Game-theoretic reinforcement learning for multi-intersection control with oversaturated traffic

Renxin Zhong, Qinzhou Ma, Andy H.F. Chow, Zicheng Su, Enming Liang, Xiaotian Qin

APPENDIX A

We summarize the key notations in the following table.

TABLE A.1: List of key notations.

Notation	Definition
$s_i(t)$	Traffic states of intersection i at time t
$a_i(t)$	Action of intersection i at time t
$r_i(t)$	Reward of intersection i at time t
$f_{ij,c}(t)$	Traffic flow of cell c at time t
$k_{ij,c}(t)$	Traffic density of cell c at time t
$v_{ij,c}(t)$	Speed of cell c at time t
$y_{ij,c}(t)$	Exogenous demand of cell c at time t
$h_{ij,c}$	Length of cell c
$d_{ij}(t)$	Delay of lane j at time t
θ	Neural network parameters
ϕ	Potential functions
Q_i	Q-function of intersection i
δ_i	TD error of intersection i
$L(\theta)$	Loss function of neural networks
γ	Discount rate
κ	Index of training epoch
ψ	Batch size
ξ	Learning rate

APPENDIX B

This section provides the proof of Proposition 2 in the main text as follows.

Proof.

$$\sum_{t\in\mathbb{T}}\sum_{i\in\mathbb{I}} -r_i'(t) \tag{1}$$

$$=\sum_{t\in\mathbb{T}}\sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{L}_i^{in}} \left(d_{ij}(t)-\lambda_1 f_{ij}(t)\Delta t\right)$$

$$=\sum_{t\in\mathbb{T}}\sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{L}_i^{in}} q_{ij}(t)(\Delta t-\hat{b}_{ij}(t)/v_{ij}^*)-\lambda_1\sum_{t\in\mathbb{T}}\sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{L}_i^{in}} f_{ij}(t)\Delta t \tag{2}$$

Given the fixed total number of vehicles and pre-defined planning routes for each vehicle, the total number of times all intersections are crossed by all vehicles over the entire planning horizon is a constant N defined by:

$$N=\sum_{t\in\mathbb{T}}\sum_{i\in\mathbb{I}}\sum_{j\in\mathbb{L}_i^{in}} f_{ij}(t)\Delta t \tag{3}$$

Combining (3) into (2), we obtain:

$$\sum_{t\in\mathbb{T}}\sum_{i\in\mathbb{I}} -r_i'(t)=\sum_{n\in\mathbb{N}}[T_n-T_n^*]-\lambda_1 N$$

We now complete the proof that the regularizer $f_{ij}(t)\Delta t$ does not affect the optimality of the original objective function.

Remark 1. *For a corridor with two intersections, if the total traffic demand over the planning horizon is 1, with planning routes that involve crossing both intersections, then $N=2$.*

APPENDIX C

In this section, we examine the Markovian property of traffic dynamics under oversaturated traffic.

Lemma C (Chapman-Kolmogorov equation [1, pp. 346])**.** *The necessary condition for the system dynamics to be Markovian is that its transition function P satisfies the Chapman-Kolmogorov (CK) equation as follows:*

$$\begin{aligned}&P\left(X_{g+t}=\Gamma\mid X_0=x\right)\\&=\sum_y P\left(X_{g+t}=\Gamma\mid X_g=y\right)\quad\cdot P\left(X_g=y\mid X_0=x\right)\end{aligned}$$

for every $t\geq 0$ and $g\geq 0$, where X represents the stochastic variable.

The C-K equation, as presented in (4), describes the probability of transitioning from the initial state x to state Γ after $g+t$ steps. This probability is obtained by summing up the probabilities of transitioning from state x to an intermediate state y at step g and then to state Γ, considering all possible intermediate states y at step g.

The necessary condition presented in Lemma C can be assessed by examining the absolute errors between the left-hand side and the right-hand side of the C-K equation (4), denoted as $\Xi(t)$. The definition of $\Xi(t)$ is as follows:

$$\begin{aligned}&\Xi(t)\\&=|P\left(X_{g+t}=\Gamma\mid X_0=x\right)-\\&\quad\sum_y P\left(X_{g+t}=\Gamma\mid X_g=y\right)\cdot P\left(X_g=y\mid X_0=x\right)\end{aligned} \tag{4}$$

The Markovian property holds when $\Xi(t)=0$, otherwise it is violated. We examine Lemma C on an intersection in

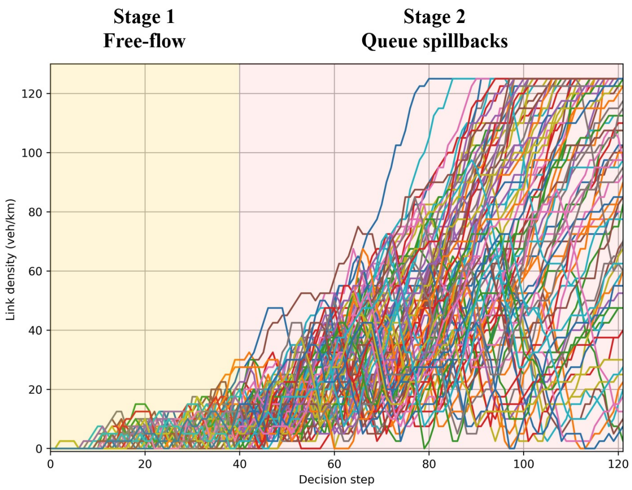

Fig. 1: Profiles of upstream link density with queue spillbacks.

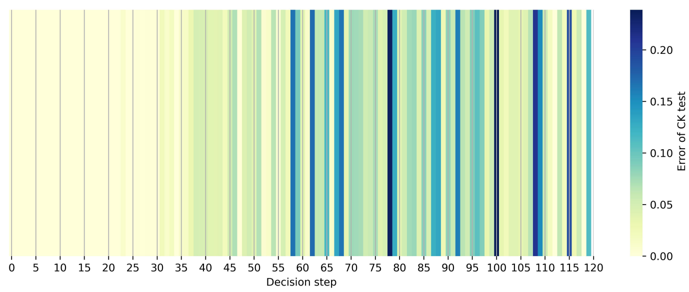

Fig. 2: Absolute errors of the C-K equation along the simulation horizon.

the Hangzhou network under oversaturated traffic conditions with 100 simulation runs. We set the number of transition steps g as 5 decision intervals and discretize the link density with an interval of 10 veh/km. Figure 1 depicts the profiles of upstream link density and Figure 2 presents the absolute errors $\Xi(t)$ along the simulation horizon. Note that during the first 40 steps, the majority of trajectories experience free-flow conditions and therefore $\Xi(t)$ equals 0. This observation suggests that the traffic dynamics under the free-flow condition conform to a Markov chain. After that, as queues from the downstream section begin to spill over, the upstream link densities start to accumulate. Consequently, $\Xi(t)$ exceeds 0 and reaches a maximum value of 0.25, which indicates a violation of the Markovian property. With the above validation of the C-K equation, we demonstrate that the fundamental MDP assumption in RL approaches to traffic signal control is no longer valid when there is queue spillover.

References

[1] W. Feller, *An introduction to probability theory and its applications, second edition*. John Wiley & Sons, 1971, vol. 2.