Learning in Quantum Common-Interest Games and the Separability Problem

Wayne Lin[1], Georgios Piliouras[1], Ryann Sim[1], and Antonios Varvitsiotis[1,2,3]

[1]Singapore University of Technology and Design, Singapore
[2]Centre for Quantum Technologies, National University of Singapore, Singapore
[3]Archimedes/Athena RC, Greece

Learning in games has emerged as a powerful tool for machine learning with numerous applications. Quantum games model interactions between strategic players who have access to quantum resources, and several recent works have studied learning in the competitive regime of quantum zero-sum games. Going beyond this setting, we introduce quantum common-interest games (CIGs) where players have density matrices as strategies and their interests are perfectly aligned. We bridge the gap between optimization and game theory by establishing the equivalence between KKT (first-order stationary) points of an instance of the Best Separable State (BSS) problem and the Nash equilibria of its corresponding quantum CIG. This allows learning dynamics for the quantum CIG to be seen as decentralized algorithms for the BSS problem. Taking the perspective of learning in games, we then introduce non-commutative extensions of the continuous-time replicator dynamics and the discrete-time best response dynamics/linear multiplicative weights update for learning in quantum CIGs. We prove analogues of classical convergence results of the dynamics and explore differences which arise in the quantum setting. Finally, we corroborate our theoretical findings through extensive experiments.

1 Introduction

The study of game theory has long revolved around the characterization and computation of Nash equilibria in various classes of games. A popular perspective is that of learning in games, where agents play repeated rounds of a game and update their strategies using information obtained from past interactions. A natural question which arises is when these learning processes give rise to Nash equilibria, and it turns out that in classical two-player zero-sum games and common-interest games (or the more general class of potential games), simple learning dynamics suffice to give rise to Nash equilibria. This connection between learning in games and game-theoretic equilibria has been leveraged in machine learning applications, such as solving large-scale zero-sum games like Poker [1] and Go [2] and understanding the behavior of Generative Adversarial Networks [3, 4, 5]. Many of these success stories have been in zero-sum games, but the focus has recently begun shifting toward understanding learning in games where agents have aligned interests [6, 7]. For instance, many recent works in multi-agent reinforcement learning have explored the challenging frontier of multi-agent coordination [8, 9, 10, 11]. As such, studying the convergence of learning dynamics to equilibria in identical/common-interest games, where players share a common utility function, is an important endeavor with many potential applications.

More recently, the study of learning in quantum games where players have access to quantum resources has also begun to gain traction. In classical (normal-form) games, mixed strategies

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

correspond to probability simplex vectors that capture classical randomness over a finite set of pure strategies; quantum games are a natural extension of the classical game theory framework and have many formulations, see e.g. [12, 13, 14, 15]. The study of learning in quantum games has so far primarily taken place in the two-player *zero-sum* setting where players select density matrices as strategies and one player obtains a payoff from a bilinear utility function which corresponds to a measurement on the joint state, while the other player gets the negative payoff [16, 17, 18]. Beyond the two-player zero-sum setting, [19, 20] have studied continuous- and discrete-time learning in general-sum quantum games. However, learning in quantum games where agents share a common utility function has yet to be explored in the literature.

In this work, we introduce quantum common-interest games (quantum CIGs) and study learning dynamics—i.e., dynamics that the players use to update their strategies over repeated rounds of play—in this class of games. In a (two-player) quantum CIG, there are two agents—which for concreteness we name Alice and Bob—who control quantum registers $\mathcal{A}$ and $\mathcal{B}$ and have strategies given by density matrices in $D(\mathcal{A})$ and $D(\mathcal{B})$ respectively. Upon playing the strategy profile $(\rho, \sigma) \in D(\mathcal{A}) \times D(\mathcal{B})$, both players receive a common utility $u(\rho, \sigma) = \langle R, \rho \otimes \sigma \rangle$, where R is a Hermitian matrix assumed to be positive definite that we refer to as the *game operator*. The canonical solution concept is the *Nash equilibria* (NE) of the game, which are the strategy profiles $(\rho, \sigma) \in D(\mathcal{A}) \times D(\mathcal{B})$ such that Alice's and Bob's strategies are best responses to each other, i.e.,

$$u(\rho, \sigma) \geq u(\rho', \sigma) \; \forall \; \rho' \in D(\mathcal{A}) \quad \text{and} \quad u(\rho, \sigma) \geq u(\rho, \sigma') \; \forall \; \sigma' \in D(\mathcal{B}).$$

Beyond multi-agent interaction, common-interest games also describe decentralized optimization of a common objective function (the utility). Quantum CIGs thus have a natural link to the Best Separable State problem (BSS), which corresponds to linear optimization over the convex hull of bipartite product states $\rho \otimes \sigma$, i.e.,

$$\max\{\mathrm{Tr}(R(\rho \otimes \sigma)) : \; \rho \in D(\mathcal{A}), \, \sigma \in D(\mathcal{B})\} \tag{BSS}$$

where R is a fixed Hermitian matrix. The BSS problem is a crucial challenge in quantum information theory, closely tied to entanglement detection [21, 22, 23, 24]. Furthermore, we prove in this paper that KKT points of a BSS problem instance correspond to Nash equilibria of its corresponding quantum CIG. Hence, learning dynamics for quantum CIGs can be used as decentralized algorithms for the BSS problem.

Motivated by the above, in this paper we introduce and study natural extensions of classical game dynamics to the quantum CIG setting. One of the most important approaches for learning in games relies on the notion of regret, which quantifies the difference between the actual payoff an agent receives and the optimal payoff they could have secured in hindsight by playing a single fixed action [25]. No-regret algorithms are algorithms for which agents accrue sublinear regret over time, ensuring that the decisions made are not significantly worse than the best fixed strategy in hindsight. A key result in the special case of two-player zero-sum games is that if players use no-regret algorithms to update their strategies, then the limit points of their time-averaged strategy profiles are approximate Nash equilibria. However, beyond the setting of zero-sum games, the limit points of the time-averaged strategy profile of players using no-regret algorithms are approximate coarse correlated equilibrium [26], which may still include strategies that no rational players will play [27, 28].

Beyond no-regret, a line of research based on *Lyapunov*-type analysis [29, 30, 31, 32] has been able to establish pointwise convergence of learning dynamics to Nash equilibria in the specific setting of common-interest games. The common utility that players receive serves as a natural choice of Lyapunov function—i.e., a function whose value increases along trajectories of the dynamics—in common-interest games because it inherently aligns all players' incentives towards a shared objective. An example of dynamics in common-interest games for which the common utility is a Lyapunov function is the *best response dynamics*, where the players make sequential unilateral updates to the pure action that maximizes the common utility [26, 30, 33]. Since at every step this common utility increases and there is only a finite number of pure strategy profiles, this process must necessarily terminate in finite time at a strategy profile from which no player can improve the common utility via unilateral deviation, which is thus a Nash equilibrium. A smoother update that maintains a distribution over actions (instead of jumping between pure actions) is the *linear multiplicative weights update*, which plays each of the pure actions with a probability proportional to a weight that is updated according to how it performed. In common-interest games, it has been shown that under the assumption that fixed points are isolated, any trajectory of the linear multiplicative weights update that is initialized in the interior converges to Nash equilibria [32].

Contributions. In this paper, we introduce and study the performance of quantum analogues of both the best response update and the linear multiplicative weights update (as well as an analogue of the continuous-time replicator dynamics) in quantum common-interest games, proving analogues of classical convergence theorems and explaining points of difference in the quantum setting and why they arise.

In Section 2, we provide the necessary background on learning in classical and quantum games, which this work builds upon and extends, as well as quantum preliminaries. In Section 3, we introduce quantum common-interest games (quantum CIGs). We demonstrate that any instance of the Best Separable State problem can be interpreted as a quantum CIG, where the Karush-Kuhn-Tucker points of the BSS instance correspond to Nash equilibria of the game.

Our first section on dynamics is Section 4, where we show that best response dynamics converge to the set of Nash equilibria in two-player quantum CIGs. Subsequently, in Section 5 we introduce our proposed (continuous-time) linear quantum replicator dynamics and (discrete-time) linear matrix multiplicative weights update and study their convergence properties in quantum CIGs. We show that Nash equilibria are fixed points, and limit points of the dynamics are also fixed points (and hence a superset of the Nash equilibria). Our continuous-time dynamic is a noncommutative generalization of the celebrated replicator dynamics [34, 35], while our discrete-time dynamic is a noncommutative extension of the linear multiplicative weights update [32]. Crucially, we find that the classical result of [32] for convergence to Nash equilibria in classical CIGs does *not* extend to the quantum setting.

Throughout the paper, we also assess the performance of our dynamics through extensive simulations. In Section 6, we also evaluate our discrete-time algorithms on the BSS problem, demonstrating that they closely approximate the optimal value.

2 Background and Related Work on Learning in Games

2.1 Classical common-interest and potential games

In a classical normal-form game, each player selects an action or, more generally, a distribution over actions. Each player then receives a payoff based on their utility function, which maps the set of pure action profiles (tuples of actions played by each player) to the real numbers. A common-interest game (CIG) is one where every player has the same utility function.

For concreteness, we focus on the two-player case, where Alice and Bob each have a finite set of pure actions ($\mathcal{A}$ and $\mathcal{B}$, respectively) and choose as their strategies probability distributions $x \in \Delta(\mathcal{A})$ and $y \in \Delta(\mathcal{B})$ to sample their pure actions from. If they select strategies (x, y), their expected utility is

$$u(x, y) = x^\top A y = \sum_{i,j} x_i y_j A_{ij},$$

where A_{ij} is the payoff if Alice selects pure (i.e., deterministic) strategy i and Bob selects pure strategy j. common-interest games model situations where players have aligned incentives and work towards a common goal, making them applicable in various cooperative scenarios.

The canonical solution concept in games is a Nash equilibrium, i.e., a strategy profile (x^*, y^*) that is stable under unilateral player deviations. This means

$$u(x^*, y^*) \geq u(x, y^*) \quad \forall\, x \in \Delta(\mathcal{A}) \quad \text{and} \quad u(x^*, y^*) \geq u(x^*, y) \quad \forall\, y \in \Delta(\mathcal{B}).$$

In classical CIGs, where both agents are maximizing the same utility function, there is a natural connection between the game (with a common payoff matrix A) and the bilinear optimization problem

$$\max\{x^\top Ay :\ x \in \Delta(\mathcal{A}),\, y \in \Delta(\mathcal{B})\},$$

over the product of simplices. Specifically, the optimization problem's KKT points correspond to Nash equilibria of the game [34].

A closely related and very expressive class of games are potential games [29], characterized by the existence of a potential function V that tracks each player's change in utility due to unilateral deviations, i.e.,

$$u_i(s, s_{-i}) - u_i(s', s_{-i}) = V(s, s_{-i}) - V(s', s_{-i}) \quad \forall\ s, s', s_{-i},$$

where u_i is the utility function of the i-th player. Potential games are one of the most studied game models and have extensive applications including Cournot competition [29] and congestion games [36] (see., e.g., [37, 38, 39, 40] for various theoretical and engineering applications). Although the definition of CIGs is more narrow, CIGs are "equivalent" to potential games in that every potential game has a corresponding CIG with the same Nash equilibria. Additionally, for any first-order dynamic—i.e., dynamics that use only first-order information—the trajectories in the potential game and its corresponding CIG are identical. This is because each player's utility can be separated into a coordination term (which is the same for all players and equal to the potential V) and a dummy term (that only depends on the other players), i.e.,

$$u_i(s) = V(s) + D_i(s_{-i}),$$

e.g., see [41]. Due to this decomposition, for each player i, the gradients of $u_i(s)$ and $V(s)$ with respect to their own strategy are equal. Consequently, the trajectories of players' strategies under first-order learning dynamics will be identical, whether they are playing the potential game or the CIG where each player receives utility V.

2.2 Learning dynamics in classical games

Learning in games takes a dynamic perspective in which agents play a game repeatedly, in either discrete or continuous time, and "learn" over time how to adjust their strategies as the system equilibrates. In this context, the players' algorithms are called the learning dynamics, and determine the next iterates x^{t+1} and y^{t+1} in discrete time, or the rate of change of each strategy, $\dot{x}$ and $\dot{y}$, in continuous time.

The field of evolutionary game theory (see, e.g., [42]), which studies how the makeup of interacting ecological populations evolve over time, lends itself naturally to this field of learning in games due to the presence of explicit evolutionary dynamics. One property that learning dynamics coming from evolutionary game theory try to capture is that if a strategy attains more than the average utility then the probability that it is played should increase, mirroring natural selection (populations that are fitter than average should proliferate). In continuous time, which is the setting of much of evolutionary game theory (studying the relative growth of populations sizes over time), this notion is perfectly captured by the *replicator dynamics*, which is perhaps the most well-studied first-order continuous-time learning dynamic. In our two-player setting, the replicator dynamics for the first player are given by the differential equation

$$\dot{x}_\ell = x_\ell((Ay)_\ell - x^\top Ay), \tag{REP}$$

where x_ℓ is the probability assigned to the ℓ-th pure action, or via the equivalent closed-form formulation

$$x_\ell = \frac{\exp(s_\ell(t))}{\sum_i \exp(s_\ell(t))}, \quad s_\ell(t) = \int_0^t (Ay(\tau))_\ell d\tau,$$

see, e.g. [35, 43, 44, 45, 46]. The dynamics for the second player can be analogously written.

There are several approaches for discrete-time learning (i.e., through rounds of repeated play) in normal-form games, and we review the ones most relevant to us. The *linear multiplicative weights update* is given by

$$x_\ell \leftarrow x_\ell \frac{1+\eta(Ay)_\ell}{\sum_\ell x_\ell(1+\eta(Ay)_\ell)}, \tag{lin-MWU}$$

where η is an adjustable stepsize in $(0,+\infty]$. The limiting case where the stepsize $\eta \to +\infty$ gives the update

$$x_\ell \leftarrow x_\ell \frac{(Ay)_\ell}{x^\top Ay}, \tag{BE}$$

which is a special case of the Baum-Eagon (BE) update [42, 47] for polynomial optimization over the product of simplices. lin-MWU and BE both adjust the probability of playing a strategy ℓ based on the strategy's performance $(Ay)_\ell$ relative to the average performance $x^\top Ay$, and can be seen as discretizations of the REP dynamics. Similar to the REP dynamics, both BE and lin-MWU have the property that they increase the weight of a strategy when it performs better than average and decrease it when it performs worse. Since lin-MWU and BE have similar properties that are relevant to us, we shall regard BE as a special case of lin-MWU with stepsize $\eta = +\infty$ and henceforth only mention lin-MWU in our discussions.

A discretization of REP that is based on its closed-form exponential formulation is the *exponential multiplicative weights update*, defined as

$$x_\ell \leftarrow x_\ell \frac{\exp(\eta(Ay)_\ell)}{\sum_\ell x_\ell \exp(\eta(Ay)_\ell)}. \tag{exp-MWU}$$

The expression of exp-MWU can be obtained by writing the exponential closed-form formulation of REP recursively for the case where the payoffs are observed at discrete times. Both lin-MWU and exp-MWU fall within the *multiplicative weights* family of learning algorithms which adjust strategies' weights according to their performance, see e.g. [32, 48, 49].

Finally, another natural algorithmic approach to learning in games is the alternating best response dynamics

$$\begin{aligned} x(t+1) &\in \arg\max_{x\in\Delta(\mathcal{A})} u(x, y(t)), \\ y(t+1) &\in \arg\max_{y\in\Delta(\mathcal{B})} u(x(t+1), y). \end{aligned} \tag{BR}$$

see e.g. [26]. In this setting, players update their strategies in an alternating fashion: at time $t+1$, the updating player selects a strategy that is a best response to the other player's strategy at time t. In a normal-form game, the updates can always be chosen to be pure strategies as the set of maximizers of the linear utility function over the simplex will always contain a vertex of the simplex.

As a first-order check that these dynamics can be used for learning Nash equilibria of a (common-interest) game, it can be easily verified that Nash equilibria are fixed points of all of the REP, lin-MWU, exp-MWU, and BR dynamics [26, 43]. However, convergence to Nash equilibria is not as simple to achieve, and often requires additional conditions. For REP, it is known (as part of the "folk theorem" of evolutionary game theory) that the limit points of trajectories initialized in the interior of the strategy space—i.e., from mixed strategies with full support—are Nash equilibria, which has the implication that any convergent interior trajectory goes to a Nash equilibrium (see e.g., [43, 46]). In addition, it has been shown that, in all except a measure-zero set of common-interest games and initializations, convergence to (pure) Nash equilibria is achieved [50, 51].

For the discrete-time dynamics on the other hand, it is known that under the assumption that fixed points are isolated, trajectories of lin-MWU that are initialized in the interior converge to Nash equilibria [32]. Alternating BR does converge in finite time to pure Nash equilibria in common-interest games (see e.g., [26, 33]), but it is worth noting that the canonical proof of this—which we have written in our introduction—relies heavily on the finiteness of the pure strategy profile space that BR stays within.

Both lin-MWU and exp-MWU are no-regret algorithms, and thus is it known that if the players use them then the limit points of the time average of the strategy profiles are approximate coarse correlated equilibria [26]. However, as stated in the introduction, the set of coarse correlated equilibria can include non-rationalizable strategies, which are strategies that are eliminated during iterative elimination of dominated strategies and thus unsatisfactory [27, 28]. Moreover, exp-MWU (specifically in the fixed stepsize regime which we consider) has been shown to exhibit chaotic, non-convergent behavior in common-interest games [32]. It is for this reason that we focus on REP, lin-MWU, and BR as the learning dynamics whose quantum analogues we want to study as learning dynamics in quantum common-interest games.

2.3 Learning dynamics in quantum games

The majority of the literature on quantum games investigates the potential advantages of using quantum strategies over classical ones. To this end, researchers have developed quantum versions of well-known games such as the Prisoner's Dilemma [12] and Matching Pennies [52]. In addition, an increasing amount of research has focused on studying quantum notions of equilibria, i.e., states that remain stable against unilateral player deviations [15], determining their tractability [14], and obtaining structural characterizations of equilibrium sets [53, 53]. Beyond the analysis of specific games, various attempts have been made to establish more general theories of quantum games that aim to unify the existing works, see e.g., [13, 54].

In contrast, there are relatively few works that investigate learning in quantum games. Most existing results focus on a specific zero-sum setting where players select density matrices ρ and σ and receive bilinear utilities $u_i(\rho, \sigma) = \mathrm{Tr}(R_i(\rho \otimes \sigma^\top))$ subject to the constraint $u_1(\rho, \sigma) + u_2(\rho, \sigma) = 0$. (The payoffs can also be expressed explicitly as bilinear functions $u_i(\rho, \sigma) = \langle \rho, \Phi_i(\sigma) \rangle$, where R_i is the Choi matrix (1) of the super-operator Φ_i.) The study of learning in quantum games has drawn much inspiration from learning in classical games. As an initial foray, [16] proved that a noncommutative version of multiplicative weights updates called the Matrix Multiplicative Weights Update (MMWU) can be used to compute approximate Nash equilibria in two-player quantum zero-sum games, and subsequently, [17] introduced algorithms for solving quantum zero-sum games that use an extra-gradient mechanism to obtain a quadratic speedup on the rate found in [16]. Most recently, [20] showed that the noncommutative variant of discrete-time no-regret learning exhibits time-average convergence to quantum coarse correlated equilibria in general quantum games. From a continuous-time perspective, [18, 19] showed that the continuous-time limit of MMWU and generalizations thereof exhibit cyclical (i.e., non-convergent) behavior in two-player zero-sum quantum games. [19] also showed that a broad class of continuous-time Follow-The-Quantum-Regularized-Leader dynamics (a generalization of the classical Follow-The-Regularized-Leader framework) exhibits constant regret in general quantum games, and proved a modified analogue of the classical evolutionary folk theorem for replicator dynamics which draws connections between Nash equilibria of the game and "stable points" of the dynamics.

We now state two of the learning dynamics that have been used for learning in quantum games, as they are analogues of some of the classical learning dynamics shown in Section 2.2 and are closely related to the quantum learning dynamics we shall introduce and study in this paper. The first is the Matrix Multiplicative Weights Update, which (for the ρ player) is given by:

$$\rho(t+1) \leftarrow \frac{\exp\left(\eta \sum_{\tau=1}^{t} \Phi(\sigma(\tau))\right)}{\mathrm{Tr}\Big(\exp\left(\eta \sum_{\tau=1}^{t} \Phi(\sigma(\tau))\right)\Big)}, \tag{MMWU}$$

and converges (in the time-average sense) to Nash equilibria in quantum zero-sum games [16]. MMWU is a generalization of exp-MWU to the quantum setting and and was first introduced for online optimization over the set of density matrices [48, 55, 56]. MMWU has found many applications: important examples include solving SDPs [57], proving that QIP=PSPACE [58], finding balanced separators [59], and spectral sparsification [60].

Recently, [18] also introduced the exponential quantum replicator dynamics (exp-QREP), a (continuous-time) quantum generalization of the exponential formulation of the replicator dynamics (REP) given by:

$$\rho = \frac{\exp(S(t))}{\mathrm{Tr}(\exp(S(t)))}, \quad S(t) = \int_0^t \Phi(\sigma(\tau))d\tau. \tag{exp-QREP}$$

Note that these dynamics are simply called the quantum replicator dynamics in [18]. The main result in [18] is that the exp-QREP dynamics exhibit a type of periodic behavior called Poincaré recurrence when applied to quantum zero-sum games. MMWU can be obtained as a discretization of the exp-QREP dynamics, in the same manner that the discrete-time exponential MWU is obtained from the exponential variant of the continuous-time replicator dynamics (REP) in the classical regime. Thus, while noncommutative generalizations of the exponential variants of the classical learning dynamics are known and been studied for learning in quantum games, the linear variants have yet to be studied. Moreover, while works such as [19, 20, 53, 53] have studied learning in general quantum games and quantum zero-sum games, quantum common-interest games remain an important class of games which have yet to be studied in the literature.

In this paper, we complete the picture of analogues (see Table 1) of the classical replicator and multiplicative update learning dynamics we showed in Section 2.2 by introducing the Linear Quantum Replicator Dynamics (lin-QREP) and the Linear Matrix Multiplicative Weights Update (lin-MMWU) and studying their convergence properties, along with those of the best response dynamics (Quantum BR), in the class of quantum common-interest games. The analysis of MMWU remains out of the scope of this paper, since its classical counterpart (exp-MWU) has been shown to exhibit chaotic behavior in classical CIGs, and its regret-minimizing property does not suffice to guarantee convergence to Nash.

	Continuous-time	Discrete-time
Linear variant	lin-QREP (this work)	lin-MMWU (this work)
Exponential variant	exp-QREP [18]	MMWU [55, 56, 61]

Table 1: Replicator variants and discretizations thereof for learning in quantum games, in analogy to the classical dynamics discussed in Section 2.2.

2.4 Quantum preliminaries

Finally, we review some quantum preliminaries as well as geometrical properties regarding the minimal face of a point in the set of density matrices, which is the strategy space of our quantum games.

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 one. The state space of a quantum register $\mathcal{H}$ is denoted by $D(\mathcal{H})$. 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., $D(\mathcal{A} \otimes \mathcal{B})$. If the two registers are independently prepared in states described by ρ and σ, then the joint state is described by the density matrix $\rho \otimes \sigma \in \mathbb{C}^{nm \times nm}$.

To interact with a quantum register, we need to measure it. One mathematical formalism of the process of measuring a quantum system is the 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 $\rho \in D(\mathcal{H})$, 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. Note that $\langle A, B \rangle$ is a real number for any Hermitian matrices A and B, and is non-negative if A and B are positive semidefinite.

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 $\mathcal{H}$. A linear operator that maps matrices to matrices, i.e., a mapping $\Phi : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$, is called a *super-operator*. The adjoint super-operator $\Phi^\dagger : \mathrm{L}(\mathcal{A}) \to \mathrm{L}(\mathcal{B})$ is uniquely determined by the equation $\langle A, \Phi(B) \rangle = \langle \Phi^\dagger(A), B \rangle$. A super-operator $\Phi : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$ is *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 $\Phi : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$ known as the *Choi-Jamiołkowski isomorphism.* Specifically, for a super-operator Φ its *Choi matrix* is:

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

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 $\Phi_R : \mathrm{L}(\mathcal{B}) \to \mathrm{L}(\mathcal{A})$ by setting $\Phi_R(E_{i,j}) = A_{i,j}$ from which it easily follows that $C_{\Phi_R} = R$. Explicitly, we have

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

where the partial trace $\mathrm{Tr}_{\mathcal{B}} : \mathcal{L}(\mathcal{A} \otimes \mathcal{B}) \to \mathcal{L}(\mathcal{A})$ 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 Φ is completely positive (i.e., $\mathbb{1}_m \otimes \Phi$ is positive for all $m \in \mathbb{N}$) iff the Choi matrix of Φ is positive semidefinite. In particular, if the Choi matrix of the super-operator Φ is PSD, it follows that Φ is positive.

Geometry of the set of density matrices. We round off this section with some properties regarding the faces of the set of density matrices. This shall be important to us as the space of strategies available to each player in a quantum game is the set of density matrices of a given dimension.

For a given quantum register $\mathcal{A}$, the *minimal face* (see, e.g., [62]) of a density matrix $\rho \in D(\mathcal{A})$, which is the smallest face of $D(\mathcal{A})$ that contains ρ, is

$$\begin{aligned} \mathrm{face}_{D(\mathcal{A})}(\rho) &\equiv \{X \succeq 0 :\ \mathrm{tr}(X) = 1,\ \mathrm{range}(X) \subseteq \mathrm{range}(\rho)\} \\ &= \{X \succeq 0 : \mathrm{tr}(X) = 1,\ U^\dagger X U \text{ is supported on the upper } r \times r \text{ submatrix}\}, \end{aligned} \tag{3}$$

where the second equality is due to the fact that, writing the spectral decomposition $\rho = \sum_{i=1}^n \lambda_i u_i u_i^\dagger$ where $S := \{i : \lambda_i > 0\} = [r]$ and $U := \begin{pmatrix} u_1 & \dots & u_n \end{pmatrix}$ is unitary, we have for any $X \succeq 0$ that

$$\begin{aligned} \mathrm{range}(X) \subseteq \mathrm{range}(\rho) = \ker(\rho)^\perp &\iff X u_i = 0\ \forall\ i > r \\ &\iff \text{If } i > r \text{ or } j > r \text{ then } [U^\dagger X U]_{ij} = u_i^\dagger X u_j = 0 \\ &\iff U^\dagger X U \text{ is supported on the upper } r \times r \text{ submatrix.} \end{aligned}$$

The relative interior of the minimal face of ρ on $D(\mathcal{A})$ is given by

$$\mathrm{relint}\ \mathrm{face}_{D(\mathcal{A})}(\rho) = \{X \succeq 0 :\ \mathrm{tr}(X) = 1,\ \mathrm{range}(X) = \mathrm{range}(\rho)\}. \tag{4}$$

Note that these are in direct analogy to the classical setting, where the minimal face of a probability distribution $x \in \Delta(\mathcal{A})$ is the set $\{x' : x'_\ell \ge 0\ \forall\ \ell,\ \sum_\ell x'_\ell = 1,\ \mathrm{supp}(x') \subseteq \mathrm{supp}(x)\}$ and its relative interior is the set $\{x' : x'_\ell \ge 0\ \forall\ \ell,\ \sum_\ell x'_\ell = 1,\ \mathrm{supp}(x') = \mathrm{supp}(x)\}$.

3 Quantum Common-Interest Games and the BSS problem

In this section, we introduce quantum common-interest games (CIGs) and establish their connection to the BSS problem by showing that the Nash equilibria of a quantum CIG and the KKT points of the corresponding BSS problem instance coincide.

3.1 Quantum games

Quantum games. In this work we study 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 utility function of the i-th player is given by the expected value of an observable R_i on the strategy profile $\bigotimes_i \rho_i$, i.e.,

$$u_i\left(\bigotimes_i \rho_i\right) = \mathrm{Tr}\left(R_i\left(\bigotimes_i \rho_i\right)\right).$$

Quantum common-interest games. A quantum CIG is a quantum game where every player has the same utility function. For simplicity, we shall in this paper consider two-player quantum CIGs, where two agents Alice and Bob control quantum registers $\mathcal{A}$ and $\mathcal{B}$ and play strategies given by density matrices in $D(\mathcal{A})$ and $D(\mathcal{B})$ respectively. Upon playing strategy profile $(\rho, \sigma) \in D(\mathcal{A}) \times D(\mathcal{B})$ both players receive a common utility $u(\rho,\sigma) = \langle R, \rho \otimes \sigma\rangle$, where R is a Hermitian matrix that we can assume without loss of generality to be positive definite. We refer to the matrix R as the *game operator*. Equivalently, using the Choi-Jamiołkowski isomorphism defined in (1), it is useful to also express the utility function as $u(\rho,\sigma) = \langle \rho, \Phi(\sigma^\top)\rangle$, since

$$\langle \rho, \Phi(\sigma^\top)\rangle = \langle \rho, \mathrm{Tr}_{\mathcal{B}}(R(\mathbb{1}_{\mathcal{A}} \otimes \sigma)\rangle = \langle \rho \otimes \mathbb{1}_{\mathcal{B}}, R(\mathbb{1}_{\mathcal{A}} \otimes \sigma)\rangle = \langle R, \rho \otimes \sigma\rangle,$$

where R is the Choi matrix of Φ. Moreover, as R is PSD it follows that Φ is positive. In order to simplify notation throughout the rest of the paper, we will drop the transpose from the utility and express it as $u(\rho,\sigma) = \langle \rho, \Phi(\sigma)\rangle$ where appropriate. This can be seen as Bob selecting $\sigma^\top$ as his strategy, instead of σ as defined before.

A quantum CIG can also be defined as the mixed extension of a game where the players' pure strategies are complex unit vectors $x \in \mathbb{S}_{\mathbb{C}}^{n-1}, y \in \mathbb{S}_{\mathbb{C}}^{m-1}$ and the common utility is biquadratic, i.e., $u(x,y) = (x \otimes y)^\dagger R(x \otimes y)$. In particular, if the players randomize their play using finitely supported distributions $\mathcal{D}_{\mathcal{A}}, \mathcal{D}_{\mathcal{B}}$ over their pure strategy spaces, i.e., $\mathcal{D}_{\mathcal{A}}$ has support $\{x_i\}_{i=1}^k$ and $\mathrm{Prob}(x_i) = \lambda_i$ whereas $\mathcal{D}_{\mathcal{B}}$ has support $\{y_j\}_{j=1}^\ell$ and $\mathrm{Prob}(y_j) = \mu_j$, the expected payoff is bilinear in the density matrices $\rho = \sum_{i=1}^k \lambda_i x_i x_i^\dagger$ and $\sigma = \sum_{j=1}^\ell \mu_j y_j y_j^\dagger$ as

$$\mathbb{E}[(x \otimes y)^\dagger R(x \otimes y)] = \mathrm{Tr}(R(\rho \otimes \sigma)),$$

where expectation is taken over $x \sim \mathcal{D}_{\mathcal{A}}$, $y \sim \mathcal{D}_{\mathcal{B}}$.

Lastly, a classical CIG with common utility $x^\top A y$ where $x \in \Delta_n, y \in \Delta_m$ is a special case of a quantum CIG. Indeed, consider the quantum CIG with diagonal game operator $R \in \mathbb{R}^{nm \times nm}$ whose diagonal entries are $R_{ij,ij} = A_{ij}$. If we only consider diagonal densities $\rho = \sum_{i=1}^n x_i e_i e_i^\dagger$ and $\sigma = \sum_{j=1}^m y_j e_j e_j^\dagger$, it is straightforward to verify that $x^\top A y = \mathrm{Tr}(R(\rho \otimes \sigma))$.

Nash equilibria and exploitability. A strategy profile (ρ^*, σ^*) is a Nash equilibrium of the quantum CIG with common interest $u(\rho,\sigma) = \langle \rho, \Phi(\sigma)\rangle$ if both strategies are best responses to the other, i.e.,

$$\langle \rho^*, \Phi(\sigma^*)\rangle \geq \langle \rho, \Phi(\sigma^*)\rangle \ \forall\, \rho \in D(\mathcal{A}) \quad \text{and} \quad \langle \rho^*, \Phi(\sigma^*)\rangle \geq \langle \rho^*, \Phi(\sigma)\rangle \ \forall\, \sigma \in D(\mathcal{B}). \tag{NE}$$

To capture distance from the set of Nash equilibria we utilize the concept of *exploitability* (see e.g. [63]), defined as

$$\frac{1}{2}\left[\lambda_{\max}(\Phi(\sigma)) - \langle \rho, \Phi(\sigma)\rangle + \lambda_{\max}(\Phi^\dagger(\rho)) - \langle \Phi^\dagger(\rho), \sigma\rangle\right], \tag{5}$$

where $\lambda_{\max}(\Phi(\sigma))$ and $\lambda_{\max}(\Phi^\dagger(\rho))$ are the maximum eigenvalues of $\Phi(\sigma)$ and $\Phi^\dagger(\rho)$ respectively. Using the variational characterization of eigenvalues,

$$\lambda_{\max}(\Phi(\sigma)) = \max\{\langle \rho', \Phi(\sigma)\rangle : \ \rho' \in D(\mathcal{A})\},$$

the difference $\lambda_{\max}(\Phi(\sigma)) - \langle\rho, \Phi(\sigma)\rangle$ is exactly the maximum gain the ρ-player can attain by unilaterally deviating from (ρ, σ). Thus, if a profile (ρ, σ) is $\epsilon-$exploitable, then it is a $2\epsilon-$*approximate Nash equilibrium* (or simply a $2\epsilon-$Nash equilibrium), in the sense that no player can unilaterally improve their payoff by $\geq 2\epsilon$.

3.2 Relation between quantum CIGs and the BSS problem

In a quantum CIG, Alice and Bob try to jointly maximize their common utility function $u(\rho, \sigma) = \langle\rho, \Phi(\sigma)\rangle$. Analogous to the classical case, there is a strong connection between the Nash equilibria of the game and the underlying BSS optimization problem. Namely, the Karush-Kuhn Tucker (KKT) points of the BSS problem are precisely the Nash equilibria of a corresponding two player quantum CIG.

Theorem 3.1. *The Nash equilibria of a two-player quantum common-interest game with common utility function $u(\rho, \sigma) = \mathrm{Tr}(R(\rho \otimes \sigma))$ correspond to the KKT points of BSS.*

Proof. We shall prove instead that the Nash equilibria of the (transposed) two-player quantum common-interest game with common utility function $u(\rho, \sigma) = \langle\rho, \Phi(\sigma)\rangle$ correspond to the KKT points of the transposed BSS problem

$$\max\{\langle\rho, \Phi(\sigma)\rangle : \rho \in D(\mathcal{A}), \sigma \in D(\mathcal{B})\}. \qquad \text{(transposed-BSS)}$$

This correspondence is equivalent to the original correspondence we want to prove since (ρ, σ) is a Nash equilibrium of the quantum CIG with common utility $\mathrm{Tr}(R(\rho \otimes \sigma))$ iff $(\rho, \sigma^\top)$ is a Nash equilibrium of the quantum CIG with common utility $\langle\rho, \Phi(\sigma)\rangle$, and similarly (ρ, σ) is a KKT point of the BSS problem iff $(\rho, \sigma^\top)$ is a KKT point of the transposed-BSS problem.

Note that $\rho \in D(\mathcal{A})$ if and only if $\langle\rho, \mathbb{1}_\mathcal{A}\rangle = \mathrm{Tr}\ \rho = 1$ and $\rho \succeq 0$, and similarly $\sigma \in D(\mathcal{B})$ if and only if $\langle\sigma, \mathbb{1}_\mathcal{B}\rangle = \mathrm{Tr}\ \sigma = 1$ and $\sigma \succeq 0$. The Lagrangian for the transposed-BSS problem is given by

$$L = \langle\rho, \Phi(\sigma)\rangle + \lambda(1 - \langle\rho, \mathbb{1}_\mathcal{A}\rangle) + \mu(1 - \langle\sigma, \mathbb{1}_\mathcal{B}\rangle) - \langle\Lambda, \rho\rangle - \langle M, \sigma\rangle\,,$$

where the dual variables satisfy $\Lambda \preceq 0$, $M \preceq 0$. Thus, the KKT conditions for the transposed-BSS problem are

$$\nabla L = 0 : \begin{cases} \dfrac{\partial L}{\partial \rho} = \Phi(\sigma) - \lambda\,\mathbb{1}_\mathcal{A} - \Lambda = 0, & \text{(6a)} \\ \dfrac{\partial L}{\partial \sigma} = \Phi^\dagger(\rho) - \mu\,\mathbb{1}_\mathcal{B} - M = 0; & \text{(6b)} \end{cases}$$

$$\text{primal feasibility:} \begin{cases} \rho \in D(\mathcal{A}), & \text{(7a)} \\ \sigma \in D(\mathcal{B}); & \text{(7b)} \end{cases}$$

$$\text{dual feasibility:} \begin{cases} \Lambda \preceq 0, & \text{(8a)} \\ M \preceq 0; & \text{(8b)} \end{cases}$$

$$\text{complementary slackness:} \begin{cases} \langle\Lambda, \rho\rangle = 0, & \text{(9a)} \\ \langle M, \sigma\rangle = 0. & \text{(9b)} \end{cases}$$

Suppose that $(\rho, \sigma, \lambda, \mu, \Lambda, M)$ is a KKT point of the transposed-BSS problem. We show that (ρ, σ) is a Nash equilibrium. Using (6a), for any density matrix $\rho' \in D(\mathcal{A})$ we get that

$$\langle\rho', \Phi(\sigma)\rangle = \lambda\,\langle\rho', \mathbb{1}_\mathcal{A}\rangle + \langle\rho', \Lambda\rangle = \lambda + \langle\rho', \Lambda\rangle \leq \lambda,$$

where the inequality follows since $\rho \succeq 0$ and $\Lambda \preceq 0$ (recall (8a)). On the other hand, if we take the inner product of (6a) with ρ instead we have that

$$\langle\rho, \Phi(\sigma)\rangle = \lambda\,\langle\rho, \mathbb{1}_\mathcal{A}\rangle + \langle\rho, \Lambda\rangle = \lambda\,\langle\rho, I\rangle = \lambda,$$

where we used the complementary slackness condition $\langle \rho, \Lambda \rangle = 0$ (9a) . Summarizing, we have that $\langle \rho', \Phi(\sigma) \rangle \leq \lambda = \langle \rho, \Phi(\sigma) \rangle \ \forall \, \rho' \in D(\mathcal{A})$, i.e., ρ is a best response to σ. Similarly we get that σ is a best response to ρ, and thus that (ρ, σ) is a Nash equilibrium of the corresponding quantum CIG.

Next, suppose that $(\rho, \sigma) \in D(\mathcal{A}) \times D(\mathcal{B})$ is a Nash equilibrium of the quantum CIG, and consider the point $(\rho, \sigma, \lambda, \mu, \Lambda, M)$ defined by

$$\begin{aligned} \lambda = \langle \rho, \Phi(\sigma) \rangle \,, &\quad \mu = \langle \rho, \Phi(\sigma) \rangle \,, \\ \Lambda = \Phi(\sigma) - \lambda \, \mathbb{1}_{\mathcal{A}}, &\quad M = \Phi^{\dagger}(\rho) - \mu \, \mathbb{1}_{\mathcal{B}} \,. \end{aligned} \tag{10}$$

The primal feasibility constraints (7a) and (7b) are satisfied by construction. Furthermore, (6a) is immediately satisfied by the definition of Λ and λ, and similarly (6b) is satisfied by the definition of M and μ. The complementary slackness condition (9a) holds since

$$\langle \rho, \Lambda \rangle = \langle \rho, \Phi(\sigma) \rangle - \lambda \, \langle \rho, \mathbb{1}_{\mathcal{A}} \rangle = \langle \rho, \Phi(\sigma) \rangle - \lambda = 0,$$

and similarly (9b) is also satisfied. Finally, since $\rho \in \mathrm{BR}_{\mathcal{A}}(\sigma)$ we have that $\forall \, v \in \mathbb{C}^m$ with $\|v\|_2 = 1$

$$v^{\dagger} \Phi(\sigma) v = \left\langle v v^{\dagger}, \Phi(\sigma) \right\rangle \leq \langle \rho, \Phi(\sigma) \rangle = \lambda,$$

which in turn implies that $\Lambda \preceq 0$ as

$$v^{\dagger} \Lambda v = v^{\dagger} \Phi(\sigma) v - \lambda \|v\|_2^2 = v^{\dagger} \Phi(\sigma) v - \lambda \leq 0.$$

Thus (8a) is satisfied and a similar argument shows that (8b) is also satisfied. □

For a classical game, if (x, y) is a Nash equilibrium, every pure strategy that is played by Alice with positive probability is a best response to y, i.e., for each i with $x_i > 0$ we have $(Ay)_i = x^T A y$, and similarly for Bob. We shall prove the analogous statement for quantum CIGs using the notion of the minimal face.

Theorem 3.2. *Let (ρ, σ) be a Nash equilibrium of a two-player quantum CIG with common utility $u(\rho, \sigma) = \langle \rho, \Phi(\sigma) \rangle$. Then any $\rho' \in D(\mathcal{A})$ satisfying $\mathrm{range}(\rho') \subseteq \mathrm{range}(\rho)$ is a best response to σ, and similarly any $\sigma' \in D(\mathcal{B})$ satisfying $\mathrm{range}(\sigma') \subseteq \mathrm{range}(\sigma)$ is a best response to ρ. In particular, in the case that $\rho \succ 0$ we have*

$$\rho' \in \mathrm{BR}_{\mathcal{A}}(\sigma) \quad \textit{for all } \rho' \in D(\mathcal{A}),$$

and symmetrically for the σ player.

Proof. By Theorem 3.1 there exist $\lambda, \mu \in \mathbb{R}$ and Hermitian matrices $\Lambda \in \mathrm{L}(\mathcal{A})$, $M \in \mathrm{L}(\mathcal{B})$ such that $(\rho, \sigma, \lambda, \mu, \Lambda, M)$ satisfy the KKT conditions (6a)-(9b). Since $\rho \succeq 0$ and $\Lambda \preceq 0$, the complementary slackness condition (9a) (i.e., $\langle \Lambda, \rho \rangle = 0$) implies that $\mathrm{range}(\Lambda) \subseteq \ker(\rho)$. On the other hand, we have that $\Phi(\sigma) = \lambda \, \mathbb{1}_{\mathcal{A}} - \Lambda$ from the KKT condition (6a), so for any $\rho' \in D(\mathcal{A})$ satisfying $\mathrm{range}(\rho') \subseteq \mathrm{range}(\rho) \subseteq \ker(\Lambda)$ we have that $\langle \rho', \Phi(\sigma) \rangle = \lambda - \langle \rho', \Lambda \rangle = \lambda$, i.e., all such ρ' do as well as ρ against σ. Finally since ρ is a best response to σ, all such ρ' are best responses to σ.

Finally, consider the special case where $\rho \succ 0$ (the case $\sigma \succ 0$ is similar). Since $\langle \Lambda, \rho \rangle = 0$ it follows that $\Lambda = 0$. Thus, the KKT conditions imply that $\Phi(\sigma) = \langle \rho, \Phi(\sigma) \rangle \, \mathbb{1}_{\mathcal{A}}$ and consequenbtly $\langle \rho', \Phi(\sigma) \rangle = \lambda$ for all $\rho' \in \mathrm{range}(\rho) = D(\mathcal{A})$. Since ρ is one such possible ρ', we have that $\lambda = \langle \rho, \Phi(\sigma) \rangle$ and hence $\Phi(\sigma) = \langle \rho, \Phi(\sigma) \rangle \, \mathbb{1}_{\mathcal{A}}$. The arguments for the other player are completely symmetric. □

4 Best Response Dynamics

In classical potential games, the alternating best response (BR) dynamics are known to converge to pure equilibria (see e.g. [26, 33]). Generally, better response dynamics (having players alternately choose pure strategies that improves their utility) converge to pure Nash equilibria because there are a finite number of pure strategy profiles in a classical (finite) potential game, and at each timestep the potential increases, so this process must necessarily terminate. This means that, in finite time, the players reach a fixed point at which every player is already playing the best response to the others and does not need to make an update under the dynamics, which is exactly a Nash equilibrium. However, this termination condition does not occur in finite time in our setting, as there are an infinite number of pure actions (albeit in a finite-dimensional space).

Nevertheless, it turns out that in our setting alternating BR dynamics can also be defined, and indeed shown to converge to the set of Nash equilibria. Specifically, in the best response dynamics, players compute and play the strategy which maximizes their utility, given what the other player has played prior, i.e.:

$$\begin{aligned} \rho^{new} &\in \arg\max_{\rho' \in D(\mathcal{A})} \langle \rho', \Phi(\sigma) \rangle, \\ \sigma^{new} &\in \arg\max_{\sigma' \in D(\mathcal{B})} \langle \rho^{new}, \Phi(\sigma') \rangle. \end{aligned} \tag{Quantum BR}$$

We have the following result for the alternating Quantum BR dynamics in two-player quantum CIGs:

Theorem 4.1. *Alternating Quantum BR converges to the set of Nash equilibria in two-player quantum CIGs. Furthermore, for any $\epsilon > 0$, an ϵ-approximate Nash equilibrium can be found in $O(\frac{1}{\epsilon})$ iterations.*

Proof. At timestep t, say after Bob's turn, Bob has just played his best response so the exploitability is only due to Alice:

$$\text{exploitability}_t = \frac{1}{2} \left(\max_{\rho'} \langle \rho', \Phi(\sigma_t) \rangle - \langle \rho_t, \Phi(\sigma_t) \rangle \right).$$

But at timestep $t+1$, Alice has just played her best response $\rho_{t+1} \in \arg\max_{\rho'} \langle \rho', \Phi(\sigma_t) \rangle$, so

$$u_{t+1} - u_t = \langle \rho_{t+1}, \Phi(\sigma_t) \rangle - \langle \rho_t, \Phi(\sigma_t) \rangle = \max_{\rho'} \langle \rho', \Phi(\sigma_t) \rangle - \langle \rho_t, \Phi(\sigma_t) \rangle = 2 \times \text{exploitability}_t, \tag{11}$$

i.e., from timestep t to timestep $t+1$ the utility increases by twice the exploitability at time t.

Define $u_{\max} := \max_{\rho,\sigma} u(\rho, \sigma) < \infty$. The sequence $(u_t)_t$ is non-decreasing and upper bounded by $u_{\max}$, so the sequence converges, i.e., $u_t \to u_\infty$ for some $u_\infty \leq u_{\max} < \infty$. Thus,

$$2 \times \text{exploitability}_t = u_{t+1} - u_t \leq u_\infty - u_t \to 0 \qquad \text{as } t \to +\infty.$$

Furthermore, to get to an ϵ-approximate Nash equilibrium for a given $\epsilon > 0$, due to (11) we can terminate at the first timestep T at which the utility improves by $< \epsilon$, i.e., $u_{T+1} - u_T < \epsilon$. By the pigeonhole principle, we will reach such a T in at most $\frac{u_{\max} - u_0}{\epsilon}$ timesteps. □

We note that this proof of linear-time convergence, unlike the remaining proofs in this paper, is specific to the two-player setting that we focus on and not easily extendable to a higher number of players. The fact that the alternating Quantum BR dynamics converge in our setting (Theorem 4.1) recovers the well-known result that alternating BR dynamics converges to pure equilibria in classical potential games [26, 33]. Specifically, players using the BR algorithm in finite potential games are guaranteed to terminate at a pure strategy Nash equilibrium. Despite this, finding a pure Nash equilibrium in classical N-player finite CIGs (more broadly, N-player potential games) is known to be PLS-complete (see e.g [33, 64]), even when the best response can be computed in polynomial time. Comparatively, we are able to guarantee linear-time convergence to an ϵ-approximate equilibrium in the two-player quantum CIG setting. Crucially, this setting does not possess a finite number of actions (in a finite-dimensional space). We leave the derivation of a similar PLS-complexity result in the vein of [64] for N-player quantum CIGs to future work.

Exploitability experiments. In Figure 1 we plot the exploitability (as defined in Equation 5) under the alternating Quantum BR dynamics in a number of randomly-generated two-player quantum CIGs, and see that exploitability goes quickly to 0 (meaning that the limit points are Nash equilibria/KKT points), corroborating the convergence result of Theorem 4.1.

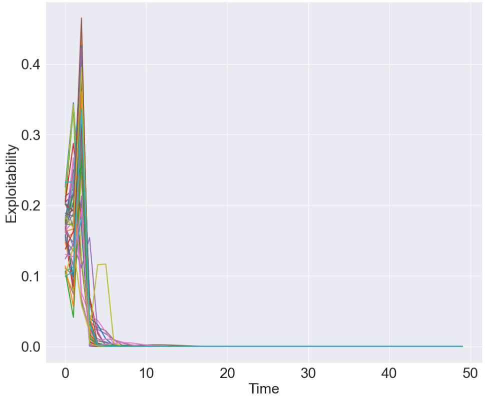


Figure 1: Exploitability of trajectories under Quantum BR. All exploitabilities go to zero quickly.

5 Linear Quantum Replicator Dynamics and Linear Matrix Multiplicative Weights Update

In this section, we introduce and study dynamics which are closely related to the classical replicator and multiplicative weights update. Consider a quantum common-interest game with utility $u(\rho, \sigma) = \langle \rho, \Phi(\sigma) \rangle$, where the Choi matrix R corresponding to the superoperator Φ is positive definite. We define the *Linear Quantum Replicator Dynamics* as:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \rho^{1/2}\Big[\Phi(\sigma) - \langle \rho, \Phi(\sigma) \rangle \mathbb{1}_{\mathcal{A}}\Big]\rho^{1/2}, \quad \text{and} \quad \frac{\mathrm{d}\sigma}{\mathrm{d}t} = \sigma^{1/2}\Big[\Phi^{\dagger}(\rho) - \langle \rho, \Phi(\sigma) \rangle \mathbb{1}_{\mathcal{B}}\Big]\sigma^{1/2}. \qquad \text{(lin-QREP)}$$

We derive the lin-QREP dynamics—and indeed, the larger family of *Linear Quantum q-Replicator Dynamics*—as a gradient flow of the common utility with respect to the quantum-Shahshahani metric. We defer this derivation to Appendix A. We show that the set of density matrices is forward-invariant under lin-QREP in Appendix B.

We also introduce a discretization of lin-QREP called the *Linear Matrix Multiplicative Weights Update*, which includes in its definition a fixed step-size $\eta \in (0, +\infty]$:

$$\rho^{new} \leftarrow \frac{\rho^{1/2}[\mathbb{1}_{\mathcal{A}} + \eta\Phi(\sigma)]\rho^{1/2}}{1 + \eta \langle \rho, \Phi(\sigma) \rangle}, \quad \text{and} \quad \sigma^{new} \leftarrow \frac{\sigma^{1/2}[\mathbb{1}_{\mathcal{B}} + \eta\Phi^{\dagger}(\rho^{new})]\sigma^{1/2}}{1 + \eta \langle \rho^{new}, \Phi(\sigma) \rangle}. \qquad \text{(lin-MMWU)}$$

Note that the updates described in lin-MMWU are performed in an alternating manner. Moreover, both lin-QREP and lin-MMWU utilize only first-order information; to perform the update, each agent only needs to know the gradient of the utility with respect to their own state.

We define lin-MMWU with infinite step-size to be the limit of the update as $\eta \to +\infty$, i.e.,

$$\rho^{new} \leftarrow \frac{\rho^{1/2}\Phi(\sigma)\rho^{1/2}}{\langle \rho, \Phi(\sigma) \rangle} \quad \text{and} \quad \sigma^{new} \leftarrow \frac{\sigma^{1/2}\Phi^{\dagger}(\rho^{new})\sigma^{1/2}}{\langle \rho^{new}, \Phi(\sigma) \rangle}, \qquad \text{(Matrix BE)}$$

which we call the *Matrix Baum-Eagon* update.

The use of a fixed step size in lin-MMWU enables finer control over the convergence process, allowing for adjustments to the rate at which updates are made. In the interest of notational brevity, we shall discuss and state results for lin-MMWU explicitly, though these results hold for Matrix BE as well. We first make the remark that these updates are all non-commutative generalizations of their classical counterparts which were discussed in Section 2.2.

Remark 5.1 (Non-commutative extensions)**.** *The lin-QREP dynamics are a non-commutative generalization of the celebrated replicator dynamics [34, 45]: specifically, the lin-QREP dynamics reduce to the usual replicator dynamics when applied to the quantum embedding of a classical game. Indeed, consider the following quantum embedding of a classical two-player CIG with common utility $x^\top Ay$: let $\{e_i\}$ and $\{f_j\}$ be orthonormal bases for $\mathcal{A}$ and $\mathcal{B}$ respectively, define the diagonal game operator $R_{ij,ij} = A_{ij}$, and consider diagonal density matrices $\rho = \sum_\ell x_\ell e_\ell e_\ell^\dagger$ and $\sigma = \sum_k y_k f_k f_k^\dagger$. Then, we have that*

$$\begin{aligned}\Phi(\sigma) &= \mathrm{Tr}_{\mathcal{B}}[R(\mathbb{1}_{\mathcal{A}} \otimes \sigma^\top)] \\ &= \mathrm{Tr}_{\mathcal{B}}\left[\left(\sum_{ij} R_{ij,ij}(e_i \otimes f_j)(e_i \otimes f_j)^\dagger\right) \cdot \left(\mathbb{1}_{\mathcal{A}} \otimes \sum_k y_k f_k f_k^\dagger\right)\right] \\ &= \sum_{ij} R_{ij,ij} y_j e_i e_i^\dagger \\ &= \sum_{ij} A_{ij} y_i e_j e_j^\dagger = \mathrm{diag}\,(Ay)\,,\end{aligned}$$

and similarly $\Phi^\dagger(\rho) = \mathrm{diag}\left(A^\top x\right)$. Consequently, if $\rho = \sum_\ell x_\ell e_k e_k^\dagger, \sigma = \sum_k y_k f_k f_k^\dagger$ are diagonal densities, we have that $\frac{\mathrm{d}\rho}{\mathrm{d}t}$ and $\frac{\mathrm{d}\sigma}{\mathrm{d}t}$ are diagonal. Specifically, the time-evolution of x is

$$\begin{aligned}\frac{\mathrm{d}x_i}{\mathrm{d}t} = \frac{\mathrm{d}\rho_{ii}}{\mathrm{d}t} &= \rho_{ii}^{1/2}\Big[\Phi(\sigma)_{ii} - \langle \rho, \Phi(\sigma)\rangle\Big]\rho_{ii}^{1/2} \\ &= x_i\left[(Ay)_i - x^\top Ay\right]\end{aligned} \tag{12}$$

and similarly $\frac{\mathrm{d}y_j}{\mathrm{d}t} = y_j\left[(A^\top x)_j - x^\top Ay\right]$, which correspond to replicator dynamics. Similarly, lin-MMWU can be understood as a non-commutative extension of lin-MWU in the sense that lin-MMWU reduces to lin-MWU when the game operator is diagonal and ρ, σ are diagonal densities. Indeed, define the diagonal game operator $R_{ij,ij} = A_{ij}$, and consider diagonal density matrices $\rho = \sum_\ell x_\ell e_\ell e_\ell^\dagger$ and $\sigma = \sum_k y_k f_k f_k^\dagger$. By (12) we have $\Phi(\sigma) = \mathrm{diag}(Ay)$. Thus

$$\rho^{new} = \frac{\rho^{1/2}[\mathbb{1}_{\mathcal{A}} + \eta\Phi(\sigma)]\rho^{1/2}}{1 + \eta\,\langle\rho, \Phi(\sigma)\rangle} = \frac{\mathrm{diag}(x \circ (1 + \eta(Ay))}{1 + \eta(x^\top Ay)}$$

where $\circ$ denotes the componentwise vector product.

We shall now provide some general properties of these dynamics in Section 5.1, prove our main theoretical results regarding their performance as learning dynamics in quantum CIGs in Section 5.2, and discuss some empirical simulations showcasing this performance in Section 5.3.

5.1 General properties of lin-QREP and lin-MMWU

In this section we prove some general properties of lin-QREP and lin-MMWU as learning dynamics in an arbitrary quantum game. Namely, we prove that the minimal face of each player's strategy is invariant along trajectories (Theorem 5.1), that lin-QREP and lin-MMWU share the same set of fixed points (Theorem 5.2) and that these fixed points can be characterized as strategies for which all strategies in their support perform equally well (Theorem 5.3). Finally, we show a partial analogue of the classical evolutionary property that the correlation with strategies that perform better than average increases under the update (Theorem 5.4).

The first property we show is that the faces (see (3)) of the set of density matrices are forward-invariant under either the continuous-time lin-QREP dynamics or the discrete-time lin-MMWU. This is an analogue of the classical property that the support of the strategy is invariant under REP and lin-MWU, and has consequences on the geometry of the dynamics' fixed points in quantum games.

Theorem 5.1. *In a quantum game, under the continuous-time dynamic lin-QREP and its discretization lin-MMWU (for any fixed step-size $\eta \in (0, +\infty]$), the minimal face of each player's* *strategy is invariant along trajectories of the dynamics. As a result, all rank-one density matrices are fixed points of the dynamics.*

Proof. Consider the lin-MMWU dynamics in a quantum game. As ρ is positive semidefinite we have that $\ker(\rho) = \ker(\rho^{1/2})$. Moreover, as Φ maps positive semidefinite matrices to positive semidefinite matrices we have that $\mathbb{1} + \eta\Phi(\sigma)$ is positive definite for all σ and $\eta > 0$. Thus, it follows that $\text{range}(\rho^{new}) = \text{range}(\rho^{1/2}(\mathbb{1} + \eta\Phi(\sigma))\rho^{1/2}) = \text{range}(\rho)$, which by (3) implies that the minimal face of the strategy ρ is invariant under the update. This holds for all players, and the proof is concluded.

A similar argument holds in the case of lin-QREP. □

We next demonstrate that lin-MMWU is a faithful discretization of lin-QREP in the sense that they share the same set of fixed points:

Theorem 5.2. *The discrete-time updates lin-MMWU for any fixed step-size $\eta \in (0, +\infty]$ and the continuous-time gradient flow lin-QREP have the same set of fixed points.*

Proof. A density matrix ρ that is stationary under the lin-QREP flow satisfies:

$$\rho^{1/2}[\Phi(\sigma) - \langle \rho, \Phi(\sigma)\rangle \mathbb{1}_{\mathcal{A}}]\rho^{1/2} = 0 \iff \rho^{1/2}\Phi(\sigma)\rho^{1/2} = \langle \rho, \Phi(\sigma)\rangle \rho.$$

On the other hand, a density matrix ρ that is stationary under a lin-MMWU update satisfies:

$$\rho = \frac{\rho^{1/2}[\mathbb{1}_{\mathcal{A}} + \eta\Phi(\sigma)]\rho^{1/2}}{1 + \eta\langle \rho, \Phi(\sigma)\rangle} \iff (1+\eta\langle \rho, \Phi(\sigma)\rangle)\rho = \rho^{1/2}[\mathbb{1}_{\mathcal{A}} + \eta\Phi(\sigma)]\rho^{1/2} \iff \langle \rho, \Phi(\sigma)\rangle \rho = \rho^{1/2}\Phi(\sigma)\rho^{1/2}.$$

Thus, a strategy is stationary under lin-QREP if and only if it is stationary under a lin-MMWU update, and hence the two dynamics have the same fixed points. □

Having shown that the fixed points of lin-QREP and lin-MMWU coincide, the following theorem then characterizes these fixed points, stating the analogue of the classical result for REP and lin-MWU that a strategy is a fixed point if and only if all the pure actions in its support perform equally well:

Theorem 5.3. *Under lin-QREP or lin-MMWU, a strategy is a fixed point if and only if all strategies whose support is included in its support (i.e., all density matrices that lie in the same face of the set of density matrices as it) do equally well. Concretely, we have for lin-QREP that*

$$\dot{\rho} = 0 \iff \langle \rho', \Phi(\sigma)\rangle = \langle \rho, \Phi(\sigma)\rangle \ \forall \ \rho' \in \text{face}_{D(\mathcal{A})}(\rho)$$

and for lin-MMWU that

$$\rho^{new} = \rho \iff \langle \rho', \Phi(\sigma)\rangle = \langle \rho, \Phi(\sigma)\rangle \ \forall \ \rho' \in \text{face}_{D(\mathcal{A})}(\rho).$$

Proof. As we have shown in Theorem 5.2 that the fixed points of the two dynamics coincide, we need only prove this result for lin-QREP. Let $\rho = \sum_{i=1}^{n} \lambda_i u_i u_i^\dagger$ where $\lambda_i > 0 \ \forall \ i \in [r]$ and $\lambda_i = 0 \ \forall \ i > r$, and $\Lambda := \text{Diag}(\lambda_1, \dots, \lambda_n)$. We first have the following characterization of a strategy ρ being a fixed point (i.e., $\dot{\rho} = 0$), where $[A]_{[r]\times[r]}$ denotes the $[r] \times [r]$ submatrix of a matrix A:

$$\begin{aligned}\dot{\rho} = 0 &\iff \rho^{1/2}\Phi(\sigma)\rho^{1/2} = \langle \rho, \Phi(\sigma)\rangle \rho \\ &\iff U^\dagger\rho^{1/2}UU^\dagger\Phi(\sigma)UU^\dagger\rho^{1/2}U = \langle \rho, \Phi(\sigma)\rangle U^\dagger\rho U \\ &\iff \Lambda^{1/2}U^\dagger\Phi(\sigma)U\Lambda^{1/2} = \langle \rho, \Phi(\sigma)\rangle \Lambda \\ &\iff [\Lambda]^{1/2}_{[r]\times[r]}[U^\dagger\Phi(\sigma)U]_{[r]\times[r]}[\Lambda]^{1/2}_{[r]\times[r]} = \langle \rho, \Phi(\sigma)\rangle [\Lambda]_{[r]\times[r]} \\ &\iff [U^\dagger\Phi(\sigma)U]_{[r]\times[r]} = \langle \rho, \Phi(\sigma)\rangle \mathbb{1}_r \,.\end{aligned} \tag{13}$$

Note that the second-to-last equivalence is due to the fact that $\Lambda^{1/2}$ is 0 in all entries outside of the $[r] \times [r]$ submatrix, so only the $[r] \times [r]$ submatrix of $U^\dagger\Phi(\sigma)U$ affects the matrix product $\Lambda^{1/2}U^\dagger\Phi(\sigma)U\Lambda^{1/2}$.

We are now ready to prove the theorem.

($\Rightarrow$) $\forall\ \rho' \in \text{face}(\rho)$ we have that

$$\langle \rho', \Phi(\sigma)\rangle = \left\langle U^\dagger \rho' U, U^\dagger \Phi(\sigma) U\right\rangle = \left\langle [U^\dagger \rho' U]_{[r]\times[r]}, [U^\dagger \Phi(\sigma) U]_{[r]\times[r]}\right\rangle = \langle \rho, \Phi(\sigma)\rangle\,,$$

where the second equality is due to (3) and the last equality is due to (13).

($\Leftarrow$) That $\langle \rho', \Phi(\sigma)\rangle = \langle \rho, \Phi(\sigma)\rangle\ \forall\ \rho' : \text{range}(\rho') \leq \text{range}(\rho)$ is equivalent to the statement

$$\left\langle U^\dagger \rho' U, U^\dagger \Phi(\sigma) U\right\rangle = \langle \rho, \Phi(\sigma)\rangle\ \forall\ \rho' : \text{range}(\rho') \leq \text{range}(\rho).$$

Letting $\rho' = u_j u_j^\dagger$ for some $j \in [r]$, we have that

$$\langle \rho, \Phi(\sigma)\rangle = \left\langle E_{jj}, U^\dagger \Phi(\sigma) U\right\rangle = [U^\dagger \Phi(\sigma) U]_{jj} \quad \forall\ j \in [r]. \tag{14}$$

Then, letting $\rho' = \frac{1}{2}(u_j + u_k)(u_j + u_k)^\dagger$ for some $j \neq k \in [r]$, we have that

$$\begin{aligned}\langle \rho, \Phi(\sigma)\rangle &= \left\langle U^\dagger \rho U, U^\dagger \Phi U\right\rangle \\ &= \left\langle \frac{1}{2}(e_j + e_k)(e_j + e_k)^\dagger, U^\dagger \Phi(\sigma) U\right\rangle \\ &= \frac{1}{2}\left([U^\dagger \Phi(\sigma) U]_{jj} + [U^\dagger \Phi(\sigma) U]_{jk} + [U^\dagger \Phi(\sigma) U]_{kj} + [U^\dagger \Phi(\sigma) U]_{kk}\right) \quad \forall\ j \neq k \in [r].\end{aligned}$$

Since (14) implies that $\langle \rho, \Phi(\sigma)\rangle = \frac{1}{2}\left([U^\dagger \Phi(\sigma) U]_{jj} + [U^\dagger \Phi(\sigma) U]_{kk}\right)$, we then have that

$$\text{Re}([U^\dagger \Phi(\sigma) U]_{jk}) = \frac{1}{2}\left([U^\dagger \Phi(\sigma) U]_{jk} + [U^\dagger \Phi(\sigma) U]_{kj}\right) = 0 \quad \forall\ j \neq k \in [r].$$

On the other hand, letting $\rho' = \frac{1}{2}(u_j + iu_k)(u_j + iu_k)^\dagger = \frac{1}{2}\left(u_j u_j^\dagger - iu_j u_k^\dagger + iu_k u_j^\dagger + u_k u_k^\dagger\right)$ gives

$$\begin{aligned}\langle \rho, \Phi(\sigma)\rangle &= \left\langle U^\dagger \rho U, U^\dagger \Phi U\right\rangle \\ &= \left\langle \frac{1}{2}\left(e_j e_j^\dagger - ie_j e_k^\dagger + ie_k e_j^\dagger + e_k e_k^\dagger\right), U^\dagger \Phi(\sigma) U\right\rangle \\ &= \frac{1}{2}\left([U^\dagger \Phi(\sigma) U]_{jj} - i[U^\dagger \Phi(\sigma) U]_{jk} + i[U^\dagger \Phi(\sigma) U]_{kj} + [U^\dagger \Phi(\sigma) U]_{kk}\right) \quad \forall\ j \neq k \in [r].\end{aligned}$$

Again, since (14) implies that $\langle \rho, \Phi(\sigma)\rangle = \frac{1}{2}\left([U^\dagger \Phi(\sigma) U]_{jj} + [U^\dagger \Phi(\sigma) U]_{kk}\right)$, we have that

$$\text{Im}([U^\dagger \Phi(\sigma) U]_{jk}) = -\frac{i}{2}\left([U^\dagger \Phi(\sigma) U]_{jk} - [U^\dagger \Phi(\sigma) U]_{kj}\right) = 0 \quad \forall\ j \neq k \in [r],$$

and putting this together with the previous result that the real part is also 0 gives

$$[U^\dagger \Phi(\sigma) U]_{jk} = 0 \quad \forall\ j \neq k \in [r].$$

Combining this with (14) gives us that $[U^\dagger \Phi(\sigma) U]_{[r]\times[r]} = \langle \rho, \Phi(\sigma)\rangle\, \mathbb{1}$, which by (13) implies that $\dot{\rho} = 0$. □

Finally, we recall a fundamental property of the classical REP and lin-MWU: for any strategy in the support, the probability of being played increases if and only if it performs better than average. This principle is rooted in evolutionary game theory, where strategies outperforming the average are more likely to proliferate, mirroring natural selection. More formally, and in discrete time for concreteness, if player i plays strategy k with positive probability $p_{ik}^{(t)} > 0$ at time step t, then we have that

$$u_i(k, p_{-i}^{(t)}) > u_i(p^{(t)}) \iff p_{ik}^{(t+1)} > p_{ik}^{(t)}. \tag{15}$$

The following theorem is a partial quantum analogue of this property for lin-QREP and lin-MMWU when looking at pure strategies in the state's eigenbasis:

Theorem 5.4. *Both lin-QREP and lin-MMWU increase the correlation of ρ with positive eigendirections of ρ that perform better than ρ (and similarly for σ). Concretely, we have that:*

For all unit eigendirections u_i of the density matrix ρ with positive eigenvalue $\lambda_i > 0$,

$$\langle \Phi(\sigma), u_i u_i^\dagger \rangle > \langle \rho, \Phi(\sigma) \rangle \iff \left\langle \dot{\rho}, u_i u_i^\dagger \right\rangle > 0 \tag{16}$$

under lin-QREP and

$$\langle \Phi(\sigma), u_i u_i^\dagger \rangle > \langle \rho, \Phi(\sigma) \rangle \iff \langle \rho^{new}, u_i u_i^\dagger \rangle > \langle \rho, u_i u_i^\dagger \rangle \tag{17}$$

under lin-MMWU with any stepsize $\eta \in (0, +\infty]$.

Proof. We first note that $\lambda_i \left\langle \Phi(\sigma), u_i u_i^\dagger \right\rangle = \left\langle \Phi(\sigma), \rho^{1/2} u_i u_i^\dagger \rho^{1/2} \right\rangle$ and $\lambda_i \langle \rho, \Phi(\sigma) \rangle = \langle \rho, \Phi(\sigma) \rangle \left\langle \rho, u_i u_i^\dagger \right\rangle$, so

$$\begin{aligned} \left\langle \Phi(\sigma), u_i u_i^\dagger \right\rangle > \langle \rho, \Phi(\sigma) \rangle &\iff \left\langle \Phi(\sigma), \rho^{1/2} u_i u_i^\dagger \rho^{1/2} \right\rangle > \langle \rho, \Phi(\sigma) \rangle \left\langle \rho, u_i u_i^\dagger \right\rangle \\ &\iff \left\langle \rho^{1/2} \Phi(\sigma) \rho^{1/2}, u_i u_i^\dagger \right\rangle > \langle \rho, \Phi(\sigma) \rangle \left\langle \rho, u_i u_i^\dagger \right\rangle . \end{aligned} \tag{18}$$

For lin-QREP, we recall that $\dot{\rho} = \rho^{1/2} \left[\Phi(\sigma) - \langle \rho, \Phi(\sigma) \rangle \mathbb{1}_{\mathcal{A}} \right] \rho^{1/2}$ so

$$\left\langle \dot{\rho}, u_i u_i^\dagger \right\rangle = \left\langle \rho^{1/2} \Phi(\sigma) \rho^{1/2}, u_i u_i^\dagger \right\rangle - \langle \rho, \Phi(\sigma) \rangle \left\langle \rho, u_i u_i^\dagger \right\rangle ,$$

which together with (18) gives (16). On the other hand, recall that lin-MMWU with stepsize $\eta > 0$ is given by

$$\rho^{new} = \frac{1}{1 + \eta \langle \rho, \Phi(\sigma) \rangle} \rho^{1/2} (\mathbb{1}_{\mathcal{A}} + \eta \Phi(\sigma)) \rho^{1/2},$$

so

$$\begin{aligned} \left\langle \rho^{new}, u_i u_i^\dagger \right\rangle > \left\langle \rho, u_i u_i^\dagger \right\rangle &\iff \left\langle \rho^{1/2} (\mathbb{1}_{\mathcal{A}} + \eta \Phi(\sigma)) \rho^{1/2}, u_i u_i^\dagger \right\rangle > (1 + \eta \langle \rho, \Phi(\sigma) \rangle) \left\langle \rho, u_i u_i^\dagger \right\rangle \\ &\iff \left\langle \rho, u_i u_i^\dagger \right\rangle + \eta \left\langle \rho^{1/2} \Phi(\sigma) \rho^{1/2}, u_i u_i^\dagger \right\rangle > \left\langle \rho, u_i u_i^\dagger \right\rangle + \eta \langle \rho, \Phi(\sigma) \rangle \left\langle \rho, u_i u_i^\dagger \right\rangle \\ &\iff \left\langle \rho^{1/2} \Phi(\sigma) \rho^{1/2}, u_i u_i^\dagger \right\rangle > \langle \rho, \Phi(\sigma) \rangle \left\langle \rho, u_i u_i^\dagger \right\rangle , \end{aligned}$$

which together with (18) gives (17). □

We note the following generalization to Theorem 5.4: suppose that ρ has spectral decomposition $\rho = \sum_i \lambda_i u_i u_i^\dagger$, and let S be the set of indices corresponding to eigendirections that perform better than average, i.e.,

$$S = \left\{ i : \lambda_i > 0,\ \left\langle u_i u_i^\dagger, \Phi(\sigma) \right\rangle > \langle \rho, \Phi(\sigma) \rangle \right\} .$$

Then for any $W = \sum_{i \in S} \mu_i u_i u_i^\dagger$ with $\mu_i \geq 0 \, \forall \, i$ we have that $\langle W, \rho^{new} \rangle > \langle W, \rho \rangle$ under lin-MMWU. More generally though, if the direction of comparison is not an eigendirection of ρ, then it is possible for it to be in the support of ρ, perform better than ρ, and yet have the update ρ^{new} under lin-MMWU be less correlated with it than ρ is. Concretely, there exist state ρ and σ, game operator Φ, and unit vector v such that $\left\langle vv^\dagger, \Phi(\sigma) \right\rangle > \langle \rho, \Phi(\sigma) \rangle$ and $\left\langle vv^\dagger, \rho \right\rangle > 0$ but $\left\langle vv^\dagger, \rho^{new} \right\rangle < \left\langle vv^\dagger, \rho \right\rangle$ under lin-MMWU with any stepsize $\eta \in (0, +\infty]$, and one such example is the following:

$$v = \begin{pmatrix} \sqrt{\frac{2}{3}} \\ \sqrt{\frac{1}{3}} \end{pmatrix}, \qquad \rho = \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}, \qquad \Phi(\sigma) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{19}$$

A simple calculation shows that $\left\langle vv^\dagger, \Phi(\sigma) \right\rangle = \frac{1}{3} > \frac{1}{4} = \langle \rho, \Phi(\sigma) \rangle$ and $\left\langle vv^\dagger, \rho \right\rangle = \frac{7}{12} > 0$. Consequently

$$\left\langle vv^\dagger, \rho^{1/2} \Phi(\sigma) \rho^{1/2} \right\rangle = \frac{1}{12} < \frac{7}{48} = \left\langle vv^\dagger, \rho \right\rangle \langle \rho, \Phi(\sigma) \rangle .$$

which means that $\left\langle vv^\dagger, \rho^{new} \right\rangle < \left\langle vv^\dagger, \rho \right\rangle$.

5.2 Learning in quantum CIGs using lin-QREP and lin-MMWU

In this section we provide theoretical results regarding the use of lin-QREP and lin-MMWU as learning dynamics for quantum common-interest games. Namely, we prove two main results: we first relate fixed points of the lin-QREP and lin-MMWU dynamics with Nash equilibria, showing that Nash equilibria are fixed points (Theorem 5.5); we then relate limit points of the dynamics' trajectories with the dynamics' fixed points, showing by a Lyapunov-type argument that limit points of dynamics are fixed points (Theorem 5.6). These two theorems together imply that limit points of the dynamics form a superset of the Nash equilibria of the game. Finally, we round the section off with a discussion on why the discrete-time update lin-MMWU fails to replicate the performance of its classical analogue lin-MWU in converging to Nash equilibria under certain assumptions.

The first main result we show for lin-QREP and lin-MMWU in quantum CIGs is that Nash equilibria are fixed points. This is the first-order check for a learning dynamic that we want to go to Nash equilibria, since it means that if we are at a Nash equilibrium, we will stay there. We can also prove a partial converse result that interior fixed points are Nash equilibria, and together these make up the first main theorem of this section, which relate Nash equilibria and fixed points of lin-QREP and lin-MMWU:

Theorem 5.5. *In a quantum CIG, the continuous-time dynamic* lin-QREP *and its discretization* lin-MMWU *(for any fixed step-size $\eta \in (0, +\infty]$) exhibit the following properties:*

(1) Nash equilibria of the CIG are fixed points of the dynamics.

(2) Interior fixed points of the dynamics are Nash equilibria of the CIG.

Proof. We utilize the results from Theorems 3.2, which characterizes Nash equilibria in quantum CIGs, and Theorem 5.3, which characterizes fixed points of lin-QREP and lin-MMWU in quantum games.

(1) If (ρ, σ) is a Nash equilibrium, we get from Theorem 3.2 that $\langle \rho', \Phi(\sigma) \rangle = \langle \rho, \Phi(\sigma) \rangle \ \forall \ \rho' \in \text{face}(\rho)$ and $\langle \rho, \Phi(\sigma') \rangle = \langle \rho, \Phi(\sigma') \rangle \ \forall \ \sigma' \in \text{face}(\sigma)$. Then by Theorem 5.3, (ρ, σ) is a fixed point.

(2) Let (ρ, σ) be an interior fixed point of the lin-QREP dynamics. As ρ is invertible and $\dot{\rho} = 0$ we immediately get that $\Phi(\sigma) = \langle \rho, \Phi(\sigma) \rangle \mathbb{1}_{\mathcal{A}}$. In turn, this implies that $\rho \in \text{BR}_{\mathcal{A}}(\sigma)$ and similarly, $\sigma \in \text{BR}_{\mathcal{B}}(\rho)$. Thus, (ρ, σ) is an interior NE. A similar argument holds for lin-MMWU. □

Next, we give the second main result of this section, which describes the convergence properties of the continuous and discrete-time dynamics introduced. Namely, we show that that they converge to the set of fixed points, which as we showed in Theorem 5.5 are a superset of the set of Nash equilibria:

Theorem 5.6. *The continuous-time dynamic* lin-QREP *and its discretization* lin-MMWU *(for any fixed step-size $\eta \in (0, +\infty]$) have the following properties:*

(1) The common utility $u(\rho, \sigma)$ is strictly increasing along trajectories, except at fixed points.

(2) The set of ω-limit points of a trajectory is a compact, connected set of fixed points of the dynamics that all attain the same utility.

Proof. **Property (1): lin-QREP.** We first focus on the continuous-time dynamic lin-QREP and show that it is a gradient flow according to a specific geometry. (See Appendix A for basic definitions of gradient flow dynamics.) Specifically, we will show that lin-QREP is a gradient flow of the scalar field $u : \mathcal{M} \to \mathbb{R}, \ (\rho, \sigma) \mapsto \langle \rho, \Phi(\sigma) \rangle$ corresponding to the common utility function with respect to the *quantum Shahshahani metric* (also known as the intrinsic Riemannian metric, see e.g. [65])

$$\langle A, B \rangle_\rho := \text{Tr}\left[\rho^{-\frac{1}{2}} A \rho^{-\frac{1}{2}} B\right], \tag{QShah}$$

defined on the product manifold $\mathcal{M} = D(\mathcal{A}) \times D(\mathcal{B})$. At any point $(\rho, \sigma) \in \mathcal{M}$, we want to find $\mathbf{grad}u(\rho, \sigma)$, defined as the unique vector $g = (g_{\mathcal{A}}, g_{\mathcal{B}}) \in T_{(\rho,\sigma)}\mathcal{M} = T_\rho D(\mathcal{A}) \times T_\sigma D(\mathcal{B})$ satisfying

$$D_{(\rho,\sigma)}u(\xi_{\mathcal{A}}, \xi_{\mathcal{B}}) = \langle (g_{\mathcal{A}}, g_{\mathcal{B}}), (\xi_{\mathcal{A}}, \xi_{\mathcal{B}})\rangle_{(\rho,\sigma)}, \quad \text{for all } \xi = (\xi_{\mathcal{A}}, \xi_{\mathcal{B}}) \in T_{(\rho,\sigma)}\mathcal{M}.$$

Expanding the above we immediately get that

$$D_{(\rho,\sigma)}u(\xi_{\mathcal{A}}, \xi_{\mathcal{B}}) = \langle g_{\mathcal{A}}, \xi_{\mathcal{A}}\rangle_\rho + \langle g_{\mathcal{B}}, \xi_{\mathcal{B}}\rangle_\sigma = \mathrm{Tr}\Big[\rho^{-\frac{1}{2}} g_{\mathcal{A}} \rho^{-\frac{1}{2}} \xi_{\mathcal{A}}\Big] + \mathrm{Tr}\Big[\sigma^{-\frac{1}{2}} g_{\mathcal{B}} \sigma^{-\frac{1}{2}} \xi_{\mathcal{B}}\Big], \tag{20}$$

while on the other hand, as the Euclidean gradient $\nabla u(\rho, \sigma) = \big(\Phi(\sigma), \Phi^\dagger(\rho)\big)$, we have that

$$D_{(\rho,\sigma)}u(\xi_{\mathcal{A}}, \xi_{\mathcal{B}}) = \langle \nabla u(\rho, \sigma), (\xi_{\mathcal{A}}, \xi_{\mathcal{B}})\rangle = \mathrm{Tr}[\Phi(\sigma)\xi_{\mathcal{A}}] + \mathrm{Tr}\big[\Phi^\dagger(\rho)\xi_{\mathcal{B}}\big]. \tag{21}$$

Equating (20) and (21), we then have that $\mathbf{grad}u(\rho, \sigma)$ is the unique element $(g_{\mathcal{A}}, g_{\mathcal{B}})$ in the product of the tangent spaces $T_\rho D(\mathcal{A}) \times T_\sigma D(\mathcal{B})$ with the following properties:

- $g_{\mathcal{A}}$ is the unique element in $T_\rho D(\mathcal{A})$ such that

$$\mathrm{Tr}\Big[\rho^{-\frac{1}{2}} g_{\mathcal{A}} \rho^{-\frac{1}{2}} \xi_{\mathcal{A}}\Big] = \mathrm{Tr}[\Phi(\sigma)\xi_{\mathcal{A}}], \quad \forall\, \xi_{\mathcal{A}} \in T_\rho D(\mathcal{A}).$$

- $g_{\mathcal{B}}$ is the unique element in $T_\sigma D(\mathcal{B})$ such that

$$\mathrm{Tr}\Big[\sigma^{-\frac{1}{2}} g_{\mathcal{B}} \sigma^{-\frac{1}{2}} \xi_{\mathcal{B}}\Big] = \mathrm{Tr}\big[\Phi^\dagger(\rho)\big]\xi_{\mathcal{B}}, \quad \forall\, \xi_{\mathcal{B}} \in T_\sigma D(\mathcal{B}).$$

A straightforward computation shows that for any constant c we have

$$\begin{aligned} \mathrm{Tr}[\Phi(\sigma)\xi_{\mathcal{A}}] &= \mathrm{Tr}[(\Phi(\sigma) - c\mathbb{1}_{\mathcal{A}})\xi_{\mathcal{A}}] \\ &= \mathrm{Tr}\left[\rho^{-\frac{1}{2}} \underbrace{\left(\rho^{\frac{1}{2}}(\Phi(\sigma) - c\mathbb{1}_{\mathcal{A}})\rho^{\frac{1}{2}}\right)}_{g_{\mathcal{A}}} \rho^{-\frac{1}{2}} \xi_{\mathcal{A}}\right] \end{aligned}$$

where for the first equality we used that all elements in the tangent space of $T_\rho D(\mathcal{A})$ have trace equal to zero (see Appendix B). Lastly, to make $g_{\mathcal{A}}$ traceless we need to select the constant c so that

$$\mathrm{Tr}\Big(\rho^{\frac{1}{2}}(\Phi(\sigma) - c\mathbb{1}_{\mathcal{A}})\rho^{\frac{1}{2}}\Big) = 0 \iff c = \frac{\mathrm{Tr}[\rho\Phi(\sigma)]}{\mathrm{Tr}[\rho]}.$$

Summarizing, we have established that

$$g_{\mathcal{A}} = \rho^{\frac{1}{2}} \left[\Phi(\sigma) - \frac{\mathrm{Tr}[\rho\Phi(\sigma)]}{\mathrm{Tr}[\rho]}\mathbb{1}_{\mathcal{A}}\right] \rho^{\frac{1}{2}}$$

and symmetrically we also get that

$$g_{\mathcal{B}} = \sigma^{\frac{1}{2}} \left[\Phi^\dagger(\rho) - \frac{\mathrm{Tr}\big[\sigma\Phi^\dagger(\rho)\big]}{\mathrm{Tr}[\sigma]}\mathbb{1}_{\mathcal{B}}\right] \sigma^{\frac{1}{2}}.$$

Thus, the gradient flow on the product manifold $D(\mathcal{A}) \times D(\mathcal{B})$ endowed with the quantum Shahshahani metric is given by

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = g_{\mathcal{A}} = \rho^{\frac{1}{2}} \left[\Phi(\sigma) - \frac{\mathrm{Tr}[\rho\Phi(\sigma)]}{\mathrm{Tr}[\rho]}\mathbb{1}_{\mathcal{A}}\right] \rho^{\frac{1}{2}}, \quad \frac{\mathrm{d}\sigma}{\mathrm{d}t} = g_{\mathcal{B}} = \sigma^{\frac{1}{2}} \left[\Phi^\dagger(\rho) - \frac{\mathrm{Tr}\big[\sigma\Phi^\dagger(\rho)\big]}{\mathrm{Tr}[\sigma]}\mathbb{1}_{\mathcal{B}}\right] \sigma^{\frac{1}{2}}.$$

Property (1)**:** lin-MMWU**.** We show that the common utility is strictly increasing under a round of sequential updates, unless the strategy profile is a fixed point. Firstly, we show that under a ρ-update of lin-MMWU, we have that

$$\langle \rho^{new}, \Phi(\sigma)\rangle \geq \langle \rho, \Phi(\sigma)\rangle\,, \tag{22}$$

with equality iff $\rho^{new} = \rho$ and similarly obtain that $\langle \rho, \Phi(\sigma^{new})\rangle \geq \langle \rho, \Phi(\sigma)\rangle$ iff $\sigma^{new} = \sigma$. Secondly, we show that

$$\langle \rho^{new}, \Phi(\sigma)\rangle = \langle \rho, \Phi(\sigma)\rangle \iff \rho^{new} = \rho, \tag{23}$$

and similarly for the σ-update. Putting these two properties together we get that $\langle \rho, \Phi(\sigma)\rangle$ is strictly increasing under lin-MMWU updates unless at a fixed point.

Substituting ρ^{new} in (22) and clearing denominators, it is equivalent to

$$\left\langle \rho^{1/2}(\mathbb{1}_{\mathcal{A}} + \eta\Phi(\sigma))\rho^{1/2}, \Phi(\sigma)\right\rangle \geq \langle \rho, \Phi(\sigma)\rangle\,(1 + \eta\,\langle \rho, \Phi(\sigma)\rangle). \tag{24}$$

Expanding the left-hand side we get

$$\langle \rho, \Phi(\sigma)\rangle + \eta \left\langle \rho^{1/2}\Phi(\sigma)\rho^{1/2}, \Phi(\sigma)\right\rangle \geq \langle \rho, \Phi(\sigma)\rangle\,(1 + \eta\,\langle \rho, \Phi(\sigma)\rangle).$$

To prove this, it suffices to show the inequality

$$\left\langle \rho^{1/2}\Phi(\sigma)\rho^{1/2}, \Phi(\sigma)\right\rangle \geq \langle \rho, \Phi(\sigma)\rangle^2\,.$$

To see this, setting $\|A\| = \sqrt{\langle A, A\rangle}$, we have that

$$\left\langle \rho^{1/2}\Phi(\sigma)\rho^{1/2}, \Phi(\sigma)\right\rangle = \left\|\rho^{1/4}\Phi(\sigma)\rho^{1/4}\right\|^2 = \left\|\rho^{1/4}\Phi(\sigma)\rho^{1/4}\right\|^2\left\|\rho^{1/2}\right\|^2 \geq \left\langle \rho^{1/4}\Phi(\sigma)\rho^{1/4}, \rho^{1/2}\right\rangle^2 = \langle \rho, \Phi(\sigma)\rangle^2\,,$$

where the inequality is due to Cauchy-Schwarz and the fact that $\left\|\rho^{1/2}\right\|^2 = \mathrm{Tr}(\rho) = 1$.

The second step is to prove (23). For this note that the Cauchy-Schwarz equality used above holds with equality iff $\rho^{1/4}\Phi(\sigma)\rho^{1/4}$ is a scaling of $\rho^{1/2}$, i.e.,

$$\langle \rho^{new}, \Phi(\sigma)\rangle = \langle \rho, \Phi(\sigma)\rangle \iff \rho^{1/4}\Phi(\sigma)\rho^{1/4} = c\rho^{1/2} \text{ for some } c \in \mathbb{R}.$$

However, note that

$$\rho^{1/4}\Phi(\sigma)\rho^{1/4} = c\rho^{1/2} \iff \rho^{1/2}\Phi(\sigma)\rho^{1/2} = c\rho.$$

Indeed, if the LHS is true we get that

$$\rho^{1/2}\Phi(\sigma)\rho^{1/2} = \rho^{1/4}(\rho^{1/4}\Phi(\sigma)\rho^{1/4})\rho^{1/4} = \rho^{1/4}(c\rho^{1/2})\rho^{1/4} = c\rho.$$

Conversely if $\rho^{1/2}\Phi(\sigma)\rho^{1/2} = c\rho$, then letting $Q := \sum_i f(\lambda_i)v_iv_i^\dagger$ where $\rho^{1/4}$ has spectral decomposition $\sum_i \lambda_i v_i v_i^\dagger$ and $f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0},\ f : x \mapsto \begin{cases} x^{-1} & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$, we have that

$$\rho^{1/4}\Phi(\sigma)\rho^{1/4} = Q\rho^{1/2}\Phi(\sigma)\rho^{1/2}Q = cQ\rho Q = c\rho^{1/2}.$$

Finally, note that the condition $\rho^{1/2}\Phi(\sigma)\rho^{1/2} = c\rho$ is equivalent to $\rho^{new} = \rho$.

Property (2): This is a direct extension of a fundamental convergence result by Losert and Akin [66] for classical games to the the compact set of density matrices (see Theorems C.1 and C.2 in the Appendix). It relies on the already established fact that the utility function is a Lyapunov function for both dynamics. □

Discussion on the non-convergence of lin-MMWU to Nash equilibria. The immediate consequence of Theorem 5.6 in tandem with Theorem 5.5 is that the interior ω-limits of any trajectory of the dynamics are Nash equilibria, but this is not sufficient to conclude that the dynamics in general converge to a Nash equilibrium. Indeed, as we shall see from our experiments in Section 5.3, while the continuous-time lin-QREP dynamics do appear to converge to Nash equilibria, there are cases where the discrete-time lin-MMWU converges to states which are not Nash equilibria. In the following discussion, we shall show why the convergence result of lin-MWU in classical common-interest games does *not* carry over to the quantum setting.

In the classical case, it was shown in [32] that, *under the assumption that fixed points are isolated, any trajectory of lin-MWU (with fixed stepsize) initialized in the interior converges to a Nash equilibrium of a classical common-interest game.* This result was proven in two steps: 1) if fixed points are isolated, then every trajectory converges, and 2) if a trajectory that begins in the interior converges, then it converges to a Nash equilibrium. The proof of Step 1 was a direct consequence of the analogous statement to our Theorem 5.6 that limit points are a compact, connected set of fixed points. However, while this assumption often holds in classical games, it is vacuous in quantum games as all rank-one density matrices are fixed points of lin-MMWU (Theorem 5.1).

This means that the best we could have hoped to achieve in our setting is the statement that *if a trajectory initialized in the interior converges, then it converges to a Nash equilibrium* (Step 2 of the proof of the classical theorem in [32]). Classically the proof of this was due to:

(1) the property that lin-MWU increases the weights of strategies that do better than average (Equation (15)), and

(2) the fact that lin-MWU preserves the support of strategies.

Together, these properties give us a result that any convergent trajectory initialized in the interior must converge to Nash equilibrium via the following proof by contradiction: if $p^{(t)}$ converges as $t \to \infty$ to a point p^* that is not a Nash equilibrium, then there exists player i and pure strategy k such that $u_i(k, p^*_{-i}) > u_i(p^*)$. By continuity of the utility function there then exists a neighborhood U containing p^* such that $u_i(k, p_{-i}) > u_i(p^*)$ for all $p \in U$, which means that after some time T for which $p^{(t)} \in U$ for all $t \geq T$, we have by (15) that $p^{(t)}_{ik} > p^{(T)}_{ik} > 0$ for all $t \geq T$, and hence $p^*_{ik} > 0$. (The fact that $p^{(T)}_{ik} > 0$ is due to the trajectory being initialized in the interior and the support-preserving property of lin-MWU.) However, this is a contradiction since p^* is a fixed point of lin-MWU (since it is the limit point of a trajectory, so by Lyapunov analysis it is a fixed point by the analogous statement to our Theorem 5.6) and hence it must be that all of the strategies in the support of p^*_i perform equally well, i.e., $u_i(k', p^*_{-i}) = u_i(p^*)$ for all k' for which $p^*_{ik} > 0$, which means, from the assumption that $u_i(k, p^*_{-i}) > u_i(p^*)$, that $p^*_{ik} = 0$.

In our setting of quantum common-interest games, we similarly get that if an interior trajectory $(\rho(t), \sigma(t)) \to (\rho^*, \sigma^*)$ but (ρ^*, σ^*) is not a Nash equilibrium, i.e., without loss of generality there exists a unit vector v satisfying $\langle vv^\dagger, \Phi(\sigma^*)\rangle > \langle \rho^*, \Phi(\sigma^*)\rangle$, then there exists a neighborhood U containing $(\rho^* \sigma^*)$ for which $\langle vv^\dagger, \Phi(\sigma)\rangle > \langle \rho, \Phi(\sigma)\rangle$ for all $(\rho, \sigma) \in U$ and a time T after which all future iterates $(\rho(t), \sigma(t)) \in U$. It also similarly holds from the fact that (ρ^*, σ^*) is a fixed point of lin-MMWU by Theorem 5.6 that $\langle vv^\dagger, \rho^*\rangle = 0$ due to the characterization that all strategies in the support of a fixed point perform equally well (Theorem 5.3).

However, the desired contradiction cannot be reached because, unlike in the classical case, the fact that $\langle vv^\dagger, \Phi(\sigma(t))\rangle > \langle \rho, \Phi(\sigma(t))\rangle$ for all $t \geq T$ is not enough to show that the quantity $\langle vv^\dagger, \Phi(\sigma(t))\rangle$ increases at each t. This is because the direction of comparison v is not necessarily an eigendirection of $\rho(t)$, and we only have the theorem that "correlation with directions that do better than average increases under lin-MMWU" for eigendirections (Theorem 5.4), but not for general directions (see the negative example in (19)). As the direction of comparison v is not guaranteed to be an eigenvector of the iterates $\rho(t)$ and, moreover, the eigenbasis of the iterates can, in general, change over time (Theorem 5.1 only says that lin-MMWU preserves the *support*, i.e., the span of the positive eigenvectors, of the strategies), the classical argument no longer holds. Indeed, as we shall see in the following experiments, lin-MMWU can converge to points that are not Nash equilibria.

5.3 Empirical simulations

Convergence of lin-QREP to the set of Nash equilibria. In order to corroborate our theoretical results, we experimentally test the convergence of lin-QREP to the set of Nash equilibria. In Figure 2 we plot the exploitability (as defined in Equation 5) of lin-QREP in 100 randomly generated $\mathcal{H}_2 \otimes \mathcal{H}_2$ quantum CIG instances with uniform initialization, where $\mathcal{H}_n$ denotes an n-level quantum system. In all runs, the exploitabilities of lin-QREP go to zero, meaning that their limit points are KKT points/Nash equilibria of the problem.

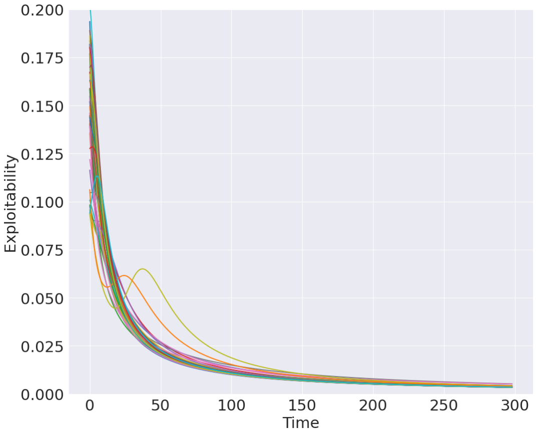


Figure 2: Exploitability of trajectories under lin-QREP. The exploitability of all trajectories goes to zero.

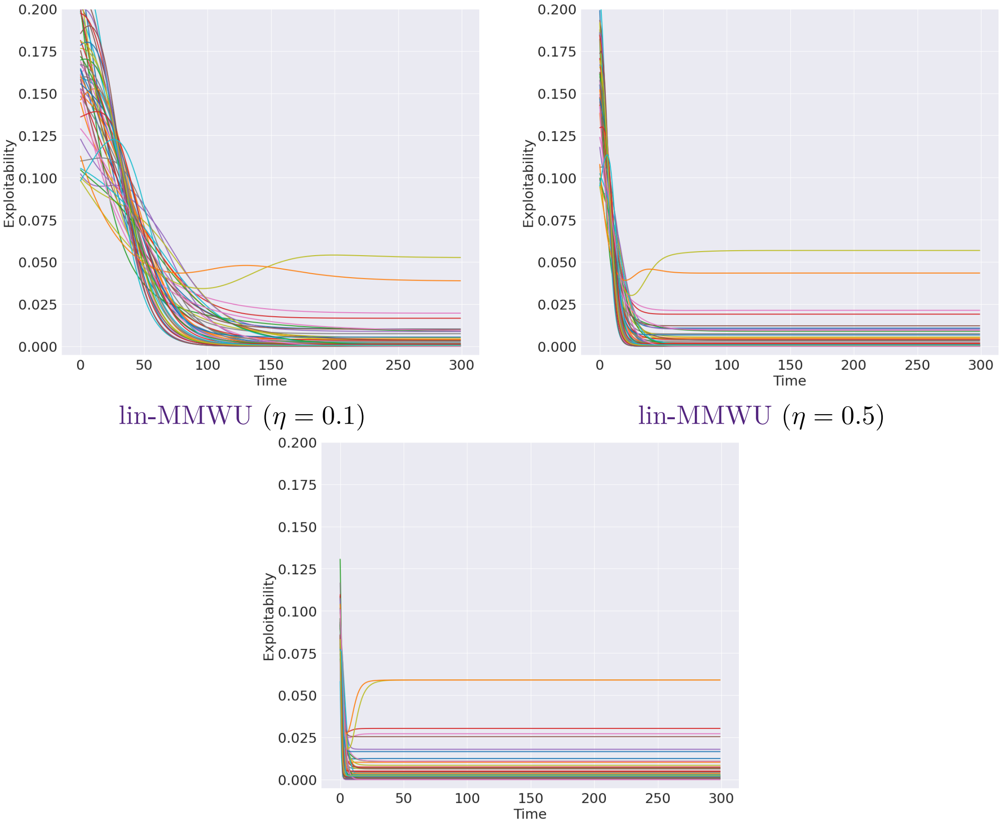
Exploitability experiments for lin-MMWU. Similar to the continuous lin-QREP dynamics, we empirically compute the exploitability of both Matrix BE and lin-MMWU. Figure 3 compares the exploitability of lin-MMWU with different stepsizes and Matrix BE, showing that some runs clearly attain a lower exploitability when using lin-MMWU. In other words, while Matrix BE is lin-MMWU with infinite stepsize η, there is a qualitative difference in the states converged to from an exploitability standpoint when using a slower/smoother update. Notice that in both lin-MMWU and Matrix BE, there exist examples of non-convergence to a pure Nash equilibrium (positive exploitability).

From the discussion at the end of Section 5.2, it stands to reason that in our setting, the trajectories which do not converge to Nash equilibria could still be converging to fixed points. Indeed, we show in Figure 4 that it is possible to find such a trajectory. We plot the Frobenius norm between the dynamics at each time step and the next iterate. Intuitively, if the log of the Frobenius norm decreases over time, the dynamics stabilize and do not exhibit any oscillating behavior. We see that these trajectories, though convergent to a point, do not have zero exploitability, meaning that the fixed point is not a Nash equilibrium.

6 Experiments for the BSS Problem

With the connection between Nash equilibria and KKT points of the BSS problem established in Section 3, we are able to empirically test our discrete-time dynamics as decentralized methods to approximately solve the BSS problem.

Recall that classically, both best response dynamics and linear multiplicative updates with constant step-size converge to Nash equilibria. However, we have shown in Section 5.2 that the classical proof

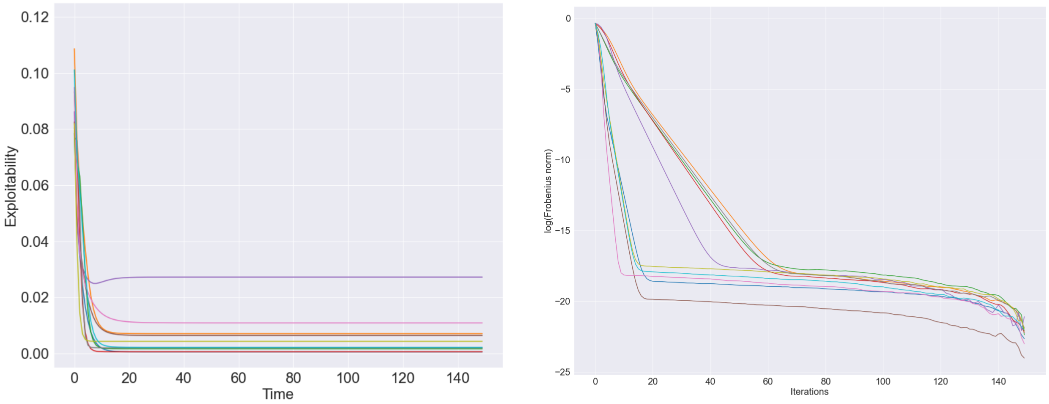


lin-MMWU ($\eta = 0.1$)

lin-MMWU ($\eta = 0.5$)

Matrix BE (lin-MMWU with $\eta = +\infty$)

Figure 3: Comparing exploitability of lin-MMWU *with different stepsizes η. Trajectories with the same color were for the same game, with uniform initialization.*



Exploitability.

log (Frobenius norm).

Figure 4: Counterexamples showing that convergence to a fixed point (low Frobenius norm) does not imply zero exploitability. All of the runs of the experiment converge to a fixed point, but several remain bounded away from zero exploitability.

technique does not carry over to lin-MMWU. Moreover, simulations in Section 5.3 experimentally confirm that lin-MMWU might fail to converge to Nash equilibria. In this section, we will compare the performance of Quantum BR to lin-MMWU variants for solving the BSS problem.

Moreover, recall that the BSS problem corresponds to linear optimization over the set of separable states, which is in general hard to compute. In order to benchmark the performance of Quantum BR

Table 2: Empirical performance of Quantum BR (uniform initialization) for the BSS problem.

Problem Dimensions	**Runs**	**Accuracy**		**Average Iterations to Convergence**
		Mean	**Std. Dev.**	
$\mathcal{H}_2 \otimes \mathcal{H}_2$	100	0.998	0.007	8.81
$\mathcal{H}_2 \otimes \mathcal{H}_3$	100	0.995	0.020	9.61

and lin-MMWU, we first identify instances of the BSS problem that can be solved to optimality. For this, we rely on the *Positive Partial Transpose* (PPT) criterion for separability [67, 68]. Specifically, any density matrix ρ describing the joint system $\mathcal{A} \otimes \mathcal{B}$ where $n = \dim \mathcal{A}$ and $m = \dim \mathcal{B}$, can be written as a block matrix

$$\rho = \begin{pmatrix} A_{11} & \dots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & \dots & A_{nn} \end{pmatrix},$$

where each block is an $m \times m$ matrix. The partial transpose of ρ with respect to $\mathcal{B}$ is the matrix obtained from ρ transposing each block A_{ij}, namely

$$\rho^{T_{\mathcal{B}}} = \begin{pmatrix} A_{11}^T & \dots & A_{1n}^T \\ \vdots & & \vdots \\ A_{n1}^T & \dots & A_{nn}^T \end{pmatrix},$$

and analogously we can also define the partial transpose of ρ with respect to $\mathcal{A}$. The PPT criterion states that a *necessary condition* for ρ to be separable is that the partial transpose $\rho^{T_{\mathcal{B}}}$ is positive semidefinite. Moreover, [68], based on previous work from [69] also show that the PPT criterion is *necessary and sufficient* when both $\mathcal{A}$ and $\mathcal{B}$ are qubit systems (i.e., $\mathcal{H}_2 \otimes \mathcal{H}_2$) or when one of them is a qubit system and the other a qutrit (i.e., $\mathcal{H}_2 \otimes \mathcal{H}_3$). Consequently, in these two regimes, the BSS problem corresponds to the following Semidefinite Program:

$$\max\{\langle R, \rho\rangle : \rho^{T_B} \succeq 0, \rho \succeq 0,\ \mathrm{Tr}(\rho) = 1\}, \tag{25}$$

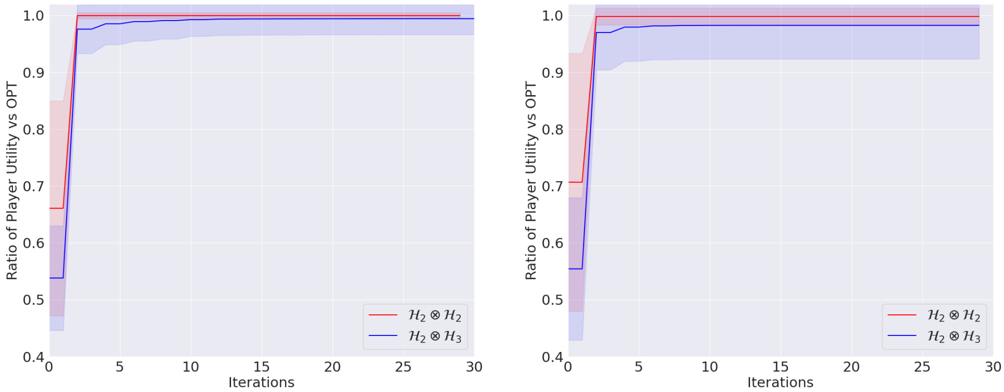
and consequently, it can be efficiently solved to optimality. Hence, we can benchmark the performance of Quantum BR and lin-MMWU in the qubit vs. qubit or qubit vs. qutrit regimes by first computing the ground truth by solving the SDP (25) which we then compare to the last iterate of the respective dynamic used.

In each run of the experiments, we randomly generate a Hermitian positive definite matrix R and standardize a uniform diagonal initialization (i.e. $\mathbb{1}_{\mathcal{H}}/n$) for the dynamic. Subsequently, we run the dynamic until convergence, which can be detected by checking the moving average (window size $= 5$) of the players' utility and we terminate the update if the moving average stabilizes for several iterations. As a benchmarking metric, we report the mean relative accuracy of the output of the dynamic compared to the optimal solution for the problem instance (25) (denoted `OPT` and computed using CVXPY [70, 71]) across 100 runs. We also report the average number of iterations needed to find a fixed point/solution, along with the standard deviation of the accuracy across the 100 runs.

In the first set of experiments, we focus on alternating Quantum BR dynamics and show that the dynamic converges very quickly to nearly optimal solutions to the BSS problem. All these results are summarized in Table 2. Figure 5 visualizes our results, and we also include a version of the experiment where the initializations for each player are random density matrices instead of uniform diagonal matrices.

In the subsequent experiment, we focus on Matrix BE, which is lin-MMWU with stepsize $\eta = +\infty$. All these results are summarized in Table 3 and visualized in Figure 6.

To explore the performance of lin-MMWU with different stepsizes, we perform the same series of experiments on lin-MMWU with a smaller stepsize of 0.9. The results of these experiments are shown in Table 4 and Figure 7.



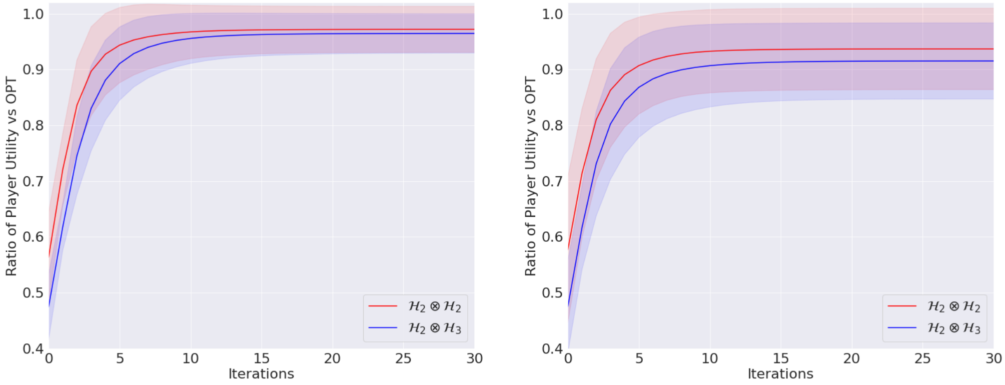
Random games, uniform initialization

Random games, random initializations

Figure 5: Ratio of the utility attained using Quantum BR vs `OPT`*, averaged over 100 random BSS problem instances. Shaded region represents ± 1 standard deviation from the mean, and each iteration represents alternating updates for ρ and σ.*

Table 3: Empirical performance of Matrix BE (lin-MMWU with $\eta = +\infty$, uniform initialization) for the BSS problem.

Problem Dimensions	Runs	Accuracy		Average Iterations to Convergence
		Mean	**Std. Dev.**	
$\mathcal{H}_2 \otimes \mathcal{H}_2$	100	0.972	0.0410	14.12
$\mathcal{H}_2 \otimes \mathcal{H}_3$	100	0.965	0.0351	16.97



Random games, uniform initialization

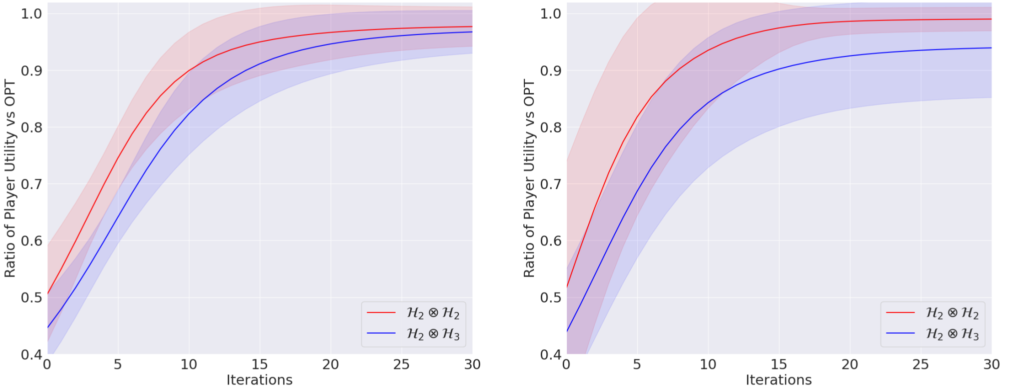
Random games, random initializations

Figure 6: Ratio of the utility attained using Matrix BE (lin-MMWU with $\eta = +\infty$) vs `OPT`*, averaged over 100 random BSS problem instances. Shaded region represents ± 1 standard deviation from the mean, and each iteration represents alternating updates for ρ and σ.*

Table 4: Empirical performance of lin-MMWU (stepsize $= 0.9$, uniform initialization) for the BSS problem.

Problem Dimensions	Runs	Accuracy		Average Iterations to Convergence
		Mean	**Std. Dev.**	
$\mathcal{H}_2 \otimes \mathcal{H}_2$	100	0.977	0.0343	29.76
$\mathcal{H}_2 \otimes \mathcal{H}_3$	100	0.972	0.0324	35.69

In our experiments, it is clear that best response dynamics perform well for the small, verifiable BSS problem instances. However, we observe that on average, the lin-MMWU dynamics also perform



Random games, uniform initialization

Random games, random initializations

Figure 7: Ratio of the utility attained using lin-MMWU vs OPT, averaged over 100 random BSS problem instances. Shaded region represents ±1 standard deviation from the mean.

competitively on average. Beyond benchmarking average case performance via SDPs, we are also interested in comparing the per-iteration performance of lin-MMWU to Matrix BE. For this, we consider the metrics of exploitability and utility attained. We performed 1000 runs of lin-MMWU and Matrix BE in small BSS instances, and found that in 93.1% of runs, lin-MMWU with stepsize $= 0.9$ obtained lower exploitability than Matrix BE, while in 99.6% of runs lin-MMWU obtained higher utility than Matrix BE. The average percentage decrease in exploitability is 36.3%, while the percentage improvement in utility is merely 0.635%. Thus, we observe empirically that lin-MMWU consistently finds less exploitable fixed points than Matrix BE for solving the BSS problem. We note that the stepsize of 0.9 is selected in order to balance performance and convergence at a reasonable rate.

7 Additional Experiments

In this section, we present a suite of additional experimental results that provide new insights into the empirical behavior of our dynamics, primarily focusing on lin-MMWU. First, we explore the potential for Matrix BE (lin-MMWU with $\eta = +\infty$) as an algorithm for biquadratic optimization, showing that it converges to rank-1 densities. To complete our experiments for Matrix BE, we provide some larger scale examples that show our convergence results in higher dimensions. Finally, we compare the lin-MMWU update with smaller stepsizes to Matrix BE, showing that they achieve comparable performance, though perturbation can be used to improve the performance of lin-MMWU.

Matrix BE as an algorithm for biquadratic optimization. Classically, replicator dynamics and their discretizations converge to pure equilibria in almost all generic common-interest games [50, 72, 73, 74]. We experimentally verify that this behavior carries over to the quantum setting. In the quantum CIG setting, the players' strategies are density matrices, which (via the SVD) correspond to distributions over rank-1 densities. Consequently, a quantum CIG can be viewed as the mixed extension of a common-interest game where players choose unit vectors and share a biquadratic utility. Our experiments suggest that when the players in a quantum CIG with game operator R use Matrix BE, their states converge to rank-1 density matrices, an intriguing analogue of the result that classical CIGs converge to pure equilibria, and an empirical confirmation of our result in Theorem 5.1 that rank-one density matrices are fixed points of the dynamics. Specifically, for a fixed, randomly generated game instance (i.e. a 4×4 Hermitian $R \succ 0$), we run Matrix BE on 100 randomly generated $\mathcal{H}_2 \otimes \mathcal{H}_2$ games with uniform initialization for both players and visualize them on the Bloch sphere (Figure 8), which is a standard technique for visualizing 2×2 density matrices and achieved using the QuTiP package [75].

In addition to the specific game instance with random initializations in Figure 8, we also explore

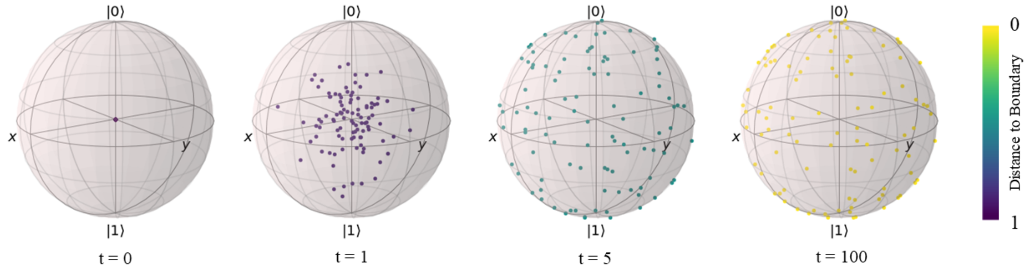


Figure 8: Matrix BE trajectories going to the boundary of the Bloch-sphere in 100 random game instances with uniform density initializations. Points are color-coded based on distance to boundary, with yellow denoting points that are close the the boundary.

lin-MMWU for a fixed game instance. In particular, we run Matrix BE (lin-MMWU with $\eta = +\infty$) on a fixed, randomly generated game with 100 randomly generated density initialization. The game generated is in $\mathcal{H}_2 \otimes \mathcal{H}_2$. We then observe a similar phenomenon as before, with lin-MMWU generating trajectories that converge to the boundary of the Bloch sphere (Figure 9) for each initialization.

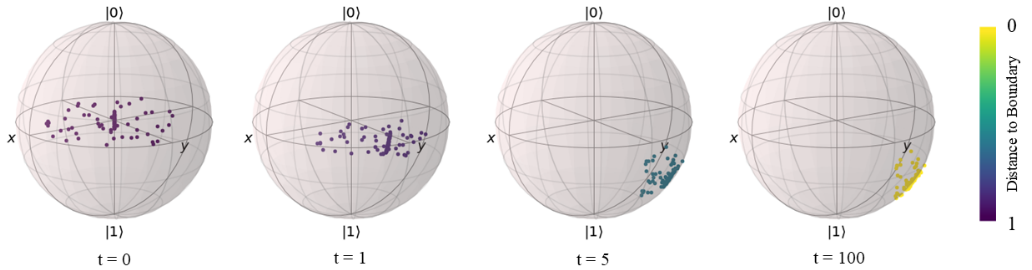


Figure 9: Matrix BE (lin-MMWU with $\eta = +\infty$) trajectories going to the boundary of the Bloch sphere in a fixed game instance with 100 randomized density initializations. Points are color-coded based on distance to boundary, with yellow denoting points that are close the the boundary.

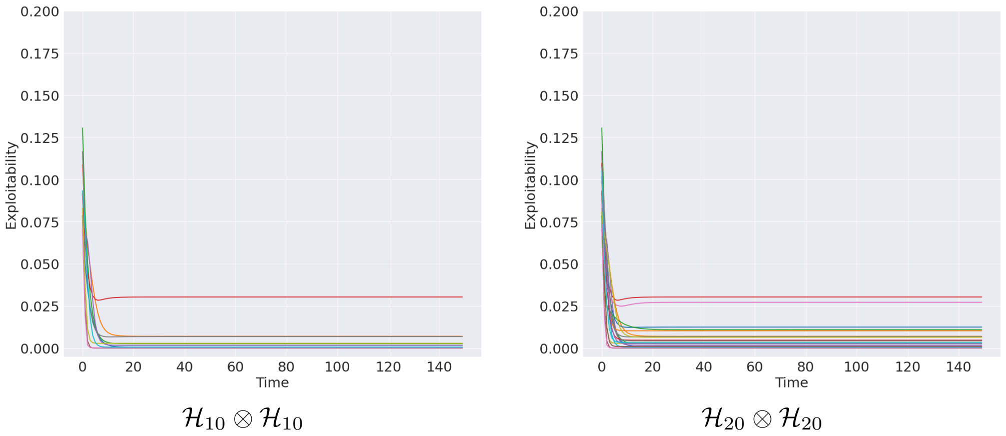
Since rank-1 densities in the quantum CIG correspond to unit vectors in the biquadratic optimization problem over the product of unit spheres $\max\{(x \otimes y)^\dagger R(x \otimes y) : \|x\|_2 = 1, \|y\|_2 = 1\}$, this means that Matrix BE can be interpreted as a learning algorithm for solving the biquadratic problem.

Large-scale experiments. Next, we present larger-scale experiments on Matrix BE and lin-MMWU which show that our results for convergence to fixed points still holds in systems of larger dimensions.

Thus far we have focused on problems of dimension $\mathcal{H}_2 \otimes \mathcal{H}_2$ and $\mathcal{H}_2 \otimes \mathcal{H}_3$ since we can efficiently compute the optimal value. In order to test the efficacy of Matrix BE for larger-scale problems, we run Matrix BE in randomly generated problems of size $\mathcal{H}_{10} \otimes \mathcal{H}_{10}$ and $\mathcal{H}_{20} \otimes \mathcal{H}_{20}$. In Figure 10 we see that the dynamics converge to fixed points, like in the smaller scale experiments.

Despite the low exploitability, it is not a guarantee that the dynamics have fully stabilized. Hence, for each run of the simulation, we additionally visualize the Frobenius norm between the dynamics at each timestep and the next iterate of the dynamics. Notice that in Figure 11, the logarithm of the Frobenius norm generally decreases steadily over time, implying the dynamics stabilize and do not exhibit any oscillating behaviour.

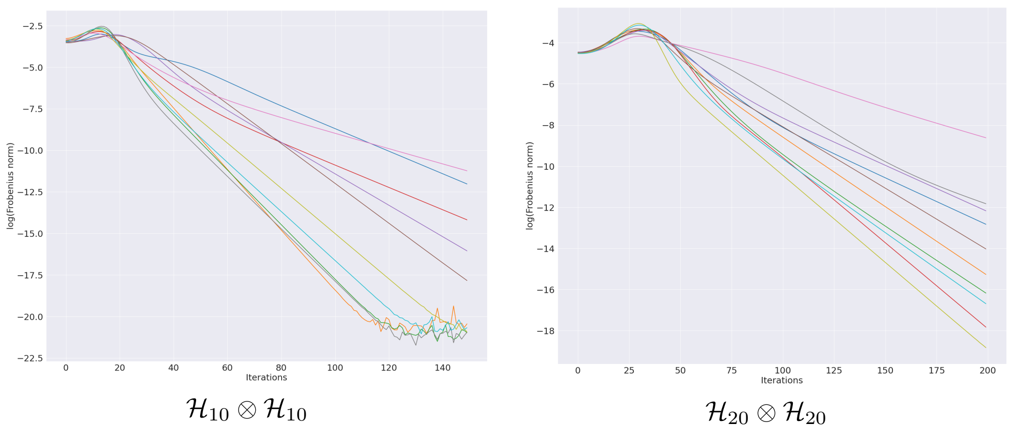
Improving empirical performance using perturbation. Notice that in Figure 3, there are two game instances that perform poorly (i.e., high exploitability) for both Matrix BE and lin-MMWU. This indicates that the dynamics converge to a sub-optimal fixed point. In order



$\mathcal{H}_{10} \otimes \mathcal{H}_{10}$

$\mathcal{H}_{20} \otimes \mathcal{H}_{20}$

Figure 10: Exploitability of Matrix BE (lin-MMWU with $\eta = +\infty$) when applied to 10 randomly generated common-interest games.



$\mathcal{H}_{10} \otimes \mathcal{H}_{10}$

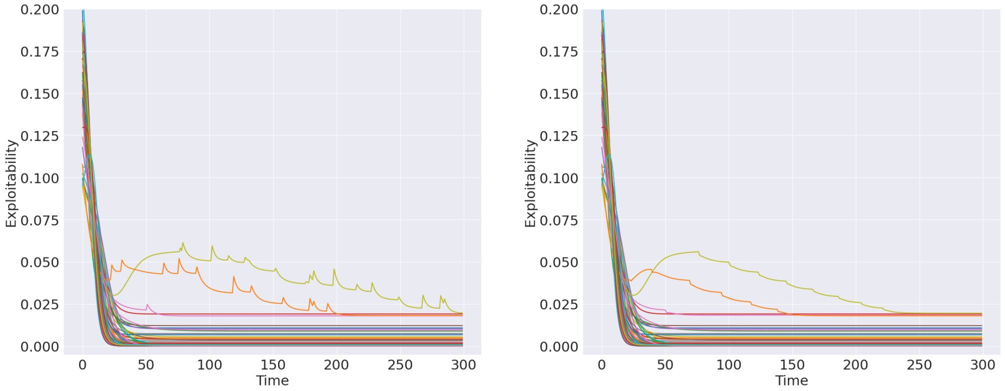
$\mathcal{H}_{20} \otimes \mathcal{H}_{20}$

Figure 11: Frobenius norm between Matrix BE (lin-MMWU with $\eta = +\infty$) at each time step and the next iterate.

to improve performance, we introduce the following modification of both algorithms: whenever a sub-optimal stationary point is reached, the players apply a random perturbation to their strategy and perform the subsequent update using this perturbed strategy. In principle, players may choose to perturb in the direction of their best response, but if the game matrix is unknown to players they can also perturb randomly, which suffices to drastically reduce exploitability (see Figure 12). We also note that while the fixed points converged to in these two game examples are entirely different after perturbation, they are still rank-1 density matrices. We provide further examples and explanation of the perturbation in Appendix D.

8 Conclusion

In this paper, we introduce the class of quantum common-interest games (CIGs), which serve as cooperative counterparts to the quantum zero-sum game formulation studied in previous works. Toward studying the behavior of learning dynamics in quantum CIGs, we introduce non-commutative extensions of continuous and discrete dynamics used for learning in classical common-interest games, completing the picture of analogues of the classical replicator, best-response and multiplicative weight update learning dynamics (see Table 1), and study their convergence properties. Along the way, we bridge game theory and optimization by establishing an equivalence between the first-order stationary points of an instance of the Best Separable State problem and the Nash equilibria of a corresponding quantum CIG, and experimentally show that our dynamics are able to converge close to the optimal solution of randomly generated BSS instances.



Random perturbation

Best-response perturbation

Figure 12: Exploitability of lin-MMWU *($\eta = 0.9$) after perturbation. Both random perturbations and best-response perturbations can help the dynamics escape points of high exploitability. However, neither kind of perturbation is sufficient to bring the exploitability to zero.*

This work opens up several exciting new research directions, the first of which is to theoretically corroborate the experimental findings for convergence of lin-QREP to Nash equilibria. Another intriguing open question is to what extent and under what conditions can the lin-MMWU dynamics be provably shown to converge to a rank-1 matrix for both players in quantum CIGs in analogy to the aforementioned classical results [28, 50, 73]. More broadly, our research contributes to the general theory for learning in quantum games by studying analogues of well-known classical dynamics in the important class of quantum common-interest games, exploring which results carry over, and understanding why some results can break down in the quantum setting. Furthering the development of learning in general quantum games would require new notions of equilibration and convergence in quantum games, and the careful exploration of analogues of other results from learning in classical games [25, 26, 76, 77].

Acknowledgments

This research is supported 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, the MOE Tier 2 Grant (MOE-T2EP20223-0018), the National Research Foundation, Singapore, under its QEP2.0 programme (NRF2021-QEP2-02-P05), the CQT++ Core Research Funding Grant (SUTD) (RS-NRCQT-00002) and partially by Project MIS 5154714 of the National Recovery and Resilience Plan, Greece 2.0, funded by the European Union under the NextGenerationEU Program. Wayne Lin and Ryann Sim acknowledge support from the SUTD President's Graduate Fellowship (SUTD-PGF). The authors also thank anonymous referees for insightful comments that greatly improved the presentation of the paper.

References

[1] Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisỳ, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. "Deepstack: Expert-level artificial intelligence in heads-up no-limit poker". Science **356**, 508–513 (2017).

[2] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. "Mastering the game of Go with deep neural networks and tree search". Nature **529**, 484–489 (2016).

[3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial nets". In Proceedings

of the 27th International Conference on Neural Information Processing Systems - Volume 2. Pages 2672–2680. NIPS'14Cambridge, MA, USA (2014). MIT Press.

[4] 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).

[5] Panayotis Mertikopoulos, Bruno Lecouat, Houssam Zenati, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. "Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile" (2018). arXiv:1807.02629.

[6] Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson, and Thore Graepel. "Cooperative AI: machines must learn to find common ground". Nature (2021).

[7] Allan Dafoe, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R McKee, Joel Z Leibo, Kate Larson, and Thore Graepel. "Open problems in cooperative AI" (2020). arXiv:2012.08630.

[8] Nolan Bard, Jakob N Foerster, Sarath Chandar, Neil Burch, Marc Lanctot, H Francis Song, Emilio Parisotto, Vincent Dumoulin, Subhodeep Moitra, Edward Hughes, et al. "The Hanabi challenge: A new frontier for AI research". Artificial Intelligence **280**, 103216 (2020).

[9] Hengyuan Hu, Adam Lerer, Alex Peysakhovich, and Jakob Foerster. ""Other-play" for zero-shot coordination". In International Conference on Machine Learning. Pages 4399–4410. PMLR (2020). arXiv:2003.02979.

[10] DJ Strouse, Kevin McKee, Matt Botvinick, Edward Hughes, and Richard Everett. "Collaborating with humans without human data". Advances in Neural Information Processing Systems **34**, 14502–14515 (2021). arXiv:2110.08176.

[11] Stefanos Leonardos, Will Overman, Ioannis Panageas, and Georgios Piliouras. "Global convergence of multi-agent policy gradient in Markov potential games" (2021). arXiv:2106.01969.

[12] Jens Eisert, Martin Wilkens, and Maciej Lewenstein. "Quantum games and quantum strategies". Physical Review Letters **83**, 3077 (1999).

[13] 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).

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

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

[16] 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).

[17] Francisca Vasconcelos, Emmanouil-Vasileios Vlatakis-Gkaragkounis, Panayotis Mertikopoulos, Georgios Piliouras, and Michael I Jordan. "A quadratic speedup in finding nash equilibria of quantum zero-sum games" (2023). arXiv:2311.10859.

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

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

[20] Wayne Lin, Georgios Piliouras, Ryann Sim, and Antonios Varvitsiotis. "No-regret learning and equilibrium computation in quantum games". Quantum **8**, 1569 (2024).

[21] Martin Grötschel, László Lovász, and Alexander Schrijver. "Geometric algorithms and combinatorial optimization". Volume 2. Springer Science & Business Media. (2012).

[22] Lawrence M Ioannou. "Computational complexity of the quantum separability problem". Quantum Information & Computation **7**, 335–370 (2007).

[23] Leonid Gurvits. "Classical deterministic complexity of Edmonds' problem and quantum entanglement". In Proceedings of the thirty-fifth annual ACM symposium on Theory of computing. Pages 10–19. (2003).

[24] S. Gharibian. "Strong NP-hardness of the quantum separability problem". Quantum Information and Computation **10**, 343–360 (2010).

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

[26] Tim Roughgarden. "Algorithmic game theory". Communications of the ACM **53**, 78–86 (2010).

[27] Yannick Viossat and Andriy Zapechelnyuk. "No-regret dynamics and fictitious play". Journal of Economic Theory **148**, 825–842 (2013).

[28] Amélie Heliou, Johanne Cohen, and Panayotis Mertikopoulos. "Learning with bandit feedback in potential games". Advances in Neural Information Processing Systems**30** (2017). url: `https://dl.acm.org/doi/abs/10.5555/3295222.3295384`.

[29] Dov Monderer and Lloyd S Shapley. "Potential games". Games and economic behavior **14**, 124–143 (1996).

[30] Brian Swenson, Ryan Murray, and Soummya Kar. "On best-response dynamics in potential games". SIAM Journal on Control and Optimization **56**, 2734–2767 (2018).

[31] Walid Krichene, Benjamin Drighès, and Alexandre M Bayen. "Online learning of Nash equilibria in congestion games". SIAM Journal on Control and Optimization **53**, 1056–1081 (2015). arXiv:1408.0017.

[32] Gerasimos Palaiopanos, Ioannis Panageas, and Georgios Piliouras. "Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos". Advances in Neural Information Processing Systems**30** (2017). arXiv:1703.01138.

[33] Eva Tardos and Tom Wexler. "Network formation games and the potential function method". Algorithmic Game TheoryPages 487–516 (2007).

[34] William H Sandholm. "Population games and evolutionary dynamics". MIT press. (2010). url: `https://mitpress.mit.edu/9780262195874/`.

[35] Peter D Taylor and Leo B Jonker. "Evolutionary stable strategies and game dynamics". Mathematical biosciences **40**, 145–156 (1978).

[36] Robert W Rosenthal. "A class of games possessing pure-strategy Nash equilibria". International Journal of Game Theory **2**, 65–67 (1973).

[37] Jason R Marden, Gürdal Arslan, and Jeff S Shamma. "Cooperative control and potential games". IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **39**, 1393–1407 (2009).

[38] Jun Zeng, Qiaoqiao Wang, Junfeng Liu, Jianlong Chen, and Haoyong Chen. "A potential game approach to distributed operational optimization for microgrid energy management with renewable energy and demand response". IEEE Transactions on Industrial Electronics **66**, 4479–4489 (2018).

[39] Qiang He, Guangming Cui, Xuyun Zhang, Feifei Chen, Shuiguang Deng, Hai Jin, Yanhui Li, and Yun Yang. "A game-theoretical approach for user allocation in edge computing environment". IEEE Transactions on Parallel and Distributed Systems **31**, 515–529 (2019).

[40] Demia Della Penda, Andrea Abrardo, Marco Moretti, and Mikael Johansson. "Potential games for subcarrier allocation in multi-cell networks with D2D communications". In 2016 IEEE International Conference on Communications (ICC). Pages 1–6. IEEE (2016).

[41] Quang Duy Lã, Yong Huat Chew, and Boon-Hee Soong. "Potential game theory: Applications in radio resource allocation". Springer. (2016).

[42] Josef Hofbauer and Karl Sigmund. "Evolutionary game dynamics". Bulletin of the American mathematical society **40**, 479–519 (2003).

[43] Josef Hofbauer, Karl Sigmund, et al. "Evolutionary games and population dynamics". Cambridge university press. (1998).

[44] Immanuel M Bomze. "Lotka-Volterra equation and replicator dynamics: a two-dimensional classification". Biological cybernetics **48**, 201–211 (1983).

[45] Jörgen W Weibull. "Evolutionary game theory". MIT press. (1997). url: `https://mitpress.mit.edu/9780262731218/`.

[46] Ross Cressman and Yi Tao. "The replicator equation and other game dynamics". Proceedings of the National Academy of Sciences **111**, 10810–10817 (2014).

[47] Leonard E Baum and John Alonzo Eagon. "An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology". Bulletin of the American Mathematical Society **73**, 360–363 (1967).

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

[49] Yoav Freund and Robert E Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting". Journal of computer and system sciences **55**, 119–139 (1997).

[50] Robert Kleinberg, Georgios Piliouras, and Éva Tardos. “Multiplicative updates outperform generic no-regret learning in congestion games”. In Proceedings of the forty-first annual ACM symposium on Theory of computing. Pages 533–542. (2009).

[51] Ioannis Panageas and Georgios Piliouras. “Average case performance of replicator dynamics in potential games via computing regions of attraction”. In Proceedings of the 2016 ACM Conference on Economics and Computation. Pages 703–720. (2016).

[52] David A Meyer. “Quantum strategies”. Physical Review Letters **82**, 1052 (1999).

[53] Constantin Ickstadt, Thorsten Theobald, and Elias Tsigaridas. “Semidefinite games”. International Journal of Game TheoryPages 1–31 (2024).

[54] Giulio Chiribella, Giacomo Mauro D’Ariano, and Paolo Perinotti. “Theoretical framework for quantum networks”. Physical Review A **80**, 022339 (2009).

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

[56] 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.

[57] 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).

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

[59] Lorenzo Orecchia, Sushant Sachdeva, and Nisheeth K Vishnoi. “Approximating the exponential, the Lanczos method and an O(m)-time spectral algorithm for balanced separator”. In Proceedings of the forty-fourth annual ACM symposium on Theory of computing. Pages 1141–1160. (2012).

[60] 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).

[61] Sanjeev Arora, Elad Hazan, and Satyen Kale. “Fast algorithms for approximate semidefinite programming using the multiplicative weights update method”. In 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05). Pages 339–348. IEEE (2005).

[62] Alexander Barvinok. “A course in convexity”. Volume 54. American Mathematical Soc. (2002).

[63] 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).

[64] Alex Fabrikant, Christos Papadimitriou, and Kunal Talwar. “The complexity of pure nash equilibria”. In Proceedings of the thirty-sixth annual ACM symposium on Theory of computing. Pages 604–612. (2004).

[65] Rajendra Bhatia. “Positive definite matrices”. Princeton University Press. (2009).

[66] V Losert and Ethen Akin. “Dynamics of games and genes: Discrete versus continuous time”. Journal of Mathematical Biology **17**, 241–251 (1983).

[67] Asher Peres. “Separability criterion for density matrices”. Physical Review Letters **77**, 1413 (1996).

[68] Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki. “On the necessary and sufficient conditions for separability of mixed quantum states”. Phys. Lett. A**223** (1996).

[69] Stanisław Lech Woronowicz. “Positive maps of low dimensional matrix algebras”. Reports on Mathematical Physics **10**, 165–183 (1976).

[70] Steven Diamond and Stephen Boyd. “CVXPY: A Python-embedded modeling language for convex optimization”. Journal of Machine Learning Research **17**, 1–5 (2016). arXiv:1603.00943.

[71] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. “A rewriting system for convex optimization problems”. Journal of Control and Decision **5**, 42–60 (2018).

[72] Panayotis Mertikopoulos and William H Sandholm. “Learning in games via reinforcement and regularization”. Mathematics of Operations Research **41**, 1297–1324 (2016).

[73] Ioannis Panageas, Georgios Piliouras, and Xiao Wang. “Multiplicative weights updates as a distributed constrained optimization algorithm: Convergence to second-order stationary points almost always”. In International Conference on Machine Learning. Pages 4961–4969. PMLR (2019). arXiv:1810.05355.

[74] Ruta Mehta, Ioannis Panageas, and Georgios Piliouras. “Natural selection as an inhibitor of genetic diversity”. In Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science. Page 73. ITCS ’15New York, NY, USA (2015). Association for Computing Machinery.

[75] J Robert Johansson, Paul D Nation, and Franco Nori. “QuTiP: An open-source python framework for the dynamics of open quantum systems”. Computer Physics Communications **183**, 1760–1772 (2012).

[76] James P. Bailey and Georgios Piliouras. “Multiplicative weights update in zero-sum games”. In Proceedings of the 2018 ACM Conference on Economics and Computation. Page 321–338. EC ’18New York, NY, USA (2018). Association for Computing Machinery.

[77] Drew Fudenberg and David K Levine. “The theory of learning in games”. Volume 2. MIT press. (1998). url: `https://mitpress.mit.edu/9780262529242`.

[78] Andre Wibisono, Ashia C Wilson, and Michael I Jordan. “A variational perspective on accelerated methods in optimization”. Proceedings of the National Academy of Sciences **113**, E7351–E7358 (2016).

[79] Arkadij Semenovič Nemirovskij and David Borisovich Yudin. “Problem complexity and method efficiency in optimization”. A Wiley-Interscience publication. Wiley. (1983).

[80] Panayotis Mertikopoulos and William H Sandholm. “Riemannian game dynamics”. Journal of Economic Theory **177**, 315–364 (2018).

[81] Siavash Shahshahani. “A new mathematical framework for the study of linkage and selection”. American Mathematical Soc. (1979).

[82] Walter Rudin. “Real and complex analysis, 3rd ed.”. McGraw-Hill, Inc. USA (1987). url: `https://dl.acm.org/doi/abs/10.5555/26851`.

[83] Ralph Tyrrell Rockafellar. “Clarke’s tangent cones and the boundaries of closed sets in $\mathbb{R}^n$”. Nonlinear Analysis: theory, methods and applications **3**, 145–154 (1979).

A Linear Quantum q-Replicator Dynamics

Gradient flow dynamics. While the implementation of learning in game theory often requires algorithms in discrete time, past work has shown that continuous-time dynamics can give rise to families of discrete dynamics. The most relevant such examples to our work are in the context of gradient-based optimization algorithms [78, 79] and evolutionary game dynamics [80, 81]. Our definition of quantum replicator dynamics (lin-QREP) can be generalized to a broader family of gradient flow dynamics which arise from the generalized family of Riemannian metrics on the PSD manifold, parametrized by $q \in \mathbb{R}$.

Consider a differentiable manifold $\mathcal{M}$ equipped with a differentiable scalar field $u : \mathcal{M} \to \mathbb{R}$ and a symmetric, positive-definite inner product $\langle \cdot, \cdot \rangle_p : T_p\mathcal{M} \times T_p\mathcal{M} \to \mathbb{R}_{\geq 0}$ defined at all $p \in \mathcal{M}$. (Here $T_p\mathcal{M}$ is the tangent space of $\mathcal{M}$ at p.) By the Riesz Representation Theorem (see, e.g., [82]), at each $p \in \mathcal{M}$ there exists a *unique* vector $\mathbf{grad}\, u(p) \in T_p\mathcal{M}$ with

$$D_p u(\xi) = \langle \mathbf{grad}\, u(p), \xi \rangle_p \quad \forall\, \xi \in T_p\mathcal{M}, \tag{26}$$

where $D_p u(\xi) : T_p\mathcal{M} \to \mathbb{R}$ is the directional derivative of u at the point p in direction ξ, i.e., $D_p u(\xi) = \langle \nabla u(p), \xi \rangle$ where $\nabla u(p)$ is the usual Euclidean gradient of u at p and $\langle \cdot, \cdot \rangle$ is the Euclidean inner product. Equation (26) allows us to associate to each point $p \in \mathcal{M}$ a vector $\mathbf{grad}\, u(p) \in T_p\, \mathcal{M}$, or in other words, to define a gradient flow on the manifold $\mathcal{M}$ given explicitly by $\dot{p} = \mathbf{grad}\, u(p)$. Moreover, simply by construction, it follows that the function $u(p)$ is nondecreasing along the trajectories of the gradient flow, i.e., $\frac{\mathrm{d}u(p)}{\mathrm{d}t} \geq 0$ since $\frac{\mathrm{d}u(p)}{\mathrm{d}t} = \langle \nabla u(p), \dot{p} \rangle = D_p u(\dot{p}) = \langle \mathbf{grad}\, u(p), \dot{p} \rangle_p = \langle \dot{p}, \dot{p} \rangle_p \geq 0$, and moreover $\frac{\mathrm{d}u(p)}{\mathrm{d}t} = 0$ if and only if $\dot{p} = 0$ (as the inner product $\langle \cdot, \cdot \rangle_p$ is positive definite), so u is in fact strictly increasing along gradient flow trajectories unless at a fixed point.

Quantum Shahshahani gradient flow. Consider a two-player quantum CIG with common utility $u(\rho, \sigma) = \langle \rho, \Phi(\sigma) \rangle$. Our goal is to provide continuous-time dynamics that improve the utility $u(\rho, \sigma)$. The state space we are operating in is the manifold $\mathcal{M} = D(\mathcal{A}) \times D(\mathcal{B})$, so all that remains is to select a metric on the manifold of density matrices which would imbue the product manifold $\mathcal{M}$ with the product metric, giving a gradient flow. To accomplish this, we consider the generalized family of Riemannian metrics on the manifold of PSD matrices parametrized by $q \in \mathbb{R}$, which we call the *quantum q-Shahshahani metric*:

$$\langle A, B \rangle_\rho^{(q)} := \mathrm{Tr}\Big[\rho^{-\frac{q}{2}} A \rho^{-\frac{q}{2}} B\Big]. \tag{QShah}$$

Indeed, in the case of diagonal matrices this family of metrics reduces to the q-Shahshahani family of metrics on the simplex (see e.g. [80]). On the PSD manifold, $q = 0$ gives the Euclidean inner product $\mathrm{Tr}[AB]$, while $q = 2$ gives the intrinsic Riemannian metric (see e.g. [65]). In addition, $q = 1$ reduces to the Shahshahani metric on the simplex in the case of diagonal matrices.

Theorem A.1 (Linear quantum q-replicator dynamics)**.** *Consider a quantum CIG with utility function $u(\rho, \sigma) = \langle \rho, \Phi(\sigma) \rangle$ where $\rho \in D(\mathcal{A}), \sigma \in D(\mathcal{B})$. The dynamics*

$$\begin{aligned} \frac{\mathrm{d}\rho}{\mathrm{d}t} &= \rho^{\frac{q}{2}} \left[\Phi(\sigma) - \frac{\mathrm{Tr}[\rho^q \Phi(\sigma)]}{\mathrm{Tr}[\rho^q]} \mathbb{1}_{\mathcal{A}} \right] \rho^{\frac{q}{2}}, \\ \frac{\mathrm{d}\sigma}{\mathrm{d}t} &= \sigma^{\frac{q}{2}} \left[\Phi^\dagger(\rho) - \frac{\mathrm{Tr}\big[\sigma^q \Phi^\dagger(\rho)\big]}{\mathrm{Tr}[\sigma^q]} \mathbb{1}_{\mathcal{B}} \right] \sigma^{\frac{q}{2}} \end{aligned} \tag{lin-QREP$_q$}$$

define a gradient flow of the utility function $u(\rho, \sigma)$ on the product manifold $D(\mathcal{A}) \times D(\mathcal{B})$ imbued with the quantum q-Shahshahani metric. Moreover, the utility $u(\rho, \sigma)$ is strictly increasing along the trajectories of the lin-QREP$_q$ *dynamics, except at fixed points.*

Proof. To derive the lin-QREP$_q$ dynamics as a gradient flow with respect to the quantum q-Shahshahani metric $\langle A, B \rangle_\rho^{(q)} := \mathrm{Tr}\big[\rho^{-\frac{q}{2}} A \rho^{-\frac{q}{2}} B\big]$ given in (QShah), we operate on the product manifold $\mathcal{M} = D(\mathcal{A}) \times D(\mathcal{B})$ endowed with the product metric and use the scalar field $u : \mathcal{M} \to \mathbb{R},\ (\rho,\sigma) \mapsto \langle \rho, \Phi(\sigma)\rangle$ corresponding to the common utility function. At any point $(\rho,\sigma) \in \mathcal{M}$, we want to find $\mathbf{grad}^{(q)}u(\rho,\sigma)$, defined as the unique vector $g = (g_{\mathcal{A}}, g_{\mathcal{B}}) \in T_{(\rho,\sigma)}\mathcal{M} = T_\rho D(\mathcal{A}) \times T_\sigma D(\mathcal{B})$ satisfying

$$D_{(\rho,\sigma)}u(\xi_{\mathcal{A}}, \xi_{\mathcal{B}}) = \langle (g_{\mathcal{A}}, g_{\mathcal{B}}), (\xi_{\mathcal{A}}, \xi_{\mathcal{B}})\rangle^{(q)}_{(\rho,\sigma)}, \quad \text{for all } \xi = (\xi_{\mathcal{A}}, \xi_{\mathcal{B}}) \in T_{(\rho,\sigma)}\mathcal{M}.$$

Expanding the above we immediately get that

$$D_{(\rho,\sigma)}u(\xi_{\mathcal{A}}, \xi_{\mathcal{B}}) = \langle g_{\mathcal{A}}, \xi_{\mathcal{A}}\rangle^{(q)}_\rho + \langle g_{\mathcal{B}}, \xi_{\mathcal{B}}\rangle^{(q)}_\sigma = \mathrm{Tr}\Big[\rho^{-\frac{q}{2}} g_{\mathcal{A}} \rho^{-\frac{q}{2}} \xi_{\mathcal{A}}\Big] + \mathrm{Tr}\Big[\sigma^{-\frac{q}{2}} g_{\mathcal{B}} \sigma^{-\frac{q}{2}} \xi_{\mathcal{B}}\Big], \tag{27}$$

while on the other hand, as the Euclidean gradient $\nabla u(\rho,\sigma) = \big(\Phi(\sigma), \Phi^\dagger(\rho)\big)$, we have that

$$D_{(\rho,\sigma)}u(\xi_{\mathcal{A}}, \xi_{\mathcal{B}}) = \langle \nabla u(\rho,\sigma), (\xi_{\mathcal{A}}, \xi_{\mathcal{B}})\rangle = \mathrm{Tr}[\Phi(\sigma)\xi_{\mathcal{A}}] + \mathrm{Tr}\big[\Phi^\dagger(\rho)\xi_{\mathcal{B}}\big]. \tag{28}$$

Equating (27) and (28), we then have that $\mathbf{grad}^{(q)}u(\rho,\sigma)$ is the unique element $(g_{\mathcal{A}}, g_{\mathcal{B}})$ in the product of the tangent spaces $T_\rho D(\mathcal{A}) \times T_\sigma D(\mathcal{B})$ with the following properties:

- $g_{\mathcal{A}}$ is the unique element in $T_\rho D(\mathcal{A})$ such that
$$\mathrm{Tr}\Big[\rho^{-\frac{q}{2}} g_{\mathcal{A}} \rho^{-\frac{q}{2}} \xi_{\mathcal{A}}\Big] = \mathrm{Tr}[\Phi(\sigma)\xi_{\mathcal{A}}] \quad \forall\, \xi_{\mathcal{A}} \in T_\rho D(\mathcal{A}).$$
- $g_{\mathcal{B}}$ is the unique element in $T_\sigma D(\mathcal{B})$ such that
$$\mathrm{Tr}\Big[\sigma^{-\frac{q}{2}} g_{\mathcal{B}} \sigma^{-\frac{q}{2}} \xi_{\mathcal{B}}\Big] = \mathrm{Tr}\big[\Phi^\dagger(\rho)\big]\xi_{\mathcal{B}} \quad \forall\, \xi_{\mathcal{B}} \in T_\sigma D(\mathcal{B}).$$

A straightforward computation shows that for any constant c we have

$$\begin{aligned}\mathrm{Tr}[\Phi(\sigma)\xi_{\mathcal{A}}] &= \mathrm{Tr}[(\Phi(\sigma) - c\mathbb{1}_{\mathcal{A}})\xi_{\mathcal{A}}] \\ &= \mathrm{Tr}\left[\rho^{-\frac{q}{2}} \underbrace{(\rho^{\frac{q}{2}}(\Phi(\sigma) - c\mathbb{1}_{\mathcal{A}})\rho^{\frac{q}{2}})}_{g_{\mathcal{A}}} \rho^{-\frac{q}{2}} \xi_{\mathcal{A}}\right]\end{aligned}$$

where for the first equality we used that tangent space of $T_\rho D(\mathcal{A})$ consists of traceless matrices. Lastly, to make $g_{\mathcal{A}}$ traceless we need to select the constant c so that

$$\mathrm{Tr}\Big(\rho^{\frac{q}{2}}(\Phi(\sigma) - c\mathbb{1}_{\mathcal{A}})\rho^{\frac{q}{2}}\Big) = 0 \iff c = \frac{\mathrm{Tr}[\rho^q \Phi(\sigma)]}{\mathrm{Tr}[\rho^q]}.$$

Summarizing, we have established that

$$g_{\mathcal{A}} = \rho^{\frac{q}{2}} \left[\Phi(\sigma) - \frac{\mathrm{Tr}[\rho^q \Phi(\sigma)]}{\mathrm{Tr}[\rho^q]} I\right] \rho^{\frac{q}{2}}$$

and symmetrically we also get that

$$g_{\mathcal{B}} = \sigma^{\frac{q}{2}} \left[\Phi^\dagger(\rho) - \frac{\mathrm{Tr}\big[\sigma^q \Phi^\dagger(\rho)\big]}{\mathrm{Tr}[\sigma^q]} I\right] \sigma^{\frac{q}{2}}.$$

Thus, the gradient flow on the product manifold $D(\mathcal{A}) \times D(\mathcal{B})$ endowed with the quantum q-Shahshahani metric is given by

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = g_{\mathcal{A}} = \rho^{\frac{q}{2}} \left[\Phi(\sigma) - \frac{\mathrm{Tr}[\rho^q \Phi(\sigma)]}{\mathrm{Tr}[\rho^q]} I\right] \rho^{\frac{q}{2}}, \quad \frac{\mathrm{d}\sigma}{\mathrm{d}t} = g_{\mathcal{B}} = \sigma^{\frac{q}{2}} \left[\Phi^\dagger(\rho) - \frac{\mathrm{Tr}\big[\sigma^q \Phi^\dagger(\rho)\big]}{\mathrm{Tr}[\sigma^q]} I\right] \sigma^{\frac{q}{2}}.$$

That the utility $u = \langle \rho, \Phi(\sigma)\rangle$ is strictly increasing unless at fixed points follows directly from the fact that this dynamic is a gradient flow. □

In terms of the convergence properties of lin-QREP$_q$ we have the following result:

Corollary A.1. *The set of ω-limit points of a trajectory $\{\rho(t), \sigma(t)\}_{t\geq 0}$ of the* lin-QREP$_q$ *dynamics is a compact, connected set of fixed points of the dynamics that all attain the same utility.*

The proof of this result follows directly from an extension of the fundamental convergence theorem by [66] to general compact sets, which we prove in Theorem C.1.

B Forward invariance of lin-QREP

In this section we show that time derivative of the state (ρ_0, σ_0) under the quantum replicator dynamics (lin-QREP) is always in the product of the tangent cones of $\rho_0 \in D(\mathcal{A})$ and $\sigma_0 \in D(\mathcal{B})$, i.e., it does not point outside of the state space. To do this we first recall the notion of the cone of feasible directions and the tangent cone:

Given a closed convex set C in a Hilbert space and a point $x \in C$, the cone of feasible directions at x is given by

$$\mathrm{dir}_C(x) = \{d : x + td \in C \text{ for some } t > 0\}$$

and the tangent cone $\mathrm{T}_C(x)$ at x (see, e.g, [83]) is given by the closure of the cone of feasible directions, i.e.,

$$\mathrm{T}_C(x) = \overline{\mathrm{dir}_C(x)}.$$

We shall characterize the tangent cones of points in the set of density matrices using the fact that the tangent cone is the polar to the normal cone. We first make the following observation characterizing the normal cones to the PSD cone:

Lemma B.1. *The normal cone to a matrix X in the PSD cone, $N_{\mathrm{PSD}}(X) \equiv \{Z : \langle Z, Y - X\rangle \leq 0\, \forall\, Y \succeq 0\}$, is equal to the set $\{Z \preceq 0 : \langle Z, X\rangle = 0\}$.*

Proof. If $Z \preceq 0$ and $\langle Z, X\rangle = 0$, then $\forall\, Y \succeq 0$ we have that $\langle Z, Y - X\rangle = \langle Z, Y\rangle - \langle Z, Y\rangle = \langle Z, X\rangle \leq 0$, so $Z \in N_{\mathrm{PSD}}(X)$.

On the other hand, if $Z \in N_{\mathrm{PSD}}(X)$ so that $\langle Z, Y - X\rangle \leq 0\, \forall\, Y \succeq 0$, then we can in particular consider the PSD matrix $Y := X + vv^\dagger$ for any vector v. Then $v^\dagger Z v = \left\langle Z, vv^\dagger\right\rangle = \langle Z, Y - X\rangle \leq 0\, \forall\, v$, so we have that $Z \preceq 0$. It then follows that $\langle Z, X\rangle \leq 0$. However, taking $Y := \frac{1}{2}X$ in the definition of the normal cone, we have that $\frac{1}{2}\langle Z, X\rangle = -\left\langle Z, -\frac{1}{2}X\right\rangle = -\langle Z, Y - X\rangle \geq 0$. Thus $\langle Z, X\rangle = 0$. □

This gives us the following characterization of the tangent cones to the PSD cone:

Lemma B.2. $\mathrm{T}_{D(\mathcal{A})}(\rho) = \{W : \mathrm{tr}(W) = 0,\ u^\dagger W u \geq 0\ \forall u \in \ker\rho\}$.

Proof. Given a closed convex set C and a point $x \in C$, the tangent cone at x is the polar of the normal cone at x, i.e. $\mathrm{T}_C(x) = N_C(x)^\circ$ where the normal cone at x is given by $N_C(x) \equiv \{z : \langle z, y - x\rangle \leq 0\ \forall y \in C\}$, and the polar of a set S is given by $S^\circ = \{w : \langle w, z\rangle \leq 0\ \forall z \in S\}$ (see, e.g., [83]) .

For a Hermitian matrix X in the PSD cone, we have from Lemma B.1 that the normal cone at X is given by

$$\begin{aligned} N_{\mathrm{PSD}}(X) =& \{Z : \langle Z, Y - X\rangle \leq 0\, \forall\, Y \succeq 0\} \\ =& \{Z \preceq 0 : \langle Z, X\rangle = 0\}, \end{aligned}$$

and so the tangent cone at X is given by

$$\mathrm{T}_{\mathrm{PSD}}(X) = N_{\mathrm{PSD}}(X)^\circ = \{W : u^\dagger W u \geq 0\ \forall u \in \ker X\}.$$

Finally, we have that

$$\begin{aligned}&\mathrm{T}_{D(\mathcal{A})}(X)\\=&\mathrm{T}_{\mathrm{PSD}}(X)\cap\mathrm{T}_{\mathrm{tr}=1}(X)\\=&\{W : u^\dagger W u \geq 0\ \forall u\in\ker X\}\cap(\mathrm{tr}=0)\\=&\{W : \mathrm{tr}\,W=0,\ u^\dagger W u\geq 0\ \forall u\in\ker X\}.\end{aligned}$$

□

With this characterization of the tangent cone in hand, we are now ready to prove the theorem:

Theorem B.1. *At any point $(\rho,\sigma)\in D(\mathcal{A})\times D(\mathcal{B})$, the time derivatives $\dot\rho=\rho^{1/2}\Big[\Phi(\sigma)-\langle\rho,\Phi(\sigma)\rangle\,\mathbb{1}_{\mathcal{A}}\Big]\rho^{1/2}$, $\dot\sigma=\sigma^{1/2}\Big[\Phi^\dagger(\rho)-\langle\rho,\Phi(\sigma)\rangle\,\mathbb{1}_{\mathcal{B}}\Big]\sigma^{1/2}$ given by lin-QREP lie in the tangent cones $\mathrm{T}_{D(\mathcal{A})}(\rho)$, $\mathrm{T}_{D(\mathcal{B})}(\sigma)$ respectively.*

Proof. Firstly, the time derivative $\dot\rho$ is traceless since

$$\begin{aligned}&\mathrm{Tr}\Big(\rho^{1/2}\Big[\Phi(\sigma)-\langle\rho,\Phi(\sigma)\rangle\,\mathbb{1}_{\mathcal{A}}\Big]\rho^{1/2}\Big)\\=&\langle\rho,\Phi(\sigma)\rangle-\mathrm{Tr}(\rho\langle\rho,\Phi(\sigma)\rangle)\\=&\langle\rho,\Phi(\sigma)\rangle-\langle\rho,\Phi(\sigma)\rangle\,\mathrm{Tr}(\rho)=0.\end{aligned}$$

Furthermore, $\forall\, u\in\ker\rho$, we have that $u\in\ker\rho^{1/2}$ and so $u^\dagger\dot\rho u=0$.

Thus $\dot\rho\in\mathrm{T}_{D(\mathcal{A})}(\rho)$ by Lemma B.2, and similarly we have that $\dot\sigma\in\mathrm{T}_{D(\mathcal{B})}(\sigma)$. □

C Auxiliary Theorems and Lemmas

The Theorems C.1 and C.2 proven in this section are direct generalizations of a fundamental convergence theorem by Losert and Akin (Proposition 1 in [66], which was written only for the simplex) to general compact sets. Nevertheless, the proof employed by Losert and Akin in [66] only really required compactness (i.e., it made use of no other properties of the simplex), and thus could actually be taken wholesale to prove Theorems C.1 and C.2. We rewrite the theorem statements and proofs here for the sake of clarity and completeness.

The notation used here is standard for dynamical systems. $x(t)$ denotes the point that x evolves to after time t has elapsed (or t iterations have passed, in the case of discrete-time dynamical systems). The limit set $\Omega(x)$[1] of an orbit $\{x(t)\}_{t\geq 0}$ (or $\{x(t)\}_{t\in\mathbb{N}}$, in the case of discrete-time systems) refers to the set of ω-limits of the orbit, i.e. the set of points ω for which there exists an increasing sequence $\{t_k\}_{k\in\mathbb{N}}$ that converges to infinity and satisfies $\lim_{k\to\infty}x(t_k)=\omega$.

Theorem C.1. *Consider a continuous-time dynamical system on a compact set $\mathcal{X}$ in a metric space, obtained through the differential equation $\dot x(t)=F(x(t))$ where F is a continuous function on $\mathcal{X}$. Suppose also that the dynamical system admits a Lyapunov function u (i.e., a function $u:\mathcal{X}\to\mathbb{R}$ such that $\forall\, x\in\mathcal{X}$, $\dot u(x)\geq 0$ with equality iff $F(x)=0$).*

Then the limit set Ω of an orbit $\{x(t)\}_{t\geq 0}$

- *is a compact connected set,*
- *consists entirely of fixed points (i.e., points x for which $F(x)=0$), and*
- *has the property that the Lyapunov function u is constant over it.*[2]

[1] The limit set $\Omega(x)$ depends on the initial condition x, but to simplify notation we shall drop the dependence on x in the notation and just write Ω when the choice of initial condition is unambiguous.

[2] i.e., $u(\omega)=u(\omega')$ for any $\omega,\omega'\in\Omega$.

Proof. u **is constant over** Ω. First note that u is continuous and $\mathcal{X}$ is compact, so $u(\mathcal{X})$ is bounded. Thus, since u is non-decreasing along orbits, $\{u(x(t))\}_{t\geq 0}$ converges to $u^* := \sup\{u(x(t))\}_{t\geq 0} < \infty$. Now consider any $\omega \in \Omega$. There exists a sequence $\{x(t_k)\}_{k\in\mathbb{N}}$ (with $t_k \to \infty$ as $k \to \infty$) that converges to ω. $\{u(x(t_k))\}_k$ also converges to $\sup\{u(x(t_k))\}_k = \sup\{u(x(t))\}_{t\geq 0} = u^*$. Thus, since u is continuous, we have that $u^* = \lim_{k\to\infty} u(x(t_k)) = u(\lim_{k\to\infty} x(t_k)) = u(\omega)$.
Thus $\forall\, \omega \in \Omega$, $u(\omega) = u^*$.

Ω **consists entirely of fixed points.** Consider any $\omega = \lim_{k\to\infty} x(t_k) \in \Omega$, where $\{t_k\}_{k\in\mathbb{N}}$ is an increasing sequence that converges to infinity. $\forall\, s \geq 0$,

$$\omega(s) = \left(\lim_{k\to\infty} x(t_k)\right)(s) = \lim_{k\to\infty} x(t_k + s) \in \Omega.$$

Thus from the already-proven fact that u obtains the same value over Ω, we have that $u(\omega(s)) = u(\omega)\ \forall\, s \geq 0$. Thus ω is a fixed point of the dynamics (by the assumption that u is a Lyapunov function, i.e., that u is strictly increasing except at fixed points).

Ω **is compact and connected.** Ω can be written as

$$\Omega = \bigcap_{t\in\mathbb{R}_{\geq 0}} \overline{\{x(s) : s \geq t\}}$$

where $\overline{\{x(s) : s \geq t\}}$, which denotes the closure of the set $\{x(s) : s \geq t\}$, is compact (since it is a closed subset of the compact set $\mathcal{X}$) and connected (since it is the closure of the image of the connected set $[t, \infty)$ under a continuous mapping) for all $t \in \mathbb{R}_{\geq 0}$. Thus Ω is the decreasing intersection of compact, connected sets, and is hence itself compact and connected. □

Theorem C.2. *Consider a discrete-time dynamical system on a compact set $\mathcal{X}$ in a metric space, obtained through the update $x(t+1) = F(x(t))$ where $F : \mathcal{X} \to \mathcal{X}$ is a continuous function. Suppose also that the dynamical system admits a Lyapunov function u (i.e., a continuous function $u : \mathcal{X} \to \mathbb{R}$ such that $\forall\, x \in \mathcal{X}$, $u(F(x)) \geq u(x)$ with equality iff $F(x) = x$).*

Then the limit set Ω of an orbit $\{x(t)\}_{t\geq 0}$

- *is a compact connected set,*
- *consists entirely of fixed points (i.e., points x for which $F(x) = x$), and*
- *has the property that the Lyapunov function u is constant over it.*[3]

Proof. u **is constant over** Ω. First note that since u is continuous and $\mathcal{X}$ is compact, so $u(\mathcal{X})$ is bounded. Thus the non-decreasing sequence $\{u(x(t))\}_{t\in\mathbb{N}}$ converges to $u^* := \sup\{u(x(t))\}_{t\in\mathbb{N}} < \infty$. Now consider any $\omega \in \Omega$. There exists a subsequence $\{x(t_k)\}_{k\in\mathbb{N}}$ of the sequence of iterates $\{x(t)\}_{t\in\mathbb{N}}$ that converges to ω. $\{u(x(t_k))\}_k$ also converges to $\sup\{u(x(t_k))\}_k = \sup\{u(x(t))\}_t = u^*$. Thus, since u is continuous, we have that $u^* = \lim_{k\to\infty} u(x(t_k)) = u(\lim_{k\to\infty} x(t_k)) = u(\omega)$.
Thus $\forall\, \omega \in \Omega$, $u(\omega) = u^*$.

Ω **consists entirely of fixed points.** Consider any $\omega \in \Omega$ and let $\omega = \lim_{k\to\infty} x(t_k)$, where $\{t_k\}_{k\in\mathbb{N}}$ is an increasing sequence that converges to infinity. Where $u^* := \sup\{u(x_t)\}_t$ as previously defined, we have, by the continuity of F and u, that

$$\begin{aligned} u(F(\omega)) =& u(F(\lim_{k\to\infty} x(t_k))) \\ =& \lim_{k\to\infty} u(F(x(t_k))) \\ =& \lim_{k\to\infty} u(x(t_k + 1)) \\ =& u^* = u(\omega), \end{aligned}$$

so we must have $F(\omega) = \omega$ by the assumption that u is a Lyapunov function.

[3] i.e., $u(\omega) = u(\omega')$ for any $\omega, \omega' \in \Omega$.

Ω is compact. Ω can be written as

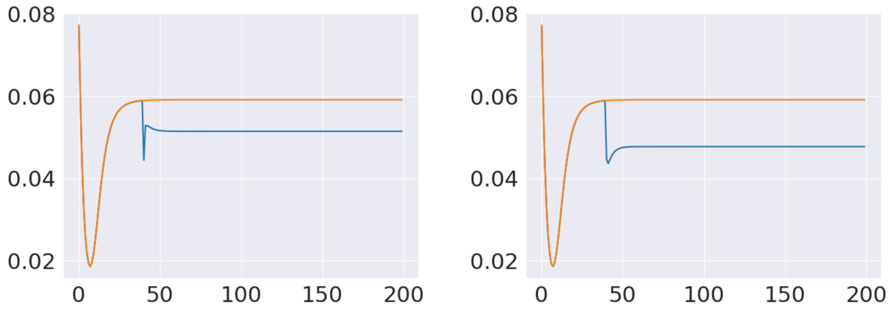
$$\Omega = \bigcap_{t\in\mathbb{N}} \overline{\{x(s) : s \in \mathbb{N}, s \geq t\}}$$

where $\overline{\{x(s) : s \in \mathbb{N}, s \geq t\}}$, which denotes the closure of the set $\{x(s) : s \in \mathbb{N}, s \geq t\}$, is compact (since it is a closed subset of the compact set $\mathcal{X}$) for all $t \in \mathbb{N}$. Thus Ω is the decreasing intersection of compact sets, and is hence itself compact.

Ω is connected. Suppose to the contrary that Ω is the disjoint union of nonempty closed sets Ω_1, Ω_2. Ω_1, Ω_2 are closed subsets of the compact set $\mathcal{X}$ and hence compact, so they are a finite distance $\eta > 0$ apart.
For $i = 1, 2$, let $V_i := \{B(\Omega_i, \frac{\eta}{4})\} \cap \{z \in \mathcal{X} : d(F(z), z) < \frac{\epsilon}{2}\}$, where $d(\cdot, \cdot)$ is the distance function and $B(\Omega_i, \frac{\epsilon}{4}) = \{z \in \mathcal{X} : d(z, \Omega_i) < \frac{\epsilon}{4}\}$ is the open set of all points which are at a distance of $< \frac{\epsilon}{4}$ to the set Ω_i.
Note that $\Omega_i \subseteq V_i$ for $i = 1, 2$ since $F(z) = z$ on Ω. Note also that V_1, V_2 are open in $\mathcal{X}$, with distance $d(V_1, V_2) \geq d(B(\Omega_1, \frac{\epsilon}{4}), B(\Omega_2, \frac{\epsilon}{4})) = \frac{\epsilon}{2}$. In particular, this means that $\forall\, z \in V_1$, $F(z) \notin V_2$ (and vice versa), since $d(F(z), z) < \frac{\epsilon}{2}\ \forall\, z \in V_1 \cup V_2$.
Now there exists T such that $x(t) \in V_1 \cup V_2\ \forall\, t \geq T$, since if not then $\exists$ subsequence $\{x(t_k)\}_k$ that lies within the compact set $\mathcal{X} - (V_1 \cup V_2)$ and hence has a limit point $x^* \in \mathcal{X} - (V_1 \cup V_2)$, which leads to a contradiction since x^* is then a limit point of $\{x(t)\}_t$ that $\notin \Omega$.
Suppose then, without loss of generality, that $x(T) \in V_1$. Then since we have established that if $x(t) \in V_1$ then $x(t+1) = F(x(t)) \notin V_2$, so we must have that $x(t) \in V_1\ \forall\, t \geq T$. But this means that no point of Ω_2 is a limit point of $\{x(t)\}_t$, which is a contradiction. Hence Ω is connected. □

D Additional Experiments

In Figure 12, we showed the resultant exploitabilities of lin-MMWU after introducing perturbations to the algorithm. In order to generate the plots, we require 3 additional parameters – maximum allowable exploitability, denoted by $\exp_{max}$ and disturbance/perturbation amount, denoted by δ. $\exp_{max}$ is an upper bound on the maximum allowable exploitability of the dynamics, and δ is a scalar in $[0, 1]$ that denotes how much each player perturbs their current strategy ρ or σ by. If randomized, we generate a random Hermitian matrix of suitable dimension, $\mathcal{A}$, and the perturbed strategy is given by $\rho^* = ([1-\delta)\rho + \delta\mathcal{A}]/\operatorname{tr}((1-\delta)\rho + \delta\mathcal{A}))$. Otherwise, the perturbed strategy is given by $\rho^* = [(1-\delta)\rho + \delta\Phi(\rho)]/\operatorname{tr}((1-\delta)\rho + \delta\mathcal{A}))$, which gives a perturbation in the direction of the best response. In Figure 13, we also show exploitability plots for lin-MMWU with only one perturbation and $\delta = 0.1$, showing that the dynamics go to a fixed point of lower exploitability. We did not perform extensive hyperparameter tuning for this experiment, but from the experiments it is clear that even a single perturbation would improve the performance of lin-MMWU.



Single random perturbation

Single best response perturbation

Figure 13: Exploitabilities of lin-MMWU with single perturbation.

As already determined in the main text, the lin-QREP dynamics are a gradient flow, and thus, the common utility function is non-decreasing along its trajectories. In Figure 14, we experimentally verify this fact for the case $q = 1$, where we run the continuous dynamics for 50 randomly generated quantum common-interest games with randomly generated initial conditions. Our experiments corroborate our theoretical findings, namely that the utility for both players (plotted is the first player's utility) is strictly increasing unless they are at a fixed point.

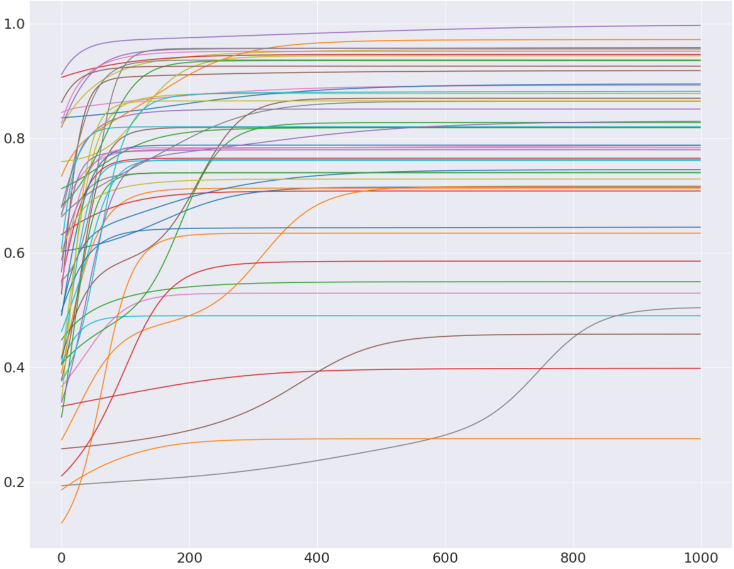
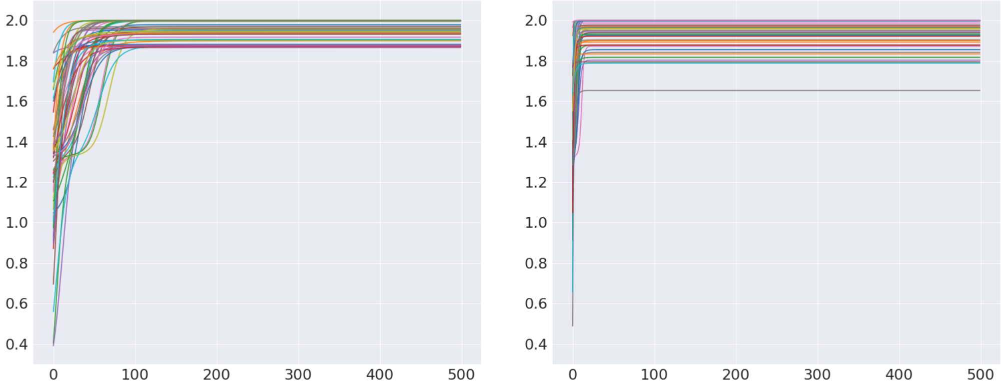


Figure 14: Utility function values for continuous lin-QREP *dynamics applied to 50 randomly generated quantum common-interest games. In all trajectories, we see that utility increases over time.*

Next, we compare the convergence properties of the continuous and discrete dynamics. Our theoretical guarantees for both lin-QREP and lin-MMWU are that their limit points are a compact connected set of fixed points (Theorem 5.6) Moreover, we know that the set of fixed points of both dynamics are equivalent (Theorem 5.2). We showcase this set of results by using a representative game example (with Nash equilibrium utility of 2, see Figure 15). Indeed, when applied to this game, both lin-QREP and lin-MMWU converge to a range of utilities which correspond to local optima of the corresponding BSS instance, and furthermore this range is similar between the continuous and discrete dynamic.

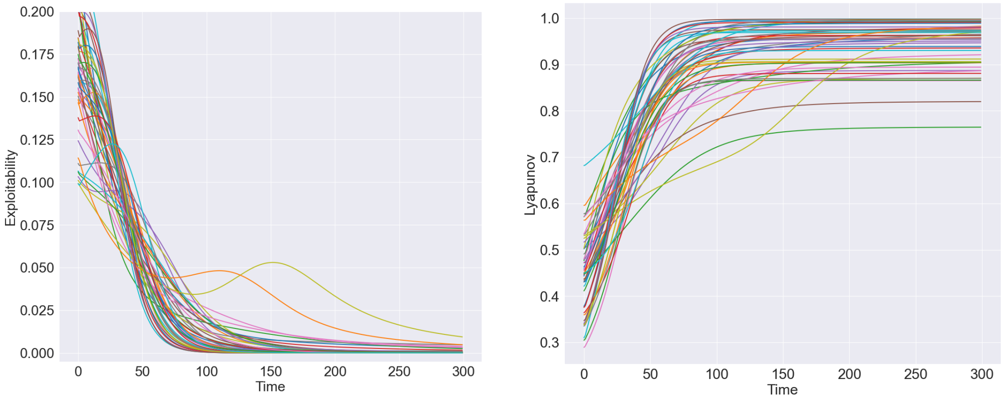


Continuous Dynamics (lin-QREP).

Discrete Dynamics (lin-MMWU).

Figure 15: Utility value of continuous vs discrete dynamics (50 random initializations) when applied to a quantum common-interest game.

Additionally, we perform experiments using MMWU (the exponential variant of lin-MMWU) with stepsize $\epsilon = 0.1$ on the same set of randomized games as in Figure 3. We show in Figure 16 that empirically, MMWU exhibits similar properties to lin-MMWU: utility is increasing over time, and exploitability goes close to zero in a large fraction of the test cases. Moreover, we see that the two "bad" cases (orange and brown trajectories) are actually able to escape the peaks of high exploitability and reach lower exploitability over time. We leave this interesting observation to future work.



Exploitability

Utility function

Figure 16: Exploitability and utility function values for MMWU dynamics applied to 50 randomly generated quantum CIGs.

Finally we compare our formulation of continuous time dynamics with the exp-QREP dynamics (which they call *quantum replicator dynamics*) derived in [18]. Their formulation is derived by taking the continuous-time limit of MMWU, which makes it distinct from our lin-QREP dynamics. In Figure 17, we run both formulations with the same randomized $\mathcal{H}_2 \otimes \mathcal{H}_2$ game and uniform initial conditions, showing diverging trajectories.

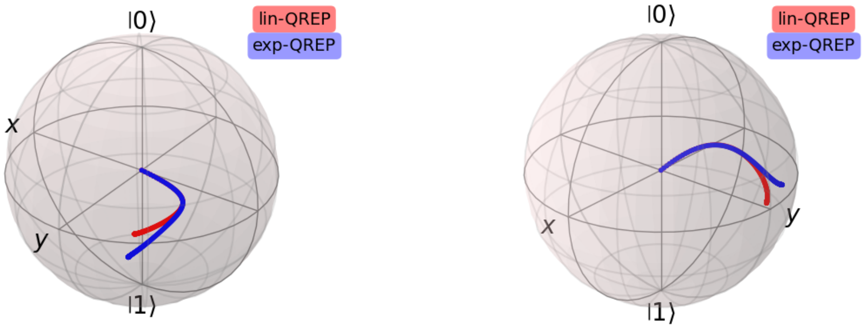


Figure 17: Example simulations where trajectories of lin-QREP and quantum replicator dynamics from [18] (exp-QREP) diverge.