sum_{i\in V}\sum_{(i,j)\in E} u_{ij}(\overline{\rho}_i \otimes \overline{\rho}_j) = 0.$$

Hence we have that

$$k\epsilon \geq \sum_{i\in V}\left(\sum_{(i,j)\in E} u_{ij}(z_i \otimes \overline{\rho}_j) - \sum_{(i,j)\in E} u_{ij}(\overline{\rho}_i \otimes \overline{\rho}_j)\right)$$

However, the sum is over non-negative numbers since the z_is are best responses. We have a sum of non-negative numbers bounded by $k\epsilon$, so for any $i \in V$,

$$k\epsilon \geq \sum_{(i,j)\in E} u_{ij}(z_i \otimes \overline{\rho}_j) - \sum_{(i,j)\in E} u_{ij}(\overline{\rho}_i \otimes \overline{\rho}_j) \geq 0.$$

Thus, for all i, if all other players j play $\overline{\rho}_j$, the payoff given by playing the best response is at most $k\epsilon$ better than the payoff obtained by playing $\overline{\rho}_i$. Hence it is a $k\epsilon$-QNE for each player j to play $\overline{\rho}_j$. □

This result gives us a decentralized way of arriving at quantum Nash equilibria in a broader class of multi-player games, i.e., that of polymatrix quantum zero-sum games. Exploring if there are other classes of multi-player games for which QNE are tractable is left to future work.

Remark 5.1. *Consider a k-player polymatrix quantum zero-sum game with utilities in $[-1, 1]$ and let n be the largest dimension of the players' registers. For any $\epsilon \leq 2k$, if each player uses MMWU with fixed stepsize $\eta = \frac{\epsilon}{2k}$, the product of their time-averaged individual sequences of play $\left(\frac{1}{T}\sum_{t=1}^{T}\rho_1^t, \dots, \frac{1}{T}\sum_{t=1}^{T}\rho_k^t\right)$ after $T = \frac{4k^2 \ln n}{\epsilon^2}$ steps is an ϵ-QNE.*

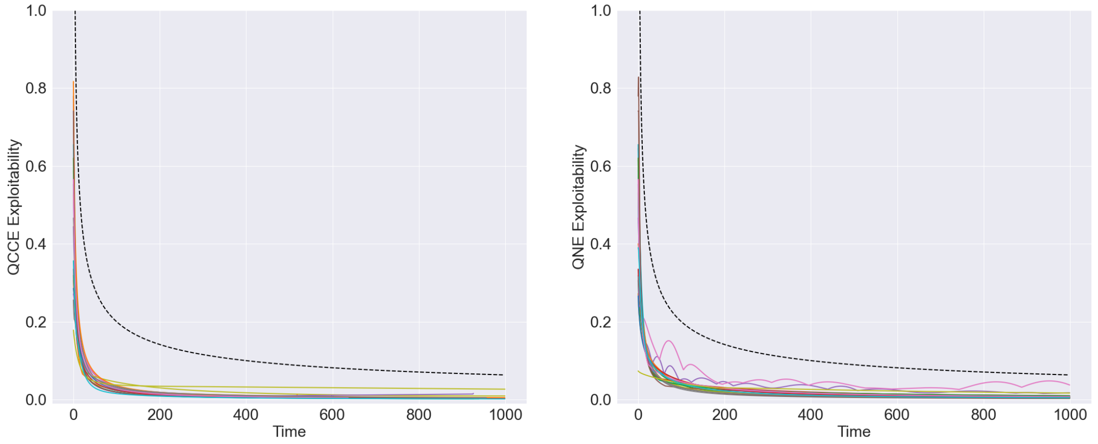
The reasoning for the above convergence rate is similar to Remark 3.1. However, since an algorithm that achieves ϵ-regret gives a $k\epsilon$-QNE, we require running the algorithm until $\frac{\epsilon}{k}$ regret is achieved instead.

6 MMWU Experiments

In this section, we consider learning using the specific no-external-regret algorithm, MMWU (Algorithm 1), and present several experiments that corroborate our theoretical results about time-averaged convergence to equilibria. For two-player zero-sum quantum games, we also present some plots showcasing the day-to-day behavior of the iterates.

First, in Figure 1 we show the exploitability (as defined in Section 2.2) of MMWU in both general and zero-sum quantum games. For the case of general games, we consider the maximum individual exploitability of the time-averaged joint strategy for both players, which we term the "QCCE-exploitability" of the players' strategies, while in the case of zero-sum games we consider the maximum individual exploitability of the product of the time-averaged individual sequences of play, which we term the "QNE-exploitability". We are concerned with the maximum over the individual exploitabilities of each of the players since if each player attains ϵ-exploitability, then all players are at an ϵ-QCCE/QNE. In both plots, we use the doubling trick to run MMWU. The maximum individual exploitabilities go to zero or remain close to zero, implying time-averaged convergence to QCCE and QNE respectively.

Next, we present some indicative examples that elucidate the behavior of MMWU in two-player quantum zero-sum games. We see that in general, the trajectories of the joint state of the players either oscillate or go to a point on the boundary, and showcase this behavior alongside the time-averaged values of the trajectories in Figures 2 and 3. In order to represent time on the Bloch sphere, we use a color gradient from green to blue (light green denotes time $t = 0$, dark blue

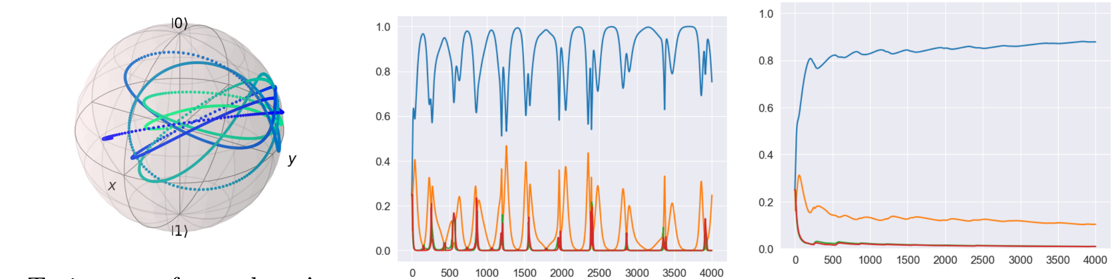


General quantum games

Zero-sum quantum games

Figure 1: Maximum individual exploitability of time-averaged strategies of players using MMWU in 20 randomly generated $\mathbb{C}^2 \otimes \mathbb{C}^2$ quantum games. The black dotted line denotes the theoretical upper-bound on the exploitability.

denotes time $t = 4000$). From the examples, even in the relatively well-studied case of MMWU, it is clear that some interesting types of behavior can be observed in quantum zero-sum games and beyond. The code used to generate our MMWU experiments can be found in the following Github repository: https://github.com/ryanndelion/No-Regret-Learning-Quantum-Games.

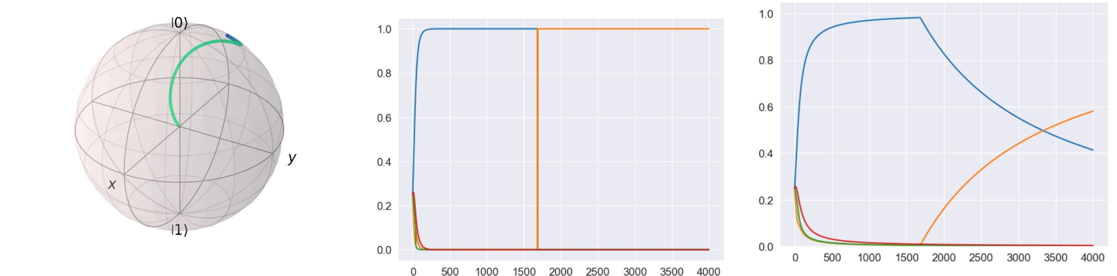


Trajectory of one player's strategy plotted on the Bloch sphere

Eigenvalues of the players' joint state over time

Eigenvalues of the time-averaged joint state over time

Figure 2: Example of oscillatory behaviour of MMWU in two-player quantum zero-sum games. Time is represented using a gradient from green to blue on the Bloch sphere.



Trajectory of one player's strategy plotted on the Bloch sphere

Eigenvalues of the players' joint state over time

Eigenvalues of the time-averaged joint state over time

Figure 3: Example of MMWU converging to the boundary (i.e., pure states) in two-player quantum zero-sum games. Time is represented using a gradient from green to blue on the Bloch sphere.

7 Discussion and Future Work

In this work we provide a general class of quantum games that fits with and subsumes prior formulations. We explore equilibrium notions in this class of games, inspired by classical solution concepts and **Φ**-regret and show an interesting analogy between deviation maps in the classical and quantum settings. We introduce quantum coarse correlated equilibria and show that for general quantum games, the set of separable QΦEs is actually the set of limits points of the time-averaged distribution of joint play when players use no-**Φ**-regret algorithms. Moreover, in the two-player and polymatrix zero-sum cases, no-regret algorithms result in convergence to quantum Nash equilibria. Overall, this indicates a rich connection between the worlds of online optimization, classical learning in games, and quantum information theory.

An interesting future direction of our work is to study general **Φ**-equilibria in other classes of quantum games, and the capability of modifications to the standard no-regret learning paradigm that can arrive at these equilibria. Specifically, designing implementable algorithms that converge to quantum correlated equilibria remains an important task, given that similar approaches have been successful in the classical regime. Additionally, the quantum game formulation allows for entangled equilibria not reachable via the standard paradigm of learning in games, examples of which were constructed in Appendix A. Investigating these entangled equilibria and how they can be computed or learnt by distributed agents is a tantalizing direction for future work. In the setting of classical games, several approaches have utilized coupled or correlated mechanisms to converge to different (and often better) equilibria than their uncoupled counterparts [40, 41]. Thus, studying a variant of no-regret learning which utilizes a mediator or correlating mechanism seems to be a reasonable initial approach to learning entangled equilibria in our formulation of quantum games.

Acknowledgments

This research is supported in part by the National Research Foundation, Singapore and the Agency for Science, Technology and Research (A*STAR) under its Quantum Engineering Programme NRF2021-QEP2-02-P05, and by the National Research Foundation, Singapore and DSO National Laboratories under its AI Singapore Program (AISG Award No: AISG2-RP-2020-016), grant PIESGP-AI-2020-01, AME Programmatic Fund (Grant No.A20H6b0151) from A*STAR and Provost's Chair Professorship grant RGEPPV2101. Wayne Lin and Ryann Sim gratefully acknowledge support from the SUTD President's Graduate Fellowship (SUTD-PGF).

References

[1] Gus Gutoski and John Watrous. "Toward a general theory of quantum games". In Proceedings of the thirty-ninth annual ACM symposium on Theory of computing. Pages 565–574. (2007).

[2] John Bostanci and John Watrous. "Quantum game theory and the complexity of approximating quantum Nash equilibria". Quantum **6**, 882 (2022).

[3] Shengyu Zhang. "Quantum strategic game theory". In Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. Pages 39–59. (2012).

[4] Jinshan Wu. "A new mathematical representation of game theory, I" (2004). arXiv:quant-ph/0404159.

[5] Faisal Shah Khan, Neal Solmeyer, Radhakrishnan Balu, and Travis S Humble. "Quantum games: a review of the history, current state, and interpretation". Quantum Information Processing **17**, 1–42 (2018).

[6] Jens Eisert and Martin Wilkens. "Quantum games". Journal of Modern Optics **47**, 2543–2556 (2000).

[7] Luca Marinatto and Tullio Weber. "A quantum approach to static games of complete information". Physics Letters A **272**, 291–303 (2000).

[8] Rahul Jain and John Watrous. "Parallel approximation of non-interactive zero-sum quantum games". In 2009 24th Annual IEEE Conference on Computational Complexity. Pages 243–253. IEEE (2009).

[9] Zhaohui Wei and Shengyu Zhang. "Full characterization of quantum correlated equilibria". Quantum Inf. Comput. **13**, 846–860 (2013).

[10] Nicolo Cesa-Bianchi and Gábor Lugosi. "Prediction, learning, and games". Cambridge university press. (2006).

[11] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. "Algorithmic game theory". Cambridge university press. (2007).

[12] Kyriakos Lotidis, Panayotis Mertikopoulos, and Nicholas Bambos. "Learning in quantum games" (2023). arXiv:2302.02333.

[13] Rahul Jain, Georgios Piliouras, and Ryann Sim. "Matrix multiplicative weights updates in quantum zero-sum games: Conservation laws & recurrence" (2022). arXiv:2211.01681.

[14] Wayne Lin, Georgios Piliouras, Ryann Sim, and Antonios Varvitsiotis. "Quantum potential games, replicator dynamics, and the separability problem" (2023). arXiv:2302.04789.

[15] Yoav Freund and Robert E Schapire. "Adaptive game playing using multiplicative weights". Games and Economic Behavior **29**, 79–103 (1999).

[16] Yang Cai and Constantinos Daskalakis. "On minmax theorems for multiplayer games". In Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete algorithms. Pages 217–234. SIAM (2011).

[17] Constantinos Daskalakis and Christos H Papadimitriou. "On a network generalization of the minmax theorem". In International Colloquium on Automata, Languages, and Programming. Pages 423–434. Springer (2009).

[18] Amy Greenwald and Amir Jafari. "A general class of no-regret learning algorithms and game-theoretic equilibria". In COLT. Volume 3, pages 2–12. (2003).

[19] Sanjeev Arora, Elad Hazan, and Satyen Kale. "The multiplicative weights update method: a meta-algorithm and applications". Theory of computing **8**, 121–164 (2012).

[20] Georgios Piliouras and Jeff S Shamma. "Optimization despite chaos: Convex relaxations to complex limit sets via Poincaré recurrence". In Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms. Pages 861–873. SIAM (2014).

[21] Victor Boone and Georgios Piliouras. "From Darwin to Poincaré and von Neumann: Recurrence and cycles in evolutionary and algorithmic game theory". In Web and Internet Economics: 15th International Conference, WINE 2019, New York, NY, USA, December 10–12, 2019, Proceedings 15. Pages 85–99. Springer (2019).

[22] Panayotis Mertikopoulos, Christos Papadimitriou, and Georgios Piliouras. "Cycles in adversarial regularized learning". In Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms. Pages 2703–2717. SIAM (2018).

[23] Geoffrey J Gordon, Amy Greenwald, and Casey Marks. "No-regret learning in convex games". In Proceedings of the 25th international conference on Machine learning. Pages 360–367. (2008).

[24] Gilles Stoltz and Gábor Lugosi. "Learning correlated equilibria in games with compact sets of strategies". Games and Economic Behavior **59**, 187–208 (2007).

[25] Constantin Ickstadt, Thorsten Theobald, and Elias Tsigaridas. "Semidefinite games" (2022). arXiv:2202.12035.

[26] Michael Johanson, Kevin Waugh, Michael Bowling, and Martin Zinkevich. "Accelerating best response calculation in large extensive games". In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume One. Page 258–265. IJCAI'11. AAAI Press (2011).

[27] Tim Roughgarden. "Twenty lectures on algorithmic game theory". Cambridge University Press. (2016).

[28] John Nash. "Non-cooperative games". Annals of Mathematics **54**, 286–295 (1951).

[29] Robert J Aumann. "Subjectivity and correlation in randomized strategies". Journal of mathematical Economics **1**, 67–96 (1974).

[30] Hervé Moulin and J P Vial. "Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon". International Journal of Game Theory **7**, 201–221 (1978).

[31] Elad Hazan et al. "Introduction to online convex optimization". Foundations and Trends® in Optimization **2**, 157–325 (2016).

[32] Satyen Kale. “Efficient algorithms using the multiplicative weights update method”. PhD thesis. Princeton University. (2007). url: www.proquest.com/dissertations-theses/efficient-algorithms-using-multiplicative-weights/docview/304824121/se-2.

[33] Koji Tsuda, Gunnar Rätsch, and Manfred K Warmuth. “Matrix exponentiated gradient updates for on-line learning and Bregman projection”. Journal of Machine Learning Research **6**, 995–1018 (2005). url: http://jmlr.org/papers/v6/tsuda05a.html.

[34] Sanjeev Arora and Satyen Kale. “A combinatorial, primal-dual approach to semidefinite programs”. In Proceedings of the thirty-ninth annual ACM symposium on Theory of computing. Pages 227–236. (2007).

[35] Rahul Jain, Zhengfeng Ji, Sarvagya Upadhyay, and John Watrous. “QIP=PSPACE”. Journal of the ACM (JACM) **58**, 1–27 (2011).

[36] Zeyuan Allen-Zhu, Zhenyu Liao, and Lorenzo Orecchia. “Spectral sparsification and regret minimization beyond matrix multiplicative updates”. In Proceedings of the forty-seventh annual ACM symposium on Theory of computing. Pages 237–245. (2015).

[37] Sergiu Hart and Andreu Mas-Colell. “A simple adaptive procedure leading to correlated equilibrium”. Econometrica **68**, 1127–1150 (2000).

[38] J v. Neumann. “Zur theorie der gesellschaftsspiele”. Mathematische annalen **100**, 295–320 (1928).

[39] Luigi Accardi and Andreas Boukas. “von Neumann’s minimax theorem for continuous quantum games” (2020). arXiv:2006.11502.

[40] Maria-Florina Balcan, Avrim Blum, and Yishay Mansour. “Circumventing the price of anarchy: Leading dynamics to good behavior”. SIAM Journal on Computing **42**, 230–264 (2013).

[41] Brian Hu Zhang, Gabriele Farina, Ioannis Anagnostides, Federico Cacciamani, Stephen Marcus McAleer, Andreas Alexander Haupt, Andrea Celli, Nicola Gatti, Vincent Conitzer, and Tuomas Sandholm. “Steering no-regret learners to a desired equilibrium” (2023). arXiv:2306.05221.

A Examples of entangled equilibria

In this section we showcase examples of entangled QCCEs which are unapproachable by the decentralized no-regret learning paradigm. The idea is to use *maximally-entangled games*, in which payoffs are assigned to states in a maximally-entangled basis instead of the standard product-state basis as one might think to do when attempting to embed a classical game in the quantum game formulation that we use (indeed, that is precisely what is done in [3]). We shall first define maximally-entangled games before characterizing the maximally-entangled QCCEs in any maximally-entangled game.

For simplicity we consider two-player games where each player has access to a qubit ($n_1 = n_2 = 2$). The Bell states

$$\begin{aligned}
|e_1\rangle = |\phi^+\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle),\\
|e_2\rangle = |\phi^-\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle),\\
|e_3\rangle = |\psi^+\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle),\\
|e_4\rangle = |\psi^-\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)
\end{aligned}$$

form a maximally entangled basis for the joint space $\mathcal{H}_1 \otimes \mathcal{H}_2$. We can define a *maximally-entangled game* as follows:

Definition A.1. *A maximally-entangled (max-ent) game is a two-player QG in which the game operators are supported on the rank-1 projectors of a maximally entangled basis $\{|e_k\rangle\}$, i.e., the game operators are given by*

$$R_1 = \sum_k a_k |e_k\rangle\langle e_k|, \qquad R_2 = \sum_k b_k |e_k\rangle\langle e_k|. \tag{19}$$

The following theorem characterizes the QCCEs in a max-ent game that are mixtures of states in the maximally-entangled basis. Crucially, coarse unilateral deviations from mixtures of maximally-entangled states set the other party's reduced state to the maximally mixed state (scaled identity), and the fact that the game operators are supported only on the rank-1 projectors of the maximally-entangled basis makes the utilities achieved by these deviations easy to compute. These two facts make characterizing when such a state is a QCCE easy.

Theorem A.1. *Fix a max-ent game on a maximally-entangled basis $\{|e_k\rangle\}$ and game operators given by (19). A mixture of states in the maximally-entangled basis, $\rho^* = \sum_k \lambda_k |e_k\rangle\langle e_k|$, is a QCCE of the max-ent game if and only if*

$$\sum_k a_k\lambda_k \geq \frac{1}{n_1 n_2}\sum_k a_k \qquad \text{and} \qquad \sum_k b_k\lambda_k \geq \frac{1}{n_1 n_2}\sum_k b_k. \tag{20}$$

Proof. For Player 1, the utility achieved from sticking to ρ^* is given by

$$\mathrm{Tr}(R_1\rho^*) = \sum_k a_k\lambda_k,$$

while the utility achieved from deviating to ρ_1' is given by

$$\begin{aligned}
\mathrm{Tr}(R_1(\rho_1' \otimes \mathrm{Tr}_{\mathcal{A}}\,\rho^*)) &= \frac{1}{n_2}\,\mathrm{Tr}(R(\rho_1' \otimes I_{\mathcal{B}}))\\
&= \frac{1}{n_2}\,\mathrm{Tr}(\mathrm{Tr}_{\mathcal{B}}(R)\rho_1')\\
&= \frac{1}{n_1 n_2}\sum_k a_k\,\mathrm{Tr}(\rho_1')\\
&= \frac{1}{n_1 n_2}\sum_k a_k,
\end{aligned}$$

where the first equality is due to the fact that

$$\operatorname{Tr}_{\mathcal{A}} \rho^* = \sum_k \lambda_k \operatorname{Tr}_{\mathcal{A}}(|e_k\rangle\!\langle e_k|) = \sum_k \lambda_k (\frac{1}{n_2} I_{\mathcal{B}}) = \frac{1}{n_2} I_{\mathcal{B}}$$

and the third equality is due to the fact that

$$\operatorname{Tr}_{\mathcal{B}}(R) = \sum_k a_k \operatorname{Tr}_{\mathcal{B}}(|e_k\rangle \langle e_k|) = \frac{1}{n_1} \sum_k a_k I_{\mathcal{A}}.$$

Thus Player 1 has no incentive to do a coarse deviation if and only if

$$\sum_k a_k \lambda_k \geq \frac{1}{n_1 n_2} \sum_k a_k.$$

We can similarly get the analogous statement for Player 2, and thus ρ^* is a QCCE if and only if both the conditions in (20) hold. □

As a corollary, we are able to construct pure, entangled QCCEs in any common-interest max-ent game.

Corollary A.1. *Fix a maximally-entangled basis $\{|e_k\rangle\}$ for the joint strategy space and suppose that on playing joint strategy ρ both players get common utility $\operatorname{Tr}(R\rho)$ where*

$$R = \sum_k a_k |e_k\rangle\!\langle e_k| .$$

Define $k^* := \arg\max_k a_k$. Then the pure, entangled joint state

$$\rho* = |e_{k^*}\rangle\!\langle e_{k^*}|$$

is a QCCE of this game.

B Additional Quantum Preliminaries

Quantum states. *Pure quantum states* correspond to (typically unit-length normalized) vectors in a Hilbert space $\mathcal{H}$. The simplest case is that of a qubit, which can be represented by a linear superposition of its two orthonormal basis states. These vectors are usually denoted as $|0\rangle = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $|1\rangle = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$ in the conventional bra–ket notation introduced by Paul Dirac and together span the qubit's two-dimensional Hilbert space. A single qubit ψ can be described by a linear combination of $|0\rangle$ and $|1\rangle$: $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ where α and β are the probability amplitudes, i.e., complex numbers such that $|\alpha|^2 + |\beta|^2 = 1$.

Quantum measurements. We utilize the generalized measurement formalism known as the positive-operator-valued measure (POVM) in the main text, but for completeness we also present several other key formalisms for quantum measurements.

Idealized von Neumann measurements. The approach codified by John von Neumann represents a measurement upon a physical system by a self-adjoint operator on that Hilbert space termed an "observable". We start by representing each observable by a Hermitian operator A. This operator will have a complete set of (normalized) eigenvectors $|\lambda_n\rangle$ and associated eigenvalues λ_n[1], thus we can write A in the form $A = \sum_n \lambda_n |\lambda_n\rangle\langle\lambda_n|$. Let's assume, for the moment and for simplicity, that all the eigenvalues are distinct. The von Neumann description then states that if we perform a measurement of A then we will find the result of the measurements to be one of the eigenvalues and the probability for finding any one of these is $P(\lambda_n) = |\langle\lambda_n|\psi\rangle|^2$. Whereas in the previous paragraph we chose the observable with $A = 0|0\rangle\langle 0| + 1|1\rangle\langle 1|$ now we can choose another

[1] That is we have $A|\lambda_n\rangle = \lambda_n|\lambda_n\rangle$.

observable $B = \lambda_1|\lambda_1\rangle\langle\lambda_1| + \lambda_2|\lambda_2\rangle\langle\lambda_2|$. With a bit of algebra, we can verify that for $n = 1, 2$: $P(\lambda_n) = |\langle\lambda_n|(\alpha|0\rangle + \beta|1\rangle)\rangle|^2 = |\alpha|^2\langle|\lambda_n|0\rangle|^2 + |\beta|^2\langle|\lambda_n|1\rangle|^2 + 2Re\{\alpha\beta^*\langle\lambda_n|0\rangle\langle\lambda_n|1\rangle^*\}$. In this case the last term, which is known as the interference term, no longer vanishes as in the case of observable A. Specifically, the expected utility/measurement will not be in agreement with that of a classical probability distribution which is in state $|0\rangle$ with probability $|\alpha|^2$ and in state $|1\rangle$ with probability $|\beta|^2$.

From quantum states to density operators/matrices. The probability of measuring eigenvalue λ_n is

$$P(\lambda_n) = |\langle\lambda_n|\psi\rangle|^2 = \langle\lambda_n|\psi\rangle\langle\psi|\lambda_n\rangle = \langle\lambda_n|\rho|\lambda_n\rangle = \mathrm{Tr}(\rho|\lambda_n\rangle\langle\lambda_n|) = \mathrm{Tr}(\rho P_n)$$

where $\rho = |\psi\rangle\langle\psi|$ is the rank-1 projection operator onto the space spanned by the state $|\psi\rangle$ and is called (probability) density of $|\psi\rangle$ and P_n is the projector on the space spanned by the eigenvector[2] $|\lambda_n\rangle$. Thus, the overall expected measurement is equal to[3]

$$\langle A\rangle_\psi = \sum_n \lambda_n P(\lambda_n) = \sum_n \lambda_n \mathrm{Tr}(\rho P_n) = \mathrm{Tr}\left(\rho \sum_n \lambda_n P_n\right) = \mathrm{Tr}(\rho A).$$

The set of projectors P_n above have the following three properties: they are Hermitian, they are positive semi-definite operators and they are complete; they sum up to the identity. These properties have physical meaning. They represent, respectively, the requirements that the projectors are observables, that they give non-negative probabilities and that the sum of the probabilities for all possible outcomes must be equal to one. Generalized measurements will correspond to a collection of such projectors without necessarily being orthonormal.

C Omitted Proofs

Lemma C.1. *The two definitions of QCCE, namely*

$$u_i(\rho) \geq u_i((\phi_i \otimes \mathbb{I}_{-i})(\rho)) \tag{dev-QCCE}$$

for all replacement channels $\phi_i : \mathrm{L}(\mathcal{H}_i) \to \mathrm{L}(\mathcal{H}_i),\ X \mapsto \mathrm{Tr}(X)\rho_i'$ for some $\rho_i' \in \mathrm{L}(\mathcal{H}_i)$ and

$$u_i(\rho) \geq u_i(\rho_i' \otimes \mathrm{Tr}_i\,\rho) \tag{QCCE}$$

where $\mathrm{Tr}_i : \mathrm{L}(\bigotimes_{i'} \mathcal{H}_{i'}) \to \mathrm{L}(\bigotimes_{i' \neq i} \mathcal{H}_{i'})$ is the partial trace with respect to player i's subsystem, are equivalent.

Proof. For any joint state $\rho \in \mathrm{D}(\otimes_i \mathcal{H}_i)$ we can write

$$\rho = \sum_k X_k \otimes Y_k$$

for some $X_k \in \mathrm{L}(\mathcal{H}_i), Y_k \in \mathrm{L}(\mathcal{H}_{-i})$ since $\rho \in \mathrm{L}(\otimes_i \mathcal{H}_i) = \otimes_i \mathrm{L}(\mathcal{H}_i)$. Then when ϕ_i is the replacement channel $\phi_i : \mathrm{L}(\mathcal{H}_i) \to \mathrm{L}(\mathcal{H}_i),\ X \mapsto \mathrm{Tr}(X)\rho_i'$,

$$\begin{aligned}(\phi_i \otimes \mathbb{I}_{-i})(\rho) = \sum_k (\phi_i \otimes \mathbb{I}_{-i})(X_k \otimes Y_k) = \sum_k \phi_i(X_k) \otimes Y_k &= \left(\sum_k \mathrm{Tr}(X_k)\rho_i'\right) \otimes Y_k \\ &= \rho_i' \otimes \sum_k \mathrm{Tr}(X_k) Y_k \\ &= \rho_i' \otimes \mathrm{Tr}_i\,\rho.\end{aligned}$$

[2] If the eigenvalues of A are degenerate, there exists a set of orthonormal eigenvectors, $|\lambda_n^j\rangle$, which correspond to the same λ_n, then $P(\lambda_n) = Tr(\rho|\sum_j |\lambda_n^j\rangle\langle\lambda_n^j|) = Tr(\rho P_n)$, i.e., we use a projector onto the set of states with eigenvalue $|\lambda_n\rangle$.

[3] Another set of useful formulas that easily follow in the case of pure states are $P(\lambda_n) = \langle\lambda_n|P_n|\lambda_n\rangle$ and $\langle A\rangle_\psi = \langle\lambda_n|A|\lambda_n\rangle$.

□

Lemma C.2. *Let $h(x, y) : X \times Y \to \mathbb{R}$ be a continuous function on the product of compact sets X, Y. Then $z : X \to \mathbb{R}$, $z(x) = \sup_{y \in Y} h(x, y)$ is continuous.*

Proof. Since h is a continuous function on a compact domain, it is uniformly continuous. In particular, given any $\epsilon > 0$ $\exists \delta > 0$ such that $\forall y$, $\forall \|x - x'\| < \delta$, $|h(x, y) - h(x', y)| < \epsilon$.

Then $\forall y$, $\forall \|x - x'\| < \delta$ we have that $h(x, y) \leq h(x', y) + \epsilon \leq z(x') + \epsilon$, which taking supremum over $y \in Y$ on both sides gives us that $z(x) \leq z(x') + \epsilon$ $\forall \|x - x'\| < \delta$.

On the other hand, we have similarly that $\forall y$, $\forall \|x - x'\| < \delta$ $h(x'y) \leq h(x, y) + \epsilon \leq z(x) + \epsilon$, so similarly taking supremum over y on both sides gives us that $z(x') \leq z(x) + \epsilon$ $\forall \|x - x'\| < \delta$.

Combining the last two results gives us that $|z(x') - z(x)| < \epsilon$ $\forall \|x - x'\| < \delta$. Thus z is continuous. □