Situational-Constrained Multi-Agent Coordination through Correlated Equilibria

Extended Abstract

Libo Zhang
The University of Auckland
Auckland, New Zealand
lzha797@aucklanduni.ac.nz

Zhirui Zeng
The University of Auckland
Auckland, New Zealand
zzen859@aucklanduni.ac.nz

Yang Chen
Shanghai Artificial Intelligence Laboratory
Shanghai, China
chenyang4@pjlab.org.cn

Jiamou Liu
The University of Auckland
Auckland, New Zealand
jiamou.liu@auckland.ac.nz

ABSTRACT

Correlated Equilibria (CE) provide a robust framework for balancing individual and collective incentives in multi-agent coordination. However, real-world applications often impose *situational constraints*, i.e., context-dependent requirements triggered only under specific conditions (e.g., emergency protocols), which challenge existing methods. To address this, we propose *Situational-Constrained Density-Based Correlated Equilibria* (SC-DBCE) within a Markov game framework. We develop *Situational-Constrained Correlated Policy Iteration* (SC-CPI), the first RL algorithm capable of solving these equilibria using a smooth Log-Sum-Exp optimization mechanism. Experiments across multiple scenarios demonstrate that SC-CPI consistently outperforms baselines in both equilibrium quality and constraint adherence.

KEYWORDS

Correlated Equilibrium; Multi-agent Coordination

ACM Reference Format:

Libo Zhang, Zhirui Zeng, Yang Chen, and Jiamou Liu. 2026. Situational-Constrained Multi-Agent Coordination through Correlated Equilibria: 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/FZYS4368

1 INTRODUCTION

Multi-agent coordination aims to align collective objectives with individual interests, often arising in settings such as multi-agent pathfinding, communication, and smart grid scenarios, where objectives are often mixed cooperative-competitive [3, 7, 11]. A canonical approach to address such coordination challenges lies in game theory. Specifically, the concept of Correlated Equilibrium (CE) serves as a powerful framework for resolving mixed cooperative-competitive dynamics, allowing agents to correlate their strategies via shared signals to achieve mutually beneficial outcomes.

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/FZYS4368

While individual objectives are effectively managed by these equilibrium concepts, modeling the necessary system-level constraints remains a significant challenge. Standard game-theoretic methods typically focus on reward maximization, often neglecting adequate modeling for auxiliary properties such as fairness and safety. Moreover, existing works that do address these requirements predominantly focus on *static constraints*, e.g., designing cost functions [1] in constrained models or imposing constraints uniformly through the entire execution [9]. Consequently, these approaches overlook *situational constraints* [10], which are critical context-dependent requirements that must be dynamically enforced only under specific conditions.

To bridge this gap, we propose *Situational-Constrained Density-Based Correlated Equilibria* (SC-DBCE), a novel solution concept that formalizes situational constraints within the coordination framework. We further develop *Situational-Constrained Correlated Policy Iteration* (SC-CPI), a reinforcement learning algorithm designed to compute such equilibria. We validate our approach through extensive experiments on the multi-agent game benchmarks[9], demonstrating that SC-CPI consistently outperforms baselines in both constraint adherence and equilibrium quality.

2 MULTI-AGENT COORDINATION

We model coordination as a Markov game [5], defined by the tuple $\langle \mathcal{S}, \{\mathcal{A}_i\}_{i=1}^N, P, \{r_i\}_{i=1}^N, \eta, \gamma \rangle$. We adopt *Correlated Equilibrium* (CE) [2] as a solution concept. A joint policy $\boldsymbol{\pi}$ is a CE if no agent benefits from unilateral deviation, ensured by non-positive *regret*—the expected gain from switching to an alternative action.

To enforce auxiliary requirements, we utilize state visitation density $\rho^{\boldsymbol{\pi}}(s) = \sum_{t=0}^{\infty} \gamma^t \Pr(s^t = s | \boldsymbol{\pi}, \eta)$ [6]. We focus on *situational constraints* $\psi : \varphi_{cond} \rightarrow \varphi_{req}$, formalized as logical implications over densities. This structure captures context-dependence via the equivalence $\neg\varphi_{cond} \vee \varphi_{req}$: the requirement φ_{req} is strictly enforced only when the context φ_{cond} holds. We denote $\Psi \equiv \wedge_{k \in K} \psi_k$ as the conjunction of k situational constraints.

We propose *Situational-Constrained Density-Based Correlated Equilibrium* (SC-DBCE) to minimize constraint violation $\sigma(\Psi, \boldsymbol{\pi})$ while maintaining equilibrium. We formulate the problem as:

$$\min_{\boldsymbol{\pi} \in \Pi} \sigma(\Psi, \boldsymbol{\pi}) \quad \text{subject to} \quad \forall i, s, a_i, a_i' : \text{Regret}_{\boldsymbol{\pi}}(s, i, a_i, a_i') \leq 0$$

This optimization ensures agents dynamically adhere to complex, context-sensitive rules without sacrificing rationality.

3 ALGORITHM

We unify the problem representation using the *occupancy measure* $\mu(s, \boldsymbol{a})$ [8], which represents discounted joint visitation frequencies. A function μ corresponds to a valid stationary policy if and only if it satisfies the *Bellman flow constraint*, which requires that the total outflow from a state matches the initial distribution η plus the discounted inflow:

$$\sum_{\boldsymbol{a} \in \mathcal{A}} \mu(s, \boldsymbol{a}) = \eta(s) + \gamma \sum_{s' \in \mathcal{S}} \sum_{\boldsymbol{a}' \in \mathcal{A}} P(s \mid s', \boldsymbol{a}') \mu(s', \boldsymbol{a}'), \quad \forall s \in \mathcal{S}.$$

Deviations from this equality are quantified as the *Bellman Flow Error* (BFErr). By enforcing BFErr = 0, we transform the SC-DBCE problem into a convex optimization task, allowing for the direct optimization of density-based situational constraints within the space of valid system dynamics.

To handle the disjunction of situational constraints $\psi \equiv \neg\varphi_{cond} \vee \varphi_{req}$, we employ the *Log-Sum-Exp* (LSE) technique. We construct a smooth, differentiable objective function:

$$\sigma(\psi, \mu) = -\frac{1}{\beta} \log \left(\exp(-\beta \cdot \sigma(\neg\varphi_{cond}, \mu)) + \exp(-\beta \cdot \sigma(\varphi_{req}, \mu)) \right)$$

where β controls approximation sharpness. This formulation stabilizes optimization by avoiding switching between different objective functions under probabilistic approaches [10]. As for Ψ, we take the sum of σ over each component ψ_k in the conjunction.

The SC-CPI algorithm (Alg. 1) computes the SC-DBCE by iterating between two phases: (1) **Occupancy Optimization**, which solves for μ by minimizing the LSE objective subject to Bellman flow and current regret constraints; and (2) **Policy Evaluation**, which updates Q-functions under the derived policy $\boldsymbol{\pi}$ to refine the regret constraints. This procedure guarantees convergence to a valid equilibrium that respects the situational constraints.

Algorithm 1 Situational-Constrained Correlated Policy Iteration (SC-CPI)

Input: Markov game, Situational Constraint Ψ.
Initialize: Occupancy Measure μ, Q-functions $\{Q_i\}$, learning rate α.
for each iteration **do**
  Solve μ by minimizing $\sigma(\Psi, \mu)$ subject to Bellman Flow and CE constraints defined by current Q_i.
  Derive joint policy $\boldsymbol{\pi}$ from μ.
  while not converged **do**
    Sample transitions under $\boldsymbol{\pi}$.
    update Q_i for each agent.
  end while
end for
Output: Joint policy $\boldsymbol{\pi}$.

4 EXPERIMENT

We evaluate SC-CPI on the multi-agent game benchmark [9] augmented with situational constraints: **Fair Gamble** (risk balancing), **Collect and Explore** (cooperative synchronization), and **Hunters** (conditional safety in mixed-motive settings).

Scenario	Method	Cons.Vio.	MaxReg	BFErr
Fair Gamble	uCE-Q	**0.00 ± 0.00**	0.02 ± 0.02	0.20 ± 0.23
	DBCPI-1	**0.00 ± 0.00**	0.00 ± 0.01	0.05 ± 0.08
	DBCPI-2	23.66 ± 24.65	0.05 ± 0.06	0.56 ± 0.63
	Prob-DBCPI	**0.00 ± 0.00**	0.00 ± 0.00	0.00 ± 0.01
	SC-CPI	**0.00 ± 0.00**	0.01 ± 0.02	0.16 ± 0.25
CaE	uCE-Q	**0.00 ± 0.00**	0.00 ± 0.00	0.00 ± 0.00
	DBCPI-1	19.83 ± 20.88	0.00 ± 0.00	0.00 ± 0.00
	DBCPI-2	**0.00 ± 0.00**	0.00 ± 0.00	0.00 ± 0.00
	Prob-DBCPI	4.44 ± 3.56	8.20 ± 8.92	1.39 ± 0.84
	SC-CPI	**0.00 ± 0.00**	0.00 ± 0.00	0.00 ± 0.00
Hunters	uCE-Q	0.72 ± 1.18	3.95 ± 4.87	0.00 ± 0.00
	DBCPI-1	**0.00 ± 0.00**	0.00 ± 0.01	0.00 ± 0.00
	DBCPI-2	8.11 ± 17.61	0.06 ± 0.13	0.01 ± 0.03
	Prob-DBCPI	3.09 ± 0.30	10.99 ± 0.96	0.17 ± 0.18
	SC-CPI	0.41 ± 0.92	0.00 ± 0.00	0.00 ± 0.00

Table 1: Experiment results in mean ± std format. Underline indicates failures in satisfying CE or Bellman Flow constraints.

Baselines. We compare SC-CPI against: (1) **uCE-Q** [4], which maximizes the sum of rewards; (2) **DBCPI** [9], evaluated as **DBCPI-1** and **DBCPI-2** by decomposing the situational constraint into independent static ones; and (3) **Prob-DBCPI** [10], which selects constraints thatare closer to satisfaction in a probabilistic manner.

Performance Metrics. We consider: (1) **Constraint Violation (Cons.Vio.)**: low Cons.Vio. indicates a better constraint satisfaction; (2) **Maximum Regret (MaxReg)**: low MaxReg indicates higher equilibrium quality; (3) **BellmanFlow Error (BFErr)**: low BFErr indicates a reliable result.

Experiment Result. Table 1 confirms SC-CPI's superiority. **uCE-Q** fails to address situational constraint, resulting in violations and high regret in *Hunters*. **DBCPI** variants prove unreliable, as preselecting a fixed disjunct often conflicts with equilibrium dynamics (e.g., DBCPI-2 in *Fair Gamble*). **Prob-DBCPI** suffers from optimization instability, leading to high regret (up to 10.99) and poor constraint adherence. In contrast, **SC-CPI** consistently achieves near-zero violations and regret, validating its ability to learn robust equilibria under complex constraints.

5 CONCLUSION AND LIMITATIONS

This paper solves multi-agent coordination problems under situational constraints. We model the problem as Markov Games, and propose Situational-Constrained Density-Based Equilibrium as a solution concept. We develop SC-CPI as the first RL algorithm capable of finding such equilibria under such context-sensitive objectives.

Limitations. Our approach still computes a joint policy, which may limit scalability in larger systems. Future work could explore decentralized methods to improve applicability in large-scale multiagent settings.

REFERENCES

[1] Eitan Altman and Adam Shwartz. 2000. Constrained markov games: Nash equilibria. In *Advances in dynamic games and applications*. Springer, 213–221.
[2] Robert J Aumann. 1987. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica: Journal of the Econometric Society* (1987), 1–18.
[3] Eli Boyarski, Ariel Felner, Roni Stern, Guni Sharon, Oded Betzalel, David Tolpin, and Eyal Shimony. 2015. Icbs: The improved conflict-based search algorithm for multi-agent pathfinding. In *Proceedings of the International Symposium on Combinatorial Search*, Vol. 6. 223–225.
[4] Amy Greenwald, Keith Hall, Roberto Serrano, et al. 2003. Correlated Q-learning. In *ICML*, Vol. 3. 242–249.
[5] Michael L Littman. 1994. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*. Elsevier, 157–163.
[6] Anders Rantzer. 2001. A dual to Lyapunov's stability theorem. *Systems & Control Letters* 42, 3 (2001), 161–168.
[7] Lijun Sun, Yijun Yang, Qiqi Duan, Yuhui Shi, Chao Lyu, Yu-Cheng Chang, Chin-Teng Lin, and Yang Shen. 2025. Multi-agent coordination across diverse applications: A survey. *arXiv preprint arXiv:2502.14743* (2025).
[8] Umar Syed, Michael Bowling, and Robert E Schapire. 2008. Apprenticeship learning using linear programming. In *Proceedings of the 25th international conference on Machine learning*. 1032–1039.
[9] Libo Zhang, Yang Chen, Toru Takisaka, Bakh Khoussainov, Michael Witbrock, and Jiamou Liu. 2023. Learning Density-Based Correlated Equilibria for Markov Games. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 652–660.
[10] Libo Zhang, Yang Chen, Toru Takisaka, Kaiqi Zhao, Weidong Li, and Jiamou Liu. 2025. Situational-constrained sequential resources allocation via reinforcement learning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence* (Montreal, Canada) *(IJCAI '25)*. Article 1014, 9 pages. https://doi.org/10.24963/ijcai.2025/1014
[11] Libo Zhang, Yuly Wu, Weidong Li, Song Yang, Yang Chen, Kaiqi Zhao, and Jiamou Liu. 2024. Optimizing Resource Distribution Towards Energy Justice in Resilient Smart Grids. In *Principle and Practice of Data and Knowledge Acquisition Workshop*. Springer, 236–245.