# Satisfaction Paths in Markov Games

Extended Abstract

Yanqing Fu
Tongji University
Shanghai, China
fuyanqing159@163.com

Chao Huang
Tongji University
Shanghai, China
csehuangchao@tongji.edu.cn

Chenrun Wang
Tongji University
Shanghai, China
1015239683@qq.com

Zhuping Wang
Tongji University
Shanghai, China
elewzp@tongji.edu.cn

## ABSTRACT

In multi-agent reinforcement learning (MARL), agents generate a sequence of joint strategy profiles $(s^t)_{t\geq 0}$, where $s^t$ is the profile at step $t$. A key update heuristic is "win-stay, lose-shift", under which an agent retains its current strategy only if it is a best response to others' choices. The resulting trajectory is called a satisfaction path. The question of whether such paths exist in Markov games had remained open. This paper introduces the notion of a grouped satisfaction path, examines the topological properties of its local minimum, establishes sufficient conditions for its existence in pure strategy games, and finally resolves the open problem for Markov games. These contributions deepen the theoretical foundation for MARL algorithm design.

## KEYWORDS

Multi-agent reinforcement learning; Satisfaction path

**ACM Reference Format:**
Yanqing Fu, Chao Huang, Chenrun Wang, and Zhuping Wang. 2026. Satisfaction Paths in Markov Games: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/KCNB4874

## 1 INTRODUCTION

From a game-theoretic perspective, multi-agent reinforcement learning (MARL) can be viewed as a repeated game in which agents iteratively select strategies based on available information, observe feedback, and adjust their behavior [9, 14]. This traces a strategy path $(s^t)_{t=0}^{T}$, where $s^t$ is the joint strategy profile at step $t$. Equilibrium concepts are central to MARL [3, 4, 7], describing stable states where no agent gains by unilaterally changing its strategy and linking learning dynamics to strategic stability [8].

A key challenge in MARL is determining whether decentralized strategy updates can drive the joint strategy profile toward equilibrium. Despite extensive study [6, 10, 13], the problem remains only partially resolved [18]. Formally, each agent $i$ updates its strategy via a revision function $s_i^{t+1} = f_i\big((s^u)_{0\leq u\leq t}\big)$, mapping the history of joint strategies to a new strategy [1, 11]. A common heuristic is the "win-stay, lose-shift" principle [2, 5, 12, 17]: an agent keeps its current strategy if it is a best response, otherwise it switches. The resulting trajectory is termed a satisfaction path [16]. Its existence means that from any initial joint strategy profile, there is a finite sequence adhering to this principle that ends in an equilibrium.

In this work, we study satisfaction paths from a topological perspective to gain new insights into their existence. Moving beyond specific revision rules, we employ a general theoretical framework to analyze the conditions for such paths. Our main objective is to lay a theoretical foundation for satisfaction paths, thereby advancing the understanding of convergence in MARL.



*Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus.* © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/KCNB4874

## 2 PRELIMINARY

A pure strategy game is defined as a tuple

$$G = (I, (S_i)_{i\in I}, (g_i)_{i\in I}),$$

where $I$ is the player set, $S_i$ is the pure strategy set of player $i$, and $g_i : \prod_{i\in I} S_i \to R$ is player $i$'s payoff function.

Definition 1. *In a pure strategy game $G$, $\epsilon$-best response ($\epsilon$-BR) correspondence for player $i$ given opponents' strategy profile $s_{-i}$ is*

$$BR_\epsilon(s_{-i}) = \left\{ s \in S_i \mid g_i(s, s_{-i}) \geq \sup_{t\in S_i} g_i(t, s_{-i}) - \epsilon \right\}. \quad (1)$$

*where $s_{-i} \in \prod_{j\neq i, j\in I} S_j$ denotes joint strategy profile of all other players. In particular, the notation $BR_\epsilon$ will be simplified to $BR$ when $\epsilon = 0$.*

Definition 2. *In a pure strategy game $G$, a joint strategy profile $s \in \prod_{i\in I} S_i$ is called an $\epsilon$-equilibrium if*

$$\forall i \in I, \quad s_i \in BR_\epsilon(s_{-i}). \quad (2)$$

A group set $P$ is a partition of the player set $I$, satisfying:

(1) For any $p \in P$, $p \neq \emptyset$.
(2) $\bigcup_{p\in P} p = I$.
(3) For any $p, q \in P$, if $p \neq q$, then $p \cap q = \emptyset$.

Thus, $P$ groups the players into disjoint, non-empty subsets, each acting as an independent decision unit.

Definition 3. *In a pure strategy game $G$ with a group set $P$, a strategy profile path $(s^t)_{t\geq 0}$ is called a grouped $\epsilon$-satisfaction path if*

*for any $t \geq 0$ and any $p \in P$,*

$$(\forall i \in p,\ s_i^t \in BR_\epsilon(s_{-i}^t)) \quad \Rightarrow \quad (\forall i \in p,\ s_i^{t+1} = s_i^t). \tag{3}$$

Definition 4. *In a pure strategy game $G$ with a group set $P$, for a joint strategy profile $s$, the $\epsilon$-BR group count with respect to $s$ is*

$$N_\epsilon(s) = |\{p \in P \mid \forall i \in p,\ s_i \in BR_\epsilon(s_{-i})\}|. \tag{4}$$

*In particular, the notation $N_\epsilon$ will be simplified to $N$ when $\epsilon = 0$.*

Definition 5. *In a pure strategy game $G$ with a group set $P$, for a joint strategy profile $s$, the admissible subsequent joint strategy profile set with respect to $s$ is*

$$T_\epsilon(s) = \left\{ s' \in \prod_{i \in I} S_i \mid \text{ path } (s, s') \text{ satisfies Equation 3} \right\}. \tag{5}$$

*$N_\epsilon(s)$ is called a local minimum (maximum) if*

$$N_\epsilon(s) \leq \inf_{t \in T_\epsilon(s)} N_\epsilon(t), \quad \left( N_\epsilon(s) \geq \sup_{t \in T_\epsilon(s)} N_\epsilon(t) \right). \tag{6}$$

*In particular, the notation $T_\epsilon$ will be simplified to $T$ when $\epsilon = 0$.*

## 3 SATISFACTION PATH THEORY

Theorem 1. *In a pure strategy game $G$ with a group set $P$, suppose that each $S_i$ is a convex compact set in a topological vector space, and each $g_i$ is analytic. Then for a joint strategy profile $s$, $N(s)$ is a local minimum, if and only if, for any joint strategy $t \in T(s)$, a group that is a BR in $s$ is still a BR in $t$.*

Theorem 1 relates $N(s)$ and $T(s)$. In essence, when $N(s)$ is a local minimum, the BR groups stop changing strategies. Thus, the theorem establishes a reduction principle for games under grouped satisfaction dynamics: once a group permanently fixes its strategy, it can be viewed as part of the environment, reducing the original game to a subgame.

Definition 6. *In a pure strategy game $G$ with a group set $P$, for a group subset $Q \subset P$ and a joint strategy profile $s$, a game $\tilde{G} = (J, (S_i)_{i \in J}, (\tilde{g}_i)_{i \in J})$ is called a sub-game if $J = \bigcup_{p \in Q} p$ and*

$$\tilde{g}_i : \prod_{j \in J} S_j \to \mathbb{R}, \quad w \mapsto g_i(s_{-J}, w).$$

*where $s_{-J} = (s_k)_{k \in I \setminus J}$ denotes the strategy profile of all players outside $J$.*

*It is said that any sub-game has an equilibrium, if for any group subset $Q \subset P$ and any joint strategy profile $s$, the sub-game $\tilde{G}$ admits at least one equilibrium.*

Theorem 2. *In a pure strategy game $G$ with a group set $P$, suppose that each $S_i$ is a convex compact set in a topological vector space, each $g_i$ is analytic, and any sub-game has an equilibrium. Then for any initial joint strategy profile $s$, there exists a finite-length grouped satisfaction path $(s^t)_{t=0}^T$ where $s^0 = s$ and $s^T$ is an equilibrium.*

Theorem 2 provides a sufficient condition for the existence of a grouped satisfaction path. Briefly, if players' payoff functions are analytic and every sub-game has an equilibrium, then a grouped satisfaction path exists from any initial joint strategy profile to an equilibrium.

A stationary mixed strategy stochastic game (sometimes called Markov games) is defined as

$$G = (I, (S_i)_{i \in I}, X, P, (g_i)_{i \in I}, (\gamma_i)_{i \in I}).$$

$X$ is the state set. $P : X \times \prod_{i \in I} S_i \to \Delta X$ is the transition probability function, mapping current state and joint strategy profile to a probability distribution over the next states. $g_i : X \times \prod_{i \in I} S_i \to R$ is the payoff function for player $i$. $\gamma_i \in [0, 1)$ is a discount factor for player $i$.

Construct a pure strategy game

$$H = ((I, X), (T_{i,x})_{(i,x) \in (I,X)}, (h_{i,x})_{(i,x) \in (I,X)}). \tag{7}$$

$(I, X) = \{(i, x) \mid i \in I,\ x \in X\}$. Because $|I| < \infty$ and $|X| < \infty$, $(I, X)$ is a finite double index set. $T_{i,x} = \Delta S_i$. Because $S_i$ is finite, $T_{i,x}$ is a simplex in some finite-dimensional Euclidean space. So $T_{i,x}$ is convex and compact. $h_{i,x}$ is

$$h_{i,x} : \prod_{(j,y) \in (I,X)} T_{j,y} \to \mathbb{R},$$

$$\prod_{(j,y) \in (I,X)} \pi_j(y) \mapsto \mathbb{E}_{(\pi_j)_{j \in I}} \left[ \sum_{t \geq 0} \gamma_i^t g_i(x^t, s^t) \mid x_0 = x \right].$$

Clearly, the virtual players in $H$ have a specific structure: all player–state pairs sharing the same player belong to a single group. This means that if a grouped satisfaction path exists in $H$, a corresponding satisfaction path exists in the original game $G$.

Theorem 3. *In a stationary mixed strategy stochastic game $G$, suppose that each $S_i$ is a finite set. Then for any initial stationary joint mixed strategy profile $\sigma$, there exists a finite-length satisfaction path $(\sigma^t)_{t=0}^T$ where $\sigma^0 = \sigma$ and $\sigma^T$ is a mixed equilibrium.*

Theorem 3 establishes the existence of satisfaction paths in Markov games, resolving an open problem from [15]. We first proved the existence of grouped satisfaction paths in pure strategy games. By constructing a connection between Markov games and pure strategy games, we extended the result of Theorem 2 to Markov games. This reduction approach elegantly circumvented the inherent difficulties in directly proving the existence of satisfaction paths in Markov games.

## 4 CONCLUSION

In brief, if a pure strategy game satisfies that *a*) each strategy set is convex and compact, *b*) each payoff function is analytic, *c*) any sub-game has an equilibrium, then for any initial joint strategy profile $s$, there exists a finite-length grouped satisfaction path $(s^t)_{0 \leq t \leq T}$ where $s^0 = s$ and $s^T$ is an equilibrium. In particular, any finite-state Markov game admits finite-length satisfaction paths from an arbitrary initial joint mixed strategy profile to some mixed equilibrium.

## REFERENCES


[1] Lawrence Edward Blume. 1993. The Statistical Mechanics of Strategic Interaction. *Games and Economic Behavior* 5, 3 (1993), 387–424.

[2] Georgios C. Chasparis, Ari Arapostathis, and Jeff S. Shamma. 2013. Aspiration Learning in Coordination Games. *SIAM Journal on Control and Optimization* 1 (2013), 465–490.

[3] Constantinos Daskalakis, Rafael Frongillo, Christos H. Papadimitriou, George Pierrakos, and Gregory Valiant. 2010. On Learning Algorithms for Nash Equilibria. In *Algorithmic Game Theory - Third International Symposium*. 114–125.

[4] Lampros Flokas, Emmanouil Vasileios Vlatakis-Gkaragkounis, Thanasis Lianeas, Panayotis Mertikopoulos, and Georgios Piliouras. 2020. No-regreet learning and mixed Nash equilibria: They do not mix. *Advances in Neural Information Processing Systems* (2020), 1380–1391.

[5] Fabrizio Germano and Gábor Lugosi. 2007. Global Nash Convergence of Foster and Young's Regret Testing. *Games and Economic Behavior* 60, 1 (2007), 135–154.

[6] Yu Guan Hsieh, Kimon Antonakopoulos, and Panayotis Mertikopoulos. 2021. Adaptive Learning in Continuous Games: Optimal Regret Bounds and Convergence to Nash Equilibrium. *Conference on Learning Theory* (2021), 2388–2422.

[7] Amir Jafari, Amy Greenwald, David Gondek, and Gunes Ercal. 2001. On No-Regret Learning, Fictitious Play, and Nash Equilibrium. *International Conference on Machine Learning* (2001), 226–233.

[8] Rida Laraki, Jérme Renault, and Sylvain Sorin. 2019. *Mathematical Foundations of Game Theory.* Springer Nature Switzerland AG, Gewerbestrasse 11, 6330 Cham, Switzerland.

[9] Shengbo Eben Li. 2023. *Reinforcement Learning for Sequential Decision and Optimal Control.* Springer Nature Singapore Pte Ltd, 152 Beach Road, 21-01/04 Gateway East, Singapore 189721, Singapore.

[10] Yulong Lu. 2023. Two-Scale Gradient Descent Ascent Dynamics Finds Mixed Nash Equilibria of Continuous Games: A Mean-Field Perspective. *International Conference on Machine Learning* (2023), 22790–22811.

[11] Jason R. Marden and Jeff S. Shamma. 2012. Revisiting log-linear learning: Asynchrony, completeness and payoff-based implementation. *Games and Economic Behavior* 75, 2 (2012), 788–808.

[12] Jason R. Marden, H. Peyton Young, Gürdal Arslan, and Jeff S. Shamma. 2009. Payoff-Based Dynamics for Multiplayer Weakly Acyclic Games. *SIAM Journal on Control and Optimization* 48, 1 (2009), 373–396.

[13] Satinder Singh, Michael Kearns, and Yishay Mansour. 2000. Nash convergence of gradient dynamics in general-sum games. *Uncertainty in artificial intelligence: Sixteenth conference on uncertainty in artificial intelligence* (2000), 541–548.

[14] Karl Tuyls and Ann Nowé. 2005. Evolutionary game theory and multi-agent reinforcement learning. *Knowledge Engineering Review* 20 (2005), 63–90.

[15] Bora Yongacoglu, Gürdal Arslan, Lacra Pavel, and Serdar Yüksel. 2024. Paths to Equilibrium in Games. *38th Conference on Neural Information Processing Systems* (2024).

[16] Bora Yongacoglu, Gürdal Arslan, and Serdar Yüksel. 2023. Satisficing Paths and Independent Multiagent Reinforcement Learning in Stochastic Games. *SIAM Journal on Mathematics of Data Science* 3 (2023), 745–773.

[17] H Peyton Young and Marden Lucy Y. Pao. 2014. Achieving Pareto Optimality Through Distributed Learning. *SIAM Journal on Control and Optimization* 5, 2753–2770.

[18] Kaiqing Zhang, Zhuoran Yang, and Tamer Baar. 2021. Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms. *Handbook of Reinforcement Learning and Control* (2021).