# Linearly Solvable Mean-Field Traffic Routing Games*

Takashi Tanaka[1] Ehsan Nekouei[2] Ali Reza Pedram[3] Karl Henrik Johansson[4]

***Abstract*—We consider a dynamic traffic routing game over an urban road network involving a large number of drivers in which each driver selecting a particular route is subject to a penalty that is affine in the logarithm of the number of drivers selecting the same route. We show that the mean-field approximation of such a game leads to the so-called linearly solvable Markov decision process, implying that its mean-field equilibrium (MFE) can be found simply by solving a finite-dimensional linear system backward in time. Based on this backward-only characterization, it is further shown that the obtained MFE has the notable property of strong time-consistency. A connection between the obtained MFE and a particular class of fictitious play is also discussed.**

## I. Introduction

The mean-field game (MFG) theory, introduced by the authors of [3] and [4] almost concurrently, provides a powerful framework to study stochastic dynamic games where (i) the number of players involved in the game is large, (ii) each individual player's impact on the network is infinitesimal, and (iii) players' identities are indistinguishable. The central idea of the MFG theory is to approximate, in an appropriate sense, the original large-population game problem by a single-player optimal control problem, in which individual player's best response to the mean field (average behavior of the population) is analyzed. Typically, the solution to the latter problem is characterized by a pair of backward Hamilton-Jacobi-Bellman (HJB) and forward Fokker-Planck-Kolmogorov (FPK) equations; the HJB equation guarantees player-by-player optimality, while the FPK equation guarantees time consistency of the solution. The coupled HJB-FPK systems, as well as alternative mathematical characterizations (e.g., McKean-Vlasov systems), have been studied extensively [4]–[6].

There has been a recent growth in the literature on MFGs and its applications. MFGs under Linear Quadratic (LQ) [7]–[9] and more general settings [10], [11] are both extensively explored. MFGs with a major agent and a large number of minor agents are studied [10] and applied to design decentralized security defense decisions in a mobile ad hoc network [12]. MFGs with multiple classes of players are investigated in [13]. The authors of [14] studied the existence of robust (minimax) equilibrium in a class of stochastic dynamic games. In [15], the authors analyzed the equilibrium of a hybrid stochastic game in which the dynamics of agents are affected by continuous disturbance as well as random switching signals. Risk-sensitive MFGs were considered in [16]. While continuous-time continuous-state models are commonly used in the references above, [17]–[21] have considered the MFG in discrete-time and/or discrete-state regime. The issues of time inconsistency in MFG and mean-field type optimal control problems are discussed in [22]–[24].

While substantial progress has been made on the MFG literature in recent years, there has been a long history of mean-field-like approaches to large-population games in the transportation research literature [25]. A well-known consequence of a mean-field-like analysis of the traffic user equilibrium is the Wardrop's first principle [26], [27], which provides the following characterization of the traffic condition at an equilibrium: *journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route*. This result, as well as a generalized concept known as *stochastic user equilibrium* (SUE) [28], has played a major role in the transportation research, including the convergence analysis of users' day-to-day routing policy adjustment process [29]–[33]. However, currently only a limited number of results are available connecting the transportation research and recent progress in the MFG theory. The work [20] considers discrete-time discrete-state mean-field route choice games. In [11], the authors modeled the interaction between drivers on a straight road as a non-cooperative game and characterized its MFE. In [34], the authors considered a continuous-time Markov chain to model the aggregated behavior of drivers on a traffic network. A Markovian framework for traffic assignment problems is introduced in [35], which is similar to the problem formulation adopted in this paper. A connection between large-population Markov Decision Processes (MDPs) and MFGs has been discussed in a recent work [36]. MFG has been applied to pedestrian crowd dynamics modeling in [37], [38].

In this paper, we apply the MFG theory to study the strategic behavior of infinitesimal drivers traveling over an urban traffic network. Specifically, we consider a discrete-time dynamic stochastic game wherein, at each intersection, each driver randomly selects one of the outgoing links as her next destination according to a randomized policy. We assume that individual drivers' dynamics are decoupled from each other, while their cost functions are coupled. In particular, we assume that the cost function for each driver is congestion-dependent, and is affine in the logarithm of the number of drivers taking the same route. We regard the congestion-dependent term in

* A preliminary version of this work has been presented at [1]. In the interest of page limitations, proofs of technical results in this paper are partly deferred to [2].

[1,3]University of Texas at Austin, TX, USA. {ttanaka, apedram}@utexas.edu. [2] City University of Hong Kong, Kowloon Tong, Hong Kong. enekouei@cityu.edu.hk. [4]KTH Royal Institute of Technology, Stockholm, Sweden. {kallej}@kth.se.

the cost function as an incentive mechanism (toll charge) imposed by the Traffic System Operator (TSO). Although the assumed structure of cost functionals is restrictive, the purpose of this paper is to show that the considered class of MFGs exhibits a *linearly solvable* nature, and requires somewhat different treatments from the standard MFG formalism. We emphasize that the computational advantages that follow from this special property are notable both from the existing MFG and the transportation research perspectives. Contributions of this paper are summarized as follows:

1) Linear solvability: We prove that the MFE of the game described above is given by the solution to a linearly solvable MDP [39], meaning that it can be computed by performing a sequence of matrix multiplications backward in time *only once*, without any need of forward-in-time computations. This offers a tremendous computational advantage over the conventional characterization of the MFE where there is a need to solve a forward-backward HJB-FPK system, which is often a non-trivial task [5].
2) Strong time-consistency: Due to the backward-only characterization, the MFE in our setting is shown to be *strongly time-consistent* [40], a stronger property than what follows from the standard forward-backward characterization of MFEs.
3) MFE and fictitious play: With an aid of numerical simulation, we show that the derived MFE can be interpreted as a limit point of the belief path of the *fictitious play* process [41] in a scenario where the traffic routing game is repeated.

The rest of the paper is organized as follows: The traffic routing game is set up in Section II and its mean field approximation is discussed in Section III. The linearly solvable MDPs are reviewed in Section IV, which is used to derive the MFE of the traffic routing game in Section V. Time consistency of the derived MFE is studied in Section VI. A connection between MFE and fictitious play is investigated in Section VII. Numerical studies are summarized in Section VIII before we conclude in Section IX.

## II. Problem Formulation

The traffic game studied in this paper is formulated as an $N$-player, $T$-stage dynamic game. Denote by $\mathcal{N} = \{1, 2, \cdots, N\}$ the set of players (drivers) and by $\mathcal{T} = \{0, 1, \cdots, T-1\}$ the set of time steps at which players make decisions.

### A. Traffic graph

The *traffic graph* is a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, ..., V\}$ is the set of nodes (intersections) and $\mathcal{E} = \{1, 2, ..., E\}$ is the set of directed edges (links). For each $i \in \mathcal{V}$, denote by $\mathcal{V}(i) \subseteq \mathcal{V}$ the set of intersections to which there is a directed link from the intersection $i$. At any given time step $t \in \mathcal{T}$, each player is located at an intersection. The node at which the $n$-th player is located at time step $t$ is denoted by $i_{n,t} \in \mathcal{V}$. At every time step, player $n$ at location $i_{n,t}$ selects her next destination $j_{n,t} \in \mathcal{V}(i_{n,t})$. By selecting $j_{n,t}$ at time $t$, the player $n$ moves to the node $j_{n,t}$ at time $t+1$ deterministically (i.e., $i_{n,t+1} = j_{n,t}$).

### B. Routing policy

At every time step $t$, each player selects her next destination according to a randomized routing policy. Let $\Delta^J$ be the $J$-dimensional probability simplex, and $Q^i_{n,t} = \{Q^{ij}_{n,t}\}_{j \in \mathcal{V}(i)} \in \Delta^{|\mathcal{V}(i)|-1}$ be the probability distribution according to which player $n$ at intersection $i$ selects the next destination $j \in \mathcal{V}(i)$. We consider the collection $Q_{n,t} = \{Q^i_{n,t}\}_{i \in \mathcal{V}}$ of such probability distributions as the *policy* of player $n$ at time $t$. For each $n \in \mathcal{N}$ and $t \in \mathcal{T}$, notice that $Q_{n,t} \in \mathcal{Q}$, where

$$\mathcal{Q} = \left\{ \{Q^i\}_{i \in \mathcal{V}} : Q^i \in \Delta^{|\mathcal{V}(i)|-1} \ \ \forall i \in \mathcal{V} \right\}$$

is the space of admissible policies. Suppose that the initial locations of players $\{i_{n,0}\}_{n \in \mathcal{N}}$ are independent and identically distributed random variables with $P_{n,0} = P_0 \in \Delta^{|\mathcal{V}|-1}$. Note that if the policy $\{Q_{n,t}\}_{t \in \mathcal{T}}$ of player $n$ is fixed, then the probability distribution $P_{n,t} = \{P^i_{n,t}\}_{i \in \mathcal{V}}$ of her location at time $t$ is computed recursively by

$$P^j_{n,t+1} = \sum_i P^i_{n,t} Q^{ij}_{n,t} \ \ \forall t \in \mathcal{T}, j \in \mathcal{V}. \tag{1}$$

If $(i_{n,t}, j_{n,t})$ is the location-action pair of player $n$ at time $t$, it has the joint distribution $P^i_{n,t} Q^{ij}_{n,t}$. We assume that location-action pairs $(i_{n,t}, j_{n,t})$ and $(i_{m,t}, j_{m,t})$ for two different players $m \neq n$ are drawn independently under individual policies $\{Q_{n,t}\}_{t \in \mathcal{T}}$ and $\{Q_{m,t}\}_{t \in \mathcal{T}}$. With a slight abuse of notation, we sometimes write $Q_n := \{Q_{n,t}\}_{t \in \mathcal{T}}$ for simplicity.

### C. Cost functional

We assume that, at each time step, the cost functional for each player has two components as specified below:

*1) Travel cost:* For each $i \in \mathcal{V}$, $j \in \mathcal{V}(i)$ and $t \in \mathcal{T}$, let $C^{ij}_t$ be a given constant representing the cost (e.g., fuel cost) for every player selecting $j$ at location $i$ at time $t$.

*2) Tax cost:* We assume that players are also subject to individual and time-varying tax penalties calculated by the TSO. The tax charged to player $n$ at time step $t$ depends not only on her own location-action pair at $t$, but also on the behavior of the entire population at that time step. Specifically, we consider the log-population tax mechanism, where the tax charged to player $n$ taking action $j$ at location $i$ at time $t$ is

$$\pi^{ij}_{N,t,n} = \alpha \left( \log \frac{K^{ij}_{N,t}}{K^i_{N,t}} - \log R^{ij}_t \right). \tag{2}$$

Here, $\alpha > 0$ is a fixed constant characterizing the "aggressiveness" of the tax mechanism. In (2), $K^i_{N,t}$ is the number of players (including player $n$) who are located at the intersection $i$ at time $t$. Likewise, $K^{ij}_{N,t}$ is the number of players (including player $n$) who takes the action $j$ at the intersection $i$ at time $t$. The parameters $R^{ij}_t > 0$ are fixed constants satisfying $\sum_j R^{ij}_t = 1$ for all $i$. We interpret $R^{ij}_t$ as the "reference" routing policy specified by the TSO in advance. Notice that (2) indicates that agent $n$ receives a positive reward by taking action $j$ at location $i$ at time $t$ if $K^{ij}_{N,t}/K^i_{N,t} < R^{ij}_t$ (i.e., the realization of the traffic flow is below the designated congestion level), while she is penalized by doing so if $K^{ij}_{N,t}/K^i_{N,t} > R^{ij}_t$. Since $K^i_{N,t}$ and $K^{ij}_{N,t}$

are random variables, $\pi^{ij}_{N,t,n}$ is also a random variable. We assume that the TSO is able to observe $K^i_{N,t}$ and $K^{ij}_{N,t}$ at every time step so that $\pi^{ij}_{N,t,n}$ is computable.[1] In what follows, we assume that each player is risk neutral. That is, each player is interested in choosing a policy that minimizes the expected sum of travel and tax costs incurred over the planning horizon $\mathcal{T}$. For player $n$ whose location-action pair at time step $t$ is $(i,j)$, the expected tax cost incurred at that time step can be expressed as

$$\Pi^{ij}_{N,n,t} \triangleq \mathbb{E}\left[\pi^{ij}_{N,n,t} \mid i_{n,t}=i, j_{n,t}=j\right]. \tag{3}$$

As we detail in [2, Appendix A, equation (22)], for each location-action pair $(i,j)$, $\Pi^{ij}_{N,n,t}$ can be expressed in terms of $Q_{-n} \triangleq \{Q_m\}_{m\neq n}$. The fact that $\Pi^{ij}_{N,n,t}$ does not depend on player $n$'s own policy will be used to analyze the optimal control problem (5) below.[2]

### D. Traffic routing game

Overall, the cost functional to be minimized by the $n$-th player in the considered game is given by

$$J\left(Q_n, Q_{-n}\right) = \sum_{t=0}^{T-1}\sum_{i,j} P^i_{n,t} Q^{ij}_{n,t}\left(C^{ij}_t + \Pi^{ij}_{N,n,t}\right). \tag{4}$$

Notice that this quantity depends not only on the $n$-th player's own policy $Q_n$ but also on the other players' policies $Q_{-n}$ through the term $\Pi^{ij}_{N,n,t}$. Equation (4) defines an $N$-player dynamic game, which we call the *traffic routing game* hereafter. We introduce the following equilibrium concepts.

**Definition 1:** The $N$-tuple of strategies $\{Q^*_n\}_{n\in\mathcal{N}}$ is said to be a *Nash equilibrium* if the inequality $J\left(Q_n, Q^*_{-n}\right) \geq J\left(Q^*_n, Q^*_{-n}\right)$ holds for each $n \in \mathcal{N}$ and $Q_n$.

**Definition 2:** The $N$-tuple of strategies $\{Q^*_n\}_{n\in\mathcal{N}}$ is said to be *symmetric* if $Q^*_1 = Q^*_2 = \cdots = Q^*_N$.

**Remark 1:** The $N$-player game described above is a *symmetric game* in the sense of [42]. Thus, [42, Theorem 3] is applicable to show that it has a symmetric Nash equilibrium.

**Remark 2:** We assume that players are able to compute a Nash equilibrium strategy $\{Q^*_n\}_{n\in\mathcal{N}}$ prior to the execution of the game based on the public knowledge $\mathcal{G}, \alpha, \mathcal{N}, \mathcal{T}, R^{ij}_t, C^{ij}_t$ and $P_0$. Often the case, it is favorable that a Nash equilibrium is *time-consistent* in that no player is given an incentive to deviate from the precomputed equilibrium routing policy after observing real-time data (such as $K^i_{N,t}$ and $K^{ij}_{N,t}$). In Section VI, we discuss a notable time consistency property of an equilibrium of the traffic routing game formulated above in the large-population limit $N \to \infty$.

## III. Mean Field Approximation

In the remainder of this paper, we are concerned with the large-population limit $N \to \infty$ of the traffic routing game.

[1] Whenever $\pi^{ij}_{N,t,n}$ is computed, we have both $K^{ij}_{N,t} \geq 1$ and $K^i_{N,t} \geq 1$ since at least player $n$ herself is counted. Hence (2) is well-defined.

[2] Although the value of $\Pi^{ij}_{N,n,t}$ for each $(i,j)$ cannot be altered by player $n$'s policy, she can minimize the total cost by an appropriate route choice (e.g., by avoiding links with high toll fees).

**Definition 3:** A set of strategies $\{Q^*_n\}_{n\in\mathcal{N}}$ is said to be an *MFE* if the following conditions are satisfied.

(a) It is symmetric, i.e., $Q^*_1 = Q^*_2 = \cdots = Q^*_N$.

(b) There exists a sequence $\epsilon_N$ satisfying $\epsilon_N \searrow 0$ as $N \to \infty$ such that for each $n \in \mathcal{N} = \{1, 2, ..., N\}$ and $Q_n$, the inequality $J(Q_n, Q^*_{-n}) + \epsilon_N \geq J(Q^*_n, Q^*_{-n})$ holds.

Now, we derive a condition that an MFE must satisfy by analyzing player $n$'s best response when all other players adopt a homogeneous routing policy $Q^* = \{Q^*_t\}_{t\in\mathcal{T}}$. Since $Q^*$ is adopted by all players other than $n$, the probability that a specific player $m(\neq n)$ is located at $i$ is given by $P^{i*}_t$, where $P^* = \{P^*_t\}_{t\in\mathcal{T}}$ is computed recursively by

$$P^{j*}_{t+1} = \sum_i P^{i*}_t Q^{ij*}_t \quad \forall j \in \mathcal{V}.$$

Player $n$'s best response is characterized by the solution to the following optimal control problem:

$$\min_{\{Q_t\}_{t\in\mathcal{T}}} \sum_{t=0}^{T-1}\sum_{i,j} P^i_t Q^{ij}_t\left(C^{ij}_t + \Pi^{ij}_{N,n,t}\right). \tag{5}$$

Here, we note that $\Pi^{ij}_{N,n,t}$ is fully determined by the homogeneous policy $Q^*$ adopted by all other players. (The detail is shown in [2, Appendix A, equation (23)].) In (5), we wrote $P_t$ and $Q_t$ in place of $P_{n,t}$ and $Q_{n,t}$ to simplify the notation.

To analyze player $n$'s best response when $N \to \infty$, we compute the quantity $\lim_{N\to\infty} \Pi^{ij}_{N,n,t}$ as follows:

**Lemma 1:** Let $\Pi^{ij}_{N,n,t}$ be defined by (3). If $Q_{m,t} = Q^*_t$ for all $m \neq n$ and $P^{i*}_t Q^{ij*}_t > 0$, then

$$\lim_{N\to\infty} \Pi^{ij}_{N,n,t} = \alpha \log \frac{Q^{ij*}_t}{R^{ij}_t}.$$

**Proof:** [2, Appendix B]. ∎

Intuitively, Lemma 1 shows that the optimal control problem (5) when $N$ is large is "close to" the optimal control problem:

$$\min_{\{Q_t\}_{t\in\mathcal{T}}} \sum_{t=0}^{T-1}\sum_{i,j} P^i_t Q^{ij}_t\left(C^{ij}_t + \alpha \log \frac{Q^{ij*}_t}{R^{ij}_t}\right). \tag{6}$$

In order for the policy $Q^*$ to constitute an MFE, the policy $Q^*$ itself needs to be the best response by player $n$. In particular, $Q^*$ must solve the optimal control problem (6). That is, the following fixed point condition must be satisfied:

$$Q^* \in \arg\min_{\{Q_t\}_{t\in\mathcal{T}}} \sum_{t=0}^{T-1}\sum_{i,j} P^i_t Q^{ij}_t\left(C^{ij}_t + \alpha \log \frac{Q^{ij*}_t}{R^{ij}_t}\right). \tag{7}$$

In the next two sections, we show that the condition (7) is closely related to the class of optimal control problems known as *linearly-solvable MDPs* [39], [43]. Based on this observation, we show that an MFE can be computed efficiently.

## IV. Linearly Solvable MDPs

In this section, we review linearly-solvable MDPs [39], [43] and their solution algorithms. For each $t \in \mathcal{T}$, let $P_t$ be the probability distribution over $\mathcal{V}$ that evolves according to

$$P^j_{t+1} = \sum_i P^i_t Q^{ij}_t \quad \forall j \in \mathcal{V} \tag{8}$$

with the initial state $P_0$. We assume $C_t^{ij}$, $R_t^{ij}$ for each $t \in \mathcal{T}, i \in \mathcal{V}, j \in \mathcal{V}$ and $\alpha$ are given positive constants. Consider the $T$-step optimal control problem:

$$\min_{\{Q_t\}_{t\in\mathcal{T}}} \sum_{t=0}^{T-1} \sum_{i,j} P_t^i Q_t^{ij} \left( C_t^{ij} + \alpha \log \frac{Q_t^{ij}}{R_t^{ij}} \right). \tag{9}$$

The logarithmic term in (9) can be written as the Kullback–Leibler (KL) divergence from the reference policy $R_t^{ij}$ to the selected policy $Q_t^{ij}$. For this reason (9) is also known as the *KL control* problem [44]. Notice the similarity and difference between the optimal control problems (6) and (9); in (6) the logarithmic term is a fixed constant ($Q^*$ is given), while in (9) the logarithmic term depends on the chosen policy $Q$. To solve (9) by backward dynamic programming, for each $t \in \mathcal{T}$, introduce the value function:

$$V_t(P_t) \triangleq \min_{\{Q_\tau\}_{\tau=t}^{T-1}} \sum_{\tau=t}^{T-1} \sum_{i,j} P_\tau^i Q_\tau^{ij} \left( C_\tau^{ij} + \alpha \log \frac{Q_\tau^{ij}}{R_\tau^{ij}} \right)$$

and the associated Bellman equation

$$V_t(P_t) = \min_{Q_t} \Big\{ \sum_{i,j} P_t^i Q_t^{ij} \left( C_t^{ij} + \alpha \log \frac{Q_t^{ij}}{R_t^{ij}} \right) + V_{t+1}(P_{t+1}) \Big\} \tag{10}$$

with the terminal condition $V_T(\cdot) = 0$. The next theorem states that the Bellman equation (10) can be linearized by a change of variables (the Cole-Hopf transformation), and thus the optimal control problem (9) is reduced to solving a linear system [39].

**Theorem 1:** Let $\{\phi_t\}_{t\in\mathcal{T}}$ be the sequence of $V$-dimensional vectors defined by the backward recursion

$$\phi_t^i = \sum_j R_t^{ij} \exp\left( -\frac{C_t^{ij}}{\alpha} \right) \phi_{t+1}^j \quad \forall i \in \mathcal{V} \tag{11}$$

with the terminal condition $\phi_T^i = 1 \ \forall i$. Then, for each $t = 0, 1, \cdots, T$ and $P_t$, the value function can be written as

$$V_t(P_t) = -\alpha \sum_i P_t^i \log \phi_t^i. \tag{12}$$

Moreover, the optimal policy for (9) is given by

$$Q_t^{ij*} = \frac{\phi_{t+1}^j}{\phi_t^i} R_t^{ij} \exp\left( -\frac{C_t^{ij}}{\alpha} \right). \tag{13}$$

**Proof:** [2, Appendix C]. ∎

We stress that (11) is linear in $\phi$ and can be computed by matrix multiplications backward in time.

## V. Mean Field Equilibrium

In this section, we investigate the relationship between the optimal control problem (9) and the fixed point condition (7) for an MFE in the traffic routing game. To this end, we introduce the value function for the optimal control problem (6), defined by

$$\tilde{V}_t(P_t) \triangleq \min_{\{Q_\tau\}_{\tau=t}^{T-1}} \sum_{\tau=t}^{T-1} \sum_{i,j} P_\tau^i Q_\tau^{ij} \left( C_\tau^{ij} + \alpha \log \frac{Q_\tau^{ij*}}{R_\tau^{ij}} \right)$$

The value function satisfies the Bellman equation:

$$\tilde{V}_t(P_t) = \min_{Q_t} \Big\{ \sum_{i,j} P_t^i Q_t^{ij} \left( C_t^{ij} + \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} \right) + \tilde{V}_{t+1}(P_{t+1}) \Big\} \tag{14}$$

with the terminal condition $\tilde{V}_T(\cdot) = 0$. We emphasize the distinction between $\tilde{V}_t(\cdot)$ and $V_t(\cdot)$. As in the previous section, $V_t(\cdot)$ is the value function associated with the KL control problem (9), whereas $\tilde{V}_t(\cdot)$ is the value function associated with the optimal control problem (6). Despite this difference, the next lemma shows an intimate connection between $V_t(\cdot)$ and $\tilde{V}_t(\cdot)$. In particular, if the parameter $Q^*$ in (6) is chosen to be the solution to the KL control problem (9), then the objective function in (6) becomes a constant that does not depend on the decision variable $\{Q_t\}_{t\in\mathcal{T}}$ (the *equalizer property*[3] of the optimal KL control policy). Moreover, under this circumstance, the value function $\tilde{V}_t(\cdot)$ for (6) coincides with the value function $V_t(\cdot)$ for the KL control problem (9).

**Lemma 2:** If $\{Q_t^*\}_{t\in\mathcal{T}}$ in (6) is fixed to be the solution to the KL control problem (9), then an arbitrary policy $\{Q_t\}_{t\in\mathcal{T}}$ with $Q_t \in \mathcal{Q}$ is an optimal solution to (6). Moreover, for each $t \in \mathcal{T}$ and $P_t$, we have

$$\tilde{V}_t(P_t) = -\alpha \sum\nolimits_i P_t^i \log \phi_t^i \tag{15}$$

where $\{\phi_t\}_{t\in\mathcal{T}}$ is the sequence calculated by (11).

**Proof:** We show (15) by backward induction. If $t = T$, the claim trivially holds due to the definition $\tilde{V}_T(P_T) = 0$ and the fact that the terminal condition for (11) is given by $\phi_T^i = 1$. Thus, for $0 \le t \le T-1$, assume that

$$\tilde{V}_{t+1}(P_{t+1}) = -\alpha \sum_j P_{t+1}^j \log \phi_{t+1}^j$$

holds. Using $\rho_t^{ij} = C_t^{ij} - \alpha \log \phi_{t+1}^j$, the Bellman equation (14) can be written as

$$\tilde{V}_t(P_t) = \min_{Q_t} \sum_{i,j} P_t^i Q_t^{ij} \left( \rho_t^{ij} + \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} \right). \tag{16}$$

Substituting $Q_t^{ij*}$ obtained by (13) into (16), we have

$$\tilde{V}_t(P_t) = \min_{Q_t} \sum\nolimits_{i,j} P_t^i Q_t^{ij} \left( -\alpha \log \phi_t^i \right) \tag{17a}$$

$$= \min_{Q_t} \sum\nolimits_i P_t^i \left( -\alpha \log \phi_t^i \right) \underbrace{\sum\nolimits_j Q_t^{ij}}_{=1} \tag{17b}$$

$$= -\alpha \sum\nolimits_i P_t^i \log \phi_t^i. \tag{17c}$$

This completes the proof of (15). The chain of equalities (17) also shows that the decision variable $Q_t$ vanishes in the "min" operator, indicating that any $Q_t \in \mathcal{Q}$ is a minimizer. This shows the equalizer property of $\{Q_t^*\}_{t\in\mathcal{T}}$. ∎

Lemma 2 provides the following insights into the MFE of the traffic routing game: Suppose that all the players except the player $n$ adopt the policy $Q^*$ (the optimal solution to (9)) and the number of players tends to infinity. Since $Q^*$ will equalize

[3]We note that the equalizer property (the term borrowed from [45]) of the minimizers of free energy functions is well-known in statistical mechanics, information theory, and robust Bayes estimation theory.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.2986195, IEEE Transactions on Automatic Control



the costs of all alternative routing policies for player $n$, any routing policy will be a best response for her. In particular, this means that the policy $Q^*$ itself will also be one of the best responses, and thus the fixed point condition (7) will be satisfied. Therefore, $Q^*$ will be an MFE of the considered traffic routing game. The following theorem, which is the main result of this paper, confirms this intuition.

**Theorem 2:** A symmetric strategy profile $Q^{ij}_{n,t} = Q^{ij*}_t$ for each $n \in \mathcal{N}, t \in \mathcal{T}$ and $i, j \in \mathcal{V}$, where $Q^{ij*}_t$ is obtained by (11)–(13), is an MFE of the traffic routing game.

**Proof:** [2, Appendix D]. ∎

Theorem 2, together with Theorem 1, provides an efficient algorithm for computing an MFE of the traffic routing game presented in Section II. In particular, we remark that the MFE can be computed by the backward-in-time recursion (11)–(13). This is in stark contrast to the standard MFG formalism in which a coupled pair of forward and backward equations must be solved to obtain an MFE.

Finally, we remark that the equalizer property of the MFE $Q^*$ characterized by Lemma 2 is a reminiscent of the *Wardrop's first principle*, stating that costs are equal on all the routes used at the equilibrium. Although the costs usually mean journey times in the literature around Wardrop's principles [26], [27], the cost in our setting is the sum of the travel costs and the tax costs as stated in (4). In this sense, Lemma 2 can be viewed as an extension of the standard description of the Wardrop's first principle.

## VI. Weak and strong time consistency

This short section presents another notable property of the MFE derived in the previous section. Let $Q_{n,t} = Q_t$ for each $n \in \mathcal{N}$ and $0 \le t \le T-1$ be a symmetric strategy profile, and $P_t$ be the probability distribution over $\mathcal{V}$ induced by $Q_t$ as in (8). For every time step $0 \le t \le T-1$, a dynamic game restricted to the time horizon $\{t, t+1, ..., T-1\}$ with the initial condition $P_t$ is called the *subgame* of the original game. The following are natural extensions of the *strong and weak time consistency* concepts in the dynamic game theory [40] to MFGs.

**Definition 4:** An MFE strategy profile $Q^*$ is said to be:

1) *weakly time-consistent* if for every $0 \le t \le T-1$, $\{Q^*_s\}_{t \le s \le T-1}$ constitutes an MFE of the subgame restricted to $\{t, t+1, ..., T-1\}$ when $\{Q_s\}_{0 \le s \le t-1} = \{Q^*_s\}_{0 \le s \le t-1}$.
2) *strongly time-consistent* of for every $0 \le t \le T-1$, $\{Q^*_s\}_{t \le s \le T-1}$ constitutes an MFE of the subgame restricted to $\{t, t+1, ..., T-1\}$ when regardless of the policy $\{Q_s\}_{0 \le s \le t-1}$ implemented in the past.

In the standard MFG formalism [3], [4] where the MFE is characterized by a forward-backward HJB-FPK system, the equilibrium policy is only weakly time-consistent in general. This is because, in the event of $P_t$ not being consistent with the distribution induced by $\{Q^*_s\}_{0 \le s \le t-1}$, the MFE of the subgame must be recalculated by solving the HJB-FPK system over $t \le s \le T-1$. In contrast, the MFE considered in this paper is characterized only by a backward equation (Theorems 1 and 2). A notable consequence of this fact is that

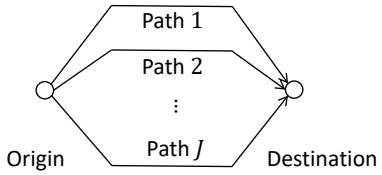


Fig. 1. Simple path choice problem.

even if the initial condition $P_t$ is inconsistent with the planned distribution, it does not alter the fact that $\{Q^*_s\}_{t \le s \le T-1}$ constitutes an MFE of the subgame restricted to $t \le s \le T-1$. Therefore, the MFE characterized by Theorems 1 and 2 is strongly time-consistent.

## VII. Mean field equilibrium and fictitious play

The equalizer property of the MFE characterized by Lemma 2 raises the following question regarding the stability of the equilibrium: If the MFE equalizes the costs of all the available route selection policies, what incentivizes individual players to stay at the MFE policy $Q^*$? In this section, we reason about the stability of MFE by relating it with the convergence of the *fictitious play* process [41] for an associated repetitive game. Convergence of fictitious play processes have been studied in depth in [41] and [46]. We also remark that fictitious play for day-to-day policy adjustments for traffic routing has been considered in [47], [48]. Fictitious play in the context of MFGs has been studied in the recent work [49].

Consider the situation in which the traffic routing game is repeated on a daily basis, and individual players update their routing policies based on their past experiences. For simplicity, we only consider a single-origin-single-destination, $N$-player traffic routing game shown in Figure 1. We assume that there are $J$ parallel routes from the origin to the destination. All players are initially located at the origin node. Each route $j$ is associated with the travel cost $c^j$ and the tax cost $\alpha \log \frac{K^j_N}{NR^j}$, where $K^j_N$ is the number of players selecting route $j$. As before, $\alpha, C^j, R^j$ are given constants. By *fictitious play*, we mean the following day-to-day policy adjustment mechanism for individual players: On day one, each player $n \in \mathcal{N}$ makes initial guesses on player $m$'s mixed strategies (for all $m \ne n$) for their route selection. Player $n$'s belief on player $m$'s policy is demoted by $Q_{n \to m}[1] \in \Delta^{J-1}$. Assuming that $Q_{n \to m}[1], \forall m \ne n$ are fixed, player $n$ selects a route with the lowest expected cost. Player $n$'s route selection is observed and recorded by all players at the end of day one. On day $\ell$, player $n$'s belief $Q_{n \to m}[\ell] \in \Delta^{J-1}$ for each $m \ne n$ is set to be equal to the vector of observed empirical frequencies of player $m$'s route choices up to day $\ell - 1$. The process is repeated on a daily basis. We call $Q_{n \to m}[\ell]$, $\ell = 1, 2, \cdots$ for all pairs $m \ne n$ the *belief paths*. The process is summarized in Algorithm 1.

In what follows, we show that Algorithm 1 converges to a unique symmetric Nash equilibrium of the $N$-player game shown in Figure 1 if the initial belief is symmetric. A numerical simulation is presented in Section VIII-B to

**Algorithm 1:** The fictitious play process for the simplified traffic routing game.

**Step 0:** On day one, each player $n$ initializes a mixed strategy in belief $Q_{n\to m}[1] \in \Delta^{J-1}$ for each $m \neq n$ according to which she believes player $m$ select routes.

**Step 1:** At the beginning of day $\ell$, each player $n$ fixes assumed mixed strategy $Q_{n\to m}[\ell] \in \Delta^{J-1}$ according to which she believes player $m$ select routes. Based on this assumption, she selects her best response $r_n[\ell] = \arg\min_j y_n^j[\ell]$, where $y_n^j[\ell]$ is the assumed cost of selecting route $j$, i.e.,

$$y_n^j[\ell] = \mathbb{E}\left(C^j + \alpha \log \frac{K_N^j}{NR^j}\right). \tag{18}$$

**Step 2:** At the end of day $k$, each player $n$ updates her belief based on observations $r_m[\ell], m \neq n$ by

$$Q_{n\to m}[\ell+1] = \frac{\ell}{\ell+1} Q_{n\to m}[\ell] + \frac{1}{\ell+1}\delta(r_m[\ell]) \tag{19}$$

where $\delta(r)$ is the indicator vector whose $r$-th entry is one and all other entries are zero. Return to Step 1.

demonstrate this convergence behavior, where we also observe that the policy obtained in the limit of the belief path is closely approximated by the MFE if $N$ is sufficiently large. This observation provides the MFE with an interpretation as a steady-state value of the players' day-to-day belief adjustment processes in a large population traffic routing game.

Convergence of Algorithm 1 is a straightforward consequence of Monderer and Shapley [41], where it is shown that every belief path for $N$-player games with identical payoff functions converges to an equilibrium. This result is directly applicable to the $N$-player traffic routing game shown in Figure 1 since it is clearly a symmetric game. The only caveat is that there is no guarantee that the belief path converges to a symmetric equilibrium if the game has multiple Nash equilibria (including non-symmetric ones). However, this difficulty can be circumvented if we impose an additional assumption that the initial belief is symmetric, i.e., $Q_{n\to m}[1] = Q[1]$ for some $Q[1] \in \Delta^{J-1}$ for all $(m,n)$ pairs. If the initial belief is symmetric, the belief path generated by Algorithm 1 remains symmetric, i.e., $Q_{n\to m}[\ell] = Q[\ell]$, $y_n[\ell] = y[\ell]$ and $r_n[\ell] = r[\ell]$ for $\ell \geq 1$. In this case, equations (18) and (19) are simplified to

$$y^i[\ell] = C^j + \alpha \sum_{k=0}^{N-1} \log\left(\frac{k+1}{NR^j}\right)\binom{N-1}{k} \times (Q^j[\ell])^k (1-Q^j[\ell])^{N-1-k}$$

and

$$Q[\ell+1] = \frac{\ell}{\ell+1} Q[\ell] + \frac{1}{\ell+1}\delta(r[\ell])$$

respectively. Combined with the convergence result by Monderer and Shapley [41], it can be concluded that every limit point of the belief path generated by Algorithm 1 with symmetric initial belief is a symmetric equilibrium.

The next lemma shows that there exists a unique symmetric equilibrium in the simple traffic routing game in Figure 1 with finite number of players.

**Lemma 3:** There exists a unique symmetric equilibrium, denoted by $Q^{(N)*}$, in the $N$-player traffic routing game shown in Figure 1.

**Proof:** [2, Appendix E]. ∎

Now, consider the limit $\lim_{N\to\infty} Q^{(N)*}$ and its relationship with the MFE $Q^*$. Notice that the MFE of the traffic routing game in Figure 1 is characterized as the unique solution to the following convex optimization problem:

$$\min_{Q\in\Delta^{J-1}} \sum_{j=1}^{J} Q^j \left(C^j + \alpha \log \frac{Q^j}{R^j}\right). \tag{20}$$

In Section VIII-B, we perform a simulation study where we observe that $Q^*$ is a good approximation of $Q^{(N)*}$ when $N$ is sufficiently large. Although the condition under which the identity $\lim_{N\to\infty} Q^{(N)*} = Q^*$ holds must be studied carefully in the future,[4] this observation suggests an important practical interpretation of the MFE: namely, it is an approximation of the limit point of the belief path (or equivalently, the empirical frequency of each player to take particular routes) of the symmetric fictitious play when $N$ is large. This provides an answer to the question regarding the stability of the MFE raised in the beginning of this section.

## VIII. Numerical Illustration

In this section, we present numerical simulations that illustrate the main results obtained in Sections V and VII.

### A. Traffic routing game and congestion control

We first illustrate the result of Theorem 2 applied to a traffic routing game shown in Fig. 2. At $t = 0$, the population is concentrated in the origin cell (indicated by "O"). For $t \in \mathcal{T}$, the travel cost for each player is

$$C_t^{ij} = \begin{cases} C_{\text{term}} & \text{if } j = i \\ 1 + C_{\text{term}} & \text{if } j \in \mathcal{V}(i) \\ 100000 + C_{\text{term}} & \text{if } j \notin \mathcal{V}(i) \text{ or } j \text{ is an obstacle} \end{cases}$$

where $\mathcal{V}(i)$ contains the north, east, south, and west neighborhood of the cell $i$. To incorporate the terminal cost, we introduce $C_{\text{term}} = 0$ if $t = 0, 1, \cdots, T-1$ and $C_{\text{term}} = 10\sqrt{\text{dist}(j, \text{D})}$ if $t = T-1$, where $\text{dist}(j, \text{D})$ is the Manhattan distance between the player's final location $j$ and the destination cell (indicated by "D"). As the reference distribution, we use $R_t^{ij} = 1/|\mathcal{V}(i)|$ (uniform distribution) for each $i \in \mathcal{V}$ and $t \in \mathcal{T}$ to incentivize players to spread over the traffic graph.

For various values of $\alpha > 0$, the backward formula (11) is solved and the optimal policy is calculated by (13). If $\alpha$ is small (e.g., $\alpha = 0.1$), it is expected that players will take the

[4]While Lemma 3 establishes the uniqueness of the symmetric equilibrium for a simple traffic routing game shown in Figure 1, its extension to the general class of traffic routing game formulated in Section II is currently unknown. The proof of the identity $\lim_{N\to\infty} Q^{(N)*} = Q^*$ (for both the simple game in Figure 1 and the general setup in Section II) must be postponed as future work.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.2986195, IEEE Transactions on Automatic Control



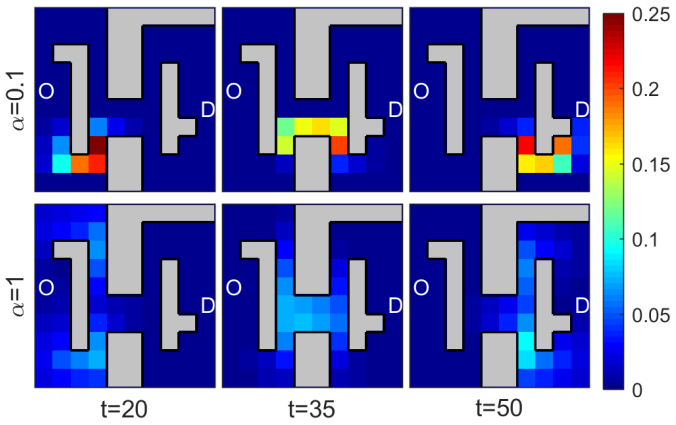


Fig. 2. Mean-field traffic routing game with $T = 70$ over a traffic graph with 100 nodes (grid world with obstacles). Plots show vehicle distribution $P_t$ at $t = 20, 35, 50$ and for $\alpha = 0.1$ and 1.

shortest path since the action cost is dominant compared to the tax cost (2). This is confirmed by numerical simulation; three figures in the top row of Fig. 2 show snapshots of the population distribution at time steps $t = 20, 35$ and 50. In the bottom row, similar plots are generated with a larger $\alpha$ ($\alpha = 1$). In this case, it can be seen that the equilibrium strategy will choose longer paths with higher probability to reduce congestion.

### B. Symmetric fictitious play

Next, we present a numerical demonstration of the symmetric fictitious play studied in Section VII. Consider a simple traffic graph in Figure 1 with three alternative paths ($J = 3$). We set travel costs $(C^1, C^2, C^3) = (2, 1, 3)$, while fixing $R^1 = R^2 = R^3 = 1/3$ and $\alpha = 1$. Figure 3 shows the belief path generated by the policy update rule (19) with the initial policy $Q[1] = (1/3, 1/3, 1/3)$. The left shows the case with 20 players ($N = 20$), while the right plot shows the case with $N = 200$. The MFE

$$Q^* = \frac{1}{\sum_j R^j \exp(-c^j)} \begin{bmatrix} R^1 \exp(-c^1) \\ R^2 \exp(-c^2) \\ R^3 \exp(-c^3) \end{bmatrix} = \begin{bmatrix} 0.245 \\ 0.665 \\ 0.090 \end{bmatrix}$$

is also shown in each plot. The plot for $N = 20$ shows that, while the belief path is convergent, there is a certain offset between its limit point and the MFE. This is because the number of players is not sufficiently large. On the other hand, when $N = 200$, the MFE $Q^*$ is a good approximate to the limit point of the belief path.

## IX. Conclusion and Future Work

In this paper, we showed that the MFE of a large-population traffic routing game under the log-population tax mechanism can be obtained via the linearly solvable MDP. Strong time consistency of the derived MFE was discussed. A connection between the MFE and fictitious play was investigated.

While this paper is restricted discrete-time discrete-state formalisms, its continuous-time continuous-state counterpart is worth investigating in the future. The interface between the existing traffic SUE theory [25] and MFG must be thoroughly studied in the future work. Convergence of fictitious play and its relationship with MFE presented in Section VII should be studied in more general settings. Linear solvability renders the proposed MFG framework attractive as an incentive mechanism for TSOs for the purpose of traffic congestion mitigation; however, questions from the perspectives of mechanism design theory, such as how to tune parameters $\alpha$ and $R$ (which are assumed given in this paper) to balance the efficiency and budget, are unexplored. Finally, generalization to non-homogeneous MFGs with multiple classes of players (which was recently studied in [50]) needs further investigation.

## Acknowledgment

The authors would like to thank Mr. Matthew T. Morris and Mr. James S. Stanesic at the University of Texas at Austin for their contributions to the numerical study in Section VIII. The first author also acknowledges valuable discussions with Dr. Tamer Başar at the University of Illinois at Urbana-Champaign.

## References

[1] T. Tanaka, E. Nekouei, and K. H. Johansson, "Linearly solvable mean-field road traffic games," *56th Annual Allerton Conference on Communication, Control, and Computing*, 2018.

[2] T. Tanaka, E. Nekouei, A. R. Pedram, and K. H. Johansson, "Linearly solvable mean-field traffic routing games," *arXiv preprint arXiv:1903.01449*, 2019.

[3] P. E. Caines, M. Huang, and R. Malhamé, "Mean field games," *in Handbook of Dynamic Game Theory (T. Başar, G. Zaccour, Eds.)*, 2018.

[4] J.-M. Lasry and P.-L. Lions, "Mean field games," *Japanese journal of mathematics*, vol. 2, no. 1, pp. 229–260, 2007.

[5] Y. Achdou and I. Capuzzo-Dolcetta, "Mean field games: Numerical methods," *SIAM Journal on Numerical Analysis*, vol. 48, no. 3, pp. 1136–1162, 2010.

[6] R. Carmona and F. Delarue, "Probabilistic analysis of mean-field games," *SIAM Journal on Control and Optimization*, vol. 51, no. 4, pp. 2705–2734, 2013.

[7] M. Huang, "Large-population LQG games involving a major player: The Nash certainty equivalence principle," *SIAM Journal on Control and Optimization*, vol. 48, no. 5, pp. 3318–3353, 2010.

[8] J. Huang, X. Li, and T. Wang, "Mean-field Linear-Quadratic-Gaussian (LQG) games for stochastic integral systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 9, pp. 2670–2675, Sept 2016.

[9] J. Moon and T. Basar, "Linear quadratic risk-sensitive and robust mean field games," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1062–1077, March 2017.

[10] M. Huang, "Mean field stochastic games with discrete states and mixed players," *International Conference on Game Theory for Networks*, pp. 138–151, 2012.

[11] G. Chevalier, J. L. Ny, and R. Malhame, "A micro-macro traffic model based on mean-field games," *2015 American Control Conference (ACC)*, pp. 1983–1988, July 2015.

[12] Y. Wang, F. R. Yu, H. Tang, and M. Huang, "A mean field game theoretic approach for security enhancements in mobile ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1616–1627, March 2014.

[13] H. Tembine and M. Huang, "Mean field difference games: Mckean-vlasov dynamics," *The 50th IEEE Conference on Decision and Control and European Control Conference*, pp. 1006–1011, Dec 2011.

[14] D. Bauso, H. Tembine, and T. Başar, "Robust mean field games," *Dynamic Games and Applications*, vol. 6, no. 3, pp. 277–303, Sep 2016.

[15] Q. Zhu, H. Tembine, and T. Basar, "Hybrid risk-sensitive mean-field stochastic differential games with application to molecular biology," *The 50th IEEE Conference on Decision and Control and European Control Conference*, pp. 4491–4497, Dec 2011.

[16] H. Tembine, Q. Zhu, and T. Başar, "Risk-sensitive mean-field games," *IEEE Transactions on Automatic Control*, vol. 59, no. 4, pp. 835–850, 2014.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2020.2986195, IEEE Transactions on Automatic Control



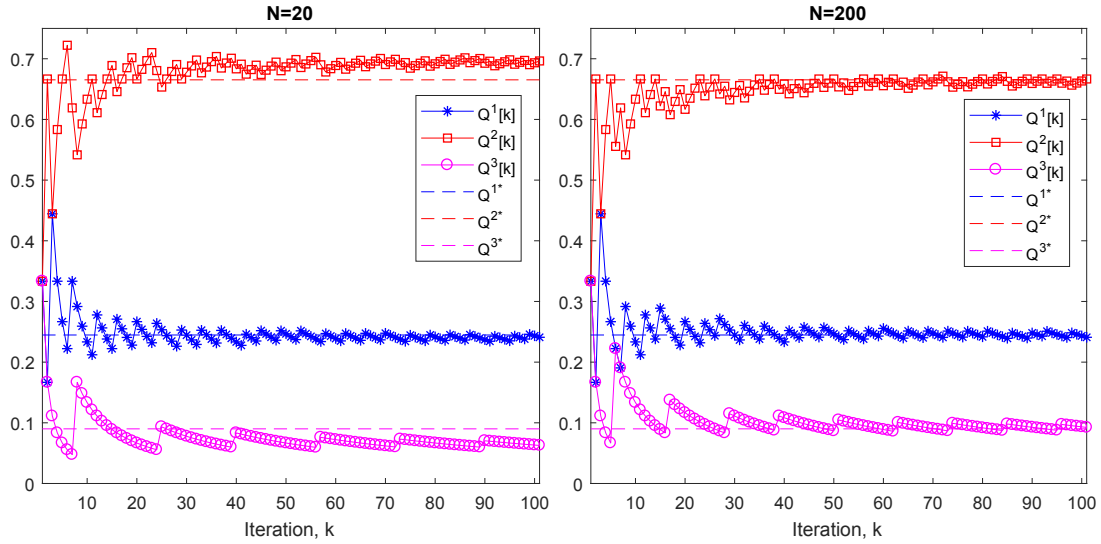


Fig. 3. Convergence of the belief path generated by the symmetric fictitious play (19) in the $N$-player single-stage traffic routing game with three ($J = 3$) alternative paths shown in Figure 1. The left plot shows the case with $N = 20$ while the right plot show the case with $N = 200$. The value of MFE $Q^*$ is also shown.


[17] B. Jovanovic and R. W. Rosenthal, "Anonymous sequential games," *Journal of Mathematical Economics*, vol. 17, no. 1, pp. 77–87, 1988.

[18] G. Y. Weintraub, L. Benkard, and B. Van Roy, "Oblivious equilibrium: A mean field approximation for large-scale dynamic games," *Advances in neural information processing systems*, pp. 1489–1496, 2006.

[19] D. A. Gomes, J. Mohr, and R. R. Souza, "Discrete time, finite state space mean field games," *Journal de mathématiques pures et appliquées*, vol. 93, no. 3, pp. 308–328, 2010.

[20] J. L. N. R. Salhab and R. P. Malhamé, "A mean field route choice game model," *The 57th IEEE Conference on Decision and Control (CDC)*, Dec 2018.

[21] N. Saldi, T. Basar, and M. Raginsky, "Markov–Nash equilibria in mean-field games with discounted cost," *SIAM Journal on Control and Optimization*, vol. 56, no. 6, pp. 4256–4287, 2018.

[22] A. Bensoussan, K. Sung, and S. C. P. Yam, "Linear–quadratic time-inconsistent mean field games," *Dynamic Games and Applications*, vol. 3, no. 4, pp. 537–552, 2013.

[23] B. Djehiche and M. Huang, "A characterization of sub-game perfect equilibria for SDEs of mean-field type," *Dynamic Games and Applications*, vol. 6, no. 1, pp. 55–81, 2016.

[24] A. K. Cissé and H. Tembine, "Cooperative mean-field type games," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 8995–9000, 2014.

[25] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis With Mathematical Programming Methods*. Prentice-Hall, 1984.

[26] J. G. Wardrop, "Some theoretical aspects of road traffic research," in *Inst Civil Engineers Proc London/UK/*, 1952.

[27] J. R. Correa and N. E. Stier-Moses, "Wardrop equilibria," *Wiley encyclopedia of operations research and management science*, 2011.

[28] C. F. Daganzo and Y. Sheffi, "On stochastic models of traffic assignment," *Transportation science*, vol. 11, no. 3, pp. 253–274, 1977.

[29] C. Fisk, "Some developments in equilibrium traffic assignment," *Transportation Research Part B: Methodological*, vol. 14, no. 3, pp. 243–255, 1980.

[30] R. B. Dial, "A probabilistic multipath traffic assignment model which obviates path enumeration," *Transportation research*, vol. 5, no. 2, pp. 83–111, 1971.

[31] W. B. Powell and Y. Sheffi, "The convergence of equilibrium algorithms with predetermined step sizes," *Transportation Science*, vol. 16, no. 1, pp. 45–55, 1982.

[32] Y. Sheffi and W. B. Powell, "An algorithm for the equilibrium assignment problem with random link times," *Networks*, vol. 12, no. 2, pp. 191–207, 1982.

[33] H. X. Liu, X. He, and B. He, "Method of successive weighted averages (MSWA) and self-regulated averaging schemes for solving stochastic user equilibrium problem," *Networks and Spatial Economics*, vol. 9, no. 4, p. 485, 2009.

[34] D. Bauso, X. Zhang, and A. Papachristodoulou, "Density flow in dynamical networks via mean-field games," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1342–1355, March 2017.

[35] J.-B. Baillon and R. Cominetti, "Markovian traffic equilibrium," *Mathematical Programming*, vol. 111, no. 1-2, pp. 33–56, 2008.

[36] Y. Yu, D. Calderone, S. H. Li, L. J. Ratliff, and B. Açıkmeşe, "A primal-dual approach to markovian network optimization," *arXiv preprint arXiv:1901.08731*, 2019.

[37] A. Lachapelle and M.-T. Wolfram, "On a mean field game approach modeling congestion and aversion in pedestrian crowds," *Transportation research part B: methodological*, vol. 45, no. 10, pp. 1572–1589, 2011.

[38] C. Dogbé, "Modeling crowd dynamics by the mean-field limit approach," *Mathematical and Computer Modelling*, vol. 52, no. 9-10, pp. 1506–1520, 2010.

[39] E. Todorov, "Linearly-solvable Markov decision problems," in *Advances in neural information processing systems*, 2007, pp. 1369–1376.

[40] T. Basar and G. Olsder, *Dynamic Noncooperative Game Theory*. Society for Industrial and Applied Mathematics, 1999.

[41] D. Monderer and L. S. Shapley, "Fictitious play property for games with identical interests," *Journal of economic theory*, vol. 68, no. 1, pp. 258–265, 1996.

[42] S.-F. Cheng, D. M. Reeves, Y. Vorobeychik, and M. P. Wellman, "Notes on equilibria in symmetric games," 2004.

[43] K. Dvijotham and E. Todorov, "A unified theory of linearly solvable optimal control," *Proceedings of Uncertainty in Artificial Intelligence (UAI)*, 2011.

[44] E. Theodorou, J. Buchli, and S. Schaal, "A generalized path integral control approach to reinforcement learning," *Journal of Machine Learning Research*, vol. 11, no. Nov, pp. 3137–3181, 2010.

[45] P. D. Grünwald and A. P. Dawid, "Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory," *the Annals of Statistics*, vol. 32, no. 4, pp. 1367–1433, 2004.

[46] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 312–327, 2005.

[47] A. Garcia, D. Reaume, and R. L. Smith, "Fictitious play for finding system optimal routings in dynamic traffic networks1," *Transportation Research Part B: Methodological*, vol. 34, no. 2, pp. 147–156, 2000.

[48] N. Xiao, X. Wang, T. Wongpiromsarn, K. You, L. Xie, E. Frazzoli, and D. Rus, "Average strategy fictitious play with application to road pricing," *American Control Conference*, pp. 1920–1925, 2013.

[49] P. Cardaliaguet and S. Hadikhanloo, "Learning in mean field games: The fictitious play," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 23, no. 2, pp. 569–591, 2017.

[50] A. Pedram and T. Tanaka, "Linearly-solvable mean-field approximation for multi-team road traffic games," *The 58th IEEE Conference on Decision and Control*, Dec 2019.