GEOMETRIC STRUCTURE AND POLYNOMIAL-TIME ALGORITHM OF GAME EQUILIBRIA

Hongbo Sun *
Shenzhen International Graduate School
Tsinghua University
Shenzhen
shb20@tsinghua.org.cn

Chongkun Xia
School of Advanced Manufacturing
Sun Yat-sen University
Shenzhen
xiachk5@mail.sysu.edu.cn

Junbo Tan
Shenzhen International Graduate School
Tsinghua University
Shenzhen
tjblql@sz.tsinghua.edu.cn

Bo Yuan
Research Institute of
Tsinghua University in Shenzhen
Tsinghua University
Shenzhen
boyuan@ieee.org

Xueqian Wang
Shenzhen International Graduate School
Tsinghua University
Shenzhen
wang.xq@sz.tsinghua.edu.cn

Bin Liang
Department of Automation
Tsinghua University
Beijing
bliang@tsinghua.edu.cn

ABSTRACT

Whether a PTAS (polynomial-time approximation scheme) exists for game equilibria has been an open question, and its absence has indications and consequences in three fields: the practicality of methods in algorithmic game theory, non-stationarity and curse of multiagency in MARL (multi-agent reinforcement learning), and the tractability of PPAD in computational complexity theory. In this paper, we introduce a geometric object called equilibrium bundle, regarding which, first, we formalize perfect equilibria of dynamic games as the zero points of its canonical section, second, we formalize a hybrid iteration of dynamic programming and interior point method as a line search on it, third, we give the existence and oddness theorems of it as an extension of those of Nash equilibria. The line search leads to any perfect equilibrium of any dynamic game, it achieves a weak approximation (approximating to an ϵ-equilibrium) in fully polynomial time, and it achieves a strong approximation (approximating to an ϵ-neighborhood of an actual equilibrium) with dependent time complexity. Our method is an FPTAS (fully PTAS) for the PPAD-complete weak approximation problem of game equilibria, implying PPAD=FP. As intermediate results, we introduce two concepts called unbiased barrier problem and unbiased KKT conditions to make the interior point method to approximate Nash equilibria, and introduce a concept called policy cone to give the sufficient and necessary condition for dynamic programming to converge to perfect equilibria. In experiment, the line search process is animated, and the method is tested on 2000 randomly generated dynamic games where it converges to a perfect equilibrium in every single case.

***K*eywords** game theory · equilibrium · dynamic programming · interior point method · polynomial-time approximation scheme

MSC codes: 90C39, 90C51, 91A15

*For discussing technical details, please contact this author.

1 Introduction

1.1 PTAS for game equilibria

Whether game equilibria can be efficiently solved has been an open question since Nash equilibrium[22] is proposed. The absence of such a polynomial-time algorithm has indications and consequences in three different fields.

- In algorithmic game theory, the most widely used algorithms for approximating equilibria are no-regret[14, 28], self-play[2, 15], and their variants.
  - Neither of them converges to Nash equilibria in general static games. No-regret converges to coarse correlated equilibria, and self-play converges to strict Nash equilibria in static games that satisfy the fictitious play property.
  - When applied to dynamic games, neither of them guarantees the approximated equilibria to be perfect, namely, being a Nash equilibrium at every stage of the game, and thus non-perfect equilibria are not optimal in dynamic games.
- In MARL, there are problems known as non-stationarity and curse of multiagency[10], and these two problems are related to problems in algorithmic game theory, such that there is currently no algorithm that can converge to Nash equilibria or perfect equilibria in polynomial time for general games.
  - Non-stationarity means that the policies are hard to converge when reinforcement learning agents simultaneously maximize their utilities.
  - Curse of multiagency means that the computation needed for the policies of all agents to achieve optimal is exponential to the number of agents.
- In computational complexity theory, there are the following results about approximating Nash equilibria.
  - There are two different approximations: weak approximation and strong approximation[26].
    - Weak approximation is the approximation to an ϵ-equilibrium, namely, a policy profile such that every player is at most ϵ away from its maximum utility.
    - Strong approximation is the approximation to an ϵ-neighborhood of an actual equilibrium, namely, a policy profile whose distance from an actual equilibrium is less than ϵ.
    - A weak approximation of an actual equilibrium could be far from the actual equilibrium.
  - For static games with any number of players, weakly approximate computation of Nash equilibria is PPAD-complete[7], if ϵ is inversely proportional to a polynomial in the game size[4].
  - For static games with any number of players, weakly approximate computation of Nash equilibria for fixed ϵ is also PPAD-complete[25].
  - For static games with two players, exact computation of Nash equilibria is PPAD-complete[5].
  - For static games with three or more players, both exact and strongly approximate computation of Nash equilibria are FIXP-complete[9].

Complexity class PPAD is defined as all the problems that reduce to an *End-Of-The-Line* problem in polynomial time[24], and *End-Of-The-Line* is pretty convincingly intractable in polynomial time, making PPAD believed to contain hard problems, that is, it is believed to be unlikely that PPAD=FP.

In *End-Of-The-Line*, there is a directed graph DG, and a polynomial-time computable function f given by a boolean circuit. DG consists of vertices that are connected by arrows, where each vertex has at most one predecessor and at most one successor. Each vertex is encoded in n bits, and f takes the bits of a vertex as input to output its predecessor and successor, either of which may be none. A vertex is called an unbalanced vertex if exactly one of its predecessor and successor is none. *End-Of-The-Line* is the problem that: given an unbalanced vertex, find another unbalanced vertex. In DG, every vertex is either on a chain or isolated, and the problem seems to let us follow a potentially exponentially long chain from a given source to a sink, making *End-Of-The-Line* pretty convincingly intractable in polynomial time.

PPAD stands for Polynomial Parity Arguments on Directed graphs, where the parity argument refers to that given an unbalanced vertex, there exist an odd number of other unbalanced vertices. The oddness theorem[13] of Nash equilibria stating that there are an odd number of Nash equilibria for almost all[2] static games is also a parity argument. The Sperner's lemma in combinatorics, which is equivalent to Brouwer's fixed point theorem that is used to prove the existence[23] of Nash equilibria, is another parity argument. All these three parity arguments are based on the fact that the solutions are connected in pairs.

[2]Almost all means other cases form a null set in all the cases, that is, the probability of encountering them is zero.

1.2 Problem definition and major results

In this paper, we deal with fully observable dynamic games, and all the results of dynamic games hold for static games as single-state degenerations. Notations generally follow Einstein summation convention[8] [3] in order to simplify tensor operations in expressions, such as $V_s^i = \pi_A^s u_A^{si}$ represents the summation over index A of the product, while the product over index s is element-wise without summation, and $\pi_a^{si} \circ r_a^{si}$ represents element-wise product with no summation over any index. And we may use different index symbols to represent the same tensor, such as μ_a^i and $\mu_{a''}^k$ represent the same tensor. And $\mathbf{1}_s$ represents the tensor whose elements are all 1. In addition, there are a few unconventional notations to further simplify the expressions. For policy π_a^{si}, let

$$\pi_A^s := \left(\prod_{k\in N} \pi_{a_k}^{sk}\right)_A^s, \pi_{Aa}^{si-} := \left(I_{a_i a} \prod_{k\in N-\{i\}} \pi_{a_k}^{sk}\right)_{Aa}^{si}, \pi_{Aaa'}^{sij-} := \left([i\neq j]\cdot I_{a_i a} I_{a_j a'} \prod_{k\in N-\{i,j\}} \pi_{a_k}^{sk}\right)_{Aaa'}^{sij},$$

where A represents $(a_i)_{i\in N}$ and $[\cdot]$ is the Iverson bracket that outputs 1 if the input is true and outputs 0 otherwise, and let $\max_a^{si}$ represent the maximum with respect to index a for every index s and i. Finally, in dealing with static games, we may drop the index s when one state is referred and put it on when all states are referred, such as π_a^i represents $\pi_a^{si}(x)$ for some state x.

Definition 1 (Dynamic game)**.** *A dynamic game Γ is a tuple $(N, \mathcal{S}, \mathcal{A}, T, u, \gamma)$, where N is a set of players, $\mathcal{S}$ is a state space, $\mathcal{A}$ is an action space, $T : \mathcal{S} \times \prod_{i\in N} \mathcal{A} \to \Delta(\mathcal{S})$ is a transition function, $\Delta(\mathcal{S})$ is a space of probability distributions on $\mathcal{S}$, $u : \mathcal{S} \times \prod_{i\in N} \mathcal{A} \to \prod_{i\in N} \mathbb{R}$ is an utility function, $\mathbb{R}$ is the set of real numbers, $\gamma \in [0, 1)$ is a discount factor. A static game is a dynamic game with only one element in its state space $\mathcal{S}$.*

Definition 2 (Perfect equilibrium)**.** *A perfect equilibrium of dynamic game Γ is a policy $\pi : \mathcal{S} \to \prod_{i\in N} \Delta(\mathcal{A})$ such that $V_s^i = \max_a^{si} \pi_{Aa}^{si-}(u_A^{si} + \gamma T_{s'A}^s V_{s'}^i)$, where $V : \mathcal{S} \to \prod_{i\in N} \mathbb{R}$ is a value function that satisfies $V_s^i = \pi_A^s(u_A^{si} + \gamma T_{s'A}^s V_{s'}^i)$. A Nash equilibrium is a perfect equilibrium of a static game.*

Dynamic game and perfect equilibrium inherit the major characteristics of stochastic game and Markov perfect equilibrium[27, 21] in existing research, and they are defined in the most intuitive and general way. A dynamic game is a game where each player $i \in N$ chooses its action $a_i \in \mathcal{A}$ on every state $s \in \mathcal{S}$, then each player earns its utility $u(s, A)$ and the game randomly transits to a new state in probability space $T(s, A)$. A perfect equilibrium is a policy where all the players simultaneously earn their maximum cumulative utilities discounted by factor $\gamma \in [0, 1)$ at every state. And static games and Nash equilibria are the single-state degenerations of dynamic games and perfect equilibria. Denote $\mathcal{P} = \{\pi_a^{si} : \mathcal{S} \to \prod_{i\in N} \Delta(\mathcal{A})\}$ as the policy space, and $\mathcal{V} = \{V_s^i : \mathcal{S} \to \prod_{i\in N} \mathbb{R}\}$ as the value space.

In this paper, we introduce a geometric object called equilibrium bundle, along with unbiased barrier problem, unbiased KKT conditions, and Brouwer function, and the major results are centered around the equilibrium bundle. They are respectively introduced as we go through the following sections, but we put the definition of equilibrium bundle here only to describe our major results.

Definition 3 (Equilibrium bundle of dynamic games)**.** *An equilibrium bundle is the tuple $(E, \mathcal{P}, \alpha : E \to \mathcal{P})$ given by the following equations, where $U_{\pi A}^{si} = u_A^{si} + \gamma T_{s'A}^s V_{\pi s'}^i$, and $V_{\pi s}^i$ that depends on π_a^{si} is the unique solution of $V_s^i = D_\pi(V_s^i)$.*

$$\begin{aligned} E &= \bigcup_{\pi_a^{si}\in\mathcal{P}} \{\pi_a^{si}\} \times B(\pi_a^{si}) \\ B(\pi_a^{si}) &= \left\{v^{si} \circ \pi_a^{si} + \bar{\mu}_a^{si}(\pi_a^{si}) | v^{si} \geq 0\right\} \\ \bar{\mu}_a^{si}(\pi_a^{si}) &= \pi_a^{si} \circ \left(\max_a^{si} \pi_{Aa}^{si-} U_{\pi A}^{si} - \pi_{Aa}^{si-} U_{\pi A}^{si}\right) \\ \alpha\left((\pi_a^{si}, \mu_a^{si})\right) &= \pi_a^{si} \end{aligned} \tag{1}$$

The map $\bar{\mu}_a^{si} : \mathcal{P} \to \{\mu_a^{si} | \min_a^{si} \mu_a^{si} = \mathbf{0}^{si}\}$ is called the canonical section of the equilibrium bundle.

The equilibrium bundle is a geometric object called fiber bundle[4] in differential geometry, its base space is policy space $\mathcal{P}$, its total space is E consisting of pairs (π_a^{si}, μ_a^{si}), its fiber over $\pi_a^{si} \in \mathcal{P}$ is $\alpha^{-1}(\pi_a^{si}) = \{\pi_a^{si}\} \times B(\pi_a^{si})$, and canonical section $\bar{\mu}_a^{si}$ is its section that maps each $\pi_a^{si} \in \mathcal{P}$ to the least element in $B(\pi_a^{si})$. Conventionally, we can

[3] Please refer to the numpy.einsum function in NumPy[11] that our implementation in experiment mainly based on for a good illustration of Einstein summation convention.

[4] We only use the basic idea that the total space is the disjoint union of the fibers over the base space to formalize a geometric structure, no deeper theories are involved.

also call E as the equilibrium bundle, and call $B(\pi_a^{si})$ as its fiber over π_a^{si}. In particular, the equilibrium bundle of static games is also an algebraic variety [5] in algebraic geometry, such that it is the solution space of some polynomial equation system, which is the unbiased KKT conditions. The major results are several facts about the equilibrium bundle, which we show as we go through the following sections.

- Equilibrium bundle, unbiased barrier problem, unbiased KKT conditions, and Brouwer function all lie in the joint space of policy and barrier parameter with certain geometric structure.
  - Being a point on the equilibrium bundle, a global optimal point of the unbiased barrier problem, a solution of the unbiased KKT conditions, and a fixed point of the Brouwer function are all equivalent.
  - Perfect equilibria of dynamic games are the zero points of the canonical section of the equilibrium bundle, where the canonical section depicts the global distribution of perfect equilibria.
  - The equilibrium bundle of dynamic games admits the existence theorem and the equilibrium bundle of static games admits the oddness theorem, as an extension of those of Nash equilibria. These two theorems are implied respectively by Brouwer's fixed point theorem[23] and the uniqueness of analytic continuations of algebraic curves[13].
- There is a hybrid iteration of dynamic programming and interior point method formalized as a line search method on the equilibrium bundle.
  - The method consists of two levels of iteration: updating onto the equilibrium bundle by alternating the steps of projected gradient descent of the unbiased barrier problem and dynamic programming, and hopping across the fibers of the equilibrium bundle to a zero point of the canonical section.
  - For any perfect equilibrium of any dynamic game, there are infinite many search path on the equilibrium bundle that lead exactly to it. The method achieves a weak approximation, namely, the approximation to an ϵ-approximate perfect equilibrium, in fully polynomial time, and the time complexity of achieving a strong approximation, namely, the approximation to an ϵ-neighborhood of an actual perfect equilibrium, depends on the gradient of the canonical section near the actual perfect equilibrium.
  - The computation of the algorithm is based on a variant of the expected utility instead of a specific game model, and thus the algorithm works with any game model that has a polynomial-time algorithm for the expected utility problem, and also with model-free cases.
  - As for intermediate results:
    * In dealing with static games, we introduce two equivalent concepts called unbiased barrier problem and unbiased KKT conditions, making the interior point method to satisfy a so-called primal-dual unbiased condition using these two concepts, so that it can approximate Nash equilibria by locally optimizing a regret minimization problem.
    * In dealing with dynamic games, we introduce a concept called policy cone, giving the iteration properties of dynamic programming and the equivalent conditions of equilibria in the context of policy cone, and then use them to give the sufficient and necessary condition for dynamic programming to converge to perfect equilibria.

Considering the research status and our major contributions, there are three direct impacts of our discovery.

- Our discovery leads to a fundamental leap in the understanding of game equilibria through the equilibrium bundle.
- There is previously no efficient algorithms solving game equilibria with comprehensive applicability, but now there is, and it is general enough to derive an efficient algorithm for any problem involving strategic interactions.
- PPAD=FP, where PPAD class is previously considered to contain hard problems, but our discovery shows it is actually not, meaning that there is an efficient algorithm for every problem in the PPAD class to be discovered.

2 Technical route and related work

There are three major challenges in approximating perfect equilibria of dynamic games in fully polynomial time, which are respectively addressed in three sections.

[5] We only use the uniqueness of analytic continuations of algebraic curves (one-dimensional algebraic varieties), no deeper theories are involved.

- Section 3: guaranteeing the policy to converge to a Nash equilibrium in a static game.
  - **First problem:** current methods like no-regret and self-play fundamentally lack the ability to converge to Nash equilibria.
    - According to the purification theorem[12], a policy is a Nash equilibrium if and only if for every player, every non-zero weighted action has the same utility.
    - As a consequence, mixed Nash equilibria are not stable equilibrium points for any method based on optimizing the utility of the updated mixed strategy, such as no-regret and self-play.
      - In no-regret[14], regret is a vector of differences between the utility of the updated mixed strategy and the utilities of every action, and the regret is minimized.
      - In self-play[2], the utility of the updated mixed strategy is maximized.
    - Our method is based on the regret minimization problem (2), where the regret is a vector of differences between the maximum utility and the utilities of every action plus a scalar. Thus, when the regret is minimized, for every player, every non-zero weighted action has the same utility. In other words, Nash equilibria are equivalently the global optimal points of problem (2).
    - **Another problem arises:** the global optimal points of problem (2) are Nash equilibria, but we intend to use local optimization methods that are polynomial-time.
      - We introduce a primal-dual unbiased condition, such that Nash equilibria are equivalently the primal-dual unbiased local extreme points of problem (2).
      - We choose the interior point method as our local optimization method, where there are two levels of iteration: updating onto the central path by locally optimizating the barrier problem, and updating along the central path to a local extreme point by gradually reducing the barrier parameter to zero.
  - **Another problem arises:** we want the interior point method to always lead to a primal-dual unbiased local extreme point of problem (2), instead of just an ordinary local extreme point.
    - We introduce two concepts called (primal-dual) unbiased barrier problem (4) and (primal-dual) unbiased KKT conditions (5), such that a global optimal point of (4) equals a solution of (5), and equals a primal-dual unbiased local extreme point of the barrier problem of (2). Thus, the equivalent result of the primal-dual unbiased condition is generalized from problem (2) to its barrier problem.
    - Then we transform the interior point method into a primal-dual unbiased interior point method that keeps the update on a (primal-dual) unbiased central path.
      - In the first iteration level, usually, the barrier problem is numerically optimized by SQP (sequential quadratic programming)[3], which is a numerical second-order method. But for unbiased barrier problem, there exists an analytical expression of its projected gradient. So we use projected gradient (6) to update onto the unbiased central path, and this analytical first-order method is much more simple.
      - In the second iteration level, we reduce the barrier parameter to update along the unbiased central path.
    - There are three assumptions for the primal-dual unbiased interior point method to work.
      - Starting point: there has to be a starting point sufficiently close to the unbiased central path.
      - Differentiability: the updated point has to move an infinitesimal step along the unbiased central path with an infinitesimal reduction of the barrier parameter.
      - Convexity: unbiased barrier problem has to be strictly locally convex for projected gradient descent (6) to update onto the unbiased central path.
  - **Another problem arises:** we need to find a way to satisfy the three assumptions in the primal-dual unbiased interior point method.
    - We introduce a geometric object called equilibrium bundle, along with its canonical section and singular point. Equilibrium bundle is a fiber bundle in differential geometry.
      - Its base space is the policy space.
      - Its total space is the solution space of unbiased KKT conditions (5) that consists of pairs of policy and barrier parameter.
      - Its fiber is the solution subspace of (5) with a given policy, which consists of barrier parameters.
      - Its canonical section maps every policy to the least barrier parameter on its fiber.
      - Its singular point is the point where the Jacobian matrix is singular.
      - Nash equilibria are the zero points of its canonical section.
    - There are several facts about the equilibrium bundle that respectively address the three assumptions.
      - Starting point: for any given policy, a large enough barrier parameter on its fiber is nearly on the equilibrium bundle.

        - Differentiability: on non-singular points of the equilibrium bundle, policy moves an infinitesimal step with an infinitesimal step of barrier parameter.
        - Convexity: on non-singular points of the equilibrium bundle, unbiased barrier problem (4) is strictly locally convex.
        - Singularity: for any given policy, a large enough barrier parameter on its fiber is guaranteed to be a non-singular point pairing with the policy. Note that moving along a fiber changes neither the policy nor the canonical section, so this singular avoidance does not affect the line search to Nash equilibria.
      * We combine the canonical section and the singular avoidance to obtain a canonical section descent formula (9) to update the barrier parameter.
      * We then transform the primal-dual unbiased interior point method into **a line search method on the equilibrium bundle**, which consists of two levels of iteration: updating onto the equilibrium bundle by projected gradient descent (6), and hopping across the fibers of the equilibrium bundle to a zero point of its canonical section by canonical section descent (9).
  - So far, **no more problems**, and we have constructed a line search method on the equilibrium bundle, such that the method guarantees to converge to any Nash equilibrium of any static game.
  - To show the significance of the equilibrium bundle, we additionally extend the existence and oddness theorems of Nash equilibria onto the equilibrium bundle, such that the two theorems of Nash equilibria are implied by degenerating the two theorems of the equilibrium bundle.
      * For any barrier parameter, there exists a policy such that it is on the equilibrium bundle, implied by Brouwer's fixed point theorem[23].
      * For any barrier parameter, there are almost always an odd number of policies such that they are on the equilibrium bundle, implied by the uniqueness of analytic continuations of algebraic curves[13].
- Section 4: guaranteeing the policy to converge in a dynamic game.
  - **First problem:** current methods that directly use the Bellman operator in dynamic games fundamentally lack the ability to converge at all.
      * Value iteration, which uses the Bellman operator for iteration, is a polynomial-time exact algorithm for computing optimal policies of MDPs (Markov decision process)[1, 20]. The iterative convergence of Bellman operator is contraction mapping convergence. And it is also directly used to solve perfect equilibria of dynamic games in early attempts[19, 16], showing that they generally do not converge.
      * We introduce a new dynamic programming operator and only require it to have monotone convergence, where contraction mapping convergence can be considered as a convergence in every direction, and monotone convergence can be considered as a convergence in only a certain direction. Meantime, the Bellman operator is a special case of our dynamic programming operator, allowing us to analyze why Bellman operator does not converge in dynamic games later.
  - **Another problem arises:** find the region that our dynamic programming operator achieves monotone convergence.
      * We introduce a concept called policy cone, along with a best response cone.
        - The policy cone is a region in the value function space with the best response cone being its subregion, with respect to a given policy.
        - The policy cone has an apex that is the value function of its policy.
        - Every value function plus a large enough scalar is in the best response cone.
      * For our dynamic programming operator, the policy cone is exactly its monotonic domain, and a point that is initially in the policy cone never leaves the policy cone under its iteration. Thus, our dynamic programming operator guarantees monotone convergence to the apex of a policy cone for fixed policy. In addition, adding a scalar at each iteration step does not affect the convergence.
      * A policy is a perfect equilibrium if and only if the apex of its policy cone is in its best response cone.
  - **Another problem arises:** we want our dynamic programming operator to maintain the monotone convergence as the policy varies, not just for fixed policies, and converge to a perfect equilibrium value function, not just an ordinary apex.
      * We first study whether Bellman operator has the monotone convergence property. It is shown that in multi-player case, Bellman operator achieves neither contraction mapping convergence nor monotone convergence, and in single-player case, it achieves both contraction mapping convergence and monotone convergence, explaining why it cannot be simply generalized from MDPs to dynamic games.

    - Then we study our dynamic programming operator.
      - Not only is the policy cone bounded below, but the apex of it is also bounded below with respect to a varying policy. Thus, monotone convergence can be maintain by letting the iteration stay in the policy cone by adding a large enough scalar and letting this scalar converge to zero.
      - The value function is now guaranteed to converge to some apex by iteration, and the apex needs to be in the best response cone to guarantee it to be a perfect equilibrium value function. Thus, we further require the iteration to stay in the best response cone.
    - Summarizing the above, we obtain the sufficient and necessary condition to make dynamic programming iteration (14) to converge to perfect equilibrium value functions. A sequence of policies that converges to a Nash equilibrium at every state of the dynamic game is sufficient to satisfy this condition.
  - So far, we have achieved two things. First, we find a dynamic programming method that guarantees convergence in dynamic game. Second, we transform the convergence to a perfect equilibrium to the convergence to a Nash equilibrium at every state, which is solved in section 3.
- Section 5: guaranteeing the converged policy to be a Nash equilibrium at every stage of the dynamic game, namely, to be a perfect equilibrium.
  - **First problem:** current methods that directly use variants of the static game method in dynamic games fundamentally lack the ability to converge to perfect equilibria.
    - For every dynamic game, there is a corresponding static game, in which the behavioral strategies of the dynamic game serve as the static strategies of the corresponding static game.
    - When the variants of no-regret[28] and self-play[15] are used in dynamic games, they deal with interactions of behavioral strategies, instead of static strategies at each stage of the dynamic game. This means that their resulting equilibria are at most Nash equilibria of the corresponding static game of behavioral strategy interactions, while not guaranteed to be a Nash equilibrium at every stage of the dynamic game.
    - We can solve this problem simply by combining the results in section 4 and section 3.
  - **Another problem arises:** dynamic programming is not an iteration that moves an infinitesimal step everytime except for when it is nearly converged, which would cause the updated point to be too far from the equilibrium bundle to update back.
    - This is solved by setting the utility of the fiber over every policy as the one generated by the value function of this policy on the equilibrium bundle, and putting dynamic programming (14) at the same level of iteration as the projected gradient descent (6).
      - Dynamic programming is always nearly converged near the equilibrium bundle.
      - The combined method is still formalized as a line search on the equilibrium bundle.
  - So far, **no more problems**, and we formalize a hybrid iteration of dynamic programming and interior point method as a line search on equilibrium bundle, which leads to any perfect equilibrium of any dynamic game, then we show that the method achieves weak approximation in fully polynomial time. In addition, all the results apply to static games as single-state degenerations.

3 Interior point method on static game

3.1 Approximating Nash equilibrium by local optimization

This section mainly deals with static games, so we drop the state index s as we state before. We first give a regret minimization problem (2) and explain why current approximation methods for Nash equilibrium lack convergence guarantee using the purification theorem. Then we give Theorem 1 that asserts two equivalent conditions of Nash equilibrium, one is being a global optimal point of problem (2), the other is being a local extreme point that also satisfies a condition called primal-dual unbiased condition. Finally, the second equivalent condition would allow us to approximate Nash equilibria by local optimization, which is chosen as the interior point method.

Definition 4 (Regret minimization problem)**.** *Let G be a static game, where the utility function is $U : \prod_{i\in N} \mathcal{A} \to \prod_{i\in N} \mathbb{R}$. A regret minimization problem of G is optimization problem* (2) *with $\mu_a^i = 0$, where (π_a^i, r_a^i, v^i) is a tuple of policy, regret, and value, and μ_a^i is a barrier parameter.*

$$\begin{aligned} \min_{(\pi_a^i, r_a^i, v^i)} \quad & \pi_a^i r_a^i - \mu_a^i \ln r_a^i - \mu_a^i \ln \pi_a^i \\ \textit{s.t.} \quad & r_a^i - v^i + \pi_{Aa}^{i-} U_A^i = 0 \\ & \mathbf{1}_a \pi_a^i - \mathbf{1}^i = 0 \end{aligned} \tag{2}$$

The case where $\mu_a^i > 0$ is a barrier problem of (2)*.*

According to the purification theorem[12], π_a^i is a Nash equilibrium if and only if for every player i, every action a that satisfies $\pi_a^i > 0$ has the same utility. Thus, mixed Nash equilibria are not stable equilibrium points for any method based on optimizing the utility of the updated mixed strategy, such as no-regret[14, 28] and self-play[2, 15]. Note that the meanings of regret are different in no-regret methods and in this paper. The regret in no-regret methods is a vector of differences between the utility of the updated mixed strategy and the utilities of every action, while the regret in regret minimization problem (2) is a vector of differences between the maximum utility and the utilities of every action plus a scalar. The difference upon the quantity being optimized is why our method has convergence guarantee while existing methods based on no-regret or self-play do not, and consequently exhibit non-stationarity when used in MARL.

Theorem 1. *Let G be a static game and (π_a^i, r_a^i, v^i) be a tuple of policy, regret and value of G. Then the following statements are equivalent.*

(i) *(π_a^i, v^i) is a Nash equilibrium of G.*

(ii) *(π_a^i, r_a^i, v^i) is an optimal point of regret minimization problem* (2)*.*

(iii) *There exist Lagrangian multipliers $(\bar{\lambda}_a^i, \tilde{\lambda}^i, \hat{\pi}_a^i, \hat{r}_a^i)$ that satisfies two conditions: the KKT conditions shown by equation* (3) *with $\mu_a^i = 0$, and the primal-dual unbiased condition $\pi_a^i = \hat{\pi}_a^i$.*

$$\begin{bmatrix} \bar{\lambda}_a^j \pi_{Aaa'}^{ij-} U_A^i + \tilde{\lambda}^i \mathbf{1}_{a'} + r_{a'}^i - \hat{r}_{a'}^i \\ \bar{\lambda}_a^i + \pi_a^i - \hat{\pi}_a^i \\ -\mathbf{1}_a \bar{\lambda}_a^i \\ r_a^i \circ \hat{\pi}_a^i - \mu_a^i \\ \pi_a^i \circ \hat{r}_a^i - \mu_a^i \\ r_a^i - v^i + \pi_{Aa}^{i-} U_A^i \\ \mathbf{1}_a \pi_a^i - \mathbf{1}^i \end{bmatrix} = 0 \tag{3}$$

The case where $\mu_a^i > 0$ of equation (3) *is the perturbed KKT conditions of* (2)*.*

Furthermore, when these statements hold, the objective function $\pi_a^i r_a^i$ of (2) *is* 0*, and $(\bar{\lambda}_a^i, \tilde{\lambda}^i, \pi_a^i - \hat{\pi}_a^i, r_a^i - \hat{r}_a^i) = 0$.*

Theorem 1 points out that the Nash equilibrium is not only equivalent to the global optimal point of regret minimization problem (2), but also equivalent to its local extreme point that satisfies the primal-dual unbiased condition, where the local extreme point is namely the point that satisfies the KKT conditions[17] in (iii). In previous research[18, 6], linear programming, quadratic programming, linear complementarity problem, and Nash equilibrium of bimatrix games have been unified, and the unification is actually a degeneration of $(i) \Leftrightarrow (ii)$ into two-player static games. $(i) \Leftrightarrow (ii)$ points out the connection between our method and existing methods, but it is not necessary in constructing our method. $(i) \Leftrightarrow (iii)$ makes it possible to use local optimization that leads to local extreme points where primal-dual bias $\pi_a^i - \hat{\pi}_a^i$ is 0 to compute global optimal points of (2), and equivalently Nash equilibria.

We choose **interior point method** to perform the local optimization, where **central path** is a path in the solution space of perturbed KKT conditions (3) with a scalar barrier parameter μ, the **first iteration level** is to update onto the central path by locally optimizating the barrier problem (2), the **second iteration level** is to update along the central path by reducing the scalar barrier parameter μ, and the central path leads to a local extreme point of the orginal problem as μ reduces to 0. In our case, we additionally intend to keep the update on a particular central path on which primal-dual bias $\pi_a^i - \hat{\pi}_a^i$ is 0.

3.2 Unbiased barrier problem and unbiased KKT conditions

Note that Theorem 1 only applies to the $\mu_a^i = 0$ case, but we have to study the $\mu_a^i > 0$ case to find a way to keep the update on the central path where primal-dual bias is 0. Thus, we introduce two concepts called unbiased barrier problem and unbiased KKT conditions, and give Theorem 2 to extend our equivalent result about the $\mu_a^i = 0$ case in Theorem 1 (iii) onto the subject of interior point method where $\mu_a^i > 0$, so that we can construct a primal-dual unbiased interior point method. In addition, we use these two concepts to give an extension of the existence theorem of Nash equilibria.

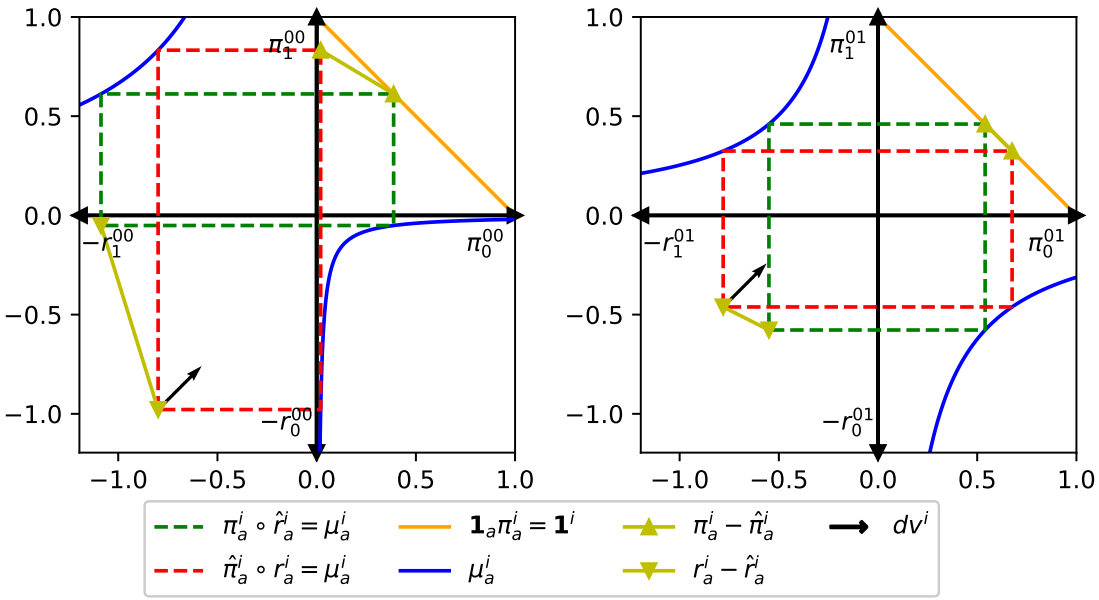

Fig. 1 Graph of unbiased barrier problem. This figure is plotted with a dynamic game where $N = \mathcal{A} = \{0, 1\}, \mathcal{S} = \{0\}$. The graph is based on the joint space of policy π_a^i and regret r_a^i. The positive half of the two axes represent two action indexes $a \in \{0, 1\}$ of π_a^i, the negative half of the two axes represent two action indexes $a \in \{0, 1\}$ of r_a^i, and the two subfigures represent two player indexes $i \in \{0, 1\}$. Plotting π_a^i and $\hat{\pi}_a^i$ on the all positive orthant, r_a^i and $\hat{r}_a^i$ on the all negative orthant, and μ_a^i between positive half and negative half of the axes as hyperbolas, $\hat{\pi}_a^i \circ r_a^i = \mu_a^i$ and $\pi_a^i \circ \hat{r}_a^i = \mu_a^i$ have rectangular shapes, and $(\pi_a^i - \hat{\pi}_a^i, r_a^i - \hat{r}_a^i)$ is the bias of two rectangles. dv^i is the direction r_a^i can move with fixed π_a^i and within the constraint $r_a^i = v^i - \pi_{Aa}^{i-} U_A^i$ of unbiased barrier problem (4). With r_a^i moving in direction dv^i and the other two corners of the rectangle fixed on the hyperbolas, the figure illustrates how there is an unique v^i to let $\hat{\pi}_a^i$ satisfy $\mathbf{1}_a \hat{\pi}_a^i = \mathbf{1}^i$ as stated in Theorem 3 (i), and the right subfigure shows a case where it is satisfied.

Definition 5 (Unbiased barrier problem)**.** *An unbiased barrier problem is the optimization problem*

$$\begin{aligned} \min_{(\pi_a^i, r_a^i, v^i)} \quad & \left(\pi_a^i - \hat{\pi}_a^i\right)\left(r_a^i - \hat{r}_a^i\right) \\ s.t. \quad & r_a^i - v^i + \pi_{Aa}^{i-} U_A^i = 0 \\ & \mathbf{1}_a \pi_a^i - \mathbf{1}^i = 0 \end{aligned} \tag{4}$$

parameterized by $\hat{\pi}_a^i$ and $\hat{r}_a^i$, and $\hat{\pi}_a^i = \mu_a^i / r_a^i$ and $\hat{r}_a^i = \mu_a^i / \pi_a^i$, which are called dual policy and dual regret respectively, and the tuple $(\pi_a^i - \hat{\pi}_a^i, r_a^i - \hat{r}_a^i)$ is called primal-dual bias.

Definition 6 (Unbiased KKT conditions)**.** *Unbiased KKT conditions are simultaneous equations*

$$\begin{bmatrix} \hat{\pi}_a^i \circ r_a^i - \mu_a^i \\ r_a^i - v^i + \pi_{Aa}^{i-} U_A^i \\ \mathbf{1}_a \hat{\pi}_a^i - \mathbf{1}^i \end{bmatrix} = 0, \tag{5a}$$

$$\pi_a^i = \hat{\pi}_a^i. \tag{5b}$$

Definition 7 (Brouwer function)**.** *Brouwer function is the family of maps $M : \{\mu_a^i | \mu_a^i \geq 0\} \to (\mathcal{P} \to \mathcal{P})$ parameterized by μ_a^i such that $M(\mu_a^i)(\pi_a^i) = \hat{\pi}_a^i$ satisfies equation* (5a)*.*

Theorem 2. *Given $\mu_a^i > 0$, for the tuple (π_a^i, r_a^i, v^i), the following properties satisfy $(i) \Leftrightarrow (ii) \Leftrightarrow (iii) \Leftrightarrow ((iv) \wedge \pi_a^i = \hat{\pi}_a^i)$ and $(iv) \Leftrightarrow (v) \Leftrightarrow (vi)$.*

(i) *Being a fixed point of Brouwer function $\hat{\pi}_a^i = M(\mu_a^i)(\pi_a^i)$.*

(ii) *Being a global optimal point of unbiased barrier problem* (4)*.*

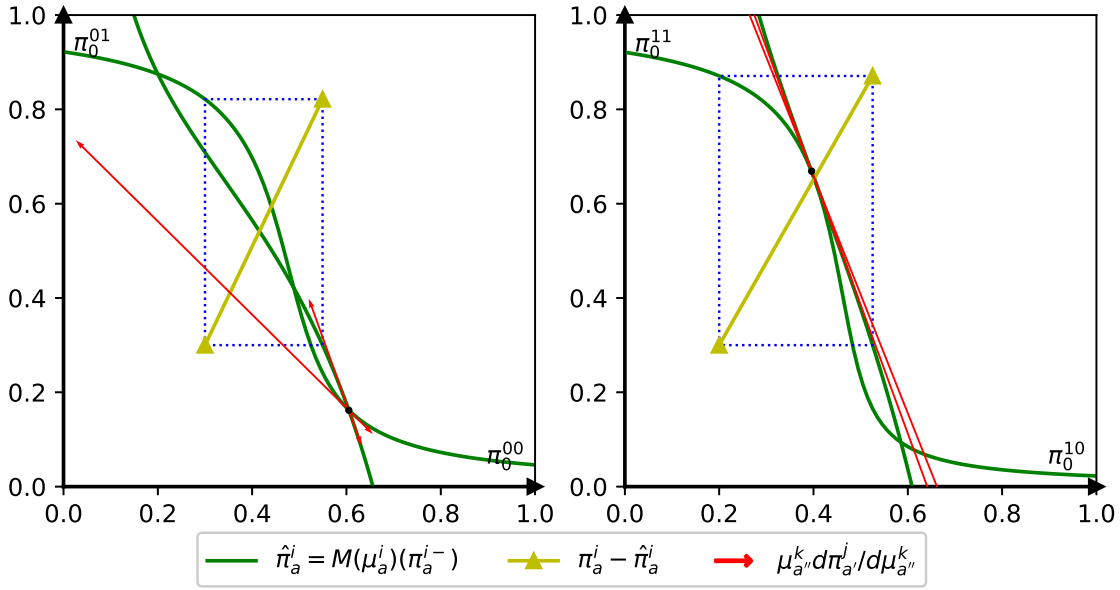

Fig. 2 Graph of unbiased KKT conditions. This figure is plotted with a dynamic game where $N = \mathcal{S} = \mathcal{A} = \{0, 1\}$. The graph is based on the policy space, where the two axes represent two player indexes $i \in \{0, 1\}$ of π_a^{si}, the two subfigures represent two state indexes $s \in \{0, 1\}$ of π_a^{si}, and only one of the two action indexes $a = 0$ is needed to represent π_a^{si} since π_a^{si} sums to 1 over action indexes. $\hat{\pi}_a^i = M(\mu_a^i)(\pi_a^{i-})$ is a set of hypersurfaces indexed by $i \in N$ and induced by the Brouwer function $M(\mu_a^i)$ for a given μ_a^i, $\pi_a^i - \hat{\pi}_a^i$ shows the mapping of $M(\mu_a^i)$, and the intersections of the hypersurfaces are fixed points of $M(\mu_a^i)$. There is at least one intersection of the hypersurfaces according to Theorem 3 (iii), and there are almost always an odd number of intersections according to Theorem 6, as extensions of the existence and oddness theorems of Nash equilibria. Differential $(d\pi_{a'}^j/\pi_{a'}^j)/(d\mu_{a''}^k/\mu_{a''}^k)$ illustrates that a intersection π_a^i moves with the i-th hypersurface as μ_a^i varies on index i, since the i-th hypersurface is only relevant to μ_a^i on index i. The right subfigure shows a singular point in Definition 8, where the differential grows infinite large.

(iii) *Being a solution of unbiased KKT conditions* (5).

(iv) *Being a solution of perturbed KKT conditions* (3) *for some* $(\bar{\lambda}_a^i, \tilde{\lambda}^i, \hat{\pi}_a^i, \hat{r}_a^i)$.

(v) *Being a KKT point of unbiased barrier problem* (4).

(vi) *Being a local extreme point of barrier problem* (2).

In particular, π_a^i *is a Nash equilibrium if and only if* $(\pi_a^i, \mathbf{0}_a^i)$ *is a solution of unbiased KKT conditions* (5).

There are two equivalence formulas in Theorem 2. The first one $(i) \Leftrightarrow (ii) \Leftrightarrow (iii) \Leftrightarrow ((iv) \wedge \pi_a^i = \hat{\pi}_a^i)$ establishs the equivalence between the unbiased barrier problem, Brouwer function, unbiased KKT conditions, and the $\mu_a^i > 0$ case in Theorem 1 (iii), allowing us the construct a primal-dual unbiased interior point method. The Brouwer function connects unbiased barrier problem and unbiased KKT conditions, such that fixed points of the Brouwer function are solutions of the unbiased KKT conditions, and the unbiased barrier problem depicts the approximation of fixed points of the Brouwer function. Actually, we only need the first formula to make subsequent constructions, while the second formula $(iii) \Leftrightarrow (iv) \Leftrightarrow (v)$ only illustrates the connection between our method and interior point method, that is, a path that consists of solutions of the unbiased KKT conditions really is a central path in interior point methods on which the primal-dual bias is 0, ignoring the difference that the barrier parameter in our method is actually a vector instead of a scalar.

Theorem 3. *The following properties about unbiased barrier problem* (4) *and unbiased KKT conditions* (5) *hold.*

(i) *For any* π_a^i *and* μ_a^i*, there exists an unique* v^i *that satisfies equation* (5a).

(ii) *If* $\mathbf{1}_a\hat{\pi}_a^i - \mathbf{1}^i = 0$*, then the projected gradient of* (4) *is*

$$\mathrm{pg}_{a''}^j := \left(I_{a'a''} - \frac{\mathbf{1}_{a'}\mathbf{1}_{a''}}{|\mathcal{A}|}\right)\left(\left(r_{a'}^j - \hat{r}_{a'}^j\right) - \left(\pi_a^i - \hat{\pi}_a^i\right)\pi_{Aaa'}^{ij-}U_A^i\right). \tag{6}$$

(iii) (Existence theorem) *For every* μ_a^i*, there is at least one solution* (π_a^i, μ_a^i) *of* (5)*.*

Theorem 3 (i) shows two facts. First, (i) shows that there is an unique v^i to let dual policy $\hat{\pi}_a^i$ satisfy constraint $\mathbf{1}_a\pi_a^i = \mathbf{1}^i$. As a consequence, (ii) shows that when $\mathbf{1}_a\hat{\pi}_a^i = \mathbf{1}^i$, $\pi_a^i - \hat{\pi}_a^i$ and dv^i are orthogonal, and the projected gradient of unbiased barrier problem (4) has the form of an analytical expression as equation (6) shows. Second, (i) shows that for every policy π_a^i, there is an unique dual policy $\hat{\pi}_a^i$ that satisfies (5a). As a consequence, the Brouwer function $\hat{\pi}_a^i = M(\mu_a^i)(\pi_a^i)$ is indeed a map, then by Brouwer's fixed point theorem, (iii) asserts the existence of a solution of unbiased KKT conditions (5) for every μ_a^i, as an extension of the existence of Nash equilibria[23].

In summary, we've transformed the interior point method into a **primal-dual unbiased interior point method**, where **(primal-dual) unbiased central path** is a path in the solution space of unbiased KKT conditions (5), the **first iteration level** is to update onto the unbiased central path by projected gradient descent (6), the **second iteration level** is to update along the unbiased central path by reducing barrier parameter μ_a^i, and the unbiased central path leads to a Nash equilibrium as μ_a^i reduces to 0. However, there are three assumptions for the primal-dual unbiased interior point method to work. **Starting point**: the iteration has to start from a point that is sufficiently close to a known point on the unbiased central path. **Differentiability**: in the second iteration level, the policy π_a^i has to move an infinitesimal step along the unbiased central path with an infinitesimal reduction of the barrier parameter μ_a^i, so that the first iteration level can update back onto the unbiased central path. **Convexity**: in the first iteration level, unbiased barrier problem (4) has to be strictly locally convex for projected gradient descent (6) to update onto the unbiased central path. Settling these assumptions is beyond the capacity of mathematical optimization, it turns out to be a geometric problem.

3.3 Equilibrium bundle

We first introduce a geometric object called equilibrium bundle to structure the solution space of unbiased KKT conditions (5), where the unbiased central paths lie. Then we settle the three assumptions using the geometric properties of the equilibrium bundle, transforming our primal-dual unbiased interior point method into a line search method on the equilibrium bundle. Finally, we give the oddness theorem of the equilibrium bundle as an extension of that of Nash equilibria.

Definition 8 (Equilibrium bundle)**.** *An equilibrium bundle of a static game* G *is the tuple* $(E, \mathcal{P}, \alpha : E \to \mathcal{P})$ *given by the following equations, where* $\mathcal{P} = \prod_{i\in N}\Delta(\mathcal{A})$ *is the policy space.*

$$\begin{aligned}
E &= \bigcup_{\pi_a^i\in\mathcal{P}} \{\pi_a^i\} \times B(\pi_a^i)\\
B(\pi_a^i) &= \left\{v^i \circ \pi_a^i + \bar{\mu}_a^i(\pi_a^i) | v^i \geq 0\right\}\\
\bar{\mu}_a^i(\pi_a^i) &= \pi_a^i \circ \left(\max_a{}^i\, \pi_{Aa}^{i-}U_A^i - \pi_{Aa}^{i-}U_A^i\right)\\
\alpha\left((\pi_a^i, \mu_a^i)\right) &= \pi_a^i
\end{aligned} \tag{7}$$

The map $\bar{\mu}_a^i : \mathcal{P} \to \{\mu_a^i | \min_a^i \mu_a^i = \mathbf{0}^i\}$ *is called the canonical section of the equilibrium bundle. The point where* $\mathrm{C}_{(j,a',l)}^{(i,a,m)}$ *is singular is called a singular point of it, where* $\mathrm{C}_{(j,a',l)}^{(i,a,m)}$ *is given by the coefficient matrix of equation* (8)*, .*

Theorem 4. *Let* $(E, \mathcal{P}, \alpha : E \to \mathcal{P})$ *be the equilibrium bundle of static game* G*, then the following statements hold.*

(i) $(\pi_a^i, \mu_a^i) \in E$ *if and only if* (π_a^i, μ_a^i) *is a solution of unbiased KKT conditions* (5)*.*

(ii) *Given* π_a^i*, then* $\mu_a^i \in B(\pi_a^i)$ *if and only if* (π_a^i, μ_a^i) *is a solution of unbiased KKT conditions* (5)*.*

(iii) *Given* π_a^i*, then for any* $\mu_a^i \in B(\pi_a^i)$*,* $\mu_a^i \geq \bar{\mu}_a^i(\pi_a^i)$*.*

(iv) $\bigcup_{\pi_a^i\in\mathcal{P}} B(\pi_a^i) = \{\mu_a^i | \mu_a^i \geq 0\}$*.*

(v) π_a^i *is a Nash equilibrium if and only if the canonical section* $\bar{\mu}_a^i(\pi_a^i) = \mathbf{0}_a^i$*.*

Definition 8 and Theorem 4 show the relation between the equilibrium bundle, the solution space of unbiased KKT conditions (5), and the joint space $\mathcal{P} \times \{\mu_a^i | \mu_a^i \geq 0\}$ of policy and barrier parameter. First, the equilibrium bundle is

the solution space of unbiased KKT conditions (5), and a fiber $B(\pi_a^i)$ is the solution subspace relating to a given π_a^i. Second, $B(\pi_a^i)$ is the part of a $|N|$-dimensional affine space that is in $\{\mu_a^i|\mu_a^i \geq 0\}$, the canonical section $\bar{\mu}_a^i(\pi_a^i)$ is the least element in $B(\pi_a^i)$ and is on the boundary of $\{\mu_a^i|\mu_a^i \geq 0\}$, and the union of all the $B(\pi_a^i)$ is exactly $\{\mu_a^i|\mu_a^i \geq 0\}$. Finally, Theorem 4 (v) shows that Nash equilibria are exactly the zero points of the canonical section $\bar{\mu}_a^i$.

Theorem 4 (v) has two implications. First, unlike the primal-dual unbiased interior point method, μ_a^i does not have to decrease to 0 to reach a Nash equilibrium, instead, μ_a^i can be any value on the fiber as long as the canonical section $\bar{\mu}_a^i(\pi_a^i)$ decreases to 0. Second, the canonical section $\bar{\mu}_a^i(\pi_a^i)$ depicts the global distribution of Nash equilibria, which can be used to search the policy space globally for the entire set of Nash equilibria.

In summary, the equilibrium bundle is exactly the structured solution space of unbiased KKT conditions (5) where the unbiased central paths lie, making it possible to leverage the geometric properties of the equilibrium bundle to study the three assumptions left to settle in the primal-dual unbiased interior point method, which gives the following theorem.

Theorem 5. *The following properties about equilibrium bundle E hold.*

(i) *For any $\hat{\mu}_a^i > 0$, the algebraic curve $\{(\pi_a^i, \hat{\mu}_a^i\mu') \in E|\mu' > 0\}$ satisfies*

$$\lim_{\mu' \to +\infty} \pi_a^i = \frac{\hat{\mu}_a^i}{\mathbf{1}_a \hat{\mu}_a^i}.$$

(ii) *The differential $(d\pi_{a'}^j/\pi_{a'}^j)/(d\mu_{a''}^k/\mu_{a''}^k)$ at $(\pi_a^i, \mu_a^i) \in E$ satisfies*

$$\begin{bmatrix} H_{(j,a')}^{(i,a)} & \hat{B}_l^{(i,a)} \\ \check{B}_{(j,a')}^m & \mathbf{0}_l^m \end{bmatrix} \begin{bmatrix} \left(\frac{\mu_{a''}^k d\pi_{a'}^j}{\pi_{a'}^j d\mu_{a''}^k}\right)_{(k,a'')}^{(j,a')} \\ \left(\frac{\mu_{a''}^k dv^l}{d\mu_{a''}^k}\right)_{(k,a'')}^l \end{bmatrix} = \begin{bmatrix} \mathrm{Diag}\left(\left(\mu_a^i/\pi_a^i\right)_{(i,a)}\right) \\ \mathbf{0}_{(k,a'')}^m \end{bmatrix}, \tag{8}$$

where

$$H_{(j,a')}^{(i,a)} = \mathrm{Diag}\left(\left(\mu_a^i/\pi_a^i\right)_{(i,a)}\right) - \left(\pi_{Aaa'}^{ij-} U_A^i \circ \pi_{a'}^j\right)_{(j,a')}^{(i,a)},$$
$$\hat{B}_l^{(i,a)} = \left(I^{il}\mathbf{1}_a\right)_l^{(i,a)}, \check{B}_{(j,a')}^m = \pi_{(j,a')} \circ \left(I^{jm}\mathbf{1}_{a'}\right)_{(j,a')}^m.$$

Denote the coefficient matrix of equation (8) *as* $\mathrm{C}_{(j,a',l)}^{(i,a,m)}$.

(iii) *Unbiased barrier problem* (4) *is locally strictly convex at $(\pi_a^i, \mu_a^i) \in E$ if $\mathrm{C}_{(j,a',l)}^{(i,a,m)}$ is non-singular.*

(iv) *For any π_a^i, there exists $\check{\mu}_a^i$ on its fiber $B(\pi_a^i)$ such that for every $\mu_a^i > \check{\mu}_a^i$, $\mathrm{C}_{(j,a',l)}^{(i,a,m)}$ is non-singular on (π_a^i, μ_a^i).*

Theorem 5 settles all the three assumptions with the context transformed from the unbiased central path to the equilibrium bundle. **Starting point**: (i) shows that we can choose any policy as the starting policy $\pi_{a,0}^i$ simply by setting $\mu_{a,0}^i = \mu'\pi_{a,0}^i$, and then $(\pi_{a,0}^i, \mu_{a,0}^i)$ is sufficiently close to the equilibrium bundle when μ' is sufficiently large. **Differentiability**: (ii) shows that the equilibrium bundle is differentiable as long as the coefficient matrix $\mathrm{C}_{(j,a',l)}^{(i,a,m)}$ is **non-singular** according to the implicit function theorem, and the differential $(d\pi_{a'}^j/\pi_{a'}^j)/(d\mu_{a''}^k/\mu_{a''}^k)$ points out the direction to update along the equilibrium bundle as μ_a^i decreases. **Convexity**: (iii) shows that unbiased barrier problem (4) is strictly locally convex as long as the coefficient matrix $\mathrm{C}_{(j,a',l)}^{(i,a,m)}$ is **non-singular**. Finally, (iv) shows that the point where $\mathrm{C}_{(j,a',l)}^{(i,a,m)}$ is **non-singular** can always be easily found.

A point where $\mathrm{C}_{(j,a',l)}^{(i,a,m)}$ is singular is a singular point of the equilibrium bundle as defined in Definition 8. Theorem 5 (ii) and (iii) and the experiments show that singular points may block the iteration, and thus they must be avoided. (iv) shows that this is always possible simply by adding $\beta^i \circ \pi_a^i$ to μ_a^i for a sufficiently large β^i. Normally, it is harmless to pick any β^i to perform the addition at any time, but note that the differential $(d\pi_{a'}^j/\pi_{a'}^j)/(d\mu_{a''}^k/\mu_{a''}^k)$ approaches identity matrix I as $\mu_a^i \to \infty$. In other words, on a point $(\pi_a^i, \mu_a^i) \in E$ where μ_a^i is too large, the equilibrium bundle is flat and the convergence is slow. Considering all the above, we give the canonical section descent (9) to update μ_a^i, where $\eta^i \in [0, 1)^i$ is the step length that is small enough, and $\beta^i \circ \pi_a^i$ is the term for singular point avoidance.

$$\mu_{a,t+1}^i = (1 - \eta_t^i) \circ \mu_{a,t}^i + \beta_t^i \circ \pi_{a,t}^i \tag{9}$$

In summary, we've transformed the primal-dual unbiased interior point method into a **line search on the equilibrium bundle**, where **equilibrium bundle** is the structured solution space of unbiased KKT conditions (5), the **first iteration level** is to update onto the equilibrium bundle by projected gradient descent (6), the **second iteration level** is to hop across the fibers of the equilibrium bundle by canonical section descent (9) and differential $(d\pi^j_{a'}/\pi^j_{a'})/(d\mu^k_{a''}/\mu^k_{a''})$, and the line search leads to a Nash equilibrium as the canonical section $\bar{\mu}^i_a(\pi^i_a)$ reduces to 0. There is also a method to search globally for the entire set of Nash equilibria using the canonical section. Specifically, we can sample policies in the policy space and calculate the canonical sections $\bar{\mu}^i_a(\pi^i_a)$ on them, then we can set the samples where $\bar{\mu}^i_a(\pi^i_a)$ are close to 0 as the starting points and search for Nash equilibria near them. The method is given by Algorithm 1 as a single-state degeneration.

Theorem 6 (Oddness theorem)**.** *For every μ^i_a, there are almost always an odd number of solutions of unbiased KKT conditions* (5)*.*

Theorem 6 gives an extension of the oddness theorem of Nash equilibria[13], and it is implied by the uniqueness of analytic continuations of algebraic curves and that almost every point on an algebraic variety is non-singular. For any algebraic curve $\mathrm{AC} = \{(\pi^i_a, \hat{\mu}^i_a\mu') \in E | \mu' > \mu''\}$ given by $\hat{\mu}^i_a > 0$ and $\mu'' \geq 0$, denoting EP as the set of points on AC as $\mu' \to \mu''$. First, there is exactly one point in EP that is connected by a branch of AC to the unique starting point of AC stated in Theorem 5 (i) as $\mu' \to \infty$, and all the other points in EP are connected in pairs by the rest branches of AC, and thus $|\mathrm{EP}|$ is odd. Second, EP equals to $\{(\pi^i_a, \hat{\mu}^i_a\mu'') \in E\}$ if every point in EP is non-singular, and this is almost always the case, where $\{(\pi^i_a, \hat{\mu}^i_a\mu'') \in E\}$ is the entire set of solutions of (5) at $\hat{\mu}^i_a\mu''$,

4 Dynamic programming on dynamic games

4.1 Policy cone and best response cone

In single-player case, value iteration, which uses the Bellman operator $\max_a(u^s_a + \gamma T^s_{s'a}V_{s'})$ to iterate value function V_s, is a polynomial-time exact algorithm for computing optimal policies of MDPs. But the similar operator generally does not converge in the multi-player dynamic games, which is explained later in this section. We define two dynamic programming operators $D_\pi : \mathcal{V} \to \mathcal{V}$ and $\hat{D}_\pi : \mathcal{V} \to \mathcal{V}$, both of which are maps parameterized by policy π^{si}_a such that

$$D_\pi(V^i_s) = \pi^s_A(u^{si}_A + \gamma T^s_{s'A}V^i_{s'}), \quad \hat{D}_\pi(V^i_s) = \max_a{}^{si} \pi^{si-}_{Aa}(u^{si}_A + \gamma T^s_{s'A}V^i_{s'}).$$

The convergence of Bellman operator is contraction mapping convergence, but we only require D_π and $\hat{D}_\pi$ to achieve monotone convergence. Contraction mapping convergence can be considered as a convergence in every direction, and monotone convergence can be considered as a convergence in only a certain direction, and we later show that the direction is $\mathbf{1}_s$.

We first introduce a concept called policy cone, along with a best response cone, which is used to bridge between the dynamic programming operators and the equilibria, where Theorem 8 shows that the iteration of D_π monotonically converges to the apex of the policy cone for fixed π^{si}_a, and Theorem 9 shows the equivalent conditions for a policy to be a perfect equilibrium and for a policy to be a Nash equilibrium at a given state.

Definition 9 (Policy cone and best response cone)**.** *Let π^{si}_a be a policy in dynamic game Γ. A policy cone C_π and a best response cone $\hat{C}_\pi$ are regions in value function space $\mathcal{V}$ such that*

$$C_\pi = \left\{V^i_s \in \mathcal{V} | V^i_s \geq D_\pi(V^i_s)\right\}, \quad \hat{C}_\pi = \left\{V^i_s \in \mathcal{V} | V^i_s \geq \hat{D}_\pi(V^i_s)\right\}.$$

Proposition 7. *Let π^{si}_a be a policy in dynamic game Γ, and then the following properties hold.*

(i) *There is an unique value function $V^i_{\pi s}$ that satisfies*

$$V^i_{\pi s} = D_\pi(V^i_{\pi s}). \tag{10}$$

In addition, $V^i_{\pi s}$ satisfies $V^i_{\pi s} \in C_\pi$, and for any $V^i_s \in C_\pi$, $V^i_s \geq V^i_{\pi s}$.

(ii) *For any $V^i_s \in \mathcal{V}$ and $x \in \mathcal{S}$, there exists an unique pair $(Y^i_{xs}, d^i_x) \in \mathcal{V} \times \mathbb{R}^N$, as well as an unique pair $(\hat{Y}^i_{xs}, \hat{d}^i_x) \in \mathcal{V} \times \mathbb{R}^N$, such that*

$$\left(Y^i_{xs} - D_\pi(Y^i_{xs})\right)(x) = 0 \quad \wedge \quad V^i_s - Y^i_{xs} = d^i_x\mathbf{1}_s, \tag{11a}$$

$$\left(\hat{Y}^i_{xs} - \hat{D}_\pi(\hat{Y}^i_{xs})\right)(x) = 0 \quad \wedge \quad V^i_s - \hat{Y}^i_{xs} = \hat{d}^i_x\mathbf{1}_s. \tag{11b}$$

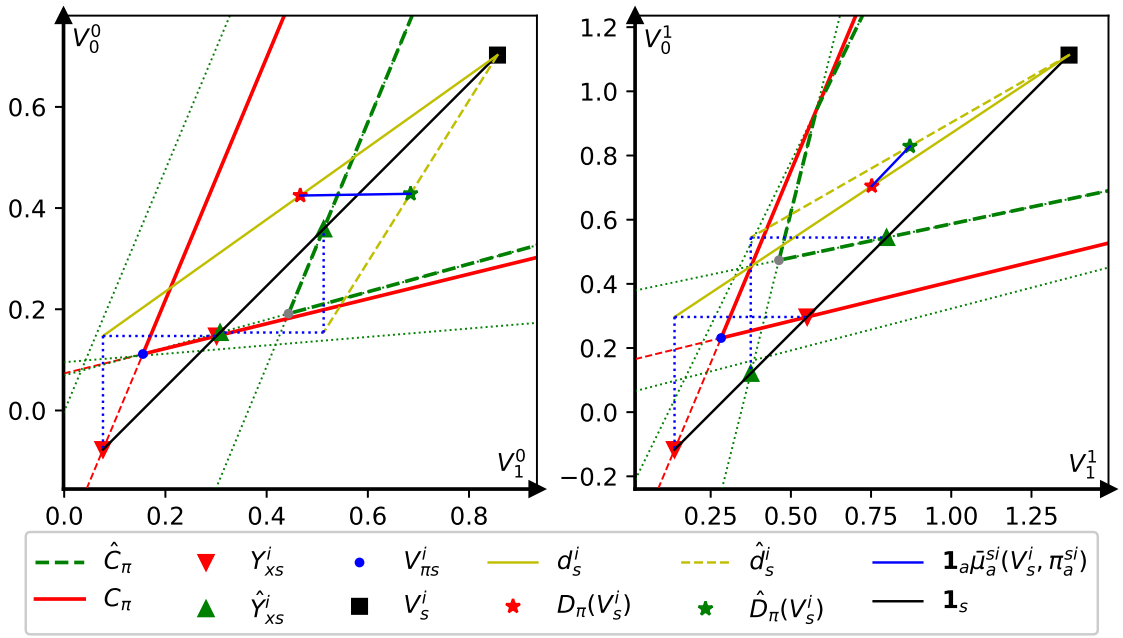

Fig. 3 Graph of policy cone. This figure is plotted with a dynamic game where $N = \mathcal{S} = \mathcal{A} = \{0,1\}$. The graph is based on the value function space $\mathcal{V}$, where the two axes represent two state indexes $s \in \{0,1\}$ of V^i_s, and the two subfigures represent two player indexes $i \in \{0,1\}$ of V^i_s. As Proposition 7 shows, C_π is a set of hyperplane-surrounded cone-shaped regions indexed by $i \in N$, with $V^i_{\pi s}$ being its apexes, and with $\hat{C}_\pi$ contained in it. $\mathbf{1}_s$ is the monotone convergence direction, which induces unique pairs (Y^i_{xs}, d^i_x) and $(\hat{Y}^i_{xs}, \hat{d}^i_x)$, and satisfies that $V^i_s + m^i\mathbf{1}_s$ lies in $\hat{C}_\pi$ for any V^i_s and sufficiently large m^i. Theorem 8 states that the residuals satisfy $V^i_s - D_\pi(V^i_s) = (1-\gamma)d^i_s$ and $V^i_s - \hat{D}_\pi(V^i_s) = (1-\gamma)\hat{d}^i_s$. Theorem 9 states that $\pi^{si}_a(x)$ is a Nash equilibrium if and only if the corresponding pair of Y^i_{xs} and $\hat{Y}^i_{xs}$ coincide, and π^{si}_a is a perfect equilibrium if and only if $V^i_{\pi s}$ lie in $\hat{C}_\pi$, in which case Y^i_{xs}, $\hat{Y}^i_{xs}$, and $V^i_{\pi s}$ all coincide. Equation (15) shows that the relation between the canonical section and the two dynamic programming operators is $\mathbf{1}_a\bar{\mu}^{si}_a(V^i_s,\pi^{si}_a) = \hat{D}_\pi(V^i_s) - D_\pi(V^i_s)$.

(iii) *For any $V^i_s \in \mathcal{V}$, there exists an $M^i > 0$ such that for any $m^i > M^i$, $V^i_s + m^i\mathbf{1}_s \in \hat{C}_\pi \subseteq C_\pi$.*

According to Definition 9 and Proposition 7 (i), policy cone C_π is a set of cone-shaped regions indexed by $i \in N$, each surrounded by $|\mathcal{S}|$ hyperplanes in $|\mathcal{S}|$-dimensional value function space $\mathcal{V}$, with $V^i_{\pi s}$ being a set of their apexes, and best response cone $\hat{C}_\pi$ is a subset of policy cone C_π for every $i \in N$, with each subset being an intersect of $|\mathcal{A}|$ cone-shaped regions. $V^i_{\pi s}$ is the value function in Definition 2 of perfect equilibrium, and $V^i_{\pi s}(x)$ is exactly the expected utility of the sequence generated by π^{si}_a with x as the initial state in dynamic game Γ for any $x \in \mathcal{S}$. Proposition 7 (ii) and (iii) show that both bottoms of C_π and $\hat{C}_\pi$ expand towards infinity along $\mathbf{1}_s$ such that $V^i_s + m^i\mathbf{1}_s$ always lies in both of them for sufficiently large m^i. and there is always an unique intersection Y^i_{xs} between the line in direction $\mathbf{1}_s$ passing through $V^i_s \in \mathcal{V}$ and each hyperplane of C_π indexed by x and i, and the same goes for $\hat{Y}^i_{xs}$ and $\hat{C}_\pi$.

Theorem 8 (Iterative properties)**.** *Let π^{si}_a be a policy in dynamic game Γ, and then the following properties of dynamic programming operator D_π hold.*

(i) *$V^i_s \geq D_\pi(V^i_s)$ if and only if $V^i_s \in C_\pi$.*

(ii) *For any $V^i_s \in C_\pi$, $D_\pi(V^i_s) \in C_\pi$.*

(iii) *For any $V^i_s \in C_\pi$, if $D_\pi(V^i_s) = \hat{D}_\pi(V^i_s)$, then $D_\pi(V^i_s) \in \hat{C}_\pi$.*

(iv) *For any $V^i_s \in C_\pi$, the residual*

$$V^i_s - D_\pi(V^i_s) = (1-\gamma)d^i_s, \tag{12a}$$

$$V^i_s - \hat{D}_\pi(V^i_s) = (1-\gamma)\hat{d}^i_s, \tag{12b}$$

where d^i_s and $\hat{d}^i_s$ are given by formula (11a) *and* (11b) *respectively.*

(v) *Let* V_s^i *iterates by* $V_{s,k+1}^i = D_\pi(V_{s,k}^i + m^i \mathbf{1}_s)$*, where* $m^i > 0$ *is a constant, and initial value function* $V_{s,0}^i \in C_\pi$*. Then*

$$\lim_{k\to\infty} V_{s,k}^i = V_{\pi s}^i + \frac{\gamma}{1-\gamma} m^i \mathbf{1}_s \quad \wedge \quad \lim_{k\to\infty} \left(V_{s,k}^i - D_\pi(V_{s,k}^i)\right) = \gamma m^i \mathbf{1}_s. \tag{13}$$

Theorem 8 uses policy cone C_π and best response cone $\hat{C}_\pi$ to describe iterative properties of dynamic programming operator D_π for fixed policy π_a^{si}. (i), (ii), and (iii) suggest that C_π is the monotonic and closed domain for D_π, and $\hat{C}_\pi$ is closed if $\pi_a^{si}(x)$ is a Nash equilibrium for every state x. That is, for any iteration starting from value function V_s^i within C_π, the value function decreases monotonically as iterates and never leaves C_π, which means the iteration converges to the apex $V_{\pi s}^i$ of C_π by the monotone convergence theorem. (iv) illustrates that iteration residual $V_s^i - D_\pi(V_s^i)$ of D_π can be expressed by the distance d_x^i pairing with the unique intersection Y_{xs}^i described in Proposition 7 (ii), where the set of distances d_x^i indexed by state x is used as a single vector d_s^i, and the same goes for $\hat{D}_\pi$, $\hat{d}_x^i$, and $\hat{Y}_{xs}^i$. In addition to value function V_s^i iteratively converging to apex $V_{\pi s}^i$, (v) shows that if a scaled $\mathbf{1}_s$ is added in every iteration, not only does V_s^i converge to $V_{\pi s}^i$ adding a scaled $\mathbf{1}_s$, but $V_s^i - D_\pi(V_s^i)$ also converge to a scaled $\mathbf{1}_s$, indicating that $\mathbf{1}_s$ is the direction of the monotone convergence of D_π.

Theorem 9 (Equilibrium conditions)**.** *The following equivalent conditions of equilibrium hold.*

(i) π_a^{si} *is a perfect equilibrium if and only if* $V_{\pi s}^i \in \hat{C}_\pi$*.*

(ii) *Given value function* V_s^i*, for any* $x \in \mathcal{S}$*,* $\pi_a^{si}(x)$ *is a Nash equilibrium of the static game with utility function* $(u_A^{si} + \gamma T_{s'A}^s V_{s'}^i)(x)$ *if and only if* $Y_{xs}^i = \hat{Y}_{xs}^i$*, where* Y_{xs}^i *and* $\hat{Y}_{xs}^i$ *are given by formula* (11a) *and* (11b) *respectively.*

Theorem 9 uses C_π and $\hat{C}_\pi$ to describe equivalent conditions of equilibria. (ii) shows that policy $\pi_a^{si}(x)$ is a Nash equilibrium for state x if and only if the unique intersects Y_{xs}^i and $\hat{Y}_{xs}^i$ coincide for state x. (i) shows that policy π_a^{si} is a perfect equilibrium if and only if the apex $V_{\pi s}^i$ of C_π is in $\hat{C}_\pi$. Note that when π_a^{si} is a perfect equilibrium, Y_{xs}^i, $\hat{Y}_{xs}^i$, and $V_{\pi s}^i$ all coincide, which is exactly the definition of perfect equilibrium such that $V_{\pi s}^i$ is the value function of π_a^{si} and $\pi_a^{si}(x)$ is a Nash equilibrium for every state x.

4.2 Iteration in policy cone

Theorem 8 shows that for fixed policy, the iteration of dynamic programming operator D_π in the policy cone converge monotonically to the apex, and Theorem 9 shows equivalent condition for the apex of policy cone to be a perfect equilibrium value function. We next study the possibility for iterative methods constructed by D_π to converge to a perfect equilibrium value function. First, we give Proposition 10 to explain why single-player dynamic programming cannot be simply generalized to multi-player dynamic games, then we give Theorem 11 that asserts the sufficient and necessary condition for D_π to converge to a perfect equilibrium value function.

Proposition 10. *Let* V_s^i *iterates by* $V_{s,k+1}^i = D(V_{s,k}^i) := D_{\pi_k}(V_{s,k}^i)$*, where* $\pi_{a,k}^{si}$ *satisfies* $D_{\pi_k}(V_{s,k}^i) = \hat{D}_{\pi_k}(V_{s,k}^i)$*. If* $V_{s,0}^i \in C_{\pi_0}$*, then the following statements satisfy* $(i) \Rightarrow (ii) \Rightarrow (iii) \Leftrightarrow (iv)$*.*

(i) $V_s^i \le Y_s^i \to D(V_s^i) \le D(Y_s^i)$ *for every* $V_s^i, Y_s^i \in O(\hat{V}_s^i)$*, where* $O(\hat{V}_s^i)$ *is a set such that* $V_{s,k}^i \in O(\hat{V}_s^i)$ *for all* k*.*

(ii) $V_{s,k}^i$ *converges to a perfect equilibrium value function by contraction mapping* D*, and* $V_{s,k}^i \in C_{\pi_k}$ *for all* k*.*

(iii) $V_{s,k}^i$ *monotonically converges to a perfect equilibrium value function.*

(iv) $V_{s,k+1}^i \le V_{s,k}^i \to D(V_{s,k+1}^i) \le D(V_{s,k}^i)$ *for all* k*.*

Using Theorem 8 (iii), it can be inferred that Proposition 10 (iv) implies $V_{s,k+1}^i \in \hat{C}_{\pi_k} \to V_{s,k+1}^i \in C_{\pi_{k+1}}$. It can be verified on the graph of policy cone that this implication formula generally does not hold, since π_k and π_{k+1} generally does not have a strong relation. Consequently, none of the four statements hold in dynamic games. In particular, neither does the first half of (ii) that operator D is a contraction mapping hold on its own, since previous research already shows that operator D fails to converge in dynamic games[19, 16]. However, as the single-player degeneration of operator D, the Bellman operator D in MDPs satisfies (i), which can be used to prove the first half of (ii) that Bellman operator D is a contraction mapping holds. Furthermore, if there is also $V_{s,0}^i \in C_{\pi_0}$ for Bellman operator

D, then all the four statements hold as Proposition 10 shows, where perfect equilibrium degenerates to optimal policy. In summary, Bellman operator D can achieve both monotone convergence and contraction mapping convergence in single-player case, but can achieve neither of the convergence in multi-player case. Proposition 10 is not necessary to construct our iterative approximation method in this paper, however, it provides a certain perspective why the Bellman operator and value iteration cannot be simply generalized to dynamic games.

Then we construct an iteration of dynamic programming operator D_π that converges to a perfect equilibrium value function. Note that not only is V_s^i bounded within C_π under the iteration of D_π for fixed π_a^{si}, but the apex $V_{\pi s}^i$ of C_π is also bounded with respect to the variable π_a^{si}. This means that if $V_{s,k}^i \in C_{\pi_k}$ for all $k \in \mathbb{N}$ and $V_{s,k+1}^i = D_{\pi_k}(V_{s,k}^i)$, then $\{V_{s,k}^i\}_{k\in\mathbb{N}}$ still satisfies the monotone convergence property. This allows us to construct the cone interior convergence conditions. First, we add a large enough $m_k^i \mathbf{1}_s$ to $V_{s,k}^i$ at every step to keep $V_{s,k}^i + m_k^i \mathbf{1}_s$ in the policy cone C_{π_k}, and let m_k^i converge to 0, so that monotone convergence property is maintained and $V_{s,k}^i$ converges to the apex of C_{π_k}. Second, the apex that $V_{s,k}^i$ converges to has to be in the best response cone $\hat{C}_{\pi_k}$ for the apex to be a perfect equilibrium value function, thus we further require $V_{s,k}^i + m_k^i \mathbf{1}_s$ be in the best repsonse cone $\hat{C}_{\pi_k}$. Summarizing these gives us Theorem 11.

Theorem 11 (Cone interior convergence conditions)**.** *Let V_s^i iterates by*

$$V_{s,k+1}^i = D_{\pi_k}(V_{s,k}^i + m_k^i \mathbf{1}_s), \tag{14}$$

where $\{\pi_{a,k}^{si}\}_{k\in\mathbb{N}}$ is a sequence of policies. Then there always exists a sequence $\{m_k^i\}_{k\in\mathbb{N}}$ sufficiently large such that $V_{s,k}^i + m_k^i \mathbf{1}_s \in \hat{C}_{\pi_k}$ for all $k \in \mathbb{N}$. Furthermore, for any such sequence $\{m_k^i\}_{k\in\mathbb{N}}$, $V_{s,k}^i$ converges to a perfect equilibrium value function if and only if $\lim_{k\to\infty} m_k^i = 0$.

Theorem 11 points out a dynamic programming method that iteratively converges to a perfect equilibrium sufficiently and necessarily, where m_k^i and $\pi_{a,k}^{si}$ are required in each iteration step. First, for simplification, we let $m_k^i \equiv m^i$, thus m^i only needs to be large enough so that $V_{s,k}^i + m^i \mathbf{1}_s \in \hat{C}_{\pi_k}$, and $V_{s,k}^i$ converges to the value function $V_{\pi s}^i$ plus a scaled $\mathbf{1}_s$ as stated in Theorem 8 (v). Second, we ensure that $\pi_{a,k}^{si}(x)$ converges to a Nash equilibrium for every state x. Then we've transformed the problem of approximating a perfect equilibrium to the problem of approximating a set of Nash equilibria.

5 FPTAS for perfect equilibria of dynamic games

5.1 Line search on the equilibrium bundle of dynamic games

In this section, our aim is to combine the two methods in the last two sections to obtain a hybrid iteration of dynamic programming and interior point method that approximates any perfect equilibrium of any dynamic game. First, we add the state index s back to the tensors in the static game case, and set the utility $U_A^{si} = u_A^{si} + \gamma T_{s'A}^s V_{s'}^i$. Then we have a family of equilibrium bundles $E_s(V_s^i)$ indexed by state $s \in \mathcal{S}$ that varies as V_s^i varies, as well as a family of canonical sections $\bar{\mu}_a^{si}(V_s^i, \pi_a^{si})$ indexed by state $s \in \mathcal{S}$ that varies as V_s^i varies, which satisfies

$$\mathbf{1}_a \bar{\mu}_a^{si}(V_s^i, \pi_a^{si}) = \hat{D}_\pi(V_s^i) - D_\pi(V_s^i), \tag{15}$$

$$\pi_a^{si} \text{ is a perfect equilibrium} \Leftrightarrow \bar{\mu}_a^{si}(V_{\pi s}^i, \pi_a^{si}) = 0 \Leftrightarrow \hat{D}_\pi(V_{\pi s}^i) = D_\pi(V_{\pi s}^i).$$

Thus, in our hybrid iteration, V_s^i iterates by dynamic programming (14), π_a^{si} iterates by projected gradient (6), and μ_a^{si} iterates by canonical section descent (9). However, note that projected gradient can update back onto the equilibrium bundle only after an infinitesimal deviation, and there is a problem that V_s^i moves more than an infinitesimal step in every iteration of the dynamic programming, except for the case where V_s^i is nearly converged to $V_{\pi s}^i$ plus a scaled $\mathbf{1}_s$. Thus second, we set $V_s^i = V_{\pi s}^i$ on the fiber over every π_a^{si}, which gives us the following definition of equilibrium bundle of dynamic games.

Definition 3 (Equilibrium bundle of dynamic games)**.** *An equilibrium bundle is the tuple $(E, \mathcal{P}, \alpha : E \to \mathcal{P})$ given by the following equations, where $U_{\pi A}^{si} = u_A^{si} + \gamma T_{s'A}^s V_{\pi s'}^i$, and $V_{\pi s}^i$ that depends on π_a^{si} is the unique solution of*

$V_s^i = D_\pi(V_s^i)$.

$$\begin{aligned} E &= \bigcup_{\pi_a^{si} \in \mathcal{P}} \{\pi_a^{si}\} \times B(\pi_a^{si}) \\ B(\pi_a^{si}) &= \left\{ v^{si} \circ \pi_a^{si} + \bar{\mu}_a^{si}(\pi_a^{si}) | v^{si} \geq 0 \right\} \\ \bar{\mu}_a^{si}(\pi_a^{si}) &= \pi_a^{si} \circ \left(\max_a^{si} \pi_{Aa}^{si-} U_{\pi A}^{si} - \pi_{Aa}^{si-} U_{\pi A}^{si} \right) \\ \alpha\left((\pi_a^{si}, \mu_a^{si})\right) &= \pi_a^{si} \end{aligned} \tag{1}$$

The map $\bar{\mu}_a^{si} : \mathcal{P} \to \{\mu_a^{si} | \min_a^{si} \mu_a^{si} = \mathbf{0}^{si}\}$ is called the canonical section of the equilibrium bundle.

The equilibrium bundle of dynamic games is still a fiber bundle, and its major difference from the equilibrium bundle of static games is that it is no longer an algebraic variety since the term $V_{\pi s}^i$ is not a polynomial with respect to π_a^{si}, and thus the equilibrium bundle of dynamic games does not satisfy the oddness theorem as that of static games. Perfect equilibria are still formalized as the zero points of its canonical section, and the hybrid iteration of dynamic programming and interior point method is still formalized as a line search on it.

Algorithm 1 A hybrid iteration of dynamic programming and interior point method

Require: Dynamic game $\Gamma = (N, \mathcal{S}, \mathcal{A}, T, u, \gamma)$.
(Optional) Sample in the policy space $\mathcal{P}$ and calculate canonical sections $\bar{\mu}_a^{si}(\pi_a^{si})$ to search for potential perfect equilibria globally.
Require: Initial policy π_a^{si}, either random or around a potential perfect equilibrium.
Set initial barrier parameter $\mu_a^{si} = \pi_a^{si}\mu'$ for sufficiently large μ', so that μ_a^{si} is around the fiber over π_a^{si} by Theorem 5 (i).
Set initial value function $V_s^i = m^i \mathbf{1}_s$ for sufficiently large m^i, so that $V_s^i \in \hat{C}_\pi$ by Proposition 7 (iii).
repeat
  repeat
    Set $U_A^{si} = u_A^{si} + \gamma T_{s'A}^s (V_{s'}^i + m^i \mathbf{1}_{s'})$ to apply dynamic programming (14).
    Calculate $\pi_{Aaa'}^{sij-} U_A^{si}$ and $\pi_{Aa}^{si-} U_A^{si}$, either model-based or model-free.
    Compute v^{si} by solving equation (5a) as stated in Theorem 3 (i).
    Calculate regret $r_a^{si} = v^{si} - \pi_{Aa}^{si-} U_A^{si}$.
    Calculate dual policy $\hat{\pi}_a^{si} = \mu_a^{si} / r_a^{si}$ and dual regret $\hat{r}_a^{si} = \mu_a^{si} / \pi_a^{si}$.
    Calculate projected gradient pg_a^{si} using equation (6).
    Calculate residual $dV_s^i := D_\pi(V_s^i + m^i \mathbf{1}_s) - V_s^i = v^{si} - \pi_a^{si} r_a^{si} - V_s^i$ of dynamic programming (14).
    Update π_a^{si} by pg_a^{si}, and update V_s^i by dV_s^i.
  until $\pi_a^{si} - \hat{\pi}_a^{si}$, $r_a^{si} - \hat{r}_a^{si}$, and $\mathrm{Angle}(dV_s^i, \mathbf{1}_s)$ all converge to 0, so that (π_a^{si}, μ_a^{si}) is on the equilibrium bundle.
  Compute differential $(d\pi_{a'}^{sj} / \pi_{a'}^{sj}) / (d\mu_{a''}^{sk} / \mu_{a''}^{sk})$ by solving equation (8).
  Calculate canonical section $\bar{\mu}_a^{si}(\pi_a^{si}) = \pi_a^{si} \circ (\max_a^{si} \pi_{Aa}^{si-} U_{\pi A}^{si} - \pi_{Aa}^{si-} U_{\pi A}^{si})$.
  Update μ_a^{si} using canonical section descent (9) to hop to another fiber in the neighborhood and move along the fiber to avoid potential singular points, and update π_a^{si} along differential $(d\pi_{a'}^{sj} / \pi_{a'}^{sj}) / (d\mu_{a''}^{sk} / \mu_{a''}^{sk})$.
until $\bar{\mu}_a^{si}(\pi_a^{si})$ converges to 0, so that the fiber over a perfect equilibrium is reached.

Algorithm 1 is still a **line search on the equilibrium bundle** that consists of two iteration levels, the **first iteration level** is to update onto the equilibrium bundle by alternating the steps of dynamic programming (14) and projected gradient descent (6), the **second iteration level** is to hop across the fibers of the equilibrium bundle by canonical section descent (9) and differential $(d\pi_{a'}^{sj} / \pi_{a'}^{sj}) / (d\mu_{a''}^{sk} / \mu_{a''}^{sk})$, and the line search leads to a perfect equilibrium as the canonical section $\bar{\mu}_a^{si}(\pi_a^{si})$ reduces to 0. Then we show that Algorithm 1 converges in polynomial time by showing the convergence rates of the three iteration formulas.

Proposition 12. *The following statements about convergence rate hold.*

(i) $\|V_s^i - D_\pi(V_s^i + m^i \mathbf{1}_s)\|_\infty$ *converge to* 0 *linearly with a rate of* γ *under the dynamic programming* (14) *for fixed* π_a^{si} *and* m.

(ii) $\|\bar{\mu}_a^i(\pi_a^i)\|_\infty$ *converge to* 0 *linearly for some* $\eta^i > 0$ *and* $\beta^i \geq 0$ *under the canonical section descent* (9).

(iii) $(\pi_a^i - \hat{\pi}_a^i)(r_a^i - \hat{r}_a^i)$ *converge to* 0 *sublinearly under the iteration of projected gradient* (6) *if the convergence point is non-singular.*

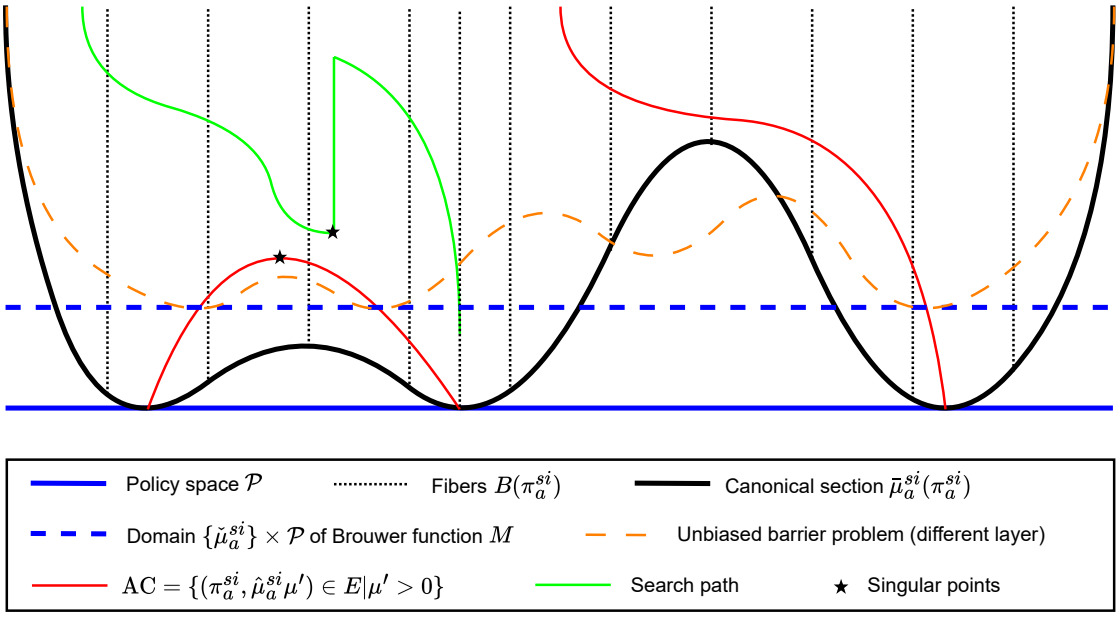

Fig. 4 Sketch graph of the equilibrium bundle. This sketch graph is based on the joint space $\mathcal{P} \times \{\mu_a^{si} | \mu_a^{si} \geq 0\}$ of policy and barrier parameter. First, equilibrium bundle E consists of the disjoint union of fibers $\{\pi_a^{si}\} \times B(\pi_a^{si})$ over each $\pi_a^{si} \in \mathcal{P}$, where each fiber $B(\pi_a^{si})$ is an affine subspace with the canonical section $\bar{\mu}_a^{si}(\pi_a^{si})$ being its least element, and perfect equilibria are zero points of the map $\bar{\mu}_a^{si}$. Second, Brouwer function $M(\mu_a^{si})$ rearranges the points in the offset policy space $\mathcal{P} \times \{\mu_a^{si}\}$, where the fixed points are the intersections between $\mathcal{P} \times \{\mu_a^{si}\}$ and the equilibrium bundle E, and unbiased barrier problem depicts the approximation to those fixed points. Third, when there is a single state in the state space, AC is an algebraic curve that satisfies the oddness theorem, such that exactly one of its endpoints is connected with the starting point as $\mu' \to \infty$, and the rest of the endpoints are connected in pairs. Finally, our method is a line search on the equilibrium bundle, which hops across the fibers to a zero point of the canonical section, and moves along a fiber to avoid singular points where the differential $d\pi_a^{si}/d\mu_{a'}^{sj}$ tend to infinity.

In Proposition 12, for a linearly converging iteration, the number of iterations needed to achieve given precision ϵ is $O(\log^m(1/\epsilon))$, and projected gradient descent is known to converge sublinearly and require $O((1/\epsilon)^n)$ iterations to achieve given precision ϵ. Note also that the elementary operations of the three iterations are all tensor operations whose complexity is $O(|N|^o|\mathcal{A}|^p|\mathcal{S}|^q)$. And Theorem 5 (iv) assures that for every dynamic game, there are always non-singular points available for the convergence so that Proposition 12 applies. Thus, Algorithm 1 converges in fully polynomial time to where the canonical section satisfies

$$\bar{\mu}_a^{si}(\pi_a^{si}) = \pi_a^{si} \circ \left(\max_a^{si} \pi_{Aa}^{si-} U_{\pi A}^{si} - \pi_{Aa}^{si-} U_{\pi A}^{si} \right) < \epsilon_a^{si}.$$

For any $(\pi_a^{si}, \mathbf{0}_a^{si}) \in E$, there are infinite many search paths on the equilibrium bundle that lead to it. Recall the two different approximations: weak approximation approximates to an ϵ-equilibrium, and strong approximation approximates to an ϵ-neighborhood of an actual equilibrium. It directly follows that on the equilibrium bundle, a strong approximation is a weak approximation, while the opposite is not true, such that an ϵ-equilibrium could be far from an actual equilibrium, depending on the nearby gradient of the canonical section, which aligns with existing results[26, 9]. For any perfect equilibrium of any dynamic game, our method achieves a weak approximation in fully polynomial time, and the time complexity for our method to achieve a strong approximation depends on the gradient of the canonical section near the actual equilibrium. Recall the complexity results that the weak approximation of Nash equilibria is PPAD-complete, and strong approximation of Nash equilibria with three or more players is FIXP-complete. This implies PPAD=FP.

5.2 Practical use

In Algorithm 1, the problem of computing $\pi_{Aa}^{si-} U_A^{si}$ and $\pi_{Aaa'}^{sij-} U_A^{si}$ is a variant of a problem called the expected utility problem, which computes $\pi_A^s U_A^{si}$. We use normal-form representation U_A^{si} to define games and deduce theorems, but note that in Algorithm 1, U_A^{si} only involves in the calculation through $\pi_{Aa}^{si-} U_A^{si}$ and $\pi_{Aaa'}^{sij-} U_A^{si}$. Thus, Algorithm 1

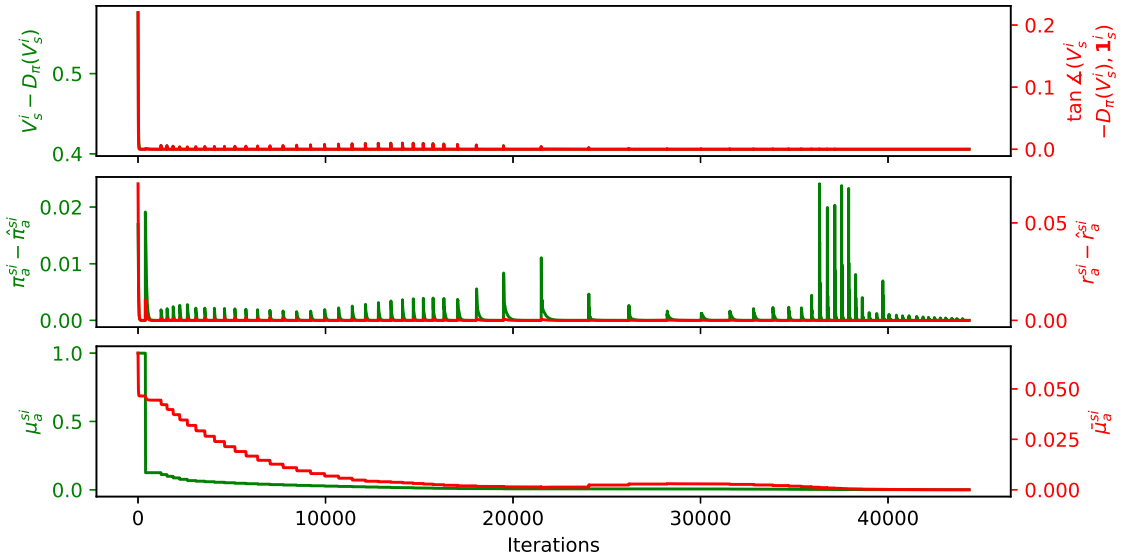

Fig. 5 Iteration curve. The figure shows the convergence of the three iterations in Proposition 12. $\mathrm{Angle}(V_s^i - D_\pi(V_s^i), \mathbf{1}_s^i)$, $(\pi_a^{si} - \hat{\pi}_a^{si}, r_a^{si} - \hat{r}_a^{si})$, and $\bar{\mu}_a^i(\pi_a^i)$ all converging to 0 indicates that the convergence point is a perfect equilibrium. In particular, $\mathrm{Angle}(V_s^i - D_\pi(V_s^i), \mathbf{1}_s^i)$ and $(\pi_a^{si} - \hat{\pi}_a^{si}, r_a^{si} - \hat{r}_a^{si})$ staying converged during the whole iteration indicates that the iteration is a line search on the equilibrium bundle.

actually works with any game representation, as long as the expected utility problem has an polynomial-time algorithm in that representation, such as those important classes of succint games like graphical games, action-graph games, and many others.

Algorithm 1 even works in model-free cases, where $\pi_{Aa}^{si-} U_A^{si}$ and $\pi_{Aaa'}^{sij-} U_A^{si}$ are estimated using sampled data collected as an self-play agent interacts with a game instance. The canonical section $\bar{\mu}_a^{si}$ can be used to guide the search of the policy space, and the term $m_k^i \mathbf{1}_s$ in dynamic programming can be used to balance exploration and exploitation of the state space. Thus, it is possible to construct a model-free MARL method that would be free from non-stationarity and curse of multiagency.

Algorithm 1 takes any dynamic game as input to output its perfect equilibrium, such as it takes a static game as a single-state dynamic game to produce its Nash equilibrium, and it takes an MDP as a single-player dynamic game to produce its optimal policy. We implement the algorithm to take any size of dynamic game as input, and we animate the line search process for dynamic games with 2 players, 2 states, and 2 actions using Fig. 3, Fig. 1, Fig. 2, and the iteration curve in Fig. 5. And we tested our method on 2000 randomly generated dynamic games of 3 players, 3 states, and 3 actions in experiment, and the iteration converges to a perfect equilibrium on every single case.

5.3 Tractability of PPAD

PPAD is defined as the complexity class of all the problems that reduces to *End-Of-The-Line* in polynomial time, and it is believed to contain hard problems because *End-Of-The-Line* is seemingly intractable in polynomial time, yet our discovery implies PPAD=FP. Thus, it is necessary that we explain how the reduction of our method can solve *End-Of-The-Line* in polynomial time.

We've introduced that *End-Of-The-Line* is believed to be intractable because it seems to let us follow a potentially exponentially long chain in directed graph DG from a given source to a sink, only by using a polynomial-time computable function f that outputs the predecessor and successor of an input vertex. However, note that it is not necessary to follow the exponentially long chain or use function f to jump over vertices one by one. The reduction of our method to *End-Of-The-Line* consists of two independent steps: computing a static game whose ϵ-approximate Nash equilibria correspond to unbalanced vertices of DG in polynomial time, and computing an ϵ-approximate Nash equilibrium in fully polynomial time. The first step is given by the reduction from *End-Of-The-Line* to *Nash*[7], and the second step is our method.

The reduction of *End-Of-The-Line* to *Nash* consists of two steps that can be done in polynomial time. First, construct a Brouwer function F on the unit cube that is given by an arithmetic circuit that consists only of addition, multiplication

and comparison, with the boolean circuit f being its subcircuit, such that every vertex of DG is encoded in the unit cube, and the unbalanced vertices except the given one are exactly the ϵ-approximate fixed points. Second, simulate every operation in arithmetic circuit F with a two-action static game to obtain a many-player two-action graphical game, and then simulate the graphical game with a three-player many-action static game, such that the weights of certain actions on ϵ'-approximate Nash equilibria are exactly the ϵ-approximate fixed points.

After the static game is constructed, our method is used to compute an ϵ-approximate Nash equilibria, and then certain components of the policy can be decoded to obtain an unbalanced vertex. Our method is a line search on the equilibrium bundle, the search path lies in the total space, and DG is encoded in the base space, where the dimension of the total space is twice that of the base space. The line search originates from the fiber over the point encoding the given unbalanced vertex, and eventually reaches the fiber over an ϵ-approximate Nash equilibria encoding another unbalanced vertex.

In summary, there are two facts that make *End-Of-The-Line* solvable in polynomial time. First, the polynomial-time computable function f is not used to jump over vertices one by one, instead, it is used to construct a static game in polynomial time, such that the directed graph is encoded in its policy space and the unbalanced vertices are encoded as Nash equilibria. Second, the search path does not follow the exponentially long chain, instead, it is completely another path that lies in a space with twice the dimension of the policy space where the chain is encoded.

6 Conclusion

In this paper, we aim to deal with fully observable dynamic games to find a polynomial-time algorithm to approximate perfect equilibria. First, we introduce the unbiased barrier problem and unbiased KKT conditions to make the interior point method to approximate Nash equilibria of static games. Second, we introduce the policy cone to give the sufficient and necessary condition for dynamic programming to converge to perfect equilibria of dynamic games. Finally, combining the two sections of results, we introduce the equilibrium bundle, such that it formalize the perfect equilibria as the zero points of its canonical section, and it formalize a hybrid iteration of dynamic programming and interior point method as a line search on it. The geometric properties of the equilibrium bundle allows us to give the existence and oddness theorems of the equilibrium bundle as an extension of those of Nash equilibria. In addition, the equilibrium bundle, unbiased barrier problem, unbiased KKT conditions, and Brouwer function all lie in the joint space of policy and barrier parameter with certain geometric structure. The hybrid iteration approximate any perfect equilibrium of any dynamic game, it achieves a weak approximation in fully polynomial time, the time complexity for it to achieve a strong approximation depends on the nearby gradient of the canonical section. This makes the method an FPTAS for the PPAD-complete weak approximation of game equilibria, implying PPAD=FP. In experiments, the line search process is animated, and the method is tested on 2000 randomly generated dynamic games where it converges to a perfect equilibrium in every single case.

Acknowledgments

Funding: This work was supported by the National Key R&D Program of China (2022YFB4701400/4701402), National Natural Science Foundation of China (No. U21B6002, 62203260, 92248304), Guangdong Basic and Applied Basic Research Foundation (2023A1515011773). **Author Contributions:** H.S. developed the theorems, implemented the method, plotted experimental results, and wrote the manuscript. X.W. supervised the research. X.W. and C.X. assisted the research with constructive discussions. C.X., J.T., and B.Y. assisted with manuscript editing. All authors read and commented the paper. **Competing interests:** The authors declare no competing interests. **Code availability:** The codes implementing our method and animating line search process that takes any dynamic game instance as input are available at `https://github.com/shb20tsinghua/PTAS_Game/tree/main`.

References

[1] Richard Bellman. “On the theory of dynamic programming”. In: *Proceedings of the National Academy of Sciences* 38.8 (1952), pp. 716–719 (cit. on p. 6).

[2] George W Brown. “Iterative solution of games by fictitious play”. In: *Activity Analysis of Production and Allocation* 13.1 (1951), p. 374 (cit. on pp. 2, 5, 8).

[3] Richard H Byrd, Mary E Hribar, and Jorge Nocedal. “An interior point algorithm for large-scale nonlinear programming”. In: *SIAM Journal on Optimization* 9.4 (1999), pp. 877–900 (cit. on p. 5).

[4] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. "Computing Nash equilibria: Approximation and smoothed complexity". In: *2006 47th Annual IEEE Symposium on Foundations of Computer Science*. Los Alamitos, CA, USA: IEEE Computer Society, 2006, pp. 603–612 (cit. on p. 2).

[5] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. "Settling the complexity of computing two-player Nash equilibria". In: *Journal of the ACM* 56.3 (2009), pp. 1–57 (cit. on p. 2).

[6] Richard W. Cottle and George B. Dantzig. "Complementary pivot theory of mathematical programming". In: *Linear Algebra and its Applications* 1.1 (1968), pp. 103–125 (cit. on p. 8).

[7] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. "The complexity of computing a Nash equilibrium". In: *Communications of the ACM* 52.2 (2009), pp. 89–97 (cit. on pp. 2, 19).

[8] Albert Einstein et al. "The foundation of the general theory of relativity". In: *Annalen Phys* 49.7 (1916), pp. 769–822 (cit. on p. 3).

[9] Kousha Etessami and Mihalis Yannakakis. "On the complexity of Nash equilibria and other fixed points". In: *SIAM Journal on Computing* 39.6 (2010), pp. 2531–2597 (cit. on pp. 2, 18).

[10] Sven Gronauer and Klaus Diepold. "Multi-agent deep reinforcement learning: a survey". In: *Artificial Intelligence Review* 55.2 (2022), pp. 895–943 (cit. on p. 2).

[11] Charles R. Harris et al. "Array programming with NumPy". In: *Nature* 585.7825 (2020), pp. 357–362 (cit. on p. 3).

[12] John C Harsanyi. "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points". In: *International Journal of Game Theory* 2.1 (1973), pp. 1–23 (cit. on pp. 5, 8).

[13] John C Harsanyi. "Oddness of the number of equilibrium points: a new proof". In: *International Journal of Game Theory* 2.1 (1973), pp. 235–250 (cit. on pp. 2, 4, 6, 13, 24).

[14] Sergiu Hart and Andreu Mas-Colell. "A simple adaptive procedure leading to correlated equilibrium". In: *Econometrica* 68.5 (2000), pp. 1127–1150 (cit. on pp. 2, 5, 8).

[15] Johannes Heinrich, Marc Lanctot, and David Silver. "Fictitious Self-Play in Extensive-Form Games". In: *Proceedings of the 32nd International Conference on Machine Learning*. Ed. by Francis Bach and David Blei. Vol. 37. Lille, France: PMLR, 2015, pp. 805–813 (cit. on pp. 2, 7, 8).

[16] Junling Hu and Michael P Wellman. "Nash Q-learning for general-sum stochastic games". In: *Journal of Machine Learning Research* 4 (2003), pp. 1039–1069 (cit. on pp. 6, 15).

[17] Harold W Kuhn and Albert W Tucker. "Nonlinear Programming". In: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. Ed. by Jerzy Neyman. Vol. 2. Berkeley and Los Angeles: University of California Press, 1951, pp. 481–492 (cit. on p. 8).

[18] Carlton E Lemke and Joseph T Howson Jr. "Equilibrium points of bimatrix games". In: *Journal of the Society for Industrial and Applied Mathematics* 12.2 (1964), pp. 413–423 (cit. on p. 8).

[19] Michael L Littman. "Value-function reinforcement learning in Markov games". In: *Cognitive Systems Research* 2.1 (2001), pp. 55–66 (cit. on pp. 6, 15).

[20] Michael L. Littman. "Markov Games as a Framework for Multi-Agent Reinforcement Learning". In: *Machine Learning Proceedings 1994*. Ed. by William W. Cohen and Haym Hirsh. San Francisco, CA, USA: Morgan Kaufmann, 1994, pp. 157–163 (cit. on p. 6).

[21] Eric Maskin and Jean Tirole. "Markov perfect equilibrium: I. Observable actions". In: *Journal of Economic Theory* 100.2 (2001), pp. 191–219 (cit. on p. 3).

[22] John F Nash Jr. "Equilibrium points in n-person games". In: *Proceedings of the National Academy of Sciences* 36.1 (1950), pp. 48–49 (cit. on p. 2).

[23] John F Nash Jr. "Non-cooperative games". In: *Annals of Mathematics* 54.2 (1951), pp. 286–295 (cit. on pp. 2, 4, 6, 11).

[24] Christos H Papadimitriou. "On the complexity of the parity argument and other inefficient proofs of existence". In: *Journal of Computer and system Sciences* 48.3 (1994), pp. 498–532 (cit. on p. 2).

[25] Aviad Rubinstein. "Inapproximability of Nash Equilibrium". In: *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*. New York, NY, USA: Association for Computing Machinery, 2015, pp. 409–418 (cit. on p. 2).

[26] Herbert Scarf. "The approximation of fixed points of a continuous mapping". In: *SIAM Journal on Applied Mathematics* 15.5 (1967), pp. 1328–1343 (cit. on pp. 2, 18).

[27] Lloyd S Shapley. "Stochastic games". In: *Proceedings of the National Academy of Sciences* 39.10 (1953), pp. 1095–1100 (cit. on p. 3).

[28] Martin Zinkevich et al. "Regret Minimization in Games with Incomplete Information". In: *Advances in Neural Information Processing Systems*. Ed. by J. Platt et al. Vol. 20. Curran Associates, Inc., 2007 (cit. on pp. 2, 7, 8).

Proofs

In this section, we provide the proofs of all the theorems and propositions in the previous sections.

Proofs in section 3

Proof of Theorem 1. First, use (i) to prove (ii) and the optimal objective function value is 0. It follows from $(\pi_a^i, r_a^i) \geq 0$ that the objective $\pi_a^i r_a^i \geq 0$. When (π_a^i, v^i) is a Nash equilibrium, $v^i = \pi_A U_A^i = \max_a^i \pi_{Aa}^{i-} U_A^i$. Then

$$\pi_a^i r_a^i = \pi_a^i \left(v^i - \pi_{Aa}^{i-} U_A^i\right) = v^i - \pi_A U_A^i = 0$$

and $r_a^i = v^i - \pi_{Aa}^{i-} U_A^i \geq 0$. Hence (π_a^i, r_a^i, v^i) is an optimal point and the optimal objective function value is 0.

Then use (ii) to prove (iii). By the guaranteed existence of Nash equilibria and the above inference, the optimal objective function value is always 0. It follows from $\pi_a^i r_a^i = 0$ and $(\pi_a^i, r_a^i) \geq 0$ that

$$\pi_a^i \circ r_a^i = 0.$$

Hence let $(\bar{\lambda}_a^i, \tilde{\lambda}^i, \hat{\pi}_a^i, \hat{r}_a^i) = (\mathbf{0}_a^i, \mathbf{0}^i, \pi_a^i, r_a^i)$, and then the equations are satisfied.

Finally, use (iii) to prove primal-dual bias $(\bar{\lambda}_a^i, \tilde{\lambda}^i, \pi_a^i - \hat{\pi}_a^i, r_a^i - \hat{r}_a^i) = 0$ and (i). Substituting $\pi_a^i = \hat{\pi}_a^i$ into $\bar{\lambda}_a^i + \pi_a^i - \hat{\pi}_a^i = 0$, we have $\bar{\lambda}_a^i = 0$. Substituting $\bar{\lambda}_a^i = 0$ into $\bar{\lambda}_a^j \pi_{Aaa'}^{ij-} U_A^i + \tilde{\lambda}^i \mathbf{1}_{a'} + r_{a'}^i - \hat{r}_{a'}^i = 0$, we have $r_a^i - \hat{r}_a^i = -\tilde{\lambda}^i \mathbf{1}_a$. Substituting $\pi_a^i = \hat{\pi}_a^i$ into $r_a^i \circ \hat{\pi}_a^i = 0$ and $\pi_a^i \circ \hat{r}_a^i = 0$, we have $\pi_a^i \circ (r_a^i - \hat{r}_a^i) = 0$. Then

$$\pi_a^i \circ \left(-\tilde{\lambda}^i \mathbf{1}_a\right) = 0.$$

Because $\mathbf{1}_a \pi_a^i = \mathbf{1}^i$, so for every index i there must exist index a such that $\pi_a^i > 0$. It follows that $\tilde{\lambda}^i = 0$ for every index i, and then $r_a^i = \hat{r}_a^i$. Hence we obtain primal-dual bias $(\bar{\lambda}_a^i, \tilde{\lambda}^i, \pi_a^i - \hat{\pi}_a^i, r_a^i - \hat{r}_a^i) = 0$.

At this time, $\pi_a^i \circ r_a^i = 0$. Considering that for every index i there must exist index a such that $\pi_a^i > 0$, then for every index i there must exist index a such that $r_a^i = v^i - \pi_{Aa}^{i-} U_A^i = 0$. Note also that $r_a^i \geq 0$, and thus

$$v^i = \max_a^i \pi_{Aa}^{i-} U_A^i.$$

Then it follows from the objective $\pi_a^i r_a^i = 0$ that $v^i = \max_a^i \pi_{Aa}^{i-} U_A^i = \pi_A U_A^i$. Hence (π_a^i, v^i) is a Nash equilibrium. □

Proof of Theorem 2. First, $(i) \Leftrightarrow (iii)$ is implied by the definition of Brouwer function M.

Then, prove $(ii) \Leftrightarrow (iii)$. In unbiased barrier problem (4), where $\hat{\pi}_a^i = \mu_a^i / r_a^i$ and $\hat{r}_a^i = \mu_a^i / \pi_a^i$,

$$\left(\pi_a^i - \hat{\pi}_a^i\right)\left(r_a^i - \hat{r}_a^i\right) = \sum_{i,a} \left(\pi_a^i \circ r_a^i + \frac{\mu_a^{i2}}{\pi_a^i \circ r_a^i} - 2\mu_a^i\right).$$

The formula takes the minimum value if and only if $\pi_a^i \circ r_a^i = \mu_a^i$. Considering the constraints $r_a^i - v^i + \pi_{Aa}^{i-} U_A^i = 0$ and $\mathbf{1}_a \hat{\pi}_a^i - \mathbf{1}^i = 0$, the simultaneous equations are exactly unbiased KKT conditions (5). Hence we obtain the equivalence.

Then, prove $(iii) \Leftrightarrow ((iv) \wedge \pi_a^i = \hat{\pi}_a^i)$. Unbiased KKT conditions (5) are exactly the simultaneous equations of perturbed KKT conditions (3) and unbiased condition $\pi_a^i = \hat{\pi}_a^i$. Hence we obtain the equivalence.

Then, prove $(iv) \Leftrightarrow (v) \Leftrightarrow (vi)$. It can be verified that KKT conditions of barrier problem (2) are perturbed KKT conditions (3), and simultaneous equations of KKT conditions and parameter $\hat{\pi}_a^i = \mu_a^i / r_a^i$ and $\hat{r}_a^i = \mu_a^i / \pi_a^i$ of unbiased barrier problem (4) are also perturbed KKT conditions (3). Note that unbiased barrier problem (4) and barrier problem (2) are both equality constrained optimization problems, and thus by the Lagrange multiplier method, their local extreme points are the points that satisfies their KKT conditions, that is, perturbed KKT conditions (3). Hence we obtain the equivalence.

Finally, $(\pi_a^i, \mathbf{0}_a^i)$ is a solution of unbiased KKT conditions (5) is equivalent to Theorem 1 (iii), and thus it is equivalent to π_a^i being a Nash equilibrium. □

Proof of Theorem 3. (i) For every index i, denote the function and its derivative

$$f(v^i) = \mathbf{1}_a \frac{\mu_a^i}{v^i - \pi_{Aa}^{i-} U_A^i} - \mathbf{1}^i, \frac{df(v^i)}{dv^i} = -\mathbf{1}_a \frac{\mu_a^i}{\left(v^i - \pi_{Aa}^{i-} U_A^i\right)^2} \leq 0.$$

There is also

$$\lim_{v^i \to \left(\max_a^i \pi_{Aa}^{i-} U_A^i\right)^+} f(v^i) = +\infty \quad \wedge \quad \lim_{v^i \to +\infty} f(v^i) = -1.$$

Thus, $f(v^i)$ monotonically decreases with respect to v^i from $+\infty$ to -1 in its domain, and hence there exists an unique v^i that satisfies $f(v^i) = 0$.

(ii) By the objective function and constraint $r_a^i - v^i + \pi_{Aa}^{i-} U_A^i = 0$, considering $\hat{\pi}_a^i$ and $\hat{r}_a^i$ are constant parameters, the differential of objective function is

$$\left(\pi_a^i - \hat{\pi}_a^i\right) dr_a^i + \left(r_a^i - \hat{r}_a^i\right) d\pi_a^i = \left(\pi_a^i - \hat{\pi}_a^i\right) \left(dv^i - \pi_{Aaa'}^{ij-} U_A^i d\pi_{a'}^j\right) + \left(r_a^i - \hat{r}_a^i\right) d\pi_a^i.$$

It follows from $\mathbf{1}_a \hat{\pi}_a^i - \mathbf{1}^i = 0$ that $(\pi_a^i - \hat{\pi}_a^i) dv^i = 0$, and thus the differential is

$$\left(\left(r_{a'}^j - \hat{r}_{a'}^j\right) - \left(\pi_a^i - \hat{\pi}_a^i\right) \pi_{Aaa'}^{ij-} U_A^i\right) d\pi_{a'}^j.$$

Finally, project the gradient regarding the constraint $\mathbf{1}_a \pi_a^i - \mathbf{1}^i = 0$, and we obtain the projected gradient

$$\mathrm{pg}_{a''}^j = \left(I_{a'a''} - \frac{\mathbf{1}_{a'} \mathbf{1}_{a''}}{|\mathcal{A}|}\right) \left(\left(r_{a'}^j - \hat{r}_{a'}^j\right) - \left(\pi_a^i - \hat{\pi}_a^i\right) \pi_{Aaa'}^{ij-} U_A^i\right).$$

(iii) (i) implies that Brouwer function $\hat{\pi}_a^i = M(\mu_a^i)(\pi_a^i)$ is indeed a map, such that for every π_a^i there is a unique $\hat{\pi}_a^i$. Note also that $\hat{\pi}_a^i = M(\mu_a^i)(\pi_a^i)$ is continuous, since unbiased KKT conditions (5) is continuous. Thus, by Brouwer's fixed point theorem, there is always a fix point that satisfies $\hat{\pi}_a^i = \pi_a^i$, that is, a point that satisfies equation (5a). □

Proof of Theorem 4. (i), (ii), (iii) It follows directly from definition of the equilibrium bundle and unbiased KKT conditions (5).

(iv) By Theorem 3 (iii), for every $\mu_a^i \geq 0$, there is at least one solution π_a^i of (5), and thus the equality follows.

(v) By Theorem 2, π_a^i is a Nash equilibrium if and only if $\mathbf{0}_a^i \in B(\pi_a^i)$, then the equivalence is directly implied.

□

Proof of Theorem 5. (i) From unbiased KKT conditions (5) we have

$$\pi_a^i \circ v^i - \pi_a^i \circ \pi_{Aa}^{i-} U_A^i = \hat{\mu}_a^i \mu'$$

and $\mathbf{1}_a \pi_a^i = \mathbf{1}^i$. As $\mu' \to +\infty$, π_a^i and $\pi_a^i \circ \pi_{Aa}^{i-} U_A^i$ is bounded, and $v^i \to +\infty$. Then

$$\lim_{\mu' \to +\infty} \pi_a^i \circ \frac{v^i}{\mu'} = \hat{\mu}_a^i.$$

Hence we obtain $\lim_{\mu' \to +\infty} \pi_a^i = \hat{\mu}_a^i / (\mathbf{1}_a \hat{\mu}_a^i)$.

(ii) Differential of unbiased KKT conditions (5) is

$$\begin{bmatrix} \pi_a^i \circ dr_a^i + d\pi_a^i \circ r_a^i - d\mu_a^i \\ dr_a^i - dv^i + \pi_{Aaa'}^{ij-} U_A^i d\pi_{a'}^j \\ \mathbf{1}_a d\pi_a^i \end{bmatrix} = 0.$$

Eliminating dr_a^i, we have

$$\begin{bmatrix} \pi_a^i \circ dv^i - \left(\pi_a^i \circ \pi_{Aaa'}^{ij-} U_A^i \circ \pi_{a'}^j\right) \frac{d\pi_{a'}^j}{\pi_{a'}^j} + \mu_a^i \circ \frac{d\pi_a^i}{\pi_a^i} \\ \mathbf{1}_a (\pi_a^i \circ \frac{d\pi_a^i}{\pi_a^i}) \end{bmatrix} = \begin{bmatrix} d\mu_a^i \\ 0 \end{bmatrix}.$$

Transform it into a linear equation system with respect to $(d\pi_{a'}^j / \pi_{a'}^j) / (d\mu_{a''}^k / \mu_{a''}^k)$ and $dv^l / (d\mu_{a''}^k / \mu_{a''}^k)$, and we obtain equation (8).

(iii) $(\pi_a^i, \mu_a^i) \in E$ is a solution of unbiased KKT conditions (5), then by Theorem 2 (ii) or (v), (π_a^i, μ_a^i) is an extreme point of unbiased barrier problem (4). When $\mathrm{C}^{(i,a,m)}_{(j,a',l)}$ is non-singular, the Jacobian matrix of unbiased KKT conditions (5) is non-singular, and then by the implicit function theorem, (π_a^i, μ_a^i) is the only solution of (5) in its neighborhood given μ_a^i. Thus, given μ_a^i, (π_a^i, μ_a^i) is the only extreme point of (4), namely, unbiased barrier problem (4) is locally strictly convex at (π_a^i, μ_a^i).

(iv) On the fiber $B(\pi_a^i)$ over π_a^i, μ_a^i/π_a^i tends to $v^i \mathbf{1}_a$ for some v^i as $\mu_a^i \to \infty$, and $\pi^{ij-}_{Aaa'} U_A^i \circ \pi_{a'}^j$ is bounded. Thus, matrix $\mathrm{C}^{(i,a,m)}_{(j,a',l)}$ tends to

$$\begin{bmatrix} \mathrm{Diag}(v^i \mathbf{1}_a) & \left(I^{il}\mathbf{1}_a\right)^{(i,a)}_{l} \\ \pi_{(j,a')} \circ \left(I^{jm}\mathbf{1}_{a'}\right)^{m}_{(j,a')} & \mathbf{0}^m_l \end{bmatrix}.$$

By elementary column transformation, the determinant of this matrix is $(-1)^{|N|}\prod_{i\in N}(v^i)^{|\mathcal{A}|-1}$. It follows that the determinant of $\mathrm{C}^{(i,a,m)}_{(j,a',l)}$ diverges as $\mu_a^i \to \infty$ in $B(\pi_a^i)$. Thus, there always exists $\breve{\mu}_a^i \in B(\pi_a^i)$ such that for every $\mu_a^i > \breve{\mu}_a^i$, the determinant of $\mathrm{C}^{(i,a,m)}_{(j,a',l)}$ is non-zero on (π_a^i, μ_a^i).

□

Proof of Theorem 6. For any $\hat{\mu}_a^i > 0$ and $\mu'' \geq 0$, there is an algebraic curve $\mathrm{AC} = \{(\pi_a^i, \hat{\mu}_a^i\mu') \in E | \mu' > \mu''\}$. Denote EP as the set points on AC as $\mu' \to \mu''$. By the implicit function theorem and PUISEUX's theorem, it can be shown that at every point on an algebraic curve, either non-singular or singular, there is an unique analytical continuation beyond that point[13]. It follows that the endpoints of an algebraic curve are always connected in pairs. And Theorem 5 (i) shows that there is an unique endpoint of algebraic curve AC as $\mu' \to \infty$, which is called the starting point. Thus, there is exactly one point in EP that is connected to the starting point of AC by a branch of AC, and all the other points in EP are connected in pairs by the rest branches of AC. In other words, there are always an odd number of points in EP.

Then we show that EP almost always equals to $\{(\pi_a^i, \hat{\mu}_a^i\mu'') \in E\}$. For a point (π_a^i, μ_a^i) where $\mathrm{C}^{(i,a,m)}_{(j,a',l)}$ is non-singular, π_a^i is the unique solution in its neighborhood for the given μ_a^i by the implicit function theorem. It follows that if $(\pi_a^i, \hat{\mu}_a^i\mu'') \in \mathrm{EP}$ is a non-singular point, then for every $\hat{\mu}_a^i > 0$, the algebraic curve $\{(\pi_a^i, \hat{\mu}_a^i\mu' + \hat{\mu}_a^i\mu'') \in E | \mu' > 0\}$ tends to $(\pi_a^i, \hat{\mu}_a^i\mu'')$ as $\mu' \to 0$. In other words, $\{(\pi_a^i, \hat{\mu}_a^i\mu'') \in E\}$ is contained in EP if every point in EP is non-singular, and thus they are equal if so. Note that almost all points on an algebraic variety are non-singular. Thus, EP almost always equals to $\{(\pi_a^i, \hat{\mu}_a^i\mu'') \in E\}$.

□

Proofs in section 4

Lemma 13. *Let $T_\pi \in \mathbb{R}^{n\times n}$ be a matrix that satisfies $T_\pi \geq 0$ and $T_\pi \mathbf{1} = \mathbf{1}$, and let $\gamma \in [0, 1)$. Then*

(i) *$(I - \gamma T_\pi)$ is invertible.*

(ii) *For any $X \in \mathbb{R}^n$, the formula $(I - \gamma T_\pi)X \geq 0 \to X \geq 0 \to (\gamma T_\pi)X \geq 0$ holds.*

Proof. (i) The eigenvalues of $(I - \gamma T_\pi)$ is given by $1 - \gamma\lambda_i$, where λ_i are the eigenvalues of T_π. Note that inequality $T_\pi X \leq \max(X)$ holds, and thus for real eigenvalue λ_i that $T_\pi X = \lambda_i X$, we have $\lambda_i \leq 1$, and then $1 - \gamma\lambda_i > 0$. Hence $(I - \gamma T_\pi)$ has no zero eigenvalues, and it's invertible.

(ii) From $T_\pi \geq 0$ and $\gamma \in [0, 1)$, we obtain $X \geq 0 \to (\gamma T_\pi)X \geq 0$. Note that inequality $T_\pi X \geq \min(X)$ holds, and it follows that

$$\min((I - \gamma T_\pi)X) \leq \min(X) - \gamma \min(T_\pi X) \leq (1 - \gamma)\min(X).$$

Hence we obtain $(I - \gamma T_\pi)X \geq 0 \to X \geq 0$. □

Proof of Proposition 7. (i) By Lemma 13 (i), for any policy π_a^{si}, there is an unique value function

$$V_{\pi s}^i = (I_{s'}^s - \gamma T_{\pi s'}^s)^{-1} u_\pi^{s'i}$$

that satisfies $V_{\pi s}^i = \pi_A^s(u_A^{si} + \gamma T_{s'A}^s V_{\pi s'}^i)$, where $T_{\pi s'}^s := \pi_A^s T_{s'A}^s$ and $u_\pi^{si} := \pi_A^s u_A^{si}$. Then by definition of C_π, we obtain $V_{\pi s}^i \in C_\pi$. It follows from $V_s^i \in C_\pi$ that

$$(I_{s'}^s - \gamma T_{\pi s'}^s)\left(V_s^i - V_{\pi s}^i\right) = V_s^i - \pi_A^s\left(u_A^{si} + \gamma T_{s'A}^s V_{s'}^i\right) \geq 0.$$

Then by Lemma 13 (ii), we have $V_s^i - V_{\pi s}^i \geq 0$, and thus for all $V_s^i \in C_\pi$, $V_s^i \geq V_{\pi s}^i$.

(ii) Considering $V_s^i - Y_{xs}^i = d_x^i \mathbf{1}_s$, there is

$$\left(Y_{xs}^i - D_\pi(Y_{xs}^i)\right)(x) = \left(V_s^i - D_\pi(V_s^i) - (1-\gamma) d_x^i \mathbf{1}_s\right)(x).$$

Then let $d_x^i = (V_s^i - D_\pi(V_s^i))/(1-\gamma)$, and thus formula (11a) is satisfied for every $x \in \mathcal{S}$. Note that d_x^i is unique, and hence Y_{xs}^i is unique.

Similarly, considering $V_s^i - \hat{Y}_{xs}^i = \hat{d}_x^i \mathbf{1}_s$, there is

$$\left(\hat{Y}_{xs}^i - \hat{D}_\pi(\hat{Y}_{xs}^i)\right)(x) = \left(V_s^i - \hat{D}_\pi(V_s^i) - (1-\gamma)\hat{d}_x^i \mathbf{1}_s\right)(x).$$

then let $\hat{d}_x^i = (V_s^i - \hat{D}_\pi(V_s^i))/(1-\gamma)$, and thus formula (11b) is satisfied for every $x \in \mathcal{S}$. Note that $\hat{d}_x^i$ is unique, and hence $\hat{Y}_{xs}^i$ is unique.

(iii) For any $V_s^i \in \mathcal{V}$, we have

$$V_s^i + m^i \mathbf{1}_s - \pi_{Aa}^{si-}\left(u_A^{si} + \gamma T_{s'A}^s \left(V_{s'}^i + m^i \mathbf{1}_{s'}\right)\right) = V_s^i - \pi_{Aa}^{si-}\left(u_A^{si} + \gamma T_{s'A}^s V_{s'}^i\right) + (1-\gamma) m^i \mathbf{1}_s,$$

where $V_s^i - \pi_{Aa}^{si-}(u_A^{si} + \gamma T_{s'A}^s V_{s'}^i)$ is constant. Then there exists an $M^i > 0$ such that for any $m^i > M^i$ the above formula is greater than 0, and hence $V_s^i + m^i \mathbf{1}_s \in \hat{C}_\pi$.

By definition, for any $V_s^i \in \hat{C}_\pi$, we have

$$V_s^i \geq \pi_{Aa}^{si-}\left(u_A^{si} + \gamma T_{s'A}^s V_{s'}^i\right) \geq \pi_A^s \left(u_A^{si} + \gamma T_{s'A}^s V_{s'}^i\right).$$

Then $V_s^i \in C_\pi$, and hence $\hat{C}_\pi \subseteq C_\pi$. □

Proof of Theorem 8. (i) It follows directly from definition of C_π.

(ii) Note that equation

$$D_\pi(V_s^i) - D_\pi\left(D_\pi(V_s^i)\right) = (I_{s'}^s - \gamma T_{\pi s'}^s)\left(V_s^i - V_{\pi s}^i - \left(V_s^i - D_\pi(V_s^i)\right)\right) = \gamma T_{\pi s'}^s \left(V_s^i - D_\pi(V_s^i)\right)$$

holds. Then if $V_s^i \in C_\pi$, by Lemma 13 (ii), we obtain $D_\pi(V_s^i) \in C_\pi$.

(iii) Note that equation

$$D_\pi(V_s^i) - \hat{D}_\pi\left(D_\pi(V_s^i)\right) = \gamma \tilde{\pi}_a^{si} \pi_{Aa}^{si-} T_{s'A}^s \left(V_s^i - D_\pi(V_s^i)\right) + D_\pi(V_s^i) - \hat{D}_\pi(V_s^i)$$

holds. By Lemma 13 (ii), if $D_\pi(V_s^i) = \hat{D}_\pi(V_s^i)$, then for any $V_s^i \in C_\pi$, $D_\pi(V_s^i) \geq \hat{D}_\pi(D_\pi(V_s^i))$, and hence $D_\pi(V_s^i) \in \hat{C}_\pi$.

(iv) It follows directly from the proof of Proposition 7 (ii).

(v) The iteration formula is equivalent to $\tilde{V}_{s,k+1}^i = D_\pi(\tilde{V}_{s,k}^i)$, where $\tilde{V}_{s,k}^i = V_{s,k}^i - \gamma m^i \mathbf{1}_s/(1-\gamma)$. It follows from $V_{s,0}^i \in C_\pi$ that $V_0' \in C(\pi)$ by (ii), and thus $\tilde{V}_{s,k}^i \geq \tilde{V}_{s,k+1}^i$ and $\tilde{V}_{s,k}^i \in C_\pi$ for all $k \in \mathbb{N}$ by (i). According to the monotone convergence theorem, $\tilde{V}_{s,k}^i$ converges as $k \to \infty$, and the limit is the unique solution $V_{\pi s}^i$ of $V_{\pi s}^i = D_\pi(V_{\pi s}^i)$. Hence we obtain

$$\lim_{k\to\infty} V_{s,k}^i = V_{\pi s}^i + \frac{\gamma}{1-\gamma} m^i \mathbf{1}_s \quad \wedge \quad \lim_{k\to\infty}\left(V_{s,k}^i - D_\pi(V_{s,k}^i)\right) = \gamma m^i \mathbf{1}_s.$$

□

Proof of Theorem 9. (i) By definition, π_a^{si} is a perfect equilibrium if and only if

$$V_{\pi s}^i = \max_a{}^{si} \pi_{Aa}^{si-}\left(u_A^{si} + \gamma T_{s'A}^s V_{\pi s'}^i\right),$$

and $V_{\pi s}^i \in \hat{C}_\pi$ if and only if

$$V_{\pi s}^i \geq \pi_{Aa}^{si-}\left(u_A^{si} + \gamma T_{s'A}^s V_{\pi s'}^i\right).$$

Note that the equality can be established in the above inequality, because $V_{\pi s}^i = \pi_A^s \left(u_A^{si} + \gamma T_{s'A}^s V_{\pi s'}^i\right)$. Then the two formulas are equivalent, and hence π_a^{si} is a perfect equilibrium if and only if $V_{\pi s}^i \in \hat{C}_\pi$.

(ii) For every $x \in \mathcal{S}$, $\pi_a^{si}(x)$ is a Nash equilibrium if and only if

$$D_\pi(V_s^i)(x) = \hat{D}_\pi(V_s^i)(x).$$

First, suppose $\pi_a^{si}(x)$ is a Nash equilibrium. Consider Y_{xs}^i and d_x^i that satisfies formula (11a), and then substituting with $V_s^i = Y_{xs}^i + d_x^i \mathbf{1}_s$, we have

$$D_\pi(Y_{xs}^i)(x) = \hat{D}_\pi(Y_{xs}^i)(x),$$

and further there is

$$\left(Y_{xs}^i - \hat{D}_\pi(Y_{xs}^i)\right)(x) = 0.$$

By the uniqueness of the pair of $\hat{Y}_{xs}^i$ and $\hat{d}_x^i$, we obtain $Y_{xs}^i = \hat{Y}_{xs}^i$.

Conversely, suppose $Y_{xs}^i = \hat{Y}_{xs}^i$, and it follows that $d_x^i = \hat{d}_x^i$. By formula (11a) and (11b) there are

$$\begin{aligned}\left(V_s^i - D_\pi(V_s^i) - (1-\gamma)d_x^i \mathbf{1}_s\right)(x) =& 0,\\ \left(V_s^i - \hat{D}_\pi(V_s^i) - (1-\gamma)\hat{d}_x^i \mathbf{1}_s\right)(x) =& 0.\end{aligned}$$

Then $D_\pi(V_s^i)(x) = \hat{D}_\pi(V_s^i)(x)$, and hence $\pi_a^{si}(x)$ is a Nash equilibrium. □

Proof of Proposition 10. Note that the limit is always a perfect equilibrium as long as $V_{s,k}^i$ converges under the assumption that $D_{\pi_k}(V_s^i) = \hat{D}_{\pi_k}(V_s^i)$, so it is suffice to show that $V_{s,k}^i$ converges.

First, use (i) to prove (ii). Consider $V_s^i - \delta \le Y_s^i \le V_s^i + \delta$, where $V_s^i, Y_s^i \in O(\hat{V}_s^i)$ and $\delta = \|V_s^i - Y_s^i\|_\infty$ is small enough. Using $V_s^i \le Y_s^i \to D(V_s^i) \le D(Y_s^i)$, there is

$$D(V_s^i) - \gamma\delta \le D(Y_s^i) \le D(V_s^i) + \gamma\delta.$$

Then we obtain $\|D(V_s^i) - D(Y_s^i)\|_\infty \le \gamma \|V_s^i - Y_s^i\|_\infty$. Thus, D is a contraction mapping on $O(\hat{V}_s^i)$, and $V_{s,k}^i \in O(\hat{V}_s^i)$ converges by the contraction mapping theorem.

By $V_s^i \le Y_s^i \to D(V_s^i) \le D(Y_s^i)$ and $V_{s,0}^i \in C_{\pi_0}$, we have $V_{s,k+1}^i \le V_{s,k}^i$ for all k, and hence we obtain $V_{s,k}^i \in C_{\pi_k}$ for all k.

Then use (ii) to prove (iii). Monotonicity follows directly from $V_{s,k}^i \in C_{\pi_k}$ for all k, and convergence already holds, and hence (iii) is obtained.

Then use (iii) to prove (iv). It is obtained directly from monotonicity.

Finally, use (iv) to prove (iii). It follows from $V_{s,k+1}^i \le V_{s,k}^i \to D(V_{s,k+1}^i) \le D(V_{s,k}^i)$ that

$$V_{s,k}^i \in C_{\pi_k} \to V_{s,k+1}^i \in C_{\pi_{k+1}},$$

and thus $V_{s,k}^i \in C_{\pi_k}$ for all k. Considering $V_{s,0}^i \in C_{\pi_0}$, it follows that $V_{s,k}^i$ monotonically decreases by Theorem 8 (i).

Note that $V_s^i < \pi_A^s(u_A^{si} + \gamma T_{s'A}^s V_{s'}^i)$ for any π_a^{si} when $V_s^i < \min_A^{si} u_A^{si}/(1-\gamma)$, that is, $\bigcup_\pi C_\pi$ is bounded, and thus there is a lower bound for $V_{s,k}^i$. Hence $V_{s,k}^i$ converges by the monotone convergence theorem. □

Proof of Theorem 11. The existence of a sequence $\{m_k^i\}_{k\in\mathbb{N}}$ such that $V_{s,k}^i + m_k^i \mathbf{1}_s \in \hat{C}_{\pi_k}$ for all $k \in \mathbb{N}$ is guaranteed by Proposition 7 (iii).

First, suppose $V_{s,k}^i$ converges to a perfect equilibrium value function $\tilde{V}_s^i$, then $\tilde{V}_s^i = D_{\tilde{\pi}}(\tilde{V}_s^i)$ for some $\tilde{\pi}_a^{si}$, and thus $\lim_{k\to\infty} m_k^i = 0$.

Conversely, suppose $\lim_{k\to\infty} m_k^i = 0$, then $V_{s,k}^i$ monotonically decreases when k is large enough since $V_{s,k}^i + m_k^i \mathbf{1}_s \in \hat{C}_{\pi_k}$. Note that $V_{\pi s}^i$ is bounded below, and thus $V_{s,k}^i$ satisfies the monotone convergence property and converges to some $\tilde{V}_s^i$. It follows from $\lim_{k\to\infty} m_k^i = 0$ that $\tilde{V}_s^i = V_{\tilde{\pi}s}^i$ and $\tilde{V}_s^i \in \hat{C}_{\tilde{\pi}}$ for some $\tilde{\pi}_a^{si}$. Hence, $\tilde{V}_s^i$ is a perfect equilibrium value function. □

Proofs in section 5

Proof of Proposition 12. (i) By the iteration formula, the residual $V_s^i - D_\pi(V_s^i + m^i \mathbf{1}_s)$ satisfies

$$D_\pi(V_s^i + m^i\mathbf{1}_s) - D_\pi(D_\pi(V_s^i + m^i\mathbf{1}_s) + m^i\mathbf{1}_s) = \gamma T^s_{\pi s'}\left(V_s^i - D_\pi(V_s^i + m^i\mathbf{1}_s)\right).$$

Note that $\gamma \in [0,1)$ and $T^s_{\pi s'}$ is constant, and $V_s^i - D_\pi(V_s^i + m^i\mathbf{1}_s)$ converges to 0 as iteration. Thus, $\|V_s^i - D_\pi(V_s^i + m^i\mathbf{1}_s)\|_\infty$ converge to 0 linearly with a rate of γ.

(ii) Denote $\tilde{\mu}_a^i(\pi_a^i) := \mu_a^i - (\mathbf{1}_a\mu_a^i)\circ\pi_a^i$, the difference of $\tilde{\mu}_a^i(\pi_a^i)$ is

$$d\tilde{\mu}_a^i(\pi_a^i) = d\mu_a^i - (\mathbf{1}_a(\mu_a^i + d\mu_a^i))\circ d\pi_a^i - (\mathbf{1}_a d\mu_a^i)\circ\pi_a^i.$$

According to the Taylor's formula of $d\pi_a^i$ with respect to $d\mu_a^i - (\mathbf{1}_a d\mu_a^i)\circ\pi_a^i$, there is

$$d\pi_a^i = \pi_a^i \circ \left(\frac{\mu^j_{a'} d\pi_a^i}{\pi_a^i d\mu^j_{a'}} \frac{d\mu^j_{a'} - (\mathbf{1}_{a'} d\mu^j_{a'})\circ\pi^j_{a'}}{\mu^j_{a'}}\right) + o(d\mu_a^i - (\mathbf{1}_a d\mu_a^i)\circ\pi_a^i).$$

Then we substitute in $d\pi_a^i$ and $d\mu_a^i = -\eta^i\circ\mu_a^i + \beta^i\circ\pi_a^i$, while noting that the part of $d\pi_a^i$ due to the term $\beta^i\circ\pi_a^i$ is 0 for any β^i on the equilibrium bundle, and obtain

$$\begin{aligned} d\tilde{\mu}_a^i(\pi_a^i) = &-\eta^j \circ \left(I - (1-\eta^i)\circ\mu_a^i\circ\frac{\mu^j_{a'} d\pi_a^i}{\pi_a^i d\mu^j_{a'}}\circ {\mu^j_{a'}}^{-1}\right)\tilde{\mu}^j_{a'}(\pi_a^i) \\ &+ \eta^j\circ(1-\eta^i)\circ\left((\mathbf{1}_a\mu_a^i)\circ o(\tilde{\mu}_a^i(\pi_a^i)) - \tilde{\mu}_a^i(\pi_a^i)\circ\frac{\mu^j_{a'} d\pi_a^i}{\pi_a^i d\mu^j_{a'}}\frac{\tilde{\mu}^j_{a'}(\pi_a^i)}{\mu^j_{a'}}\right) \end{aligned}$$

The second additive is a higher-order infinitesimal with respect to $\tilde{\mu}_a^i(\pi_a^i)$. The differential $(\mu^j_{a'} d\pi_a^i)/(\pi_a^i d\mu^j_{a'})$ of the equilibrium bundle tends to I as $\mu_a^i \to \infty$ on the fiber over π_a^i, then there always exists a large enough μ_a^i on the fiber over π_a^i such that the eigenvalues of $(I - (1-\eta^i)\circ\mu_a^i\circ((\mu^j_{a'} d\pi_a^i)/(\pi_a^i d\mu^j_{a'}))\circ {\mu^j_{a'}}^{-1})$ are all positive for a given η^i. Note that μ_a^i has a general formula under the iteration formula, and it follows that $\tilde{\mu}_a^i(\pi_a^i)$ converges to 0. Thus, $\|\tilde{\mu}_a^i(\pi_a^i)\|_\infty$ converges to 0 linearly.

Then we show $\|\bar{\mu}_a^i(\pi_a^i)\|_\infty$ also converges to 0 linearly.

$$\tilde{\mu}_a^i(\pi_a^i) = \left(\mu_a^i - (\mathbf{1}_a\mu_a^i)\circ\pi_a^i\right) = \pi_a^i\circ\left(\pi_A U^i_{\pi A} - \pi^{i-}_{Aa} U^i_{\pi A}\right)$$

$$\bar{\mu}_a^i(\pi_a^i) = \pi_a^i\circ\left(\max_a{}^i \pi^{i-}_{Aa} U^i_{\pi A} - \pi^{i-}_{Aa} U^i_{\pi A}\right) = \tilde{\mu}_a^i(\pi_a^i) + \pi_a^i\circ\left(\max_a{}^i \pi^{i-}_{Aa} U^i_{\pi A} - \pi_A U^i_{\pi A}\right)$$

Note that when $\|\tilde{\mu}_a^i(\pi_a^i)\|_\infty < \epsilon$, there exists a coefficient k irrelevant to ϵ such that either $\pi_a^i < k\epsilon$ or $|\pi_A U^i_{\pi A} - \pi^{i-}_{Aa} U^i_{\pi A}| < k\epsilon$ for every index (i,a). Then we have specifically $|\pi_A U^i_{\pi A} - \max_a{}^i \pi^{i-}_{Aa} U^i_{\pi A}| < k\epsilon$ for the corresponding index since the algorithm leads to Nash equilibria. It follows that $\|\tilde{\mu}_a^i(\pi_a^i)\|_\infty < \epsilon$ implies $\|\bar{\mu}_a^i(\pi_a^i)\|_\infty < (k+1)\epsilon$ with k irrelevant to ϵ. Hence, $\|\bar{\mu}_a^i(\pi_a^i)\|_\infty$ converges to 0 linearly.

(iii) Consider unbiased barrier problem (4) with $(\hat{\pi}_a^i, \hat{r}_a^i)$ being variables instead of constant parameters. By the objective function and constraint $r_a^i - v^i + \pi^{i-}_{Aa} U_A^i = 0$, the differential of objective function is

$$\begin{aligned} &\left(\pi_a^i - \hat{\pi}_a^i\right)\left(dr_a^i + \mu_a^i/{\pi_a^i}^2\circ d\pi_a^i\right) + \left(r_a^i - \hat{r}_a^i\right)\left(d\pi_a^i + \mu_a^i/{r_a^i}^2\circ dr_a^i\right) \\ =&\left(\left(\pi_a^i - \hat{\pi}_a^i\right)\circ M_a^i\right)\left(dv^i - \pi^{ij-}_{Aaa'} U_A^i d\pi^j_{a'}\right) + \left(\left(r_a^i - \hat{r}_a^i\right)\circ M_a^i\right) d\pi_a^i. \end{aligned}$$

It follows from $\mathbf{1}_a\hat{\pi}_a^i - \mathbf{1}^i = 0$ that $(\pi_a^i - \hat{\pi}_a^i)dv^i = 0$, and thus the differential is

$$\left(-\left(\left(\pi_a^i - \hat{\pi}_a^i\right)\circ M_a^i\right)\pi^{ij-}_{Aaa'} U_A^i + \left(\left(r^j_{a'} - \hat{r}^j_{a'}\right)\circ M^j_{a'}\right) - \delta^j_{a'}\right) d\pi^j_{a'}.$$

Finally, project the gradient regarding the constraint $\mathbf{1}_a\pi_a^i - \mathbf{1}^i = 0$, and we obtain the projected gradient

$$\mathrm{pg}^j_{a''} = N_{a'a''}\left(\left(\left(r^j_{a'} - \hat{r}^j_{a'}\right)\circ M^j_{a'}\right) - \left(\left(\pi_a^i - \hat{\pi}_a^i\right)\circ M_a^i\right)\pi^{ij-}_{Aaa'} U_A^i - \delta^j_{a'}\right),$$

where

$$N_{a'a''} = I_{a'a''} - \frac{\mathbf{1}_{a'}\mathbf{1}_{a''}}{|\mathcal{A}|}, M^j_{a'} = \mathbf{2}^j_{a'} + \frac{\mu^j_{a'} - \pi^j_{a'}\circ r^j_{a'}}{\pi^j_{a'}\circ r^j_{a'}}, \delta^j_{a'} = \frac{(\pi^j_{a'} - \hat{\pi}^j_{a'})^2}{\pi^j_{a'}}\circ\frac{dv^j}{d\pi^j_{a'}}.$$

This projected gradient and the projected gradient (6) only differ by two higher-order infinitesimals $M_{a'}^{j}$ and $\delta_{a'}^{j}$. Note also that on a non-singular point, unbiased barrier problem (4) is locally strictly convex as Theorem 5 (iv) has proved. Considering that projected gradient descent is known to converge sublinearly, projected gradient descent (6) converges sublinearly if the convergence point is non-singular.

□