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 agents 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 characteristic, 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 [2] and [3] 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 optimal reaction of individual players to the mean field (average behavior of the population) is sought. 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 [3]–[5].

In recent years, there has been a growth in the literature on MFGs and its applications. Linear quadratic MFGs [6]–[8] as well as MFGs with more general settings [9], [10] have been extensively explored. MFGs with a major agent and a large number of minor agents are studied [9] and applied to design decentralized security defense decisions in a mobile ad hoc network [11]. MFGs with multiple classes of players are investigated in [12]. The authors of [13] studied the existence of robust (minimax) equilibrium in a class of stochastic dynamic games. In [14], 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 MFG was considered in [15]. While continuous-time continuous-state models are commonly used in the references above, [16]–[20] 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 [21]–[23].

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 [24]. A well-known consequence of a mean-field-like analysis of the traffic user equilibrium is the Wardrop’s first principle [25], [26] characterizing 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) [27], has played a major role in the transportation research, including the convergence analysis of users’ day-to-day routing policy adjustment process [28]–[32]. However, currently only a limited number of papers are available to connect the transportation research and recent progress in the MFG theory. The work [19] considers a discrete-time discrete-state mean-field route choice game. In [10], the authors modeled the interaction between drivers on a straight road as a non-cooperative game and characterized the MFE of this game. In [33], the authors considered a continuous-time Markov chain to model the aggregate behavior of drivers on a traffic network. MFG has been applied to pedestrian crowd dynamics modeling in [34], [35].

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 her randomized policy. We assume that individual drivers’ dynamics are decoupled from each other, but their cost functions are coupled. In particular, we assume that the cost function for each driver has a congestion-independent term (fixed travel cost) and a congestion-dependent penalty term proportional to the logarithm of the number of drivers taking the same route. We regard the congestion-dependent term as an incentive mechanism (toll charge) imposed by the Traffic System Operator (TSO) to induce desired traffic flow. 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 [36], 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 [3].

2) Strong time-consistency: Due to the backward-only characterization, the MFE in our setting is shown to be strongly time-consistent [37], 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 [38] in a scenario where the road traffic 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. Derived MFE is studied in terms of time consistency in Section VI, and its connection to fictitious play 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 $\mathcal{N} = \{1, 2, \ldots, N\}$ the set of players (drivers) and by $\mathcal{T} = \{0, 1, \ldots, T - 1\}$ the set of time steps at which players make decisions.

A. Traffic graph

The traffic graph is a directed graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, V\}$ is the set of nodes (intersections) and $\mathcal{E} = \{1, 2, \ldots, E\}$ 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 given time step $n$, at location $i_{n,t}$ selects $j_{n,t} \in \mathcal{V}(i_{n,t})$, which indicates her next destination. 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 based on her randomized policy. Let $\Delta^i$ be the $J$-dimensional probability simplex, and $Q_{n,t}^i = \{Q_{n,t}^i \mid j \in \mathcal{V}(i) \} \in \Delta^i$ 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 $Q_{n,t} = \{Q_{n,t}^i \mid i \in \mathcal{V}\}$ 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 $Q_{n,t} \subseteq \mathcal{Q}$, where

$$\mathcal{Q} = \left\{ \left\{ Q^i \mid i \in \mathcal{V} \right\} : Q^i \in \Delta^{\mathcal{V}(i)} \right\}$$

is the space of probability distributions. Suppose that $\pi_{n,0}$ has an independent and identical distribution $P_{n,0} = P_0 \in \Delta^{\mathcal{V}}$. Note that if the policy $\{Q_{n,t}\}_{t \in \mathcal{T}}$ of player $n$ is fixed, then the probability distribution $P_{n,t} = \{P_{n,t}^i \mid i \in \mathcal{V}\}$ of her location at time $t$ is recursively computed by

$$P_{n,t+1}^i = \sum_{j \in \mathcal{V}(i)} P_{n,t}^j Q_{n,t+1}^j, \quad \forall t \in \mathcal{T}, j \in \mathcal{V}. \quad (1)$$

If $(i_{n,t}, j_{n,t})$ is the location-action pair of player $n$ at time $t$, it has a joint distribution $P_{n,t}^i Q_{n,t}^j$. If $\{Q_{n,t}\}_{t \in \mathcal{T}}$ and $\{P_{n,t}\}_{t \in \mathcal{T}}$ are fixed for $m \neq n$, we assume that $(i_{n,t}, j_{n,t})$ and $(i_{m,t}, j_{m,t})$ with $m \neq n$ are drawn independently. With a slight abuse of notation, in the sequel 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. Individual tax values depend not only on the players’ locations and actions, but also on how the entire population is distributed over the traffic graph $G$. 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

$$\pi_{n,t}^{ij} = \alpha \left( \log \frac{K_{N,t}^{ij}}{K_{N,t}^i} - \log R_{ij}^t \right). \quad (2)$$

In (2), $\alpha > 0$ is a fixed constant. The parameter $R_{ij}^t > 0$ is also a fixed constant satisfying $\sum_j R_{ij}^t = 1$ for all $i$. $R_{ij}^t$ can be interpreted as the reference policy (state transition probability) designated by the TSO in advance. $K_{N,t}^i$ is the number of players (including player $n$) who are located at the intersection $i$ at time $t$. Likewise, $K_{N,t}^{ij}$ is the number of players (including player $n$) who takes the action $j$ at the intersection $i$ at time $t$. The tax rule (2) indicates that agent $n$ receives a positive payment by taking action $j$ at location $i$ at time $t$ if $K_{N,t}^{ij} / K_{N,t}^i < R_{ij}^t$ (less congested than designated level), while she is penalized by doing so if $K_{N,t}^{ij} / K_{N,t}^i > R_{ij}^t$. Since $K_{N,t}^i$ and $K_{N,t}^{ij}$ are random variables, $\pi_{n,t}^{ij}$ is a random variable. We assume that the TSO is able to observe $K_{N,t}^i$ and $K_{N,t}^{ij}$ at every time step to compute $\pi_{n,t}^{ij}$.

1Whenever $\pi_{n,t}^{ij}$ is computed, we have both $K_{N,t}^{ij} \geq 1$ and $K_{N,t}^i \geq 1$ since at least player $n$ herself is counted. Hence (2) is well-defined.
pair at time $t$ is $(i,j)$, she is interested in minimizing the expectation of the tax penalty

$$
\Pi_{N,n,t}^{ij} \triangleq \mathbb{E} \left[ \pi_{N,n,t}^{ij} \mid i_{n,t} = i, j_{n,t} = j \right].
$$

(3)

It is possible to write $\Pi_{N,n,t}^{ij}$ in terms of $Q_{-n} \triangleq \{Q_m\}_{m \neq n}$ (Appendix A, equation (22)). In particular, the value of $\Pi_{N,n,t}^{ij}$ does not depend on player $n$’s own strategy. This fact will be used in Section III.

D. Traffic routing game

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

$$
J(Q_n, Q_{-n}) = \sum_{t=0}^{T-1} \sum_{i,j} P_{n,t}^{ij} Q_{n,t}^{ij} \left( C_{ij} + \Pi_{N,n,t}^{ij} \right).
$$

(4)

Notice that this quantity depends not only on the $n$-th player’s own strategy $Q_n$ but also on the other players’ strategies $Q_{-n}$ through the term $\Pi_{N,n,t}^{ij}$. If other players’ strategies $Q_{-n}$ are fixed, the term $\Pi_{N,n,t}^{ij}$ is fully specified. Thus, minimizing (4) can be thought of an optimal control problem for player $n$ with control action $Q_{n,t}$ and $P_{n,t}$, with the state space equation (1). If all players are strategic, (4) defines an $N$-player dynamic game. We introduce the following equilibrium concepts for the game described above.

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

**Definition 2**: The $N$-tuple of strategies $\{Q_n^*\}_{n \in N}$ 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 [39]. Thus, [39, 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 $\{Q_n^*\}_{n \in N}$ prior to the execution of the game based on the public knowledge $G, \alpha, N, T, R_{ij}^*, C_{ij}^*$ and $P_0$. Often the case, a Nash equilibrium is desired to be time consistent in that no player is given an incentive to deviate from the precomputed equilibrium routing policy after making real-time observations (such as $K_{N,t}^i$ and $K_{N,t}^N$). Time consistency of a Nash equilibrium in the large-population limit $N \to \infty$ is discussed in Section VII.

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.

**Definition 3**: A set of strategies $\{Q_n^*\}_{n \in N}$ 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 $\epsilon_N \to 0$ as $N \to \infty$ such that for each $n \in N = \{1,2,\ldots, N\}$ and $Q_{-n}$, the inequality $J(Q_n, Q_{-n}^*) + \epsilon_N \geq J(Q_n^*, Q_{-n})$ holds.

Below, we derive a condition that an MFE must satisfy by analyzing the best response of player $n$ to the mean field and combining it with the time consistency condition.

A. Best response

Let $Q^* = \{Q_t^*\}_{t \in T}$ be a fixed routing policy, and $P^* = \{P_t^*\}_{t \in T}$ be defined recursively by

$$
P_{t+1}^j = \sum_i P_t^{ij} Q_t^{ij} \quad \forall j \in V.
$$

Suppose all players other than $n$ adopt the strategy $Q^*$. Then, for each $t \in T$, the probability that the $m$-th player $(m \neq n)$ is located at $i$ is $P_{t+1}^i$. Since $\Pi_{N,n,t}^{ij}$ does not depend on player $n$’s policy, the best response by player $n$ to the mean field is characterized by the solution to the following optimal control problem:

$$
\min_{\{Q_t\}_{t \in T}} \sum_{t=0}^{T-1} \sum_{i,j} P_{t+1}^{ij} Q_t^{ij} \left( C_{ij} + \Pi_{N,n,t}^{ij} \right)
$$

(5)

In (5), we wrote $P_t$ and $Q_t$ in place of $P_{n,t}$ and $Q_{n,t}$ to simplify the notation. Here, $\Pi_{N,n,t}^{ij}$ is a constant (which can be written explicitly in terms of $Q^*$ as shown in equation (23) in Appendix A) that does not depend on the decision variable $\{Q_t\}_{t \in T}$ in (5). The next lemma shows the asymptotic limit of $\Pi_{N,n,t}^{ij}$ as $N \to \infty$ as a function of $Q^*$.

**Lemma 1**: Let $\Pi_{N,n,t}^{ij}$ be defined by (5). If $Q_{m,t} = Q_t^*$ for $m \neq n$ and $P_{t+1}^i Q_t^{ij} > 0$, then

$$
\lim_{N \to \infty} \Pi_{N,n,t}^{ij} = \alpha \log \frac{Q_t^{ij}}{R_t^{ij}}.
$$

**Proof**: Appendix B.

Thus, in the limit of $N \to \infty$, the optimal control problem (5) is well approximated by

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

(6)

B. Time consistency

If the policy $Q^*$ constitutes an MFE, the following time consistency condition must be met:

$$
Q^* \in \arg \min_{\{Q_t\}_{t \in T}} \sum_{t=0}^{T-1} \sum_{i,j} P_t^{ij} Q_t^{ij} \left( C_{ij} + \alpha \log \frac{Q_t^{ij}}{R_t^{ij}} \right).
$$

(7)

This condition ensures that if the policy $\{Q_t^*\}_{t \in T}$ is used by the population excluding player $n$, then the best response by the player $n$ is again $\{Q_t^*\}_{t \in T}$. 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 [36], [40], and can be solved very efficiently.

IV. LINEARLY SOLVABLE MDPs

In this section, we review linearly-solvable MDPs [36], [40] and their solution algorithms. For each $t \in T$, let $P_t$ be the probability distribution over $V$ that evolves according to

$$
P_{t+1} = \sum_i P_t^{ij} Q_t^{ij} \quad \forall j \in V
$$

(8)
with the initial state $P_0$. We assume $C_{ij}^t$, $R_{ij}^t$ for each $t \in T, i \in V, j \in V$ and $\alpha$ are given positive constants. Consider the $T$-step optimal control problem:

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

(9)

The logarithmic term in (9) can be written as the Kullback–Leibler (KL) divergence from the reference policy $R_{ij}^t$ to the selected policy $Q_{ij}^t$. For this reason (9) is also known as the KL control problem [41]. Notice the difference between the optimal control problems (6) and (9); in (6) the logarithmic term is a fixed constant, while in (9) the logarithmic term depends on the chosen policy. The optimal control problem (9) can be solved by standard backward dynamic programming. For each $t \in T$, introduce the value function:

$$
V_t(P_t) \equiv \min_{\{Q_t\}_{t=0}^{T-1}} \sum_{t'=t}^{T-1} \sum_{i,j} P_{t'}^i Q_{t'}^{ij} \left( C_{ij}^{t'} + \alpha \log \frac{Q_{t'}^{ij}}{R_{t'}^{ij}} \right)
$$

and the associated Bellman equation

$$
V_t(P_t) = \min_{\{Q_t\}_{t=0}^{T-1}} \left\{ \sum_{i,j} P_t^i Q_t^{ij} \left( C_{ij}^t + \alpha \log \frac{Q_t^{ij}}{R_t^{ij}} \right) + V_{t+1}(P_{t+1}) \right\}.
$$

(10)

For convenience, we define $V_T(\cdot) = 0$. The next theorem states that the above Bellman equation can be linearized by a change of variables (Cole-Hopf transform), and consequently the optimal control problem (9) is reduced to solving a linear system [36].

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

$$
\phi_t^i = \sum_j R_{ij}^t \exp \left( -\frac{C_{ij}^t}{\alpha} \right) \phi_{t+1}^j \forall i \in V
$$

(11)

with the terminal condition $\phi_{T+1}^i = 1 \forall i$. Then, for each $t = 0, 1, \ldots, T$ and $P_t$, the value function can be written as

$$
\hat{V}_t(P_t) = -\alpha \sum_i P_t^i \log \phi_t^i.
$$

(12)

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

$$
Q_{ij}^{t*} = \frac{\phi_{t+1}^j}{\phi_t^i} R_{ij}^t \exp \left( -\frac{C_{ij}^t}{\alpha} \right).
$$

(13)

**Proof:** Appendix $\Box$

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 show that if $Q^*$ is the solution to the optimal control problem (9), then it satisfies the consistency condition (7). Based on this observation, we show that $Q^*$ is in fact the MFE policy for the road traffic game.

To solve (6), we once again apply dynamic programming. For each $P_t$, define the value function

$$
\hat{V}_t(P_t) \equiv \min_{\{Q_t\}_{t=0}^{T-1}} \sum_{t'=t}^{T-1} \sum_{i,j} P_{t'}^i Q_{t'}^{ij} \left( C_{ij}^{t'} + \alpha \log \frac{Q_{t'}^{ij}}{R_{t'}^{ij}} \right) + \hat{V}_{t+1}(P_{t+1}).
$$

The value function satisfies the Bellman equation:

$$
\hat{V}_t(P_t) = \min_{Q_t} \left\{ \sum_{i,j} P_t^i Q_t^{ij} \left( C_{ij}^t + \alpha \log \frac{Q_t^{ij}}{R_t^{ij}} \right) + \hat{V}_{t+1}(P_{t+1}) \right\}.
$$

(14)

For convenience, we also define $\hat{V}_t(\cdot) = 0$. Notice that we distinguish $\hat{V}_t(\cdot)$ from $\hat{V}_t(\cdot)$, as $\hat{V}_t(\cdot)$ is associated with the optimal control problem (9) ($Q^*$ in the logarithmic term is fixed) while $V_t(\cdot)$ is associated with the auxiliary control problem (9) ($Q$ in the logarithmic term is a variable). Despite this difference, the next key lemma shows that the these value functions coincide with each other.

**Lemma 2:** Let $\{\phi_t\}_{t \in T}$ be the sequence defined by (11). Then, for each $t \in T$ and $P_t$, we have

$$
\hat{V}_t(P_t) = -\alpha \sum_i P_t^i \log \phi_t^i.
$$

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

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

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

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

(15)

Substituting $Q_t^{ij}$ obtained by (13) into (15), we have

$$
\hat{V}_t(P_t) = \min_{Q_t} \sum_{i,j} P_t^i Q_t^{ij} \left( -\alpha \log \phi_t^j \right)
$$

$$
= \min_{Q_t} \sum_i P_t^i \left( -\alpha \log \phi_t^i \right) \sum_j Q_t^{ij}
$$

(16a)

$$
= -\alpha \sum_i P_t^i \log \phi_t^i.
$$

(16b)

This completes the proof. $\Box$

The proof above shows an interesting property of the optimal control problem (6), namely, that any feasible control policy is an optimal control policy.

**Lemma 3:** If $\{Q_t\}_{t \in T}$ in (6) is fixed to be the solution of the auxiliary control problem (9), then an arbitrary sequence of control actions $\{Q_t\}_{t \in T}$ with $Q_t \in Q$ is an optimal solution to (6).

**Proof:** This result follows from the proof of Lemma 2 Since the final expression (16b) does not depend on $Q_t$, any control action $Q_t \in Q$ is a minimizer of the right hand side of the Bellman equation (14). $\Box$

Lemma 3 shows that if all the players except $n$ adopt the policy $\{Q_t\}_{t \in T}$, then any arbitrary policy is the best response by player $n$. This result is the reminiscent of the Wardrop's first principle stating that travel costs are equal on all used routes at the equilibrium.

We are now ready to state the main result of this paper. The next theorem, together with Theorem 1, provides a numerical
method to compute an MFE of the road traffic game presented in Section [1].

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

Proof: Appendix [D]

It is notable that the computational procedure shown in Theorem 1 does not involve any forward-in-time recursion. Consequently, the policy \( Q^* \) constituting the MFE does not depend on the initial state \( P_0 \) of the mean-field. These features are in the stark contrast to the standard MFG formalism, where coupled forward and backward equations must be solved to obtain an MFE.

VI. WEAK AND STRONG TIME CONSISTENCY

In this short section, we emphasize the special time consistency property of the derived MFE. Let \( Q_{n,t} = Q_t \) for each \( n \in \mathcal{N} \) and \( 0 \leq t \leq T - 1 \) 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 \( 0 \leq t \leq T - 1 \), 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 [37] to MFGs.

**Definition 4:** An MFE strategy profile \( Q^* \) is said to be:

1) **weakly time-consistent** if for every \( 0 \leq t \leq T - 1 \), \( \{Q_t^s\}_{0 \leq s \leq t} \) constitutes an MFE of the subgame restricted to \( \{t, t + 1, ..., T - 1\} \) when \( \{Q_s\}_{0 \leq s \leq t - 1} = \{Q_t^s\}_{0 \leq s \leq t - 1} \).

2) **strongly time-consistent** if for every \( 0 \leq t \leq T - 1 \), \( \{Q_t^s\}_{0 \leq s \leq t} \) constitutes an MFE of the subgame restricted to \( \{t, t + 1, ..., T - 1\} \) when regardless of the policy \( \{Q_s\}_{0 \leq s \leq t - 1} \) implemented in the past. In the standard MFG formalism [2], [3], 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 \( \{Q_t^s\}_{0 \leq s \leq t - 1} \), the MFE of the subgame must be recalculated by solving the HJB-FPK system over \( t \leq s \leq T - 1 \) with modified initial condition. In contrast, the MFE obtained for the class of linearly solvable MFGs is characterized only by a backward equation (Theorems [1] and [2]). A notable consequence of this fact is that even if the initial condition \( P_1 \) is inconsistent with the planned distribution, it does not alter the fact that \( \{Q_t^s\}_{0 \leq s \leq t-1} \) constitutes an MFE of the subgame restricted to \( t \leq s \leq T - 1 \). Thus, it can be concluded that the MFE characterized by Theorems [1] and [2] is strongly time-consistent.

VII. MEAN FIELD EQUILIBRIUM AND FICTITIOUS PLAY

To provide the MFE with a practical interpretation, in this section, we 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 restrict ourselves to the \( N \)-player, single-stage 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.

Specifically, we consider the fictitious play process studied in [38], which has already been applied to the day-to-day policy adjustment for traffic routing in [42], [43]. The process proceeds as follows: On day \( \ell \), player \( n \) assumes that player \( m(\neq n) \) adopts a mixed strategy (denoted by \( Q_{n \rightarrow m}[\ell] \in \Delta^{J-1} \)), which is simply the vector of empirical frequencies of player \( m \)'s route choices up to day \( \ell - 1 \). With \( Q_{n \rightarrow m}[\ell] \) being fixed, player \( n \) selects a route with the lowest expected cost. This process is repeated indefinitely. The process is summarized in Algorithm 1. Further details of the process can be found in [38]. More general discussion on fictitious play can be found in [44]. Fictitious play in the context of MFG has been studied in the recent work [45].

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

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

2. **Step 1:** At the beginning of day \( \ell \), each player \( n \) fixes assumed mixed strategy \( Q_{n \rightarrow m}[\ell] \in \Delta^{J-1} \) 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.,

\[
y_n^j[\ell] = E \left( C^j + \alpha \log \frac{K^j_N}{NR^j} \right).
\]  

3. **Step 2:** At the end of day \( k \), each player \( n \) updates her belief based on observations \( r_n[m], m \neq n \) by

\[
Q_{n \rightarrow m}[\ell + 1] = \frac{\ell}{\ell + 1} Q_{n \rightarrow m}[\ell] + \frac{1}{\ell + 1} \delta(r_n[m])
\]  

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

Monderer and Shapley [38] showed that in \( N \)-player games with identical payoff functions, every fictitious play process converges in beliefs to an equilibrium. This result is directly applicable to the simplified road traffic game in Figure 1 since it is a symmetric game. However, since the game has multiple
Nash equilibrium points (including non-symmetric ones) in general, there is no guarantee that the belief path generated by Algorithm 1 converges to a symmetric equilibrium. To overcome this difficulty, we restrict ourselves to the symmetric fictitious play process. The symmetric fictitious play is defined by Algorithm 1 together with the 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, no mechanism in Algorithm 1 breaks the symmetry; for all \( \ell \geq 1 \), we have \( Q_{n \to m}[\ell] = Q[\ell]. \) \( y_n[\ell] = y[\ell] \) and \( r_n[\ell] = r[\ell] \). Equations (17) and (18) are simplified to

\[
y^t[\ell] = C^j + \alpha \sum_{k=0}^{N-1} \log \binom{k+1}{N} \binom{N-1}{k} (1 - Q^j[\ell])^{N-1-k}
\]

and

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

respectively. Since the same belief \( Q[\ell] \) is shared by all the players, a limit point of the belief path is necessarily a symmetric policy. Combined with the convergence result by Monderer and Shapley [38], it can be concluded that every symmetric policy. Combined with the convergence result by Monderer and Shapley [38], it can be concluded that every limit point of the belief path generated by the symmetric fictitious play above is a symmetric equilibrium.

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

**Lemma 4:** There exists a unique symmetric equilibrium, denoted by \( Q^{(N^*)} \), in the \( N \)-player single-stage road traffic game shown in Figure 1

**Proof:** Appendix E

Let \( Q^* \) be the MFE of the simplified traffic routing game in Figure 1 which is characterized as the unique solution to the following convex optimization problem:

\[
\min_{Q \in \Delta^{J-1}} \sum_{j=1}^{J} Q^j \left( C^j + \alpha \log \frac{Q^j}{R^j} \right). \tag{19}
\]

Suppose we have \( \lim_{N \to \infty} Q^{(N^*)} = Q^* \). Then, one can interpret the MFE as 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 observation is confirmed in a numerical study in Section VIII-B. While this conclusion is currently restricted to the single stage road traffic game, it is worthwhile to investigate its generalization in future work.

### VIII. Numerical Illustration

#### A. Road traffic game and congestion control

In this section, we 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 \( t \in T \), the travel cost for each player is

\[
C^j = \begin{cases} 
C_{\text{term}} & \text{if } j = i \\
1 + C_{\text{term}} & \text{if } j \in V(i) \\
100000 + C_{\text{term}} & \text{if } j \notin V(i) \text{ or } j \text{ is an obstacle}
\end{cases}
\]

where \( V(i) \) contains the north, east, south, and west neighborhood of the cell \( i \). To incorporate the terminal cost, we introduce \( C_{\text{term}} = 0 \) if \( t = 0, 1, \ldots, T - 1 \) and \( C_{\text{term}} = 10 \sqrt{ \text{dist}(j, D) } \) if \( t = T - 1 \), where \( \text{dist}(j, D) \) is the Manhattan distance between the player’s final location \( j \) and the destination cell (indicated by “D”). As the reference distribution, we use \( R^j = 1/|V(i)| \) (uniform distribution) for each \( i \in V \) and \( t \in T \) 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 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

This subsection presents 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 (18) 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. This confirms the theoretical observation in Section [VII].

IX. CONCLUSION AND FUTURE WORK

We showed that the MFE of a large-population road traffic game under the log-population tax mechanism can be obtained via the linearly solvable MDP. Strong time consistency of the obtained MFE was discussed. A preliminary study on the MFE and its connection to fictitious play was presented.

Continuous-time, continuous-state correspondence of the results presented in this paper is worth investigating in the future. Generalization to MFGs with multiple classes of players is also worth investigating. The interface between the traffic SUE theory and MFG should be thoroughly studied. Convergence of fictitious play and its relationship with MFE presented in Section [VII] should be studied in more general context. Finally, how the results in this paper can be used for the traffic incentive design (mechanism design) for congestion mitigation is an important question to address in the future.

APPENDIX A

EXPLICIT EXPRESSION OF \([3]\)

Since player \( n \)'s probability of taking action \( j \) at location \( i \) at time step \( t \) is given by \( P_{n,t}^i Q_{n,t}^j \), the total number \( K_{N,t}^i \) of such players follows the Poisson binomial distribution

\[
\Pr(K_{N,t}^i = k) = \sum_{A_k \in F_k} \prod_{n \in A} P_{n,t}^i Q_{n,t}^j \prod_{n^* \in A^c} (1 - P_{n,t}^i Q_{n,t}^j).
\]

Here, \( F_k \) is the set of all subsets of size \( k \) that can be selected from \( N \) = \{1, 2, ..., \( N \)\}, and \( A^c = F_k \setminus A \). Similarly, the distribution of \( K_{N,t}^i \) is given by

\[
\Pr(K_{N,t}^i = k) = \sum_{A_k \in F_k} \prod_{n \in A} P_{n,t}^i Q_{n,t}^j \prod_{n^* \in A^c} (1 - P_{n,t}^i Q_{n,t}^j).
\]

Notice also that the conditional distribution of \( K_{N,t}^i \) given player \( n \)'s location-action pair \((i_{n,t}, j_{n,t}) = (i, j)\) is

\[
\Pr(K_{N,t}^i = k \mid i_{n,t} = i, j_{n,t} = j) = \sum_{A_k \in F_k} \prod_{n \in A} P_{n,t}^i Q_{n,t}^j \prod_{n^* \in A^c} (1 - P_{n,t}^i Q_{n,t}^j).
\]

Here, \( F_k \) is the set of all subsets of size \( k \) that can be selected from \( N \) \{1\}, and \( A^c = F_k \setminus A \). Similarly, the conditional distribution of \( K_{N,t}^i \) given \( i_{n,t} = i \) is

\[
\Pr(K_{N,t}^i = k + 1 \mid i_{n,t} = i) = \sum_{A_k \in F_k} \prod_{m \in A} P_{m,t}^i Q_{m,t}^j \prod_{m^* \in A^c} (1 - P_{m,t}^i Q_{m,t}^j).
\]

Therefore, given the prior knowledge that the player \( n \)'s location-action pair at time \( t \) is \((i, j)\), the expectation of her tax penalty \( \pi_{N,n,t}^{ij} \) is

\[
\Pi_{N,n,t}^{ij} = \mathbb{E} \left[ \pi_{N,n,t}^{ij} \mid i_{n,t} = i, j_{n,t} = j \right] = \sum_{k=0}^{N-1} \alpha \log \frac{k + 1}{N} \sum_{A_k \in F_k} \prod_{m \in A} P_{m,t}^i Q_{m,t}^j \prod_{m^* \in A^c} (1 - P_{m,t}^i Q_{m,t}^j) + \alpha \log R_{i,t}^j.
\]

Notice that the quantity \( \Pi_{N,n,t}^{ij} \) depends on the strategies \( Q_{m,t}^j \), but not on \( Q_m \). In other words, \( \pi_{N,n,t}^{ij} \) is a random variable whose distribution does not depend on player \( n \)'s own strategy.

To evaluate \( \Pi_{N,n,t}^{ij} \) when all players other than player \( n \) takes the same strategy (i.e., \( Q_{m,t} = Q_i^j \) for \( m \neq n \)), notice that the conditional distributions of \( K_{N,t}^i \) and \( K_{N,t}^i \) given \((i_{n,t}, j_{n,t}) = (i, j)\), provided by \( [30] \) and \( [31] \), simplify to the binomial distributions

\[
\Pr(K_{N,t}^i = k + 1 \mid i_{n,t} = i, j_{n,t} = j) = \binom{N-1}{k} (P_i^s Q_{t}^j)^k (1 - P_i^s Q_{t}^j)^{N-1-k}.
\]

Thus, the expression \( [32] \) simplifies to

\[
\Pi_{N,n,t}^{ij} = \mathbb{E} \left[ \pi_{N,n,t}^{ij} \mid i_{n,t} = i, j_{n,t} = j \right] = \sum_{k=0}^{N-1} \alpha \log \frac{k + 1}{N} \binom{N-1}{k} (P_i^s Q_{t}^j)^k (1 - P_i^s Q_{t}^j)^{N-1-k} + \alpha \log R_{i,t}^j.
\]

APPENDIX B

PROOF OF LEMMA [1]

Let \( K_{N,-n,t}^i \) denote the number of agents, except agent \( n \), which are located at intersection \( i \) at time \( t \) and let \( K_{N,-n,t}^{ij} \) denote the number of agents, except agent \( n \), which are located at intersection \( i \) at time \( t \) and select intersection \( j \) as their next destination. Thus, we have \( K_{N,-n,t}^i = \sum_{i \neq n} 1 \{i_{t,t} = i\} \) and \( K_{N,-n,t}^{ij} = \sum_{i \neq n} 1 \{i_{t,t} = i, j_{t,t} = j\} \), where \( 1 \{\cdot\} \) is the
Next we show the other direction. For $\epsilon \in (0, \frac{P_i^* Q_i^{ij*}}{2})$, we can write $E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} \right]$ as
\[
E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} \right] = E \left[ \log \left( \frac{1}{N} + E \left[ K_{ij}^{(N-1).t} \right] \right) \right] \leq \log \left( \frac{1}{N} + \frac{N - 1}{N} P_i^* Q_i^{ij*} \right)
\]
Using Jensen inequality, we have
\[
E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} \right] \leq \log \left( \frac{1}{N} + E \left[ K_{ij}^{(N-1).t} \right] \right) = \log \left( 1 + \frac{N - 1}{N} P_i^* Q_i^{ij*} \right)
\]
\[
\limsup_{N \to \infty} E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} \right] \leq \log P_i^* Q_i^{ij*}
\]
Next we show the other direction. For $\epsilon \in (0, \frac{P_i^* Q_i^{ij*}}{2})$, we can write $E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} \right]$ as
\[
E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} \right] = E \left[ \log \left( \frac{1}{N} + E \left[ K_{ij}^{(N-1).t} \right] \right) \right] \leq E \left[ \log \left( \frac{1}{N} + E \left[ K_{ij}^{(N-1).t} \right] \right) \right] = \log \left( 1 + \frac{N - 1}{N} P_i^* Q_i^{ij*} \right)
\]
Using the Hoeffding inequality, it follows that $K_{ij}^{(N-1).t}$ converges to $P_i^* Q_i^{ij*}$ in probability as $N$ becomes large. From continuous mapping theorem, we have the convergence of $\log \frac{K_{ij}^{(N-1).t}}{N}$ in probability to $\log P_i^* Q_i^{ij*}$ for $P_i^* Q_i^{ij*} > 0$. Similarly, $\log \frac{K_{ij}^{(N-1).t}}{N}$ converges to $1$ in probability. Thus, from Slutsky’s Theorem, we have $\lim \log \frac{1 + K_{ij}^{(N-1).t}}{N} 1 \{ K_{ij}^{(N-1).t} > \epsilon \} = \log P_i^* Q_i^{ij*}$ in distribution. Using Fatou’s lemma and the fact that $\log \frac{1 + K_{ij}^{(N-1).t}}{N} 1 \{ K_{ij}^{(N-1).t} > \epsilon \} \geq \log \epsilon$, we have
\[
\liminf_{N \to \infty} E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} 1 \{ K_{ij}^{(N-1).t} > \epsilon \} \right] \geq \log P_i^* Q_i^{ij*}
\]
We also have
\[
\left| E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} 1 \{ K_{ij}^{(N-1).t} \leq \epsilon \} \right] \right| \leq \log (N) Pr \left( \frac{K_{ij}^{(N-1).t}}{N} \leq \epsilon \right)
\]
Using the Hoeffding inequality, it is straightforward to show that $Pr \left( \frac{K_{ij}^{(N-1).t}}{N} \leq \epsilon \right)$ decays to zero exponentially in $N$ which implies that
\[
\lim_{N \to \infty} E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} 1 \{ K_{ij}^{(N-1).t} \leq \epsilon \} \right] = 0
\]
Thus, we have
\[
\liminf_{N \to \infty} E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} \right] \geq \log P_i^* Q_i^{ij*}
\]
which implies that $\lim_{N \to \infty} E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} \right] = \log P_i^* Q_i^{ij*}$. Following similar steps, it is straightforward to show that $\lim_{N \to \infty} E \left[ \log \frac{1 + K_{ij}^{(N-1).t}}{N} \right] = \log P_i^*$ which completes the proof.
APPENDIX C
PROOF OF THEOREM

Due to the choice of the terminal condition \( \phi_P = 1 \),
\[
V_T(P_t) = \sum_i P_i t^i C_t = -\alpha \sum_i P_i t^i \log \phi_t = 0
\]
holds. To complete the proof by backward induction, assume that \( V_{t+1}(P_{t+1}) = -\alpha \sum_i P_{t+1} \log \phi_{t+1}^i \) holds for some \( 0 \leq t \leq T - 1 \). Then, due to the Bellman equation (10), we have
\[
V_t(P_t) = \min_{q_t} \left\{ \sum_{i,j} P_i t^i Q_t^{ij} \left( \rho_t^{ij} + \alpha \log \frac{Q_t^{ij}}{R_t^{ij}} \right) \right\}
\]
where \( \rho_t^{ij} = C_t^{ij} - \alpha \log \phi_t^{i+1} \) is a constant. It is elementary to show that the minimum is attained by
\[
Q_t^{ij*} = \frac{R_t^{ij}}{\phi_t} \exp \left( -\frac{\rho_t^{ij}}{\alpha} \right)
\]
from which (13) follows. By substitution, the optimal value is shown to be \( V_t(P_t) = -\alpha \sum_i P_i t^i \log \phi_t \). This completes the induction proof.

APPENDIX D
PROOF OF THEOREM

Let the policies \( Q_{n,t}^{ij} = Q_t^{ij*} \) for \( m \neq n \) be fixed. It is sufficient to show that there exists a sequence \( \epsilon_N \to 0 \) such that the cost of adopting a strategy \( Q_{n,t}^{ij} = Q_{n,t}^{ij*} \) for player \( n \) is no greater than \( \epsilon_N \) plus the cost of adopting any other policy. Since
\[
\Pi_{N,n,t}^{ij} \to \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} \text{ as } N \to \infty,
\]
there exists a sequence \( \delta_N \downarrow 0 \) such that
\[
\Pi_{N,n,t}^{ij} + \delta_N > \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} \quad \forall i,j,t.
\]
Now, for all policy \( \{ Q_{n,t}\} \in \mathcal{T} \) of player \( n \) and the induced distributions \( P_{n,t}^j \), we have
\[
\begin{align*}
\sum_{t=0}^{T-1} \sum_{i,j} P_{n,t}^j Q_{n,t}^{ij} \left( C_t^{ij} + \Pi_{N,n,t}^{ij} \right) &> \sum_{t=0}^{T-1} \sum_{i,j} P_{n,t}^j Q_{n,t}^{ij} \left( C_t^{ij} + \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} - \delta_N \right) \\
&= \sum_{t=0}^{T-1} \sum_{i,j} P_{n,t}^j Q_{n,t}^{ij} \left( C_t^{ij} + \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} \right) - TV^2 \delta_N \\
&\geq \min_{\{Q_{n,t}\} \in \mathcal{T}} \sum_{t=0}^{T-1} \sum_{i,j} P_{n,t}^j Q_{n,t}^{ij} \left( C_t^{ij} + \alpha \log \frac{Q_t^{ij*}}{R_t^{ij}} \right) - TV^2 \delta_N.
\end{align*}
\]
Notice that the minimization in the last line is attained by adopting \( Q_{n,t} = Q_t^{ij*} \). Since \( \epsilon_N \triangleq TV^2 \delta_N \downarrow 0 \), this completes the proof.

APPENDIX E
PROOF OF LEMMA

For each \( j = 1, 2, \ldots, J \), define \( f^j_N : [0, 1] \to \mathbb{R} \) by
\[
f^j_N(Q^j) \triangleq \frac{c^j + \alpha \sum_{n=0}^{N-1} \log \left( \frac{Q^j}{n} (N-1)^{N-1-n} \right)}{J}.
\]
Notice that \( f^j_N(Q) \) is a continuous and strictly increasing function. If \( Q^* \in \Delta^{J-1} \) is a symmetric Nash equilibrium of the \( N \)-player single-stage road traffic game, it satisfies the condition
\[
Q^* \in \arg \min_{Q \in \Delta^{J-1}} \sum_j Q^j f^j_N(Q^j). \quad (26)
\]
From the KKT condition, \( Q^* \in \Delta^{J-1} \) satisfies (26) if and only if there exists \( \lambda \in \mathbb{R} \) such that
\[
f_N^j(Q^*) = \lambda \text{ for all } j \text{ such that } Q^j > 0 \quad (27a)
\]
\[
f_N^j(0) \geq \lambda \text{ for all } j \text{ such that } Q^j = 0. \quad (27b)
\]
Alternatively, the above condition can directly be obtained from the Wardrop’s first principle (25). Thus, it is sufficient to show that there exists a unique \( Q^* \in \Delta^{J-1} \) satisfying the condition (27). For each \( j = 1, 2, \ldots, N \), define \( g^j : \mathbb{R} \to [0, 1] \) by
\[
g^j(\lambda) = \begin{cases} 
0 & \text{if } \lambda \leq f^j(0) \\
(\lambda - f^j(0))^{-1} & \text{if } f^j(0) < \lambda \leq f^j(1) \\
1 & \text{if } f^j(1) < 1.
\end{cases} \quad (28)
\]
The next claim follows from the intermediate value theorem and the monotonicity of \( g^j \) for each \( j \).

Claim 1: There exists \( \lambda \in \mathbb{R} \) such that \( g^j(\lambda) = 1 \). Moreover, if there exist \( \lambda_1, \lambda_2 \in \mathbb{R} \) such that \( g^j(\lambda_1) = g^j(\lambda_2) \), then \( g^j(\lambda_1) = g^j(\lambda_2) \) for each \( j \).

Claim 2: If \( Q^* \in \Delta^{J-1} \) and \( \lambda \in \mathbb{R} \) satisfy (27), then it is necessary that \( g^j(\lambda) = 1 \).

Proof: Assume (27) and \( g^j(\lambda) \neq 1 \). Then
\[
\sum_{j=1}^N Q^j = \sum_{j:Q^j>0} Q^j \quad (29a)
\]
\[
= \sum_{j:Q^j>0} g^j(\lambda) \quad (29b)
\]
\[
= \sum_{j:Q^j>0} g^j(\lambda) + \sum_{j:Q^j=0} g^j(\lambda) \quad (29c)
\]
\[
= \sum_j g^j(\lambda) \quad (29d)
\]
\[
= g(\lambda) \neq 1. \quad (29e)
\]

Equality (29b) follows from (27a) and (28). Equality (29c) holds since, for \( j \) such that \( Q^j = 0 \), we have \( f_N^j(0) \geq \lambda \) by (27b) and thus, by definition (28), \( g^j(\lambda) = 0 \). However, the chain of equalities (29) is a contradiction to \( Q^* \in \Delta^{J-1} \).

Now, pick any \( \lambda \in \mathbb{R} \) such that \( g^j(\lambda) = 1 \), and construct \( Q^* \) by \( Q^* = g^j(\lambda) \). By Claim 1, such a \( Q^* \) is unique. It is easy to check that \( Q^* \in \Delta^{J-1} \) and the condition (27) is satisfied. This construction, together with Claim 2, shows that there exists a unique \( Q^* \in \Delta^{J-1} \) satisfying the condition (27).
ACKNOWLEDGMENT

The authors would like to thank Mr. Matthew T. Morris and Mr. James S. Stanesci 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] P. E. Caines, M. Huang, and R. Malhamé, “Mean field games,” in Handbook of Dynamic Game Theory (T. Başar, G. Zaccour, Eds.), 2018.

[3] J. L. Lasserre and P. L. Lions, “Mean field games,” Japanese Journal of Mathematics, vol. 2, no. 1, pp. 229–260, 2007.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

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

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

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

[17] G. Y. Weinfraub, 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.

[18] 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.

[19] 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.

[20] 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.

[21] A. Bensoussan, K. Sung, and S. C. P. Yam, “Linear–quadratic time-inconsistent mean field games,” Dynamic Games and Applications, vol. 4, pp. 537–552, 2013.

[22] 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.