

Riemannian Optimization for Multi-Player Quantum Games on Product Unitary Manifolds

ALIREZA HABIBI^{*†}, SETAREH MAGHSUDI^{†‡}

[†]Faculty of Electrical Engineering and Information Technology, Ruhr University Bochum, Bochum, Germany,

[‡] Faculty of Computer Science, Ruhr University Bochum, Bochum, German

^{*}Corresponding author: Alireza Habibi (email: alireza.habibi@ruhr-uni-bochum.de)

Abstract—Quantum game theory is an extension of classical game theory that uses quantum principles in game theory. The Eisert–Wilkens–Lewenstein (EWL) quantum game is an early example of the two-player classical Prisoner’s Dilemma transformed into a quantum Prisoner’s Dilemma. In the EWL game, the players choose pure quantum strategies represented by unitary matrices. This extension can resolve the classical dilemma by enabling cooperative equilibrium with higher payoff. In this paper, we first discuss the Extended EWL (EEWL) for multiplayer quantum games with mixed strategies. In EEWL, each player controls a set of unitary operators as quantum actions and uses a classical mixed strategy over these actions. The payoffs are defined as expectation values of Hermitian reward operators acting on a shared quantum state, which is generated and measured according to the EEWL protocol. We then propose the Unitary Strategy Matrix Exponential Algorithm (USMEA), a geometry-aware sequential algorithm for the EEWL mixed-strategy setting, in which each player jointly learns a trainable set of local unitary actions and the associated classical mixing probabilities. Thereby it acts as a learning-and-control layer for multi-agent quantum decision systems. We analyze the convergence properties of USMEA under standard smoothness and step-size conditions and validate the theory with numerical experiments. These results show how classical optimization methods can be systematically integrated into the design and analysis of engineered quantum strategic interactions.

I. INTRODUCTION

Decision-making is a fundamental challenge in different fields like economics, biology, management, and artificial intelligence. The typical decision-making process includes multiple steps, like analyzing available information, predicting possible outcomes, and selecting the best strategy from different choices. Game theory studies the decision-making process in which the payoff to a player depends on its own strategies as well as those of others. [1], [2]. When a player chooses a single action, this is called the pure strategy. A mixed strategy means that a player selects between multiple strategies based on a probability distribution. In competitive or uncertain situations where uncertainty exists, mixed strategies can offer a strategic advantage [3], [4].

One of the most popular games is prisoner’s dilemma. The classical prisoner’s dilemma shows how two rational players might choose not to cooperate, even when mutual cooperation would result in a better outcome for both [5]–[7]. In the EWL quantum game, introducing entanglement and allowing players to use 2×2 unitary operators with two free parameters as pure

quantum strategies change the nature of the equilibrium point in the prisoner’s dilemma. The resulting equilibrium becomes more efficient and fairer for the players. Under these conditions, the Nash equilibrium aligns with the Pareto optimal outcome, which resolves the dilemma that exists in the classical version of the game [8]. However, when players are allowed to use unitary operators with three free parameters, the situation changes. In this case, a Nash equilibrium no longer exists, since each action can be countered by an opposing move that reduces the expected payoff. When the EWL game is extended to include mixed quantum strategies, the Nash equilibrium reemerges. This new equilibrium can yield a higher expected payoff than its classical counterpart [9]. In this scenario, in the mixed-strategy quantum setting, the equilibrium or fixed-point structure may fail to be isolated. Since then, the EWL game has been adapted for multiplayer games and integrated into larger quantum strategic environments. Researchers have also investigated alternative forms of quantum games beyond the initial formulation [10]–[12]. Flitney *et al.* examined the impact of decoherence in the EWL quantum game and showed that the quantum advantage is diminished as decoherence increases [13]. Xu *et al.* implemented the EWL quantum game in the IBM quantum computer quantum hardware and showed that the difference between classical simulation and quantum execution was approximately 1% in this game [14]. The EWL quantum game has been used in many different areas like finance, quantum communication, cryptography, algorithmic trading, cognitive modeling, and economics [15]–[21]. With the growing interest in quantum computing and quantum information science in recent years, the exploration of quantum game theory has gained significant attention [22]–[24]. While quantum hardware continues to advance and become more accessible, theoretical research in this field is still important. This kind of study helps to design new protocols, cover fundamental quantum advantages, and guide the creation of future quantum algorithms and applications in quantum game theory.

In game theory, gradient-based methods are widely used to enable players to iteratively refine their actions. However, when unitary matrices represent actions, optimization becomes challenging. Classical gradient descent with constraints in Euclidean space is inefficient and difficult to scale for unitary matrices [25], [26]. In these cases, the optimization in the Rie-

mannian manifold is a suitable tool to address the challenge. Riemannian gradient descent projects gradients onto tangent spaces and maps updates back to the manifold. Such a method has seen success in low-rank matrix learning, signal processing, and deep learning with orthogonality constraints [27]. The application of this idea to quantum games with mixed strategies has not been investigated. Most existing approaches use a fixed set of unitary actions or predefine the classical probabilities for mixing strategies [9], [25], [28].

We consider an extension of EWL quantum game in which each player jointly learns a set of local unitary actions and a classical mixed strategy over these actions. This leads to a hybrid decision space given by probability simplices and a product of unitary manifolds. We then study USMEA as a geometry-aware sequential method for the coupled multi-agent dynamics induced by this formulation. In addition, this formulation is a coupled multi-agent problem that is non-convex, generally non-monotone, and not, in general, a potential game. Moreover, due to coupling and symmetry, the equilibrium and fixed-point structure can be nontrivial, and the fixed points may not be isolated, which makes even local convergence analysis meaningful in this setting.

Our contributions: Our main contributions in this paper are as follows.

- We formulate the EEWL framework for multiplayer quantum games with mixed strategies, where each player jointly learns a trainable set of local unitary actions and a classical mixing distribution over these actions.
- We propose USMEA as a geometry-aware sequential learning scheme for jointly updating the unitary actions and the mixing probabilities.
- We show how standard mixed-strategy equilibrium-existence arguments apply to the EEWL model, and we derive the corresponding player-wise gradients, smoothness bounds, and step-size conditions for the update dynamics.
- In the experimental results, we study the convergence behavior of USMEA.

In this work, all learning and optimization steps of USMEA are carried out on a classical computer. From an engineering perspective, the proposed framework can be viewed as a learning-and-control layer sitting on top of a quantum system where multiple agents act locally on a shared entangled state [29]–[31]. The EEWL game shows how local unitary actions on a shared state can result to payoffs, while USMEA gives a systematic approach to adjust these actions and their mixing probabilities using Riemannian optimization. This links quantum game theory to the development of multi-agent controllers and learning-based protocols in emerging quantum technologies.

In Section II, we introduce the notation and basic concepts from quantum information, quantum game theory, and Riemannian manifold that we use in the rest of the paper. Section III presents the EEWL model for multi-player quantum games with a general strategy space. Section IV introduces the USMEA update rules. In Section V, we study

the convergence and performance of the proposed algorithm. Section VI provides experimental results for the algorithm's performance in quantum games. In Section VII, we summarize the main findings of this work, discuss their implications, and outline potential directions for future works.

In addition, for readers with a game-theoretic background and limited familiarity with quantum basics, we provided a concise overview of the quantum mechanics used in this paper in Supplementary Material, Section S-I. The Supplementary Material also contains additional theorems and all proofs in Section S-II, as well as detailed descriptions of the experimental setups.

II. NOTATIONS

This section introduces the basic notations of quantum game and the Riemannian manifold structure for unitary operators.

A. Quantum game notations

Let $\mathcal{H} \cong \mathbb{C}^d$ be a finite-dimensional complex Hilbert space, where d is the dimension of the Hilbert space [32]. For an N -player quantum game, each player i controls a local Hilbert space $\mathcal{H}^{(i)}$ of dimension d_i . The global Hilbert space of this composite system is given by the tensor product of the local Hilbert spaces,

$$\mathcal{H} = \bigotimes_{i=1}^N \mathcal{H}^{(i)}, \quad d = \dim \mathcal{H} = \prod_{i=1}^N d_i, \quad (1)$$

where $d_i = \dim(\mathcal{H}^{(i)})$ is the dimension of local Hilbert space i . Quantum operators are denoted by bold capital letters. In particular, $\mathbf{U}^{(i)} \in \mathcal{U}(d_i)$ denotes the unitary operator used by player i as its pure quantum strategy, and its dimension is $d_i \times d_i$. To reflect the tensor-product structure, we can reindex the matrix entries $O_{m,n}$ of operator $\mathbf{O}$ using multi-indices $m \mapsto (j_1, \dots, j_N)$ and $n \mapsto (j'_1, \dots, j'_N)$ where $j_k, j'_k = 1, \dots, d_k$. In this case the operator becomes a $2N$ -index tensor with entries $O_{j_1, \dots, j_N, j'_1, \dots, j'_N}$.

In this paper, we employ Positive Operator-Valued Measurements (POVMs). Let us consider a set of possible measurement outcomes Ω . For each outcome $\omega \in \Omega$, we associate a positive semi-definite operator $\mathbf{P}_\omega : \mathcal{H} \rightarrow \mathcal{H}$ satisfying the completeness relation $\sum_{\omega \in \Omega} \mathbf{P}_\omega = \mathbf{I}$. The probability of obtaining a specific measurement outcome ω when the system is in the quantum state described by the density matrix ρ is given by $p_\omega = \text{Tr}(\mathbf{P}_\omega \rho)$. In addition, the completeness relation ensures that the total probability across all outcomes sums to one ($\sum_{\omega \in \Omega} p_\omega = 1$). In a quantum game, a set of outcomes is used to compute and allocate payoffs to players.

Small capital letters are used for vectors. We use the shorthand $(\mathbf{j}) \equiv (j_1, j_2, \dots, j_N)$, where j_i denotes the index corresponding to the player i . The notation $(j'_i; \mathbf{j}_{-i}) = (j_1, \dots, j'_i, \dots, j_N)$ shows that the index j_i for player i has been replaced by j'_i . The term $(\mathbf{j}_{-i})$ refers to the tuple $(j_1, j_2, \dots, j_{i-1}, j_{i+1}, \dots, j_N)$, which omits the entry for player i . In addition, we use d_{-i} as the product of all subsystem dimensions except for i , that is $d_{-i} = \prod_{k \neq i} d_k$. The

softmax function of a vector $\mathbf{a} = (a_1, \dots, a_m)$ at temperature T is defined as

$$\sigma_T(\mathbf{a})_j = \frac{e^{a_j/T}}{\sum_{k=1}^m e^{a_k/T}}, \quad (2)$$

with $0 \leq \sigma_T(a)_j \leq 1$ for all j . The $\sigma_T(\mathbf{a})$ is a probability vector with $\sum_{j=1}^m \sigma_T(a)_j = 1$.

We use the superscript (s) to indicate a strategy profile of local strategies applied independently by each player in product space.

$$\mathbf{U}^{(s)} = \mathbf{U}^{(1)} \otimes \mathbf{U}^{(2)} \otimes \dots \otimes \mathbf{U}^{(N)}, \quad (3a)$$

$$\begin{aligned} \mathbf{U}^{(s,i)} &= \mathbf{I}^{(1)} \otimes \dots \otimes \mathbf{U}^{(i)} \otimes \dots \otimes \mathbf{I}^{(N)} \\ &\equiv \mathbf{U}^{(i)} \otimes \mathbf{I}^{(-i)}, \end{aligned} \quad (3b)$$

$$\mathbf{U}^{(s,-i)} = \mathbf{U}^{(1)} \otimes \dots \otimes \mathbf{I}^{(i)} \otimes \dots \otimes \mathbf{U}^{(N)}, \quad (3c)$$

$$\begin{aligned} (\mathbf{U}'^{(i)}; \mathbf{U}^{(s,-i)}) &= (\mathbf{U}^{(1)} \otimes \dots \otimes \mathbf{U}'^{(i)} \otimes \dots \otimes \mathbf{U}^{(N)}) \\ &\equiv \mathbf{U}'^{(i)} \otimes \mathbf{U}^{(s,-i)}, \end{aligned} \quad (3d)$$

where $\dim(\mathbf{I}^{(i)}) = \dim(\mathbf{U}^{(i)})$. For the complex conjugate of a number a , we use the notation $\bar{a}$. The trace operation is defined as

$$\text{Tr}(\mathbf{A}) = \sum_{\mathbf{j}} A_{\mathbf{j},\mathbf{j}}, \quad (4a)$$

$$\text{Tr}_{-i}(\mathbf{A}) \Big|_{j_i, j'_i} = \sum_{\mathbf{j}_{-i}} A_{(j_i, \mathbf{j}_{-i}), (j'_i, \mathbf{j}_{-i})}, \quad (4b)$$

where Tr_{-i} denotes the partial trace over all indices except for subsystem i .

For any Hermitian operator $\mathbf{A} \in \mathbb{C}^{d \times d}$ with real eigenvalues $\{\lambda_i\}_{i=1}^d$, the operator (spectral) norm and the Frobenius (Hilbert–Schmidt) norm are defined as follows (see, e.g., [33]–[35]):

$$\|\mathbf{A}\|_{\text{op}} = \sup_{\|\mathbf{v}\|_2=1} \|\mathbf{A}\mathbf{v}\|_2 = \max_j |\lambda_j|, \quad (5)$$

$$\|\mathbf{A}\|_F = \sqrt{\text{Tr}(\mathbf{A}^\dagger \mathbf{A})} = \left(\sum_{i=1}^d \lambda_i^2 \right)^{1/2}, \quad (6)$$

where $\mathbf{v} \in \mathbb{C}^d$ is a unit vector. These norms satisfy the standard inequalities

$$\|\mathbf{A}\|_{\text{op}} \leq \|\mathbf{A}\|_F \leq \sqrt{d} \|\mathbf{A}\|_{\text{op}}, \quad \|\mathbf{A}\mathbf{B}\|_F \leq \|\mathbf{A}\|_{\text{op}} \|\mathbf{B}\|_F. \quad (7)$$

B. Riemannian Manifold

A Riemannian manifold $(\mathcal{M}, g)$ is a finite-dimensional smooth manifold $\mathcal{M}$ of dimension n (with tangent bundle $(T\mathcal{M})$ equipped with a Riemannian metric g). A Riemannian metric g on $\mathcal{M}$ is a smoothly varying field of inner products $x \mapsto g_x(\cdot, \cdot)$ on the tangent spaces $T_x \mathcal{M}$ for each $x \in \mathcal{M}$. In this paper, we consider $\mathcal{M}_{\mathcal{U}}$ to be the unitary group $\mathcal{U}(n)$ and equip it with the standard bi-invariant Riemannian metric derived from the Frobenius inner product. At a point $\mathbf{U} \in \mathcal{U}(d_i)$, the tangent space is $T_{\mathbf{U}} \mathcal{U}(d_i) = \{\mathbf{X}\mathbf{U} | \mathbf{X}^\dagger = -\mathbf{X}\}$, where $\mathbf{X}$ is skew-Hermitian. For any skew-Hermitian $\mathbf{X}$, the curve $\gamma(t) = \exp(t\mathbf{X})\mathbf{U}$ (with $t \in \mathbb{R}$) traces a

path along the manifold. In practice, all updates to unitary actions are performed using exponential retractions of the form $\exp(\eta \mathbf{X})\mathbf{U}$. The authors recommend references [26], [36], [37] for a more comprehensive introduction.

III. MODEL

In this section, we first provide the definition of the EEWL quantum game. Next, we define the best response (BR) and the Nash equilibrium. Finally, we discuss the existence of Nash equilibrium points within this quantum game.

Definition 1: An EEWL quantum game is a tuple $\mathcal{Q} = \langle \mathcal{N}, \mathcal{H}, \mathcal{S}, r \rangle$ defined as,

- $\mathcal{N} = \{1, 2, \dots, N\}$: A finite set of N players.
- $\mathcal{H} = \mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)} \otimes \dots \otimes \mathcal{H}^{(N)}$: The Hilbert space of the game. Each player $i \in \mathcal{N}$ access a local Hilbert space $\mathcal{H}^{(i)} \cong \mathbb{C}^{d_i}$.
- $\mathcal{S}^{(i)} = \left\{ \mathbf{U}_{j_i}^{(i)} \right\}_{j_i=1}^{m_i}$: The set of actions (pure strategies) available to player i . Each action $\mathbf{U}_{j_i}^{(i)} \in \mathcal{U}(d_i)$ acts on $\mathcal{H}^{(i)}$. The $m_i \geq 1$ is the total number of actions that player i can choose.
- $\mathcal{S} = \mathcal{S}^{(1)} \otimes \mathcal{S}^{(2)} \otimes \dots \otimes \mathcal{S}^{(N)}$: The joint strategy space. A pure joint quantum strategy profile is the tensor product,

$$\mathbf{U}_{\mathbf{j}}^{(s)} = \mathbf{U}_{j_1}^{(1)} \otimes \mathbf{U}_{j_2}^{(2)} \otimes \dots \otimes \mathbf{U}_{j_N}^{(N)}. \quad (8)$$

- The reward or payoff function r : Each player $i \in \mathcal{N}$ receives a payoff determined by an function $r^{(i)} : \rho_{\mathbf{j}} \rightarrow \mathbb{R}$. The density matrix $\rho_{\mathbf{j}} = \mathbf{U}_{\mathbf{j}}^{(s)} \rho_0 \mathbf{U}_{\mathbf{j}}^{(s)\dagger}$ is the final state of the system when all players use the joint strategy profile $\mathbf{U}_{\mathbf{j}}^{(s)}$, and ρ_0 denotes the initial state of the quantum system. To calculate the payoff, one can use Hermitian payoff operator $\mathbf{R}_i : \mathcal{H} \rightarrow \mathcal{H}$, such that

$$r_{\mathbf{j}}^{(i)} = \text{Tr}(\mathbf{R}_i \rho_{\mathbf{j}}). \quad (9)$$

Mixed Strategy Setup: When players adopt mixed strategies, they select among actions according to a classical probability distribution. This probabilistic choice is described by the vector

$$\mathbf{p}^{(i)} = (p_1^{(i)}, p_2^{(i)}, \dots, p_{m_i}^{(i)}) \in \Delta^{m_i-1}, \quad (10)$$

where Δ^{m_i-1} is the (m_i-1) -dimensional probability simplex. The set of (mixed quantum) strategies of player i is

$$z^{(i)} = \left(\{\mathbf{U}_{j_i}^{(i)}\}, \mathbf{p}^{(i)} \right) = \left\{ (\mathbf{U}_{j_i}^{(i)}, p_{j_i}^{(i)}) \mid \sum_{j_i=1}^{m_i} p_{j_i}^{(i)} = 1 \right\}. \quad (11)$$

The joint mixed (quantum) strategies for all players is given by $z = (z^{(1)}, \dots, z^{(N)})$, which abbreviates the collection of all action sets and probability distributions. We use $z^{(-i)} = (\mathbf{U}^{(-i)}, \mathbf{p}^{(-i)})$ as the collection of action sets and probability distributions of all players except player i . The expected payoff for player i is

$$\bar{r}^{(i)}(z) = \sum_{\mathbf{j}} p_{\mathbf{j}} r^{(i)}(\mathbf{U}_{\mathbf{j}}^{(s)}), \quad (12)$$

where $p_j = \Pi_i p_{j_i}^{(i)}$ is the joint classical probability of the strategy profile. We rewrite the expected payoff of player i as $r^{(i)} = \sum_{j_i} p_{j_i}^{(i)} \ell_{j_i}^{(i)}$, where $\ell_{j_i}^{(i)}$ is the per-action payoff and defined as the expected payoff obtained by player i when using the pure strategy $\mathbf{U}_{j_i}^{(i)}$ while the other players follow their current mixed strategies,

$$\ell_{j_i}^{(i)} = \sum_{\mathbf{j}_{-i}} p_{\mathbf{j}_{-i}} r^{(i)} \left(\mathbf{U}_{j_i \mathbf{j}_{-i}}^{(s)} \right). \quad (13)$$

The vector of per-action payoffs is denoted by $\ell^{(i)} = (\ell_1^{(i)}, \dots, \ell_{m_i}^{(i)})$.

Remark 1: The expected payoff satisfies the following properties:

- 1) **Phase invariance:** The expected payoff is invariant under global phase shifts.
- 2) **Boundedness:** The payoff is bounded both from below and above by the extremal eigenvalues of the payoff operator $\mathbf{R}_i$,

$$\lambda_{\min}(\mathbf{R}_i) \leq r^{(i)} \leq \lambda_{\max}(\mathbf{R}_i), |r^{(i)}| \leq \|\mathbf{R}_i\|_{\text{op}}, \quad (14)$$

where $\lambda_{\min}(\mathbf{R}_i)$ and $\lambda_{\max}(\mathbf{R}_i)$ are the smallest and largest eigenvalues of $\mathbf{R}_i$ [33]. These bounds also hold for the expected payoff under a mixed strategy.

The payoff operators and density matrices are defined on compact spaces. The payoff functions are formulated over a compact strategy space consisting of unitary matrices. This strategy space is composed of differentiable operators, such as unitary transformations and inner products, which ensure finite outputs, smooth gradients, and controlled variations. Thanks to these well-defined mathematical conditions, we can systematically study equilibria and strategic behavior in quantum games more easily.

Best Response (BR): When other players' strategies $z^{(-i)}$, player i 's best response set is

$$\text{BR}^{(i)}(z^{(-i)}) = \arg \max_{z^{(i)}} \bar{r}^{(i)}(z^{(i)}; z^{(-i)}). \quad (15)$$

Any $z^{(i)*} \in \text{BR}^{(i)}(z^{(-i)})$ is a best response to $z^{(-i)}$ which maximizes player i 's expected payoff while the other players use $z^{(-i)}$.

Nash Equilibrium: A joint mixed strategy z^* is a Nash equilibrium if for every player i ,

$$z^{(i)*} = \arg \max_{z^{(i)}} \bar{r}^{(i)}(z^{(i)}; z^{(-i)*}). \quad (16)$$

which means no player can improve their expected payoff by unilaterally changing $z^{(i)*}$.

Theorem 1: In the mixed strategy EEWL game defined in Definition 1, there exists at least one mixed strategy Nash equilibrium.

Remark 2: If each player is restricted to a finite set of fixed set of non-equivalent actions, the game reduces to a finite game, and the existence of a mixed strategy Nash equilibrium follows directly from Nash's classical theorem [38].

Remark 3: If the set of actions contains only a single action for each player ($m_i = 1$), the player can still adjust their

unitary strategy to search for an optimal solution. However, in this case, Theorem 1 no longer applies, and the existence of a Nash equilibrium is no longer guaranteed.

IV. USMEA ALGORITHM

In this section, we first define the decision manifold and the loss function. Next, we describe our approach to optimizing player strategies. Finally, we introduce the USMEA algorithm, which minimizes these loss functions.

The strategies live on the product manifold as the decision manifold in the EEWL quantum game,

$$\mathcal{M} = \underbrace{\prod_{i=1}^N \prod_{j_i=1}^{m_i} \mathcal{U}(d_i)}_{\mathcal{M}_{\mathcal{U}}} \times \underbrace{\prod_{i=1}^N \Delta^{m_i-1}}_{\mathcal{P}}, \quad (17)$$

with two geometries: (i) a Riemannian geometry on the unitary manifold $\mathcal{M}_{\mathcal{U}}$, (ii) a simplex geometry on the probability simplex $\mathcal{P}$. Since both $\mathcal{M}_{\mathcal{U}}$ and $\mathcal{P}$ are compact, their product space $\mathcal{M}$ is also compact. Thus $\mathcal{M}$ is compact.

The payoff function in EEWL quantum games is non-convex, and multiple fixed points may exist. Consequently, standard global guarantees from convex optimization do not apply. The obtained solutions can depend on both the initialization and the chosen regularization parameters.

For optimization, we use a loss function with an entropy regularization term to ensure a proper balance between exploration and exploitation during learning. The learning objective is to minimize the entropy-regularized loss,

$$L^{(i)}(z) = -\bar{r}^{(i)}(z) - TH(\mathbf{p}^{(i)}), \quad (18)$$

where $H(\mathbf{p}^{(i)}) = -\sum_{j_i} p_{j_i}^{(i)} \log p_{j_i}^{(i)}$ is the Shannon entropy, and $T > 0$ is the temperature parameter. From a game-theoretic perspective, T acts as a bounded-rationality or decision-noise parameter. It controls the trade-off between exploration and exploitation in the mixed-strategy update. When T is large, the player's response becomes smoother and more exploratory. In this regime, the player is less sensitive to payoff differences, so its strategy appears more random. When T is small, the update becomes more sharply concentrated on higher-payoff actions and is closer to a best response.

In USMEA, each player first updates their set of actions and then updates their mixed strategy using a softmax function applied to the current payoffs. For temperature $T > 0$ and step size $\eta_i > 0$, the USMEA block map $\mathcal{T}_i = \mathcal{T}_i^P \circ \mathcal{T}_i^U$ updates $z^{(i)}$ as

$$\mathcal{T}_i^U : \quad \mathbf{U}_{j_i}^{(i)[t+1]} = \exp(\eta_i \mathbf{G}_{j_i}^{(i)[t]}) \mathbf{U}_{j_i}^{(i)[t]}, \quad \forall j_i, \quad (19a)$$

$$\mathcal{T}_i^P : \quad \mathbf{p}^{(i)[t+1]} = \sigma_T(\ell^{(i)[t+1]}), \quad (19b)$$

where $\mathbf{G}_{j_i}^{(i)}$ is the Riemannian gradient of the expected payoff for player i over action j_i is

$$\mathbf{G}_{j_i}^{(i)} = \sum_{\mathbf{j}_{-i}} p_{(j_i; \mathbf{j}_{-i})} \text{Tr}_{-i} \left([\mathbf{R}_i, \mathbf{U}_{j_i \mathbf{j}_{-i}}^{(s)} \rho_0 \mathbf{U}_{j_i \mathbf{j}_{-i}}^{(s)\dagger}] \right). \quad (20)$$

In Eq. (19a), the Riemannian gradient is skew-Hermitian, $\mathbf{G}_{j_i}^{(i)[t]} = -\mathbf{G}_{j_i}^{(i)[t]\dagger}$. Therefore, $\exp(\eta_i \mathbf{G}_{j_i}^{(i)[t]})$ is unitary. Since the product of two unitary matrices is also unitary, the updated action remains in $\mathcal{U}(d_i)$. The Eq. (19b) is the softmax update with $p_{j_i}^{(i)[t+1]} \geq 0$ and $\sum_{j_i} p_{j_i}^{(i)[t+1]} = 1$. Therefore, the action-update map $\mathcal{T}_i^U$ preserves the unitary manifold, and the probability-update map $\mathcal{T}_i^P$ preserves the simplex. Thus, $\mathcal{T}_i : \mathcal{M} \rightarrow \mathcal{M}$ is a well-defined self-map. As a result, the algorithm requires no additional constraints to preserve the unitarity of actions or the normalization of probabilities.

The algorithm can be used to compute both BRs and fixed-point equilibrium. A BR iteration updates $z^{(i)}$ and fixed $z^{(-i)}$. To compute the fixed-point equilibrium, the algorithm follows a Gauss–Seidel update scheme, where players are updated sequentially within each iteration. It is shown that the sequential update is often more locally stable than simultaneous updates of all players at once [39], [40]. In the sequential updates, at iteration t and intermediate step i we have $z^{[t,i]} = \mathcal{T}_i(z^{[t,i-1]})$, with $z^{[t+1,0]} = z^{[t,N]}$, where $z^{[t+1,0]}$ represents the beginning of iteration $t+1$. Equivalently, the composition $\mathcal{T} = \mathcal{T}_N \circ \dots \circ \mathcal{T}_1$ is called the iteration map or the one-sweep map. One iteration satisfies $z^{[t+1,0]} = \mathcal{T}(z^{[t,0]})$.

In a standard Riemannian gradient descent combined with softmax (RGD+Softmax), all variables are updated at once. By contrast, USMEA uses a player-wise sequential block-update scheme, in which the players are updated in Gauss–Seidel order. This distinction matters in the coupled multi-agent EEWL game, because each player tries to minimize its own loss, while updating one player's block can also change the losses of the others through the coupled dynamics.

We present the procedure for finding the best response $z^{(i)*}$ in Algorithm 1 and the fixed point z^* in Algorithm 2.

Algorithm 1 USMEA algorithm to find BR

Require: η_i : learning parameter, ϵ : convergence tolerance, i : player index, T : initial temperature for classical probabilities, $\mathbf{U}_{j_{-i}}^{(-i)}$: Set of all players' action except player i

- 1: Initialize a random set of actions $\{\mathbf{U}_1^{(i)[0]}, \mathbf{U}_2^{(i)[0]}, \dots, \mathbf{U}_{m_i}^{(i)[0]}\}$ and corresponding probabilities $\{p_1^{(i)[0]}, p_2^{(i)[0]}, \dots, p_{m_i}^{(i)[0]}\}$ for player i
 - 2: **repeat** ($k = 0, 1, \dots$: each step)
 - 3: Calculate $\{\mathbf{G}_{j_i}^{(i)[k]}\}_{j_i=1}^{m_i}$
 - 4: **for every** $j_i \in \{1, 2, \dots, m_i\}$ **do**
 - 5: $\mathbf{U}_{j_i}^{(i)[k+1]} \leftarrow e^{\eta_i \mathbf{G}_{j_i}^{(i)[k]}} \mathbf{U}_{j_i}^{(i)[k]}$
 - 6: **end for**
 - 7: Calculate $\ell^{(i)[k+1]}$
 - 8: $\mathbf{p}^{(i)[k+1]} \leftarrow \sigma_T(\ell^{(i)[k+1]})$
 - 9: Update T
 - 10: **until** $\left\| \mathbf{U}_{j_i}^{(i)[k+1]} - \mathbf{U}_{j_i}^{(i)[k]} \right\| \leq \epsilon$ and $\left\| p_{j_i}^{(i)[k+1]} - p_{j_i}^{(i)[k]} \right\| \leq \epsilon$ for all i, j_i
-

Algorithm 2 USMEU algorithm to find fixed points

Require: $\{\eta_i\}$: learning parameter for USMEU, ϵ : convergence tolerance, T : initial temperature for classical probabilities, ϵ : convergence tolerance

- 1: Initialize a random set of quantum actions $\{\mathbf{U}_1^{(i)[0]}, \mathbf{U}_2^{(i)[0]}, \dots, \mathbf{U}_{m_i}^{(i)[0]}\}$ and corresponding probabilities $\{p_1^{(i)[0]}, p_2^{(i)[0]}, \dots, p_{m_i}^{(i)[0]}\}$ for each player $i \in \mathcal{N}$
 - 2: **repeat** ($k = 0, 1, \dots$: each step)
 - 3: **for every** player $i \in \mathcal{N}$ **do**
 - 4: Calculate $\{\mathbf{G}_{j_i}^{(i)[k]}\}_{j_i=1}^{m_i}$ according to the GS update rule.
 - 5: **for every** $j_i \in \{1, 2, \dots, m_i\}$ **do**
 - 6: $\mathbf{U}_{j_i}^{(i)[k+1]} \leftarrow e^{\eta_i \mathbf{G}_{j_i}^{(i)[k]}} \mathbf{U}_{j_i}^{(i)[k]}$
 - 7: **end for**
 - 8: Calculate $\ell^{(i)[k+1]}$
 - 9: $\mathbf{p}^{(i)[k+1]} \leftarrow \sigma_T(\ell^{(i)[k+1]})$
 - 10: **end for**
 - 11: Update T
 - 12: **until** $\left\| \mathbf{U}_{j_i}^{(i)[k+1]} - \mathbf{U}_{j_i}^{(i)[k]} \right\| \leq \epsilon$ and $\left\| p_{j_i}^{(i)[k+1]} - p_{j_i}^{(i)[k]} \right\| \leq \epsilon$ for all i, j_i
-

V. THEORETICAL ANALYSIS

In this section, we present the theoretical analysis of the proposed USMEA algorithm. First, we derive the Riemannian gradient from the payoff function. We then compute the Lipschitz coefficients and identify a safe interval for the learning rate. Finally, we prove the results for the BR updates and extend them to the full sequential setting to obtain fixed-point equilibria of the joint dynamics.

Each unitary action j_i is updated on the Riemannian manifold using the exponential retraction defined in Eq. (19a). The term $\mathbf{G}_{j_i}^{(i)} = \text{grad}_{\mathbf{U}_{j_i}^{(i)}} L^{(i)}$ represents the Riemannian gradient on the unitary group manifold [36].

Theorem 2: Let $\mathbf{G}_{j_i}^{(i)}$ be the Riemannian gradient associated with the unitary updates. Under the USMEA update rules, we have

- 1) $\text{grad}_{\mathbf{U}_{j_i}^{(i)}} L^{(i)} = -\text{grad}_{\mathbf{U}_{j_i}^{(i)}} \bar{r}^{(i)}$.
- 2) The explicit form of the Riemannian gradient is given by Eq. (20).

The payoff operator $\mathbf{R}_i$ is bounded, $\|\mathbf{R}_i\|_{\text{op}} < \infty$. The map $\rho \mapsto \bar{r}^{(i)}$ is smooth over the manifold $\mathcal{M}$. Furthermore, since $\mathbf{R}_i$, ρ_0 , and the unitary operators $\mathbf{U}_{j_i}^{(i)}$ are analytic, the expected payoff $\bar{r}^{(i)}$ is itself an analytical function. Since $\mathcal{M}$ is compact and $\bar{r}^{(i)}$ is analytic, its derivative is bounded on $\mathcal{M}$. Therefore $\bar{r}^{(i)}$ is Lipschitz continuous on the manifold $\mathcal{M}$. Estimating the Lipschitz coefficients of the payoff gradients on $\mathcal{M}$ lets us choose safe learning rates and bound the per-iteration change in $L^{(i)}$. To begin with, we define the distance between two action sets on $\mathcal{M}$. Then, we compute the corresponding Lipschitz coefficients.

Definition 2: The geodesic distance between unitary matrices and the distance between probability values are defined as follows

- 1) For player i , the block (geodesic) distance between two unitary matrices $\mathbf{U}_{j_i}^{(i)}$ and $\mathbf{U}_{j_i}'^{(i)}$ is [41]

$$d_{\mathcal{U}}(\mathbf{U}_{j_i}^{(i)}, \mathbf{U}_{j_i}'^{(i)}) = \left\| \log \left(\mathbf{U}_{j_i}^{(i)\dagger} \mathbf{U}_{j_i}'^{(i)} \right) \right\|_F. \quad (21)$$

- 2) For player i , the distance between two unitary sets $\{\mathbf{U}_{j_i}^{(i)}\}_{j_i=1}^{m_i}$ and $\{\mathbf{U}_{j_i}'^{(i)}\}_{j_i=1}^{m_i}$, and the probability vectors $\mathbf{p}^{(i)}$ and $\mathbf{p}'^{(i)}$ is

$$\Delta_{\mathcal{U}}^{(i)}(\{\mathbf{U}_{j_i}^{(i)}\}, \{\mathbf{U}_{j_i}'^{(i)}\}) = \sum_{j_i=1}^{m_i} d_{\mathcal{U}}(\mathbf{U}_{j_i}^{(i)}, \mathbf{U}_{j_i}'^{(i)}), \quad (22a)$$

$$\Delta_p^{(i)}(\mathbf{p}^{(i)}, \mathbf{p}'^{(i)}) = \left\| \mathbf{p}^{(i)} - \mathbf{p}'^{(i)} \right\|_2. \quad (22b)$$

The total distance for the player i is then defined by combining these via the Euclidean norm

$$\Delta^{(i)} = \left\| (\Delta_{\mathcal{U}}^{(i)}, \Delta_p^{(i)}) \right\| = \sqrt{(\Delta_{\mathcal{U}}^{(i)})^2 + (\Delta_p^{(i)})^2}. \quad (23)$$

- 3) When all players update their actions and probability distributions from z to z' , the distances are

$$\Delta_{\mathcal{U}}^2 = \sum_{i=1}^N \Delta_{\mathcal{U}}^{(i)}(\{\mathbf{U}_{j_i}^{(i)}\}_{j_i=1}^{m_i}, \{\mathbf{U}_{j_i}'^{(i)}\}_{j_i=1}^{m_i}), \quad (24a)$$

$$\Delta_p^2 = \sum_{i=1}^N \left\| \mathbf{p}^{(i)} - \mathbf{p}'^{(i)} \right\|_2^2. \quad (24b)$$

The total joint distance is then given by

$$\Delta(z, z') = \sqrt{\Delta_{\mathcal{U}}^2 + \Delta_p^2}. \quad (25)$$

Theorem 3: Suppose that the player i modifies only the unitary action j_i , then

- 1) The gradient $\mathbf{G}_{j_i}^{(i)}$ is Lipschitz continuous along the unitary geodesic with constant A_i as

$$\left\| \mathbf{G}_{j_i}^{(i)}(\mathbf{U}_{j_i}^{(i)}) - \mathbf{G}_{j_i}^{(i)}(\mathbf{U}_{j_i}'^{(i)}) \right\|_F \leq A_i d_{\mathcal{U}}(\mathbf{U}_{j_i}^{(i)}, \mathbf{U}_{j_i}'^{(i)}), \quad (26)$$

where

$$A_i = 4\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\rho_0\|_F. \quad (27)$$

- 2) The per-action payoff $\ell_{j_i}^{(i)}$ is Lipschitz continuous with constant M_i as

$$\left| \ell_{j_i}^{(i)}(\mathbf{U}_{j_i}^{(i)}) - \ell_{j_i}^{(i)}(\mathbf{U}_{j_i}'^{(i)}) \right| \leq M_i d_{\mathcal{U}}(\mathbf{U}_{j_i}^{(i)}, \mathbf{U}_{j_i}'^{(i)}), \quad (28)$$

where

$$M_i = 2\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\rho_0\|_F. \quad (29)$$

When the actions are updated according to USMEA, Eq. (19a), the previous probability distribution no longer

minimizes the loss function. Instead, the new minimizer of the loss is given by the softmax distribution in Eq. (19b).

Theorem 4: Fix $z^{(-i)}$. $z^{(i)} = (\{\mathbf{U}_{j_i}^{(i)}\}, \mathbf{p}^{(i)})$ to $z^{(i)'} = (\{\mathbf{U}_{j_i}'^{(i)}\}, \mathbf{p}'^{(i)})$ using USMEA. If the step size satisfies $\eta_i \leq 1/A_i$, then the update is monotonic in $L^{(i)}$, and we have the following guaranteed improvement

$$\begin{aligned} L^{(i)}(z^{(i)'}; z^{(-i)}) &\leq L^{(i)}(z^{(i)}; z^{(-i)}) \\ &\quad - \eta_i \left(1 - \frac{1}{2} \eta_i A_i \right) \Delta_{\mathcal{U}}^{(i)2} \\ &\quad - \frac{T}{2} \Delta_p^{(i)2} \end{aligned} \quad (30)$$

In particular, the right-hand side is greater unless simultaneously $\mathbf{G}_{j_i}^{(i)} = 0$ for all j_i and $\mathbf{p}'^{(i)} = \mathbf{p}^{(i)}$.

Remark 4: The bound in Eq. (30) is fully explicit in terms of (η_i, T) and the defined block constants. It shows that the USMEA update, when $z^{(-i)}$, constitutes a descent step for $L^{(i)}$. In the limit $T \rightarrow 0$, the softmax gain vanishes, and the update reduces to the payoff-based ascent term.

Theorem 5: Fix $z^{(-i)}$, and assume that $\eta_i \in (0, 1/A_i]$. Then the USMEA updates for player i converge to a fixed point $z^{(i)*} = (\mathbf{U}_{j_i}^{(i)*}, \mathbf{p}^{(i)*})$. At this point, for every j_i we have

$$\mathbf{G}_{j_i}^{(i)}(z^{(i)*}; z^{(-i)}) = 0, \quad \mathbf{p}^{(i)*} = \sigma_T(\ell^{(i)*}), \quad (31)$$

where $\ell_{s_i}^{(i)*}$ denotes the per-action expected payoff at this fixed point.

Remark 5: The step-size condition $\eta_i \in (0, 1/A_i]$ is a guaranteed safe interval for the USMEA BR update. Since A_i is derived from a worst-case Lipschitz bound on the Riemannian gradient, this interval is not necessarily sharp. In practice, larger step sizes may still work empirically.

Remark 6: The $z^{(i)*}$ is a BR-stationary point of $L^{(i)}$. It is not always locally attracting. If $z^{(i)*}$ is a strict local minimizer of $L^{(i)}$, then it is locally asymptotically stable for gradient flow and locally attracting for the BR update when the step size is sufficiently small. If it corresponds to a nondegenerate saddle of $L^{(i)}$, then it is not locally attracting. As $T \rightarrow 0$, an attracting BR fixed point corresponds to a local best response to $z^{(-i)}$ [42]–[44].

Remark 7: Convergence of $L^{(i)}$ does not necessarily imply convergence of the expected payoff $\bar{r}^{(i)}$. Indeed, it is possible for the entropy term $H^{[t]}$ and the expected payoff sequence $\bar{r}^{(i)[t]}$ to vary in such a way that their combination keeps $L^{(i)}$ constant. Therefore, convergence of $L^{(i)}$ alone is not a reliable stopping criterion for the BR update.

Corollary 1: Fix $z^{(-i)}$. Suppose $L^{(i)[t]}$ converges to a fixed point. Then the expected payoff sequence $\{\bar{r}^{(i)[t]}\}$ satisfies the oscillation bound,

$$\limsup_t \bar{r}^{(i)[t]} - \liminf_t \bar{r}^{(i)[t]} \leq T \log m_i. \quad (32)$$

By Corollary 1, the expected payoff sequence $\bar{r}^{(i)}$ may oscillate within a band of width at most $T \log m_i$. Based on Remark 7 and Corollary 1, the convergence of $L^{(i)}$ does not imply the convergence of $\bar{r}^{(i)}$ for $T > 0$. Therefore, it

is not advisable to use the convergence of $L^{(i)}$ as the sole stopping criterion in the algorithm. As $T \rightarrow 0$, this oscillation window shrinks to zero. Consequently, when $z^{(-i)}$ is fixed, the sequence $\{\bar{r}_t^{(i)}\}$ converges to the payoff of a (local) BR fixed point for player i . Now, we turn to the joint dynamics in which all players' actions and mixing strategies are updated sequentially.

Theorem 6: Consider all players update their actions and probabilities using iteration map $\mathcal{T} = \mathcal{T}_N \circ \dots \circ \mathcal{T}_1$. Then there exists at least one fixed point $z^* \in \mathcal{M}$ such that $\mathcal{T}(z^*) = z^*$. In other words,

$$\mathbf{g}_{j_i}^{(i)}(z^*) = 0, \quad p_{j_i}^{(i)*} = \sigma_T(\ell^{(i)*})_{j_i}, \quad \forall j_i, i. \quad (33a)$$

Remark 8: With a single action per player ($m_i = 1$), the entropy term in the loss function vanishes and USMEA reduces to unitary updates. In this case, a Nash equilibrium may not exist (Remark 3), but fixed points still do. To verify whether a fixed point is a Nash equilibrium, one can compute the BR at that point.

Remark 9: In the mixed strategy setting ($m_i > 1$), a fixed point at temperature $T > 0$ is a quantal-response (logit) equilibrium. For a fixed point with at least two non-equivalent actions per player, as the temperature is annealed to zero, the corresponding equilibrium path converges to a Nash equilibrium [44], [45].

When a player updates $z^{(i)}$ to improve its loss, the change influences not only its own loss but also the losses of other players. A single block update affects the entire game. As a result, convergence should be studied collectively through the full Gauss-Seidel one-sweep map, viewed as a discrete dynamical system on $\mathcal{M}$ [46]. The relevant local object is the differential $DT(z^*) : Tz^*\mathcal{M} \rightarrow Tz^*\mathcal{M}$ at a fixed point z^* . Since fixed points of $\mathcal{T}$ need not be isolated and may form a C^1 embedded submanifold of $\mathcal{M}$, local convergence requires conditions that permit neutral directions tangent to the fixed-point set and contraction in the transverse directions.

Theorem 7 (Local convergence of USMEA near a non-isolated fixed-point): Let $\mathcal{M}^{(\text{FP})} \subset \mathcal{M}$ be a C^1 embedded submanifold of fixed points of $\mathcal{T}$, that is, $\mathcal{T}(z) = z, \forall z \in \mathcal{M}^{(\text{FP})}$. Fix $z^* \in \mathcal{M}^{(\text{FP})}$. Assume that:

- 1) $\dim \mathcal{M}^{(\text{FP})} = l$.
- 2) The eigenvalue 1 of $DT(z^*)$ has geometric multiplicity l , and the eigenvalue 1 is semisimple.
- 3) Every other eigenvalue λ of $DT(z^*)$ satisfies $|\lambda| < 1$.

Then there exists a neighborhood $\mathcal{V} \subset \mathcal{M}$ of z^* such that, for every $z^{[0]} \in \mathcal{V}$, the iterates $z^{[k+1]} = \mathcal{T}(z^{[k]})$ remain in $\mathcal{V}$ and converge to a limit $z^\infty \in \mathcal{M}^{(\text{FP})}$. In particular, the USMEA dynamics converge locally to a fixed point in the embedded fixed-point submanifold $\mathcal{M}^{(\text{FP})}$.

Remark 10 (Isolated fixed point): When the fixed point z^* is isolated, the fixed-point submanifold $\mathcal{M}^{(\text{FP})}$ is locally zero-dimensional, that is, $l = 0$. In this case, there are no tangent neutral directions, and the eigenvalue 1 does not appear in $DT(z^*)$. Therefore, the assumptions of Theorem 7 reduce to the standard local asymptotic stability condition that every

eigenvalue λ of $DT(z^*)$ satisfies $|\lambda| < 1$. Equivalently, the spectral radius of $DT(z^*)$ is strictly smaller than 1.

To compute the differential of the one-sweep map at a fixed point z^* , we use the chain rule on smooth manifolds, as

$$DT(z^*) = DT_N(z^*) \cdots DT_1(z^*). \quad (34)$$

Similarly, for each block update $\mathcal{T}_i$, we have $DT_i(z^*) = DT_i^p(z^*)DT_i^u(z^*)$. Therefore, the computation of $DT(z^*)$ reduces to computing $DT_i^u(z^*)$ and $DT_i^p(z^*)$ for all players, forming each block differential $DT_i(z^*)$, and multiplying them in the Gauss-Seidel order. The eigenvalues used in the local convergence analysis are then obtained from the resulting linear map $DT(z^*)$. The detailed calculations of $DT_i^u(z^*)$ and $DT_i^p(z^*)$ are provided in Section S-III of the Supplementary Material.

Remark 11: For any permutation κ of $\{1, 2, \dots, N\}$, define $\mathcal{T}_\kappa = \mathcal{T}_{\kappa(N)} \circ \dots \circ \mathcal{T}_{\kappa(1)}$. All theorems and lemmas in this paper apply to every $\mathcal{T}_\kappa$. None of our arguments depend on the specific update order. However, in general, since $DT_\kappa(z^*) \neq DT(z^*)$, so the corresponding eigenvalues may also differ. Therefore, the local spectral conditions in Theorem 7 must be verified for the chosen update order.

VI. EXPERIMENTAL ANALYSIS

In this section, we test the performance of our proposed algorithm in finding BRs and fixed points. First, to study the stability of the Riemannian component of the algorithm in the quantum Prisoner's Dilemma, we consider the case where each player uses a single unitary operator, that is, $m_i = 1$ for all i . In this case, there is no softmax update, and only the unitary (Riemannian) step is applied. Next, we analyze the algorithm in the quantum Prisoner's Dilemma with mixed strategies. Finally, we examine its performance in larger strategy spaces as well as in a three-player Prisoner's Dilemma.

In the plots, we use $(\text{BR}^{(i)}; z^*)$ to show the BR to the fixed point z^* by player i . To evaluate the convergence behavior of the learning algorithm, we use the relative error defined as

$$\varepsilon_{\text{rel}} = \frac{|\bar{r}^{(i)}(z^{[k]}) - \bar{r}^{(i)}(z^*)|}{\bar{r}^{(i)}(z^*)}. \quad (35)$$

We use $\bar{\varepsilon}_{\text{rel}}$ to denote the average relative error over n configurations initialized with different random seeds. The annealing temperature is adjusted according to $T = T_0 \alpha^k$, where T_0 is the initial temperature.

Since each unitary action is not unique, we only present the corresponding payoffs and BRs. To determine whether a fixed point is a Nash equilibrium, we compute the BR for all players at that fixed point. If any player can adopt an alternative strategy that increases their payoff, the fixed point cannot be classified as a Nash equilibrium.

In these experiments, we consider four different quantum games as follows

- 1) **Game 1:** Two-player quantum Prisoner's Dilemma.
- 2) **Game 2:** Two-player quantum game with asymmetric strategy spaces ($d_1 = 2$ and $d_2 = 3$).
- 3) **Game 3:** Two-player quantum game with $d_1 = d_2 = 3$.

Profile	Payoffs	Profile	Payoffs
(C,C)	(3,3)	(D,C)	(5,0)
(C,D)	(0,5)	(D,D)	(1,1)

(a) Two-player Prisoner's Dilemma.

Profile	Payoffs	Profile	Payoffs
(C,C,C)	(3,3,3)	(D,C,C)	(5,2,2)
(C,C,D)	(2,2,5)	(D,C,D)	(4,0,4)
(C,D,C)	(2,5,2)	(D,D,C)	(4,4,0)
(C,D,D)	(0,4,4)	(D,D,D)	(1,1,1)

(b) Three-player Prisoner's Dilemma.

TABLE I: Payoff table for classical Prisoner's Dilemma [47].

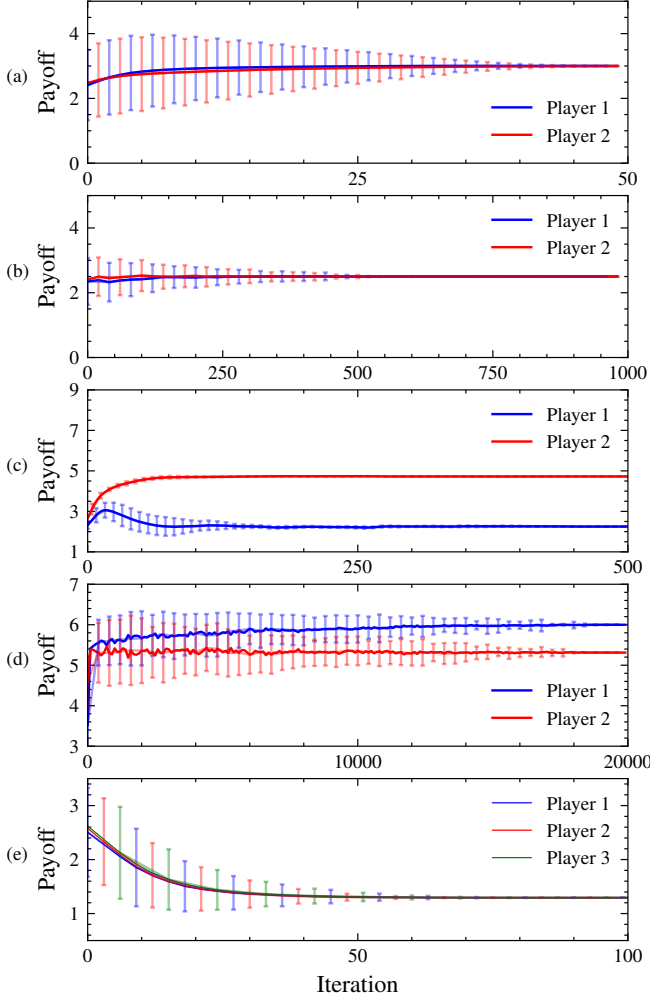


Fig. 1: Convergence of the expected payoffs with $n = 200$ and $\eta = 0.05$. (a) Game 1: Prisoner's Dilemma with pure strategies and $\gamma = \pi/2$. (b) Game 1: Prisoner's Dilemma with mixed strategies, with $T_0 = 10$, $\alpha = 0.99995$, and $\gamma = \pi/2$. (c) Game 2, with $T_0 = 1$, $\alpha = 0.9995$, and $\gamma = \pi/2$. (d) Game 3, with $T_0 = 1$, $\alpha = 0.99995$, and $\gamma = \pi/2$. (e) Game 4, with $\gamma = \pi/8$.

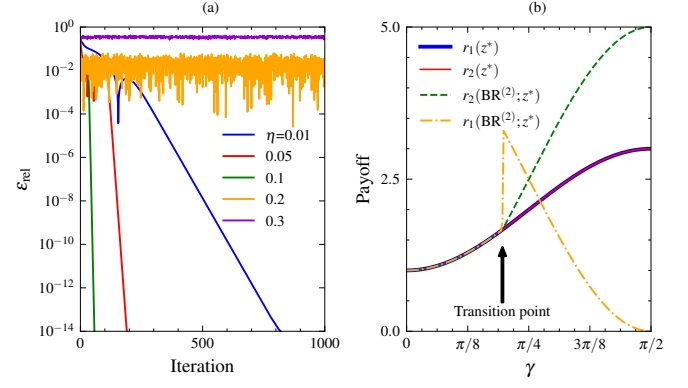


Fig. 2: Quantum Prisoner's Dilemma with pure quantum strategy. (a) Average relative payoff error with $n = 100$. (b) Payoff at the fixed point for different γ . The transition occurs is $\gamma_{TP} = 0.616 \pm 0.001$. For $\gamma < \gamma_{TP}$, a pure-strategy Nash equilibrium set exists, while for $\gamma > \gamma_{TP}$ no pure-strategy Nash equilibrium exists. Beyond the transition, players can find BRs that counter their opponents' strategies.

4) Game 4: Three-player quantum Prisoner's Dilemma.

For each quantum game, the details of the initial states, outcome sets, and the corresponding outcome payoffs are provided in Supplementary Material, Section S-IV. We use $\gamma \in [0, \pi/2]$ as the entanglement parameter. When $\gamma = 0$, the state is unentangled (separable) and the outcomes reduce to the classical game. When $\gamma = \pi/2$, the state is maximally entangled [8]. In this scenario, separable initial states regain classical strategic behavior, while entangled initial states introduce quantum correlations that change the equilibrium structure.

The payoff tables for the classical two-player and three-player Prisoner's Dilemma are shown in Table I. In the classical Prisoner's Dilemma, the Nash equilibrium occurs when all players choose to defect, which gives them a payoff of 1.

In Fig. 1, the trajectories of the expected payoffs for Games 1-4 are shown as they converge to their respective fixed points. As the strategy space expands and the number of players increases, the algorithm converges more slowly.

Nash equilibrium and BRs in the quantum Prisoner's Dilemma:

Fig. 2 shows the convergence of USMEA for the quantum Prisoner's Dilemma with pure strategies. Due to the symmetry between the two players in this game, the payoff remains identical for both of them throughout the learning process. In Fig. 2-(a), the convergence behavior of the payoff for different learning rates is examined. When the learning rate is very small, the convergence is slow and requires many iterations to stabilize. As we increase the learning rate, convergence becomes faster. However, if the learning rate becomes too large, the algorithm fails to converge to the optimal solution. Instead, it may overshoot or oscillate around potential solutions, failing to reach the optimal payoff.

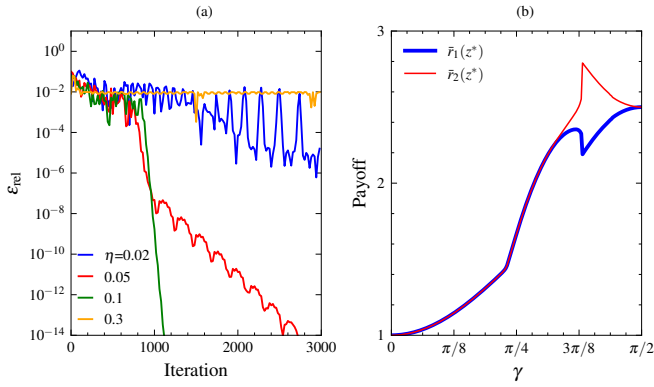


Fig. 3: Quantum Prisoner's Dilemma with mixed strategies ($m_i = 2$), $T_0 = 10$, and $\alpha = 0.99995$. (a) Average relative payoff error with $n = 100$. (b) Expected payoff at the fixed point for different γ .

In Fig. 2-(b), the payoffs at the fixed point and the corresponding best-response payoffs are shown as functions of the entanglement parameter γ . When γ is small, the game owns a pure-strategy Nash equilibrium set. However, as γ increases, the players' strategies become more strongly coupled through nonclassical correlations. We observe a transition point at $\gamma_{\text{TP}} = 0.616 \pm 0.001$, beyond which the pure-strategy Nash equilibrium disappears. Beyond the transition point, at least one player can find a counter-strategy that increases its own payoff and lowers the opponent's payoff. The dashed curves show the payoffs when Player 2 uses the best response to the fixed point. Beyond the transition point, Player 2 increase its own payoff and reduce Player 1's payoff. Hence, the fixed point is no longer a mutual best response. Physically, γ_{TP} is the entanglement threshold at which the game enters an entanglement-dominated strategic regime. Below this threshold, pure-strategy equilibria survive, whereas above it the nonclassical correlations are strong enough to destroy the pure-strategy equilibrium. We emphasize that this is a phase-transition-like change in equilibrium structure, not a thermodynamic phase transition [48], [49]. At the maximum entanglement level, $\gamma = \pi/2$, the optimal strategy leads to a cooperative payoff of (3,3), even though the BR function allows one player to obtain a payoff of 5 and decrease the other player's payoff. Our experiments further reveal beyond the transition point, the total payoff remains constant at $r_1 + r_2 = 5$. These behaviors agree with Ref. [47] and related studies. For more than 95% of random action initializations, the algorithm converged to these fixed points in this model.

The convergence behavior of the algorithm for the two-player quantum Prisoner's Dilemma with mixed strategies is shown in Fig. 3. For mixed strategies and $\gamma = \pi/2$, the expected payoff converges to 2.5 for each player. This point is a Nash equilibrium and yields a higher expected payoff than the classical Prisoner's Dilemma Nash equilibrium, which is 1. Fig. 3-(a) shows the convergence of the expected payoff for different learning rates η . Fig. 3-(b) shows the expected

m_i	γ	l	semisimple at $\lambda = 1$	$r_{\perp}$
1	0	2	yes	0.9500
1	$\pi/8$	1	yes	0.9709
1	$\pi/2$	3	yes	0.9350
2	0	4	yes	0.9750
2	$\pi/8$	1	yes	0.9927
2	$\pi/2$	4	yes	0.9875

TABLE II: Local spectral conditions for the two-player EEWL Prisoner's Dilemma at $T = 1$ and $\eta = 0.05$, evaluated at selected numerically computed fixed points. Here, $r_{\perp} = \max_{\lambda \neq 1} |\lambda|$ is the largest modulus among the non-neutral eigenvalues.

payoff at the fixed point for different values of γ . We start from $\gamma = \pi/2$ and let the algorithm converge to a fixed point. Since changing γ continuously shifts the fixed point, we decrease γ gradually and use the previous fixed point as the initial condition for the next run. Because this convergence point is attracting, the algorithm can easily find the new fixed point with this approach. This method helps us track a specific fixed point when the system has multiple fixed points, especially when there is symmetry between different players.

To assess whether the observed local behavior is consistent with Theorem 7, we next examine the spectrum of $DT(z^*)$ at a few representative fixed points computed numerically. Table II reports the corresponding local spectral quantities.

To further examine the convergence of the proposed update rule, we compare USMEA with the RGD+Softmax under the same initialization. Fig. 4 shows payoff trajectories of player 1 in game 1 for different temperatures and step sizes. In Fig. 4-(a), both methods converge for sufficiently small step sizes. For larger step sizes in the tested range, however, USMEA remains stable and converges faster, whereas RGD+Softmax exhibits stronger oscillations and may fail to converge (see Fig. 4-(b) and Fig. 4-(c)). Across many random initializations (not shown here), USMEA converged in a larger fraction of runs than RGD+Softmax. Both methods can lose convergence when the step size is chosen outside their stable regime. Fig. 4-(b) and Fig. 4-(d), which use the same step size but different temperatures, shows that the lower temperature makes the oscillatory behavior of RGD+Softmax more pronounced. Overall, these experiments suggest that USMEA has better empirical robustness to initialization, step size, and temperature in the tested mixed-strategy setting.

Complexity and Convergence Discussion:

The general form of an $n \times n$ unitary matrix $\mathbf{U} \in \mathcal{U}(n)$ is given by

$$\mathbf{U} = \begin{pmatrix} U_{11} & U_{12} & \cdots & U_{1n} \\ U_{21} & U_{22} & \cdots & U_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ U_{n1} & U_{n2} & \cdots & U_{nn} \end{pmatrix}, \quad (36)$$

where $U_{ij} \in \mathbb{C}$. The unitary group $\mathcal{U}(n)$ has n^2 degrees of freedom. However, ignoring the global phase invariance,

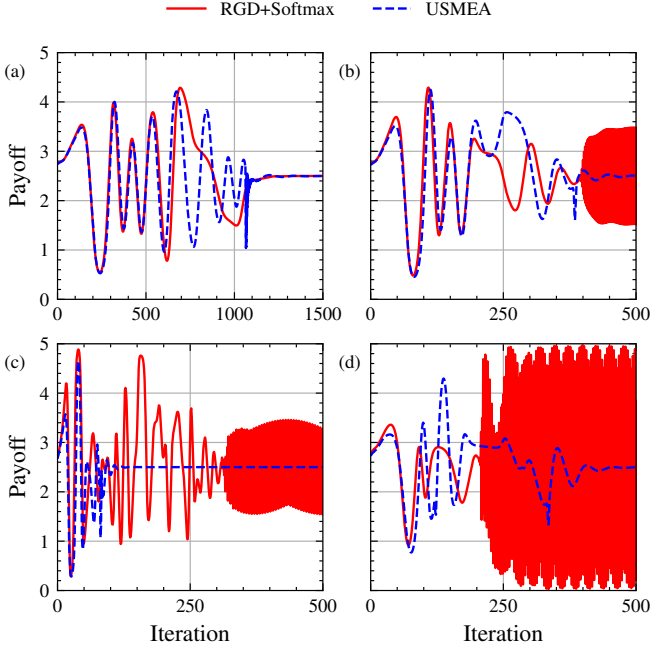


Fig. 4: Comparison between USMEA and the RGD+Softmax baseline in the mixed-strategy Quantum Prisoner's Dilemma ($m_i = 2$), using the same initialization and showing the payoff trajectory of player 1 in Game 1. (a) $T = 1$ and $\eta = 0.01$ (b) $T = 1$ and $\eta = 0.03$ (c) $T = 1$ and $\eta = 0.1$ (d) $T = 0.5$ and $\eta = 0.03$

there are effectively $n^2 - 1$ independent complex parameters. In the EEWL quantum game, the dimension of the overall Hilbert space grows exponentially with the number of players and corresponding strategy space, scaling as $O(\prod_{i \in \mathcal{N}} d_i)$. However, due to the product (or structured) nature of the joint strategy operator $\mathbf{U}^{(s)}$, the number of parameters that actually need to be optimized remains much smaller. The number of trainable decision variables scales as $O(\sum_{i \in \mathcal{N}} m_i(d_i^2 - 1))$, which reflects the local parameterization of the EEWL strategy space. For a dense-matrix implementation, joint-space payoff and gradient evaluation becomes the dominant computational cost as the problem size grows. In particular, the most expensive operations are the repeated construction of joint operators, matrix multiplications, and full/partial trace computations. As a result, the dominant per-sweep cost scales as $O(N \prod_{i \in \mathcal{N}} m_i d_i^3)$. When calculating the BR, as $T \rightarrow 0$ and under the conditions of Theorem 4, the updates become monotonic. In this regime, convergence is guaranteed. In contrast, when computing a fixed point z^* , Theorem 7 provides a local convergence guarantee under spectral conditions on the differential $D\mathcal{T}(z^*)$ at that fixed point. These assumptions are properties of the fixed point itself, not of the initialization. Accordingly, the theorem ensures local convergence only for iterates started sufficiently close to a fixed point satisfying these conditions.

Implementation on quantum hardware and simulators:

So far, we have used a general form of the EEWL quantum game with arbitrary local dimensions d_i and applied tools from quantum mechanics to extend classical game theory in a fully mathematical way. For the implementation of the quantum games on quantum hardware and simulators, we specialize to $d_i = 2^{n_i}$, where n_i is the number of qubits controlled by player i , and the local Hilbert space is $\mathcal{H}^{(i)} \cong (\mathbb{C}^2)^{\otimes n_i}$. Each local unitary $\mathbf{U}_{j_i}^{(i)}$ can be implemented as a parameterized quantum circuit built from qubit gates with parameter vectors $\theta_{j_i}^{(i)}$. In this hybrid setting, the optimization algorithm still runs on a classical computer, while all required quantum probabilities are estimated on a quantum device. Since changing $\theta_{j_i}^{(i)}$ preserves unitarity by construction, USMEA updates only the parameters $\theta_{j_i}^{(i)}$ in the first step. In the second step, it updates the classical probabilities using per-action payoffs estimated from measurement statistics returned by the quantum hardware. The hardware cost in this setting is governed by three factors. (1) the total number of qubits $\sum_i n_i$, (2) the circuit depth required to approximate each $\mathbf{U}_{j_i}^{(i)}$ from its parameters $\theta_{j_i}^{(i)}$, and (3) the number of measurement shots needed to estimate the payoffs with sufficient accuracy. This hybrid classical-quantum computing scheme shows that the proposed framework can, in principle, be scaled from purely classical simulations of small games to experimental implementations on near-term quantum devices, where noise and hardware constraints may affect the learned strategies.

Reproducibility Statement: All experiments in this paper are designed to be fully reproducible. Comprehensive details of the experimental setup, including implementation steps, algorithmic procedures, and parameter choices, are provided in the main text and in the supplementary material. ChatGPT and QuillBot were used to polish the language of this paper.

VII. CONCLUSION

In this paper, we presented the EEWL framework for multiplayer quantum games with mixed quantum strategies. We proposed USMEA as a learning algorithm for both unitary action sets and classical mixing probabilities, using an entropy-based loss function. We believe that USMEA and the EEWL formulation provide a structured basis for further algorithmic development and stronger theoretical frameworks in quantum game dynamics. We studied the existence of solutions in the EEWL setting and analyzed the convergence of USMEA under standard conditions for non-convex manifolds. In the experimental results, we examined the convergence of our algorithm in four different quantum models. While USMEA is effective as a learning procedure, it does not enumerate all Nash equilibria. For future work, we plan to replace the softmax function with alternative models for probability assignment. An important direction is to hybridize USMEA with methods such as homotopy or path-following to refine candidate fixed points. One could also employ adaptive or stochastic Riemannian gradient methods to train the actions. Beyond Nash equilibria, one can study correlated equilibria and regret in the EEWL setting. Finally, it remains important

to understand how noise and hardware limitations in quantum devices affect the learned strategies.

REFERENCES

- [1] W. Edwards, "The theory of decision making," *Psychological bulletin*, vol. 51, p. 380, 1954.
- [2] C. F. Camerer, "Advances in behavioral economics," *Russel Sage Foundation*, 2004.
- [3] J. Von Neumann and O. Morgenstern, "Theory of games and economic behavior: 60th anniversary commemorative edition," in *Theory of games and economic behavior*. Princeton university press, 2007.
- [4] D. Fudenberg and J. Tirole, *Game theory*. MIT press, 1991.
- [5] A. W. Tucker, "A two-person dilemma," *Prisoner's Dilemma*, 1950.
- [6] W. Poundstone, *Prisoner's dilemma*. Anchor, 2011.
- [7] A. Rapoport, "Prisoner's dilemma," in *The New Palgrave Dictionary of Economics*. Springer, 2018, pp. 10749–10753.
- [8] J. Eisert, M. Wilkens, and M. Lewenstein, "Quantum games and quantum strategies," *Physical Review Letters*, vol. 83, p. 3077, 1999.
- [9] J. Eisert and M. Wilkens, "Quantum games," *Journal of Modern Optics*, vol. 47, pp. 2543–2556, 2000.
- [10] S. C. Benjamin and P. M. Hayden, "Multiplayer quantum games," *Physical Review A*, vol. 64, p. 030301, 2001.
- [11] A. P. Flitney and D. Abbott, "Quantum version of the monty hall problem," *Physical Review A*, vol. 65, p. 062318, 2002.
- [12] C. Schmid, A. P. Flitney, W. Wiecek, N. Kiesel, H. Weinfurter, and L. C. Hollenberg, "Experimental implementation of a four-player quantum game," *New Journal of Physics*, vol. 12, p. 063031, 2010.
- [13] A. P. Flitney and D. Abbott, "Quantum games with decoherence," *Journal of Physics A: Mathematical and General*, vol. 38, p. 449, 2004.
- [14] K. Xu and Z. Wu, "Experimental implementation of quantum prisoner dilemma on ibm quantum computers," in *2022 15th International Conference on Advanced Computer Theory and Engineering (ICACTE)*. IEEE, 2022, pp. 13–18.
- [15] T. Holtfort, A. Horsch *et al.*, "Quantum economics: A systematic literature review," *SocioEconomic Challenges*, vol. 8, pp. 62–77, 2024.
- [16] A. Li and X. Yong, "Entanglement guarantees emergence of cooperation in quantum prisoner's dilemma games on networks," *Scientific reports*, vol. 4, p. 6286, 2014.
- [17] M. Ullah, G. R. Lakhan, A. Channa, and S. Gul, "Game theory and stock investment," *Multicultural Education*, vol. 7, pp. 40–44, 2021.
- [18] M. Hanauske, J. Kunz, S. Bernius, and W. König, "Doves and hawks in economics revisited: An evolutionary quantum game theory based analysis of financial crises," *Physica A: Statistical Mechanics and its Applications*, vol. 389, pp. 5084–5102, 2010.
- [19] A. Khrennikov, "Quantum-like modeling: cognition, decision making, and rationality," *Mind & Society*, vol. 19, pp. 307–310, 2020.
- [20] F. S. Khan and N. Bao, "Quantum prisoner's dilemma and high frequency trading on the quantum cloud," *Frontiers in Artificial Intelligence*, vol. 4, p. 769392, 2021.
- [21] K. Ikeda and S. Aoki, "Theory of quantum games and quantum economic behavior," *Quantum Information Processing*, vol. 21, p. 27, 2022.
- [22] R. P. Antón, J. I. L. Sánchez, and A. C. Bellot, "The game theory in quantum computers: A review," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, pp. 6–14, 2024.
- [23] J. Bostanci and J. Watrous, "Quantum game theory and the complexity of approximating quantum nash equilibria," *Quantum*, vol. 6, p. 882, 2022.
- [24] J. Preskill, "Quantum computing in the nisc era and beyond," *Quantum*, vol. 2, p. 79, 2018.
- [25] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE transactions on signal processing*, vol. 50, pp. 635–650, 2002.
- [26] I. A. Luchnikov, M. E. Krechetov, and S. N. Filippov, "Riemannian geometry and automatic differentiation for optimization problems of quantum physics and quantum technologies," *New Journal of Physics*, vol. 23, p. 073006, 2021.
- [27] T. Abrudan, J. Eriksson, and V. Koivunen, "Optimization under unitary matrix constraint using approximate matrix exponential," in *Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers*, 2005. IEEE, 2005, pp. 242–246.
- [28] A. P. Flitney and D. Abbott, "An introduction to quantum game theory," *Fluctuation and Noise Letters*, vol. 2, pp. R175–R187, 2002.
- [29] C. Song, Y. Liu, D. Dong, and H. Yonezawa, "Fast state stabilization using deep reinforcement learning for measurement-based quantum feedback control," *IEEE Transactions on Quantum Engineering*, 2025.
- [30] L. Le and T. N. Nguyen, "Dqra: Deep quantum routing agent for entanglement routing in quantum networks," *IEEE Transactions on Quantum Engineering*, vol. 3, pp. 1–12, 2022.
- [31] A. Abane, M. Cubeddu, V. S. Mai, and A. Battou, "Entanglement routing in quantum networks: A comprehensive survey," *IEEE Transactions on Quantum Engineering*, 2025.
- [32] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*. Cambridge university press, 2010.
- [33] J. Watrous, *The theory of quantum information*. Cambridge university press, 2018.
- [34] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.
- [35] R. Bhatia, *Matrix analysis*. Springer Science & Business Media, 2013, vol. 169.
- [36] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.
- [37] R. Wiersema and N. Killoran, "Optimizing quantum circuits with riemannian gradient flow," *Physical Review A*, vol. 107, p. 062421, 2023.
- [38] J. F. Nash Jr, "Equilibrium points in n-person games," *Proceedings of the national academy of sciences*, vol. 36, pp. 48–49, 1950.
- [39] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of optimization theory and applications*, vol. 109, pp. 475–494, 2001.
- [40] L. Grippo and M. Sciandrone, "On the convergence of the block nonlinear gauss–seidel method under convex constraints," *Operations research letters*, vol. 26, pp. 127–136, 2000.
- [41] P. Neff, Y. Nakatsukasa, and A. Fischle, "A logarithmic minimization property of the unitary polar factor in the spectral and frobenius norms," *SIAM Journal on Matrix Analysis and Applications*, vol. 35, pp. 1132–1154, 2014.
- [42] P.-A. Absil and K. Kurdyka, "On the stable equilibrium points of gradient systems," *Systems & Control Letters*, vol. 55, pp. 573–577, 2006.
- [43] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht, "Gradient descent converges to minimizers," in *Conference on Learning Theory*, 2016, pp. 1246–1257.
- [44] R. D. McKelvey and T. R. Palfrey, "Quantal response equilibria for normal form games," *Games and economic behavior*, vol. 10, pp. 6–38, 1995.
- [45] A. Balter, J. M. Schumacher, and N. Schweizer, "Decision under ambiguity, composed optimization, and quantal response equilibria," *Nikolaus, Decision under ambiguity, composed optimization, and quantal response equilibria (August 03, 2024)*, 2024.
- [46] K. Nipp and D. Stoffer, "Attractive invariant manifolds for maps: Existence, smoothness and continuous dependence on the map," *Seminar für Angewandte Mathematik, ETH Zürich, Tech. Rep. Research Report No. 92-11*, 1992. [Online]. Available: https://www.sam.math.ethz.ch/sam_reports/reports_final/reports1992/1992-11.pdf
- [47] J. Du, X. Xu, H. Li, X. Zhou, and R. Han, "Playing prisoner's dilemma with quantum rules," *Fluctuation and Noise Letters*, vol. 2, pp. R189–R203, 2002.
- [48] J. Du, H. Li, X. Xu, X. Zhou, and R. Han, "Phase-transition-like behaviour of quantum games," *Journal of Physics A: Mathematical and General*, vol. 36, pp. 6551–6562, 2003.
- [49] S. Bugu, "Entanglement as a strategic resource in adversarial quantum games," *arXiv preprint arXiv:2510.22444*, 2025.
- [50] K. B. Petersen, M. S. Pedersen *et al.*, "The matrix cookbook," *Technical University of Denmark*, vol. 7, p. 510, 2008.
- [51] S. Haykin, "Adaptive filter theory," *Prentice Hall google schola*, vol. 2, pp. 333–346, 2002.
- [52] B. Gao and L. Pavel, "On the properties of the softmax function with application in game theory and reinforcement learning," *arXiv preprint arXiv:1704.00805*, 2017.
- [53] R. Y. Rubinstein, "Optimization of computer simulation models with rare events," *European Journal of Operational Research*, vol. 99, pp. 89–112, 1997.
- [54] I. L. Glicksberg, "A further generalization of the kakutani fixed point theorem, with application to nash equilibrium points," *Proceedings of the American Mathematical Society*, vol. 3, pp. 170–174, 1952.

- [55] E. Malvetti, C. Arenz, G. Dirr, and T. Schulte-Herbrüggen, “Randomized gradient descents on riemannian manifolds: Almost sure convergence to global minima in and beyond quantum optimization,” *arXiv preprint arXiv:2405.12039*, 2024.
- [56] H. Attouch, J. Bolte, and B. F. Svaiter, “Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized gauss–seidel methods,” *Mathematical programming*, vol. 137, pp. 91–129, 2013.
- [57] J. Eldering, *Normally Hyperbolic Invariant Manifolds: The Noncompact Case*, ser. Atlantis Studies in Dynamical Systems. Atlantis Press, 2018, vol. 2. [Online]. Available: <https://www.jaapeldering.nl/files/NHIM-noncompact-book.pdf>
- [58] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge University Press, 2013.
- [59] W. Rudin, *Functional Analysis*, 2nd ed., ser. International Series in Pure and Applied Mathematics. McGraw-Hill, 1991.

Supplementary Materials:

Riemannian Optimization for Multi-Player Quantum Games on Product Unitary Manifolds

S-I. BASICS OF QUANTUM MECHANICS FOR GAME THEORISTS

To describe the quantum state of a system, we use a density matrix $\rho : \mathcal{H} \rightarrow \mathcal{H}$. The density matrix allows for the computation of outcome probabilities in quantum measurements. A valid density matrix is a Hermitian, positive semidefinite, and satisfies

$$\rho^\dagger = \rho, \quad \text{Tr}(\rho) = 1, \quad \frac{1}{d} \leq \text{Tr}(\rho^2) \leq 1, \quad (\text{S1})$$

The value $\text{Tr}(\rho^2)$ provides insight about the purity of the state. For a pure quantum state, $\text{Tr}(\rho^2) = 1$, while for a mixed quantum state, $\text{Tr}(\rho^2) < 1$. In particular, for the maximally mixed state, one has $\text{Tr}(\rho^2) = \frac{1}{d}$. In a finite-dimensional Hilbert space, an operator $\mathbf{A} : \mathcal{H} \rightarrow \mathcal{H}$ is a linear map and can be represented as a $d \times d$ complex matrix. In quantum mechanics, operators represent either physical observables or transformations of the quantum state. Unitary operators ($\mathbf{U}\mathbf{U}^\dagger = \mathbf{I}$, where $\mathbf{I}$ denotes the identity operator) can be used for the evolution of quantum states as $\rho' = \mathbf{U}\rho\mathbf{U}^\dagger$. Hermitian operators ($\mathbf{O} = \mathbf{O}^\dagger$) correspond to measurable quantities (observables) and are guaranteed to have real eigenvalues. The commutator of two operators is defined as $[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA}$. The expectation value of an observable $\mathbf{O}$ with respect to a quantum state ρ is computed as

$$\langle \mathbf{O} \rangle = \text{Tr}(\mathbf{O}\rho). \quad (\text{S2})$$

For a pure quantum state, we can also use the Dirac notation. In Dirac notation, a pure quantum state ψ is represented by a ket, $|\psi\rangle$, in a Hilbert space spanned by a set of orthonormal basis kets $\{|j\rangle\}$. Any state can be written as $|\psi\rangle = \sum_j c_j |j\rangle$, where the complex coefficients c_j are probability amplitudes. The quantity $|c_j|^2$ gives the probability of finding the system in basis state $|j\rangle$. Each ket $|\psi\rangle$ has a vector counterpart in the chosen basis, often written as a column vector of complex amplitudes $(c_1, c_2, \dots)^T$, and its dual bra $\langle\psi| = |\psi\rangle^\dagger$ corresponds to the conjugate transpose row vector. The inner product between two states is written as $\langle\psi'|\psi\rangle$. The outer product $|j'\rangle\langle j|$ forms a basis for the space of linear operators. For composite systems, the joint state space is constructed using the Kronecker product of the subsystems' bases. For example, for two subsystems A and B , the $|j_A j_B\rangle = |j_A\rangle \otimes |j_B\rangle$ describes the joint basis. For a mixed quantum state, the density matrix can be written in terms of pure states as

$$\rho = \sum_j \alpha_j |\psi_j\rangle\langle\psi_j|, \quad (\text{S3})$$

where α_j represents the classical probability of the system being in the pure state $|\psi_j\rangle$.

Separable and Entangled states: In quantum mechanics, when dealing with a composite system composed of two or more subsystems, the overall state can be either separable or entangled. Consider that we have a system made up of subsystems A and B with the density matrix $\rho^{(AB)}$. The system is separable if its state can be expressed as

$$\rho^{(AB)} = \sum_j p_j \rho_j^{(A)} \otimes \rho_j^{(B)}, \quad (\text{S4})$$

where p_j is classical probabilities (with $\sum_j p_j = 1$), and $\rho_j^{(A)}$ and $\rho_j^{(B)}$ are the density matrices of subsystems A and B , respectively. The symbol $\otimes$ denotes the Kronecker product with the properties

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}). \quad (\text{S5})$$

In a separable state, the total system can be viewed as a statistical mixture of independent states. If no such decomposition exists, the state $\rho^{(AB)}$ is entangled. Entangled states show a non-classical correlation between subsystems. This quantum correlation cannot be explained by classical probability theory or any local hidden variable model.

S-II. PROOFS AND ADDITIONAL RESULTS

In this section, we provide some additional theorems and missing proofs from the main part of our paper.

Theorem S-1 (Wirtinger's Calculus for Real-Valued Functions [50], [51]): Let z be a complex variable, and let $f : \mathbb{C} \rightarrow \mathbb{R}$ be a real-valued function. Then, the gradient of f with respect to z is given by

$$\nabla_z f(z, \bar{z}) = 2 \frac{\partial f(z, \bar{z})}{\partial \bar{z}}. \quad (\text{S6})$$

Proof: For a detailed derivation, see Refs. [50], [51]. In Eq. (S6), the factor of two arises from Wirtinger's calculus for real-valued functions with complex-valued parameters. This method treats the complex variable z and its conjugate $\bar{z}$ as independent variables.

Lemma S-1: The softmax function is L -Lipschitz with respect to the $\|\cdot\|_2$ norm, with Lipschitz constant $L = 1/T$ as

$$\|\sigma_T(\ell) - \sigma_T(\ell')\|_2 \leq \frac{1}{T} \|\ell - \ell'\|_2. \quad (\text{S7})$$

Proof: For a detailed proof, see [52].

Lemma S-2: For any operator $\mathbf{X}$ on $\mathcal{H}^{(i)} \otimes \mathcal{H}^{(-i)}$,

$$\|\text{Tr}_{-i}(\mathbf{X})\|_F \leq \sqrt{d_{-i}} \|\mathbf{X}\|_F. \quad (\text{S8})$$

Proof: The Hilbert–Schmidt inner product between two operators is given by $\langle \mathbf{A}, \mathbf{B} \rangle_{\text{HS}} = \text{Tr}(\mathbf{A}^\dagger \mathbf{B})$. For any operator $\mathbf{Y}^{(i)} : \mathcal{H}^{(i)} \rightarrow \mathcal{H}^{(i)}$, we have

$$\langle \text{Tr}_{-i}(\mathbf{X}), \mathbf{Y}^{(i)} \rangle_{\text{HS}} = \text{Tr}_i \left((\text{Tr}_{-i}(\mathbf{X}))^\dagger (\mathbf{Y}^{(i)} \otimes \mathbf{I}^{(-i)}) \right) = \langle \mathbf{X}, \mathbf{Y}^{(i)} \otimes \mathbf{I}^{(-i)} \rangle_{\text{HS}}. \quad (\text{S9})$$

By duality, we have

$$\begin{aligned} \|\text{Tr}_{-i}(\mathbf{X})\|_F &= \sup_{\|\mathbf{Y}^{(i)}\|_F=1} |\langle \text{Tr}_{-i}(\mathbf{X}), \mathbf{Y}^{(i)} \rangle_{\text{HS}}| \\ &= \sup_{\|\mathbf{Y}^{(i)}\|_F=1} |\langle \mathbf{X}, \mathbf{Y}^{(i)} \otimes \mathbf{I}^{(-i)} \rangle_{\text{HS}}| \\ &\leq \|\mathbf{X}\|_F \sup_{\|\mathbf{Y}^{(i)}\|_F=1} \|\mathbf{Y}^{(i)} \otimes \mathbf{I}^{(-i)}\|_F. \end{aligned} \quad (\text{S10})$$

Since Frobenius norm is multiplicative on tensor products [35], we have

$$\|\mathbf{Y}^{(i)} \otimes \mathbf{I}^{(-i)}\|_F = \|\mathbf{Y}^{(i)}\|_F \|\mathbf{I}^{(-i)}\|_F = \sqrt{d_{-i}}. \quad (\text{S11})$$

Combining Eq. (S10) and Eq. (S11), we obtain

$$\|\text{Tr}_{-i}(\mathbf{X})\|_F \leq \sqrt{d_{-i}} \|\mathbf{X}\|_F. \quad (\text{S12})$$

Proposition S-1: Consider that all actions are fixed, and player i updates its probabilities by minimizing its loss function $L^{(i)}$ at temperature $T > 0$. Let $\mathbf{p}'^{(i)}$ be the probability vector that minimizes $L^{(i)}$, given by the softmax function [53]. Then, for any other probability vector $\mathbf{p}^{(i)}$ in the simplex, the following improvement bound holds

$$L^{(i)}(\mathbf{p}'^{(i)}) - L^{(i)}(\mathbf{p}^{(i)}) \leq -\frac{T}{2} \left\| \mathbf{p}'^{(i)} - \mathbf{p}^{(i)} \right\|_2^2. \quad (\text{S13})$$

Proof: At a minimum, $L^{(i)}$ satisfies

$$L^{(i)}(\mathbf{p}'^{(i)}) = -T \log \sum_{j_i} \exp \left(\ell_{j_i}^{(i)} / T \right). \quad (\text{S14})$$

For any $\mathbf{p}^{(i)} \in \Delta^{m_i-1}$, we compute the loss function difference as

$$L^{(i)}(\mathbf{p}'^{(i)}) - L^{(i)}(\mathbf{p}^{(i)}) = -\log \sum_{j_i} \exp \left(\ell_{j_i}^{(i)} / T \right) + \sum_{j_i} p_{j_i}^{(i)} \ell_{j_i}^{(i)} - T \sum_{j_i} p_{j_i}^{(i)} \log p_{j_i}^{(i)}. \quad (\text{S15})$$

Using softmax equation, $\ell_{j_i}^{(i)}$ can be expressed as

$$\ell_{j_i}^{(i)} = T \left(\log \sum_{j_i} \exp \left(\ell_{j_i}^{(i)} / T \right) - \log p_{j_i}'^{(i)} \right), \quad (\text{S16})$$

we substitute the above equation into the loss function difference and obtain

$$L^{(i)}(\mathbf{p}'^{(i)}) - L^{(i)}(\mathbf{p}^{(i)}) = -T \text{KL} \left(\mathbf{p}^{(i)} \parallel \mathbf{p}'^{(i)} \right), \quad (\text{S17})$$

where $\text{KL}(\cdot \parallel \cdot)$ denotes the Kullback–Leibler (KL) divergence

$$\text{KL} \left(\mathbf{p}^{(i)} \parallel \mathbf{p}'^{(i)} \right) = \sum_{j_i} p_{j_i}^{(i)} \log \left(\frac{p_{j_i}^{(i)}}{p_{j_i}'^{(i)}} \right). \quad (\text{S18})$$

Applying Pinsker's inequality and using $\|\cdot\|_1 \geq \|\cdot\|_2$, we obtain

$$\text{KL} \left(\mathbf{p}^{(i)} \parallel \mathbf{p}'^{(i)} \right) \geq \frac{1}{2} \left\| \mathbf{p}^{(i)} - \mathbf{p}'^{(i)} \right\|_1^2 \geq \frac{1}{2} \left\| \mathbf{p}^{(i)} - \mathbf{p}'^{(i)} \right\|_2^2, \quad (\text{S19})$$

we conclude

$$L^{(i)}(\mathbf{p}'^{(i)}) - L^{(i)}(\mathbf{p}^{(i)}) \leq -\frac{T}{2} \left\| \mathbf{p}'^{(i)} - \mathbf{p}^{(i)} \right\|_2^2. \quad (\text{S20})$$

Proof of Remark 1:

(1) Set $\mathbf{U}'^{(i)} = e^{i\theta_i} \mathbf{U}^{(i)}$. Then

$$\mathbf{U}'^{(s)} \rho_0 \mathbf{U}'^{(s)\dagger} = e^{i \sum_i \theta_i} \mathbf{U}^{(s)} \rho_0 \mathbf{U}^{(s)\dagger} e^{-i \sum_i \theta_i} = \mathbf{U}^{(s)} \rho_0 \mathbf{U}^{(s)\dagger}, \quad (\text{S21})$$

so the payoff is unchanged.

(2) Let $\rho = \mathbf{U}^{(s)} \rho_0 \mathbf{U}^{(s)\dagger}$. ρ is positive definite and satisfies $\text{Tr}(\rho) = 1$. The map $\rho \mapsto \text{Tr}(\mathbf{R}_i \rho)$ is linear, and the set of density matrices forms the convex hull of rank-1 projectors. Because $\text{Tr}(\mathbf{R}_i \rho)$ is a convex combination of eigenvalues of $\mathbf{R}_i$ in an appropriate basis, we have

$$\lambda_{\min}(\mathbf{R}_i) \leq \text{Tr}(\mathbf{R}_i \rho) \leq \lambda_{\max}(\mathbf{R}_i). \quad (\text{S22})$$

For the operator-norm bound, we use the Schatten norm properties of operators and we obtain

$$|\text{Tr}(\mathbf{R}_i \rho)| \leq \|\mathbf{R}_i\|_{\text{op}} \|\rho\| = \|\mathbf{R}_i\|_{\text{op}} \text{Tr}(\rho) = \|\mathbf{R}_i\|_{\text{op}}. \quad (\text{S23})$$

Both bounds are tight. Equality happens when choosing ρ as the projector onto an eigenvector of $\mathbf{R}_i$ with eigenvalue $\lambda_{\max}$ or $\lambda_{\min}$. For mixed strategies, the averaged state

$$\bar{\rho} = \sum_j p_j \mathbf{U}_j^{(s)} \rho_0 \mathbf{U}_j^{(s)\dagger} \quad (\text{S24})$$

is again a density matrix (convex combination of density matrices). Replacing ρ by $\bar{\rho}$ yields the same bounds for $\bar{r}^{(i)}$.

Proof of Theorem 1:

The unitary group $\mathcal{U}(d_i)$ is compact (closed and bounded in finite dimension). The set of mixed strategies for player i is convex, closed, and compact. The expected payoff function is continuous, bounded, and linear in each player's own mixed strategy. These properties satisfy the hypotheses of Glicksberg's generalization of Nash's theorem (compact action sets, continuous payoffs, and linearity in one's own mixture) [54]. Therefore, there is at least one mixed strategy Nash equilibrium in the EEWL quantum game.

Proof of Theorem 2:

(1) In the first update of USMEA, $\mathbf{p}^{(i)}$ is fixed, so $H(\mathbf{p}^{(i)})$ remains constant and its derivative with respect to $\mathbf{U}_{j_i}^{(i)}$ vanishes. Therefore, when computing the gradient of $L^{(i)}$, only the term $-\bar{r}^{(i)}$ contributes. Since Riemannian gradients are linear, we obtain (i).

(2) The payoff function is real-valued but depends on complex-valued parameters. First, we calculate the Euclidean gradient of $r^{(i)}(\mathbf{U}^{(s)})$ with respect to $\mathbf{U}^{(i)}$. We use the Wirtinger derivatives (Theorem S-1) for real-valued functions with complex variables. We compute the gradient with respect to $\bar{U}_{m,n}^{(i)}$. We treat $U_{m,n}^{(i)}$ and $\bar{U}_{m,n}^{(i)}$ as independent variables. For convenience, we define $\mathbf{X} = \mathbf{R}_i \mathbf{U} \rho_0$. The derivative of the real-valued function $r^{(i)}$ with respect to $\bar{U}_{m,n}^{(i)}$ is then given by

$$\begin{aligned} \frac{\partial r^{(i)}(\mathbf{U}^{(s)})}{\partial \bar{U}_{m,n}^{(i)}} &= \frac{\partial}{\partial \bar{U}_{m,n}^{(i)}} \left(\sum_{\substack{j_1, j_2, \dots, j_N \\ j'_1, j'_2, \dots, j'_N}} X_{j_1, \dots, j_i, \dots, j_N, j'_1, \dots, j'_i, \dots, j'_N} \bar{U}_{j_1, j'_1}^{(1)} \bar{U}_{j_2, j'_2}^{(2)} \dots \bar{U}_{j_N, j'_N}^{(N)} \right) \\ &= \sum_{\substack{j_1, j_2, \dots, j_N \\ j'_1, j'_2, \dots, j'_N}} X_{j_1, \dots, j_i, \dots, j_N, j'_1, \dots, j'_i, \dots, j'_N} \frac{\partial}{\partial \bar{U}_{m,n}^{(i)}} \left(\bar{U}_{j_1, j'_1}^{(1)} \bar{U}_{j_2, j'_2}^{(2)} \dots \bar{U}_{j_N, j'_N}^{(N)} \right). \end{aligned} \quad (\text{S25})$$

By applying the chain rule, we have

$$\begin{aligned} \frac{\partial r^{(i)}(\mathbf{U}^{(s)})}{\partial \bar{U}_{m,n}^{(i)}} &= \sum_{\substack{j_1, j_2, \dots, j_N \\ j'_1, j'_2, \dots, j'_N}} R_{j_1, \dots, j_i, \dots, j_N, j'_1, \dots, j'_i, \dots, j'_N} \sum_{k=1}^N \left(\frac{\partial}{\partial \bar{U}_{m,n}^{(i)}} \bar{U}_{j_k, j'_k}^{(k)} \right) \prod_{l \neq k} \bar{U}_{j_l, j'_l}^{(l)} \\ &= \sum_{\substack{j_1, j_2, \dots, j_N \\ j'_1, j'_2, \dots, j'_N}} X_{j_1, \dots, j_i, \dots, j_N, j'_1, \dots, j'_i, \dots, j'_N} \sum_{k=1}^N \left(\delta_{m, j_k} \delta_{n, j'_k} \delta_{i, k} \right) \prod_{l \neq k} \bar{U}_{j_l, j'_l}^{(l)} \\ &= \sum_{\substack{\{j_1, \dots, j_{m-1}, j_{m+1}, \dots, j_N\} \\ \{j'_1, \dots, j'_{n-1}, j'_{n+1}, \dots, j'_N\}}} X_{j_1, \dots, m, \dots, j_N, j'_1, \dots, n, \dots, j'_N} \prod_{l \neq i} \bar{U}_{j_l, j'_l}^{(l)}. \end{aligned} \quad (\text{S26})$$

This expression simplifies to the (m, n) -th element of the partial trace and can be expressed as

$$\frac{\partial r^{(i)}(\mathbf{U}^{(s)})}{\partial \bar{U}_{mn}^{(i)}} = \left[\text{Tr}_{-i} \left(\mathbf{X} \mathbf{U}_{-i}^{(s)\dagger} \right) \right]_{mn}. \quad (\text{S27})$$

The derivative with respect to $\mathbf{U}^{(i)}$ is given by

$$\frac{\partial r^{(i)}(\mathbf{U}^{(s)})}{\partial \bar{\mathbf{U}}^{(i)}} = \text{Tr}_{-i} \left(\mathbf{X} \mathbf{U}_{-i}^{(s)\dagger} \right). \quad (\text{S28})$$

By substituting $\mathbf{X} = \mathbf{R}_i \mathbf{U} \rho_0$ into the equation and using Theorem S-1, we have

$$\nabla_{\mathbf{U}^{(i)}} r^{(i)} = 2 \frac{\partial r^{(i)}(\mathbf{U}^{(s)})}{\partial \bar{\mathbf{U}}^{(i)}} = 2 \text{Tr}_{-i} \left(\mathbf{R}_i \mathbf{U}^{(s)} \rho_0 \mathbf{U}_{-i}^{(s)\dagger} \right). \quad (\text{S29})$$

The Riemannian gradient of the payoff function with respect to the unitary operator $\mathbf{U}^{(i)}$ is given by (Ref. [27])

$$2\mathbf{g}^{(i)} = \left(\nabla_{\mathbf{U}^{(i)}} r^{(i)} \right) \mathbf{U}^{(i)\dagger} - \mathbf{U}^{(i)} \left(\nabla_{\mathbf{U}^{(i)}} r^{(i)} \right)^\dagger, \quad (\text{S30})$$

which is an antisymmetric matrix. By performing straightforward calculations, we obtain

$$\mathbf{g}^{(i)} = \text{Tr}_{-i} \left([\mathbf{R}_i, \mathbf{U}^{(s)} \rho_0 \mathbf{U}^{(s)\dagger}] \right). \quad (\text{S31})$$

A similar calculation for the mixed strategy leads to

$$\mathbf{G}_{j_i}^{(i)} = \sum_{\mathbf{j}_{-i}} p_{j_i; \mathbf{j}_{-i}} \text{Tr}_{-i} \left([\mathbf{R}_i, \mathbf{U}_{j_i; \mathbf{j}_{-i}}^{(s)} \rho_0 \mathbf{U}_{j_i; \mathbf{j}_{-i}}^{(s)\dagger}] \right). \quad (\text{S32})$$

Proof of Theorem 3:

Consider two unitary actions $\mathbf{U}_{j_i}^{(i)}$ and $\mathbf{U}_{j_i}'^{(i)}$ for player i with action j_i . These actions can be connected via a geodesic path: $\mathbf{U}^{(i)}(t) = \exp(t\mathbf{X}^{(i)})\mathbf{U}^{(i)}$, where $\mathbf{X}^{(i)} = -\mathbf{X}^{(i)\dagger}$ such that $\|\mathbf{X}^{(i)}\|_F = d_{\mathcal{U}}(\mathbf{U}^{(i)}, \mathbf{U}'^{(i)})$ and $t \in [0, 1]$. For each t , let $\mathbf{U}^{(s)}(t)$ the joint strategy profile and the evolved density matrix $\rho(t) = \mathbf{U}^{(s)}(t) \rho_0 \mathbf{U}^{(s)\dagger}(t)$. The Riemannian gradient of the payoff over $\mathbf{U}^{(i)}(t)$ (block i) at time t is

$$\mathbf{g}^{(i)}(t) = \text{Tr}_{-i}([\mathbf{R}_i, \rho(t)]). \quad (\text{S33})$$

The Riemannian gradient of expected reward over action j_i is then given by

$$\mathbf{G}_{j_i}^{(i)}(t) = \sum_{\mathbf{j}_{-i}} p_{j_i; \mathbf{j}_{-i}} \mathbf{g}_{j_i; \mathbf{j}_{-i}}^{(i)}(t). \quad (\text{S34})$$

(1) By differentiating $\mathbf{g}^{(i)}(t)$ along the geodesic, we obtain:

$$\begin{aligned} \frac{d}{dt} \mathbf{g}^{(i)}(t) &= \text{Tr}_{-i} \left([\mathbf{R}_i, \mathbf{X}^{(i)} \rho(t)] + [\mathbf{R}_i, \rho(t) \mathbf{X}^{(i)\dagger}] \right) \\ &= \text{Tr}_{-i} \left([\mathbf{R}_i, \mathbf{X}^{(i)} \rho(t)] - [\mathbf{R}_i, \rho(t) \mathbf{X}^{(i)}] \right) \\ &= \text{Tr}_{-i} \left([\mathbf{R}_i, [\mathbf{X}^{(i)}, \rho(t)]] \right). \end{aligned} \quad (\text{S35})$$

By Lemma S-2 and the bound $\|[\mathbf{A}, \mathbf{B}]\|_F \leq 2\|\mathbf{A}\|_{\text{op}}\|\mathbf{B}\|_F$, we get

$$\begin{aligned} \left\| \frac{d}{dt} \mathbf{g}^{(i)}(t) \right\|_F &\leq \sqrt{d_{-i}} \left\| [\mathbf{R}_i, [\mathbf{X}^{(i)}, \rho(t)]] \right\|_F \\ &\leq 2\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \left\| [\mathbf{X}^{(i)}, \rho(t)] \right\|_F \\ &\leq 4\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\mathbf{X}^{(i)}\|_{\text{op}} \|\rho(t)\|_F \\ &\leq 4\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\mathbf{X}^{(i)}\|_{\text{op}} \|\rho_0\|_F \\ &\leq 4\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\mathbf{X}^{(i)}\|_F \|\rho_0\|_F. \end{aligned} \quad (\text{S36})$$

Using Jensen's inequality for expectations, $\|\mathbb{E}(\mathbf{A})\|_F \leq \mathbb{E}(\|\mathbf{A}\|_F)$, and integrating over $t \in [0, 1]$, we obtain

$$\begin{aligned} \left\| \mathbf{G}_{j_i}^{(i)}(1) - \mathbf{G}_{j_i}^{(i)}(0) \right\|_F &\leq \int_0^1 \left\| \frac{d}{dt} \mathbf{G}_{j_i}^{(i)}(t) \right\|_F dt \\ &\leq 4\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\rho_0\|_F \|\mathbf{X}^{(i)}\|_F. \end{aligned} \quad (\text{S37})$$

Finally, recalling the distance defined in Eq. (21) as $d_{\mathcal{U}} = \|\mathbf{X}^{(i)}\|_F$, we conclude the proof of part (i) as

$$\left\| \mathbf{G}_{j_i}^{(i)}(\mathbf{U}_{j_i}^{(i)}) - \mathbf{G}_{j_i}^{(i)}(\mathbf{U}_{j_i}'^{(i)}) \right\|_F \leq A_i d_{\mathcal{U}} \left(\mathbf{U}_{j_i}^{(i)}, \mathbf{U}_{j_i}'^{(i)} \right), \quad (\text{S38})$$

where

$$A_i = 4\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\rho_0\|_F. \quad (\text{S39})$$

(2) The per-action payoff $\ell_{j_i}^{(i)}(t)$ can be expressed as

$$\begin{aligned} \ell_{j_i}^{(i)}(t) &= \sum_{\mathbf{j}_{-i}} p_{j_i, \mathbf{j}_{-i}} r^{(i)} \left(\mathbf{U}_{j_i, \mathbf{j}_{-i}}^{(s)}(t) \right) \\ &= \sum_{\mathbf{j}_{-i}} p_{j_i, \mathbf{j}_{-i}} \text{Tr} \left(\mathbf{R}_i \rho_{j_i, \mathbf{j}_{-i}}(t) \right). \end{aligned} \quad (\text{S40})$$

Differentiating $\ell_{j_i}^{(i)}(t)$ along the geodesic yields

$$\begin{aligned} \frac{d}{dt} \ell_{j_i}^{(i)}(t) &= \sum_{\mathbf{j}_{-i}} p_{j_i, \mathbf{j}_{-i}} \text{Tr} \left(\mathbf{R}_i \left[\mathbf{X}^{(i)}, \rho_{j_i, \mathbf{j}_{-i}}(t) \right] \right) \\ &= - \sum_{\mathbf{j}_{-i}} p_{j_i, \mathbf{j}_{-i}} \text{Tr} \left(\left[\mathbf{R}_i, \rho_{j_i, \mathbf{j}_{-i}}(t) \right] \mathbf{X}^{(i)} \right) \\ &= \sum_{\mathbf{j}_{-i}} p_{j_i, \mathbf{j}_{-i}} \text{Tr} \left(\mathbf{X}^{(i)\dagger} \left[\mathbf{R}_i, \rho_{j_i, \mathbf{j}_{-i}}(t) \right] \right) \\ &= \left\langle \mathbf{X}^{(i)}, \sum_{\mathbf{j}_{-i}} p_{j_i, \mathbf{j}_{-i}} \text{Tr}_{-i} \left(\left[\mathbf{R}_i, \rho_{j_i, \mathbf{j}_{-i}}(t) \right] \right) \right\rangle_{\text{HS}} \\ &= \left\langle \mathbf{X}^{(i)}, \mathbf{G}_{j_i}^{(i)}(t) \right\rangle_{\text{HS}}. \end{aligned} \quad (\text{S41})$$

Applying the Cauchy–Schwarz inequality gives

$$\left| \frac{d}{dt} \ell_{j_i}^{(i)}(t) \right| = \left| \left\langle \mathbf{X}^{(i)}, \mathbf{G}_{j_i}^{(i)}(t) \right\rangle_{\text{HS}} \right| \leq \|\mathbf{G}_{j_i}^{(i)}(t)\|_F \|\mathbf{X}^{(i)}\|_F. \quad (\text{S42})$$

To bound $\|\mathbf{G}_{j_i}^{(i)}(t)\|_F$, we first bound the pure-strategy case. For each $\mathbf{j}_{-i}$, Lemma S-2 and commutator norms imply

$$\begin{aligned} \left\| \text{Tr}_{-i} \left(\left[\mathbf{R}_i, \rho_{j_i, \mathbf{j}_{-i}}(t) \right] \right) \right\|_F &\leq \sqrt{d_{-i}} \left\| \left[\mathbf{R}_i, \rho_{j_i, \mathbf{j}_{-i}}(t) \right] \right\|_F \\ &\leq 2\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\rho_{j_i, \mathbf{j}_{-i}}(t)\|_F \\ &\leq 2\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\rho_0\|_F. \end{aligned} \quad (\text{S43})$$

Taking expectation over $\mathbf{j}_{-i}$ and using Jensen's inequality

$$\left\| \mathbf{G}_{j_i}^{(i)}(t) \right\|_F \leq 2\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\rho_0\|_F. \quad (\text{S44})$$

By integrating Eq. (S42) from $t = 0$ to $t = 1$, and applying the bound from Eq. (S44), we obtain

$$\begin{aligned} \left| \ell_{j_i}^{(i)}(1) - \ell_{j_i}^{(i)}(0) \right| &\leq \int_0^1 \left\| \mathbf{G}_{j_i}^{(i)}(t) \right\|_F \|\mathbf{X}^{(i)}\|_F dt \\ &\leq 2\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\rho_0\|_F \|\mathbf{X}^{(i)}\|_F. \end{aligned} \quad (\text{S45})$$

Finally, substituting $d_{\mathcal{U}} = \|\mathbf{X}^{(i)}\|_F$ and defining

$$M_i = 2\sqrt{d_{-i}} \|\mathbf{R}_i\|_{\text{op}} \|\rho_0\|_F, \quad (\text{S46})$$

we conclude Eq. (28).

Proof of Theorem 4

First, fix $\mathbf{p}^{(i)}$ and modify $\{\mathbf{U}_{j_i}^{(i)}\}$. Using [55, Lemma 3.7], we have

$$L^{(i)}\left(\left(\{\mathbf{U}_{j_i}^{(i)}\}, \mathbf{p}^{(i)}\right); z^{(-i)}\right) - L^{(i)}\left(\left(\{\mathbf{U}_{j_i}^{(i)}\}, \mathbf{p}^{(i)}\right); z^{(-i)}\right) \leq -\eta_i \left(1 - \frac{1}{2}\eta_i A_i\right) \Delta_{\mathcal{U}}^{(i)2}, \quad (\text{S47})$$

Next, update $\mathbf{p}^{(i)}$ while keeping the $\{\mathbf{U}_{j_i}^{(i)}\}$ fixed. Using Proposition S-1, we get

$$L^{(i)}\left(\left(\{\mathbf{U}_{j_i}^{(i)}\}, \mathbf{p}'^{(i)}\right); z^{(-i)}\right) - L^{(i)}\left(\left(\{\mathbf{U}_{j_i}^{(i)}\}, \mathbf{p}^{(i)}\right); z^{(-i)}\right) \leq -\frac{T}{2} \left\| \mathbf{p}'^{(i)} - \mathbf{p}^{(i)} \right\|_2^2. \quad (\text{S48})$$

Adding inequalities Eqs. (S47) and (S48) gives the bound in Eq. (30). Since both terms on the right-hand side are nonnegative for $\eta_i \leq 1/A_i$, monotonicity of $L^{(i)}$ follows.

Proof of Theorem 5:

Fix $z^{(-i)}$. The function $L^{(i)}(z^{(i)}; z^{(-i)})$ is bounded as

$$\begin{aligned} \left\| L^{(i)}\left(z^{(i)}; z^{(-i)}\right) \right\| &\leq \left\| \sum_{j_i} p_{j_i}^{(i)} \ell_{j_i}^{(i)} \right\| + \left\| T \sum_{j_i} p_{j_i}^{(i)} \log p_{j_i}^{(i)} \right\| \\ &\leq \sum_{j_i} p_{j_i}^{(i)} \|R_i\|_{\text{op}} + \left\| T \sum_{j_i} p_{j_i}^{(i)} \log p_{j_i}^{(i)} \right\| \\ &\leq \|R_i\|_{\text{op}} + T \log m_i. \end{aligned} \quad (\text{S49})$$

Moreover, by Theorem 4, if $\eta_i \leq 1/A_i$, then after each update at step k ,

$$\begin{aligned} L^{(i)}\left(z^{(i)[t+1]}; z^{(-i)}\right) - L^{(i)}\left(z^{(i)[t]}; z^{(-i)}\right) &\leq -\eta_i \left(1 - \frac{1}{2}\eta_i A_i\right) \sum_{j=1}^{m_i} \left\| \mathbf{G}_{j_i}^{(i)[t]} \right\|_F^2 - \frac{T}{2} \left\| \mathbf{p}^{(i)[t+1]} - \mathbf{p}^{(i)[t]} \right\|_2^2 \\ &\leq \eta_i \left(1 - \frac{1}{2}\eta_i A_i\right) \sum_{j=1}^{m_i} \left\| \mathbf{G}_{j_i}^{(i)[t]} \right\|_F^2. \end{aligned} \quad (\text{S50})$$

Since the last expression is non-positive, $L^{(i)}(z^{(i)[t]}; z^{(-i)})$ is non-decreasing. Summing Eq. (S50) over k and using the boundedness of $L^{(i)}$, we obtain

$$\sum_{k=0}^{\infty} \sum_{j_i=1}^{m_i} \left\| \mathbf{G}_{j_i}^{(i)[t]} \right\|_F^2 < \infty, \quad \Rightarrow \quad \liminf_{k \rightarrow \infty} \max_{j_i} \left\| \mathbf{G}_{j_i}^{(i)[t]} \right\|_F = 0. \quad (\text{S51})$$

Since $L^{(i)}$ is real-analytic on the compact manifold $\mathcal{M}$, it satisfies the Kurdyka-Łojasiewicz (KL) property. By the KL convergence principle [36], [56], the sequence $\{z^{(i)[t]}\}_t$ converges to a critical point $z^{(i)*}$ of $L^{(i)}$, i.e.,

$$\mathbf{G}_{j_i}^{(i)}\left(\mathbf{U}_{j_i}^{(i)*}, \mathbf{p}^{(i)*}\right) = 0, \quad \forall j_i. \quad (\text{S52})$$

Consequently, the optimal probabilities are given by the softmax.

Proof of Corollary 1:

Since $H^{[t]} \in (0, \log m_i)$ for all k , for any k and l we have

$$\begin{aligned} \left| \bar{r}^{(i)[t]} - \bar{r}^{(i)[l]} \right| &= \left| L^{(i)[t]} - L^{(i)[l]} \right| + T \left| H^{[t]} - H^{[l]} \right| \\ &\leq \left| L^{(i)[t]} - L^{(i)[l]} \right| + T \left| H^{[t]} - H^{[l]} \right|. \end{aligned} \quad (\text{S53})$$

If $\{L^{(i)[t]}\}$ converges, the first term vanishes. Thus,

$$\limsup_k \bar{r}^{(i)[t]} - \liminf_k \bar{r}^{(i)[t]} \leq T \left(\max H - \min H \right) \leq T \log m_i. \quad (\text{S54})$$

Proof of Theorem 6

Let $\{z^{[t,0]}\}_{t \geq 0}$ be the sequence of iterates generated by the iteration map, initialized at $z^{[0,0]}$. We first prove the existence of accumulation points (i.e., limit points). Since the manifold $\mathcal{M}$ is compact and the iteration sequence $\{z^{[t,0]}\} \subset \mathcal{M}$ is infinite, by sequential compactness there exists a convergent subsequence $\{z^{[t,0]}\}_{t \geq 0}$ and a point $z^* \in \mathcal{M}$ such that $\lim_{t \rightarrow \infty} z^{[t,0]} = z^*$, as guaranteed by standard results in manifold optimization [36]. Therefore, z^* is an accumulation point of the full sequence $\{z^{[t,0]}\}$. Next, we argue that the map $\mathcal{T} = \mathcal{T}_N \circ \dots \circ \mathcal{T}_1$ is continuous. This follows directly from Theorem 3, which ensures that the Riemannian gradient update over the unitary group is Lipschitz continuous, and the softmax update over the simplex

is smooth. As a result, each block update map $\mathcal{T}_i$ is continuous, and so is their composition $\mathcal{T}$ is also continuous.

Finally, we prove that any accumulation point must be a fixed point. Assume, for contradiction, that z^* is not a fixed point of $\mathcal{T}$, meaning that for some player i , the block update at z^* is not stationary. That is, either $\nabla_{U_{j_i}^{(i)}} L^{(i)}(z^*) \neq 0$ or $p^{(i)} \neq \sigma_T(\ell^{(i)})$. From Theorem 4, it then follows that there exists a neighborhood around such a point $\tilde{z}_i^*$ in which player i 's update yields a strict decrease in loss of at least some fixed constant $\delta > 0$.

Because $\{z^{[t,0]}\}$ converges to z^* , for sufficiently large t , $z^{[t+1,i-1]}$ lies within this neighborhood. As a result, the loss satisfies

$$L^{(i)}(z^{[t+1,i]}) \leq L^{(i)}(z^{[t+1,i-1]}) - \delta, \quad \text{for large } t. \quad (\text{S55})$$

This contradicts the convergence of the loss sequence $L^{(i)}(z^{[t,i]})$ as established in Theorem 5. Hence, each local update at the accumulation point must be block-stationary. That is,

$$\mathcal{T}_i(z_{i-1}^*) = \tilde{z}_i^*, \quad \text{with } \tilde{z}_0^* = z^*. \quad (\text{S56})$$

Recursively applying this for all $i \in \mathcal{N}$ yields $\mathcal{T}(z^*) = \mathcal{T}_N \circ \dots \circ \mathcal{T}_1(z^*) = \tilde{z}_N^* = z^*$, which shows that z^* is a fixed point of $\mathcal{T}$.

Proof of Theorem 7:

The proof proceeds in three steps. First, we identify the neutral directions associated with the eigenvalue 1 and separate them from the remaining spectrum. Second, we introduce local coordinates near the fixed-point submanifold and derive a suitable local representation of $\mathcal{T}$. Third, we show convergence of the iterates.

Step 1: Since $\mathcal{M}^{(\text{FP})}$ is a manifold of fixed points, for any smooth curve $z(t) \subset \mathcal{M}^{(\text{FP})}$ with $z(0) = z^*$ we have $\mathcal{T}(z(t)) = z(t)$. Differentiating at $t = 0$ gives $D\mathcal{T}(z^*)z'(0) = z'(0)$. Hence every tangent vector to $\mathcal{M}^{(\text{FP})}$ at z^* is an eigenvector of $D\mathcal{T}(z^*)$ associated with the eigenvalue 1. Therefore, $T_{z^*}\mathcal{M}^{(\text{FP})} \subseteq \mathcal{E}_1$, where $\mathcal{E}_1$ denotes the eigenspace of $D\mathcal{T}(z^*)$ corresponding to $\lambda = 1$. By assumption (1), $\dim T_{z^*}\mathcal{M}^{(\text{FP})} = l$. By assumption (2), the eigenspace $\mathcal{E}_1$ also has dimension l . Thus, $T_{z^*}\mathcal{M}^{(\text{FP})} = \mathcal{E}_1$. Therefore, the neutral directions of the linearized USMEA dynamics are exactly the tangent directions of the fixed-point submanifold $\mathcal{M}^{(\text{FP})}$. Since the eigenvalue 1 is semisimple and all other eigenvalues satisfy $|\lambda| < 1$, the tangent space $T_{z^*}\mathcal{M}$ admits the invariant decomposition

$$T_{z^*}\mathcal{M} = \mathcal{E}_1 \oplus \mathcal{E}_s, \quad (\text{S57})$$

where $\mathcal{E}_s$ is the invariant subspace corresponding with the eigenvalues strictly inside the unit disk. Since $\mathcal{E}_1 = T_{z^*}\mathcal{M}^{(\text{FP})}$, the subspace $\mathcal{E}_s$ can be regarded as the contracting normal space.

Step 2: Choose local coordinates (u, v) near z^* such that $T_{z^*}\mathcal{M} = \mathcal{E}_1 \oplus \mathcal{E}_s$, where $u \in \mathbb{R}^l$ parametrizes directions tangent to $\mathcal{M}^{(\text{FP})}$ and $v \in \mathbb{R}^r$ parametrizes directions in $\mathcal{E}_s$. Since $\mathcal{M}^{(\text{FP})}$ is a C^1 embedded submanifold, these coordinates may be chosen so that $\mathcal{M}^{(\text{FP})} = \{(u, v) : v = 0\}$. In these coordinates, the local representative of $\mathcal{T}$ can be written as [46], [57]

$$\tilde{\mathcal{T}}(u, v) = (u + a(u, v), Bv + b(u, v)), \quad (\text{S58})$$

with $a(u, 0) = 0, b(u, 0) = 0$, and where the spectrum of B consists of the eigenvalues of $D\mathcal{T}(z^*)$ other than 1. By assumption (3), every eigenvalue of B lies strictly inside the unit disk. Hence the spectral radius satisfies $r(B) < 1$. Since the space is finite-dimensional, there exists an equivalent norm $\|\cdot\|$ on $\mathbb{R}^r$ and a constant $q \in (0, 1)$ such that [58], [59]

$$\|Bv\| \leq q\|v\|, \quad \forall v \in \mathbb{R}^r. \quad (\text{S59})$$

Moreover, since $\tilde{\mathcal{T}}$ is smooth, after shrinking the neighborhood, there exists a constant $c > 0$ such that

$$\|a(u, v)\| \leq c\|v\|, \quad \|b(u, v)\| \leq c\|v\|^2. \quad (\text{S60})$$

Step 3: Convergence of the iterates:

Let

$$(u^{[k+1]}, v^{[k+1]}) = \tilde{\mathcal{T}}(u^{[k]}, v^{[k]}). \quad (\text{S61})$$

Then

$$v^{[k+1]} = Bv^{[k]} + b(u^{[k]}, v^{[k]}), \quad (\text{S62})$$

and therefore

$$\|v^{[k+1]}\| \leq q\|v^{[k]}\| + c\|v^{[k]}\|^2. \quad (\text{S63})$$

Choose $\theta \in (q, 1)$. By shrinking the neighborhood once more, we may assume that $c\|v\| \leq \theta - q$ throughout the neighborhood. Hence,

$$\|v^{[k+1]}\| \leq \theta \|v^{[k]}\|. \quad (\text{S64})$$

It follows that $v^{[k]} \rightarrow 0$ geometrically. For the tangential component, we have

$$u^{[k+1]} - u^{[k]} = a(u^{[k]}, v^{[k]}), \quad (\text{S65})$$

so

$$\|u^{[k+1]} - u^{[k]}\| \leq c\|v^{[k]}\|. \quad (\text{S66})$$

Since $\|v^{[k]}\|$ decays geometrically, the series

$$\sum_{k=0}^{\infty} \|u^{[k+1]} - u^{[k]}\| \quad (\text{S67})$$

converges. Therefore, $(u^{[k]})$ is a Cauchy sequence, and hence $u^{[k]} \rightarrow u^\infty$ for some u^∞ . Combining this with $v^{[k]} \rightarrow 0$, we obtain

$$(u^{[k]}, v^{[k]}) \rightarrow (u^\infty, 0) \in \mathcal{M}^{(\text{FP})}. \quad (\text{S68})$$

Returning to the original manifold coordinates yields $z^{[k]} \rightarrow z^\infty \in \mathcal{M}^{(\text{FP})}$. This proves the result.

S-III. DERIVATION OF THE BLOCK DIFFERENTIALS $D\mathcal{T}_i^U(z^\star)$ AND $D\mathcal{T}_i^P(z^\star)$

In this section, we compute the block differentials $D\mathcal{T}_i^U(z^\star)$ and $D\mathcal{T}_i^P(z^\star)$ by linearizing them at the fixed point $z^\star$.

Construction of $D\mathcal{T}_i^U(z^\star)$

Since $\mathcal{T}_i^U$ modifies only the unitary actions of player i , its differential is the identity on all blocks except the unitary-action block associated with player i .

At the fixed point $z^\star$, the linearization takes the form

$$\delta z_U^{(i)+} = \delta z_U^{(i)} + \eta_i \left(\sum_{i'=1}^N \mathbf{H}_{ii'} \delta z_U^{(i')} + \sum_{i'=1}^N \mathbf{B}_{ii'} \delta \mathbf{p}^{(i')} \right), \quad (\text{S69})$$

where $\delta z_U^{(i)}$ collects the perturbations of all unitary actions of player i , $\mathbf{H}_{ii'}$ denotes the block assembled from the matrices $\mathbf{H}_{ji,j_i'}^{(i,i')}$, and $\mathbf{B}_{ii'}$ denotes the block assembled from the vectors $\mathbf{B}_{ji,j_i'}^{(i,i')}$.

Therefore, $D\mathcal{T}_i^U(z^\star)$ is the identity on every block except the row corresponding to the unitary-action variables of player i . On that row, it is given by

$$[\eta_i \mathbf{H}_{i1}, \eta_i \mathbf{B}_{i1}, \dots, \mathbf{I} + \eta_i \mathbf{H}_{ii}, \eta_i \mathbf{B}_{ii}, \dots, \eta_i \mathbf{H}_{iN}, \eta_i \mathbf{B}_{iN}]. \quad (\text{S70})$$

In the following, we construct the blocks $\mathbf{H}_{ii'}$ and $\mathbf{B}_{ii'}$ in detail.

1) *Construction of $\mathbf{H}_{ii'}$:* We now compute the derivative of the action gradient with respect to the action variable $\mathbf{U}_{j_i'}^{(i')}$. For a tangent perturbation $\mathbf{X}^{(i')} \in \mathcal{SU}(d_{i'})$, define

$$\mathbf{H}_{ji,j_i'}^{(i,i')}[\mathbf{X}^{(i')}] := D_{\mathbf{U}_{j_i'}^{(i')}} \mathbf{G}_{ji}^{(i)}(z^\star)[\mathbf{X}^{(i')}]. \quad (\text{S71})$$

To evaluate this derivative, we perturb the action $\mathbf{U}_{j_i'}^{(i')}$ along the tangent direction $\mathbf{X}^{(i')}$ according to $\mathbf{U}_{j_i'}^{(i')}(t) = \exp(t\mathbf{X}^{(i')})\mathbf{U}_{j_i'}^{(i')}(0)$, and differentiate with respect to t at $t = 0$.

Using Eq. (S35) and the mixed-strategy weights, we obtain the following expression at the fixed point $z^\star$ for $i \neq i'$:

$$\mathbf{H}_{ji,j_i'}^{(i,i')}[\mathbf{X}^{(i')}] = p_{ji}^{(i)} p_{j_i'}^{(i')} \sum_{\mathbf{j}_{-(i,i')}} \left(\prod_{k \neq i,i'} p_{j_k}^{(k)} \right) \text{Tr}_{-i} \left(\left[\mathbf{R}_i, \left[\mathbf{X}^{(i')} \otimes \mathbf{I}^{(-i')}, \rho_{j_i,j_i',\mathbf{j}_{-(i,i')}}^\star \right] \right] \right). \quad (\text{S72})$$

Similarly, for $i = i'$, we have

$$\mathbf{H}_{j_i, j_i'}^{(i, i)}[\mathbf{X}^{(i)}] = \delta_{j_i, j_i'} p_{j_i}^{(i)} \sum_{\mathbf{j}_{-i}} \left(\prod_{k \neq i} p_{j_k}^{(k)} \right) \text{Tr}_{-i} \left(\left[\mathbf{R}_i, \left[\mathbf{X}^{(i)} \otimes \mathbf{I}^{(-i)}, \rho_{j_i, \mathbf{j}_{-i}}^* \right] \right] \right). \quad (\text{S73})$$

To obtain a matrix representation of this linear operator, we choose orthonormal bases

$$\{\mathbf{E}_1^{(i)}, \dots, \mathbf{E}_{d_i^2-1}^{(i)}\} \subset \mathcal{SU}(d_i), \quad \forall i, \quad (\text{S74})$$

with respect to the chosen inner product $\langle \mathbf{A}, \mathbf{B} \rangle$. Then the matrix entries of the operator $\mathbf{H}_{j_i, j_i'}^{(i, i')}$ are given by

$$[\mathbf{H}_{j_i, j_i'}^{(i, i')}]_{ab} = \langle \mathbf{E}_a^{(i)}, \mathbf{H}_{j_i, j_i'}^{(i, i')} [\mathbf{E}_b^{(i')}] \rangle = -\text{Tr} \left(\mathbf{E}_a^{(i)} \mathbf{H}_{j_i, j_i'}^{(i, i')} [\mathbf{E}_b^{(i')}] \right), \quad (\text{S75})$$

where the last identity uses the trace inner product on $\mathcal{SU}(d_i)$. After stacking over all indices, one obtains the block $\mathbf{H}_{ii'}$.

2) *Construction of $\mathbf{B}_{ii'}$* : We now compute the derivative of the action gradient with respect to the probability variable $p_{j_i'}^{(i')}$. Define

$$\mathbf{B}_{j_i, j_i'}^{(i, i')} := D_{p_{j_i'}^{(i')}} \mathbf{G}_{j_i}^{(i)}(z^*). \quad (\text{S76})$$

For $i' \neq i$, differentiating with respect to $p_{j_i'}^{(i')}$ gives

$$\mathbf{B}_{j_i, j_i'}^{(i, i')} = p_{j_i}^{(i)} \sum_{\mathbf{j}_{-(i, i')}} \left(\prod_{k \neq i, i'} p_{j_k}^{(k)} \right) \text{Tr}_{-i} \left(\left[\mathbf{R}_i, \rho_{j_i, j_i', \mathbf{j}_{-(i, i')}}^* \right] \right). \quad (\text{S77})$$

For $i' = i$, only the prefactor $p_{j_i}^{(i)}$ is differentiated, and therefore

$$\mathbf{B}_{j_i, j_i'}^{(i, i)} = \delta_{j_i, j_i'} \sum_{\mathbf{j}_{-i}} \left(\prod_{k \neq i} p_{j_k}^{(k)} \right) \text{Tr}_{-i} \left(\left[\mathbf{R}_i, \rho_{j_i, \mathbf{j}_{-i}}^* \right] \right). \quad (\text{S78})$$

The coordinates of $\mathbf{B}_{j_i, j_i'}^{(i, i')}$ with respect to the basis $\{\mathbf{E}_a^{(i)}\}$ are

$$[\mathbf{B}_{j_i, j_i'}^{(i, i')}]_a = \langle \mathbf{E}_a^{(i)}, \mathbf{B}_{j_i, j_i'}^{(i, i')} \rangle = -\text{Tr} \left(\mathbf{E}_a^{(i)} \mathbf{B}_{j_i, j_i'}^{(i, i')} \right). \quad (\text{S79})$$

Finally, after stacking over all indices, one obtains the block $\mathbf{B}_{ii'}$.

Construction of $D\mathcal{T}_i^p(z^*)$

Since $\mathcal{T}_i^p$ modifies only the probability vector of player i , its differential is the identity on all blocks except the probability-variable block of player i . Because the probability update is defined by the softmax function, its linearization at the fixed point z^* is

$$\delta \mathbf{p}^{(i)+} = \mathbf{S}_i \left(\sum_{i'=1}^N \mathbf{C}_{ii'} \delta z_U^{(i')} + \sum_{i'=1}^N \mathbf{D}_{ii'} \delta \mathbf{p}^{(i')} \right), \quad (\text{S80})$$

where

$$\mathbf{S}_i = \frac{1}{T} \left(\text{Diag}(\mathbf{p}^{(i)*}) - \mathbf{p}^{(i)*} \mathbf{p}^{(i)*T} \right), \quad (\text{S81})$$

and

$$\mathbf{C}_{ii'} = D_{z_U^{(i')}} \ell^{(i)}(z^*), \quad \mathbf{D}_{ii'} = D_{\mathbf{p}^{(i')}} \ell^{(i)}(z^*). \quad (\text{S82})$$

Therefore, $D\mathcal{T}_i^p(z^*)$ is the identity on every block except the row corresponding to the probability variables of player i . On that row, it is given by

$$[\mathbf{S}_i \mathbf{C}_{i1}, \mathbf{S}_i \mathbf{D}_{i1}, \dots, \mathbf{S}_i \mathbf{C}_{iN}, \mathbf{S}_i \mathbf{D}_{iN}]. \quad (\text{S83})$$

In the following, we derive the blocks $\mathbf{C}_{ii'}$ and $\mathbf{D}_{ii'}$ in detail.

3) *Construction of $\mathbf{C}_{ii'}$* : The derivative of the per-action payoff $\ell_{j_i}^{(i)}$ used in the probability update with respect to the action variable $\mathbf{U}_{j_{i'}}^{(i')}$, evaluated at the fixed point $\mathbf{z}^*$, is a scalar-valued linear functional of the perturbation $\mathbf{X}^{(i')}$. For $i \neq i'$,

$$C_{j_i, j_{i'}}^{(i, i')}[\mathbf{X}^{(i')}] = p_{j_{i'}}^{(i')} \sum_{\mathbf{j}_{-(i, i')}} \left(\prod_{k \neq i, i'} p_{j_k}^{(k)} \right) \text{Tr} \left(\mathbf{R}_i \left[\mathbf{X}^{(i')} \otimes \mathbf{I}^{(-i')}, \rho_{j_i, j_{i'}, \mathbf{j}_{-(i, i')}}^* \right] \right). \quad (\text{S84})$$

Similarly, for $i' = i$,

$$C_{j_i, j_i}^{(i, i)}[\mathbf{X}^{(i)}] = \delta_{j_i, j_i} \sum_{\mathbf{j}_{-i}} \left(\prod_{k \neq i} p_{j_k}^{(k)} \right) \text{Tr} \left(\mathbf{R}_i \left[\mathbf{X}^{(i)} \otimes \mathbf{I}^{(-i)}, \rho_{j_i, \mathbf{j}_{-i}}^* \right] \right). \quad (\text{S85})$$

Since $C_{j_i, j_{i'}}^{(i, i')}$ is a scalar-valued linear functional of the perturbation, it is represented with respect to the orthonormal basis $\{\mathbf{E}_b^{(i')}\}$ by the row vector

$$[C_{j_i, j_{i'}}^{(i, i')}]_b = C_{j_i, j_{i'}}^{(i, i')}[\mathbf{E}_b^{(i')}]. \quad (\text{S86})$$

Stacking these row vectors over all actions j_i and $j_{i'}$ yields the block $\mathbf{C}_{ii'}$ at the fixed point.

4) *Construction of $\mathbf{D}_{ii'}$* : The derivative of the per-action payoff $\ell_{j_i}^{(i)}$ used in the probability update with respect to the probability variable $p_{j_{i'}}^{(i')}$ is

$$D_{j_i, j_{i'}}^{(i, i')} := D_{p_{j_{i'}}^{(i')}} \ell_{j_i}^{(i)}(\mathbf{z}^*). \quad (\text{S87})$$

For $i' \neq i$, we have

$$D_{j_i, j_{i'}}^{(i, i')} = \sum_{\mathbf{j}_{-(i, i')}} \left(\prod_{k \neq i, i'} p_{j_k}^{(k)} \right) \text{Tr} \left(\mathbf{R}_i \rho_{j_i, j_{i'}, \mathbf{j}_{-(i, i')}}^* \right), \quad (\text{S88})$$

which is a scalar. Since $\ell_{j_i}^{(i)}$ does not depend on the mixed probabilities of player i , we have $D_{j_i, j_{i'}}^{(i, i')} = 0$. Stacking these scalars over the indices yields the block $\mathbf{D}_{ii'}$ at the fixed point.

Note that when $m_i = 1$, then the probability simplex of player i is trivial, so the probability update $\mathcal{T}_i^p$ is absent and $D\mathcal{T}_i(z^*) = D\mathcal{T}_i^U(z^*)$. In particular, if $m_i = 1$ for all players, then the one-sweep differential is constructed only from the action-step differentials, and no probability-related blocks exist.

S-IV. ADDITIONAL INFORMATION ON EXPERIMENTAL RESULTS

In our experiments, we consider an outcome set $\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$, where m represents the total number of outcomes, which remains the same for all players. We use the projective probability operator $\mathbf{P}_{\omega_j} = |\omega_j\rangle \langle \omega_j|$. Each player i has a payoff vector $\mathbf{r}_i = (R(\omega_1), R(\omega_2), \dots, R(\omega_m))$, which assigns a real-valued payoff to every possible outcome. By combining these payoff vectors with the outcome set, we construct the payoff operator for each game, as

$$\mathbf{R}_i = \sum_{\omega \in \Omega} R_i(\omega) \mathbf{P}_{\omega}. \quad (\text{S89})$$

We use the initial density matrix $\rho_0 = |\psi_0\rangle \langle \psi_0|$, where $|\psi_0\rangle$ is the initial state vector defined separately for each game and include entanglement parameters. Detailed descriptions of each game used in the experimental results are provided below

Game 1: In the two player quantum Prisoner's Dilemma, we follow Ref. [8] with $N = 2$, $d_i = 2$. The initial states, payoff outcomes are

$$|\psi_0\rangle = \cos(\gamma/2) |00\rangle + i \sin(\gamma/2) |11\rangle, \quad (\text{S90a})$$

$$(C, C): |\omega_1\rangle = \cos(\gamma/2) |00\rangle + i \sin(\gamma/2) |11\rangle, \quad (\text{S90b})$$

$$(C, D): |\omega_2\rangle = \cos(\gamma/2) |01\rangle + i \sin(\gamma/2) |10\rangle, \quad (\text{S90c})$$

$$(D, C): |\omega_3\rangle = i \sin(\gamma/2) |01\rangle + \cos(\gamma/2) |10\rangle, \quad (\text{S90d})$$

$$(D, D): |\omega_4\rangle = i \sin(\gamma/2) |00\rangle + \cos(\gamma/2) |11\rangle, \quad (\text{S90e})$$

The payoff vectors corresponding to players 1 and 2 are

$$\mathbf{r}_1 = (3, 0, 5, 1), \quad \mathbf{r}_2 = (3, 5, 0, 1). \quad (\text{S91})$$

Game 2: In this two player quantum game we use

$$|\psi_0\rangle = \cos(\gamma/2)|00\rangle + i \sin(\gamma/2)|12\rangle, \quad (\text{S92a})$$

$$|\omega_1\rangle = \cos(\gamma/2)|00\rangle + i \sin(\gamma/2)|12\rangle, \quad (\text{S92b})$$

$$|\omega_2\rangle = \cos(\gamma/2)|01\rangle + i \sin(\gamma/2)|11\rangle, \quad (\text{S92c})$$

$$|\omega_3\rangle = \cos(\gamma/2)|02\rangle + i \sin(\gamma/2)|10\rangle, \quad (\text{S92d})$$

$$|\omega_4\rangle = i \sin(\gamma/2)|02\rangle + \cos(\gamma/2)|10\rangle, \quad (\text{S92e})$$

$$|\omega_5\rangle = i \sin(\gamma/2)|01\rangle + \cos(\gamma/2)|11\rangle, \quad (\text{S92f})$$

$$|\omega_6\rangle = i \sin(\gamma/2)|00\rangle + \cos(\gamma/2)|12\rangle. \quad (\text{S92g})$$

The corresponding payoff vectors for the two players are given by

$$\mathbf{r}_1 = (4, 5, 0.5, 1, 1.15, 1.25), \quad \mathbf{r}_2 = (4.25, 0.52, 5.2, 1.1, 1.55, 1.9). \quad (\text{S93})$$

Game 3:

$$|\psi_0\rangle = \cos(\gamma/2)|00\rangle + i \sin(\gamma/2)|22\rangle, \quad (\text{S94a})$$

$$|\omega_1\rangle = \cos(\gamma/2)|00\rangle + i \sin(\gamma/2)|22\rangle, \quad (\text{S94b})$$

$$|\omega_2\rangle = \cos(\gamma/2)|01\rangle + i \sin(\gamma/2)|21\rangle, \quad (\text{S94c})$$

$$|\omega_3\rangle = \cos(\gamma/2)|02\rangle + i \sin(\gamma/2)|20\rangle, \quad (\text{S94d})$$

$$|\omega_4\rangle = \cos(\gamma/2)|10\rangle + i \sin(\gamma/2)|12\rangle, \quad (\text{S94e})$$

$$|\omega_5\rangle = e^{i\gamma/2}|11\rangle, \quad (\text{S94f})$$

$$|\omega_6\rangle = i \sin(\gamma/2)|10\rangle + \cos(\gamma/2)|12\rangle, \quad (\text{S94g})$$

$$|\omega_7\rangle = i \sin(\gamma/2)|02\rangle + \cos(\gamma/2)|20\rangle, \quad (\text{S94h})$$

$$|\omega_8\rangle = i \sin(\gamma/2)|01\rangle + \cos(\gamma/2)|21\rangle, \quad (\text{S94i})$$

$$|\omega_9\rangle = i \sin(\gamma/2)|00\rangle + \cos(\gamma/2)|22\rangle, \quad (\text{S94j})$$

and the corresponding payoff vectors for the two players are given by

$$\mathbf{r}_1 = (4, 5, 0.5, 1, 1.15, 1.25, 2, 11, 4), \quad \mathbf{r}_2 = (4.25, 11, 5.2, 6.1, 1.55, 1.9, 3, 2, 4). \quad (\text{S95})$$

Game 4: Three-Player Prisoner's Dilemma:

In the three-player quantum Prisoner's Dilemma game with $N = 3$ and $d_i = 2$ for all players, the initial states and payoff outcomes are as follows

$$|\psi_0\rangle = \cos(\gamma/2)|000\rangle + i \sin(\gamma/2)|111\rangle, \quad (\text{S96a})$$

$$|\omega_1\rangle = \cos(\gamma/2)|000\rangle + i \sin(\gamma/2)|111\rangle, \quad (\text{S96b})$$

$$|\omega_2\rangle = \cos(\gamma/2)|001\rangle + i \sin(\gamma/2)|110\rangle, \quad (\text{S96c})$$

$$|\omega_3\rangle = \cos(\gamma/2)|010\rangle + i \sin(\gamma/2)|101\rangle, \quad (\text{S96d})$$

$$|\omega_4\rangle = \cos(\gamma/2)|011\rangle + i \sin(\gamma/2)|100\rangle, \quad (\text{S96e})$$

$$|\omega_5\rangle = i \sin(\gamma/2)|011\rangle + \cos(\gamma/2)|100\rangle, \quad (\text{S96f})$$

$$|\omega_6\rangle = i \sin(\gamma/2)|010\rangle + \cos(\gamma/2)|101\rangle, \quad (\text{S96g})$$

$$|\omega_7\rangle = i \sin(\gamma/2)|001\rangle + \cos(\gamma/2)|110\rangle, \quad (\text{S96h})$$

$$|\omega_8\rangle = i \sin(\gamma/2)|000\rangle + \cos(\gamma/2)|111\rangle. \quad (\text{S96i})$$

The payoff vectors are

$$\mathbf{r}_1 = (3, 2, 2, 0, 5, 4, 4, 1), \quad \mathbf{r}_2 = (3, 2, 5, 4, 2, 0, 4, 1), \quad \mathbf{r}_3 = (3, 5, 2, 4, 2, 4, 0, 1). \quad (\text{S97})$$