Matrix Multiplicative Weights Updates in Quantum Zero-Sum Games: Conservation Laws & Recurrence

Rahul Jain[*]
rahul@comp.nus.edu.sg

Georgios Piliouras[†]
georgios@sutd.edu.sg

Ryann Sim[†]
ryann_sim@mymail.sutd.edu.sg

Abstract

Recent advances in quantum computing and in particular, the introduction of quantum GANs, have led to increased interest in quantum zero-sum game theory, extending the scope of learning algorithms for classical games into the quantum realm. In this paper, we focus on learning in quantum zero-sum games under *Matrix Multiplicative Weights Update* (a generalization of the multiplicative weights update method) and its continuous analogue, *Quantum Replicator Dynamics*. When each player selects their state according to quantum replicator dynamics, we show that the system exhibits conservation laws in a quantum-information theoretic sense. Moreover, we show that the system exhibits Poincaré recurrence, meaning that almost all orbits return arbitrarily close to their initial conditions infinitely often. Our analysis generalizes previous results in the case of classical games [43, 49].

1 Introduction

The Nash equilibrium has always been a central concept in non-cooperative game theory. While the existence of Nash equilibria is well known in finite games where mixed strategies are allowed [45], it is less clear how to efficiently compute these equilibria themselves. Indeed, much work revolves around finding methods to solve for Nash equilibria in different contexts, as well as analyzing the complexity of solving for the Nash [15, 34].

Despite the dominance of equilibrium computation in classical theory, recent works have begun to move towards attempting to understand the nature of learning in games [23, 52]. The family of regret-minimizing algorithms known as Follow-The-Regularized-Leader (FTRL) have been studied extensively, both for discrete time [7, 13, 46, 50] and continuous time [21, 43, 48, 49] dynamics. These algorithms have found many applications in the domain of machine learning - one such example is that of Generative Adversarial Networks (GANs) [25], which have been successfully modeled using zero-sum game theory. This rich connection has led to a stronger understanding of adversarial learning and improved practical performance of GANs in various settings [21, 33, 37].

On the other side of the coin, quantum computing is a field which has garnered much interest in the machine learning community. Many quantum machine learning algorithms have been proposed, extending standard classical algorithms. One such example is quantum GANs [12, 14, 40]. While such works describe the architecture of quantum GANs, we still do not fully understand the behavior of learning algorithms in the quantum setting. In this paper, we provide an initial analysis of learning in zero-sum quantum games. We do so by studying the learning behaviour of the ubiquitous Multiplicative Weights Update (MWU) update rule [5, 22, 38], a specific instance of the FTRL framework, when applied to such games.

There have been several important results studying MWU and its continuous counterpart, *replicator dynamics*. [7] showed that classical MWU does not converge in day-to-day behaviour to the mixed Nash equilibrium when applied

---

[*]Centre for Quantum Technologies and Department of Computer Science, National University of Singapore and MajuLab, UMI3654, Singapore.

[†]Engineering Systems and Design, Singapore University of Technology and Design, Singapore

to two player zero-sum games. [43] showed that the orbits of players in zero-sum games exhibit Poincaré recurrence when using replicator dynamics. However, quantum systems can oftentimes behave in radically different ways to their classical counterparts. Indeed, allowing for quantum strategies can create situations wherein players can gain greater payoffs than if they were playing using only classical strategies (see e.g. the Mermin–Peres magic square game [41, 47]). As such, our extension of the classical results to the quantum realm has to be treated with care.

We focus on a matrix generalization of MWU known as *Matrix Multiplicative Weights Update* (MMWU), which changes the paradigm of standard MWU from cost vectors to cost matrices, and from probability vectors to density matrices. This generalization has been independently discovered and studied by [55] as Matrix Exponentiated Gradient Updates and [5] as the Matrix Multiplicative Weights algorithm. Applications of the MMWU algorithm include solving semi-definite programs (SDPs) [6] and obtaining bounds on the sample complexity for learning problems in quantum computing [31]. From a game theoretic standpoint, [29] analyzed the MMWU algorithm for quantum zero-sum games, proving time-average convergence to an approximate Nash equilibrium. We extend this result by studying the learning behaviour of MMWU in quantum zero-sum games, which could lay the groundwork for developing novel algorithms that achieve convergence to the Nash in day-to-day behaviour in quantum zero-sum games, similar to the case of their classical analogues [16, 42].

Contributions. In this paper, we first study the dynamics of MMWU in quantum zero-sum games, utilizing tools from information theory and classical game theory (Section 3). We provide bounds on the rate of change of the total quantum relative entropy between a fully mixed Nash equilibrium and the evolving state of the system. We then study the continuous counterpart of MMWU, which we call quantum replicator dynamics (Section 4). We show that the aforementioned total quantum relative entropy is a constant of motion. Furthermore, the dynamics do not converge to equilibrium (in the day-to-day sense) but rather exhibit a weaker form of approximate periodicity known as *Poincaré recurrence*. Our proof of this result departs from the standard classical method, representing our main technical novelty. Finally, we present several simulations which corroborate our theoretical results (Section 5).

2 Preliminaries and Definitions

2.1 Quantum Theory

Basic concepts. Similarly to [29] and in order to make the presentation self-contained, we will start by introducing some basic concepts of quantum information and game theory.

First, we refer to a quantum *register* as a collection of qubits representing a message that is transferred from one party to another. We associate a vector space $\mathcal{H} = \mathbb{C}^n$ with any quantum register. This intuitively represents the maximum number of distinct classical states that can be stored in the register without error. The *state* of a quantum register is represented by a *density matrix*, which is an $n \times n$ positive semi-definite matrix with trace 1. We will use $D(\mathcal{H})$ to denote the set of all density matrices associated with a register that is described by $\mathcal{H}$. One can naturally view such density matrices as linear operators acting on $\mathcal{H}$.

When two registers with associated spaces $\mathcal{A} = \mathbb{C}^n$ and $\mathcal{B} = \mathbb{C}^m$ are considered as a joint register, the associated space is the tensor product $\mathcal{A} \otimes \mathcal{B} = \mathbb{C}^{nm}$. If the two registers are independently prepared in states described by ρ and σ, then the joint state is described by the $nm \times nm$ density matrix $\rho \otimes \sigma$.

Next, for a given vector space $\mathcal{H} = \mathbb{C}^n$, we define $\mathrm{L}(\mathcal{H})$ as the set of all $n \times n$ complex matrices. Furthermore, we denote the subset of $\mathrm{L}(\mathcal{H})$ given by *Hermitian* matrices as $\mathrm{Herm}(\mathcal{H})$. A Hermitian matrix A satisfies the equality $A = A^\dagger$, where $A^\dagger$ denotes the *adjoint* (or conjugate transpose) of matrix A. Subsequently, we define $\mathrm{Pos}(\mathcal{H})$ as the subset of $\mathrm{Herm}(\mathcal{H})$ which consists of all positive semi-definite $n \times n$ matrices.

Finally, the *Hilbert-Schmidt inner product* on $\mathrm{L}(\mathcal{H})$ is defined as $\langle A, B \rangle = \mathrm{Tr}(A^\dagger B)$ for all $A, B \in \mathrm{L}(\mathcal{H})$. Note that $\langle A, B \rangle$ is a real number for any Hermitian matrices A and B, and is also additionally non-negative if A and B are positive semi-definite.

Measurements and observables. In the context of game theory, we are also interested in the concepts of quantum *measurements* and *observables*. An observable is simply a property of the quantum system which is measurable. The measurement of a register having associated vector space $\mathcal{H} = \mathbb{C}^n$ is a collection of linear operators $\{P_i : 1 \leq i \leq k\} \subset \mathrm{Pos}(\mathcal{H})$ which satisfies $\sum_{i=1}^{k} P_i = \mathbb{1}_{\mathcal{H}}$, where $\mathbb{1}_{\mathcal{H}}$ is the identity matrix on $\mathcal{H}$. If the register corresponding to $\mathcal{H}$ is in a state defined by density matrix ρ and the measurement described by P_i is performed, each outcome i will be observed with probability $\langle P_i, \rho \rangle$. An important note is that two mixed states with the same density matrix are indistinguishable from each other by any measurement.

Additional notation and definitions. A linear mapping of the form $\Phi : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$ is called a super-operator. The adjoint super-operator to Φ is $\Phi^\dagger : \mathrm{L}(\mathcal{A}) \to \mathrm{L}(\mathcal{B})$ and is uniquely determined by the condition:

$$\langle A, \Phi(B) \rangle = \langle \Phi^\dagger(A), B \rangle$$

A super-operator $\Phi : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$ is said to be positive if $\Phi(P)$ is positive semi-definite for every choice of positive semi-definite operator $P \in \mathrm{Pos}(\mathcal{B})$. In addition, $\Phi^\dagger$ is positive if and only if Φ is positive. There is a one-to-one and linear correspondence between the collection of operators of the form $R \in \mathrm{L}(\mathcal{A} \otimes \mathcal{B})$ and the collection of super-operators $\Phi : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$ defined above. Specifically, for each super-operator Φ we can define an operator R as follows:

$$R = \sum_{1 \leq i,j \leq m} \Phi(E_{i,j}) \otimes E_{i,j} \tag{1}$$

where $E_{i,j}$ is the matrix with a 1 in entry (i, j) and 0 elsewhere. Conversely, given an operator $R \in \mathrm{L}(\mathcal{A} \otimes \mathcal{B})$, we can define:

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

This correspondence is linear and one can translate back and forth between the two as needed for an given application. R is assumed to be positive semi-definite, so the corresponding super-operator Φ is also positive.

2.2 Quantum Game Theory

A quantum system which can be manipulated by any number of agents, and where the utility of the moves is well defined, quantified and ordered can be conceived as a quantum game. In particular, a 2-player zero-sum quantum game sees players Alice and Bob each sending a quantum state to a referee, who then performs a measurement on these two states to determine their payoffs. Let $\mathcal{A} = \mathbb{C}^n$ and $\mathcal{B} = \mathbb{C}^m$ be the vector spaces that correspond to the state that Alice and Bob send to the referee.

In order to determine the payoffs of the players' actions, the referee performs a joint measurement where Alice's and Bob's states are viewed as a single register. Thus, the referee's measurement can be described by a collection

$$\{R_a : a \in \mathcal{S}\} \subset \mathrm{Pos}(\mathcal{A} \otimes \mathcal{B}) \tag{3}$$

which satisfies the condition $\sum_{a=1}^{k} R_a = \mathbb{1}_{\mathcal{A} \otimes \mathcal{B}}$. We associate each possible measurement outcome a with a payoff for each player. Since we are only considering zero-sum games, if the payoff that Alice receives from the referee is $v(a)$, then Bob's corresponding payoff will be $-v(a)$. Henceforth, we refer to the states sent by Alice and Bob to the referee as ρ and σ respectively. For a given choice of ρ and σ, Alice's expected payoff is:

$$u(\rho, \sigma) = \sum_{a=1}^{k} v(a) \langle R_a, \rho \otimes \sigma \rangle = \langle R, \rho \otimes \sigma \rangle \tag{4}$$

where $R = \sum_{a=1}^{k} v(a) R_a$. Likewise, Bob's corresponding expected payoff is $-u(\rho, \sigma) = -\langle R, \rho \otimes \sigma \rangle$. R is referred to throughout the rest of the paper as a *payoff observable*. A necessary and sufficient condition for matrix R acting on $\mathcal{A} \otimes \mathcal{B}$ to be obtained from some real valued payoff function v is that R is Hermitian. From Equation 2, we can equivalently define the expected payoff of Alice as $u(\rho, \sigma) = \langle \rho, \Phi(\sigma) \rangle$. We will use the latter formulation for the remainder of the paper.

A key notion of equilibrium in classical game theory is the Nash equilibrium. In the quantum setting, we define the pair of quantum states (ρ^*, σ^*) as a Nash equilibrium of R if

$$u(\rho^*, \sigma^*) \geq u(\rho, \sigma^*) \quad \text{and} \quad u(\rho^*, \sigma^*) \geq u(\rho^*, \sigma) \tag{5}$$

for all ρ, σ. That is, neither Alice nor Bob would prefer to unilaterally deviate from playing ρ^* and σ^* respectively. On another note, since the set of available strategies (density matrices) for both agents is convex and compact and the utility function of Alice is bilinear, standard extensions of the of von Neumann's Min-Max Theorem apply [20].

Finally, we define the notion of a 'fully mixed' Nash equilibrium. An equilibrium (ρ^*, σ^*) is fully mixed if ρ^*, σ^* are full rank.

2.3 Information Theory

We also introduce several information theoretic concepts which will be referenced throughout the paper.

Shannon entropy. The Shannon entropy of a random variable X where each strategy x is obtained with probability $p(x)$ is given by $H(X) = -\sum_x p(x) \log p(x)$. This intuitively is a measure of randomness or uncertainty in the system. A natural generalization of the Shannon entropy to a quantum context is the von Neumann entropy. For a quantum mechanical system defined by density matrix ρ, the von Neumann entropy is given by $S(\rho) = -\text{Tr}(\rho \log \rho)$.

Bregman divergence. We are also interested in the notion of Bregman divergence, which measures the distance between two points. Let $x^* \in \mathcal{X}$ be a Nash equilibrium and let $x \in \mathcal{X}$ be an arbitrary strategy profile. Also, let F be a continuously-differentiable, strictly convex function. The Bregman divergence from x^* to x is given by $D_F(x^* \| x) = F(x^*) - F(x) - \langle \nabla F(x), x^* - x \rangle$.

In the context of quantum games, the Bregman divergence is defined using matrix notation. We define the quantum relative entropy between two quantum states using the von Neumann entropy $S(\rho)$. In particular, the quantum relative entropy between two quantum states ρ and σ is defined as $S(\rho \| \sigma) = \text{Tr}\left(\rho(\log \rho - \log \sigma)\right)$.

2.4 Dynamical Systems

Flows. Consider a differential equation $\dot{p} = f(p)$ on a topological space $\mathcal{P}$. The existence and uniqueness theorem for ordinary differential equations guarantees that we can write the unique solution to the differential equation as a continuous map $\phi : \mathcal{P} \times \mathcal{H} \to \mathcal{P}$. This is referred to as the *flow of the differential equation* such that for any point $p \in \mathcal{P}$, $\phi(p, -)$ defines a function of time corresponding to the trajectory of p. Conversely, fixing a time t provides a map $\phi^t \equiv \phi(-, t) : \mathcal{P} \to \mathcal{P}$. In Section 4, we introduce the notion of quantum replicator dynamics, which are Lipschitz continuous differential equations. Hence, a unique flow ϕ of these replicator dynamics exists.

Liouville's theorem. Liouville's theorem can be applied to any system of ordinary differential equations with a continuously differentiable vector field ξ on an open domain $\mathcal{Y} \in \mathbb{R}^d$. The divergence of ξ at $y \in \mathcal{Y}$ is the trace of the Jacobian at y: div $\xi(y) = \sum_{i=1}^d \frac{\partial \xi_i}{\partial y_i}(y)$. Because the divergence is continuous, it is integrable on Lebesgue measurable subsets of $\mathcal{Y}$. Given any such set C, define the image of C under flow ϕ at time t as $C(t) = \{\phi(c, t) : c \in C\}$. $C(t)$ is measurable and of volume $\text{vol}[C(t)] = \int_{C(t)} d\mu$. Liouville's formula states that the time derivative of the volume $\text{vol}[C(t)]$ exists and links it to the divergence of ξ:

$$\frac{d}{dt} \left[\text{vol } C(t)\right] = \int_{C(t)} \text{div}(\xi) d\mu \tag{6}$$

If div $\xi(y)$ is null at any $y \in \mathcal{Y}$, then the volume is conserved. Since div ξ is continuous, the converse statement is also true - if the volume is conserved on any open set, div $\xi(y)$ has to be null at any point $y \in \mathcal{Y}$.

Diffeomorphisms of flows. A function $\mathbf{f}$ between two topological spaces is called a diffeomorphism if i) $\mathbf{f}$ is a bijection, ii) $\mathbf{f}$ is continuously differentiable, iii) $\mathbf{f}$ has a continuously differentiable inverse. Two flows $\Phi^t : A \to A$ and $\Psi^t : B \to B$ are diffeomorphic if there exists a diffeomorphism $\mathbf{f} : A \to B$ such that for each $x \in A$ and $t \in \mathbb{R}$, $\mathbf{f}(\Phi^t(p)) = \Psi^t(\mathbf{f}(p))$. For the purpose of our analysis, the replicator dynamics defined in Equations 17 are translated via a diffeomorphism from the interior of $\mathcal{P}$ to a space $\mathcal{C} = \Pi_{i \in V} \mathbb{R}^{n-1}$, which allows us to show certain desirable properties.

Poincaré recurrence. The concept of Poincaré recurrence arises from Henri Poincaré's celebrated 1890 work regarding the three body problem [51]. He proved that if a dynamical system preserves volume and always remains bounded in its orbits, almost all trajectories return arbitrarily close to their initial position, and do so infinitely often.

Theorem 1 (Poincaré recurrence). *If a flow preserves volume and has only bounded orbits then for each open set there exist orbits that intersect the set infinitely often.*

3 MMWU in Quantum Zero-Sum Games

In [29], the MMWU algorithm for zero-sum games is shown to exhibit time-average convergence to an approximate Nash equilibrium in two-player quantum zero-sum games. The MMWU algorithm is shown in Algorithm 1.

Note that here we focus specifically on two-player games, and utilize the expected payoffs defined via super-operators $\Phi : \text{L}(\mathcal{A}) \to \text{L}(\mathcal{B})$ as seen in Equation 2. Moreover, μ is the step-size in the quantum algorithm.

In this section, we examine closely the update steps for each player in the MMWU algorithm (Algorithm 1) and analyze the limiting behaviour of the total quantum relative entropy in the system. First, we introduce two useful facts which will aid in the analysis.

Algorithm 1: Parallel Approximation of Equilibrium Point

Let $\mu = \epsilon/8$ and let $N = \lceil 64 \log{(nm)}/\epsilon^2 \rceil$.
Initialize: $A_0 = \mathbb{1}_{\mathcal{A}}$, $\rho_0 = A_0/\text{Tr}(A_0)$, $B_0 = \mathbb{1}_{\mathcal{B}}$, and $\sigma_0 = B_0/\text{Tr}(B_0)$.
for $j = 1 \dots N$ **do**
$\quad A_j = \exp\left(\mu \sum_{i=0}^{j-1} \Phi(\sigma_i)\right)$
$\quad \rho_j = A_j/\text{Tr}(A_j)$
$\quad B_j = \exp\left(-\mu \sum_{i=0}^{j-1} \Phi^*(\rho_i)\right)$
$\quad \sigma_j = B_j/\text{Tr}(B_j)$
end for

Fact 1 (Golden-Thompson inequality [24, 54]). *Let A, B be Hermitian matrices. Then*

$$\text{Tr} \exp(A + B) \leq \text{Tr} \exp(A) \exp(B) \tag{7}$$

Fact 2. *Let $0 \leq A \leq \mathbb{1}$ be a PSD matrix and δ be a real number. Then,*

$$\exp(\delta A) \leq \mathbb{1} + \delta \exp(\delta) A \tag{8}$$

Next, we put forward two corollaries which will help us prove Theorem 2. We first define $\Delta S(\rho^*\|\rho_j) = S(\rho^*\|\rho_j) - S(\rho^*\|\rho_{j-1})$ and $\Delta S(\sigma^*\|\sigma_j) = S(\sigma^*\|\sigma_j) - S(\sigma^*\|\sigma_{j-1})$.

Corollary 1. *The change in the sum of quantum relative entropies in a quantum zero-sum game between a fully-mixed Nash equilibrium and the players' strategies is given by:*

$$\Delta S(\rho^*\|\rho_j) + \Delta S(\sigma^*\|\sigma_j) = \log \frac{\text{Tr}A_j}{\text{Tr}A_{j-1}} + \log \frac{\text{Tr}B_j}{\text{Tr}B_{j-1}} \tag{9}$$

Corollary 2. *The following trace inequalities hold for PSD matrices A and B updated with MMWU:*

$$\Delta S(\rho^*\|\rho_j) + \Delta S(\sigma^*\|\sigma_j) \geq \mu \exp(-\mu)\text{Tr}\left(\rho_j \Phi(\sigma_{j-1})\right) - \mu \exp(\mu)\text{Tr}\left(\rho_{j-1}\Phi(\sigma_j)\right) \tag{10}$$

$$\Delta S(\rho^*\|\rho_j) + \Delta S(\sigma^*\|\sigma_j) \leq \mu \exp(\mu)\text{Tr}\left(\rho_{j-1}\Phi(\sigma_{j-1})\right) - \mu \exp(-\mu)\text{Tr}\left(\rho_{j-1}\Phi(\sigma_{j-1})\right) \tag{11}$$

The proof of Corollary 2 relies on Facts 1 and 2, as well as the equality shown in Corollary 1.

In many practical scenarios, one would use decreasing step-sizes when running MMWU. As such, we utilize Corollary 2 and take the limit as step-size μ goes to 0 in order to show the following theorem:

Theorem 2. *The change in the sum of quantum relative entropies between a fully mixed Nash equilibrium and the player's strategies and in a two-player zero-sum quantum game tends to zero when step-size $\mu \to 0$. Specifically,*

$$\lim_{\mu \to 0} \frac{1}{\mu}\left(\Delta S(\rho^*\|\rho_j) + \Delta S(\sigma^*\|\sigma_j)\right) = 0 \tag{12}$$

The proof of Theorem 2, along with Corollaries 1 and 2 can be found in Appendix C.

Theorem 2 further motivates an investigation into the continuous time variant of MMWU, which we call quantum/matrix replicator dynamics. In particular, we show that in the continuous case, the sum of quantum relative entropies is invariant. We introduce and study quantum replicator dynamics in detail in Section 4.

4 Replicator Dynamics in Quantum Zero-Sum Games

We have seen that in the case of discrete dynamics (MMWU), as the step-size becomes infinitesimal, the total quantum relative entropy between the Nash equilibrum and the system state tends to stabilize. A natural question would then be: does the same result hold in continuous time? In order to explore this question in greater detail, we first need to define the continuous analogue of MMWU, the *quantum replicator dynamics*, which has well known classical analogues [52].

We start by rewriting the MMWU update steps, but now defined over a continuous time interval $[0, t]$:

$$A(t) = \int_0^t \Phi(\sigma(\tau))d\tau \tag{13}$$

$$\rho(t) = \exp(A(t))/\mathrm{Tr}(\exp(A(t))) \tag{14}$$

$$B(t) = -\int_0^t \Phi^\dagger(\rho(\tau))d\tau \tag{15}$$

$$\sigma(t) = \exp(B(t))/\mathrm{Tr}(\exp(B(t))) \tag{16}$$

Note here that we shift the exponential terms from the definition of $A(t)$ and $B(t)$ to the corresponding $\rho(t)$ and $\sigma(t)$ terms. This will help simplify some of the proof techniques later on in the paper. Furthermore, in the rest of the paper we will typically drop from the notation the explicit dependence on t to ease with the notational burden.

It is important to note the following observation, which will be helpful in our later analysis.

Observation 1. *The discrete-time trajectories ρ_j and σ_j defined in Algorithm 1 are a standard Euler discretization (with step μ) of the continuous-time trajectories $\rho(t)$ and $\sigma(t)$ defined in Equations 14 and 16.*

We now define the *quantum replicator dynamics* as:

$$d\rho/dt = \frac{d}{dt}\left(\frac{\exp(A)}{\mathrm{Tr}(\exp(A))}\right), \quad d\sigma/dt = \frac{d}{dt}\left(\frac{\exp(B)}{\mathrm{Tr}(\exp(B))}\right) \tag{17}$$

It is worth noting that in the classical/commuting setting, one can write the replicator equations in a form that describes the relative utility that one agent obtains as compared to the average utility overall. However, in the quantum case this is not possible in general, since it relies on the assumption that $\int_0^t \Phi(\sigma(\tau))d\tau$ and $\Phi(\sigma(t))$, and respectively $\int_0^t \Phi^\dagger(\rho(\tau))d\tau$ and $\Phi^\dagger(\rho(t))$ commute.

As a consequence of Observation 1, we can also conclude that the dynamical system described by the replicator dynamics defined in Equations 17 is a limit case of the dynamical system described by the MMWU algorithm as $\mu \to 0$.

We are now able to state the main theorem for quantum relative entropy in quantum replicator dynamics.

Theorem 3. *When applying matrix/quantum replicator dynamics in a quantum zero-sum game with a fully-mixed Nash equilibrium (ρ^*, σ^*), the sum of quantum relative entropies between the fully-mixed Nash equilibrium and the state of the system $(\rho(t), \sigma(t))$ is invariant on every system trajectory, i.e.:*

$$\frac{d\big(S(\rho^*\|\rho(t)) + S(\sigma^*\|\sigma(t))\big)}{dt} = 0 \tag{18}$$

4.1 Poincaré Recurrence in Quantum Zero-Sum Games

Now that we have described analytical results surrounding the day-to-day behaviour of quantum replicator dynamics, we seek to understand the *dynamics* of the trajectories. After all, invariance of quantum relative entropy does not fully describe how the system moves over time. We show that for any two-player zero-sum quantum game updated with replicator dynamics, the system exhibits *Poincaré recurrence*, insofar as the game is zero-sum and has a fully-mixed Nash equilibria. As introduced in Section 2, the notion of Poincaré recurrence is a weaker version of periodicity. To be precise, for almost all initial conditions $\rho_0 \in \mathcal{P}$, the replicator dynamics return arbitrarily close to ρ_0 infinitely often.

Theorem 4. *The quantum replicator dynamics given in Equations 17 are Poincaré recurrent in any two player zero-sum game which has a fully-mixed Nash equilibrium.*

The proof of this main theorem involves carefully piecing together several auxiliary results, which we will describe in the rest of the section. Furthermore, we stress that due to the non-commutative nature of quantum systems, the standard (classical) approach of differentiating the discrete-time dynamics in the primal space of probability distributions does not apply directly unless we have the highly unlikely situation where $\int_0^t \Phi(\sigma(\tau))d\tau$ and $\Phi(\sigma(t))$ (resp. $\int_0^t \Phi^\dagger(\rho(\tau))d\tau$ and $\Phi^\dagger(\rho(t))$) commute. This problem with carrying over the standard approach is explicitly discussed in Appendix D.

For the proof in the quantum setting, we first define a *canonical transformation* on the space of the matrices $A(t)$ and $B(t)$, which will be crucial in proving the theorem.

Definition 1 (Canonical transformation). *We define the canonical transformation of $A'(t)$ and $B'(t)$ to be a mapping of $A(t)$ and $B(t)$ as defined by Equations 13 and 15. In particular, we define*

$$\begin{aligned} A'(t) &= A(t) - (v^\dagger A(t)v)\mathbb{1} \\ B'(t) &= B(t) - (v^\dagger B(t)v)\mathbb{1} \end{aligned} \tag{19}$$

where v is a fixed vector defined as $v = [1, 0 \ldots 0]^\top$, such that the values of $v^\dagger A(t)v$ and $v^\dagger B(t)v$ are real numbers corresponding to the $(1,1)$-th element of matrices $A(t)$ and $B(t)$ for all t. Notice that this creates matrices $A'(t)$ and $B'(t)$ which have 0 as the $(1,1)$-th entry.

Under the transformation in Definition 1, the vector fields $\dot{A}'(t) = F(A')$ and $\dot{B}'(t) = F(B')$ are given by:

$$\dot{A}'(t) = \Phi(\sigma(t)) - (v^\dagger \Phi(\sigma(t))v)\mathbb{1}$$
$$\dot{B}'(t) = -\Phi^\dagger(\rho(t)) + (v^\dagger \Phi^\dagger(\rho(t))v)\mathbb{1}$$

(20)

where $\frac{d}{dt}\left(v^\dagger A(t)v\right)$ is given by $v^\dagger \frac{dA(t)}{dt}v$.

Moreover, the values of $\rho'(t)$ and $\sigma'(t)$ are defined as:

$$\rho'(t) = \exp(A'(t))/\text{Tr}(\exp(A'(t)))$$
$$\sigma'(t) = \exp(B'(t))/\text{Tr}(\exp(B'(t)))$$

(21)

Proposition 1. *The dynamics of $\rho(t)$ and $\sigma(t)$ remain the same after undergoing the canonical transformation. Equivalently, $A'(t)$ and $A(t)$ (resp. $B'(t)$ and $B(t)$) admit the same strategy $\rho(t)$ (resp. $\sigma(t)$).*

Proposition 2. *The mappings $A'(t)$ and $\rho(t)$ (resp. $B'(t)$ and $\sigma(t)$) are diffeomorphic to one another.*

Proposition 2 will be of crucial importance to our proof technique, since we first prove recurrence for the system described by $A'(t)$ and $B'(t)$, then recover recurrence in $\rho(t)$ and $\sigma(t)$.

To show Poincaré recurrence of Equations 17, we first prove two key properties: (i) the flow of $\dot{A}'$ is volume preserving, meaning that the trace of the Jacobian of the respective vector fields $\dot{A}'(t) = F(A')$ and $\dot{B}'(t) = F(B')$ are zero, and (ii) A' and B' have bounded orbits from any interior initial condition. Then, Poincaré recurrence of A' and B' follows from Poincaré's recurrence theorem.

Volume Conservation. We introduce a lemma which shows that in two-player zero-sum quantum replicator dynamics, the canonical transformation produces a dynamical system which preserves volume.

Lemma 1. *For two-player zero-sum quantum replicator dynamics, the vector fields that arise as a result of the canonical transformation in Definition 1 are volume preserving.*

The proof of Lemma 1 follows from considering the flows of $A'(t)$ and $B'(t)$ and showing that the divergences of the vector fields defined in Equations 20 are equal to zero. A straightforward application of Liouville's theorem completes the proof.

Bounded Orbits. We now show that the transformed dynamical system always has bounded orbits when initialized on the interior of the space of probability density matrices.

Lemma 2. *For any finite initial points $A(0)$ and $B(0)$, the dynamics mapped to $A'(t)$ and $B'(t)$ via the transformation in Definition 1 have bounded orbits.*

The proof of Lemma 2 relies on Theorem 3, and leverages the hermicity of the matrices involved. Moreover, we use the fact that the canonically transformed matrices $A'(t)$ and $B'(t)$ have zero as the $(1,1)$-th element to bound the eigenvalues of $A'(t)$ and $B'(t)$ away from infinity.

Now we are ready to prove Theorem 4 using Lemmas 1 and 2.

Proof of Theorem 4. By Lemmas 1 and 2, as well as the Poincaré recurrence theorem introduced in Section 2, we immediately see that the system of replicator equations given by dA'/dt and dB'/dt are Poincaré recurrent since they are volume preserving and have bounded orbits. Since the flows of $A'(t)$ and $\rho(t)$ are diffeomorphic to one another (likewise for $B'(t)$ and $\sigma(t)$), $d\rho/dt$ and $d\sigma/dt$ are also Poincaré recurrent. This concludes the proof. $\square$

All proofs of the results in this section are provided in Appendix D.

5   Experimental Results

To corroborate the theoretical results presented in prior sections, we performed relevant simulations of quantum games using both discrete MMWU and replicator dynamics. In the rest of this section, we standardize the use of quantum
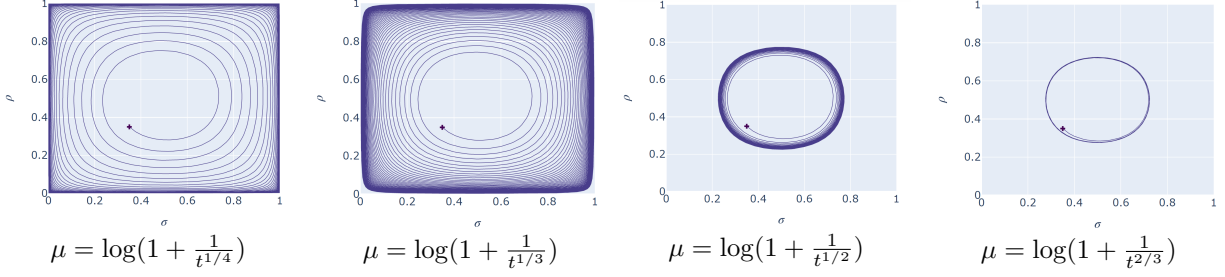
$$\mu = \log(1 + \tfrac{1}{t^{1/4}}) \qquad \mu = \log(1 + \tfrac{1}{t^{1/3}}) \qquad \mu = \log(1 + \tfrac{1}{t^{1/2}}) \qquad \mu = \log(1 + \tfrac{1}{t^{2/3}})$$

Figure 1: Eigenvalue trajectories for quantum Matching Pennies game with decreasing μ values.

game matrices obtained via basis transform (described in more detail in Appendix E). This effectively allows us to transform classical games to the matrix setting.

First, we show the trajectories of the first eigenvalue of each player in a quantum Matching Pennies game, obtained using the discrete MMWU algorithm. We see that in accordance to Theorem 2, the rate of divergence of the trajectories from the uniform Nash goes to zero for cases with rapidly decreasing learning rate μ.

In the case of replicator dynamics, we present Bloch sphere representations of the trajectories in a quantum Matching Pennies game. The Bloch sphere is a unit 2-sphere representation of a qubit, and we utilize it to visualize the orbits of the replicator dynamics. In particular, the density matrix representing the strategy of each player at each time-step is given as a point within the sphere, and we plot the movement of these orbits over time. According to Theorems 3 and 4, we expect the trajectories of the replicator dynamics to stay on the interior of the Bloch sphere, since the surface of the sphere corresponds to the pure states of the system. We see from Figure 2 that over time, the system never reaches the boundary of the sphere, which experimentally agrees with our theory.

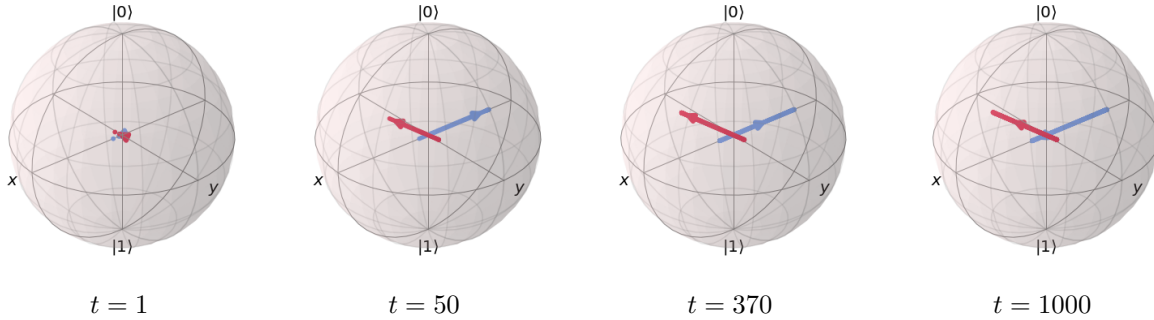$$t = 1 \qquad\qquad t = 50 \qquad\qquad t = 370 \qquad\qquad t = 1000$$

Figure 2: Bloch sphere trajectories for quantum Matching Pennies game between Alice (blue) and Bob (red). The arrowheads represent the current state of each player at each time-step. Notice that over time, the orbits oscillate within the interior of the Bloch sphere.

Finally, in the experiments above we have only studied single qubit systems. We provide further experiments which show that our results hold beyond the single qubit setting in Appendix E.2.

6 Conclusion

In this paper, we studied the properties of Matrix Multiplicative Weights Update and its continuous analogue, quantum replicator dynamics, in the context of two-player zero-sum quantum games. First, we provide a formulation of quantum replicator dynamics which arises from MMWU. Then, we show that such systems exhibit quantum information-theoretic constant of motions. Finally, we show that in quantum replicator systems with interior Nash equilibria, the dynamics exhibit Poincaré recurrence.

This work constitutes an initial step towards analyzing learning behaviour in games with quantum information. In the classical world, showing that conservation laws and recurrence holds has led to a better understanding of game dynamics in increasingly complex settings [44, 48, 53]. Similar work is now potentially possible in the quantum setting.

Moreover, one key implication which can be derived from our recurrence results is to encourage novel discretization methods that preserve the connections to volume preservation and more generally, conservative dynamical systems (e.g.

Hamiltonians). This line of research has received increased interest in recent years [17, 57, 58]. It is a natural question to explore whether such techniques can lead to similar advantages in the quantum setting as well.

Finally, some other interesting directions for future work include:

- extending our results to multi-agent network generalizations of quantum zero-sum games,
- understanding the potential behavior of quantum GANs using our results about learning in quantum zero-sum games, and
- understanding the quantum setting for different classes of games, e.g., potential games.

References

[1] Aaronson, S., Chen, X., Hazan, E., Kale, S., and Nayak, A. (2019). Online learning of quantum states. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124019.

[2] Accardi, L. and Boukas, A. (2020). Von neumann's minimax theorem for continuous quantum games. *Journal of Stochastic Analysis*, 1(2).

[3] Allen-Zhu, Z. and Li, Y. (2017). Follow the compressed leader: Faster online learning of eigenvectors and faster mmwu. In *International Conference on Machine Learning*, pages 116–125. PMLR.

[4] Allen-Zhu, Z., Liao, Z., and Orecchia, L. (2015). 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.

[5] Arora, S., Hazan, E., and Kale, S. (2012). The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164.

[6] Arora, S. and Kale, S. (2016). A combinatorial, primal-dual approach to semidefinite programs. *J. ACM*, 63(2).

[7] Bailey, J. P. and Piliouras, G. (2018). Multiplicative weights update in zero-sum games. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 321–338.

[8] Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., and Lloyd, S. (2017). Quantum machine learning. *Nature*, 549(7671):195–202.

[9] Bondesan, R. and Welling, M. (2021). The hintons in your neural network: a quantum field theory view of deep learning.

[10] Bostanci, J. and Watrous, J. (2021). Quantum game theory and the complexity of approximating quantum nash equilibria. *arXiv preprint arXiv:2102.00512*.

[11] Brandão, F. G., Kalev, A., Li, T., Lin, C. Y.-Y., Svore, K. M., and Wu, X. (2017). Quantum sdp solvers: Large speed-ups, optimality, and applications to quantum learning. *arXiv preprint arXiv:1710.02581*.

[12] Chakrabarti, S., Yiming, H., Li, T., Feizi, S., and Wu, X. (2019). Quantum wasserstein generative adversarial networks. *Advances in Neural Information Processing Systems*, 32.

[13] Cheung, Y. K. (2018). Multiplicative weights updates with constant step-size in graphical constant-sum games. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

[14] Dallaire-Demers, P.-L. and Killoran, N. (2018). Quantum generative adversarial networks. *Physical Review A*, 98(1):012324.

[15] Daskalakis, C., Goldberg, P. W., and Papadimitriou, C. H. (2009). The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259.

[16] Daskalakis, C., Ilyas, A., Syrgkanis, V., and Zeng, H. (2018). Training GANs with optimism. In *ICLR*.

[17] Diakonikolas, J. and Jordan, M. I. (2021). Generalized momentum-based methods: A hamiltonian perspective. *SIAM Journal on Optimization*, 31(1):915–944.

[18] Dong, D., Chen, C., Li, H., and Tarn, T.-J. (2008). Quantum reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 38(5):1207–1220.

[19] Dunjko, V., Taylor, J. M., and Briegel, H. J. (2017). Advances in quantum reinforcement learning. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*.

[20] Fan, K. (1953). Minimax theorems. *Proceedings of the National Academy of Sciences*, 39(1):42–47.

[21] Flokas, L., Vlatakis-Gkaragkounis, E.-V., and Piliouras, G. (2019). Poincaré recurrence, cycles and spurious equilibria in gradient-descent-ascent for non-convex non-concave zero-sum games.

[22] Freund, Y. and Schapire, R. E. (1999). Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29(1-2):79–103.

[23] Fudenberg, D. and Levine, D. K. (1998). *The theory of learning in games*, volume 2. MIT press.

[24] Golden, S. (1965). Lower bounds for the helmholtz function. *Phys. Rev.*, 137:B1127–B1128.

[25] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

[26] Hazan, E., Kale, S., and Shalev-Shwartz, S. (2012). Near-optimal algorithms for online matrix prediction. In *Conference on Learning Theory*, pages 38–1. JMLR Workshop and Conference Proceedings.

[27] Hopkins, S. B., Li, J., and Zhang, F. (2020). Robust and heavy-tailed mean estimation made simple, via regret minimization. *arXiv preprint arXiv:2007.15839*.

[28] Jain, R., Ji, Z., Upadhyay, S., and Watrous, J. (2011). QIP = PSPACE. *J. ACM*, 58(6).

[29] Jain, R. and Watrous, J. (2009). Parallel approximation of non-interactive zero-sum quantum games. In *2009 24th Annual IEEE Conference on Computational Complexity*, pages 243–253. IEEE.

[30] Johansson, J. R., Nation, P. D., and Nori, F. (2012). Qutip: An open-source python framework for the dynamics of open quantum systems. *Computer Physics Communications*, 183(8):1760–1772.

[31] Kale, S. (2007). *Efficient algorithms using the multiplicative weights update method*. Princeton University.

[32] Khan, F. S., Solmeyer, N., Balu, R., and Humble, T. S. (2018). Quantum games: a review of the history, current state, and interpretation. *Quantum Information Processing*, 17(11).

[33] Kodali, N., Abernethy, J., Hays, J., and Kira, Z. (2017). On convergence and stability of gans. *arXiv preprint arXiv:1705.07215*.

[34] Lemke, C. E. and Howson, Jr., J. T. (1964). Equilibrium points of bimatrix games. *Journal of the Society for Industrial and Applied Mathematics*, 12(2):413–423.

[35] Li, T., Chakrabarti, S., and Wu, X. (2019). Sublinear quantum algorithms for training linear and kernel-based classifiers. In *International Conference on Machine Learning*, pages 3815–3824. PMLR.

[36] Li, T., Wang, C., Chakrabarti, S., and Wu, X. (2020). Sublinear classical and quantum algorithms for general matrix games. *arXiv preprint arXiv:2012.06519*.

[37] Lin, T., Jin, C., and Jordan, M. (2020). On gradient descent ascent for nonconvex-concave minimax problems. In *International Conference on Machine Learning*, pages 6083–6093. PMLR.

[38] Littlestone, N. and Warmuth, M. K. (1994). The weighted majority algorithm. *Information and computation*, 108(2):212–261.

[39] Lloyd, S., Mohseni, M., and Rebentrost, P. (2013). Quantum algorithms for supervised and unsupervised machine learning. *arXiv preprint arXiv:1307.0411*.

[40] Lloyd, S. and Weedbrook, C. (2018). Quantum generative adversarial learning. *Physical review letters*, 121(4):040502.

[41] Mermin, N. D. (1990). Simple unified form for the major no-hidden-variables theorems. *Physical review letters*, 65(27):3373.

[42] Mertikopoulos, P., Lecouat, B., Zenati, H., Foo, C.-S., Chandrasekhar, V., and Piliouras, G. (2019). Optimistic mirror descent in saddle-point problems: Going the extra(-gradient) mile. In *ICLR*.

[43] Mertikopoulos, P., Papadimitriou, C. H., and Piliouras, G. (2018). Cycles in adversarial regularized learning. In *SODA '18 - Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2703–2717, New Orleans, United States.

[44] Nagarajan, S. G., Balduzzi, D., and Piliouras, G. (2020). From chaos to order: Symmetry and conservation laws in game dynamics. In *International Conference on Machine Learning*, pages 7186–7196. PMLR.

[45] Nash, J. (1951). Non-cooperative games. *Annals of Mathematics*, 54(2):286–295.

[46] Palaiopanos, G., Panageas, I., and Piliouras, G. (2017). Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos.

[47] Peres, A. (1990). Incompatible results of quantum measurements. *Physics Letters A*, 151(3-4):107–108.

[48] Perolat, J., Munos, R., Lespiau, J.-B., Omidshafiei, S., Rowland, M., Ortega, P., Burch, N., Anthony, T., Balduzzi, D., Vylder, B. D., Piliouras, G., Lanctot, M., and Tuyls, K. (2020). From poincaré recurrence to convergence in imperfect information games: Finding equilibrium via regularization.

[49] Piliouras, G. and Shamma, J. (2014). Optimization despite chaos: Convex relaxations to complex limit sets via poincaré recurrence. In *SODA*.

[50] Piliouras, G., Sim, R., and Skoulakis, S. (2021). Optimal no-regret learning in general games: Bounded regret with unbounded step-sizes via clairvoyant mwu. *arXiv preprint arXiv:2111.14737*.

[51] Poincaré, H. (1890). Sur le problème des trois corps et les équations de la dynamique. *Acta mathematica*, 13(1):A3–A270.

[52] Sandholm, W. H. (2010). *Population Games and Evolutionary Dynamics*. The MIT Press.

[53] Skoulakis, S., Fiez, T., Sim, R., Piliouras, G., and Ratliff, L. (2020). Evolutionary game theory squared: Evolving agents in endogenously evolving zero-sum games. *arXiv preprint arXiv:2012.08382*.

[54] Thompson, C. J. (1965). Inequality with applications in statistical mechanics. *Journal of Mathematical Physics*, 6(11):1812–1813.

[55] Tsuda, K., Rätsch, G., and Warmuth, M. K. (2005). Matrix exponentiated gradient updates for on-line learning and bregman projection. *Journal of Machine Learning Research*, 6(Jun):995–1018.

[56] van Apeldoorn, J. and Gilyén, A. (2019). Quantum algorithms for zero-sum games. *arXiv preprint arXiv:1904.03180*.

[57] Wibisono, A., Tao, M., and Piliouras, G. (2022). Alternating mirror descent for constrained min-max games. *arXiv preprint arXiv:2206.04160*.

[58] Wibisono, A., Wilson, A. C., and Jordan, M. I. (2016). A variational perspective on accelerated methods in optimization. *proceedings of the National Academy of Sciences*, 113(47):E7351–E7358.

[59] Youssry, A., Ferrie, C., and Tomamichel, M. (2019). Efficient online quantum state estimation using a matrix-exponentiated gradient method. *New Journal of Physics*, 21(3):033006.

[60] Zhang, S. (2012). Quantum strategic game theory. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 39–59.

Appendices

This supplementary material contains an overview of additional related work in Appendix A, additional examples and definitions in Appendix B, proofs omitted from the paper for space considerations in Appendix C and D, and further experimental results and details in Appendix E.

A   Additional Related Work

We now describe a broader class of related work. In particular, we focus on work regarding $a)$ the MMWU algorithm and its applications as well as $b)$ quantum algorithms in game theory.

MMWU Algorithm and Applications.   MMWU as an algorithm has been widely studied in the past. For instance, MMWU has been applied to the problem of solving semidefinite programs in [11]. The central idea behind this line of work is the 'quantization' of standard SDP solvers using the matrix multiplicative weights method. Furthermore, MMWU has also been applied to the problem of learning a quantum state, also known as state tomography. [1, 59] have utilized MMWU as a framework for online learning in this setting.

In a broader sense, MMWU can also be considered a member of the class of Matrix Follow-The-Regularized-Leader (FTRL) algorithms. This paradigm has been studied in works such as [4, 26, 27], which admits a wider range of applications such as spectral sparsification of graphs. In addition, some more recent works such as [3] study extensions to FTRL within the regret minimization framework.

Quantum Computing.   Quantum information theory has proved to be a useful tool in the fields of reinforcement learning and deep learning [9, 18, 19]. The basic results of classical game theory, such as von Neumann's minimax theorem and more have been shown to hold in the quantum context [2, 32], while recent work has shown analogous complexity and equilibrium results between classical and quantum games [10, 60].

There is also a rich literature towards solving and analyzing games using quantum algorithms. [56] derives a sublinear time quantum algorithm for solving two-player zero-sum games, while [36] derives a similar algorithm for general matrix games. Finally, quantum algorithms have proven to be useful in many machine learning problems such as classification [35] and GANs [40]. For a general overview of quantum algorithms in machine learning see [39] and [8].

B   Additional Preliminaries

B.1   An Example: Quantum Matching Pennies

To provide a clearer picture of quantum games, we illustrate the differences between the classical and quantum versions of the famous Matching Pennies game. In classical game theory, Matching Pennies is defined by the payoff matrix $P = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$. In the terminology defined above, the payoff observable R would then simply be a diagonal 4×4 matrix with the elements of the classical payoff matrix on the diagonal:

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Let us first consider the case where Alice and Bob play randomized mixed strategies in a classical fashion. Alice and Bob will then simultaneously select *vectors* which represent a probability distribution over the strategies *heads* or *tails*. For example, Alice could return the vector $x_A = [1/2, 1/2]$ and Bob could return the vector $x_B = [1/3, 2/3]$. The expected utility for each player is then determined using the payoff matrix:

$$u_A = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}^\top \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/3 \end{bmatrix} = 0, \quad u_B = \begin{bmatrix} 1/3 \\ 2/3 \end{bmatrix}^\top \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} = 0$$

For the quantum case, Alice and Bob will instead send density matrices that represent their probabilities of playing a given strategy to a referee. For example, Alice might send the matrix $x_A = \begin{bmatrix} 1/2 & 1/2i \\ -1/2i & 1/2 \end{bmatrix}$ and Bob could return the

matrix $x_B = \begin{bmatrix} 1/3 & 1/2i \\ -1/2i & 2/3 \end{bmatrix}$. The payoffs to each player are then calculated using the payoff observable R as defined in Equation 4.

B.2 Classical MWU Algorithm

In the paper we focus on the matrix version of MWU, with particular focus on the quantum interpretation of the positive semi-definite matrices used in Algorithm 1. However, many of the results we show are analogous to results shown in classical settings.

For two-player zero-sum games (i.e. $A = -B^\top$, where A and B are the payoff matrices for players a and b.), the classical MWU algorithm is shown in Algorithm 2 [22]:

Algorithm 2: MWU for Two-Player Zero-Sum Games

Initialize: $x_{i,a}^0 = 1/n$ and $x_{i,b}^0 = 1/n$ for all i (both players, uniform)
for t=1...T **do**
  Players a and b choose action $i \in \mathcal{S}$ with probability $x_{i,a}^t$ and $x_{i,b}^t$ respectively.
  for each action i **do**
  $$x_{i,a}^t = \frac{x_{i,a}^{t-1}(1+\epsilon)^{(Ax_b^{t-1})_i}}{\sum_{j \in \mathcal{S}} x_{j,a}^{t-1}(1+\epsilon)^{(Ax_b^{t-1})_j}}$$
  $$x_{i,b}^t = \frac{x_{i,b}^{t-1}(1+\epsilon)^{(Bx_a^{t-1})_i}}{\sum_{j \in \mathcal{S}} x_{j,b}^{t-1}(1+\epsilon)^{(Bx_a^{t-1})_j}}$$
  end for
end for

Furthermore $(Ax_b^{t-1})_i$ and $(Bx_a^{t-1})_i$ represent the expected utility obtained by players a and b respectively for playing strategy i.

C Proofs from Section 3

In this section we present proofs of results in Section 3 which were omitted from the main paper due to space considerations.

C.1 Proof of Corollary 1

Proof. We use the definition of quantum relative entropy. First, we consider the change in quantum relative entropy from timestep $j-1$ to j.
$$S(\rho^*\|\rho_j) - S(\rho^*\|\rho_{j-1}) = \text{Tr}\left(\rho^*(\log\rho_{j-1} - \log\rho_j)\right)$$
$$= \text{Tr}\left(\rho^*\left(\log\frac{A_{j-1}}{\text{Tr}A_{j-1}} - \log\frac{A_j}{\text{Tr}A_j}\right)\right)$$
$$= \text{Tr}\left(\rho^*\left(\log A_{j-1} - \log A_j + \log\text{Tr}A_j - \log\text{Tr}A_{j-1}\right)\right)$$
$$= \text{Tr}\left(\rho^*\left(-\mu\Phi(\sigma_{j-1}) + \log\text{Tr}A_j - \log\text{Tr}A_{j-1}\right)\right)$$
Hence,
$$S(\rho^*\|\rho_j) - S(\rho^*\|\rho_{j-1}) = -\mu\text{Tr}(\rho^*\Phi(\sigma^*)) + \text{Tr}(\rho^*(\log\text{Tr}A_j - \log\text{Tr}A_{j-1})) \tag{22}$$
Similarly, we know that
$$S(\sigma^*\|\sigma_j) - S(\sigma^*\|\sigma_{j-1}) = -\mu\text{Tr}(\sigma^*\Phi^\dagger(\rho^*)) + \text{Tr}(\sigma^*(\log\text{Tr}B_j - \log\text{Tr}B_{j-1})) \tag{23}$$
Summing up equations 22 and 23:
$$\Delta S(\rho^*\|\rho_j) + \Delta S(\sigma^*\|\sigma_j)$$
$$= \mu\left[-\text{Tr}(\rho^*\Phi(\sigma^*)) + \text{Tr}(\sigma^*\Phi^\dagger(\rho^*))\right] + \log\frac{\text{Tr}A_j}{\text{Tr}A_{j-1}} + \log\frac{\text{Tr}B_j}{\text{Tr}B_{j-1}}$$
$$= \log\frac{\text{Tr}A_j}{\text{Tr}A_{j-1}} + \log\frac{\text{Tr}B_j}{\text{Tr}B_{j-1}}$$
$$\square$$

C.2   Proof of Corollary 2

Proof. We first show the upper bound on the sum $\log\left(\frac{\mathrm{Tr}A_j}{\mathrm{Tr}A_{j-1}}\right) + \log\left(\frac{\mathrm{Tr}B_j}{\mathrm{Tr}B_{j-1}}\right)$. Note that the definition of A_j is:

$$A_j = \exp\left(\mu\sum_{i=0}^{j-1}\Phi(\sigma_i)\right)$$

We can alternatively write:

$$A_j = \exp\left(\log(A_{j-1}) + \mu\Phi(\sigma_{j-1})\right) \tag{24}$$

Hence, by taking the trace of A_{j-1} and applying Facts 1 and 2:

$$
\begin{aligned}
\mathrm{Tr}(A_{j-1}) &\leq \mathrm{Tr}\left[A_j\exp(-\mu\Phi(\sigma_{j-1}))\right] &&\text{(Fact 1)}\\
&\leq \mathrm{Tr}\left[A_j(\mathbb{1} - \mu\exp(-\mu)\Phi(\sigma_{j-1}))\right] &&\text{(Fact 2)}\\
&= (\mathrm{Tr}(A_j))(1 - \mu\exp(-\mu)\mathrm{Tr}\left(\rho_j\Phi(\sigma_{j-1})\right)\\
&\leq (\mathrm{Tr}(A_j))\exp(-\mu\exp(-\mu)\mathrm{Tr}\left(\rho_j\Phi(\sigma_{j-1})\right)). &&(1 + x \leq \exp(x))
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathrm{Tr}(B_{j-1}) &\leq \mathrm{Tr}\left[B_j\exp(\mu\Phi^\dagger(\rho_{j-1}))\right]\\
&\leq \mathrm{Tr}\left[B_j(\mathbb{1} + \mu\exp(\mu)\Phi^\dagger(\rho_{j-1}))\right]\\
&= (\mathrm{Tr}(B_j))(1 + \mu\exp(\mu)\mathrm{Tr}\left(\rho_{j-1}\Phi(\sigma_j)\right))\\
&\leq (\mathrm{Tr}(B_j))\exp(\mu\exp(\mu)\mathrm{Tr}\left(\rho_{j-1}\Phi(\sigma_j)\right)).
\end{aligned}
$$

By rearranging and taking matrix logarithm on both sides, we obtain:

$$\log\left(\frac{\mathrm{Tr}A_j}{\mathrm{Tr}A_{j-1}}\right) \geq \mu\exp(-\mu)\mathrm{Tr}\left(\rho_j\Phi(\sigma_{j-1})\right) \qquad \log\left(\frac{\mathrm{Tr}B_j}{\mathrm{Tr}B_{j-1}}\right) \geq -\mu\exp(\mu)\mathrm{Tr}\left(\rho_{j-1}\Phi(\sigma_j)\right) \tag{25}$$

The inequality in Equation 10 follows by summing up the two inequalities obtained above and applying Corollary 1. Likewise, we can perform similar calculations on $\mathrm{Tr}(A_j)$ and $\mathrm{Tr}(B_j)$ to obtain the statement of Equation 11.

$$
\begin{aligned}
\mathrm{Tr}A_j &\leq \mathrm{Tr}\left[A_{j-1}\exp(\mu\Phi(\sigma_{j-1}))\right]\\
&\leq \mathrm{Tr}\left[A_{j-1}(\mathbb{1} + \mu\exp(\mu)\Phi(\sigma_{j-1}))\right]\\
&= (\mathrm{Tr}A_{j-1})(1 + \mu\exp(\mu)\mathrm{Tr}\left(\rho_{j-1}\Phi(\sigma_{j-1})\right))\\
&\leq (\mathrm{Tr}A_{j-1})\exp(\mu\exp(\mu)\mathrm{Tr}\left(\rho_{j-1}\Phi(\sigma_{j-1})\right))
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathrm{Tr}B_j &\leq \mathrm{Tr}\left[B_{j-1}\exp(-\mu\Phi^\dagger(\rho_{j-1}))\right]\\
&\leq \mathrm{Tr}\left[B_{j-1}(\mathbb{1} - \mu\exp(-\mu)\Phi^\dagger(\rho_{j-1}))\right]\\
&= (\mathrm{Tr}B_{j-1})(1 - \mu\exp(-\mu)\mathrm{Tr}\left(\rho_{j-1}\Phi(\sigma_{j-1})\right))\\
&\leq (\mathrm{Tr}B_{j-1})\exp(-\mu\exp(-\mu)\mathrm{Tr}\left(\rho_{j-1}\Phi(\sigma_{j-1})\right))
\end{aligned}
$$

Again rearranging and taking matrix logarithm on both sides:

$$\log\frac{\mathrm{Tr}A_j}{\mathrm{Tr}A_{j-1}} \leq \mu\exp(\mu)\mathrm{Tr}\left(\rho_{j-1}\Phi(\sigma_{j-1})\right) \qquad \log\frac{\mathrm{Tr}B_j}{\mathrm{Tr}B_{j-1}} \leq -\mu\exp(-\mu)\mathrm{Tr}\left(\rho_{j-1}\Phi(\sigma_{j-1})\right) \tag{26}$$

Thus, an upper bound on $\Delta S(\rho^*\|\rho_j) + \Delta S(\sigma^*\|\sigma_j)$ is:

$$\Delta S(\rho^*\|\rho_j) + \Delta S(\sigma^*\|\sigma_j) \leq \mu\exp(\mu)\mathrm{Tr}\left(\rho_{j-1}\Phi(\sigma_{j-1})\right) - \mu\exp(-\mu)\mathrm{Tr}\left(\rho_{j-1}\Phi(\sigma_{j-1})\right)$$

$$\square$$

C.3 Proof of Theorem 2

In order to prove Theorem 2, we utilize Corollary 2 and take the limit of the inequalities as $\mu \to 0$.

Proof. Consider the MMWU update from time $j-1$ to j. As $\mu \to 0$, ρ_j and σ_j do not change more than $O(\mu)$. Indeed, since all payoffs in the game are bounded in the MMWU update from time $j-1$ to j, all entries in the numerators increase by at most $\exp(O(\mu)) = 1 + O(\mu)$. Likewise, the denominator is at least as large, but also upper bounded by the previous value of the denominator multiplied by $(1 + O(\mu))$. Hence, every entry in the outputs ρ_j and σ_j are at most $O(\mu)$ from ρ_{j-1} and σ_{j-1}. Moreover, we have that $Tr(\rho_j \Phi(\sigma_{j-1}))$, $Tr(\rho_{j-1}\Phi(\sigma_j))$ and $Tr(\rho_{j-1}\Phi(\sigma_{j-1}))$ are all within $O(\mu)$ of each other. Using Taylor expansion, both the upper bounds and lower bounds given in Equations 10,11 are of the order of $O(\mu^2)$ and the theorem follows.

$\square$

D Proofs and Additional Results from Section 4

In this section, we focus on the key results shown in Section 4. In particular, we first present some connections between the quantum and classical replicator dynamics. Then, we show the comprehensive proof of Theorem 4, which relies on Lemmas 1 and 2. In addition, the proofs of all other results can also be found in this section.

D.1 Proof of Observation 1

Proof. It suffices to show that at the limit, the integrals $A(t)$ and $B(t)$ defined in 13 and 15 can be written in the discrete summation form as seen in Algorithm 1. First we write the Riemann sum for $A(t)$ by taking j infinitesimal intervals in time interval $[0, t]$, each of width μ:

$$A(t) = \int_0^t \Phi(\sigma(\tau))d\tau = \sum_{i=0}^{j-1} \left(\frac{t}{j}\right) \Phi(\sigma_i)$$

However $t = \mu j$, so the above can be written as $\sum_{i=0}^{j-1} \mu \Phi(\sigma_i)$. Taking limit as $\mu \to 0$, it is clear to see that the matrix exponent of the continuous-time trajectory $A(t)$ is equal to its discrete-time counterpart A_j. Hence, the trajectories $\rho(t)$ and ρ_j are equivalent at the limit. A similar argument holds for $\sigma(t)$ and σ_j at the limit. $\square$

D.2 Connections between Quantum and Classical Replicator Dynamics

In this section we show that in the special case where commutativity holds, the quantum replicator dynamics presented in Section 4 are equivalent to that of the classical replicator equations. Recall that we used the integral form of the replicator dynamics in the main text. However, typically in the classical case one can write the replicator dynamics in a form that represents the utilities obtained by each player:

$$\dot{x} = x \left(Ay - \left(x^\top Ay\right) \vec{\mathbb{1}} \right)$$

where x and y are n-dimensional probability vectors representing the strategies of each player, A is the payoff matrix and $\vec{\mathbb{1}}$ is the n-dimensional all-one vector. In the quantum setting, we require a few additional results and assumptions in order to write the Equations 17 in the classical form. The following lemma arrives directly as a result of the series definition of matrix exponential.

Lemma 3. *If matrices P and $\frac{dP}{dt}$ commute then*

$$\frac{d(\exp(P))}{dt} = \frac{dP}{dt} \cdot \exp(P) = \exp(P) \cdot \frac{dP}{dt}$$

Proof.

$$\begin{aligned}
\frac{d}{dt}(\exp(P)) &= \frac{d}{dt}\left[\mathbb{1} + P + \frac{1}{2}P^2 + \frac{1}{3!}P^3 + \cdots\right] \\
&= \frac{dP}{dt}\left[\mathbb{1} + P + \frac{1}{2}P^2 + \cdots\right] \\
&= \frac{dP}{dt} \cdot \exp(P) = \exp(P) \cdot \frac{dP}{dt} \qquad \text{(by commutativity)}
\end{aligned}$$

$\square$

By leveraging Lemma 3 and assuming that commutativity holds, we can state the following lemma:

Lemma 4. *If $\int_0^t \Phi(\sigma(\tau))d\tau$ & $\Phi(\sigma(t))$, and $\int_0^t \Phi^\dagger(\rho(\tau))d\tau$ & $\Phi^\dagger(\rho(t))$ commute, then the replicator system when applied to a two-player zero-sum quantum game is equivalent to:*

$$
\begin{aligned}
d\rho/dt &= \rho[\Phi(\sigma) - \mathrm{Tr}(\rho\Phi(\sigma))\mathbb{1}] & (27) \\
d\sigma/dt &= \sigma[-\Phi^\dagger(\rho) + \mathrm{Tr}(\sigma\Phi^\dagger(\rho))\mathbb{1}] & (28)
\end{aligned}
$$

Proof. By definition we have that $\rho = \exp(A)/\mathrm{Tr}(\exp(A))$ where $A = \int_0^t \Phi(\sigma(\tau))d\tau$. By applying Lemma 3 we have that if $\int_0^t \Phi(\sigma(\tau))d\tau$ and $\Phi(\sigma(t))$ commute then $d\exp(A)/dt = \exp(A)\Phi(\sigma) = \Phi(\sigma)\exp(A)$.

$$
\begin{aligned}
d\rho/dt &= \frac{(d\exp(A)/dt)\mathrm{Tr}(\exp(A)) - \exp(A)d(\mathrm{Tr}(\exp(A)))/dt}{(\mathrm{Tr}\exp(A))^2} \\
&= \frac{(d\exp(A)/dt)\mathrm{Tr}(\exp(A)) - \exp(A)\mathrm{Tr}(d\exp(A)/dt)}{(\mathrm{Tr}\exp(A))^2} \\
&= \frac{\exp(A)\Phi(\sigma)\mathrm{Tr}(\exp(A)) - \exp(A)\mathrm{Tr}(\exp(A)\Phi(\sigma))}{(\mathrm{Tr}\exp(A))^2} \\
&= \rho[\Phi(\sigma) - \mathrm{Tr}(\rho\Phi(\sigma))\mathbb{1}]
\end{aligned}
$$

The proof in the case of $d\sigma/dt$ is similar. $\qquad\square$

We can see that Equations 27 and 28 take a familiar form - indeed, the quantum replicator dynamics model the difference in utility obtained by each player compared to the average utility in much the same way as the classical case. In general, commutativity holds only in the case when all matrices $\Phi^\dagger(\rho(t))$ and $\Phi(\sigma(t))$ are diagonal, which is precisely the classical setting. The assumption that commutativity holds in the quantum setting is a very strong one, and in general commutativity does not hold once we enter the realm of quantum information.

Reminiscent of a similar result in classical evolutionary game theory, one can write a proof for the invariance of total entropy directly using the formulation of replicator dynamics. However given the commutativity assumption, this 'proof' only holds in the case where we are writing classical probability vectors and game payoff matrices in the quantum notation.

Proof of Theorem 3 (Classical case). We will prove that $d\mathrm{Tr}\big(\rho^*(\log\rho^* - \log\rho(t)) + \sigma^*(\log\sigma^* - \log\sigma(t))\big)/dt = 0$. To do that we will focus on the terms related to the first agent, i.e., $d\mathrm{Tr}\big(\rho^*(\log\rho^* - \log\rho(t))\big)/dt = -d\mathrm{Tr}\big(\rho^*\log\rho(t))\big)/dt$.

$$
\begin{aligned}
d\mathrm{Tr}\big(\rho^*\log\rho(t))\big)/dt &= \mathrm{Tr}\big(\rho^*d(\log\rho(t))/dt\big) \\
&= \mathrm{Tr}\big(\rho^*\rho(t)^{-1}d\rho(t)/dt\big) \quad \text{(by Lemma 3}^3\text{)} \\
&= \mathrm{Tr}\big(\rho^*\rho^{-1}\rho[\Phi(\sigma) - \mathrm{Tr}(\rho\Phi(\sigma))I]\big) \quad \text{(by Lemma 4)} \\
&= \mathrm{Tr}\big(\rho^*[\Phi(\sigma) - \mathrm{Tr}(\rho\Phi(\sigma))I]\big) \\
&= \mathrm{Tr}\big(\rho^*\Phi(\sigma)\big) - \mathrm{Tr}\big(\mathrm{Tr}(\rho\Phi(\sigma))\rho^*]\big) \\
&= \mathrm{Tr}\big(\rho^*\Phi(\sigma)\big) - \mathrm{Tr}(\rho\Phi(\sigma)) \quad \text{(Since } \mathrm{Tr}\big(\rho^*\big) = 1\text{)} \\
&= \mathrm{Tr}\big(\rho^*\Phi(\sigma^*)\big) - \mathrm{Tr}(\rho\Phi(\sigma))
\end{aligned}
$$

The last line comes from the assumption that the equilibrium (ρ^*, σ^*) is 'fully mixed', i.e. the payoff of an agent when deviating from the Nash equilibrium to any other strategy remains exactly equal to her equilibrium payoff (e.g., Rock-Paper-Scissors). A similar analysis for the second agent results in:

$$
d\mathrm{Tr}\big(\sigma^*\log\sigma(t))\big)/dt = -\mathrm{Tr}\big(\sigma^*\Phi^\dagger(\rho^*)\big) + \mathrm{Tr}(\sigma\Phi^\dagger(\rho))
$$

These terms cancel out and the theorem follows. $\qquad\square$

---

[3]Note here that $\rho(t)^{-1}$ exists since the exponential of a matrix is always an invertible matrix.

However in general we cannot use the above formulation, but rather need an argument tailored to the quantum setting using Theorem 2 and Observation 1. The general proof of Theorem 3 is presented here:

D.3 Proof of Theorem 3

In the proof of this theorem, we leverage Observation 1 and Theorem 2.

Proof. First we note that by expanding the limit definition of the derivative we obtain:

$$
\frac{d\big(S(\rho^*\|\rho(t)) + S(\sigma^*\|\sigma(t))\big)}{dt}
$$
$$
= \lim_{h\to 0} \frac{\big(S(\rho^*\|\rho(t+h)) + S(\sigma^*\|\sigma(t+h))\big) - \big(S(\rho^*\|\rho(t)) + S(\sigma^*\|\sigma(t))\big)}{h}
$$

Here, we apply Observation 1 in the following manner: If we perform the substitution $j \to t+h$, $j-1 \to t$ and $\mu \to h$, we obtain a reformulation of the same limit in the language of stepsizes of duration μ. As $\mu \to 0$, the difference between the continuous time flow solution and its discrete-time Euler approximation is of order $O(\mu^2)$. Thus, we can compute this limit evaluated along the points of the discrete-time MMWU trajectory. This limit by Theorem 3.5 is equal to zero.

$\square$

D.4 Proof of Proposition 1

Proof. We first consider the definition of $\rho'(t)$.

$$
\rho'(t) = \frac{\exp(A'(t))}{\mathrm{Tr}(\exp(A'(t)))}
$$
$$
= \frac{\exp\left(\int_0^t \Phi(\sigma(\tau))d\tau - (v^\dagger A(t)v)\mathbb{1}\right)}{\mathrm{Tr}\left(\exp\left(\int_0^t \Phi(\sigma(\tau))d\tau - (v^\dagger A(t)v)\mathbb{1}\right)\right)}
$$
$$
= \frac{\exp\left(\int_0^t \Phi(\sigma(\tau))d\tau\right) \cdot \exp\left(-(v^\dagger A(t)v)\mathbb{1}\right)}{\mathrm{Tr}\left(\exp\left(\int_0^t \Phi(\sigma(\tau))d\tau\right) \cdot \exp\left(-(v^\dagger A(t)v)\mathbb{1}\right)\right)}
$$

Note that the denominator can be written as the trace of a matrix where each diagonal entry is the corresponding value of A multiplied by $\exp(-v^\dagger A(t)v)$. Thus, the above can be rewritten as

$$
\rho'(t) = \frac{\exp(A(t)) \cdot \exp\left(-(v^\dagger A(t)v)\mathbb{1}\right)}{\exp(-v^\dagger A(t)v) \cdot \mathrm{Tr}\left(\exp(A(t))\right)}
$$
$$
= \frac{\exp(A(t))}{\mathrm{Tr}(\exp(A(t)))} \cdot \mathbb{1}
$$
$$
= \rho(t)
$$

We can perform similar computations to show the same holds true for $\sigma'(t)$ and $\sigma(t)$. $\square$

D.5 Proof of Proposition 2

Proof. Consider the map f between A' and ρ. We have shown in Proposition 1 that the map in one direction is $\rho = f(A') = \frac{\exp(A')}{\mathrm{Tr}(\exp(A'))}$. This is continuously differentiable. Now let us consider the inverse map f^{-1}. In particular, we show that the inverse mapping exists and is equal to $A' = f^{-1}(\rho) = \log(\rho) - (v^\dagger \log(\rho)v)\mathbb{1}$.

Indeed,

$$f(f^{-1}(\rho)) = \frac{\exp(\log(\rho) - (v^\dagger \log(\rho)v)\mathbb{1})}{\text{Tr}(\exp(\log(\rho) - (v^\dagger \log(\rho)v)\mathbb{1}))}$$

$$= \frac{\exp(\log(\rho)) \cdot \exp\left(-(v^\dagger \log(\rho)v)\mathbb{1}\right)}{\text{Tr}\left(\rho \cdot \exp\left(-(v^\dagger \log(\rho)v)\mathbb{1}\right)\right)}$$

$$= \frac{\rho \cdot \exp\left(-(v^\dagger \log(\rho)v)\mathbb{1}\right)}{\exp\left(-(v^\dagger \log(\rho)v)\right)}$$

$$= \rho \cdot \mathbb{1} = \rho$$

where we have used the fact that $\text{Tr}(\rho) = 1$. Furthermore, note that the inverse mapping $f^{-1}(\rho) = \log(\rho) - (v^\dagger \log(\rho)v)\mathbb{1}$ is smooth.

Hence, since f is a bijection and the inverse map is differentiable, $A'(t)$ and $\rho(t)$ are diffeomorphic to each other. Similarly, $B'(t)$ and $\sigma(t)$ are diffeomorphic to each other. $\qquad\square$

Lemmas 1 and 2 are crucial to proving our recurrence result. Typically, volume conservation is not difficult to prove, since the canonical transformation guarantees that the dynamical system becomes separable, which then allows for a simple application of Liouville's theorem. However, proving that the orbits of the dynamical system are bounded is more challenging. We carefully consider the definitions of $A'(t)$ and $B'(t)$ and utilize properties regarding their eigenvalues to show this property.

D.6   Proof of Lemma 1

Proof. Let $\Psi_{A'}$ denote the flow of $A'(t)$ and $\Psi_{B'}$ denote the flow of $B'(t)$. First let us consider $A'(t)$.

$$\dot{A'} = \frac{dA'}{dt} = \Phi(\sigma(t)) - (v^\dagger\Phi(\sigma(t))v)\mathbb{1}$$

$$= \Phi\left(\frac{\exp(B(t))}{\text{Tr}(\exp(B(t)))}\right) - \left(v^\dagger\Phi\left(\frac{\exp(B(t))}{\text{Tr}(\exp(B(t)))}\right)v\right)\mathbb{1}$$

Now, we derive the first order partial-derivatives of $F(A')$ with respect to $A'(t)$:

$$\frac{\partial F(A')}{\partial A'} = \frac{\partial}{\partial A'}\left(\Phi\left(\frac{\exp(B(t))}{\text{Tr}(\exp(B(t)))}\right) - \left(v^\dagger\Phi\left(\frac{\exp(B(t))}{\text{Tr}(\exp(B(t)))}\right)v\right)\mathbb{1}\right)$$

$$= \frac{\partial}{\partial A'}\left(\Phi\left(\frac{\exp(B'(t))}{\text{Tr}(\exp(B'(t)))}\right) - \left(v^\dagger\Phi\left(\frac{\exp(B'(t))}{\text{Tr}(\exp(B'(t)))}\right)v\right)\mathbb{1}\right)$$

$$= 0$$

Clearly, the partial derivative $\frac{\partial F(A')}{\partial A'}$ only depends on the value of $B'(t)$ and not $A'(t)$. This implies that the vector field is separable, and so the first order partial derivative with respect to $A'(t)$ is zero. By a similar argument, the partial derivative $\frac{\partial F(B')}{\partial B'}$ is also zero. By definition, the diagonal of the Jacobian matrix describing the vector fields is zero, and hence the divergence (which is the trace of the Jacobian) is zero as well. We can then directly apply Liouville's theorem to conclude that the flows $\Psi_{A'}$ and $\Psi_{B'}$ are volume preserving. $\qquad\square$

D.7   Proof of Lemma 2

Proof. First, $\rho(0) = \frac{\exp(A(0))}{\text{Tr}(\exp(A(0)))}$ (resp. $\sigma(0)$) is bounded away from the boundary, since we assumed that since $A(0)$ and $B(0)$ are finite matrices. Moreover, $A'(0)$ and $B'(0)$ are also finite. The sum of quantum relative entropies at time 0 is thus:

$$S(\rho^*\|\rho(0)) + S(\sigma^*\|\sigma(0)) = \text{Tr}(\rho^*(\log\rho^* - \log\rho(0))) + \text{Tr}(\sigma^*(\log\sigma^* - \log\sigma(0)))$$

This sum is finite since $\rho(0), \sigma(0)$ have full support. Indeed, the support of ρ^* and σ^* are contained in the support of $\rho(0)$ and $\sigma(0)$ respectively. By Theorem 3, the sum of quantum relative entropies $S(\rho^*\|\rho(t)) + S(\sigma^*\|\sigma(t))$ is bounded above by a fixed value C for all time t. Now, let $\lambda_{\min}(\rho^*)$ denote the minimum eigenvalue of ρ^*. We have that $\rho^* \geq \lambda_{\min}(\rho^*)\mathbb{1}$. Note the following:

$$\text{Tr}(\rho^*(\log\rho^*)) - \text{Tr}(\rho^*(\log\rho(t))) < C$$

$$-\text{Tr}(\rho^*(\log\rho(t))) < D$$

$$-\text{Tr}(\lambda_{\min}(\rho^*)\mathbb{1}(\log\rho(t))) < D$$

where C and D are positive and finite real numbers. Using the fact that the trace of a matrix is basis independent, we use a basis where $\rho(t)$ is diagonal. Here, the minimum eigenvalue of $\rho(t)$ cannot go to zero, otherwise $-\text{Tr}(\lambda_{\min}(\rho^*)\mathbb{1}(\log \rho(t)))$ goes to $+\infty$, a clear contradiction to the inequality above. A similar argument holds for $\sigma(t)$. This is equivalent to saying that all the eigenvalues of $\rho(t)$ and $\sigma(t)$ lie in the interval $[\epsilon, 1-\epsilon]$ for some small $\epsilon > 0$.

Note that the value of $A'_{1,1}(t)$ is always zero by design. We also know by construction that $A'(t)$ is Hermitian, since $\sigma(t)$ is Hermitian, Φ constitutes a hermicity preserving map and the integral of Hermitian matrices remains Hermitian. Hence, the smallest eigenvalue of $A'(t)$, $\lambda_{\min}(A'(t))$ is upper bounded by the smallest element on the diagonal of $A'(t)$, namely $\lambda_{\min}(A'(t)) \leq 0$, since $A'_{1,1}(t) = 0$. If the largest eigenvalue $\lambda_{\max}(A'(t))$ goes to $+\infty$, then the corresponding smallest eigenvalue of $\rho(t) = \frac{\exp(A'(t))}{\text{Tr}(\exp(A'(t)))}$ goes to 0. Likewise if the smallest eigenvalue $\lambda_{\min}(A'(t))$ goes to $-\infty$, the smallest eigenvalue of $\rho(t)$ goes to 0 as well. We have previously shown that this cannot occur, since the eigenvalues of $\rho(t)$ are all bounded away from zero. Hence, it follows that all eigenvalues of $A'(t)$ are bounded away from infinity.

Now, consider the scalar $v^\dagger A'(t)w$ for arbitrary bounded vectors v and w. $A'(t)$ is Hermitian, so by the spectral theorem it has an orthonormal eigenbasis. Now we express v and w in terms of this n-dimensional eigenbasis (which we denote by $\mathbb{e}$):

$$v = \sum_{i=1}^{n} a_i \mathbb{e}_i \quad \text{and} \quad w = \sum_{i=1}^{n} b_i \mathbb{e}_i$$

where a_i and b_i are scalars. Thus,

$$v^\dagger A'(t)w = \left(\sum_{i=1}^{n} a_i \mathbb{e}_i\right)^\dagger A'(t) \left(\sum_{i=1}^{n} b_i \mathbb{e}_i\right)$$
$$= \sum_{i=1}^{n} a_i b_i \lambda_i$$

where λ_i are the eigenvalues of $A'(t)$ corresponding to $\mathbb{e}_i$. Here a_i, b_i and λ_i are all bounded, so it follows that all elements of matrix $A'(t)$ are bounded away from infinity as well. Likewise, the elements of matrix $B'(t)$ remain bounded from infinity and thus the dynamical system described by $A'(t)$ and $B'(t)$ have bounded orbits.

$\square$

E   Experiments

In order to formalize a method for simulating and understanding quantum games, we propose a method to generate complex R matrices with the same eigenvalues as the values of the classical payoff matrix. This means that for any standard eigenbasis of a classical game, we create a new, complex eigenbasis where the eigenvalues are the same, but the eigenvectors are complex. Intuitively, this can be viewed as mapping a classical game matrix to the Hilbert space, effectively allowing generalizations of classically studied games to the semi-definite context. We introduce the following lemma, which describes a method to obtain a basis transformation to Hilbert space.

Lemma 5. *For any $n \times n$ two player zero-sum game with classical payoff matrix P, let $\lambda_{i,j} = P_{i,j}$ and let V and W be unitary matrices of size $n \times n$. Then, the matrix R defined as*

$$R = \sum_{i,j} \lambda_{i,j}(V_i \otimes W_j)(V_i^\dagger \otimes W_j^\dagger) \tag{29}$$

produces a transformation from the classical payoff space $\mathcal{P}$ to the density matrix space $\mathcal{C}$. In particular, R is a complex $n^2 \times n^2$ matrix which satisfies the following properties:

- *The matrix R is Hermitian.*

- *The eigenvalues of R correspond to the elements of the classical payoff matrix P.*

Proof. First note that Equation 29 is equivalent to taking the following basis transformation:

$$\sum_{i,j} \lambda_{i,j}(V_i \otimes W_j)(V_i^\dagger \otimes W_j^\dagger) = (V \otimes W)\Lambda(V^\dagger \otimes W^\dagger)$$

19

where $\Lambda = \begin{bmatrix} \lambda_{1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{n,n} \end{bmatrix}$ is the $n^2 \times n^2$ diagonal matrix formed by placing the elements of classical payoff matrix P on the diagonal. Moreover, V and W are both unitary, so the tensor products $V \otimes W$ and $V^\dagger \otimes W^\dagger$ are also unitary. This holds because:

$$(V \otimes W)(V \otimes W)^\dagger = (V \otimes W)(V^\dagger \otimes W^\dagger)$$
$$= VV^\dagger \otimes WW^\dagger$$
$$= \mathbb{1} \otimes \mathbb{1} = \mathbb{1}$$

This implies that R can be written as a unitary diagonalization of the form $PDP^\dagger$, where $P = (V \otimes W)$ and $D = \Lambda$. Hence R is Hermitian and has real eigenvalues by the spectral theorem. In addition, by construction the eigenvalues of R are the elements of the classical payoff matrix P. $\qquad\square$

Moreover, the initial conditions of a simulation of a two-player quantum zero-sum game can be similarly obtained by using the same orthonormal bases as in Lemma 5 and applying a change of basis to the initial conditions in the standard basis.

E.1 Additional Experimental Details (Matching Pennies)

In the main text, we showed several simulations that illustrate the theoretical results in Sections 3 and 4. Here we present detailed descriptions of the simulations, as well as additional experiments which serve to corroborate auxiliary results in the paper.

To generate Figure 1, we used a modified version of Algorithm 1, where instead of constant step-size μ we define $\mu = \log(1 + \frac{1}{t^a})$, where a is an exponent which determines the rate of decrease of the step-size, and t is the time-step of the simulation. Intuitively, this means that for higher values of a, the step-sizes go to zero at a faster rate. Empirically, we observe that if the exponent is greater than $\frac{1}{2}$, then we can clearly see the phenomenon described in Theorem 2, since the quantum relative entropy in the system increases when the trajectories move away from the fully mixed Nash equilibrium.

In the case of replicator dynamics, we show experimentally (Figure 3) that the sum of quantum relative entropies for both players is a constant of motion as shown in Theorem 3. We first use Lemma 5 to obtain the R matrix of a quantum game given some complex basis and classical payoff matrix (Matching Pennies, for instance). Then, we run the discrete Algorithm 1 with decreasing step-size to compute the interior ϵ-approximate Nash of the quantum game. The replicator equations are then solved given a set of initial conditions for each player. Thereafter, we compute the quantum relative entropy between each player's strategy at each discretized time step and the ϵ-approximate Nash equilibrium.

Notice that although the quantum relative entropies of the first player (pink) and second player (purple) are oscillating over time, their sum remains approximately constant.
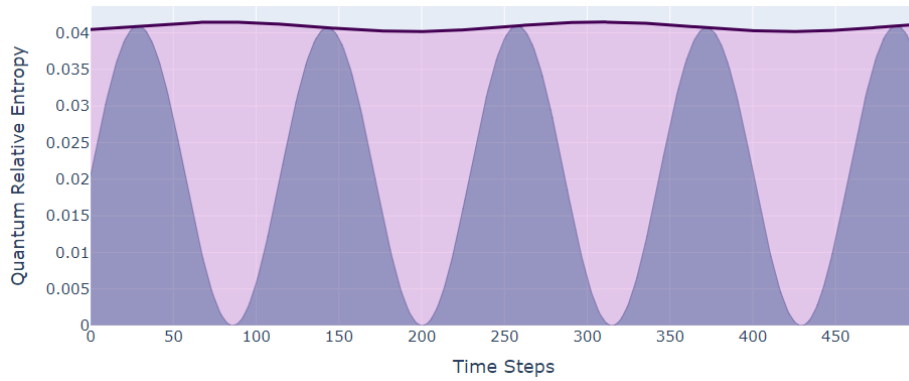

Figure 3: Approximately constant sum of quantum relative entropy values for two-player quantum Matching Pennies game.

Finally, in order to generate Figure 2, we perform a similar simulation to the above. However, we also transform the trajectories of each player using the Pauli matrices, which are given as:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{30}$$

The expectations of the density matrices under this transformation (commonly referred to as the Bloch vector) produce the coordinates on the x, y and z axes within the Bloch sphere. We plot these coordinates for each time step and hence obtain the trajectories for each player in the Bloch sphere.

E.2 Larger-Scale Experiments

In the main text, we study a single-qubit system in the form of a Matching Pennies game. However, our results extend to games with larger strategy sets, some of which can be interpreted as multi-qubit systems. We study a 16×16 game with a payoff matrix in $[-1, 1]$ which can be played via 2 qubits. Moreoever, this game is selected because it has an interior (and uniform) Nash equilibrium. Similarly, we generate an 64×64 generalization of the 16×16 game to represent an example of a 3-qubit game with interior Nash. In Figure 4, we plot the Frobenius norm between the initial condition and the system state when updated using quantum replicator dynamics (Equations 17). Notice that despite the relatively erratic behaviour of the strategies, they eventually return to the initial condition, implying that recurrence holds (Theorem 4).

Furthermore, to confirm the volume conservation property shown in Theorem 3 we perform the same simulation as in Figure 3 for the same 2 and 3 qubit systems as above, obtaining Figure 5. However, a minor note is that we perform these simulations without the basis transform of Lemma 5 to avoid potential numerical errors since the matrices are of larger dimension. The initial conditions are random density matrices, as before. Note that the sum of quantum relative entropies between each player and the interior Nash equilibrium is constant as expected from our theoretical results.

E.3 Reproducibility Details

All experiments performed for this work were done using Python 3.7 and have been compiled into a Jupyter notebook for ease of viewing. Running the code requires basic scientific computing packages such as NumPy, SciPy and Cython, as well as data visualization packages such as Matplotlib and Plotly. In addition, we use QuTip [30], a quantum simulation package. Most of our code has been edited such that it can be easily executed on a standard computer in a matter of minutes.
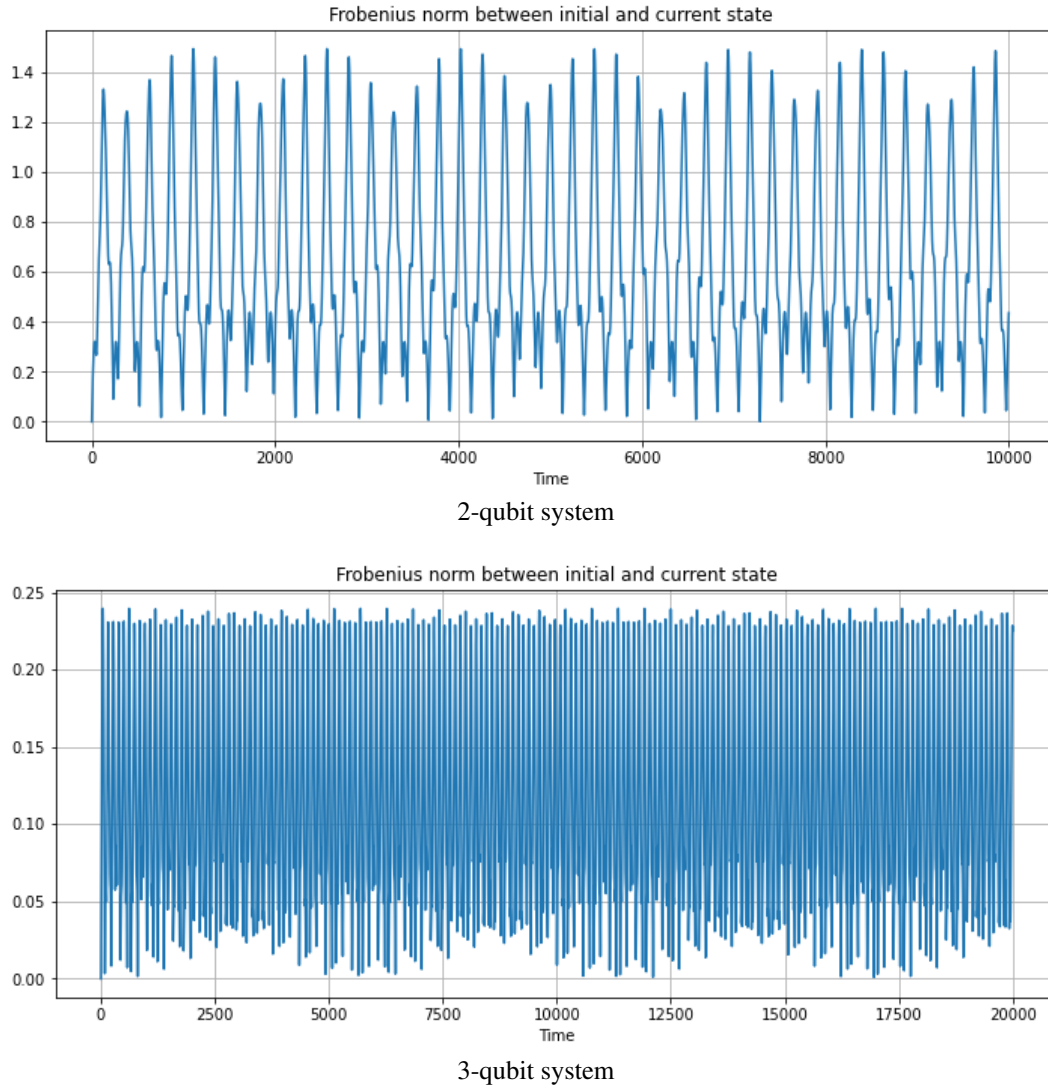
2-qubit system

3-qubit system

Figure 4: Frobenius norm between the initial condition and current system state for 2 and 3 qubit systems with interior Nash equilibrium. Note that at approx. 7200 iterations the 2-qubit system returns arbitrarily close to 0, and the same occurs at approx. 12000 iterations for the 3-qubit system.
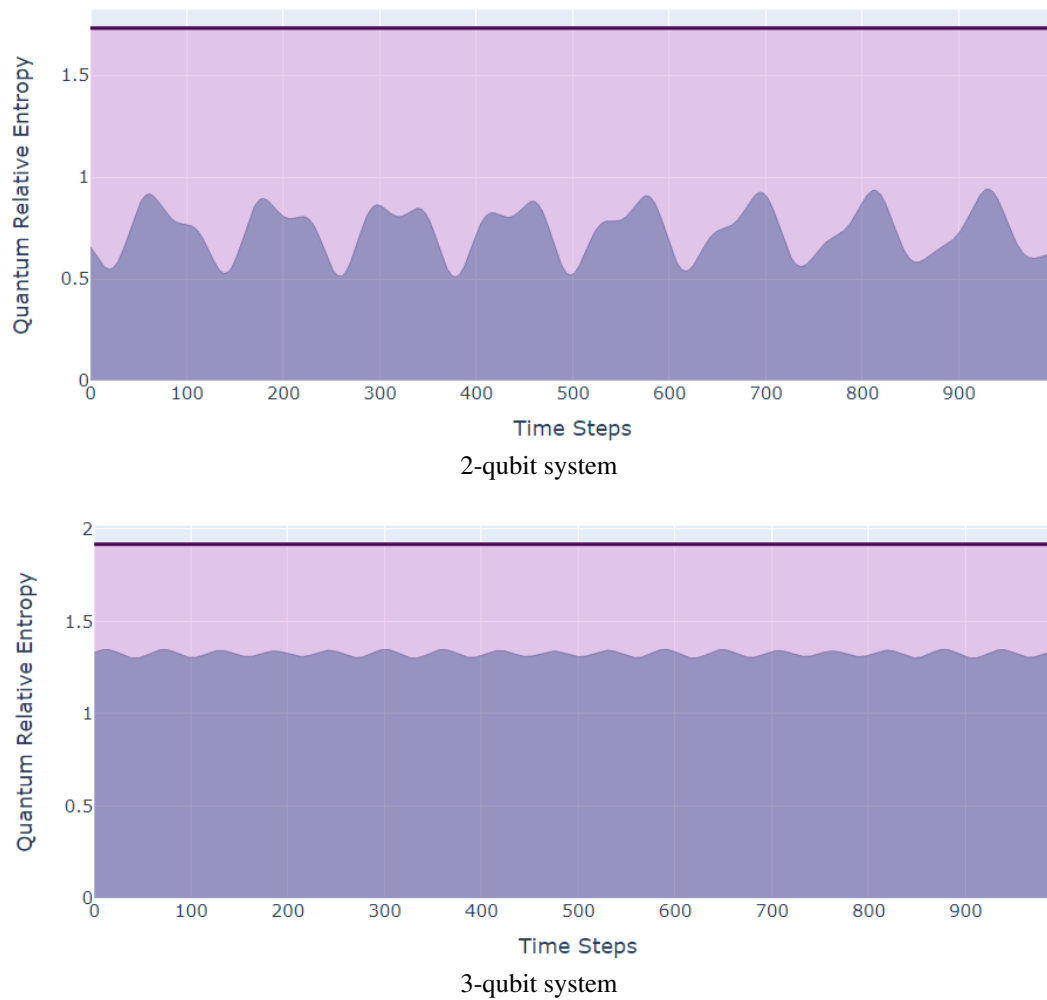
2-qubit system

3-qubit system

Figure 5: Sum of quantum relative entropies of each player to the Nash equilibrium for 2 and 3 qubit systems with interior Nash.