

# LEARNING IN QUANTUM GAMES

KYRIAKOS LOTIDIS\*, PANAYOTIS MERTIKOPOULOS<sup>‡</sup>, AND NICHOLAS BAMBOS<sup>‡</sup>

ABSTRACT. In this paper, we introduce a class of learning dynamics for general quantum games, that we call “*follow the quantum regularized leader*” (FTQL), in reference to the classical FTRL template for finite games. We show that the induced quantum state dynamics decompose into (i) a classical, commutative component which governs the dynamics of the system’s eigenvalues in a way analogous to the evolution of mixed strategies under FTRL; and (ii) a non-commutative component for the system’s eigenvectors which has no classical counterpart. Despite the complications that this non-classical component entails, we find that the FTQL dynamics incur no more than constant regret in *all* quantum games. Moreover, adjusting classical notions of stability to account for the nonlinear geometry of the state space of quantum games, we show that only *pure quantum equilibria* can be stable and attracting under FTQL while, as a partial converse, pure equilibria that satisfy a certain “variational stability” condition are always attracting. Finally, we show that the FTQL dynamics are *Poincaré recurrent* in quantum min-max games, extending in this way a very recent result for the quantum replicator dynamics.

## 1. INTRODUCTION

The advent of quantum information theory – and, with it, the associated “quantum advantage” [4, 36, 50] – has had a profound impact on computer science and machine learning, from quantum cryptography and shadow tomography [1], to quantum generative adversarial networks (QGANs) and adversarial learning [7, 11, 27]. At a high level, the advantages of quantum-based computing are owed to the possibility of preparing superpositions of binary-state quantum systems known as *qubits*: classical bits cannot lie in superposition, so the calculations that can be performed by classical computers are *de facto* limited by their binary alphabet and memory structure. In light of this, quantum computing has the potential to greatly accelerate the development of artificial intelligence algorithms and models, with Google’s “Sycamore” 54-qubit processor training an autonomous vehicle model in less than 200 seconds [4].

In a similar manner, when such models are deployed in a multi-agent context – e.g., as in the case of QGANs or autonomous vehicles – the landscape changes drastically relative to classical non-cooperative frameworks. The main reason for this is again the “quantum advantage”: due to the intricacies of decoherence and entanglement – two quantum notions that have no classical counterpart – quantum players can have a distinct advantage over “classical” players, achieving higher payoffs at equilibrium than would otherwise be possible [15, 33]. This is again owed to the fact that probabilistic mixing works differently in the quantum and classical worlds: in classical games, a mixed *strategy* is a probabilistic convex combination of

---

\* DEPARTMENT OF MANAGEMENT SCIENCE & ENGINEERING, STANFORD UNIVERSITY.

‡ UNIV. GRENoble ALPES, CNRS, INRIA, GRENoble INP, LIG, 38000 GRENoble, FRANCE.

‡ DEPARTMENT OF ELECTRICAL ENGINEERING AND MANAGEMENT SCIENCE & ENGINEERING, STANFORD UNIVERSITY.

*E-mail addresses:* [klotidis@stanford.edu](mailto:klotidis@stanford.edu), [panayotis.mertikopoulos@imag.fr](mailto:panayotis.mertikopoulos@imag.fr), [bambos@stanford.edu](mailto:bambos@stanford.edu).

2020 *Mathematics Subject Classification.* Primary 91A81, 37N40, 68Q32; secondary 68T05, 81Q93, 91B80.

*Key words and phrases.* Quantum games; Nash equilibrium; regularized learning; asymptotic stability.

the constituent pure strategies; in quantum games, a mixed *state* is a probabilistic mixture of the quantum *projectors* associated to each constituent state. Because of this, a mixed quantum state can return payoffs that lie outside the convex hull of classical mixed strategies, thus providing a tangible advantage to players with access to quantum technologies – e.g., the ability to encode their action in a qubit register, which is then submitted to a “referee” (the natural mechanism determining the payoffs of a quantum game).

Of course, the extent to which the advantage of quantum players manifests itself is contingent on the players’ actually reaching an equilibrium. The recent work of Bostanci & Watrous [5] has shown that the problem of computing an approximate Nash equilibrium of a quantum game is included in PPAD, so, by the seminal work of Daskalakis et al. [12, 13], it must be complete for this class (since computing a quantum equilibrium is at least as hard as computing a classical one). Thus, given that the dimensionality of a quantum game is exponential in the number of qubits available to each player, computing a Nash equilibrium of a quantum game quickly becomes an intractable affair, in all but the smallest games. On that account, it seems more reasonable to turn to an online learning paradigm where each player seeks to minimize their individual regret, and instead ask:

*Are all equilibrium outcomes equally likely under a quantum no-regret learning scheme?  
Is there a class of equilibria with an inherent selection bias – either for or against?*

**Our contributions.** First, to achieve no-regret in a quantum setting, we introduce a flexible model for learning in general  $N$ -player quantum games based on the popular “follow the regularized leader” (FTRL) template for *finite* games [41, 42]. The resulting model, which we call “follow the quantum regularized leader” (FTQL), contains as a special case the matrix multiplicative / exponential weights (MMW) dynamics that have been used extensively in quantum games and matrix learning [2, 22–24, 46], and which give rise to the quantum replicator dynamics [20, 23]. Importantly, as we show in Section 3, the mixed-state dynamics of FTQL decompose into a “classical” part (eigenvalues evolve as the FTRL dynamics in finite games), plus a “quantum” component capturing the evolution of the system’s eigenfunctions (and which has no classical analogue).

In terms of regret minimization, all FTQL dynamics incur at most constant regret, so they represent a compelling choice from a learning standpoint. However, deriving the dynamics’ equilibrium convergence properties is significantly more difficult because of the nonlinear geometry of the game’s state space. Specifically, in contrast to finite games (where pure strategies are isolated extreme points), the pure states of a quantum game form a continuous manifold of stationary points (all of them extreme), so the study of stability and convergence questions becomes a highly involved affair. Nonetheless, despite these topological complications, we show that FTQL enjoys the following fundamental properties: (i) Nash equilibria are stationary; (ii) limits of interior trajectories and Lyapunov stable states are Nash; (iii) only *pure quantum equilibria* can be stable and attracting under FTQL (up to the exclusion of trivial stationary states); and (iv) as a partial converse, we show that pure states that satisfy a certain “variational stability” condition are attracting, irrespective of the chosen regularizer. On that account, our results lead to an implicit quantum “purification” principle: under FTQL, *mixed states are inherently fragile, and only pure quantum states can be consistently attracting*.<sup>1</sup>

Finally, we complement our results with a closer look at two-player, zero-sum quantum games, where we show that FTQL exhibits a cycling property known as *Poincaré recurrence*:

---

<sup>1</sup>In the classical world, a version of the above collection of results is sometimes referred to as the “folk theorem” of evolutionary game theory [10, 17, 21, 30].

almost all trajectories of play return arbitrarily close to their starting point infinitely often. This result is a broad generalization of a recent result by Jain et al. [23], who established this property for the MMW dynamics. In this regard, our result shows that the MMW result is not a coincidence: despite the “quantum advantage”, perfect competition cannot be resolved by the dynamics of regularized learning.

To simplify the presentation, we focus throughout on models that evolve in continuous time. This allows us to sidestep issues having to do with hyperparameter tuning and the like, and instead spotlight the essential aspects of the theory.

## 2. PRELIMINARIES

We start by briefly reviewing some basics of quantum game theory and introducing the necessary context for our results.

**Notation.** Given a (complex) Hilbert space  $\mathcal{H}$ , we will use Dirac’s bra-ket notation to distinguish between an element  $|\psi\rangle$  of  $\mathcal{H}$  and its adjoint  $\langle\psi|$ ; otherwise, when a basis is implied by the context, we will use the dagger notation “ $\dagger$ ” to denote the Hermitian transpose  $\psi^\dagger$  of  $\psi$ . We will also write  $\mathbb{H}^d$  for the space of  $d \times d$  Hermitian matrices, and  $\mathbb{H}_+^d$  for the cone of positive-semidefinite matrices in  $\mathbb{H}^d$ . Finally, given a real function  $f: \mathbb{R} \rightarrow \mathbb{R}$  and a Hermitian matrix  $\mathbf{X} \in \mathbb{H}^d$  with unitary eigen-decomposition  $\mathbf{X} = \sum_{\alpha=1}^d x_\alpha \mathbf{u}_\alpha \mathbf{u}_\alpha^\dagger$ , we will write  $f(\mathbf{X})$  for the (likewise Hermitian) matrix  $f(\mathbf{X}) = \sum_{\alpha=1}^d f(x_\alpha) \mathbf{u}_\alpha \mathbf{u}_\alpha^\dagger$ .

**Quantum games.** Following [15, 19], a *quantum game* consists of the following primitives:

- (1) A finite set of *players*  $i \in \mathcal{N} = \{1, \dots, N\}$ .
- (2) Each player  $i \in \mathcal{N}$  has access to a complex Hilbert space  $\mathcal{H}_i \cong \mathbb{C}^{d_i}$  describing the set of (pure) *quantum states* available to the player (typically a discrete register of qubits). In more detail, a quantum state is an element  $\psi_i$  of  $\mathcal{H}_i$  with unit norm, so the set of all such states is the unit sphere  $\Psi_i := \{\psi_i \in \mathcal{H}_i : \|\psi_i\| = 1\}$  of  $\mathcal{H}_i$ . We will also write  $\Psi := \prod_i \Psi_i$  for the space of all ensembles  $\psi = (\psi_1, \dots, \psi_N)$  of pure states  $\psi_i \in \Psi_i$  that are independently prepared by each player  $i \in \mathcal{N}$ .
- (3) The players’ rewards are determined by their individual *payoff functions*  $u_i: \Psi \rightarrow \mathbb{R}$ . These payoff functions are not arbitrary, but are obtained from a joint *positive operator-valued measure* (POVM) quantum measurement process that unfolds as follows [8]: First, we assume given a finite set of possible *measurement outcomes*  $\omega \in \Omega$  that a referee can observe from the players’ quantum states (e.g., measure a player-prepared qubit to be “up” or “down”). Each such outcome  $\omega \in \Omega$  is associated to a positive semi-definite operator  $\mathbf{P}_\omega: \mathcal{H} \rightarrow \mathcal{H}$  that acts on the tensor product  $\mathcal{H} := \bigotimes_i \mathcal{H}_i$  of the players’ individual state spaces; we further assume that  $\sum_{\omega \in \Omega} \mathbf{P}_\omega = \mathbf{I}$  so the joint probability of observing  $\omega \in \Omega$  when the system is at state  $\psi \in \Psi$  is

$$P_\omega(\psi) = \langle \psi_1 \otimes \dots \otimes \psi_N | \mathbf{P}_\omega | \psi_1 \otimes \dots \otimes \psi_N \rangle \quad (1)$$

The payoff to each player  $i \in \mathcal{N}$  is given by the outcome of this measurement process via a *payoff observable*  $U_i: \Omega \rightarrow \mathbb{R}$ ; specifically, in this context,  $u_i(\psi)$  denotes the player’s expected payoff at state  $\psi \in \Psi$ , viz.

$$u_i(\psi) := \langle U_i \rangle \equiv \sum_{\omega} P_\omega(\psi) U_i(\omega). \quad (2)$$

A *quantum game* is then defined as a tuple  $\mathcal{Q} \equiv \mathcal{Q}(\mathcal{N}, \Psi, u)$  with players, quantum states, and payoff functions as above.

**Mixed states.** In addition to pure states, each player  $i \in \mathcal{N}$  can also prepare probabilistic mixtures thereof, known as *mixed states*. In contrast to mixed strategies in classical, finite games, these mixed states are *not* convex combinations of their pure counterparts; instead, given a family of pure quantum states  $\psi_{i\alpha_i} \in \Psi_i$  indexed by  $\alpha_i \in \mathcal{A}_i$ , a mixed state is described by a *density matrix* of the form

$$\mathbf{X}_i = \sum_{\alpha_i \in \mathcal{A}_i} x_{i\alpha_i} |\psi_{i\alpha_i}\rangle \langle \psi_{i\alpha_i}| \quad (3)$$

where  $x_{i\alpha_i} \geq 0$  is the mixing weight of  $\psi_{i\alpha_i}$ , and we assume that  $\text{tr } \mathbf{X}_i = 1$  (the states  $\psi_{i\alpha_i}$  are not assumed to be orthogonal in this context). By Born's rule, this means that if each player  $i \in \mathcal{N}$  prepares a mixed state according to  $\mathbf{X}_i$ , the probability of observing  $\omega \in \Omega$  under  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_N)$  will be  $P_\omega(\mathbf{X}) = \sum_{\alpha} x_{\alpha} \langle \psi_{\alpha} | \mathbf{P}_{\omega} | \psi_{\alpha} \rangle$ , where, in multi-index notation,  $\alpha = (\alpha_1, \dots, \alpha_N)$ ,  $x_{\alpha} = \prod_i x_{i\alpha_i}$ , and  $\psi_{\alpha} = \bigotimes_i \psi_{i\alpha_i}$ . Thus, in a slight abuse of notation, the expected payoff to player  $i \in \mathcal{N}$  under  $\mathbf{X}$  will be

$$\begin{aligned} u_i(\mathbf{X}) &= \sum_{\omega \in \Omega} \sum_{\alpha \in \mathcal{A}} x_{\alpha} P_{\omega}(\psi_{\alpha}) U_i(\omega) = \sum_{\alpha \in \mathcal{A}} x_{\alpha} u_i(\psi_{\alpha}) \\ &= \sum_{\alpha_1 \in \mathcal{A}_1} \cdots \sum_{\alpha_N \in \mathcal{A}_N} x_{\alpha_1} \cdots x_{\alpha_N} u_i(\psi_{\alpha_1}, \dots, \psi_{\alpha_N}). \end{aligned} \quad (4)$$

**Contrasting to other classes of games.** The expression (4) for a player's expected payoff under a mixed state is reminiscent of mixed extensions of classical finite games, but this association is very tenuous. From a conceptual standpoint, the principal differences are as follows:

- (1) There is an infinite continuum of pure states  $\psi \in \Psi$ , not a finite number thereof (as is the case in finite games).
- (2) The decomposition (3) of a density matrix into pure states is not unique; generically, there may be a continuum of (non-equivalent) families of pure states and mixing weights giving rise to the same density matrix.
- (3) The convex superposition  $\lambda\psi + (1 - \lambda)\psi'$  of two pure states  $\psi$  and  $\psi'$  may give rise to quantum interference terms of the form  $|\psi\rangle\langle\psi'|$  and  $|\psi'\rangle\langle\psi|$  in the induced payoff; these cross-terms have no analogue in finite games.

Because of the above, treating a quantum game as a ‘‘tensorial’’ extension of a finite game can be misleading. Instead, it would be more appropriate to view a quantum game as a *continuous game* where each player  $i \in \mathcal{N}$  controls a matrix variable  $\mathbf{X}_i$  drawn from the ‘‘spectraplex’’

$$\mathcal{X}_i = \{\mathbf{X}_i \in \mathbb{H}_+^{d_i} : \text{tr } \mathbf{X}_i = 1\} \quad (5)$$

and the player's payoff function  $u_i: \mathcal{X} \equiv \prod_j \mathcal{X}_j \rightarrow \mathbb{R}$  is linear in every player's density matrix  $\mathbf{X}_j \in \mathcal{X}_j$ ,  $j \in \mathcal{N}$ .

**Nash equilibrium.** In our quantum setting, the classical solution concept of a *Nash equilibrium* (NE) characterizes mixed quantum states  $\mathbf{X}^* \in \mathcal{X}$  which discourage unilateral deviations in the sense that

$$u_i(\mathbf{X}^*) \geq u_i(\mathbf{X}_i; \mathbf{X}_{-i}^*) \quad \text{for all } \mathbf{X}_i \in \mathcal{X}_i, i \in \mathcal{N} \quad (\text{NE})$$

where we write  $(\mathbf{X}_i; \mathbf{X}_{-i}) = (\mathbf{X}_1, \dots, \mathbf{X}_i, \dots, \mathbf{X}_N)$  for the choice of player  $i$  relative to all other players. Since  $\mathcal{X}_i$  is convex and  $u_i$  is linear in  $\mathbf{X}_i$ , the existence of Nash equilibria follows from the seminal theorem of Debreu [14].

Now, letting

$$\mathbf{V}_i(\mathbf{X}) = \nabla_{\mathbf{X}_i^\top} u_i(\mathbf{X}) \quad (6)$$

denote the individual payoff gradient of player  $i$ , standard arguments [16, 40] show that the Nash equilibria of a quantum game  $\mathcal{Q}$  are precisely the solutions of the variational inequality

$$\text{tr}[\mathbf{V}(\mathbf{X}^*)(\mathbf{X} - \mathbf{X}^*)] \leq 0 \quad \text{for all } \mathbf{X} \in \mathcal{X} \quad (\text{VI})$$

where  $\mathbf{V}(\mathbf{X}) = (\mathbf{V}_1(\mathbf{X}), \dots, \mathbf{V}_N(\mathbf{X}))$ . We note here that, since  $u_i$  is linear in  $\mathbf{X}_i$ , the  $i$ -th block  $\mathbf{V}_i(\mathbf{X})$  of  $\mathbf{V}(\mathbf{X})$  does not depend on  $\mathbf{X}_i$ , and we have

$$u_i(\mathbf{X}_i; \mathbf{X}_{-i}) = \text{tr}[\mathbf{X}_i \mathbf{V}_i(\mathbf{X})] \quad \text{for all } \mathbf{X} \in \mathcal{X}. \quad (7)$$

We will use these properties freely in the sequel.

**Regret.** Complementing the notion of a Nash equilibrium, an important rationality requirement in dynamic environments is the minimization of a player's *regret*, i.e., the performance gap between the player's expected cumulative payoff over time versus the payoff of the best fixed state in hindsight. Formally, the regret of the  $i$ -th player against the trajectory of play  $\mathbf{X}(t) \in \mathcal{X}$ ,  $t \geq 0$ , is defined as

$$\text{Reg}_i(T) = \max_{\mathbf{X}'_i \in \mathcal{X}_i} \int_0^T [u_i(\mathbf{X}'_i; \mathbf{X}_{-i}(t)) - u_i(\mathbf{X}(t))] dt \quad (8)$$

and we say that player  $i$  has *no regret* if  $\text{Reg}_i(T) = o(T)$ . In the rest of the paper, we will focus on learning dynamics that incur no regret, and we will examine their convergence properties relative to the game's Nash equilibria.

### 3. LEARNING DYNAMICS

**Learning via quantum regularization.** In classical, finite games, the most widely studied class of no-regret dynamics is the so-called “*follow the regularized leader*” (FTRL) family of algorithms [30, 41, 42]. The main idea behind this popular template is the following: at each instance  $t \geq 0$ , every player  $i \in \mathcal{N}$  plays a mixed strategy that maximizes the player's cumulative payoff minus a certain regularization penalty. In this way, strategies that perform consistently better tend to be preferred over their underperforming counterparts, while the “regularization penalty” introduces a certain degree of exploration to avoid getting stuck.

In the quantum regime, the role of mixed strategies is played by the game's mixed quantum states, so the reinforcement mechanism behind FTRL leads to the matrix-valued “*follow the quantum regularized leader*” (FTQL) dynamics

$$\mathbf{X}_i(t) = \arg \max_{\mathbf{X}_i \in \mathcal{X}_i} \left\{ \int_0^t u_i(\mathbf{X}_i; \mathbf{X}_{-i}(s)) ds - h_i(\mathbf{X}_i) \right\} \quad (9)$$

where  $h_i: \mathcal{X}_i \rightarrow \mathbb{R}$  denotes the *penalty function* – or *regularizer* – of player  $i$  (discussed in detail below). As stated, the FTQL dynamics (9) are in integral form, which is not particularly well-suited for our analysis. Instead, to obtain a more concrete, autonomous reformulation, consider as a first step the *regularized best response maps*

$$\mathbf{Q}_i(\mathbf{Y}_i) := \arg \max_{\mathbf{X}_i \in \mathcal{X}_i} \{ \text{tr}[\mathbf{Y}_i \mathbf{X}_i] - h_i(\mathbf{X}_i) \} \quad (10)$$

defined for all Hermitian  $\mathbf{Y}_i \in \mathcal{Y}_i := \mathbb{H}^{d_i}$ . Then, in view of (7), the integral dynamics (9) can be recast in differential form as

$$\dot{\mathbf{Y}}_i(t) = \mathbf{V}_i(\mathbf{X}(t)) \quad \mathbf{X}_i(t) = \mathbf{Q}_i(\mathbf{Y}_i(t)). \quad (\text{FTQL})$$

The dynamics (FTQL) will be the basis of our analysis, so some remarks are in order. First, in terms of interpretation, (FTQL) can be seen as a gradient-following process coupled with a regularized state selection scheme – the dynamics  $\dot{\mathbf{Y}}_i = \mathbf{V}_i$  and the mapping  $\mathbf{Q}_i: \mathbf{Y}_i \mapsto \mathbf{X}_i$  respectively. In this regard, the regularizer  $h_i$  which underlies the definition of  $\mathbf{Q}_i$  plays a crucial role, and different choices of  $h_i$  may yield very different dynamics. For concreteness, we will only assume in the sequel that  $h_i$  is a trace function of the form  $h_i(\mathbf{X}_i) = \text{tr}[\theta_i(\mathbf{X}_i)]$  where  $\theta_i: [0, 1] \rightarrow \mathbb{R}$  is continuous on  $[0, 1]$  and has  $\inf_{x \in (0, 1]} \theta_i''(x) > 0$ . We will also say that  $\theta_i$  is *steep* when  $\lim_{x \rightarrow 0^+} \theta_i'(x) = -\infty$ , and, for normalization purposes, we will assume that  $\theta_i(0) = 0$ .

Suppressing player indices for simplicity, some standard examples of regularizers are as follows:

**Example 3.1** ( $L^2$  regularization). If  $\theta(x) = x^2/2$ , the players' penalty function is the squared Frobenius norm  $h(\mathbf{X}) = (1/2)\|\mathbf{X}\|_F^2$ , in which case Eq. (10) gives the orthogonal projector  $\mathbf{Q}(\mathbf{Y}) = \Pi_{\mathcal{X}}(\mathbf{Y}) \equiv \arg \min_{\mathbf{X} \in \mathcal{X}} \|\mathbf{X} - \mathbf{Y}\|_F$ . This choice leads to the (Frobenius) projection dynamics:

$$\dot{\mathbf{Y}} = \mathbf{V}(\mathbf{X}) \quad \mathbf{X} = \Pi_{\mathcal{X}}(\mathbf{Y}) \quad (\text{PD})$$

**Example 3.2** (Von Neumann regularization). Another standard choice is  $\theta(x) = x \log x$  which yields the (negative) von Neumann entropy  $h(\mathbf{X}) = \text{tr}[\mathbf{X} \log \mathbf{X}]$ . By a standard calculation, this choice of regularizer gives rise to the *matrix multiplicative / exponential weights* (MMW) dynamics

$$\dot{\mathbf{Y}} = \mathbf{V}(\mathbf{X}) \quad \mathbf{X} = \frac{\exp(\mathbf{Y})}{\text{tr}[\exp(\mathbf{Y})]} \quad (\text{MMW})$$

A discrete-time version of these dynamics was introduced in the context of kernel learning by Tsuda et al. [46] and Kakade et al. [24]; for a series of more recent developments in the context of quantum learning see [2, 22, 23].

**Example 3.3** (Tsallis regularization). Interpolating between the above, the *Tsallis regularizer* is given by  $\theta(x) = [q(1 - q)]^{-1}(x - x^q)$  for some  $q > 0$  (with the continuity convention  $(x - x^q)/(1 - q) = x \log x$  for  $q = 1$ ). When  $q \leftarrow 2$ , we recover the projection dynamics (PD); by contrast, the choice  $q \leftarrow 1$  gives rise to (MMW); finally, the choice  $q \leftarrow 1/2$  is particularly popular in the context of bandit online learning, cf. [24, 51] and references therein.

**The mixed-state dynamics of FTQL.** Under (FTQL), the evolution of the players' mixed states  $\mathbf{X}(t)$  is described *implicitly* via that of the auxiliary score matrix  $\mathbf{Y}(t)$ . On the other hand, obtaining an *explicit* expression for the dynamics of  $\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t))$  is considerably more difficult because the rules of matrix calculus do not provide an analytic expression for the tensor derivative  $\nabla_{\mathbf{Y}} \mathbf{Q}(\mathbf{Y})$  of  $\mathbf{Q}$ , even when the latter is available in closed form.

To circumvent this difficulty, we will work with a unitary eigendecomposition of  $\mathbf{X}$  of the form

$$\mathbf{X} = \sum_{\alpha=1}^d x_{\alpha} \mathbf{u}_{\alpha} \mathbf{u}_{\alpha}^{\dagger} \quad (11)$$

where  $x_{\alpha} \geq 0$ ,  $\alpha = 1, \dots, d$ , is an enumeration of the eigenvalues of  $\mathbf{X}$ ,  $\mathbf{u}_{\alpha} \in \mathcal{H}$  is a unit-norm eigenvector of  $\mathbf{X}$  corresponding to  $x_{\alpha}$ , and player indices have again been suppressed (to lighten notation). Since, in general,  $\mathbf{X}$  does not commute with  $\mathbf{V}$  (and hence with  $\dot{\mathbf{X}}$ ), the eigenvalues and eigenvectors of  $\mathbf{X}$  will evolve in a coupled, concurrent manner; our first result below provides an explicit expression for this co-evolution:

**Theorem 1.** Let  $\mathbf{X}(t) = \sum_{\alpha=1}^d x_{\alpha}(t) \mathbf{u}_{\alpha}(t) \mathbf{u}_{\alpha}^{\dagger}(t)$  be an eigendecomposition of  $\mathbf{X}(t)$  as per (11), and suppose that  $h$  is steep. Then, under (FTQL), the entries  $[\dot{\mathbf{X}}]_{\alpha\beta} = \mathbf{u}_{\alpha}^{\dagger} \dot{\mathbf{X}} \mathbf{u}_{\beta}$  of  $\dot{\mathbf{X}}$  follow the quantum state dynamics:

$$[\dot{\mathbf{X}}]_{\alpha\beta} = \begin{cases} \frac{V_{\alpha\alpha}}{\theta''(x_{\alpha})} - \frac{\sum_{\kappa} V_{\kappa\kappa} / \theta''(x_{\kappa})}{\sum_{\kappa} \theta''(x_{\alpha}) / \theta''(x_{\kappa})} & \alpha = \beta \\ \frac{x_{\beta} - x_{\alpha}}{\theta'(x_{\beta}) - \theta'(x_{\alpha})} V_{\alpha\beta} & \alpha \neq \beta \end{cases} \quad (\text{QD})$$

where  $V_{\alpha\beta} = \mathbf{u}_{\alpha}^{\dagger} \mathbf{V}(\mathbf{X}) \mathbf{u}_{\beta}$  and we are using the continuity convention  $(x - y) / [\theta'(x) - \theta'(y)] = 1 / \theta''(x)$  when  $y \rightarrow x$ .

The proof of Theorem 1 is based on the fact that, if a Hermitian matrix  $\mathbf{A}$  with eigendecomposition  $\mathbf{A} = \sum_{\alpha} \lambda_{\alpha} \mathbf{u}_{\alpha} \mathbf{u}_{\alpha}^{\dagger}$  follows the dynamics  $\dot{\mathbf{A}} = \mathbf{M}$ , a differentiation of the eigenvalue equation  $\mathbf{A} \mathbf{u}_{\alpha} = \lambda_{\alpha} \mathbf{u}_{\alpha}$  yields

$$\dot{\lambda}_{\alpha} \delta_{\alpha\beta} = (\lambda_{\alpha} - \lambda_{\beta}) \mathbf{u}_{\alpha}^{\dagger} \dot{\mathbf{u}}_{\beta} + \mathbf{u}_{\alpha}^{\dagger} \mathbf{M} \mathbf{u}_{\beta} \quad \text{for all } \alpha, \beta. \quad (12)$$

This identity allows us to analyze and derive an expression for  $\dot{\mathbf{X}}$  by solving the Lagrangian associated with the maximization problem (10). The full proof is relegated to the appendix; instead, we focus here on a representative example.

**Example 3.4** (The quantum replicator dynamics). An important special case of the dynamics (QD) is obtained by the von Neumann regularizer  $h(\mathbf{X}) = \text{tr}[\mathbf{X} \log \mathbf{X}]$  of Example 3.2. This yields the *quantum replicator dynamics*

$$[\dot{\mathbf{X}}]_{\alpha\beta} = \begin{cases} x_{\alpha} [V_{\alpha\alpha} - \sum_{\kappa} x_{\kappa} V_{\kappa\kappa}] & \text{for } \alpha = \beta \\ \frac{x_{\beta} - x_{\alpha}}{\log x_{\beta} - \log x_{\alpha}} V_{\alpha\beta} & \text{for } \alpha \neq \beta \end{cases} \quad (\text{QRD})$$

for the diagonal and off-diagonal elements of  $\mathbf{X}$  respectively. The diagonal part of (QRD) is formally analogous to the replicator dynamics of evolutionary game theory [39, 45, 48] and captures the evolution of the eigenvalues of  $\mathbf{X}(t)$ . Thus, taken together with its off-diagonal component, (QRD) provides an explicit expression for the evolution of mixed states under (MMW). Alternatively, by applying Fréchet's differentiation formula to (MMW) directly, (QRD) can be rewritten in basis-free notation as

$$\dot{\mathbf{X}} = \int_0^1 \mathbf{X}^{1-s} \mathbf{V}(\mathbf{X}) \mathbf{X}^s ds - \text{tr}[\mathbf{X} \mathbf{V}(\mathbf{X})] \mathbf{X} \quad (13)$$

The dynamics (QRD) and the coordinate-free expression (13) agree with the dynamics of Jain et al. [23] (who derived an equivalent expression under the assumption that  $\mathbf{X}$  and  $\dot{\mathbf{X}}$  commute), but not with the dynamics of Hidalgo [20] that follow a different, unrelated quantization paradigm. We provide the relevant calculations in Appendix B.  $\diamond$

In the sections that follow, we will examine in detail how the classical and quantum components of (FTQL) interface with each other to determine the player's long-run behavior.

#### 4. REGRET MINIMIZATION

We begin our analysis of (FTQL) with a result concerning the dynamics' regret minimization properties. To provide the necessary context, it is known that the FTRL dynamics incur at most constant regret in classical, *finite* games [25, 32]. As we illustrate below, despite taking place over a *continuum* of pure states, the matrix-valued dynamics (FTQL) enjoy the

same regret minimization guarantees in quantum games. Formally, we have the following result.

**Proposition 1.** *Let  $\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t))$ ,  $t \geq 0$ , be a trajectory of play induced by (FTQL). Then, for all  $T \geq 0$ , we have*

$$\text{Reg}_i(T) \leq |d_i \cdot \theta_i(1/d_i) - \theta_i(1)|. \quad (14)$$

The proof of [Proposition 1](#) builds on the general theory of [\[25\]](#) and is presented in detail in [Appendix C](#). Instead, we only note here that the dependence of the bound [\(14\)](#) on the dimensionality of the game depends crucially on the choice of regularizer: (i) the Euclidean regularizer  $\theta(x) = x^2/2$  gives an  $\mathcal{O}(1)$  bound; (ii) the von Neumann entropy leads to an  $\mathcal{O}(\log d)$  dependence; and, finally, (iii) the bound for the Tsallis regularizer  $\theta(x) \propto x - x^q$  is  $\mathcal{O}(d^{1-q})$  for all  $q \in (0, 1)$ . This should be contrasted to *discrete-time* models of online learning, where quadratic regularization leads to suboptimal results relative to *both* the entropic and Tsallis variants [\[26, 44, 51\]](#). The reason for this discrete-to-continuous gap has to do with the fact that (FTQL) admits an *exact* energy function, the so-called *Fenchel coupling*

$$F_i(\mathbf{X}_i, \mathbf{Y}_i) = h_i(\mathbf{X}_i) + h_i^*(\mathbf{Y}_i) - \text{tr}[\mathbf{X}_i \mathbf{Y}_i] \quad (15)$$

where  $\mathbf{X}_i \in \mathcal{X}_i$ ,  $\mathbf{Y}_i \in \mathcal{Y}_i$ , and

$$h_i^*(\mathbf{Y}_i) = \max_{\mathbf{X}_i \in \mathcal{X}_i} \{\text{tr}[\mathbf{X}_i \mathbf{Y}_i] - h_i(\mathbf{X}_i)\} \quad (16)$$

denotes the convex conjugate of  $h_i$ . A version of this primal-dual coupling was first introduced by [\[30\]](#) in the setting of finite games, and it has the following fundamental property:

**Lemma 1.** *Let  $\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t))$ ,  $t \geq 0$ , be a trajectory of play induced by (FTQL). Then, for all  $\mathbf{P} \in \mathcal{X}$ , we have*

$$\frac{d}{dt} F_i(\mathbf{P}_i, \mathbf{Y}_i(t)) = \text{tr}[\mathbf{V}_i(\mathbf{X}(t)) (\mathbf{X}_i(t) - \mathbf{P}_i)]. \quad (17)$$

The importance of this lemma (which is proved in [Appendix A](#)) lies in that the RHS of [\(17\)](#) is precisely the integrand of the regret, so [Proposition 1](#) is obtained by applying [Lemma 1](#) to the state that witnesses the maximum in the definition [\(8\)](#) of the regret of player  $i$ . We defer the relevant calculations to [Appendix C](#).

## 5. CONVERGENCE, STABILITY, AND THE FOLK THEOREM

In view of the strong regret minimization guarantees of [Proposition 1](#), the matrix-valued dynamics (FTQL) emerge as a very compelling choice from a learning standpoint. At the same time, even in the case of classical finite games, it is known that regret minimization *does not* suffice to exclude non-rationalizable outcomes: for example, as was shown by Viossat & Zapechelnyuk [\[47\]](#), the players' empirical frequency of play under a no-regret policy may still end up assigning positive selection probability to (strictly) dominated strategies, and *only* dominated ones. On that account, our aim in the rest of this section will be to take a closer look at the convergence and stability properties of (FTQL) relative to the game's Nash equilibria.



**Notions of stability and convergence.** Our analysis will require some basic concepts from the theory of dynamical systems, which we quickly discuss below. To begin with, recall that a *flow* on an abstract metric space  $\mathcal{Z}$  is a continuous map  $\phi: \mathbb{R} \times \mathcal{Z} \rightarrow \mathcal{Z}$  such that (a)  $\phi_0(z) = z$ ; and (b)  $\phi_{t+s}(z) = \phi_t(\phi_s(z))$  for all  $t, s \in \mathbb{R}$  and all  $z \in \mathcal{Z}$ . Informally, a flow is usually generated by the *solution orbits* of a system of well-posed ordinary differential equations (ODEs), such as (FTQL): in this interpretation,  $\phi_t(z)$  simply denotes the position at time  $t$  of the ODE solution that starts at  $z$  at time  $t = 0$ .

With this in mind, the following notions of invariance and stability will play a key role in our analysis. Given a point  $p \in \mathcal{Z}$ , we will say that:

- (1)  $p$  is *stationary* if  $\phi_t(p) = p$  for all  $t \in \mathbb{R}$ .
- (2)  $p$  is *Lyapunov stable* – or just *stable* – if, for every neighborhood  $\mathcal{U}$  of  $p$  in  $\mathcal{Z}$ , there exists some (smaller) neighborhood  $\mathcal{U}'$  of  $p$  in  $\mathcal{Z}$  such that  $\phi_t(\mathcal{U}') \subseteq \mathcal{U}$  for all  $t \in \mathbb{R}$ . In other words,  $p$  is stable if any orbit that starts close enough to  $p$  remains close enough.
- (3)  $p$  is *attracting* if it admits a neighborhood  $\mathcal{U}$  such that  $\lim_{t \rightarrow \infty} \phi_t(z) = p$  for all  $z \in \mathcal{U}$ . In other words,  $p$  is attracting if all nearby orbits converge to  $p$ .
- (4)  $p$  is *asymptotically stable* if it is stable and attracting.

In what follows, we will seek to characterize precisely the stable and/or attracting states of (FTQL).

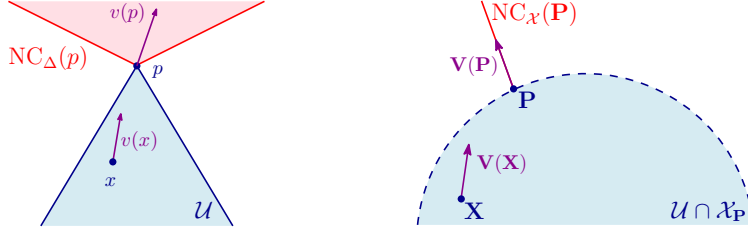
**The classical regime.** To set the stage for the analysis to come, it will be useful to revisit the classical regime of learning in classical *finite* games. Focusing for concreteness on the standard case of the replicator dynamics (cf. Examples 3.2 and 3.4 above), the stability and convergence landscape for general finite games can be encoded in the so-called “folk theorem” of evolutionary game theory, which states the following [10, 21]:

- (1) Nash equilibria are stationary.
- (2) Limits of interior orbits are Nash equilibria.
- (3) Lyapunov stable states are Nash equilibria.
- (4) A state is asymptotically stable if and only if it is a strict Nash equilibrium (that is, every player has a unique best response at equilibrium).

Modulo some technicalities, these properties extend to the entire class of FTRL dynamics for learning in classical *finite* games, cf. [17, 30], and references therein. However, the nonlinear geometry of the players’ state space in quantum games places severe structural limitations on which of these properties transfer over to the non-commutative, matrix-valued setting of (FTQL). We explore this issue below.

**Regularized learning in the spectraplex.** A quick look at the quantum replicator dynamics (QRD) reveals the following structural property: an eigenvalue of  $\mathbf{X}$  that is initially zero in (QRD) will always remain zero; likewise, an eigenvalue that is initially positive, will always remain positive. Formally, this means that the kernel  $\ker(\mathbf{X})$  of  $\mathbf{X}$  remains invariant under (QRD); hence, given that the linear span of a Hermitian matrix is the orthocomplement of its kernel, the same holds for  $\text{im}(\mathbf{X})$  as well.

The fact that the kernel – or, equivalently, the image – of a density matrix remains invariant under (QRD) is the quantum analogue of the fact that the support of a mixed strategy profile remains invariant under the standard replicator dynamics. In the context of finite games, an immediate consequence of this invariance is that *all* pure strategy profiles are stationary (as zero-dimensional faces of the simplex). This property extends to (QRD)



**Figure 1:** The fundamental geometric discrepancy between the classical and quantum regimes (left and right respectively). In classical finite games, the normal cone  $NC_{\Delta}(p)$  to the simplex at a pure strategy  $p$  has nonempty topological interior, so pure Nash equilibria are generically *robust*: if  $v(p)$  is normal to  $\Delta$  at  $p$ , it will remain normal to  $\Delta$  after a small perturbation. On the other hand, in quantum games, the normal cone  $NC_{\mathcal{X}}(\mathbf{P})$  to the spectraplex at a pure state  $\mathbf{P}$  is a ray, so pure Nash equilibria cannot be robust in this sense. We also note the different geometry of pure states: in the simplex, pure strategy profiles are isolated extreme points; in the spectraplex, pure states form a continuous manifold.

and, in fact, to the entire class of mixed-state dynamics under study: formally, under (QD), *all pure quantum states are stationary*.

That being said, the major qualitative difference between the quantum and classical regimes is that, in quantum games, there is a *continuum* of pure states, namely the entire manifold  $\Psi$  of rank 1 density matrices (a product of spheres). By contrast, in finite games, the pure states are the corners of the simplex  $\Delta$  spanned by the player's pure strategies, so they are finite in number and *isolated*. As a result, in classical finite games, a pure strategy profile *can* be asymptotically stable; in quantum games, since every pure state is surrounded by other invariant states, *it cannot*.

A second major difference is that, in finite games, strict Nash equilibria are *robust*: a small perturbation of the payoffs of the game does not change the game's strict equilibria. In quantum games, this robustness disappears: indeed, the variational characterization (VI) of Nash equilibria means that  $\mathbf{V}(\mathbf{X}^*)$  must be an element of the normal cone to  $\mathcal{X}$  at  $\mathbf{X}^*$ ; however, the normal cone to the spectraplex at a matrix of rank 1 has empty topological interior, so the required membership property cannot be robust (for a graphical illustration, see Fig. 1). In particular, any generic perturbation to the payoffs of a quantum game, no matter how small, may lead to a small displacement of the equilibrium in question on the manifold of pure states  $\Psi$ .

**Consistency and variational stability.** In view of the above, we can already draw two major conclusions for the quantum setting:

- (1) Any concept of asymptotic stability must also include a notion of *consistency*: a state cannot be accessed if it is absent from the linear span of the dynamics' initial state.
- (2) Any concept of robustness must likewise incorporate a notion of *variational stability*: small perturbations to an equilibrium state must tend to reinstate it.

We formalize these two ideas as follows:

**Definition 1.** Fix a state  $\mathbf{P} \in \mathcal{X}$  and let

$$\mathcal{X}_{\mathbf{P}} := \{\mathbf{X} \in \mathcal{X} : \ker(\mathbf{X}) \leq \ker(\mathbf{P})\} \quad (18)$$

denote the *domain of consistency* of  $\mathbf{P}$  in  $\mathcal{X}$ , i.e., the set of mixed states whose linear span contains that of  $\mathbf{P}$ . Then, given a flow  $\chi: \mathbb{R} \times \mathcal{X} \rightarrow \mathcal{X}$ , we will say that:

- (1)  $\mathbf{P}$  is *consistently attracting* if it admits a neighborhood  $\mathcal{U}$  such that  $\lim_{t \rightarrow \infty} \chi_t(\mathbf{X}) = \mathbf{P}$  for all  $\mathbf{X} \in \mathcal{U} \cap \mathcal{X}_{\mathbf{P}}$ . In other words,  $\mathbf{P}$  is consistently attracting if it attracts all nearby consistent initializations.
- (2)  $\mathbf{P}$  is *consistently asymptotically stable* if it is Lyapunov stable and consistently attracting.

**Definition 2.** We say that  $\mathbf{X}^* \in \mathcal{X}$  is (*locally*) *variationally stable* if there exists a neighborhood  $\mathcal{U}$  of  $\mathbf{X}^*$  in  $\mathcal{X}$  such that

$$\text{tr}[\mathbf{V}(\mathbf{X})(\mathbf{X} - \mathbf{X}^*)] < 0 \quad \text{for all } \mathbf{X} \in \mathcal{U} \setminus \{\mathbf{X}^*\}. \quad (\text{VS})$$

Intuitively, [Definition 1](#) captures precisely the accessibility condition that we discussed above:  $\mathcal{X}_{\mathbf{P}}$  is a dense convex set consisting of the relative interior of all faces of  $\mathcal{X}$  that contain  $\mathbf{P}$  (including  $\mathcal{X}$  itself). As for [Definition 2](#), variational stability should be seen as an equilibrium refinement in the spirit of the seminal concept of *evolutionary stability* [28, 29] that underlies the “folk theorem” for finite games; for a detailed discussion of (VS) in the context of continuous games, see [31].

*Remark.* In [Definition 1](#), the notion of Lyapunov stability does not have a “consistency” caveat tacked on. As we discuss in [Appendix D](#), the reason for this is that, in the case of Lyapunov stability, the two notions end up coinciding, so it is not necessary to make this distinction.  $\diamond$

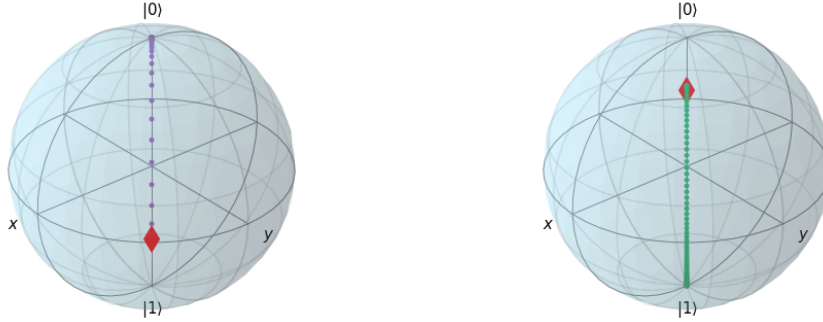
**A quantum “folk theorem”.** We are now in a position to state the main result of this section. To simplify the presentation, we assume below that (FTQL) is run with a steep regularizer and, as per [Theorem 1](#), the quantum state dynamics (QD) refer to the flow induced by (FTQL) on  $\mathcal{X}$ .

**Theorem 2.** Let  $\mathcal{Q} \equiv \mathcal{Q}(\mathcal{N}, \Psi, u)$  be a quantum game, fix some state  $\mathbf{X}^* \in \mathcal{X}$ , and let  $\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t))$  be a trajectory of play induced by (FTQL) with steep regularizers. Then:

- (1) If  $\mathbf{X}^*$  is a Nash equilibrium, it is a rest point of (QD).
- (2) If  $\mathbf{X}(t) \rightarrow \mathbf{X}^*$  as  $t \rightarrow \infty$ ,  $\mathbf{X}^*$  is a Nash equilibrium of  $\mathcal{Q}$ .
- (3) If  $\mathbf{X}^*$  is stable under (QD), it is a Nash equilibrium of  $\mathcal{Q}$ .
- (4) If  $\mathbf{X}^*$  is consistently asymptotically stable, then it is pure.
- (5) If  $\mathbf{X}^*$  satisfies (VS), it is consistently asymptotically stable.

Before discussing the proof of [Theorem 2](#), some remarks are in order. Perhaps the most important one concerns the asymptotic stability part of the theorem (which is arguably the most salient point of the classical folk theorem as well). Here, even though the standard notion of asymptotic stability is ruled out by the geometry of the game’s state space, (FTQL) achieves the next best thing: by definition, states that are consistently asymptotically stable attract all but a measure zero of nearby initial conditions, and [Theorem 2](#) shows that *only* pure states can have this property. This selection result has important implications for quantum games because it shows that regularized learning essentially “collapses” an initial mixed state to a *specific* pure state – and this, despite the fact that any mixed state can be prepared by an infinitum of combinations of pure states.

On the flip side of all this, the implication that variationally stable states are also (consistently) asymptotically stable provides a relevant convergence criterion for (FTQL) and indicates an inherent robustness to variations of player beliefs and predictions. In particular,



**Figure 2:** Convergence to a variationally stable equilibrium in a quantum anti-coordination game. Each player’s trajectory is represented in their individual Bloch sphere; see [Appendix F](#) for the details.

since (VS) only involves the primitives of the underlying game, the fact that such states are attracting under *all* FTQL dynamics means that they can be seen as universal attractors – and since only pure states can have this property, we also infer indirectly that variationally stable states are *a fortiori* pure.

**Proof sketch.** We conclude this section with a short outline of the proof of the last two parts of [Theorem 2](#) (the full proof is presented in [Appendix D](#)). First, the implication that only pure states can be stable attractors hinges on the following volume preservation principle:

**Proposition 2.** *Let  $\mathcal{W} \subseteq \mathcal{Y}$  be an open set of initial conditions of (FTQL), and let  $\mathcal{W}_t = \psi_t(\mathcal{W})$ ,  $t \geq 0$ , denote the evolution of  $\mathcal{W}$  under the flow  $\psi: \mathbb{R} \times \mathcal{Y} \rightarrow \mathcal{Y}$  of (FTQL). Then,  $\text{vol}(\mathcal{W}_t) = \text{vol}(\mathcal{W})$  for all  $t \geq 0$ .*

[Proposition 2](#) is central to our work in several aspects. First, this volume-preservation property applies to *all* quantum games and *all* FTQL dynamics, generalizing in this way the very recent analysis of Jain et al. [23] for (MMW) / (QRD) in two-player, zero-sum games. This degree of generality is due to the fact that each player’s payoff function is individually linear in the player’s own density matrix, so the individual gradient fields  $\mathbf{V}_i$  do not depend on  $\mathbf{X}_i$ ; [Proposition 2](#) then follows by an application of Liouville’s formula.

In this regard, [Proposition 2](#) can be seen as a quantum analogue of the universal volume-preservation property of FTRL dynamics in *finite* games [9, 17]. However, as in the classical world, it is worth noting that [Proposition 2](#) concerns the flow of (FTQL) on  $\mathcal{Y}$ , not the induced flow on  $\mathcal{X}$ ; in particular,  $\mathbf{Q}(\mathcal{W}_t)$  could still collapse to a boundary state of  $\mathcal{X}$  if  $\mathcal{W}_t$  escapes to infinity. Because of this, although [Proposition 2](#) precludes convergence to full-rank equilibria relatively easily, excluding lower-rank equilibria requires more delicate arguments, where the notion of consistency plays a major role (and has no classical counterpart).

Finally, the last part of [Theorem 2](#) leverages an energy argument in the spirit of Lyapunov’s direct method. In tune with the primal-dual nature of (FTQL), a natural choice of energy function is the Fenchel coupling  $E(t) = F(\mathbf{X}^*, \mathbf{Y}(t))$  of (15), which is in turn linked to (VS) via [Lemma 1](#). Putting these two elements together readily yields that  $\lim_{t \rightarrow \infty} E(t)$  exists; then, by a trapping argument, it can be shown that there exists a sequence of times  $t_n \rightarrow \infty$  such that  $\mathbf{X}(t_n) \rightarrow \mathbf{X}^*$ , which allows us to conclude that  $E(t) \rightarrow 0$  and ultimately yields our claim.

## 6. POINCARÉ RECURRENCE IN ZERO-SUM GAMES

In this last section, our aim is to study the long-run behavior of (FTQL) in zero-sum games, a class of games that arise frequently in applications – and is possibly the most widely studied family of quantum games.

In classical finite games, it is well known that the replicator dynamics – and, more generally, all FTRL dynamics – produce trajectories that cycle around interior equilibria, leading to a behavior known as *Poincaré recurrence* [32, 35]. Formally, in our setting, the dynamics (FTQL) are said to be Poincaré recurrent if, for almost every initial condition  $\mathbf{X}(0) = \mathbf{Q}(\mathbf{Y}(0)) \in \mathcal{X}$  (in the sense of Lebesgue), there exists an increasing sequence of times  $t_n \rightarrow \infty$  such that  $\mathbf{X}(t_n) \rightarrow \mathbf{X}(0)$  as  $n \rightarrow \infty$ , i.e., almost all trajectories of (FTQL) return infinitely close to their starting point infinitely often.

In our notation, a two-player zero-sum quantum game  $\mathcal{Q} \equiv \mathcal{Q}(\mathcal{N}, \Psi, u)$  corresponds to the case where  $\mathcal{N} = \{1, 2\}$  and  $u_1 = -u_2 = -\mathcal{L}$  for some min-max merit function  $\mathcal{L}: \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathbb{R}$ . In this context, Jain et al. [23] recently showed that the matrix multiplicative / exponential weights dynamics (MMW) are Poincaré recurrent in all two-player, zero-sum games that admit a full-rank equilibrium. As we show below, this property extends to the *entire* class of regularized learning dynamics under study.

**Theorem 3.** *Let  $\mathcal{Q} \equiv \mathcal{Q}(\mathcal{N}, \Psi, u)$  be a 2-player zero-sum game admitting a full-rank equilibrium  $\mathbf{X}^* \in \text{ri } \mathcal{X}$ . Then, almost every solution orbit  $\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t))$  of (FTQL) is Poincaré recurrent: there exists an increasing sequence of times  $t_n \rightarrow \infty$  such that  $\lim_{n \rightarrow \infty} \mathbf{X}(t_n) = \mathbf{X}(0)$ .*

The proof of Theorem 3 comprises three main ingredients: The first is Proposition 2, which shows that the flow induced by (FTQL) on  $\mathcal{Y}$  is volume-preserving. The second is that the Fenchel coupling (15) that played the role of a local energy function in Theorem 2 becomes a *constant of motion* in zero-sum games (this is also where the full-rank assumption comes into play). The third is that, modulo a certain quotient process that does not affect the dynamics’ trajectories on  $\mathcal{X}$ , the level sets of the Fenchel coupling are bounded; this is the most challenging part of the proof, and the point where the non-commutativity of the various processes involved complicates things considerably. With these basic ingredients in place, our result follows by an application of Poincaré’s theorem, which states that volume-preserving flows with bounded trajectories are recurrent. We provide the relevant details in Appendix E.

## 7. CONCLUDING REMARKS

When quantum computing models are deployed in a multi-agent context – from autonomous vehicles to quantum GANs – the players’ interaction landscape changes dramatically relative to classical interactions. The study of game-theoretic learning in this quantum setting is still in its infancy, so it is not clear at this stage what can be expected by quantum players with bounded rationality. In this regard, the study of the FTQL dynamics provides the following important insights: the geometric structure of quantum state space leads to an inflation of “learning traps” (stationary states) that have no classical counterpart; nonetheless, the only states that can be stable and attracting under FTQL are the game’s pure quantum equilibria. Solidifying our understanding of the limits of quantum game-theoretic learning is a particularly fruitful research direction with potentially far-reaching implications for the deployment of multi-agent quantum computing systems.

## APPENDIX A. REGULARIZED BEST RESPONSES AND THE FENCHEL COUPLING

In this appendix, we introduce the basic properties of the regularized best response map  $\mathbf{Q}$  and the Fenchel coupling. To simplify notation, we will suppress player indices throughout; formally, this means that, for example,  $\mathbf{X}$  stands for the profile  $\mathbf{X} = (\mathbf{X}_i)_{i \in \mathcal{N}}$ ; the ensemble mirror map  $\mathbf{Q}: \mathcal{Y} \rightarrow \mathcal{X}$  denotes the product map  $\mathbf{Q} = \prod_i \mathbf{Q}_i$ ; the aggregate regularizer  $h$  on  $\mathcal{X}$  represents the sum  $h = \sum_i h_i$ , etc. The proper substitutions are easily inferred from the context, so there is no danger of confusion.

We will also assume that  $\mathcal{V}$  is endowed with some abstract norm  $\|\cdot\|$ , not necessarily induced by the Hilbert structure of the quantum state space  $\mathcal{H}$ . Also, for technical reasons, it will also be convenient to view  $h$  as an extended-real-valued function  $h: \mathcal{V} \rightarrow \mathbb{R} \cup \{\infty\}$  defined over the entire space  $\mathcal{V} \equiv \mathbb{H}^d$  of  $d \times d$  Hermitian matrices by assigning the value  $h(\mathbf{X}) = +\infty$  to all  $\mathbf{X} \in \mathcal{V} \setminus \mathcal{X}$ . Following Rockafellar [37], this allows us to define the *subdifferential* of  $h$  at  $\mathbf{X} \in \mathcal{X}$  as

$$\partial h(\mathbf{X}) := \{\mathbf{Y} \in \mathcal{Y} : h(\mathbf{X}') \geq h(\mathbf{X}) + \text{tr}[\mathbf{Y}(\mathbf{X}' - \mathbf{X})] \text{ for all } \mathbf{X}' \in \mathcal{V}\}, \quad (\text{A.1})$$

where  $\mathcal{Y} \equiv \mathcal{V}$  plays here the role of the algebraic dual of  $\mathcal{V}$ . The *domain of subdifferentiability* of  $h$  is then defined as

$$\text{dom } \partial h := \{\mathbf{X} \in \text{dom } h : \partial h \neq \emptyset\} \quad (\text{A.2})$$

and the *convex conjugate* of  $h$  is given by the expression

$$h^*(\mathbf{Y}) = \max_{\mathbf{X} \in \mathcal{X}} \{\text{tr}[\mathbf{Y}\mathbf{X}] - h(\mathbf{X})\} \quad \text{for all } \mathbf{Y} \in \mathcal{Y}. \quad (\text{A.3})$$

Since  $h$  is a trace function of the form  $h(\mathbf{X}) = \text{tr}[\theta(\mathbf{X})]$  with  $\theta$  strongly convex over  $[0, 1]$ , it readily follows that  $h$  is also  $K$ -strongly convex relative to  $\|\cdot\|$  for some positive constant  $K > 0$ . We then have the following basic results:

**Lemma A.1.** *Let  $h$  be a  $K$ -strongly convex regularizer on  $\mathcal{X}$  as above, and let  $\mathbf{Q}: \mathcal{Y} \rightarrow \mathcal{X}$  be its induced mirror map. Then:*

- (1)  $\mathbf{Q}$  is single-valued on  $\mathcal{Y}$ : in particular, for all  $\mathbf{X} \in \mathcal{X}$ ,  $\mathbf{Y} \in \mathcal{Y}$ , we have  $\mathbf{X} = \mathbf{Q}(\mathbf{Y}) \iff \mathbf{Y} \in \partial h(\mathbf{X})$ .
- (2) The prox-domain  $\mathcal{X}_h := \text{im } \mathbf{Q}$  of  $h$  satisfies  $\text{ri } \mathcal{X} \subseteq \mathcal{X}_h \subseteq \mathcal{X}$ .
- (3)  $\mathbf{Q}$  is  $(1/K)$ -Lipschitz continuous and  $\mathbf{Q} = \nabla h^*$ .

Variants of these properties can be found in several points in the literature – see e.g., [31] and references therein – but we provide a few quick pointers here for completeness.

*Proof.* For the first property of  $\mathbf{Q}$ , note that the maximum in (10) is attained for all  $\mathbf{Y} \in \mathcal{Y}$  because  $h$  is lower-semicontinuous (l.s.c.) and strongly convex. Furthermore, by Fermat's theorem for stationarity points,  $\mathbf{X}$  solves (10) if and only if  $\mathbf{Y} - \partial h(\mathbf{X}) \ni 0$ , i.e., if and only if  $\mathbf{Y} \in \partial h(\mathbf{X})$ . The above also shows that  $\mathcal{X}_h = \text{dom } \partial h$ ; since  $\text{ri } \mathcal{X} \subseteq \text{dom } \partial h \subseteq \mathcal{X}$  [37, Chap. 26], our second claim follows. Finally, for our third claim, the expression  $\mathbf{Q} = \nabla h^*$  is an immediate consequence of Danskin's theorem [43], while the Lipschitz continuity of  $\mathbf{Q}$  follows from standard variational analysis results, [see e.g., 38, Theorem 12.60(b)]. ■

The next properties of  $\mathbf{Q}$  are more specialized and are intimately related to the structure of the spectraplex:

**Lemma A.2.** *With notation and assumptions as in Lemma A.1, we have:*

- (1)  $\partial h(\mathbf{X}) = \theta'(\mathbf{X}) + \text{NC}(\mathbf{X})$ , where  $\text{NC}(\mathbf{X})$  denotes the normal cone to  $\mathcal{X}$  at  $\mathbf{X}$ ; in particular, for all  $\mathbf{X} \in \text{ri } \mathcal{X}$ , we have  $\partial h(\mathbf{X}) = \{\theta'(\mathbf{X}) + \lambda \mathbf{I} : \lambda \in \mathbb{R}\}$ .

- (2)  $\mathbf{Y}$  and  $\mathbf{Q}(\mathbf{Y})$  commute for all  $\mathbf{Y} \in \mathcal{Y}$ .
- (3) If  $\text{eig}(\mathbf{Y}) = \{y_\alpha\}_{\alpha=1}^d$  and  $\text{eig}(\mathbf{X}) = \{x_\alpha\}_{\alpha=1}^d$  respectively denote the eigenvalues of  $\mathbf{Y} \in \mathcal{Y}$  and  $\mathbf{X} = \mathbf{Q}(\mathbf{Y}) \in \mathcal{X}$ , we have  $x_\alpha \rightarrow 0$  whenever  $y_\alpha - y_\beta \rightarrow -\infty$  for some  $\beta \neq \alpha$ .

*Proof.* For our first claim, let  $\delta_{\mathcal{X}}$  denote the (convex) indicator of  $\mathcal{X}$ , viz.  $\delta_{\mathcal{X}}(\mathbf{X}) = 0$  for all  $\mathbf{X} \in \mathcal{X}$  and  $\delta_{\mathcal{X}}(\mathbf{X}) = \infty$  for all  $\mathbf{X} \in \mathcal{V} \setminus \mathcal{X}$ . By standard convex analysis arguments [37], we have  $\partial\delta_{\mathcal{X}}(\mathbf{X}) = \text{NC}(\mathbf{X})$  for all  $\mathbf{X} \in \mathcal{X}$ , so

$$\partial h(\mathbf{X}) = \partial[\text{tr}[\theta(\mathbf{X})] + \delta_{\mathcal{X}}(\mathbf{X})] = \theta'(\mathbf{X}) + \text{NC}(\mathbf{X}), \quad (\text{A.4})$$

and our assertion follows.

For our second claim, let  $\mathbf{Y} = \mathbf{U}\Psi\mathbf{U}^\dagger$  be an eigen-decomposition of  $\mathbf{Y}$ . Then the objective of (10) can be rewritten as

$$\begin{aligned} \text{tr}[\mathbf{Y}\mathbf{X}] - h(\mathbf{X}) &= \text{tr}[\mathbf{U}\Psi\mathbf{U}^\dagger\mathbf{X}] - \text{tr}[\theta(\mathbf{X})] = \text{tr}[\Psi\mathbf{U}^\dagger\mathbf{X}\mathbf{U}] - \text{tr}[\mathbf{U}^\dagger\theta(\mathbf{X})\mathbf{U}] \\ &= \text{tr}[\Psi\mathbf{U}^\dagger\mathbf{X}\mathbf{U}] - \text{tr}[\theta(\mathbf{U}^\dagger\mathbf{X}\mathbf{U})] \end{aligned} \quad (\text{A.5})$$

since  $\mathbf{X}$  and  $\mathbf{U}^\dagger\mathbf{X}\mathbf{U}$  have the same eigenvalues and  $\text{tr}[\theta(\cdot)]$  only depends on the eigenvalues of its argument. Since  $\mathcal{X}$  remains invariant under conjugation by any unitary matrix (i.e.,  $\mathbf{U}^\dagger\mathcal{X}\mathbf{U} = \mathcal{X}$  for every unitary  $\mathbf{U}$ ), it follows that

$$\max_{\mathbf{X} \in \mathcal{X}} \{\text{tr}[\mathbf{Y}\mathbf{X}] - h(\mathbf{X})\} = \max_{\mathbf{X} \in \mathcal{X}} \{\text{tr}[\Psi\mathbf{X} - \theta(\mathbf{X})]\} = \max_{x \in \Delta_d} \left\{ \sum_{\alpha=1}^d [y_\alpha x_\alpha - \theta(x_\alpha)] \right\} \quad (\text{A.6})$$

where, in obvious notation,  $x_\alpha$  and  $y_\alpha$  respectively denote the eigenvalues of  $\mathbf{X}$  and  $\mathbf{Y}$ . This expression shows that the maximum value of (10) is attained when  $\mathbf{Y}$  and  $\mathbf{X}$  have a common basis of eigenvectors, which in turn means that they commute. Since  $\theta$  is strongly convex by assumption, (10) admits a unique solution, and our claim follows.

Finally, for our third claim, let  $\mathbf{X}_n = \mathbf{Q}(\mathbf{Y}_n)$  for some sequence  $\mathbf{Y}_n \in \mathcal{Y}$ , and write  $x_{\alpha,n}$  and  $y_{\alpha,n}$  for the eigenvalues of  $\mathbf{X}_n$  and  $\mathbf{Y}_n$  respectively. We seek to show that  $x_{\alpha,n} \rightarrow 0$  if there exists some  $\beta$  such that  $y_{\alpha,n} - y_{\beta,n} \rightarrow -\infty$ .

Now, by descending to a subsequence if necessary, we may assume without loss of generality that there exists some  $\varepsilon > 0$  such that  $x_{\alpha,n} \geq \varepsilon > 0$  for all  $n$ . Then, by the defining relation (10) of  $\mathbf{Q}(\mathbf{Y})$ , we have

$$\sum_{\alpha=1}^d [y_{\alpha,n} x_{\alpha,n} - \theta(x_{\alpha,n})] \geq \sum_{\alpha=1}^d [y_{\alpha,n} x'_{\alpha,n} - \theta(x'_{\alpha,n})] \quad (\text{A.7})$$

for all  $x' \in \Delta_d$ . Therefore, taking  $x'_n = x_n + \varepsilon(e_\beta - e_\alpha)$ , we readily obtain

$$\varepsilon(y_{\alpha,n} - y_{\beta,n}) \geq \sum_{\kappa} [\theta(x_{\kappa,n}) - \theta(x'_{\kappa,n})] \geq \min h - \max h \quad (\text{A.8})$$

which contradicts our original assumption that  $y_{\alpha,n} - y_{\beta,n} \rightarrow -\infty$ . With  $\mathcal{X}$  compact, the above shows that  $x_\alpha^* = 0$  for any limit point  $x^*$  of  $x_n$ , as claimed.  $\blacksquare$

With all this in hand, we proceed below with the basic properties of the Fenchel coupling – so named to account for the fact that it collects all terms of Fenchel's inequality, cf. [31]. The two main properties we will need are encoded in the following proposition

**Proposition A.1.** *With notation and assumptions as in Lemma A.1, we have:*

- (1)  $F(\mathbf{P}, \mathbf{Y}) \geq (K/2) \|\mathbf{Q}(\mathbf{Y}) - \mathbf{P}\|^2$  for all  $\mathbf{P} \in \mathcal{X}$ ,  $\mathbf{Y} \in \mathcal{Y}$ .
- (2)  $\nabla_{\mathbf{Y}^\top} F(\mathbf{P}, \mathbf{Y}) = \mathbf{Q}(\mathbf{Y}) - \mathbf{P}$  for all  $\mathbf{P} \in \mathcal{X}$ ,  $\mathbf{Y} \in \mathcal{Y}$ .

*Proof.* For our first claim, let  $\mathbf{X} = \mathbf{Q}(\mathbf{Y})$  so  $h^*(\mathbf{Y}) = \text{tr}[\mathbf{Y}\mathbf{X}] - h(\mathbf{X})$  by the definition (A.3) of  $h^*$ . The strong convexity of  $h$  then yields

$$h(\mathbf{X}) + t \text{tr}[\mathbf{Y}(\mathbf{P} - \mathbf{X})] \leq h(\mathbf{X} + t(\mathbf{P} - \mathbf{X})) \leq th(\mathbf{P}) + (1-t)h(\mathbf{X}) - \frac{1}{2}Kt(1-t)\|\mathbf{X} - \mathbf{P}\|^2, \quad (\text{A.9})$$

leading to the bound

$$\frac{1}{2}K(1-t)\|\mathbf{X} - \mathbf{P}\|^2 \leq h(\mathbf{P}) - h(\mathbf{X}) - \text{tr}[\mathbf{Y}(\mathbf{P} - \mathbf{X})] = F(\mathbf{P}, \mathbf{Y}) \quad (\text{A.10})$$

for all  $t \in (0, 1]$ . Our assertion then follows by letting  $t \rightarrow 0^+$  in (A.10).

As for our second claim, Lemma A.1 readily yields

$$\nabla F(\mathbf{P}, \mathbf{Y}) = \nabla_{\mathbf{Y}^\top} h^*(\mathbf{Y}) - \nabla_{\mathbf{Y}^\top} \text{tr}[\mathbf{Y}\mathbf{P}] = \mathbf{Q}(\mathbf{Y}) - \mathbf{P} \quad (\text{A.11})$$

as claimed.  $\blacksquare$

We are now in a position to prove Lemma 1 on the evolution of the Fenchel coupling under the dynamics (FTQL); for convenience, we restate the result below.

**Lemma 1.** *Let  $\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t))$ ,  $t \geq 0$ , be a trajectory of play induced by (FTQL). Then, for all  $\mathbf{P} \in \mathcal{X}$ , we have*

$$\frac{d}{dt} F_i(\mathbf{P}_i, \mathbf{Y}_i(t)) = \text{tr}[\mathbf{V}_i(\mathbf{X}(t)) (\mathbf{X}_i(t) - \mathbf{P}_i)]. \quad (17)$$

*Proof.* By definition, we have

$$\begin{aligned} \frac{d}{dt} F_i(\mathbf{P}_i, \mathbf{Y}_i(t)) &= \frac{d}{dt} [h_i(\mathbf{P}_i) + h_i^*(\mathbf{Y}_i(t)) - \text{tr}[\mathbf{Y}_i(t)\mathbf{P}_i]] \\ &= \text{tr}[\dot{\mathbf{Y}}_i(t) (\mathbf{X}_i(t) - \mathbf{P}_i)] = \text{tr}[\mathbf{V}_i(\mathbf{X}(t)) (\mathbf{X}(t) - \mathbf{P})], \end{aligned} \quad (\text{A.12})$$

where, in the last line, we used Proposition A.1.  $\blacksquare$

The last result we will need is a ‘‘reciprocity’’ property in the spirit of [31] which shows that the convergence topology induced by  $F$  on  $\mathcal{X}$  is compatible with the ordinary one. While this property is relatively easy to establish in linear polytopes with decomposable regularizers, the matrix setting complicates things considerably.

**Proposition A.2.** *Let  $\mathbf{Y}_n$ ,  $n = 1, 2, \dots$ , be a sequence in  $\mathcal{Y}$ . Then  $\mathbf{X}_n = \mathbf{Q}(\mathbf{Y}_n)$  converges to  $\mathbf{P} \in \mathcal{X}$  if and only if  $\lim_{n \rightarrow \infty} F(\mathbf{P}, \mathbf{Y}_n) = 0$ .*

*Proof.* We begin by showing the direct implication ‘‘ $\implies$ ’’. Indeed, assuming that  $\mathbf{X}_n \rightarrow \mathbf{P}$ , the definition of  $F$  gives

$$\begin{aligned} F(\mathbf{P}, \mathbf{X}_n) &= h(\mathbf{P}) + h^*(\mathbf{Y}_n) - \text{tr}[\mathbf{Y}_n\mathbf{P}] \\ &= h(\mathbf{P}) - h(\mathbf{X}_n) + \text{tr}[\mathbf{Y}_n(\mathbf{X}_n - \mathbf{P})] \\ &= h(\mathbf{P}) - h(\mathbf{X}_n) + \text{tr}[\theta'(\mathbf{X}_n)(\mathbf{X}_n - \mathbf{P})] \end{aligned} \quad (\text{A.13})$$

where, in the last equality, we used the first part of Lemma A.2. Since  $\mathbf{X}_n \rightarrow \mathbf{P}$ , the term  $h(\mathbf{P}) - h(\mathbf{X}_n)$  vanishes as  $n \rightarrow \infty$ , so it suffices to show the same for the second term as well. To that end, if we let  $\mathbf{X}_n = \sum_{\alpha} x_{\alpha,n} \mathbf{u}_{\alpha,n} \mathbf{u}_{\alpha,n}^\dagger$  be an eigen-decomposition of  $\mathbf{X}_n$  and we set  $p_{\alpha,n} = \mathbf{u}_{\alpha,n}^\dagger \mathbf{P} \mathbf{u}_{\alpha,n}$  and  $p_{\alpha} = \lim_{n \rightarrow \infty} p_{\alpha,n}$  (the limit exists because  $\mathbf{X}_n$  converges to  $\mathbf{P}$  by



assumption), we readily get

$$\begin{aligned} \text{tr}[\theta'(\mathbf{X}_n)(\mathbf{X}_n - \mathbf{P})] &= \sum_{\alpha=1}^d \theta'(x_{\alpha,n})(x_{\alpha,n} - p_{\alpha,n}) \\ &= \sum_{\alpha \in \text{supp}(p)} \theta'(x_{\alpha,n})(x_{\alpha,n} - p_{\alpha,n}) + \sum_{\alpha: p_{\alpha}=0} \theta'(x_{\alpha,n})(x_{\alpha,n} - p_{\alpha,n}) \end{aligned} \quad (\text{A.14})$$

Since  $\mathbf{X}_n$  converges to  $\mathbf{P}$  the first sum above vanishes: this is because  $\lim_{n \rightarrow \infty} x_{\alpha,n} > 0$ , so  $\theta'(x_{\alpha,n})$  has a finite limit for all  $\alpha \in \text{supp}(p)$ . As for the second, since  $x_{\alpha,n} - p_{\alpha,n} \leq x_{\alpha,n}$ , it suffices to show that  $z\theta'(z) \rightarrow 0$  as  $z \rightarrow 0^+$ ; we prove that this is so in [Lemma A.3](#) below, for  $f = \theta'$ .

Finally, as for the converse implication “ $\Leftarrow$ ”, our assertion follows immediately from the first part of [Proposition A.1](#).  $\blacksquare$

**Lemma A.3.** *Let  $f : (0, 1] \rightarrow \mathbb{R}$  be differentiable such that (i)  $\lim_{x \rightarrow 0^+} f(x) = -\infty$ , (ii)  $\inf_{x \in (0, 1]} f'(x) > 0$ , and (iii)  $\int_0^1 |f(x)| dx < \infty$ . Then,  $\lim_{x \rightarrow 0^+} xf(x) = 0$ .*

*Proof.* For convenience, let  $g = -f$ . Then, we readily get (i)  $\lim_{x \rightarrow 0^+} g(x) = +\infty$ , (ii)  $\inf_{x \in (0, 1]} g'(x) < 0$ , and (iii)  $\int_0^1 |g(x)| dx < \infty$ . Since  $\inf_{x \in (0, 1]} g'(x) < 0$ , we get that  $g$  is strictly decreasing. Now,  $\lim_{x \rightarrow 0^+} g(x) = +\infty$  implies that  $g$  is strictly positive close to 0, i.e.,  $g(x) > 0$  in  $(0, c)$  for some  $c > 0$ . Let  $\varepsilon \in (0, c)$ . Since  $g$  is decreasing and positive in  $(0, c)$ , we have:

$$0 \leq \varepsilon g(\varepsilon) \leq \int_0^\varepsilon g(x) dx \quad (\text{A.15})$$

We will now show that  $\lim_{\varepsilon \rightarrow 0^+} \int_0^\varepsilon g(x) dx = 0$ . For this, we can write  $\int_0^\varepsilon g(x) dx$  as:

$$\int_0^\varepsilon g(x) dx = \int \mathbb{1}_{(0, \varepsilon)} g(x) dx \quad (\text{A.16})$$

where  $\mathbb{1}_A(x) = 1$  if  $x \in A$  and 0, otherwise. Let  $\{\varepsilon_n\}_{n \in \mathbb{N}}$  be a sequence of positive reals with  $\varepsilon_n \rightarrow 0$  as  $n \rightarrow \infty$ . Then, it holds:

$$\mathbb{1}_{(0, \varepsilon_n)} g(x) \xrightarrow{n \rightarrow \infty} 0 \quad \text{for all } x \in (0, 1] \quad (\text{A.17})$$

since, for fixed  $x \in (0, 1]$ , we have that  $\varepsilon_n < x$  for all  $n$  large enough, which implies that  $\mathbb{1}_{(0, \varepsilon_n)} g(x) = 0$ . Moreover, for all  $x \in (0, 1]$ , it holds:

$$|\mathbb{1}_{(0, \varepsilon_n)} g(x)| \leq |g(x)| \quad (\text{A.18})$$

with  $\int_0^1 |g(x)| dx < \infty$ . Hence, by the dominated convergence theorem [\[18\]](#), we get that:

$$\lim_{n \rightarrow \infty} \int_0^1 \mathbb{1}_{(0, \varepsilon_n)} g(x) dx = 0 \quad (\text{A.19})$$

and, since  $\{\varepsilon_n\}_{n \in \mathbb{N}}$  was arbitrary, we conclude that  $\lim_{\varepsilon \rightarrow 0^+} \int_0^\varepsilon g(x) dx = 0$ . Hence, combining it with [\(A.15\)](#), we get that  $\lim_{x \rightarrow 0^+} xg(x) = 0$ , i.e.,  $\lim_{x \rightarrow 0^+} xf(x) = 0$ .  $\blacksquare$

## APPENDIX B. GENERAL PROPERTIES OF THE DYNAMICS

**Theorem 1.** *Let  $\mathbf{X}(t) = \sum_{\alpha=1}^d x_\alpha(t) \mathbf{u}_\alpha(t) \mathbf{u}_\alpha^\dagger(t)$  be an eigendecomposition of  $\mathbf{X}(t)$  as per [\(11\)](#), and suppose that  $h$  is steep. Then, under [\(FTQL\)](#), the entries  $[\dot{\mathbf{X}}]_{\alpha\beta} = \mathbf{u}_\alpha^\dagger \dot{\mathbf{X}} \mathbf{u}_\beta$  of  $\dot{\mathbf{X}}$  follow*

the quantum state dynamics:

$$[\dot{\mathbf{X}}]_{\alpha\beta} = \begin{cases} \frac{V_{\alpha\alpha}}{\theta''(x_\alpha)} - \frac{\sum_{\kappa} V_{\kappa\kappa}/\theta''(x_\kappa)}{\sum_{\kappa} \theta''(x_\alpha)/\theta''(x_\kappa)} & \alpha = \beta \\ \frac{x_\beta - x_\alpha}{\theta'(x_\beta) - \theta'(x_\alpha)} V_{\alpha\beta} & \alpha \neq \beta \end{cases} \quad (\text{QD})$$

where  $V_{\alpha\beta} = \mathbf{u}_\alpha^\dagger \mathbf{V}(\mathbf{X}) \mathbf{u}_\beta$  and we are using the continuity convention  $(x - y)/[\theta'(x) - \theta'(y)] = 1/\theta''(x)$  when  $y \rightarrow x$ .

*Proof.* First of all, according to the dynamics described in Section 3,  $\mathbf{X}(t)$  is obtained as a regularized best-response, i.e., a solution of the maximization problem:

$$\mathbf{Q}(\mathbf{Y}) = \arg \max_{\mathbf{X}' \in \mathcal{X}} \{\text{tr}[\mathbf{Y}\mathbf{X}'] - \text{tr}[\theta(\mathbf{X}')]\}. \quad (\text{B.1})$$

Letting  $\mathbf{X}(t) = \sum_{\alpha=1}^d x_\alpha(t) \mathbf{u}_\alpha(t) \mathbf{u}_\alpha^\dagger(t)$  be an eigendecomposition of  $\mathbf{X}(t)$ , since  $h$  is steep, we readily obtain that  $\mathbf{X}(t) \in \text{ri } \mathcal{X}$ , which implies that  $x_\alpha(t) > 0$  for all  $\alpha$ . Since  $\inf_{x \in (0,1]} \theta''(x) > 0$ , the function  $x \mapsto \theta(x)$  is strictly convex, and so is  $\mathbf{X} \mapsto \text{tr}[\theta(\mathbf{X})]$ , see [6, Theorem 2.10]. Hence, (B.1) has a unique solution in  $\mathcal{X}$ .

By the KKT conditions, the dual variables associated with the positive semi-definiteness constraints,  $x_\alpha(t) \geq 0$  for  $\alpha = 1, \dots, d$ , are equal to zero in the optimal solution, since the inequalities are strict, as argued before. Hence, it is enough to consider the “reduced” Lagrangian:

$$\mathcal{L}(\mathbf{X}; \lambda) = \text{tr}[\mathbf{Y}\mathbf{X}] - \text{tr}[\theta(\mathbf{X})] - \lambda(\text{tr}[\mathbf{X}] - 1) \quad (\text{B.2})$$

where  $\lambda \in \mathbb{R}$  is the dual variable associated with the constraint  $\text{tr}[\mathbf{X}] = 1$ . Then, differentiating  $\mathcal{L}$  with respect to  $\mathbf{X}^\top$ , the solutions need to satisfy  $\nabla_{\mathbf{X}^\top} \mathcal{L}(\mathbf{X}; \lambda) = \mathbf{Y} - \theta'(\mathbf{X}) - \lambda \mathbf{I} = 0$ , i.e.,

$$\theta'(\mathbf{X}) = \mathbf{Y} - \lambda \mathbf{I} \quad (\text{B.3})$$

where we used that  $\nabla_{\mathbf{X}^\top} \text{tr}[\theta(\mathbf{X})] = \theta'(\mathbf{X})$ , see [34]. Differentiating (B.3) with respect to  $t$ , and invoking that  $\dot{\mathbf{Y}} = \mathbf{V}$ , we obtain:

$$\begin{aligned} \frac{d}{dt} \theta'(\mathbf{X}) &= \dot{\mathbf{Y}} - \dot{\lambda} \mathbf{I} \\ &= \mathbf{V} - \dot{\lambda} \mathbf{I} \end{aligned} \quad (\text{B.4})$$

Writing  $\theta'(\mathbf{X})$  in the same eigenbasis as  $\mathbf{X}$ , i.e.,  $\theta'(\mathbf{X}) = \sum_{\alpha=1}^d \theta'(x_\alpha) \mathbf{u}_\alpha \mathbf{u}_\alpha^\dagger$ , the above equation becomes:

$$\sum_{k=1}^d \theta''(x_k) \dot{x}_k \mathbf{u}_k \mathbf{u}_k^\dagger + \sum_{k=1}^d \theta'(x_k) \dot{\mathbf{u}}_k \mathbf{u}_k^\dagger + \sum_{k=1}^d \theta'(x_k) \mathbf{u}_k \dot{\mathbf{u}}_k^\dagger = \mathbf{V} - \dot{\lambda} \mathbf{I} \quad (\text{B.5})$$

Applying  $\mathbf{u}_\alpha^\dagger$  on the left and  $\mathbf{u}_\beta$  on the right of (B.5), we obtain:

$$\sum_{k=1}^d \theta''(x_k) \dot{x}_k \mathbf{u}_\alpha^\dagger \mathbf{u}_k \mathbf{u}_k^\dagger \mathbf{u}_\beta + \sum_{k=1}^d \theta'(x_k) \mathbf{u}_\alpha^\dagger \dot{\mathbf{u}}_k \mathbf{u}_k^\dagger \mathbf{u}_\beta + \sum_{k=1}^d \theta'(x_k) \mathbf{u}_\alpha^\dagger \mathbf{u}_k \dot{\mathbf{u}}_k^\dagger \mathbf{u}_\beta = \mathbf{u}_\alpha^\dagger \mathbf{V} \mathbf{u}_\beta - \dot{\lambda} \mathbf{u}_\alpha^\dagger \mathbf{u}_\beta \quad (\text{B.6})$$

and using that  $\mathbf{u}_k^\dagger \mathbf{u}_\ell = \delta_{k\ell}$ , the above relation becomes:

$$\theta''(x_\alpha) \dot{x}_\alpha \delta_{\alpha\beta} + \theta'(x_\beta) \mathbf{u}_\alpha^\dagger \dot{\mathbf{u}}_\beta + \theta'(x_\alpha) \dot{\mathbf{u}}_\alpha^\dagger \mathbf{u}_\beta = \mathbf{V}_{\alpha\beta} - \dot{\lambda} \delta_{\alpha\beta} \quad (\text{B.7})$$

Now, we observe that since  $\mathbf{u}_\alpha^\dagger \mathbf{u}_\beta = \delta_{\alpha\beta}$ , differentiating it with respect to  $t$ , we get:

$$\dot{\mathbf{u}}_\alpha^\dagger \mathbf{u}_\beta + \mathbf{u}_\alpha^\dagger \dot{\mathbf{u}}_\beta = 0 \quad (\text{B.8})$$

and, hence, (B.7) becomes:

$$\theta''(x_\alpha)\dot{x}_\alpha\delta_{\alpha\beta} + (\theta'(x_\beta) - \theta'(x_\alpha))\mathbf{u}_\alpha^\dagger\dot{\mathbf{u}}_\beta = \mathbf{V}_{\alpha\beta} - \dot{\lambda}\delta_{\alpha\beta} \quad (\text{B.9})$$

With the above equation in hand, we proceed to the final steps of the proof. Following the same procedure as before, the  $\alpha\beta$ -entry of  $\dot{\mathbf{X}}$ , can be written as:

$$[\dot{\mathbf{X}}]_{\alpha\beta} = \dot{x}_\alpha\delta_{\alpha\beta} + (x_\beta - x_\alpha)\mathbf{u}_\alpha^\dagger\dot{\mathbf{u}}_\beta \quad (\text{B.10})$$

(1) For  $\alpha = \beta$ , equation (B.10) gives:

$$[\dot{\mathbf{X}}]_{\alpha\alpha} = \dot{x}_\alpha \quad (\text{B.11})$$

and, equation (B.9) becomes:

$$\theta''(x_\alpha)\dot{x}_\alpha = \mathbf{V}_{\alpha\alpha} - \dot{\lambda} \quad (\text{B.12})$$

Since  $x_\alpha > 0$ , by the hypothesis on  $\theta$ , we have  $\theta''(x_\alpha) > 0$ , and therefore:

$$\dot{x}_\alpha = \frac{\mathbf{V}_{\alpha\alpha}}{\theta''(x_\alpha)} - \frac{\dot{\lambda}}{\theta''(x_\alpha)} \quad (\text{B.13})$$

Summing the above for  $\beta = 1, \dots, d$ , and using the fact that  $\sum_{\beta=1}^d \dot{x}_\beta = 0$  (since  $\sum_{\beta=1}^d x_\beta = 1$ ), we obtain:

$$\dot{\lambda} = \frac{\sum_{\beta=1}^d \mathbf{V}_{\beta\beta}/\theta''(x_\beta)}{\sum_{\beta=1}^d 1/\theta''(x_\beta)} \quad (\text{B.14})$$

Hence, combining (B.14) with (B.11) and (B.13), we obtain:

$$[\dot{\mathbf{X}}]_{\alpha\alpha} = \frac{\mathbf{V}_{\alpha\alpha}}{\theta''(x_\alpha)} - \frac{\sum_{\beta} \mathbf{V}_{\beta\beta}/\theta''(x_\beta)}{\sum_{\beta} \theta''(x_\alpha)/\theta''(x_\beta)} \quad (\text{B.15})$$

(2) For  $\alpha \neq \beta$ , equation (B.10) gives:

$$[\dot{\mathbf{X}}]_{\alpha\beta} = (x_\beta - x_\alpha)\mathbf{u}_\alpha^\dagger\dot{\mathbf{u}}_\beta \quad (\text{B.16})$$

and, equation (B.9) becomes:

$$(\theta'(x_\beta) - \theta'(x_\alpha))\mathbf{u}_\alpha^\dagger\dot{\mathbf{u}}_\beta = \mathbf{V}_{\alpha\beta} \quad (\text{B.17})$$

Expressing  $\dot{\mathbf{u}}_\beta$  in the basis  $\mathbf{u}_1, \dots, \mathbf{u}_d$ , we write it as  $\dot{\mathbf{u}}_\beta = \sum_{k=1}^d B_{\beta k}\mathbf{u}_k$  with the coefficients  $B_{\beta k}$ 's to be determined. Hence, (B.17) can be written as:

$$(\theta'(x_\beta) - \theta'(x_\alpha)) \sum_{k=1}^d B_{\beta k}\mathbf{u}_\alpha^\dagger\mathbf{u}_k = \mathbf{V}_{\alpha\beta} \quad (\text{B.18})$$

and, since  $\mathbf{u}_k^\dagger\mathbf{u}_\ell = \delta_{k\ell}$ , we readily get:

$$(\theta'(x_\beta) - \theta'(x_\alpha))B_{\beta\alpha} = \mathbf{V}_{\alpha\beta} \quad (\text{B.19})$$

or, equivalently:

$$B_{\beta\alpha} = \frac{\mathbf{V}_{\alpha\beta}}{\theta'(x_\beta) - \theta'(x_\alpha)} \quad (\text{B.20})$$

Now, it is easy to see that  $\mathbf{u}_\alpha^\dagger\dot{\mathbf{u}}_\beta = B_{\beta\alpha}$ . Therefore, combining it with (B.16) and (B.20), we conclude that:

$$[\dot{\mathbf{X}}]_{\alpha\beta} = \frac{x_\beta - x_\alpha}{\theta'(x_\beta) - \theta'(x_\alpha)} \mathbf{V}_{\alpha\beta} \quad (\text{B.21})$$

This concludes the proof. ■

**Proposition B.1.** *The quantum replicator dynamics (QRD) can be written in the form:*

$$\dot{\mathbf{X}} = \int_0^1 \mathbf{X}^{1-s} \mathbf{V}(\mathbf{X}) \mathbf{X}^s ds - \text{tr}[\mathbf{XV}(\mathbf{X})] \mathbf{X} \quad (\text{B.22})$$

*Proof.* In the case of the von Neumann regularizer, we obtain the replicator dynamics:

$$\dot{\mathbf{Y}} = \mathbf{V}(\mathbf{X}) \quad \mathbf{X} = \frac{\exp(\mathbf{Y})}{\text{tr}[\exp(\mathbf{Y})]} \quad (\text{B.23})$$

Hence, differentiating with respect to  $t$  and using the chain rule, we obtain:

$$\begin{aligned} \dot{\mathbf{X}} &= \frac{1}{\text{tr}[\exp(\mathbf{Y})]} \frac{d}{dt} \exp(\mathbf{Y}) + \exp(\mathbf{Y}) \frac{d}{dt} \frac{1}{\text{tr}[\exp(\mathbf{Y})]} \\ &= \frac{1}{\text{tr}[\exp(\mathbf{Y})]} \frac{d}{dt} \exp(\mathbf{Y}) - \frac{\exp(\mathbf{Y})}{\text{tr}[\exp(\mathbf{Y})]^2} \frac{d}{dt} \text{tr}[\exp(\mathbf{Y})] \\ &= \frac{1}{\text{tr}[\exp(\mathbf{Y})]} \frac{d}{dt} \exp(\mathbf{Y}) - \frac{\exp(\mathbf{Y})}{\text{tr}[\exp(\mathbf{Y})]^2} \text{tr} \left[ \frac{d}{dt} \exp(\mathbf{Y}) \right] \end{aligned} \quad (\text{B.24})$$

By the Fréchet derivative [49], it holds:

$$\begin{aligned} \frac{d}{dt} \exp(\mathbf{Y}) &= \int_0^1 e^{(1-s)\mathbf{Y}} \dot{\mathbf{Y}} e^{s\mathbf{Y}} ds \\ &= \int_0^1 e^{(1-s)\mathbf{Y}} \mathbf{V}(\mathbf{X}) e^{s\mathbf{Y}} ds \\ &= \text{tr}[\exp(\mathbf{Y})] \int_0^1 \mathbf{X}^{1-s} \mathbf{V}(\mathbf{X}) \mathbf{X}^s ds \end{aligned} \quad (\text{B.25})$$

and, using (B.25) in (B.24), we get:

$$\begin{aligned} \dot{\mathbf{X}} &= \frac{1}{\text{tr}[\exp(\mathbf{Y})]} \text{tr}[\exp(\mathbf{Y})] \int_0^1 \mathbf{X}^{1-s} \mathbf{V}(\mathbf{X}) \mathbf{X}^s ds \\ &\quad - \frac{\exp(\mathbf{Y})}{\text{tr}[\exp(\mathbf{Y})]^2} \text{tr} \left[ \text{tr}[\exp(\mathbf{Y})] \int_0^1 \mathbf{X}^{1-s} \mathbf{V}(\mathbf{X}) \mathbf{X}^s ds \right] \\ &= \int_0^1 \mathbf{X}^{1-s} \mathbf{V}(\mathbf{X}) \mathbf{X}^s ds - \frac{\exp \mathbf{Y}}{\text{tr}[\exp(\mathbf{Y})]} \text{tr} \left[ \int_0^1 \mathbf{X}^{1-s} \mathbf{V}(\mathbf{X}) \mathbf{X}^s ds \right] \\ &= \int_0^1 \mathbf{X}^{1-s} \mathbf{V}(\mathbf{X}) \mathbf{X}^s ds - \frac{\exp \mathbf{Y}}{\text{tr}[\exp(\mathbf{Y})]} \int_0^1 \text{tr}[\mathbf{X}^{1-s} \mathbf{V}(\mathbf{X}) \mathbf{X}^s] ds \\ &= \int_0^1 \mathbf{X}^{1-s} \mathbf{V}(\mathbf{X}) \mathbf{X}^s ds - \frac{\exp \mathbf{Y}}{\text{tr}[\exp(\mathbf{Y})]} \int_0^1 \text{tr}[\mathbf{XV}(\mathbf{X})] ds \\ &= \int_0^1 \mathbf{X}^{1-s} \mathbf{V}(\mathbf{X}) \mathbf{X}^s ds - \text{tr}[\mathbf{XV}(\mathbf{X})] \mathbf{X} \end{aligned} \quad (\text{B.26})$$

as asserted. ■

### APPENDIX C. REGRET MINIMIZATION

Our aim in this appendix is to prove that the dynamics (FTQL) incur at most constant regret. For convenience, we restate the relevant result below.

**Proposition 1.** *Let  $\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t))$ ,  $t \geq 0$ , be a trajectory of play induced by (FTQL). Then, for all  $T \geq 0$ , we have*

$$\text{Reg}_i(T) \leq |d_i \cdot \theta_i(1/d_i) - \theta_i(1)|. \quad (14)$$

Our proof essentially follows the analysis of [25]; the specialization of the proof occurs in deriving the exact value of the regret bound (14) but, for completeness, we provide the entire proof.

*Proof.* Let  $\mathbf{P}_i \in \mathcal{X}$  be a best fixed action in hindsight, i.e.,:

$$\mathbf{P}_i \in \arg \max_{\mathbf{X}'_i \in \mathcal{X}_i} \int_0^T u_i(\mathbf{X}'_i; \mathbf{X}_{-i}(t)) dt \quad (C.1)$$

and define the function  $E_i(t) := F_i(\mathbf{P}_i, \mathbf{Y}_i(t))$  for  $t \geq 0$ . By Lemma 1, we have that:

$$\frac{d}{dt} E_i(t) = \text{tr}[\mathbf{V}_i(\mathbf{X}(t))(\mathbf{X}_i(t) - \mathbf{P}_i)] = u_i(\mathbf{X}(t)) - u_i(\mathbf{P}_i; \mathbf{X}_{-i}(t)) \quad (C.2)$$

Integrating over time, and combining (8) with (C.1), we obtain:

$$\begin{aligned} \text{Reg}_i(T) &= E_i(0) - E_i(T) \\ &= h_i^*(\mathbf{Y}_i(0)) - \text{tr}[\mathbf{P}_i \mathbf{Y}_i(0)] - h_i^*(\mathbf{Y}_i(T)) + \text{tr}[\mathbf{P}_i \mathbf{Y}_i(T)] \\ &\leq h_i^*(\mathbf{Y}_i(0)) - \text{tr}[\mathbf{P}_i \mathbf{Y}_i(0)] + h_i(\mathbf{P}_i) \\ &= h_i(\mathbf{P}_i) - h_i(\mathbf{Q}_i(\mathbf{Y}_i(0))) \\ &= \text{tr}[\theta_i(\mathbf{P}_i)] - \text{tr}[\theta_i(\mathbf{Q}_i(\mathbf{Y}_i(0)))] \\ &\leq \max_{\mathbf{X}'_i \in \mathcal{X}_i} \text{tr}[\theta_i(\mathbf{X}'_i)] - \min_{\mathbf{X}'_i \in \mathcal{X}_i} \text{tr}[\theta_i(\mathbf{X}'_i)] \end{aligned} \quad (C.3)$$

where we used the fact that

$$h_i^*(\mathbf{Y}_i) = \max_{\mathbf{X}'_i \in \mathcal{X}_i} \{\text{tr}[\mathbf{X}'_i \mathbf{Y}_i] - h_i(\mathbf{X}'_i)\} \geq \text{tr}[\mathbf{P}_i \mathbf{Y}_i] - h_i(\mathbf{P}_i) \quad (C.4)$$

with equality if and only if  $\mathbf{X}'_i = \mathbf{Q}_i(\mathbf{Y}_i)$ . The precise bound (14) then follows by noting that the maximum difference in the values of  $h$  is attained between the barycenter of the spectraplex and any pure density matrix of rank 1, which thus yields the bound

$$\max h_i - \min h_i = |d_i \cdot \theta_i(1/d_i) - \theta_i(1)|, \quad (C.5)$$

and our assertion follows.  $\blacksquare$

#### APPENDIX D. THE FOLK THEOREM

In this appendix, our aim is to prove Theorem 2. To lay the necessary groundwork, we begin with a series of helper lemmas. The first is related to the verification parts of Theorem 2, and concerns the behavior of the dynamics near its trapping regions.

**Lemma D.1.** *Fix some  $\mathbf{X}^* \in \mathcal{X}$ , and assume that every neighborhood  $\mathcal{U}$  of  $\mathbf{X}^*$  in  $\mathcal{X}$  admits a solution trajectory  $\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t))$  of (FTQL) such that  $\mathbf{X}(t) \in \mathcal{U}$  for all  $t \geq 0$ . Then  $\mathbf{X}^*$  is a Nash equilibrium.*

*Proof.* We argue by contradiction. Indeed, if  $\mathbf{X}^*$  is not Nash, then, by the variational characterization (VI) of Nash equilibria, there exists some deviation  $\mathbf{P} \in \mathcal{X}$  such that  $\text{tr}[\mathbf{V}(\mathbf{X}^*)(\mathbf{P} - \mathbf{X}^*)] > 0$ . Then, by continuity, there exists a sufficiently small compact neighborhood  $\mathcal{C}$  of  $\mathbf{X}^*$  in  $\mathcal{X}$  such that  $\text{tr}[\mathbf{V}(\mathbf{X})(\mathbf{P} - \mathbf{X})] > 0$  for all  $\mathbf{X} \in \mathcal{C}$ , implying in turn that  $c := \inf_{\mathbf{X} \in \mathcal{C}} \text{tr}[\mathbf{V}(\mathbf{X})(\mathbf{P} - \mathbf{X})] > 0$ . Now, by assumption, there exists a solution orbit

$\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t))$  of (FTQL) such that  $\mathbf{X}(t) \in \mathcal{U}$  for all  $t \geq 0$  so, by Lemma 1, we readily obtain

$$\frac{d}{dt}F(\mathbf{P}, \mathbf{Y}(t)) = \text{tr}[\mathbf{V}(\mathbf{X}(t))(\mathbf{X}(t) - \mathbf{P})] \leq -c < 0 \quad \text{for all } t \geq 0. \quad (\text{D.1})$$

A simple integration then yields  $F(\mathbf{P}, \mathbf{Y}(t)) \rightarrow -\infty$  as  $t \rightarrow \infty$ , a contradiction that proves our claim.  $\blacksquare$

The second intermediate result we will need is a uniform convergence lemma for states that are consistently asymptotically stable:

**Lemma D.2.** *Suppose that  $\mathbf{X}^* \in \text{ri } \mathcal{X}$  is consistently asymptotically stable under (FTQL). Then, for every sufficiently small compact neighborhood  $\mathcal{U}$  of  $\mathbf{X}^*$  in  $\text{ri } \mathcal{X}$ , we have*

$$\lim_{t \rightarrow \infty} \sup_{\mathbf{X} \in \mathcal{U}} \|\chi_t(\mathbf{X}) - \mathbf{X}^*\| = 0. \quad (\text{D.2})$$

*Proof.* First of all, since  $\mathbf{X}^* \in \text{ri } \mathcal{X}$ , we have that  $\ker(\mathbf{X}^*) = \{0\}$ , which implies that  $\mathcal{X}_{\mathbf{X}^*} = \text{ri } \mathcal{X}$ . Hence, for any open set  $O$  in  $\text{ri } \mathcal{X}$ , it holds that  $O \cap \mathcal{X}_{\mathbf{X}^*} = O$ . Now, let  $\mathcal{U}_0$  be the basin of attraction of  $\mathbf{X}^*$ , according to the definition of the consistent asymptotic stability in Definition 1, and let  $\mathcal{U} \subseteq \mathcal{U}_0$  be a compact neighborhood of  $\mathbf{X}^*$ . Suppose, for the sake of contradiction, that  $\sup_{\mathbf{X} \in \mathcal{U}} \|\chi_t(\mathbf{X}) - \mathbf{X}^*\| \not\rightarrow 0$ . This implies that there exists  $\varepsilon > 0$ , a sequence  $\{t_n\}_{n \in \mathbb{N}}$  with  $t_n \rightarrow \infty$  as  $n \rightarrow \infty$ , and  $\mathbf{X}_n \in \mathcal{U}$  for  $n \in \mathbb{N}$ , such that

$$\|\chi_{t_n}(\mathbf{X}_n) - \mathbf{X}^*\| \geq \varepsilon \quad \text{for all } n \in \mathbb{N} \quad (\text{D.3})$$

Since  $\mathcal{U}$  is compact, we may assume (by taking subsequences, if necessary) that  $\mathbf{X}_n$  converges to some limit point  $\mathbf{X}_\infty \in \mathcal{U}$ . Now, since  $\mathbf{X}^*$  is Lyapunov stable, there exists a neighborhood  $\mathcal{U}'$  of  $\mathbf{X}^*$ , such that the trajectory  $\chi_t(\mathbf{X})$  remains within  $\varepsilon/2$ -distance of  $\mathbf{X}^*$ , if  $\mathbf{X} \in \mathcal{U}'$ , or, equivalently:

$$\|\chi_t(\mathbf{X}) - \mathbf{X}^*\| < \frac{\varepsilon}{2} \quad \text{for all } t \geq 0 \text{ and } \mathbf{X} \in \mathcal{U}' \quad (\text{D.4})$$

Define the hitting time  $\tau := \inf\{t \geq 0 : \chi_t(\mathbf{X}_\infty) \in \mathcal{U}'\}$ , as the first time that the trajectory enters  $\mathcal{U}'$  when starting from  $\mathbf{X}_\infty$ . It is easy to see that  $\tau < \infty$ , since  $\mathbf{X}_\infty \in \mathcal{U} \subseteq \mathcal{U}_0$ , and, thus,  $\|\chi_t(\mathbf{X}_\infty) - \mathbf{X}^*\| \rightarrow 0$  as  $t \rightarrow \infty$  by the asymptotic stability of  $\mathbf{X}^*$ .

By continuity of  $\chi$ , we readily get that there exists a neighborhood  $\mathcal{D}$  of  $\mathbf{X}_\infty$  such that  $\chi_\tau(\mathcal{D}) \subseteq \mathcal{U}''$ , where  $\mathcal{U}''$  is a neighborhood of  $\mathbf{X}^*$  with  $\|\chi_t(\mathbf{X}) - \mathbf{X}^*\| < \varepsilon$  for all  $t \geq 0$  and  $\mathbf{X} \in \mathcal{U}''$ .

Since  $\mathbf{X}_n \rightarrow \mathbf{X}_\infty$ , we conclude that  $\mathbf{X}_n \in \mathcal{D}$  for all sufficiently large  $n$ , which, in turn, implies that  $\chi_\tau(\mathbf{X}_n) \in \mathcal{U}''$ . Moreover, by definition of the sequence  $\{t_n\}_{n \in \mathbb{N}}$ , we have that  $t_n \rightarrow \infty$ , which gives that  $t_n > \tau$  for  $n$  sufficiently large, since  $\tau < \infty$ , as argued before.

Therefore, for all  $n$  sufficiently large, we have  $\|\chi_{t_n}(\mathbf{X}_n) - \mathbf{X}^*\| < \varepsilon$ , which contradicts (D.3).  $\blacksquare$

Our last step before proving Theorem 2 is a comparison between the notions of Lyapunov stability induced by the flow  $\psi$  on  $\mathcal{X}$  via  $\mathbf{Q}$ , and the corresponding notion relative to the conjugate flow  $\chi$  on  $\mathcal{X}$ . Formally, we have the following definition.

**Definition 3.** We say that  $\mathbf{X}^* \in \mathcal{X}$  is:

- (1)  $\chi$ -stable, if for all neighborhoods  $\mathcal{U}$  of  $\mathbf{X}^* \in \mathcal{X}$ , there exists a neighborhood  $\mathcal{U}'$  of  $\mathbf{X}^* \in \mathcal{X}$ , such that  $\chi_t(\mathbf{X}) \in \mathcal{U}$  for all  $\mathbf{X} \in \mathcal{U}'$  and all  $t \geq 0$ .
- (2)  $\psi$ -stable, if for all neighborhoods  $\mathcal{U}$  of  $\mathbf{X}^* \in \mathcal{X}$ , there exists a neighborhood  $\mathcal{U}'$  of  $\mathbf{X}^* \in \mathcal{X}$ , such that  $\chi_t(\mathbf{X}) \in \mathcal{U}$  for all  $\mathbf{X} \in \mathcal{U}' \cap \text{im } \mathbf{Q}$  and all  $t \geq 0$ .

**Lemma D.3.** *A point  $\mathbf{X}^* \in \mathcal{X}$  is  $\chi$ -stable if and only if it is  $\psi$ -stable.*

*Proof.* The “only if” part is trivial, so we focus on the “if” part. To that end, suppose that  $\mathbf{X}^*$  is  $\psi$ -stable but not  $\chi$ -stable. Then, there exists a neighborhood  $\mathcal{U}$  of  $\mathbf{X}^*$  in  $\mathcal{X}$  such that for all neighborhoods  $\mathcal{U}'$  of  $\mathbf{X}^*$ , there exists  $\mathbf{X} \in \mathcal{U}'$  such that  $\chi_t(\mathbf{X}) \notin \mathcal{U}$  for some  $t \geq 0$ . Now, let  $\mathcal{U}_0$  be a neighborhood of  $\mathbf{X}^*$  such that  $\text{cl}(\mathcal{U}_0) \subseteq \mathcal{U}$ . Then, by  $\psi$ -stability of  $\mathbf{X}^*$ , there exists a neighborhood  $\mathcal{U}'_0$  of  $\mathbf{X}^*$  such that  $\chi_t(\mathbf{X}) \in \mathcal{U}_0$  for all  $\mathbf{X} \in \mathcal{U}'_0 \cap \text{im } \mathbf{Q}$  and all  $t \geq 0$ . But, since  $\mathbf{X}^*$  is not  $\chi$ -stable, there exists  $\mathbf{X} \in \mathcal{U}'_0 \cap (\text{im } \mathbf{Q})^c$  such that the hitting time  $\tau = \inf\{t \geq 0 : \chi_t(\mathbf{X}) \notin \mathcal{U}\}$  is finite. Moreover, let  $\{\mathbf{X}_n\}_{n \in \mathbb{N}} \subseteq \mathcal{U}'_0 \cap \text{im } \mathbf{Q}$  be a sequence converging to  $\mathbf{X}$  (such a sequence exists, because  $\text{cl}(\text{im } \mathbf{Q}) = \mathcal{X}$ ). Since  $\mathbf{X}^*$  is  $\psi$ -stable, it holds that  $\chi_\tau(\mathbf{X}_n) \in \mathcal{U}_0$  for all  $n$ , and, therefore, using the continuity of  $\chi$ , we get:

$$\chi_\tau(\mathbf{X}) = \lim_{n \rightarrow \infty} \chi_\tau(\mathbf{X}_n) \quad (\text{D.5})$$

which implies that  $\chi_\tau(\mathbf{X}) \in \text{cl}(\mathcal{U}_0) \subseteq \mathcal{U}$ , as a limit point of  $\mathcal{U}_0$ . But, since  $\chi_\tau(\mathbf{X}) \notin \mathcal{U}$  by assumption, we obtain a contradiction that completes our proof. ■

Thanks to the above proposition, we do not need to distinguish between “Lyapunov stable” and “consistently Lyapunov stable” states, as the two notions are equivalent. Thus, with this issue settled, we are in a position to prove our quantum version of the “folk theorem”, which we restate for convenience below:

**Theorem 2.** *Let  $\mathcal{Q} \equiv \mathcal{Q}(\mathcal{N}, \Psi, u)$  be a quantum game, fix some state  $\mathbf{X}^* \in \mathcal{X}$ , and let  $\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t))$  be a trajectory of play induced by (FTQL) with steep regularizers. Then:*

- (1) *If  $\mathbf{X}^*$  is a Nash equilibrium, it is a rest point of (QD).*
- (2) *If  $\mathbf{X}(t) \rightarrow \mathbf{X}^*$  as  $t \rightarrow \infty$ ,  $\mathbf{X}^*$  is a Nash equilibrium of  $\mathcal{Q}$ .*
- (3) *If  $\mathbf{X}^*$  is stable under (QD), it is a Nash equilibrium of  $\mathcal{Q}$ .*
- (4) *If  $\mathbf{X}^*$  is consistently asymptotically stable, then it is pure.*
- (5) *If  $\mathbf{X}^*$  satisfies (VS), it is consistently asymptotically stable.*

We prove the various parts of the theorem below.

*Proof of Part 1.* Suppose that  $\mathbf{X}^*$  is a Nash equilibrium of  $\mathcal{Q}$ . If  $\text{rank}(\mathbf{X}^*) = 1$ , there is nothing to show, so it suffices to establish our claim for the case  $\text{rank}(\mathbf{X}^*) > 1$ . In that case, by restricting the dynamics if necessary to the relative interior of the face  $\mathcal{X}^* = \{\mathbf{X} \in \mathcal{X} : \ker \mathbf{X} \geq \ker \mathbf{X}^*\}$  of  $\mathcal{X}$  whose relative interior contains  $\mathbf{X}^*$ , we may assume without loss of generality that  $\mathbf{X}^*$  has, in fact, full rank. With this in mind, fix some  $\mathbf{Y} \in \mathbf{Q}^{-1}(\mathbf{X}^*)$ , and consider the trajectory  $\mathbf{Y}(t) = \mathbf{Y} + t\mathbf{V}(\mathbf{X}^*)$  starting at  $\mathbf{Y}$  at time  $t = 0$ . Since  $\mathbf{V}(\mathbf{X}^*) \in \text{NC}(\mathbf{X}^*)$  by the variational characterization (VI) of Nash equilibria, we conclude by Lemma A.1 that  $\mathbf{Q}(\mathbf{Y}(t)) = \mathbf{Q}(\mathbf{Y})$  for all  $t \geq 0$ . In turn, this shows that  $\dot{\mathbf{Y}}(t) = \mathbf{V}(\mathbf{X}^*) = \mathbf{V}(\mathbf{Q}(\mathbf{Y}(t)))$  for all  $t \geq 0$ , i.e.,  $\mathbf{Y}(t)$  is a solution orbit of (FTQL). By the Picard-Lindelöf theorem, this implies that  $\mathbf{Y}(t)$  is the unique solution thereof; hence, given that  $\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t)) = \mathbf{X}^*$  for all  $t \geq 0$ , our claim follows. ■

*Proof of Parts 2 and 3.* Convergence to and Lyapunov stability of  $\mathbf{X}^*$  both imply the assumption of Lemma D.1, so our assertions follow immediately by the conclusion of said lemma. ■

*Proof of Part 4.* The proof of this part consists of three steps. First, we will construct a measure  $\mu$  on  $\text{ri } \mathcal{X}$  that is invariant under the flow  $\chi$ , i.e.,  $\mu(A) = \mu(\chi_t(A))$  for all measurable sets  $A$  and  $t \geq 0$ , where by  $\chi_t(A)$  we denote the image of  $A$  after time  $t$ . Next, we will show that if  $\mathbf{X}^* \in \text{ri } \mathcal{X}$ , it cannot be consistently asymptotically stable. Finally, we will conclude

that if  $\mathbf{X}^* \in \mathcal{X}$ , is consistently asymptotically stable, then it is pure with a simple reduction argument.

**Step 1.** By the definition of  $\mathbf{Q}$ , we have  $\mathbf{Y} = \nabla_{\mathbf{X}^\top} \theta(\mathbf{X}) + c \cdot \mathbf{I}$ , i.e.,  $\mathbf{Q}^{-1}(\mathbf{X}) = \{\nabla_{\mathbf{X}^\top} \theta(\mathbf{X}) + c \cdot \mathbf{I} : c \in \mathbb{R}\}$ . Let  $\mathcal{Y}_0 := \mathcal{Y} / \sim$  be the quotient space induced by the equivalence relation  $\sim$ , where  $\mathbf{Y} \sim \mathbf{Y}'$  if  $\mathbf{Y} - \mathbf{Y}' = c \cdot \mathbf{I}$  for some  $c \in \mathbb{R}$ , and let  $q : \mathcal{Y} \rightarrow \mathcal{Y}_0$  be the corresponding quotient map, defined as  $q(\mathbf{Y}) = \{\mathbf{Y} + c \cdot \mathbf{I} : c \in \mathbb{R}\}$ . Thus, by descending to the quotient space  $\mathcal{Y}_0$ , the mirror map  $\mathbf{Q}$  factors through  $\mathcal{Y}_0$  as  $\mathbf{Q}_0 : \mathcal{Y}_0 \rightarrow \mathcal{X}$  defined as  $\mathbf{Q}_0 := \mathbf{Q} \circ q$ . Now, let  $\lambda$  denote the Lebesgue measure on  $\mathcal{Y}_0$ , and let  $\mu = \mathbf{Q}_0 \# \lambda$  denote the pushforward of  $\lambda$  to  $\mathcal{X}$  via  $\mathbf{Q}_0$ , i.e.,  $\mu(A) = \lambda(\mathbf{Q}_0^{-1}(A))$  for all measurable sets  $A$  in  $\mathcal{X}$ . Finally, by [Proposition E.1](#), we readily get that  $\mu(\chi_t(A)) = \mu(A)$  for all measurable sets  $A \subseteq \mathcal{X}$  and all  $t \geq 0$ .

**Step 2.** For the sake of contradiction, suppose  $\mathbf{X}^* \in \text{ri } \mathcal{X}$  is consistently asymptotically stable, and let  $\mathcal{U}$  be a sufficiently small compact neighborhood of  $\mathbf{X}^*$ , as per [Lemma D.2](#). Then, it holds  $\lim_{t \rightarrow \infty} \sup_{\mathbf{X} \in \mathcal{U}} \|\chi_t(\mathbf{X}) - \mathbf{X}^*\| = 0$ , which implies that:

$$\lim_{t \rightarrow \infty} \mu(\chi_t(\mathcal{U})) = \mu(\{\mathbf{X}^*\}) = 0 < \mu(\mathcal{U}). \quad (\text{D.6})$$

However, by *Step 1* of the proof, we have that  $\lim_{t \rightarrow \infty} \mu(\chi_t(\mathcal{U})) = \mu(\mathcal{U})$ , since  $\mu(\chi_t(\mathcal{U})) = \mu(\mathcal{U})$  for all  $t \geq 0$ . Hence, combining it with [\(D.6\)](#), we arrive at a contradiction.

**Step 3.** Suppose, finally, that  $\mathbf{X}^*$  is consistently asymptotically stable with  $r := \text{rank}(\mathbf{X}^*) > 1$ . Note that if  $\mathbf{X}^*$  is full-rank, it cannot be consistently asymptotically stable, as shown at the *Step 2* of the proof. Let  $\mathcal{A}_{\mathbf{X}^*} := \{\mathbf{X} \in \mathcal{X} : \ker(\mathbf{X}) \geq \ker(\mathbf{X}^*)\}$ , which is convex. Indeed, let  $\mathbf{X}, \mathbf{X}' \in \mathcal{A}_{\mathbf{X}^*}$  and  $\lambda \in (0, 1)$ . For  $y \in \ker(\mathbf{X}^*)$ , we have:

$$\begin{aligned} [\lambda \mathbf{X} + (1 - \lambda) \mathbf{X}'] y &= \lambda \mathbf{X} y + (1 - \lambda) \mathbf{X}' y \\ &= 0 \end{aligned} \quad (\text{D.7})$$

i.e.,  $y \in \ker(\lambda \mathbf{X} + (1 - \lambda) \mathbf{X}')$ , and, thus,  $\ker(\lambda \mathbf{X} + (1 - \lambda) \mathbf{X}') \geq \ker(\mathbf{X}^*)$ .

Hence, considering the restriction of the dynamics on  $\mathcal{A}_{\mathbf{X}^*}$ , *Step 2* of the proof shows that a point in  $\text{ri } \mathcal{A}_{\mathbf{X}^*}$  cannot be consistently asymptotically stable on the induced topology. Therefore, the restriction of  $\mathbf{X}^*$  on  $\mathcal{A}_{\mathbf{X}^*}$ , denoted  $\mathbf{X}^*|_{\mathcal{A}_{\mathbf{X}^*}}$  cannot be consistently asymptotically stable on  $\mathcal{A}_{\mathbf{X}^*}$ , since it has full-rank on  $\mathcal{A}_{\mathbf{X}^*}$ . So, it remains to show that  $\mathbf{X}^*$  cannot be consistently asymptotically stable on  $\mathcal{X}$ .

For the sake of contradiction, suppose it is. Then, according to [Definition 1](#), there exist  $\mathcal{U}$  of  $\mathbf{X}^*$  in  $\mathcal{X}$  such that  $\lim_{t \rightarrow \infty} \chi_t(\mathbf{X}) = \mathbf{X}^*$  for all  $\mathbf{X} \in \mathcal{U} \cap \mathcal{A}_{\mathbf{X}^*}$ . But, since  $\mathcal{U}$  is a neighborhood of  $\mathbf{X}^*$  in  $\mathcal{X}$ ,

$$(\mathcal{U} \cap \mathcal{A}_{\mathbf{X}^*}) \cap \mathcal{A}_{\mathbf{X}^*} = \mathcal{U} \cap \{\mathbf{X}' \in \mathcal{X} : \ker(\mathbf{X}') = \ker(\mathbf{X}^*)\} \supseteq \{\mathbf{X}^*\} \quad (\text{D.8})$$

and, thus, the restriction of  $\mathcal{U} \cap \{\mathbf{X}' \in \mathcal{X} : \ker(\mathbf{X}') = \ker(\mathbf{X}^*)\}$  on  $\mathcal{A}_{\mathbf{X}^*}$  is a neighborhood of  $\mathbf{X}^*$  in  $\mathcal{A}_{\mathbf{X}^*}$ . Therefore, by our previous argument on the induced dynamics on  $\mathcal{A}_{\mathbf{X}^*}$ , there exists  $\mathbf{X}_0 \in \mathcal{U} \cap \{\mathbf{X}' \in \mathcal{X} : \ker(\mathbf{X}') = \ker(\mathbf{X}^*)\}$  such that  $\chi_t(\mathbf{X}_0) \not\rightarrow \mathbf{X}^*$ , as  $t \rightarrow \infty$ . This contradicts our original assumption, so our proof is complete.  $\blacksquare$

*Proof of Part 5.* Suppose that  $\mathbf{X}^*$  is (locally) variationally stable, i.e.,  $\text{tr}[\mathbf{V}(\mathbf{X})(\mathbf{X} - \mathbf{X}^*)] < 0$  for all  $\mathbf{X} \in \mathcal{U}$ , where  $\mathcal{U}$  is a neighborhood of  $\mathbf{X}^*$ . Then, by Lyapunov's direct method for the energy function  $E(t) := F(\mathbf{X}^*, \mathbf{Y}(t))$ , and invoking [Lemma D.3](#), we get that  $\mathbf{X}^*$  is stable, thus, there exists neighborhood  $\mathcal{U}'$  of  $\mathbf{X}^*$ , such that  $\mathbf{X}(t) \in \mathcal{U}$  for all  $t \geq 0$ , if  $\mathbf{X}(0) \in \mathcal{U}'$ . This means that if  $\mathbf{X}(0) \in \mathcal{U}'$ , then  $\text{tr}[\mathbf{V}(\mathbf{X}(t))(\mathbf{X}(t) - \mathbf{X}^*)] < 0$  for all  $t \geq 0$ . In what follows, we will show that  $\mathbf{X}(t) \rightarrow \mathbf{X}^*$  if  $\mathbf{X}(0) \in \mathcal{U}'$ .



Now, let  $\mathbf{X}(0) \in \mathcal{U}'$ . Since  $\mathbf{X}(t) \in \mathcal{U}$  for all  $t \geq 0$ , and using [Lemma 1](#), we have that:

$$\frac{d}{dt}E(t) = \text{tr}[\mathbf{V}(\mathbf{X}(t))(\mathbf{X}(t) - \mathbf{X}^*)] < 0 \quad (\text{D.9})$$

where the last inequality comes from [\(VS\)](#). Hence,  $E(t)$  is decreasing if  $\mathbf{X}(0) \in \mathcal{U}$ .

Now, we will show that  $\mathbf{X}^*$  is an  $\omega$ -limit of the flow, i.e., there exists a sequence  $\{t_n\}_{n \in \mathbb{N}}$  such that  $\mathbf{X}(t_n) \rightarrow \mathbf{X}^*$  as  $n \rightarrow \infty$ . For the sake of contradiction, suppose such a sequence does not exist. Then,  $\|\mathbf{X}(t) - \mathbf{X}^*\|$  is bounded away from zero, which implies that there exists  $c > 0$  such that  $\text{tr}[\mathbf{V}(\mathbf{X}(t))(\mathbf{X}(t) - \mathbf{X}^*)] < -c$ , i.e.,  $\frac{d}{dt}E(t) < -c$ . Integrating over time, we get

$$E(t) - E(0) < -ct$$

which implies that  $E(t) \rightarrow -\infty$  as  $t \rightarrow \infty$ . This contradicts the non-negativity property of the Fenchel coupling, thus, we conclude that there exists a sequence  $\{t_n\}_{n \in \mathbb{N}}$  such that  $\mathbf{X}(t_n) \rightarrow \mathbf{X}^*$  as  $n \rightarrow \infty$ .

Finally, by [Proposition A.2](#), we get that  $F(\mathbf{X}^*, \mathbf{Y}(t_n)) \rightarrow 0$  as  $n \rightarrow \infty$ , i.e.,  $E(t_n) \rightarrow 0$ . Therefore, since  $E(t)$  is decreasing and  $E(t_n) \rightarrow 0$ , we conclude that  $E(t) \rightarrow 0$  as  $t \rightarrow \infty$ , i.e.,  $F(\mathbf{X}^*, \mathbf{Y}(t)) \rightarrow 0$ . Again, using [Proposition A.2](#), we conclude that  $\mathbf{X}(t) \rightarrow \mathbf{X}^*$ , i.e.,  $\mathbf{X}^*$  is consistently asymptotically stable.  $\blacksquare$

## APPENDIX E. POINCARÉ RECURRENCE

Our aim in this appendix will be to prove our recurrence result for min-max games admitting a full-rank equilibrium. For ease of reference, we restate the result below:

**Theorem 3.** *Let  $\mathcal{Q} \equiv \mathcal{Q}(\mathcal{N}, \Psi, u)$  be a 2-player zero-sum game admitting a full-rank equilibrium  $\mathbf{X}^* \in \text{ri } \mathcal{X}$ . Then, almost every solution orbit  $\mathbf{X}(t) = \mathbf{Q}(\mathbf{Y}(t))$  of [\(FTQL\)](#) is Poincaré recurrent: there exists an increasing sequence of times  $t_n \rightarrow \infty$  such that  $\lim_{n \rightarrow \infty} \mathbf{X}(t_n) = \mathbf{X}(0)$ .*

As we briefly discussed in the main part of the paper, our proof strategy will consist of the following steps:

- (1) First, we will need to collapse the space of dual space  $\mathcal{Y}$  of  $\mathcal{V} \equiv \mathbb{H}^d$  to a subspace  $\mathcal{Z}$  with constant trace. The reason for this is that the mirror map  $\mathbf{Q}: \mathcal{Y} \rightarrow \mathcal{X}$  is not injective, even when  $h$  is steep: for all  $\lambda \in \mathbb{R}$ , we have  $\mathbf{Q}(\mathbf{Y} + \lambda \mathbf{I}) = \mathbf{Q}(\mathbf{Y})$ , which means that the inverse image of any point in  $\text{im } \mathbf{Q}$  always contains a copy of the real line. This collapse is intended to “quotient out” this redundancy so as to enable volume comparisons later on.
- (2) The second step is to show that the dynamics [\(FTQL\)](#) remain well-defined on the quotient space  $\mathcal{Z}$ , and that they are incompressible, that is, the volume of a set of initial conditions is maintained over time. This is encoded by [Proposition 2](#) in the main text.
- (3) The next step in our proof will be to show that the induced dynamics on  $\mathcal{Z}$  admit a *constant of motion*. This will be provided by the Fenchel coupling relative to the game’s interior equilibrium (and it is also the point where the full-rank assumption is needed).
- (4) Finally, the last – and most challenging – part of the proof is to show that the level sets of the Fenchel coupling are bounded in  $\mathcal{Z}$  (by contrast, they are *not* bounded in  $\mathcal{Y}$ ).

Once these steps are complete, [Theorem 3](#) will follow from Poincaré’s theorem. In what follows, we encode each of these steps in a series of lemmas and proposition. To lighten

notation, we will suppress player indices throughout unless it is absolutely necessary to include them.

**Step 1: Descending to the quotient.** To quotient out the redundancy mentioned above, consider the transformed matrix variables

$$\mathbf{Z} = \mathbf{Y} - (1/d) \operatorname{tr}[\mathbf{Y}]\mathbf{I} \quad (\text{E.1})$$

so  $\operatorname{tr} \mathbf{Z} = 0$  for all  $\mathbf{Y} \in \mathcal{Y}$ . Formally, this transformation can be represented via the map  $\mathbf{\Pi}: \mathcal{Y} \rightarrow \mathcal{Z}$ , where

$$\mathcal{Z} = \{\mathbf{Z} \in \mathcal{Y} : \operatorname{tr} \mathbf{Z} = 0\} \quad (\text{E.2})$$

and  $\mathbf{\Pi}: \mathbf{Y} \mapsto \mathbf{\Pi}(\mathbf{Y}) = \mathbf{Z}$  is defined via (E.1) above. As such,  $\mathbf{Z}$  can be seen as a representative of the equivalence relation  $\mathbf{Y} \sim \mathbf{Y} + \lambda \mathbf{I}$ ,  $\lambda \in \mathbb{R}$ , under which light  $\mathcal{Z}$  can be identified with the quotient  $\mathcal{Y}/\sim$ .

By construction and the properties of  $\mathbf{Q}$  (Lemma A.1 in particular), we have  $\mathbf{Q}(\mathbf{Y}) = \mathbf{Q}(\mathbf{\Pi}(\mathbf{Y})) = \mathbf{Q}(\mathbf{Z})$  for all  $\mathbf{Y} \in \mathcal{Y}$ . Accordingly, (FTQL) readily gives

$$\begin{aligned} \dot{\mathbf{Z}} &= \frac{d}{dt}[\mathbf{Y} - (1/d) \operatorname{tr}[\mathbf{Y}]\mathbf{I}] = \dot{\mathbf{Y}} - (1/d) \operatorname{tr}[\dot{\mathbf{Y}}]\mathbf{I} \\ &= \mathbf{V}(\mathbf{Q}(\mathbf{Y})) - (1/d) \operatorname{tr}[\mathbf{V}(\mathbf{Q}(\mathbf{Y}))]\mathbf{I} \\ &= \mathbf{V}(\mathbf{Q}(\mathbf{Z})) - (1/d) \operatorname{tr}[\mathbf{V}(\mathbf{Q}(\mathbf{Z}))]\mathbf{I} \end{aligned} \quad (\text{E.3})$$

so we can rewrite (FTQL) in terms of  $\mathbf{Z}$  as

$$\dot{\mathbf{Z}} = \mathbf{V}(\mathbf{X}) - (1/d) \operatorname{tr}[\mathbf{V}(\mathbf{X})]\mathbf{I} \quad \mathbf{X} = \mathbf{Q}(\mathbf{Z}). \quad (\text{FTQL}_{\mathcal{Z}})$$

Since  $\operatorname{tr}[\dot{\mathbf{Z}}] = \operatorname{tr}[\mathbf{V}(\mathbf{X})] - (1/d) \operatorname{tr}[\mathbf{V}(\mathbf{X})] \operatorname{tr} \mathbf{I} = 0$ , it follows that any traceless initial condition of (FTQL<sub>z</sub>) will remain traceless – and hence remain in  $\mathcal{X}$  for all  $n \geq 0$ . We thus conclude that (FTQL<sub>z</sub>) is a well-posed dynamical system on  $\mathcal{Z}$  whose induced flow  $\zeta: \mathbb{R} \times \mathcal{Z} \rightarrow \mathcal{Z}$  conjugates  $\psi$  in the sense that  $\zeta_t \circ \mathbf{\Pi} = \mathbf{\Pi} \circ \psi_t$  for all  $t \geq 0$ .

**Step 2: Volume preservation.** Our second step is to show that the dynamics (FTQL<sub>z</sub>) are volume-preserving. Formally, we establish the following direct analogue of Proposition 2:

**Proposition E.1.** *Let  $\mathcal{W} \subseteq \mathcal{Z}$  be an open set of initial conditions of (FTQL<sub>z</sub>), and let  $\mathcal{W}_t = \zeta_t(\mathcal{W})$ ,  $t \geq 0$ , denote the evolution of  $\mathcal{W}$  under the flow  $\zeta: \mathbb{R} \times \mathcal{Z} \rightarrow \mathcal{Z}$  of (FTQL<sub>z</sub>). Then,  $\operatorname{vol}(\mathcal{W}_t) = \operatorname{vol}(\mathcal{W})$  for all  $t \geq 0$ .*

*Remark.* In the above (as in Proposition 2), the volume form  $\operatorname{vol}(\cdot)$  stands for the ordinary Lebesgue measure on  $\mathcal{Z}$ .

*Proof of Proposition E.1.* The key observation here is that the dynamics (FTQL<sub>z</sub>) are *incompressible* – that is, *divergence-free*. Indeed, since  $\mathbf{V}_i$  does not depend on  $\mathbf{X}_i$  (it is not possible to suppress player indices here), we readily have  $\nabla_{\mathbf{Z}_i^\top} \mathbf{V}_i(\mathbf{Q}(\mathbf{Y})) = 0$  for all  $i \in \mathcal{N}$ . This immediately implies that the field

$$\mathbf{W}_i(\mathbf{Z}) = \mathbf{V}_i(\mathbf{Q}(\mathbf{Z})) - (1/d_i) \operatorname{tr}[\mathbf{V}_i(\mathbf{Q}(\mathbf{Z}))]\mathbf{I} \quad (\text{E.4})$$

has  $\operatorname{div}_{\mathbf{Z}}(\mathbf{W}(\mathbf{Z})) = 0$ , so (FTQL<sub>z</sub>) is incompressible. Our claim then follows from Liouville's formula [3].  $\blacksquare$

Proposition 2 is proved in exactly the same way, so we omit its proof (which is not needed for the sequel, anyway).

**Step 3: The Fenchel coupling as a constant of motion.** We now proceed to show that the dynamics (FTQL $\mathcal{Z}$ ) admit a constant of motion, namely the Fenchel coupling relative to the game's full-rank equilibrium. Indeed, letting  $E(t) = F(\mathbf{X}^*, \mathbf{Z}(t))$ , Lemma 1 readily gives:

$$\begin{aligned} \frac{dE}{dt} &= \text{tr}[\mathbf{V}_1(\mathbf{X})(\mathbf{X}_1 - \mathbf{X}_1^*)] + \text{tr}[\mathbf{V}_2(\mathbf{X})(\mathbf{X}_2 - \mathbf{X}_2^*)] \\ &= u_1(\mathbf{X}_1, \mathbf{X}_2) - u_1(\mathbf{X}_1^*, \mathbf{X}_2) + u_2(\mathbf{X}_1, \mathbf{X}_2) - u_2(\mathbf{X}_1, \mathbf{X}_2^*) \\ &= -\mathcal{L}(\mathbf{X}_1, \mathbf{X}_2) + \mathcal{L}(\mathbf{X}_1^*, \mathbf{X}_2) + \mathcal{L}(\mathbf{X}_1, \mathbf{X}_2) - \mathcal{L}(\mathbf{X}_1, \mathbf{X}_2^*) = \mathcal{L}(\mathbf{X}_1^*, \mathbf{X}_2) - \mathcal{L}(\mathbf{X}_1, \mathbf{X}_2^*) = 0 \end{aligned} \quad (\text{E.5})$$

where, in the last line, we used the assumption that  $\mathbf{X}^*$  is a full-rank equilibrium; see also the relevant discussion in [22, 23].

**Step 4: Boundedness of the trajectories.** We are left to show that the trajectories of (FTQL $\mathcal{Z}$ ) are bounded. The key to this will be the following technical lemma:

**Lemma E.1.** *Let  $\mathbf{Z}_n$ ,  $n = 1, 2, \dots$ , be a sequence in  $\mathcal{Z}$  such that  $F(\mathbf{P}, \mathbf{Z}_n)$  remains constant for some interior  $\mathbf{P} \in \mathcal{X}$ . Then the eigenvalues  $\{z_{\alpha,n}\}_{\alpha=1,\dots,d}$  of  $\mathbf{Z}_n$  are bounded, i.e.,  $\sup_{\alpha,n} |z_{\alpha,n}| < \infty$ .*

*Proof.* We argue by contradiction. Indeed, assume that  $F(\mathbf{P}, \mathbf{Z}_n)$  is constant but  $\sup_{\alpha,n} z_{\alpha,n} = \infty$ . Then, letting  $z_n^+ = \max_{\alpha} z_{\alpha,n}$  and  $z_n^- = \min_{\alpha} z_{\alpha,n}$ , we will also have  $\limsup_{n \rightarrow \infty} (z_n^+ - z_n^-) = \infty$  by the fact that  $\mathbf{Z}_n \in \mathcal{Z}$  (so  $\text{tr} \mathbf{Z}_n = 0$ ). Then, by relabeling indices and descending to a finer subsequence if necessary, we may assume without loss of generality that the following properties hold concurrently:

- (1)  $\mathbf{X}_n = \mathbf{Q}(\mathbf{Z}_n)$  converges to some  $\mathbf{X} \in \mathcal{X}$  (by the compactness of  $\mathcal{X}$ ).
- (2) All elements of the eigen-decomposition  $\mathbf{X}_n = \sum_{\alpha} x_{\alpha,n} \mathbf{u}_{\alpha,n} \mathbf{u}_{\alpha,n}^\dagger$  of  $\mathbf{X}_n$  also converge (by the compactness of the unitary group  $U(d)$  and the unit simplex).
- (3)  $z_{d,n} = z_n^+$  and  $z_{1,n} = z_n^-$  for all  $n = 1, 2, \dots$
- (4) The index set  $\{1, \dots, d\}$  can be partitioned into two nonempty sets  $\mathcal{S}^+$  and  $\mathcal{S}^-$  such that a)  $\sup_n |z_n^+ - z_{\alpha,n}| < \infty$  for all  $\alpha \in \mathcal{S}^+$ ; and b)  $z_n^+ - z_{\alpha,n} \rightarrow \infty$  for all  $\alpha \in \mathcal{S}^-$ .

Since  $\mathbf{X}_n$  and  $\mathbf{Z}_n$  commute (by Lemma A.2), we get:

$$\begin{aligned} \text{tr}[\mathbf{Z}_n \cdot (\mathbf{P} - \mathbf{X}_n)] &= \sum_{\alpha=1}^d z_{\alpha,n} (p_{\alpha,n} - x_{\alpha,n}) \\ &= \sum_{\alpha=1}^d (z_{\alpha,n} - z_n^+) (p_{\alpha,n} - x_{\alpha,n}) \\ &= \sum_{\alpha \in \mathcal{S}^+} (z_{\alpha,n} - z_n^+) (p_{\alpha,n} - x_{\alpha,n}) + \sum_{\alpha \in \mathcal{S}^-} (z_{\alpha,n} - z_n^+) (p_{\alpha,n} - x_{\alpha,n}) \end{aligned} \quad (\text{E.6})$$

where we let  $p_{\alpha,n} = \mathbf{u}_{\alpha,n}^{\text{diag}} \mathbf{P} \mathbf{u}_{\alpha,n}$ , and we used the fact that  $\text{tr} \mathbf{P} = \text{tr} \mathbf{X}_n = 1$  in the second line. Now, by the third part of Lemma A.2, we have  $x_{\alpha,n} \rightarrow 0$  for all  $\alpha \in \mathcal{S}^-$ , so  $\lim_{n \rightarrow \infty} (p_{\alpha,n} - x_{\alpha,n}) > 0$  for all  $\alpha \in \mathcal{S}^-$  (recall that  $p_{\alpha,n}$  converges and  $\mathbf{P}$  has been assumed full-rank). As a result, we get

$$\sum_{\alpha \in \mathcal{S}^-} (z_{\alpha,n} - z_n^+) (p_{\alpha,n} - x_{\alpha,n}) \rightarrow -\infty \quad \text{as } n \rightarrow \infty. \quad (\text{E.7})$$

and hence, since all quantities involved in the sum over  $\alpha \in \mathcal{S}^+$  are bounded, we conclude that  $\text{tr}[\mathbf{Z}_n \cdot (\mathbf{P} - \mathbf{X}_n)] \rightarrow -\infty$  as  $n \rightarrow \infty$ . However, by the definition of the Fenchel coupling,

we have

$$F(\mathbf{P}, \mathbf{Z}_n) = h(\mathbf{P}) + h^*(\mathbf{Z}_n) - \text{tr}[\mathbf{Z}_n \mathbf{P}] = h(\mathbf{P}) - h(\mathbf{X}_n) - \text{tr}[\mathbf{Z}_n \cdot (\mathbf{P} - \mathbf{X}_n)], \quad (\text{E.8})$$

which in turn yields  $\lim_{n \rightarrow \infty} F(\mathbf{P}, \mathbf{Z}_n) = -\infty$ , a contradiction.  $\blacksquare$

**Step 5: Putting everything together.** With all this said and done, we are finally in a position to prove [Theorem 3](#).

*Proof of Theorem 3.* By [Lemma E.1](#) and the fact that  $F(\mathbf{X}^*, \mathbf{Z}(t))$  is a constant of motion under  $(\text{FTQL}_{\mathcal{Z}})$ , it follows that the trajectories of  $(\text{FTQL}_{\mathcal{Z}})$  remain bounded. Hence, by [Proposition E.1](#) and Poincaré’s recurrence theorem, it follows that the dynamics  $(\text{FTQL}_{\mathcal{Z}})$  are recurrent, i.e., for (Lebesgue) almost every initial condition  $\mathbf{Z} \in \mathcal{Z}$ , the induced trajectory  $\mathbf{Z}(t) \equiv \zeta_t(\mathbf{Z})$  returns arbitrarily close to  $\mathbf{Z}$  infinitely many times. More precisely, this means that for every neighborhood  $\mathcal{W}$  of  $\mathbf{Z}(0)$  in  $\mathcal{Z}$ , there exists some finite return time  $\tau_{\mathcal{W}}$  such that  $\zeta_{\tau_{\mathcal{W}}}(\mathbf{Z}) \in \mathcal{W}$ . Hence, by taking a shrinking net of balls  $\mathcal{W}_n$  centered at  $\mathbf{Z}$  and with radius  $1/n$ , and letting  $t_n = \tau_{\mathcal{W}_n}$ , we deduce that  $\zeta(t_n) \rightarrow \mathbf{Z}$  as  $n \rightarrow \infty$ . Our assertion then follows by applying the above construction to some  $\mathbf{Z} \in \mathbf{Q}^{-1}(\mathbf{X}(0)) \cap \mathcal{Z}$  and noting that  $\mathbf{X}(t_n) = \chi_{t_n}(\mathbf{X}(0)) = \mathbf{Q}(\zeta_{t_n}(\mathbf{Z})) \rightarrow \mathbf{Q}(\mathbf{Z}) = \mathbf{X}(0)$ .  $\blacksquare$

## APPENDIX F. NUMERICAL RESULTS

In this appendix, we describe our simulations setup. We consider a simple two-player zero-sum game with payoff matrices  $P_1 = \begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix}$  for the row- and  $P_2 = -P_1$  for the column-player, expressed in the quantum regime. Denoting the rows as  $(\alpha_1, \alpha_2)$  and the columns as  $(\beta_1, \beta_2)$ , we have that the action profile  $(\alpha_1, \beta_2)$  is a strict NE. We run [\(MMW\)](#) on the two-player quantum game with initial conditions  $(0.2, 0.8)$  and  $(0.8, 0.2)$  for the row- and column-player, respectively, and generate the trajectories of the players’ states up to time  $t = 100$ . The way we visualize them is through each player’s Bloch sphere. Note that the red diamond “ $\blacklozenge$ ” on the figures denotes the initial mixed states of the players. As we observe, the trajectories converge to the boundary of the Bloch spheres, i.e., to pure states. To facilitate the reproducibility of our results, we used we denote that we used the openly available code from: [\[23, https://openreview.net/forum?id=FFZYhY2z3j\]](https://openreview.net/forum?id=FFZYhY2z3j)

## REFERENCES

- [1] Aaronson, S. Shadow tomography of quantum states. *SIAM Journal on Computing*, 49(5):STOC18–368–STOC18–394, January 2020.
- [2] Aaronson, S., Chen, X., Hazan, E., Kale, S., and Nayak, A. Online learning of quantum states. In *NeurIPS ’18: Proceedings of the 32nd International Conference of Neural Information Processing Systems*, 2018.
- [3] Arnold, V. I. *Mathematical Methods of Classical Mechanics*. Number 60 in Graduate Texts in Mathematics. Springer, New York, NY, 2 edition, 1989.
- [4] Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J. C., Barends, R., Biswas, R., Boixo, S., Brandao, F. G., Buell, D. A., Burkett, B., Chen, Y., Chen, Z., Chiaro, B., Collins, R., Courtney, W., Dunsworth, A., Farhi, E., Foxen, B., Fowler, A., Gidney, C., Giustina, M., Graff, R., Guerin, K., Habegger, S., Harrigan, M. P., Hartmann, M. J., Ho, A., Hoffmann, M. R., Huang, T., Humble, T. S., Isakov, S. V., Jeffrey, E., Jiang, Z., Kafri, D., Kechedzhi, K., Kelly, J., Klimov, P. V., Knysh, S., Korotkov, A. N., Kostritsa, F., Landhuis, D., Lindmark, M., Lucero, E., Lyakh, D., Mandra, S., McClean, J. R., McEwen, M., Megrant, A., Mi, X., Michielsen, K., Mohseni, M., Mutus, J., Naaman, O., Neeley, M., Neill, C., Niu, M. Y., Ostby, E., Petukhov, A., Platt, J. C., Quintana, C., Rieffel, E. G., Roushan, P., Rubin, N. C., Sank, D., Satzinger, K. J., Smelyanskiy, V., Sung, K. J., Trevithick, M. D., Vainsencher, A., Villalonga,

- B., White, T., Yao, Z. J., Yeh, P., Zalcman, A., Neven, H., and Martinis, J. M. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 2019.
- [5] Bostanci, J. and Watrous, J. Quantum game theory and the complexity of approximating quantum Nash equilibria. *Quantum*, 6:882, 2022.
- [6] Carlen, E. A. Trace inequalities and quantum entropy: An introductory course. 2009.
- [7] Chakrabarti, S., Huang, Y., Li, T., Feizi, S., and Wu, X. Quantum Wasserstein generative adversarial networks. In *NeurIPS '19: Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019.
- [8] Chuang, I. and Nielsen, M. *Quantum Computation and Quantum Information*. Cambridge University Press, 2 edition, 2010.
- [9] Coucheney, P., Gaujal, B., and Mertikopoulos, P. Penalty-regulated dynamics and robust learning procedures in games. *Mathematics of Operations Research*, 40(3):611–633, August 2015.
- [10] Cressman, R. *Evolutionary Dynamics and Extensive Form Games*. The MIT Press, 2003.
- [11] Dallaire-Demers, P.-L. and Killoran, N. Quantum generative adversarial networks. *Physical Review A*, 98(1):012324, 2018.
- [12] Daskalakis, C., Goldberg, P. W., and Papadimitriou, C. H. The complexity of computing a Nash equilibrium. In *STOC '06: Proceedings of the 38th annual ACM symposium on the Theory of Computing*, 2006.
- [13] Daskalakis, C., Goldberg, P. W., and Papadimitriou, C. H. The complexity of computing a Nash equilibrium. *Communications of the ACM*, 52(2):89–97, 2009.
- [14] Debreu, G. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences of the USA*, 38(10):886–893, October 1952.
- [15] Eisert, J., Wilkens, M., and Lewenstein, M. Quantum games and quantum strategies. *Physical Review Letters*, 83(15):3077–3080, October 1999.
- [16] Facchinei, F. and Pang, J.-S. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Series in Operations Research. Springer, 2003.
- [17] Flokas, L., Vlatakis-Gkaragkounis, E. V., Lianas, T., Mertikopoulos, P., and Piliouras, G. No-regret learning and mixed Nash equilibria: They do not mix. In *NeurIPS '20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.
- [18] Folland, G. B. *Real Analysis*. Wiley-Interscience, 2 edition, 1999.
- [19] Gutoski, G. and Watrous, J. Toward a general theory of quantum games. In *STOC '07: Proceedings of the 39th annual ACM symposium on the Theory of Computing*, 2007.
- [20] Hidalgo, E. G. Quantum replicator dynamics. *Physica A: Statistical Mechanics and its Applications*, 369(2):393–407, September 2006.
- [21] Hofbauer, J. and Sigmund, K. Evolutionary game dynamics. *Bulletin of the American Mathematical Society*, 40(4):479–519, July 2003.
- [22] Jain, R. and Watrous, J. Parallel approximation of non-interactive zero-sum quantum games. In *CCC '09: Proceedings of the 2009 IEEE International Conference on Computational Complexity*, 2009.
- [23] Jain, R., Piliouras, G., and Sim, R. Matrix multiplicative weights updates in quantum games: Conservation law & recurrence. In *NeurIPS '22: Proceedings of the 36th International Conference on Neural Information Processing Systems*, 2022.
- [24] Kakade, S. M., Shalev-Shwartz, S., and Tewari, A. Regularization techniques for learning with matrices. *The Journal of Machine Learning Research*, 13:1865–1890, 2012.
- [25] Kwon, J. and Mertikopoulos, P. A continuous-time approach to online optimization. *Journal of Dynamics and Games*, 4(2):125–148, April 2017.
- [26] Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, Cambridge, UK, 2020.
- [27] Lloyd, S. and Weedbrook, C. Quantum generative adversarial learning. *Physical Review Letters*, 121(4):040502–040507, July 2018.
- [28] Maynard Smith, J. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, 1982.
- [29] Maynard Smith, J. and Price, G. R. The logic of animal conflict. *Nature*, 246:15–18, November 1973.
- [30] Mertikopoulos, P. and Sandholm, W. H. Learning in games via reinforcement and regularization. *Mathematics of Operations Research*, 41(4):1297–1324, November 2016.

- [31] Mertikopoulos, P. and Zhou, Z. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1-2):465–507, January 2019.
- [32] Mertikopoulos, P., Papadimitriou, C. H., and Piliouras, G. Cycles in adversarial regularized learning. In *SODA '18: Proceedings of the 29th annual ACM-SIAM Symposium on Discrete Algorithms*, 2018.
- [33] Meyer, D. A. Quantum strategies. *Physical Review Letters*, 82(5):1052–1055, February 1999.
- [34] Petersen, K. B. and Pedersen, M. S. The matrix cookbook, nov 2012.
- [35] Piliouras, G. and Shamma, J. S. Optimization despite chaos: Convex relaxations to complex limit sets via Poincaré recurrence. In *SODA '14: Proceedings of the 25th annual ACM-SIAM Symposium on Discrete Algorithms*, 2014.
- [36] Preskill, J. Quantum computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018.
- [37] Rockafellar, R. T. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.
- [38] Rockafellar, R. T. and Wets, R. J. B. *Variational Analysis*, volume 317 of *A Series of Comprehensive Studies in Mathematics*. Springer-Verlag, Berlin, 1998.
- [39] Sandholm, W. H. *Population Games and Evolutionary Dynamics*. MIT Press, Cambridge, MA, 2010.
- [40] Scutari, G., Facchinei, F., Palomar, D. P., and Pang, J.-S. Convex optimization, game theory, and variational inequality theory in multiuser communication systems. *IEEE Signal Process. Mag.*, 27(3): 35–49, May 2010.
- [41] Shalev-Shwartz, S. Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194, 2011.
- [42] Shalev-Shwartz, S. and Singer, Y. Convex repeated games and Fenchel duality. In *NIPS' 06: Proceedings of the 19th Annual Conference on Neural Information Processing Systems*, pp. 1265–1272. MIT Press, 2006.
- [43] Shapiro, A., Dentcheva, D., and Ruszczyński, A. *Lectures on Stochastic Programming*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2009.
- [44] Slivkins, A. Introduction to multi-armed bandits. *Foundations and Trends in Machine Learning*, 12(1-2):1–286, November 2019.
- [45] Taylor, P. D. and Jonker, L. B. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40(1-2):145–156, 1978.
- [46] Tsuda, K., Rätsch, G., and Warmuth, M. K. Matrix exponentiated gradient updates for on-line Bregman projection. *Journal of Machine Learning Research*, 6:995–1018, 2005.
- [47] Viossat, Y. and Zapechelnnyuk, A. No-regret dynamics and fictitious play. *Journal of Economic Theory*, 148(2):825–842, March 2013.
- [48] Weibull, J. W. *Evolutionary Game Theory*. MIT Press, Cambridge, MA, 1995.
- [49] Wilcox, R. M. Exponential operators and parameter differentiation in quantum physics. *Journal of Mathematical Physics*, 8(4):962–982, 1967.
- [50] Zhong, H.-S., Wang, H., Deng, Y.-H., Chen, M.-C., Peng, L.-C., Luo, Y.-H., Qin, J., Wu, D., Ding, X., Hu, Y., Hu, P., Yang, X.-Y., Zhang, W.-J., Li, H., Li, Y., Jiang, X., Gan, L., Yang, G., You, L., Wang, Z., Li, L., Liu, N.-L., Lu, C.-Y., and Pan, J.-W. Quantum computational advantage using photons. *Science*, 370(6523):1460–1463, December 2020.
- [51] Zimmert, J. and Seldin, Y. Tsallis-INF: An optimal algorithm for stochastic and adversarial bandits. *Journal of Machine Learning Research*, 22(28):1–49, 2021.