Population Markov Potential Game: An Alternative Framework for Markovian Traffic Assignment

Xuhang Liu, Kenan Zhang*, Rui Yao
School of Architecture, Civil and Environmental Engineering, École polytechnique fédérale de Lausanne, Lausanne, Switzerland
* Corresponding author

Extended abstract submitted for presentation at the 12th Triennial Symposium on Transportation Analysis conference (TRISTAN XII)
June 22-27, 2025, Okinawa, Japan

February 28, 2025

Keywords: traffic assignment problem, Markov potential game, population game

1 INTRODUCTION

Traffic Assignment Problem (TAP) aims to predict the traffic flows on a given road network and demand profile. Adopting the equilibrium concept from game theory, the solution to TAP is defined on a traffic flow pattern such that no one has incentives to further change their routing strategies (Sheffi, 1985). The TAP models can be categorized in different ways, while two common classifications are i) static vs dynamic: whether time-dependent traffic dynamics are considered; and ii) deterministic vs stochastic: whether route choice contains randomness or uncertainty. This study proposes a novel dynamic and stochastic TAP model, where each vehicle's routing decision is characterized by a Markov decision process (MDP). Accordingly, the traffic dynamics are Markovian while the stochastic routing behaviors are described by a state-dependent policy.

A series of Markovian TAP models have been proposed in the literature, though mostly from the perspective of route choice modeling (e.g., Bell, 1995, Akamatsu, 1997, Baillon & Cominetti, 2008, Oyama *et al.*, 2022). Specifically, these models consider route choice as a sequence of link choices and thus naturally fits it into the MDP framework, where state is the current node, action refers to the next link, and the link choice probabilities are given by a presumed discrete choice model, e.g., multinomial logit (MNL) model as per Akamatsu (1997). The link choice probabilities, combined with the demand vector, yield expected link flows, which, in turn, determine link utilities and complete the feedback loop. The existing Markovian TAP models, however, are restricted to a predefined policy (e.g., MNL) and deterministic state transitions (e.g., the agent surely moves to the end node of the selected link). The model proposed in this study is free of these constraints and thus provides a more flexible modeling framework.

Our proposed model, namely, population Markov potential game (PMPG), stands at the intersection of three classic games: population game, potential game, and Markov game. Hence, it can be seen as an extension of the population potential game (Sandholm, 2001) to Markov dynamics, a special case of Markov population games (Elokda *et al.*, 2024) with a potential function, or a special case of Markov potential game (Leonardos *et al.*, 2021) over a large population of agents. Nevertheless, the proposed model has a wide range of applications in transportation research thanks to its ideal analytical properties and efficient solution approach. Due to the page limit, we only present one example, while more applications will be included in the full paper.

2 POPULATION MARKOV POTENTIAL GAME

2.1 Model setup

Consider a Markov game defined on $\mathcal{G} = (\mathcal{N}, \mathcal{S}, \mathcal{A}, R, P, \gamma, \rho)$ with a finite horizon T, where

- $\mathcal{N}$ is the set of homogeneous agents with population size $N = |\mathcal{N}|$.
- $\mathcal{S}$ is the finite state space and $\Delta(\mathcal{S})$ refers to the probability simplex over $\mathcal{S}$. We use $s_{n,t}$ to denote the state of agent $n \in \mathcal{N}$ at time $t \in \mathcal{T} := \{0, \ldots, T-1\}$ and $\mathbf{s}_{-n,t} = (s_{m,t})_{m \neq n}$ to denote the joint states of all other agents.
- $\mathcal{A}$ is the finite action space. Similarly, $\Delta(\mathcal{A})$ denotes the probability simplex, and $a_{n,t}$ and $\mathbf{a}_{-n,t}$ refer to actions of n and other agents, respectively.
- $P : \mathcal{S}^N \times \mathcal{A}^N \times \mathcal{S} \to [0,1]$ is the state transition, for which $P(s_{n,t'}|s_{n,t}, a_{n,t}, \mathbf{s}_{-n,t}, \mathbf{a}_{-n,t})$ is the probability of agent n transitioning from $s_{n,t}$ to $s_{n,t'}$ by taking action $a_{n,t}$ given other agents' states $\mathbf{s}_{-n,t}$ and actions $\mathbf{a}_{-n,t}$.
- $R : \mathcal{S}^N \times \mathcal{A}^N \to \mathbb{R}$ is the expected reward at each time (or stage payoff) derived from the immediate reward $r(s_{n,t'}|s_{n,t}, a_{n,t})$ that is more commonly used in MDP

$$R(s_{n,t}, a_{n,t}, \mathbf{s}_{-n,t}, \mathbf{a}_{-n,t}) = \sum_{s_{n,t'} \in \mathcal{S}} r(s_{n,t'}|s_{n,t}, a_{n,t}) P(s_{n,t'}|s_{n,t}, a_{n,t}, \mathbf{s}_{-n,t}, \mathbf{a}_{-n,t}) \tag{1}$$

- $\gamma \in (0,1]$ is the discount factor.
- ρ is the initial state distribution.

Every agent n follows a policy $\pi_n = (\pi_{n,t})_{t \in \mathcal{T}}$, where $\pi_{n,t}(s,a)$ describes the probability of choosing action $a \in \mathcal{A}$ at state $s \in \mathcal{S}$. Accordingly, we denote $\pi_{-n} = (\pi_m)_{n \neq m}$ as the joint policy of all agents other than n. Each agent aims to maximize the expected total discounted payoff function given all other agents' policies:

$$V_n(\pi_n, \pi_{-n}) = \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t R(s_{n,t}, a_{n,t}, \mathbf{s}_{-n,t}, \mathbf{a}_{-n,t})\right], \tag{2}$$

with $s_{m,0} \sim \rho, a_{m,t} \sim \pi_{m,t}(s_{m,t}, \cdot), s_{m,t'} \sim P(\cdot|s_{m,t}, a_{m,t}, \mathbf{s}_{-m,t}, \mathbf{a}_{-m,t}), \forall m \in \mathcal{N}$.

2.2 Definitions and properties

When the number of agents is sufficiently large, the population can be treated as a continuum. The game then becomes a Markov population game (Elokda *et al.*, 2024), and the distribution of agent states is summarized by the population state distribution $\mu = (\mu_t)_{t \in \mathcal{T}}$, where $\mu_t : \mathcal{S} \to \Delta(\mathcal{S})$ is the population state distribution at time t. Besides, the joint state and action $(\mathbf{s}_{-n,t}, \mathbf{a}_{-n,t})$ in Eq. (1) are replaced by (μ_t, π_t) and the following assumption holds.

Assumption 1. *With a large population, the state transition is insensitive to individual policy*

$$\lim_{N \to \infty} \frac{\partial P(s'|s, a, \mu_t, \pi_t)}{\partial \pi_{n,t}(s,a)} = 0, \ \forall s' \in \mathcal{S}, \tag{3}$$

and the dynamics of population state distribution are determined by aggregate actions

$$\mu_t(s') = \frac{1}{N} \sum_{\tau < t} \sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}} P(s'|s, a, \mu_\tau, \pi_\tau) \mu_\tau(s) \sum_{n \in \mathcal{N}} \pi_{n,\tau}(s,a) \tag{4}$$

In this study, we define a special case of Markov population games that intersects with Markov potential games (MPG) (Macua *et al.*, 2018, Leonardos *et al.*, 2021).

Definition 1. *A Markov population game is a population Markov potential game (PMPG) if there exists a potential function* $\Phi(\mu, \pi)$ *such that*

$$\frac{\partial \Phi(\mu, \pi)}{\partial \pi_{n,t}(s, a)} = \gamma^t \mu_t(s) R(s, a, \mu_t, \pi_t), \quad \forall n \in \mathcal{N}, s \in \mathcal{S}, a \in \mathcal{A}. \tag{5}$$

Definition 2. *A joint policy* $\pi^* = (\pi_n^*)_{n \in \mathcal{N}}$ *is a Nash policy if for every agent* n*, it holds that*

$$V_n(\pi_n^*, \pi_{-n}^*) \geq V_n(\pi_n, \pi_{-n}^*), \quad \forall \pi_n \in \Pi, \tag{6}$$

where Π *is the feasible set of individual policies.*

We then establish the existence of Nash policy in PMPG and the equivalence to an optimization problem.

Proposition 1. *For a PMPG with potential function* $\Phi(\mu, \pi)$ *and state dynamics* $f(\mu, \pi) = 0$*, there always exists a Nash policy* π^* *that also solves the optimization problem:*

$$\max_{\pi, \mu} \Phi(\mu, \pi), \quad s.t.\ f(\mu, \pi) = 0,\ \pi_n \in \Pi. \tag{7}$$

We further derive a sufficient condition for a Markov population game to be a PMPG.

Proposition 2. *A Markov population game is a PMPG if any two agents* $m, n \in \mathcal{N}$*, the symmetric condition holds, that is,* $\forall t \in \mathcal{T}, s_m, s_n \in \mathcal{S}, a_m, a_n \in \mathcal{A}$*,*

$$\frac{\partial \mu_t(s_m) R(s_m, a_m, \mu_t, \pi_t)}{\partial \pi_{n,t}(s_n, a_n)} = \frac{\partial \mu_t(s_n) R(s_n, a_n, \mu_t, \pi_t)}{\partial \pi_{m,t}(s_m, a_m)}. \tag{8}$$

2.3 Solution algorithm

Due to Proposition 1, we develop a policy gradient method to solve the equilibrium of PMPG:

- Forward loading: Compute population state distribution μ using the current policy π.
- Policy update: Evaluate the gradient $\nabla_{\pi_n} \Phi(\mu, \pi) = \left(\frac{\partial \Phi(\mu, \pi)}{\partial \pi_{n,t}(s,a)}\right)_{t \in \mathcal{T}, s \in \mathcal{S}, a \in \mathcal{A}}$ as per Eq. (5) and update policy of representative agent according to

$$\pi_n \leftarrow \text{Proj}_{\Pi}\Big[\pi_n + \alpha \nabla_{\pi_n} \Phi(\mu, \pi)\Big], \tag{9}$$

  where α is the step-size and $\text{Proj}_{\Pi}[\cdot]$ denotes a projection on to Π.

As will be shown in the numerical experiments, the above algorithm is much more efficient than standard value iterations for Markov game (e.g., Zhang *et al.*, 2023).

3 APPLICATION

In this section, we present an example of PMPG in ride-hailing. Particularly, we show the ride-hailing vehicle routing game (RIVER) proposed in Zhang *et al.* (2023) is a PMPG and demonstrates the efficiency of our proposed solution algorithm.

RIVER considers a fleet of N ride-hailing vehicles moving across zones $\mathcal{Z}$ over $\mathcal{T}$ discrete time intervals to search and serve passengers. At the beginning of each time interval t, vacant vehicles in zone $i \in \mathcal{Z}$ can move to a neighboring zone $j \in \mathcal{Z}_i$ for passenger search. The aggregate demand-supply interactions in each zone lead to a meeting probability $m_{t,j}$, i.e., how likely a vehicle pick up a passenger in zone j by the end of interval t. Accordingly, the sequential cruising decisions of each vehicle are modeled as an MDP with the state defined as the current zone and action as the next search zone. For vehicle n with $s_{n,t} \in \mathcal{Z}_j, a_{n,t} = j$, the expected stage payoff is $R(s_{n,t}, a_{n,t}, \mu_t, \pi_t) = m_{t,j} \bar{p}_{t,j}$, where $\bar{p}_{t,j}$ is the average trip fare from zone j at time t.

Zhang *et al.* (2023) derived $m_{t,j}$ as a continuously differentiable function of the demand flow $q_{t,j}$ and vacant vehicle flow $y_{t,j}$, and the latter is computed as $y_{t,j} = \sum_{i\in\mathcal{Z}_j} \mu_t(i) \sum_{n\in\mathcal{N}} \pi_{n,t}(i,j)$. Under this setting, we prove RIVER is a PMPG using Proposition 2. Consider two agents with $s_m = i, a_m = j$ and $s_n = k, a_n = l$. If they choose the same search zone (i.e., $j = l$), then

$$\frac{\partial \mu_t(s_m) R(s_m, a_m, \mu_t, \pi_t)}{\partial \pi_{n,t}(s_n, a_n)} = \frac{\partial \mu_t(s_n) R(s_n, a_n, \mu_t, \pi_t)}{\partial \pi_{m,t}(s_m, a_m)} = \frac{\partial m_{t,j}}{\partial y_{t,j}} \mu_t(i)\mu_t(k)\bar{p}_{t,j}. \tag{10}$$

Otherwise, their rewards are independent and the partial derivatives are simply zero.

Furthermore, we show Assumption 1 holds and construct and prove the potential function as

$$\Phi(\mu, \pi) = \sum_{t\in\mathcal{T}} \sum_{j\in\mathcal{Z}} \gamma^t \int_0^{\sum_{i\in\mathcal{Z}_j} \mu_t(i) \sum_{n\in\mathcal{N}} \pi_{n,t}(i,j)} m_{t,j}(y)\bar{p}_{t,j} \mathrm{d}y. \tag{11}$$

See Zhang *et al.* (2023) for the specification of population state dynamics.

Figure 1 illustrates the convergence of two solution algorithms on a simplified scenario with seven zones, 12 time intervals, and 500 drivers. It can be clearly observed that policy gradient achieves a much faster convergence compared to value iterations.

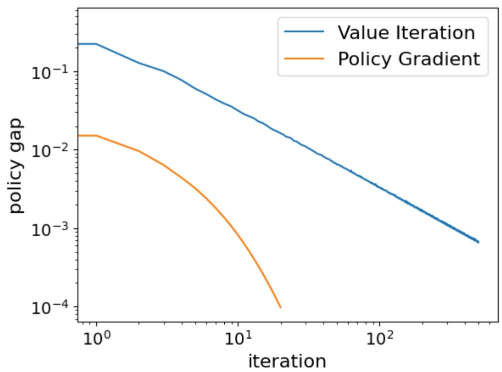

Figure 1 – *Convergence of solution algorithms.*

References

Akamatsu, Takashi. 1997. Decomposition of path choice entropy in general transport networks. *Transportation Science*, **31**(4), 349–362.

Baillon, J-B, & Cominetti, Roberto. 2008. Markovian traffic equilibrium. *Mathematical Programming*, **111**(1), 33–56.

Bell, Michael GH. 1995. Alternatives to Dial's logit assignment algorithm. *Transportation Research Part B: Methodological*, **29**(4), 287–295.

Elokda, Ezzat, Bolognani, Saverio, Censi, Andrea, Dörfler, Florian, & Frazzoli, Emilio. 2024. Dynamic Population Games: A Tractable Intersection of Mean-Field Games and Population Games. *IEEE Control Systems Letters*.

Leonardos, Stefanos, Overman, Will, Panageas, Ioannis, & Piliouras, Georgios. 2021. Global convergence of multi-agent policy gradient in markov potential games. *arXiv preprint arXiv:2106.01969*.

Macua, Sergio Valcarcel, Zazo, Javier, & Zazo, Santiago. 2018. Learning parametric closed-loop policies for Markov potential games. *arXiv e-prints*, arXiv–1802.

Oyama, Yuki, Hara, Yusuke, & Akamatsu, Takashi. 2022. Markovian traffic equilibrium assignment based on network generalized extreme value model. *Transportation Research Part B: Methodological*, **155**, 135–159.

Sandholm, William H. 2001. Potential games with continuous player sets. *Journal of Economic theory*, **97**(1), 81–108.

Sheffi, Yosef. 1985. *Urban transportation networks*. Vol. 6. Prentice-Hall, Englewood Cliffs, NJ.

Zhang, Kenan, Mittal, Archak, Djavadian, Shadi, Twumasi-Boakye, Richard, & Nie, Yu Marco. 2023. RIde-hail vehicle routing (RIVER) as a congestion game. *Transportation Research Part B: Methodological*, **177**, 102819.