Parameter Estimation in Optimal Tolling for Traffic Networks Under the Markovian Traffic Equilibrium

Chih-Yuan Chiu[1] and Shankar Sastry[1]

***Abstract*— Tolling, or congestion pricing, has emerged as an effective tool for preventing gridlock in traffic systems. However, tolls are currently mostly designed on route-based traffic assignment models (TAM), which may be unrealistic and computationally expensive. Existing approaches also impractically assume that the central tolling authority can access latency function parameters that characterize the time required to traverse each network arc (edge), as well as the entropy parameter β that characterizes commuters' stochastic arc-selection decisions on the network. To address these issues, this work formulates an online learning algorithm that simultaneously refines estimates of linear arc latency functions and entropy parameters in an arc-based TAM, while implementing tolls on each arc to induce equilibrium flows that minimize overall congestion on the network. We prove that our algorithm incurs regret upper bounded by $O(\sqrt{T}\ln(T)|A|\max\{|I|\ln(|A|/|I|), B\})$, where T denotes the total iteration count, $|A|$ and $|I|$ denote the total number of arcs and nodes in the network, respectively, and B describes the number of arcs required to construct an estimate of β (usually $\ll |I|$). Finally, we present numerical results on simulated traffic networks that validate our theoretical contributions.**

I. Introduction

Modern transportation systems are often plagued with congestion, induced by commuters who select latency-minimizing routes from their source to their destination in a self-interested manner. Tolling mechanisms, which impose additional prices on each arc (edge) in the network, offer a natural solution to this issue. By appropriately augmenting the overall cost of traveling on particularly congested arcs, effectively implemented tolls can reshape commuters' incentives, and motivate them to make arc selections that reduce the overall network congestion.

Although various traffic assignment and tolling mechanisms have been proposed to regulate congestion on transportation networks, the theoretical guarantees of these approaches, if any, are usually predicated upon unrealistic or impractical modeling assumptions. For instance, [1–3] design traffic assignment schemes or tolls using *route-based traffic assignment models* (TAMs) to capture commuters' navigation decisions, i.e., each commuter is assumed to make a single route selection at their origin, and to refrain from deviating from their selected route at intermediate nodes. Likewise, [4] presents an online learning algorithm to infer the unknown latency functions of a traffic network, while performing optimal route assignment over the network in the context of a route-based TAM. Unfortunately, route-based TAMs do not capture the behavior of commuters who reroute halfway to their destination, and can be computationally expensive, since the number of routes in a traffic network can grow exponentially with the number of arcs (edges). In contrast, [5–8] investigate commuters' decision making and tolling mechanisms in a traffic network over a stochastic *arc-based TAM*, in which commuters sequentially select among outgoing arcs at each intermediate node from source to destination. In particular, an entropy parameter $\beta > 0$ is used to characterize the degree of irrationality with which the traveler population selects arc sequences, due to the incomplete and imperfect information they possess about the latency cost of each arc. However, these approaches unrealistically assume that the central tolling authority possesses perfect knowledge of β and the network latency functions.

To address the above shortcomings, this work presents an online learning algorithm in the framework of a stochastic, arc-based traffic assignment model (TAM), to simultaneously learn the latency function and the entropy parameter, while implementing tolls that become increasingly effective at reducing overall congestion in subsequent iterations. At each iteration, we first implement tolls, constructed during the most recent iteration, on each arc in the network. We then collect the resulting equilibrium traffic flow and latency data from each arc, and apply a regularized least-squares method to update our estimates of the latency function parameters, based on the collected data. In turn, the flow data and latency function estimates can then be used to update our estimate of the entropy parameter β, using the Principle of Optimism in the Face of Uncertainty. Finally, these improved estimates of the latency function and entropy parameters are used to design an improved tolling strategy for the next iteration.

We define the stage-wise regret of our algorithm at each iteration t to be the difference between the following two quantities: (a) The overall latency in the network induced by equilibrium flows corresponding to the toll implemented at iteration t, and (b) The minimum overall latency attainable by the tolling mechanism if it possessed perfect knowledge of the entropy parameter and each arc latency function. The cumulative regret is then computed by summing the stage-wise regret across all iterations. Our algorithm incurs regret of order $O(\sqrt{T}\ln(T)|A| \cdot \max\{|I|\ln(|A|/|I|), B\})$, where T denotes the total iteration count, $|A|$ and $|I|$ denote the number of arcs and nodes in the network, respectively, and B denotes the number of arcs in the network used to construct the estimate of the entropy parameter β at each iteration.

On a technical level, our algorithm utilizes concepts famil-

Supported by NSF Grant 2031899, Collaborative Research: Transferable, Hierarchical, Expressive, Optimal, Robust, Interpretable Networks.

[1]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 (emails: {chihyuan_chiu, sastry} at berkeley dot edu).

iar to the bandits community, such as the regularized least-squares method for latency function estimation [4, 9, 10], and the Principle of Optimism in the Face of Uncertainty for entropy parameter estimation and toll design [9]. However, the problem formulation and proof methodologies considered in this work differ significantly from the above literature. First, in our problem setup, the decision maker's actions are tolls, which induce equilibrium flows through a non-convex map; in turn, the regret is defined from the overall network congestion generated by these equilibrium flows. Similarly, the unknown entropy parameter estimated in our work affects the cumulative regret in a complicated, network structure-dependent manner (see Section IV, Remark 1). These complex dependencies between the actions, unknown parameters, and regret preclude the direct use of analysis techniques in the bandit literature. Moreover, whereas the decision-maker in [4] estimates latency functions in the context of a route-based TAM and implements optimal flow assignments directly, our work estimates both the latency functions and entropy parameter β of an underlying arc-based TAM, and implements tolls, which in turn induce an equilibrium flow from which the regret is computed. In particular, to estimate the entropy parameter β, we use a novel approximation scheme beyond the methods in [4].

Likewise, various methods have investigated the problem of estimating the entropy parameter of softmax models in the context of traffic assignment models or maximum entropy inverse reinforcement learning [11–13]. However, these approaches usually use heuristic models to approximate the unknown parameter [11, 12], or assume that the overall objective can be written as a convex function of the entropy parameter [13]. These assumptions separate the above methods from our work, since our formulation involves cost and equilibrium models that are highly non-convex in the action variables (tolls) and in the unknown entropy parameter β.

The following sections are structured as follows. Section II introduces the traffic network studied throughout the remaining sections, as well as the incentive structures faced by the commuters traversing the network. Section III presents our online algorithm. An upper bound for the overall regret incurred by this algorithm is given in Section IV. Finally, Section V presents empirical evidence for the theoretical regret bounds on our algorithm, while Section VI summarizes our work and presents avenues for future research.

Notation: Below, for any $n \in \mathbb{N}$, we denote $[n] := \{1, \cdots, n\}$. For any $n \in \mathbb{N}$ and $i \in [n]$, let e_i denote the i-th standard unit vector in the Euclidean space $\mathbb{R}^n$. We set $\mathbf{1}\{\cdot\}$ to equal 1 if the input event occurs, and 0 otherwise.

II. Preliminaries

A. Setup

Let $G = (I, A)$ be a directed acyclic graph that describes a single-origin single-destination traffic network, with I and A denoting the set of nodes and the set of arcs, respectively. For each arc $a \in A$, we denote the start and end nodes of a by i_a and j_a, respectively . For each node $i \in I$, let $A_i^-, A_i^+ \subset A$ denote the set of incoming and outgoing arcs. Let $g_o \geq 0$ denote the traffic flow entering the network G at each iteration.

To traverse the network, commuters sequentially select from outgoing arcs at each intermediate node, from the origin o to the destination d. Each arc $a \in A$ is associated with a positive, strictly increasing *latency function* $s_a : [0, \infty) \to [0, \infty)$, which captures the time required to travel through arc a due to congestion produced by the traffic load $w_a \geq 0$, and a *toll* $p_a \geq 0$, the monetary value each traveler must pay to access the arc. Throughout the rest of the paper, we adopt a linear latency model, formally stated as follows[1].

*Assumption 1 (**Linear Latency Functions**):* For each arc $a \in A$, there exists a coefficient $\theta_a \in \mathbb{R}$ such that $s_a(w_a) = \theta_a w_a$.

The *cost* $c_a : [0, \infty)^3 \to [0, \infty)$ on each arc is then obtained by summing the travel time and toll:

$$\begin{aligned} c_a(\theta_a, w_a, p_a) &= s_a(w_a) + p_a \\ &= \theta_a w_a + p_a, \end{aligned}$$

while the *perceived cost* $\tilde{c}_a$ additionally includes a zero-mean stochastic error term $\delta_a \in \mathbb{R}$ that encapsulates variations in commuters' perception of travel time:

$$\begin{aligned} \tilde{c}_a(\theta_a, w_a, p_a) &= s_a(w_a) + p_a + \delta_a \\ &= \theta_a w_a + p_a + \delta_a, \end{aligned}$$

At every non-destination node $i \in I \backslash \{d\}$, commuters select among outgoing nodes $a \in A_i^+$ by computing their perceived minimum cost-to-go $\{\tilde{z}_a \in \mathbb{R} : a \in A_i^+\}$ on arc a:

$$\begin{aligned} &\tilde{z}_a(\theta, w, p) \qquad (1) \\ &:= \tilde{c}_a(\theta, w_a, p_a) + \mathbb{E}_\delta\Big[\min_{a' \in A_{j_a}^+} \tilde{z}_{a'}(\theta, w, p)\Big], \quad j_a \neq d, \\ &\tilde{z}_a(\theta, w, p) \\ &:= \tilde{c}_a(\theta, w_a, p_a), \quad j_a = d. \qquad (2) \end{aligned}$$

In this work, we adopt the *logit Markovian Model* [14, 15], under which the noise terms δ_a are described by the Gumbel distribution with scale (or, entropy) parameter $\beta > 0$. As a result, the expected cost-to-go z_a for each arc $a \in A$ admits the following closed-form expression:

$$z_a(\theta, w, p) = c_a(\theta_a, w_a, p_a) - \frac{1}{\beta} \ln\left(\sum_{a' \in A_{j_a}^+} e^{-\beta z_{a'}(\theta, w, p)}\right). \tag{3}$$

The corresponding equilibrium flow, called the *Markovian Traffic Equilibrium (MTE)* $\bar{w}^{\theta,\beta}(p) \in \mathbb{R}^{|A|}$ *corresponding to the latency function parameters* $\theta \in \mathbb{R}^{|A|}$*, entropy parameter* $\beta > 0$*, and toll vector* $p \in \mathbb{R}^{|A|}$, is the unique flow vector satisfying the following fixed point equation—For each non-destination node $i \in I \backslash \{d\}$ and outgoing arc $a \in A_i^+$:

$$\bar{w}_a^{\theta,\beta}(p) = \left(g_i + \sum_{a' \in A_i^+} \bar{w}_{a'}^{\theta,\beta}(p)\right)$$

[1]For an extension of our least-squares-based latency function estimation method to higher-degree polynomial latency functions, please see [4].

$$\cdot \frac{\exp(-\beta z_a(\theta, \bar{w}^{\theta,\beta}(p), p))}{\sum_{a' \in A_i^+} \exp(-\beta z_{a'}(\theta, \bar{w}^{\theta,\beta}(p), p))},$$
$$\bar{w}_a^{\theta,\beta}(p) \in \mathcal{W},$$

where $g_i := g_o$ if $i = o$ and $g_i = 0$ otherwise, and $\mathcal{W}$ is defined as the constraint set that enforces the conservation of traffic flow:

$$\mathcal{W} := \Bigg\{ w \in \mathbb{R}^{|A|} : \sum_{a \in A_i^+} w_a = \sum_{a \in A_i^-} w_a, \forall\, i \neq o, d, \quad (4)$$
$$\sum_{a \in A_o^+} w_a = g_o,\ w_a \geq 0, \forall\, a \in A \Bigg\}$$

B. Socially Optimal Tolls

The objective of toll implementation is to realign commuter's incentives and route selection decisions, to induce *perturbed social optimality* with respect to the logit Markovian model detailed in Section II-A, as defined below.

*Definition 1 (**Perturbed Socially Optimal Flow**):* Let the perturbed total weighted latency $L : \mathcal{W} \times \mathbb{R}^{|A|} \times \mathbb{R} \to \mathbb{R}$ be given by:

$$L(w, \theta, \beta) \quad (5)$$
$$:= \sum_{a \in A} w_a s_a(w_a) +$$
$$\frac{1}{\beta^\star} \sum_{i \in I \setminus \{d\}} \left[\sum_{a \in A_i^+} w_a \ln w_a - \left(\sum_{a \in A_i^+} w_a \right) \ln \left(\sum_{a \in A_i^+} w_a \right) \right].$$

We call $w^\star \in \mathcal{W}$ the *perturbed socially optimal flow* with latency parameters θ and entropy parameter $\beta > 0$ if it solves $\min_{w \in \mathcal{W}} L(w, \theta, \beta)$, with $\mathcal{W}$ given by (4).

In the perturbed total latency L defined above, the first component is the total latency on the network weighted by the traffic load on each arc, while the second component is a non-positive entropy term that achieves its minimum when the traffic load at each non-destination node allocates itself equally among all outgoing arcs. Thus, the entropy parameter β weights the total network latency against the tendency of commuters with imperfect information to explore among outgoing arcs at each intermediate node.

Since the minimization problem posed by Definition 1 is strictly convex, the perturbed socially optimal flow exists and is unique. Moreover, [7, 8] establish that, given a traffic network $G = (I, A)$ with latency function parameters $\theta \in \mathbb{R}^{|A|}$ and entropy parameter $\beta > 0$, there exists an *optimal toll* $\bar{p} \in \mathbb{R}^{|A|}$ whose corresponding MTE $\bar{w}^{\theta,\beta}(\bar{p})$ is perturbed socially optimal, and a dynamic tolling scheme that converges to the optimal toll. Those results, in the context of the online tolling problem considered in this work, are as summarized below. For more details, please see [8].

Proposition 1: There exists $\tilde{w} \in \mathcal{W}$ and $\bar{p} \in \mathbb{R}^{|A|}$ such that $\tilde{w} = \bar{w}^{\theta,\beta}(\bar{p})$ and $\bar{p}_a^t = \bar{w}_a^t \cdot \theta_a$ for each $a \in A$. Moreover, $\bar{w}$ is perturbed socially optimal, i.e., $\tilde{w} = arg\min_{w \in \mathcal{W}} L(w, \theta, \beta)$.

C. Online Learning Problem

Here, we pose the online learning problem that forms the central focus of this work. Let T denote the total number of iterations for which the algorithm is run. Consider a traffic network G with known node and arc structure (I, A), but unknown latency function parameters $\{\theta_a^\star : a \in A\}$ and entropy parameter $\beta^\star > 0$. We assume that $\theta^\star$ and $\beta^\star$ are bounded, as posed below.

*Assumption 2 (**Parameter Bounds**):* There exist constants $c_\theta, C_\theta, c_\beta > 0$ such that $\theta_a^\star \in [c_\theta, C_\theta]$ for each $a \in A$, and $\beta^\star > c_\beta$. The central authority has access to c_β but not necessarily c_θ or C_θ.

The above assumptions are not overly restrictive, since roads cannot be arbitrarily congestive, and travelers usually have some non-zero proclivity for selecting cost-minimizing arcs and routes. Moreover, as established in Section III, the arc latency parameter estimation errors $\|\theta^t - \theta^\star\|_2$ shrinks rapidly as t increases. This allows the true, unknown temperature parameter $\beta^\star$, and thus a lower bound for $\beta^\star$, to be estimated with increasing accuracy as more data is collected.

Now, consider ourselves in the position of a central traffic authority that wishes to minimize the perturbed total latency over the iterations $t \in [T]$, despite initially lacking knowledge of the function parameters $\theta \in \mathbb{R}^{|A|}$, and the underlying entropy parameter β. To accomplish this, at each iteration $t \in [T]$, we implement a toll vector $\hat{p}^t \in \mathbb{R}^{|A|}$, and observe the resulting MTE traffic load allocation $w^t := \bar{w}^{\theta^\star,\beta^\star}(p^t) \in \mathcal{W}$, as well as the random realizations of the travelers' latencies on each arc:

$$\ell_{a,j}^t = s_a(w_a^t) + \epsilon_{a,j}^t.$$

for each $j \in [\lfloor w^t \rfloor]$, where $\epsilon_{a,j}^t$ are independent 1-subGaussian random variables. We then use the flow data $\{w_a^t : a \in A\}$ and the latency data $\{\ell_{a,j}^t : a \in A, j \in [\lfloor w^t \rfloor]\}$ to update our estimates of the underlying, unknown latency function parameter $\theta^\star$ and entropy parameter $\beta^\star$, and correspondingly design our toll to implement at the next iteration $t + 1$. The cumulative regret R over the iterations $t \in [T]$ is thus given by:

$$R := \sum_{t=1}^{T} \left[L\big(\bar{w}^{\theta^\star,\beta^\star}(p^t), \theta^\star, \beta^\star\big) - L\big(\bar{w}^{\theta^\star,\beta^\star}(p^\star), \theta^\star, \beta^\star\big) \right] \quad (6)$$

The core tenet of the above framework is that, as we accumulate more data on the traffic flow and realized latencies, we can construct increasingly accurate estimates of $\theta^\star$ and $\beta^\star$, and consequently adapt our tolls p^t to reduce congestion in an increasingly effective manner.

III. Main Algorithm

In this section, we present the main components of our algorithm (Algorithm 1). Section III-A describes the least-squares estimator used to approximate the arc latency functions from collected flow data. Section III-B then discusses our novel approximation scheme for the unknown entropy parameter β. Finally, we present our main algorithm in Section III-C.

A. Least-Squares Estimator for Latency Function Parameters

First, we present the regularized least-squares estimator for the arc latency coefficients $\{\theta_a : a \in A\}$. At each iteration $t \in [T]$, for each arc $a \in A$, we observe the traffic flow at the current iteration, w_a^t, and latency data $\{\ell_{a,j}^k : a \in A, k \in [t], j \in \lfloor w_a^t \rfloor\}$, We then update the regularized least-squares estimate $\hat{\theta}_a^t > 0$ for the true coefficient $\theta_a^\star$, with regularizer $\lambda_a > 0$, as follows [2]:

$$\hat{\theta}_a^t := \arg\min_{\theta_a} \left(\sum_{j=1}^{t-1} \sum_{k=1}^{\lfloor w_a^j \rfloor} (\ell_{a,k}^j - \theta_a w_a^j)^2 + \lambda_a \|\theta_a\|_2^2 \right).$$

The following lemma states that these estimates, across iterations $t \in [T]$, lie within a neighborhood of the true parameter $\theta_a^\star$.

Lemma 1: [9] For each $t \in [T]$ arc $a \in A$, define:

$$V_a^t := \sum_{\tau=1}^{t} \lfloor w_a^t \rfloor (w_a^t)^2,$$

$$\gamma_a^t := \sqrt{\lambda_a} C_\theta + \sqrt{2 \ln T + 2 \ln \left(\frac{V_a^{t-1}}{\lambda_a} \right)}, \qquad (7)$$

and let the "good event" E be defined by:

$$E := \left\{ \forall\, t \in [T], \forall\, a \in A : |\hat{\theta}_a^{t-1} - \theta_a^\star| \leq \frac{\gamma_a^t}{\sqrt{V_a^{t-1}}} \right\}.$$

Then $\mathbb{P}(E) \geq 1 - \frac{|A|}{T}$.

Proof: (**Sketch**) We construct upper confidence bounds for the least square estimator using covering arguments and martingale theory, as is standard in the bandit literature (see [9], Chapter 20, and [4].) For details, please see Appendix A. ∎

In words, with probability at least $1 - \frac{|A|}{T}$, for each arc $a \in A$ at each iteration $t \in [T]$, the estimate $\hat{\theta}_a^t$ falls within the confidence interval $\left[\hat{\theta}_a^t - \frac{\gamma_a^t}{\sqrt{V_a^{t-1}}}, \hat{\theta}_a^t + \frac{\gamma_a^t}{\sqrt{V_a^{t-1}}} \right]$. Below, for convenience, we set:

$$\hat{\theta}_a^{t,-} := \hat{\theta}_a^t - \frac{\gamma_a^t}{\sqrt{V_a^t}},$$

$$\hat{\theta}_a^{t,+} := \hat{\theta}_a^t + \frac{\gamma_a^t}{\sqrt{V_a^t}}$$

B. Entropy Parameter Estimation

Intuitively, the entropy parameter governs the degree to which travelers at an intermediate node prefer to select an outgoing arc that minimizes the cost-to-go. Specifically, when $\beta \to \infty$, travelers at node i select with probability 1 an outgoing arc $a \in A_i^+$ that minimizes the cost-to-go; when $\beta \to 0$, travelers at node i select from all outgoing arcs with equal probability, essentially ignoring their cost-to-go values. As such, a natural approach for estimate β would begin by fixing a node $i^\star$, whose outgoing routes to the destination are relatively straightforward to describe. Then, we can analyze data that characterize the traffic flows and costs among its outgoing arcs at each iteration $t \in [T]$, to gain insight into the strength of the commuters' preference to minimize their cost-to-go, i.e., to estimate β.

[2]We assume $\lfloor w \rfloor \geq 1$, i.e., each arc is traversed upon by at least one commuter per iteration.

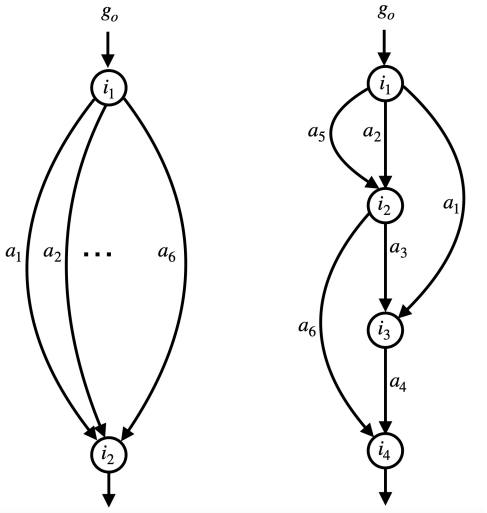

Fig. 1: (Left) A parallel 6-arc network; here, $i^\star = i_1$. (Right) A more general network with 6 arcs; here, $i^\star = i_2$, since there are two routes from i_2 to the destination i_4 which do not share an arc.

We thus begin with the following lemma, which states that regardless of the precise structure of the traffic network G, there must exist a node $i^\star \in I$ with properties desirable for estimating $\beta > 0$. For every node $i^\star$ satisfying the conditions of Lemma 2, each outgoing arc $a' \in A_{i^\star}^+$ yields exactly one route from $i^\star$ to d. Thus, the route segments from $i^\star$ to d have structure akin to a parallel-link network, allowing the estimation of the entropy parameter $\beta^\star$ from $i^\star$ to be straightforward. Examples are furnished in Figure 1.

Lemma 2: There exists a node $i^\star \in I \backslash \{d\}$ such that $|A_{i^\star}^+| \geq 2$, and for each $j \in A_{i^\star}^+$, either $j = d$, or there exists only one route from j to d.

Proof: (**Sketch**) This follows by starting from the destination d and recursively searching for the desired node $i^\star$ by moving back towards the origin o. For details, please see Appendix B. ∎

Below, we present assumptions that facilitate the estimation of the true, unknown temperature parameter $\beta^\star$. First, for each node $i^\star \in I \backslash \{d\}$, and any arc latency parameter estimate $\theta \in \mathbb{R}^{|A|}$ and temperature parameter $\beta > 0$ within a range of reasonable estimates for the true parameters $\theta^\star \in \mathbb{R}^{|A|}$ and $\beta^\star > 0$, we assume that the MTE costs of the outgoing edges $A_{i^\star}^+$ are not identical. In particular, for each such node $i^\star$, among the outgoing arcs A, there must be sufficiently differentiation, in the form of a strictly positive gap $\Delta_z > 0$, between the minimum and maximum costs-to-go. This facilitates the estimation of the temperature parameter in β, and emphasizes its role in the stochastic route choices made on the part of the travelers. Indeed, the temperature parameter β is not meaningful in networks with route segments that are virtually indistinguishable in cost.

Assumption 3: Let $\bar{p}(\hat{\theta}, \hat{\beta})$ denote the optimal toll corresponding to an arc-based TAM with entropy parameter $\hat{\beta}$, over a network with latency function parameters θ. There exists $\Delta_z > 0$, such that, for any node $i^\star \in I$ satisfying the conditions of Lemma 2, and any parameter estimates within known bounds, $\hat{\theta} \in [c_\theta, C_\theta]$ and $\hat{\beta} \in [c_\beta, \infty)$, we have:

$$\max_{a' \in A^+_{i^\star}} z_{a'}(\hat{\theta}, \bar{w}^{\theta^\star, \beta^\star}(\bar{p}(\hat{\theta}, \hat{\beta})), \bar{p}(\hat{\theta}, \hat{\beta})) - \min_{a' \in A^+_{i^\star}} z_{a'}(\theta, \bar{w}^{\theta^\star, \beta^\star}(\bar{p}(\hat{\theta}, \hat{\beta})), \bar{p}(\hat{\theta}, \hat{\beta})) \geq \Delta_z.$$

In the following lemma, we establish an estimator β^t for the temperature parameter β at each iteration t whose proximity to the true temperature parameter $\beta^\star$ is directly proportional to the gap between the under- and over-estimators $\theta^{t,-} \in \mathbb{R}^{|A|}$ and $\theta^{t,+} \in \mathbb{R}^{|A|}$ of the true arc latency parameter $\theta^\star$. The key intuition behind the estimator is that, if the true latency function parameters $\theta^\star$ on each arc were known, the underlying entropy parameter $\beta^\star$ can be perfectly recovered by comparing the flows of outgoing arcs at a non-destination node, and the ratios between the costs-to-go of these arcs. However, since the central authority lacks access to $\theta^\star$, we instead use the upper and lower bounds of the confidence interval at each iteration t, i.e., $\{\theta_a^{t,+}, \theta_a^{t,-} : a \in A\}$, to construct an estimate β^t of the underlying, unknown entropy parameter $\beta^\star$. Moreover, we construct the estimate β^t to provably under-approximate $\beta^\star$, i.e., to guarantee that $\beta^t \leq \beta^\star$. This can be viewed as an extension of the Principle of Optimism in the Face of Uncertainty, since the total latency (5) is non-decreasing in the entropy parameter β (Recall that the entropy term, to which the $1/\beta^\star$ factor is multiplied, is always non-positive).

Lemma 3: Let $i^\star \in I \backslash \{d\}$ be any node satisfying the conditions in Lemma 2, and let:

$$a^\star \in \arg\min_{a' \in A^+_{i^\star}} z_{a'}(\theta^\star, w^t, p^t). \tag{8}$$

Then there exists $\beta^t \in [c_\beta, \beta^\star]$ such that:

$$\frac{\exp(-\beta^t \cdot z_{a^\star}(\theta^{t,-}, w^t, p^t))}{\sum_{a' \in A^+_{i^\star}} \exp(-\beta^t \cdot z_{a'}(\theta^{t,+}, w^t, p^t))} = \frac{\exp(-\beta^\star \cdot z_{a^\star}(\theta^\star, w^t, p^t))}{\sum_{a' \in A^+_{i^\star}} \exp(-\beta^\star \cdot z_{a'}(\theta^\star, w^t, p^t))} \tag{9}$$

Moreover, let $A(i^\star)$ denote the set of all arcs contained in a route from $i^\star$ to d. Then:

$$|\beta^t - \beta^\star| \leq \frac{\beta^\star g_o}{\Delta_z} \cdot \sum_{a \in A(i^\star)} (\theta_a^{t,+} - \theta_a^{t,-}) w_a^t. \tag{10}$$

Proof: (**Sketch**) This follows from the monotonicity of the exponential function, as well as the monotonicity of the cost-to-go terms z_a with respect to the latency function parameters θ. For details, please see Appendix C. ∎

The upper bound (10) demonstrates that, by applying the least-squares estimator described in Section III-A, which ensures that $\|\theta^{t,+} - \theta^{t,-}\|_2 < O(1/\sqrt{t})$ as $t \to \infty$, we can likewise ensure that $|\beta^t - \beta^\star| < O(1/\sqrt{t})$ as $t \to \infty$.

C. Algorithm Overview

Armed with the estimation schemes for $\theta^\star$ and $\beta^\star$ presented in Sections III-A and III-B, we proceed to present our online learning algorithm (Algorithm 1). At each iteration t, the central authority uses latency function and entropy parameter estimates obtained in the previous round to compute the corresponding optimal toll p^t (Line 2). Observe that, for the latency function parameter, we use the lower bound $\theta^{t,-}$ of the confidence interval $(\theta^{t,-}, \theta^{t,+})$, in accordance with the Principle of Optimism in the Face of Uncertainty. Commuters then sequentially select arcs in the traffic network to minimize their average cost-to-go, resulting in the MTE traffic allocation $w^t := \bar{w}^{\theta^\star, \beta^\star}(p^t)$ (Line 3). The central authority then collects this data, and uses the regularized least-squares method in Section III-A to construct an updated estimate θ^t of the underlying latency function parameters $\theta^\star$ (Lines 5-11). Finally, we construct an update estimate β^t of the underlying entropy parameter $\beta^\star$ using the approach in Section III-B (Lines 14-15).

Algorithm 1: Simultaneous Tolling and Parameter Estimation

Data: $i^\star \in I$, $\beta^0 := c_\beta > 0$, λ_a, $V_a^0 = \lambda_a$, $Q_a^0 = 0$, and p_a^0 $\theta_a^{0,-} > 0$, $\theta_a^{0,+} > 0$ $(\forall a \in A)$

1 **for** $t = 1, \cdots, T$ **do**
2 $\quad p^t \leftarrow$ Solution to $p^t = \theta^{t-1,-} \cdot \bar{w}^{\theta^{t-1,-}, \beta^t}(p^t)$.
3 $\quad w^t \leftarrow \bar{w}^{\theta^\star, \beta^\star}(p^t)$ (Commuters' flow allocation)
4 $\quad$ **for** $a \in A$ **do**
5 $\quad\quad \ell_{a,1}^t, \cdots, \ell_{a,\lfloor w_a^t \rfloor}^t \leftarrow$ Costs collected from arc a at iteration t
6 $\quad\quad \gamma_a^t \leftarrow \sqrt{\lambda_a} C_\theta + \sqrt{2 \ln T + \ln\left(\frac{V_a^{t-1}}{\lambda_a}\right)}$
7 $\quad\quad \theta_a^{t,-} \leftarrow \max\left\{\hat{\theta}_a^{t-1} - \frac{\gamma_a^t}{\sqrt{V_a^{t-1}}}, 0\right\}$
8 $\quad\quad \theta_a^{t,+} \leftarrow \hat{\theta}_a^{t-1} + \frac{\gamma_a^t}{\sqrt{V_a^{t-1}}}$
9 $\quad\quad V_a^t \leftarrow V_a^{t-1} + \lfloor w_a^t \rfloor (w_a^t)^2$
10 $\quad\quad Q_a^t \leftarrow Q_a^{t-1} + w_a^t \cdot \sum_{k=1}^{\lfloor w_a^t \rfloor} \ell_{a,k}^t$
11 $\quad\quad \hat{\theta}_a^t \leftarrow Q_a^t / V_a^t$
12 $\quad$ **end**
13 $\quad \tilde{\beta}^t \leftarrow$ Solution to—$\forall a \in A_{i^\star}^+$:
$$\frac{w_a^t}{\sum_{a' \in A^+_{i^\star}} w_{a'}^t} = \frac{\exp(-\tilde{\beta}^t \cdot z_a(\theta^{t,-}, w^t, p^t))}{\sum_{a' \in A^+_{i^\star}} \exp(-\tilde{\beta}^t \cdot z_a(\theta^{t,+}, w^t, p^t))}.$$
15 $\quad \beta^t \leftarrow \max\{c_\beta, \tilde{\beta}^t\}$.
16 **end**

IV. Regret Analysis

Here, we upper bound the regret incurred by Algorithm 1. First, we require the following lemma, which facilitates the decomposition of the regret into tractable terms.

Lemma 4: Suppose $\theta_a^2 \geq \theta_a^1$ for each $a \in A$, and $\beta^2 \geq \beta^1$. Then, for each $w \in \mathcal{W}$:

$$L(w, \theta^1, \beta^1) \leq L(w, \theta^2, \beta^2).$$

Proof: This follows by noting that $w \geq 0$, and that the entropy term in L is non-positive. ∎

We now present our regret bound.

Theorem 1: There exists $K(\lambda, \Delta_z, c_\theta, C_\theta, c_\beta, \beta^\star) > 0$ such that for any $T \in \mathbb{N}$:

$$R \leq K g_o^2 \ln^2(g_o)|A|\sqrt{T}\ln(Tg_o)\max\left\{|I|\ln\left(\frac{|A|}{|I|}\right), B\right\},$$

where $B := |A(i^\star)|$ denotes the set of all arcs used to construct the estimates β^t.

Proof: (**Proof Sketch**) As in Algorithm 1, set $p^t \in \mathbb{R}^{|A|}$ and $p^\star \in \mathbb{R}^{|A|}$ to be the unique solutions to the following fixed-point equations:

$$p^t = \theta^{t-1,-} \cdot \bar{w}^{\theta^{t-1},\beta^{t-1}}(p^t),$$
$$p^\star = \theta^\star \cdot \bar{w}^{\theta^\star,\beta^\star}(p^\star).$$

Under the good event E described in Lemma 1:

$$\begin{aligned} L\big(\bar{w}^{\theta^{t,-},\beta^t}(p^t), \theta^{t,-}, \beta^t\big) &\leq L\big(\bar{w}^{\theta^\star,\beta^\star}(p^\star), \theta^{t,-}, \beta^t\big) \\ &\leq L\big(\bar{w}^{\theta^\star,\beta^\star}(p^\star), \theta^\star, \beta^\star\big), \end{aligned}$$

where the first inequality follows since Definition 1, Proposition 1, and the definition of p^t (Algorithm 1, Line 2) together imply that $\bar{w}^{\theta^{t,-},\beta^t}(p^t) = \arg\min_{w\in\mathcal{W}} L(w, \theta^{t,-}, \beta^\star, \beta^t)$, while the second inequality follows from Lemmas 1 and 4.

Define $\chi : \mathcal{W} \to \mathbb{R}$ to be the entropy term in C:

$$\begin{aligned} &\chi(w) \qquad (11)\\ &:= \sum_{i\in I\setminus\{d\}}\left[\sum_{a\in A_i^+} w_a \ln w_a - \left(\sum_{a\in A_i^+} w_a\right)\ln\left(\sum_{a\in A_i^+} w_a\right)\right] \end{aligned}$$

Thus, the regret R can be upper bounded as follows:

$$\begin{aligned} R &= \sum_{t=1}^T \big[L\big(\bar{w}^{\theta^\star,\beta^\star}(p^t), \theta^\star, \beta^\star\big) - L\big(\bar{w}^{\theta^\star,\beta^\star}(p^\star), \theta^\star, \beta^\star\big)\big] \\ &\leq \sum_{t=1}^T \big[L\big(\bar{w}^{\theta^\star,\beta^\star}(p^t), \theta^\star, \beta^\star\big) - L\big(\bar{w}^{\theta^{t,-},\beta^t}(p^t), \theta^{t,-}, \beta^t\big)\big] \\ &= \sum_{t=1}^T \big[L\big(\bar{w}^{\theta^\star,\beta^\star}(p^t), \theta^\star, \beta^\star\big) - L\big(\bar{w}^{\theta^\star,\beta^\star}(p^t), \theta^{t,-}, \beta^t\big)\big] \\ &\quad + \sum_{t=1}^T \big[L\big(\bar{w}^{\theta^\star,\beta^\star}(p^t), \theta^{t,-}, \beta^t\big) \\ &\qquad - L\big(\bar{w}^{\theta^{t,-},\beta^t}(p^t), \theta^{t,-}, \beta^t\big)\big] \\ &= \sum_{t=1}^T \sum_{a\in A} (\theta_a^\star - \theta_a^{t,-})\bar{w}_a^{\theta^\star,\beta^\star}(p^t)^2 \qquad (12) \\ &\quad + \sum_{t=1}^T \left(\frac{1}{\beta^\star} - \frac{1}{\beta^t}\right)\cdot\chi\big(\bar{w}_a^{\theta^\star,\beta^\star}(p^t)\big) \qquad (13) \\ &\quad + \sum_{t=1}^T \big[L\big(\bar{w}^{\theta^\star,\beta^\star}(p^t), \theta^{t,-}, \beta^t\big) \qquad (14) \\ &\qquad - L\big(\bar{w}^{\theta^{t,-},\beta^t}(p^t), \theta^{t,-}, \beta^t\big)\big], \end{aligned}$$

where, in accordance with the notation in Algorithm 1, we set $w^t := \bar{w}^{\theta^\star,\beta^\star}(p^t)$. Define the three summands (12), (13), (14) by R_1, R_2, and R_3 respectively. The convergence rate of $\theta^{t,-} \to \theta^\star$ and $\beta^t \to \beta^\star$ can then be analyzed to yield non-asymptotic bounds for R_1 and R_2, respectively. In turn, these bounds are then used to bound R_3.

For more details, please see Appendices E, F, and G. ∎

Remark 1: Compared to [4], our regret upper bound contains an extra term $\max\{|I|\ln(|A|/|I|), B\}$, due to the following unique features of our problem formulation: (1) Entropy parameter estimation, which contributes the network structure-dependent constant B, (2) The tolling authority affects the equilibrium flow allocation indirectly, through tolls, instead of directly dictating commuters' route selections, (3) Mismatch between the latency function and entropy parameter estimates $(\theta^{t,-}, \beta^t)$ used by the tolling authority to compute tolls, and the true parameters $(\theta^\star, \beta^\star)$ used by the commuters to best-respond to the implemented toll.

V. Experiments

We present numerical results on simulated traffic networks that validate the regret bounds presented in Theorem 1. We ran Algorithm 1 for $T = 2500$ iterations, with $g_o = 100$, on the parallel-arc network in Figure 1 (left), with underlying parameters $\theta^\star := (1.5, 2.5, 3.5, 4.5, 5.5, 6.5) \in \mathbb{R}^6$, and $\beta^\star = 0.25$, and on the more general network in Figure 1 (right), with underlying parameters $\theta^\star := (0.6, 0.4, 0.4, 0.4, 0.6, 0.6) \in \mathbb{R}^6$, and $\beta^\star = 0.25$. To suppress constants in the cumulative regret, we selected $\lambda_a = 0.01$ for each $a \in [6]$. For convenience, for each iteration $t \in [T]$, let $L^t := L(w^{\theta^\star,\beta^\star}(p^t), \theta^\star, \beta^\star)$ denote the cost incurred at iteration t, let $L^\star := L\big(w^{\theta^\star,\beta^\star}(p^\star), \theta^\star, \beta^\star\big)$ denote the minimum possible cost, and let $R^t := \sum_{\tau=1}^t \big[L\big(w^{\theta^\star,\beta^\star}(p^\tau), \theta^\star, \beta^\star\big) - L\big(w^{\theta^\star,\beta^\star}(p^\star), \theta^\star, \beta^\star\big)\big]$ denote the cumulative regret up to iteration t.

Figure 2 illustrates the growth of the cumulative regret $R^t - L^\star t$ as a function of the iteration count t. We also provide logarithmic plots that describe the decay of the stage-wise regret $L^t - L^\star$, the magnitude of the latency function parameter estimation error $\|\theta\|_2$, and the magnitude of the entropy parameter estimation error $|\beta^t - \beta|$. For both networks, the cumulative regret increases as a sub-linear function of t, while the cumulative regret, θ estimation error, and β estimation error decrease gracefully to 0 as t increases.

VI. Conclusion and Future Work

This work presents a novel online learning algorithm to learn the latency function and entropy parameters that characterize commuters' arc-selection decisions on a single source-single destination traffic network, while simultaneously implementing tolls to minimize the overall network congestion. We characterize a notion of regret using the accumulation across iterations of the gap between the incurred and minimum costs, and prove that our cumulative regret metric increases sub-linearly in the number of iterations t. Finally, we present numerical results illustrating the performance of our regret algorithm on simulated traffic networks.

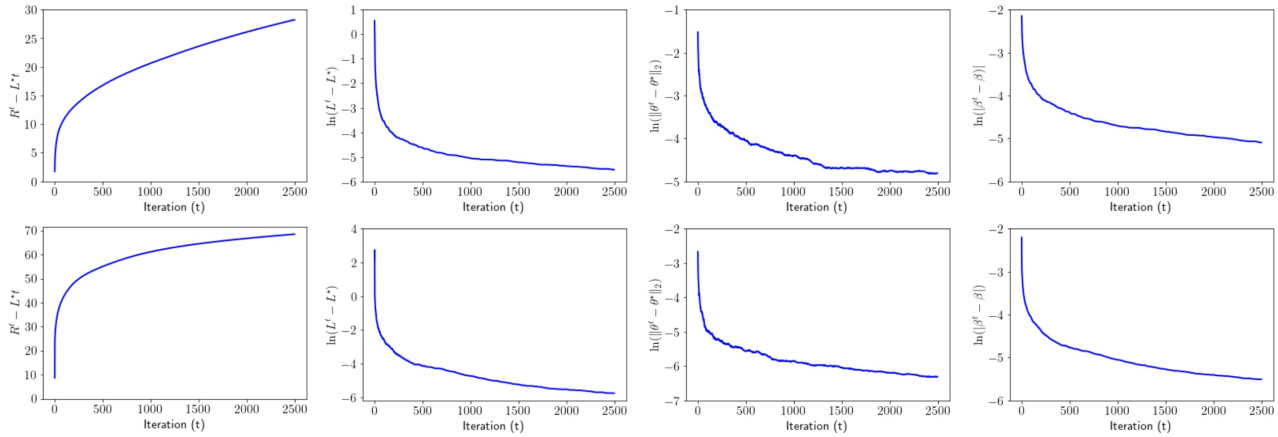

Fig. 2: (Left to right) The cumulative regret $R^t - L^\star t$, logarithm of stage-wise regret $\ln(L^t - L^\star)$, logarithm of θ-estimation error $\ln(\|\theta^t - \theta^\star\|_2)$, and logarithm of stage-wise regret $\ln(|\beta^t - \beta^\star|)$ for the parallel-arc network in Figure 1 (top) and the more general network in Figure 1 (bottom), as a function of the iteration count t. Note the sub-linear growth of the cumulative regret with respect to the iteration count, and the rapid decay of the stage-wise regret, θ-estimation error, and β-estimation error to 0.

A natural avenue of future work is to extend the results presented in this paper to traffic networks with multiple origin-destination pairs, and possibly bi-directional edges. Such settings pose particular challenges to the estimation of the entropy parameters, since each arc in the network could be shared among commuters with different travel histories and destinations. It would also be interesting to explore the relaxation of the assumption that the central authority possesses knowledge of a lower bound $c_\beta > 0$ for $\beta^\star$.

VII. Acknowledgements

The authors would like to thank Chinmay Maheshwari and Pan-Yang Su for fruitful discussions regarding the arc-based congestion game formulation considered in this work.

References

[1] Haripriya Pulyassary, Ruifan Yang, Zhanhao Zhang, and Manxi Wu. "Capacity Allocation and Pricing of High Occupancy Toll Lane Systems with Heterogeneous Travelers". In: *arXiv preprint arXiv:2304.09234* (2023).

[2] Dario Paccagnan, Rahul Chandan, Bryce L Ferguson, and Jason R Marden. "Incentivizing Efficient Use of Shared infrastructure: Optimal Tolls in Congestion Games". In: *arXiv preprint arXiv:1911.09806* (2019).

[3] José Correa, Cristóbal Guzmán, Thanasis Lianeas, Evdokia Nikolova, and Marc Schröder. "Network Pricing: How to Induce Optimal Flows Under Strategic Link Operators". In: *Operations Research* 70.1 (2022), pp. 472–489. DOI: 10.1287/opre.2020.2067.

[4] Sreenivas Gollapudi, Kostas Kollias, Chinmay Maheshwari, and Manxi Wu. "Online Learning for Traffic Navigation in Congested Networks". In: *International Conference on Algorithmic Learning Theory* (2023).

[5] Noriko Kanekoa, Daisuke Fukudab, and Qian Gec. "Optimal Congestion Tolling Problem under the Markovian Traffic Equilibrium". In: *Sustainability* (2021).

[6] Chinmay Maheshwari, Kshitij Kulkarni, Manxi Wu, and S. Shankar Sastry. "Dynamic Tolling for Inducing Socially Optimal Traffic Loads". In: *2022 American Control Conference (ACC)*. 2022, pp. 4601–4607. DOI: 10.23919/ACC53348.2022.9867193.

[7] Chih-Yuan Chiu, Chinmay Maheshwari, Pan-Yang Su, and Shankar Sastry. "Arc-based Traffic Assignment: Equilibrium Characterization and Learning". In: *62nd IEEE Conference on Decision and Control (CDC)* (2023).

[8] Chih-Yuan Chiu, Chinmay Maheshwari, Pan-Yang Su, and Shankar Sastry. "Dynamic Tolling in Arc-based Traffic Assignment Models". In: *59th Annual Allerton Conference on Communication, Control, and Computing* (2023).

[9] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020. DOI: 10.1017/9781108571401.

[10] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. "Improved Algorithms for Linear Stochastic Bandits". In: *Advances in Neural Information Processing Systems*. Ed. by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger. Vol. 24. Curran Associates, Inc., 2011.

[11] Yuki Oyama and Eiji Hato. "Prism-based Path Set Restriction for Solving Markovian Traffic Assignment Problem". In: *Transportation Research Part B: Methodological* 122 (2019), pp. 528–546.

[12] Yuki Oyama, Yusuke Hara, and Takashi Akamatsu. "Markovian Traffic Equilibrium Assignment Based on Network Generalized Extreme Value Model". In: *Transportation Research Part B: Methodological* 155 (2022), pp. 135–159.

[13] Paul B. Reverdy and Naomi Ehrich Leonard. "Parameter Estimation in Softmax Decision-Making Models With Linear Objective Functions". In: *IEEE Transactions on Automation Science and Engineering* 13 (2015), pp. 54–67.

[14] Takashi Akamatsu. "Decomposition of Path Choice Entropy in General Transport Networks". In: *Transportation Science* 31.4 (Nov. 1997), pp. 349–362. DOI: 10.1287/trsc.31.4.349.

[15] Jean-Bernard Baillon and Roberto Cominetti. "Markovian Traffic Equilibrium". In: *Mathematical Programming* (Feb. 2008). DOI: 10.1007/s10107-006-0076-2.

Appendix

Please use the following link to access a version with the appendix (https://drive.google.com/file/d/1LM7BwxI4ntOpy8J8TFyOyLXNHIBi4kL7/view?usp=sharing). The authors will make certain that this link stays active.

Below, we present proofs omitted in the main paper due to space limitations.

First, we recall the definitions of the *depth* and *height* of a graph, as defined in [7] Appendix A, and restated below for completeness.

*Definition 2 (**Depth of a DAG**):* Given a DAG $G = (I, A)$ describing a single-origin single-destination traffic network, the *depth of* G, denoted $\ell(G)$, is defined by:

$$\ell(G) := \max_{a \in A} \ell_a$$

Since the acyclic traffic graphs studied in this work have finitely many edges, we have $\ell(G) < \infty$. Below, we summarize properties of the depth of a DAG.

Proposition 2: Given a Condensed DAG $G = (I, A)$ with the route set $\mathbf{R}$:

1) For any $a \in A$, we have $\ell_a = 1$ if and only if $i_a = o$. Similarly, if $\ell_a = \ell(G)$, then $j_a = d$.
2) For any fixed $r \in \mathbf{R}$, and any $a, a' \in r$ with $\ell_{a,r} < \ell_{a',r}$, we have $\ell_a < \ell_{a'}$ i.e., arcs along a route have strictly increasing depth from the origin to the destination.
3) Fix any $a \in A$, and any $r \in \mathbf{R}$ containing a such that $\ell_{a,r} = \ell_a$. Then, for any $a' \in \mathbf{R}$ preceding a in r, we have $\ell_{a',r} = \ell_{a'}$.
4) For each depth $k \in [\ell(G)] := \{1, \cdots, \ell(G)\}$, there exists some $a \in A$ such that $\ell_a = k$.

Proof: See [7], Appendix A. ∎

Similarly, we can define and characterize the height of a DAG.

*Definition 3 (**Height of a DAG**):* Given a DAG $G = (I, A)$ describing a single-origin single-destination traffic network, the *height of* G, denoted $m(G)$, is defined by:

$$m(G) := \max_{a \in A} m_a$$

As with depth, we note that DAGs with finitely many edges have finite height, i.e., $m(G) < \infty$.

Proposition 3: Given an Condensed DAG $G = (I, A)$ with the route set $\mathbf{R}$:

1) For any $a \in A$, we have $m_a = 1$ if and only if $j_a = d$. Similarly, if $m_a = m(G)$, then $i_a = o$.
2) For any fixed $r \in \mathbf{R}$, and any $a, a' \in r$ with $m_{a,r} < m_{a',r}$, we have $m_a < m_{a'}$ i.e., arcs along a route from the origin to the destination have strictly decreasing depth.
3) Fix any $a \in A$, and any $r \in \mathbf{R}$ containing a such that $m_{a,r} = m_a$. Then, for any $a' \in \mathbf{R}$ following a in r, we have $m_{a',r} = m_{a'}$.
4) For each height $k \in [m(G)] := \{1, \cdots, m(G)\}$, there exists an arc $a \in A$ such that $m_a = k$.

Proof: See [7], Appendix A. ∎

A. Proof of Lemma 1

At each iteration $t \in [T]$, for each arc $a \in A$, the regularized least-squares estimate $\hat{\theta}_a^t > 0$ for the true coefficient $\theta_a^\star$, with regularizer $\lambda_a > 0$, is given by:

$$\hat{\theta}_a^t := \arg\min_{\theta_a} \left(\sum_{j=1}^{t-1} \sum_{k=1}^{\lfloor w_a^j \rfloor} (\ell_{a,k}^j - \theta_a w_a^j)^2 + \lambda_a \|\theta_a\|_2^2 \right).$$

Note that the cost objective in the above argmin expression is convex and quadratic. Thus, by setting the gradient to 0, we can compute the optimal parameter estimate as follows (for more details, please see Gollapudi et al. [4], Lemma 2):

$$\hat{\theta}_a^t = \left(\lambda_a + \sum_{j=1}^{t-1} (w_a^j)^3 \right)^{-1} \left(\sum_{j=1}^{t-1} w_a^j \cdot \sum_{k=1}^{\lfloor w_a^j \rfloor} \ell_{a,k}^j \right) \tag{15}$$

For convenience, we define:

$$V_a^t := \lambda_a + \sum_{j=1}^{t-1} (w_a^j)^3, \tag{16}$$

$$W_a^t := \sum_{j=1}^{t-1} (w_a^j)^3, \tag{17}$$

$$U_a^t := \sum_{j=1}^{t-1} w_a^j \cdot \sum_{k=1}^{\lfloor w_a^j \rfloor} \ell_{a,k}^j, \tag{18}$$

$$S_a^t := \sum_{j=1}^{t-1} w_a^j \cdot \sum_{k=1}^{\lfloor w_a^j \rfloor} \epsilon_{a,k}^j. \tag{19}$$

Thus, we can write (15) as:

$$\hat{\theta}_a^t = (V_a^t)^{-1} U_a^t = (V_a^t)^{-1} (W_a^t \theta_a + S_a^t). \tag{20}$$

For each arc $a \in A$, the above process generates regularized least-squares estimates $\{\hat{\theta}_a^t\}$, across iterations $t \in [T]$, for the true underlying parameter $\theta_a^\star$. The following lemma demonstrates that these estimates, across iterations $t \in [T]$, lie within a neighborhood of the true parameter $\theta_a^\star$.

Proof: (**Proof of Lemma 1**) The following proof parallels that of Gollapudi et al. [4], Lemma 3, and is included for completeness.

From (20), we have:

$$\begin{aligned}
&\sqrt{V_a^t} |\hat{\theta}_a^t - \theta_a^\star| \\
&= \sqrt{V_a^t} |(V_a^t)^{-1} (W_a^t \theta_a + S_a^t) - \theta_a| \\
&= \sqrt{V_a^t} |(V_a^t)^{-1} S_a^t + \big((V_a^t)^{-1} W_a^t - 1\big)\theta_a| \\
&= \sqrt{V_a^t} |(V_a^t)^{-1} S_a^t + \big((V_a^t)^{-1} (V_a^t - \lambda_a) - 1\big)\theta_a| \\
&= \sqrt{V_a^t} |(V_a^t)^{-1} S_a^t - \lambda_a (V_a^t)^{-1} \theta_a| \\
&= (V_a^t)^{-1/2} |S_a^t| + \sqrt{\lambda_a} \theta_a.
\end{aligned} \tag{21}$$

To bound $(V_a^t)^{-1/2} |S_a^t|$, define $M_a^t(z) := \exp\left(z S_a^t - \frac{1}{2} V_a^t z^2\right)$ for each $z \in \mathbb{R}$. Then, for any fixed $z \in \mathbb{R}$:

$$\begin{aligned}
&\mathbb{E}[M_a^t(z) | \mathcal{F}_a^t] \\
&= M_a^{t-1}(z) \cdot \mathbb{E}\left[\exp\left(w_a^t \cdot \sum_{k=1}^{\lfloor w_a^t \rfloor} \epsilon_{a,k}^t z - \frac{1}{2} \lfloor w_a^t \rfloor (w_a^t)^2 z^2 \right) \middle| \mathcal{F}_a^t \right] \\
&= M_a^{t-1}(z) \cdot \prod_{k=1}^{\lfloor w_a^t \rfloor} \mathbb{E}\left[\exp\left(w_a^t \cdot \epsilon_{a,k}^t z - \frac{1}{2} \lfloor w_a^t \rfloor (w_a^t)^2 z^2 \right) \middle| \mathcal{F}_a^t \right] \\
&\leq M_a^{t-1}(z).
\end{aligned}$$

so $M_a^t(z)$ is a supermartingale adapted to the filtration $\mathcal{F}_a^t := \sigma(w_a^1, \tilde{s}_a^1)$. Thus, so is $\tilde{M}_a^t := \mathbb{E}_{z \sim \mathcal{N}(0,1)}[M_a^t(z)]$. It thus

follows from Lattimore and Szepesvari [9], Theorem 20.4, that:

$$(V_a^t)^{-1/2}|S_a^t| \leq \sqrt{2\ln t + \ln\left(\frac{V_a^t}{\lambda_a}\right)}. \tag{22}$$

The proof now follows from (21) and (22). ∎

B. Proof of Lemma 2

Proof: (**Proof of Lemma 2**) By assumption, the graph G contains more than one route from the origin o to the destination d. Thus, there exists some $a \in A$ such that $|A_{i_a}^+| \geq 2$, so the quantity:

$$m^\star := \min\{m_a : a \in A, |A_{i_a}^+| \geq 2\}$$

is well-defined. Now, fix any $a \in A$ such that $m_a = m^\star$, and $|A_{i_a}^+| \geq 2$. It suffices to show that, for each $j \in A_{i_a}^+$, there exists only one route connecting j to the destination d. Suppose by contradiction that there exists some $j' \in A_{i_a}^+$ such that at least two distinct routes connect j' to d. Let $\bar{j} \in I\backslash\{d\}$ denote any node at which these routes diverge. Then for any $\bar{a} \in A_{\bar{j}}^\star$, we have $|A_{i_{\bar{a}}}^+| = |A_{\bar{j}}^+| \geq 2$, and:

$$m_{\bar{a}} < m_a = m^\star,$$

a contradiction to the definition of $m^\star$. This concludes the proof. ∎

C. Proof of Lemma 3

Proof: (**Proof of Lemma 3**) Fix $t \in [T]$. Define $\kappa_{a^\star}^t \in \mathbb{R}$ by:

$$\begin{aligned}\kappa_{a^\star}^t &:= \frac{\exp(-\beta^\star \cdot z_{a^\star}(\theta^\star, w^t, p^t))}{\sum_{a' \in A_{i^\star}^+} \exp(-\beta^\star \cdot z_{a'}(\theta^\star, w^t, p^t))} \\ &= \frac{w_{a^\star}^t}{\sum_{a' \in A_{i^\star}^+} w_{a'}^t},\end{aligned}$$

and let $f^t, g^t : \mathbb{R} \times \mathbb{R}^{|A|} \times \mathbb{R}^{|A|} \to \mathbb{R}$ be given as follows:

$$\begin{aligned}&f^t(\beta, \theta^+, \theta^-) \\ &:= \frac{\exp(-\beta^t \cdot z_{a^\star}(\theta^{t,-}, w^t, p^t))}{\sum_{a' \in A_{i^\star}^+} \exp(-\beta^t \cdot z_{a'}(\theta^{t,+}, w^t, p^t))}, \\ &g^t(\beta, \theta^+, \theta^-) \\ &:= \ln f^t(\beta, \theta^+, \theta^-) - \ln \kappa_a^t \\ &= -\beta \cdot z_{a^\star}(\theta^-, w^t, p^t) \\ &\qquad - \ln\left(\sum_{a' \in A_{i^\star}^+} \exp\left(-\beta \cdot z_{a'}(\theta^+, w^t, p^t)\right)\right) - \ln \kappa_{a^\star}^t\end{aligned}$$

Note that $g^t(\beta^t, \theta^+, \theta^-) = 0$ holds if and only if $f^t(\beta^t, \theta^+, \theta^-) = \kappa_a^t$. If one takes $\theta^+ = \theta^{t,+}$ and $\theta^- = \theta^{t,+}$ this becomes a restatement of (9). We note that $z_a(\theta, w, p)$ is continuously differentiable for each $a \in A$, $\theta \in \mathbb{R}^{|A|}$, $w \in \mathcal{W}$, and $p \in \mathbb{R}^{|A|}$, and the log-sum-exp function is continuously differentiable in the entropy parameter β. Thus, f^t and g^t are likewise continuously differentiable at each $\beta > 0$ and each $\theta^+, \theta^- \in \mathbb{R}^{|A|}$.

The remainder of the proof proceeds in two parts. We first prove that, given any fixed values $\theta_a^+ \geq \theta_a^\star$, $\theta_a^- \leq \theta_a^\star$ for each $a \in A$, there exists a unique fixed point solution β to the function $g^t(\beta^t, \theta^+, \theta^-) = 0$. In particular, given $\theta_a^{t,+} \geq \theta_a^\star$, $\theta_a^{t,-} \leq \theta_a^\star$ for each $a \in A$, there exists a unique entropy parameter estimate $\beta^t > 0$ that solves $g^t(\beta^t, \theta^{t,+}, \theta^{t,-}) = 0$, i.e., that satisfies (9), and $\beta^\star$ is the unique entropy parameter value that satisfies $g^t(\beta^\star, \theta^\star, \theta^\star) = 0$. We then bound the gap between $\beta^\star$ and β by bounding the difference between $\theta^{t,+}$ and $\theta^\star$, and between $\theta^{t,-}$ and $\theta^\star$.

1) Claim—Given any fixed $\theta_a^+ \geq \theta_a^\star$, $\theta_a^- \leq \theta_a^\star$ for each $a \in A$, there exists a unique fixed point solution β to the function $g^t(\beta^t, \theta^+, \theta^-) = 0$:

Proof: To show that, for any $\theta^{t,+}, \theta^{t,-} \in \mathbb{R}^{|A|}$, the fixed-point equation $g^t(\beta^\star, \theta^{t,+}, \theta^{t,-}) = 0$, has a unique solution (or equivalently that $f^t(\beta, \theta^+, \theta^-) = \kappa_a^t$ has a unique solution), we first note that:

$$\begin{aligned}\frac{1}{|A_{i^\star}^+|} &\leq \kappa_{a^\star}^t \\ &= \frac{\exp(-\beta^\star \cdot z_{a^\star}(\theta^\star, w^t, p^t))}{\sum_{a' \in A_{i^\star}^+} \exp(-\beta^\star \cdot z_{a'}(\theta^\star, w^t, p^t))} \\ &< 1.\end{aligned}$$

and that $f^t(0, \theta^+, \theta^-) = 1/|A_{i^\star}^+|$. Below, we establish that $\lim_{\beta \to \infty} f^t(\beta, \theta^+, \theta^-) = 1/|A_{i^\star}^+|$, by lower bounding $\frac{\partial g^t}{\partial \beta}$. The existence and uniqueness of a solution β to the fixed-point equation $f^t(\beta, \theta^+, \theta^-) = \kappa_{a^\star}^t$ then follows from the Intermediate Value Theorem.

To compute derivatives of g^t, we observe that, since $i_{a^\star} = i^\star$ satisfies the conditions of Lemma 2, for each $a' \in A_{i^\star}^+$, there exists exactly one route that connects $j_{a'}$ and d. As a result, $z_{a'}(\theta^+, w^t, p^t)$ equals the sum of latencies on a' and on arcs comprising that route, and therefore does not depend on the entropy parameter β. Thus:

$$\begin{aligned}&\frac{\partial g^t}{\partial \beta}(\beta, \theta^+, \theta^-) \\ &= -z_{a^\star}(\theta^-, w^t, p^t) \\ &\qquad + \frac{\sum_{a' \in A_{i^\star}^+} e^{-\beta \cdot z_{a'}(\theta^{t,-}, w^t, p^t)} \cdot z_{a'}(\theta^{t,-}, w^t, p^t)}{\sum_{a' \in A_{i^\star}^+} e^{-\beta \cdot z_{a'}(\theta^{t,-}, w^t, p^t)}} \\ &= -z_{a^\star}(\theta^-, w^t, p^t) \\ &\qquad + \sum_{\bar{a} \in A_{i^\star}^+} \frac{e^{-\beta \cdot z_{\bar{a}}(\theta^{t,-}, w^t, p^t)}}{\sum_{a' \in A_{i^\star}^+} e^{-\beta \cdot z_{a'}(\theta^{t,-}, w^t, p^t)}} \\ &\qquad\qquad \cdot z_{\bar{a}}(\theta^{t,-}, w^t, p^t) \\ &= \sum_{\bar{a} \in A_{i^\star}^+} \frac{e^{-\beta \cdot z_{\bar{a}}(\theta^{t,-}, w^t, p^t)}}{\sum_{a' \in A_{i^\star}^+} e^{-\beta \cdot z_{a'}(\theta^{t,-}, w^t, p^t)}} \\ &\qquad\qquad \cdot \left[z_{\bar{a}}(\theta^{t,-}, w^t, p^t) - z_{a^\star}(\theta^{t,-}, w^t, p^t)\right]\end{aligned}$$

$$= \sum_{\bar{a}\in A^+_{i^\star}} \frac{w^t_{\bar{a}}}{\sum_{a'\in A^+_{i^\star}} w^t_{a'}} \cdot \Big[z_{\bar{a}}(\theta^{t,-}, w^t, p^t) - z_{a^\star}(\theta^{t,-}, w^t, p^t)\Big].$$

The flow continuity equations imply that $\sum_{a'\in A^+_{i^\star}} w^t_{a'} \leq g_o$; together with the assumption that $w^t_a \geq 1$ for each $a \in A$, we have:

$$\frac{w^t_{\bar{a}}}{\sum_{a'\in A^+_{i^\star}} w^t_{a'}} \geq \frac{1}{g_o}.$$

Combining this with the definition of Δ_z, we obtain:

$$\frac{\partial g^t}{\partial \beta}(\beta, \theta^+, \theta^-) \geq \frac{\Delta_z}{g_o}. \tag{23}$$

Thus, $g^t(\beta, \theta^+, \theta^-)$ increases to $+\infty$ as $\beta \to \infty$, and therefore so does f^t.

To reiterate for emphasis, this claim establishes the unique existence of a entropy parameter estimate $\beta^t > 0$ that satisfies $g^t(\cdot, \theta^{t,+}, \theta^{t,-}) = 0$, or equivalently, (9). This claim also establishes that $\beta = \beta^\star$ is the unique solution to $g^t(\cdot, \theta^\star, \theta^\star) = 0$.

2) Claim—We have:

$$|\beta^t - \beta^\star| = \frac{\beta^\star g_o}{\Delta_z} \cdot \sum_{a\in A(i^\star)} (\theta^{t,+}_a - \theta^{t,-}_a) w^t_a.$$

Proof: For convenience, we denote $\theta^\pm := (\theta^+, \theta^-) \in \mathbb{R}^{2|A|}$. For any $\theta^\pm \in \mathbb{R}^{2|A|}$ such that $\theta^+_a > \theta^\star_a$ and $\theta^-_a < \theta^\star_a$ for each $a \in A$, let $\beta = \hat{\beta}(\theta^+, \theta^-)$ denote the unique solution to $g^t(\beta, \theta^+, \theta^-)$. Note that for any fixed $w \in \mathcal{W}$ and $p \in \mathbb{R}^{|A|}$, since $z_{a^\star}(\theta, w, p)$ is component-wise increasing in θ, we have $f^t(0, \theta^+, \theta^-) \leq \kappa^t_a \leq f^t(\beta^\star, \theta^+, \theta^-)$. It thus follows from the Intermediate Value Theorem that $\hat{\beta}(\theta^+, \theta^-) \in [0, \beta^\star]$.

By (23), we have $\frac{\partial g^t}{\partial \beta}(\beta, \theta^+, \theta^-) \neq 0$ at each $\beta > 0$. This allows us to apply the Implicit Function Theorem, which yields that $\hat{\beta}$ is continuously differentiable in $\theta^\pm$, with:

$$\frac{\partial \hat{\beta}}{\partial \theta^\pm}(\theta^\pm) = \left[\frac{\partial g}{\partial \beta}(\beta, \theta^+, \theta^-)\right]^{-1} \left[\frac{\partial g}{\partial \theta^\pm}(\beta, \theta^+, \theta^-)\right]$$

Now, define $u_+ := \theta^{t,+} - \theta^\star$ and $u_- := \theta^{t,-} - \theta^\star$. We then have:

$$\begin{aligned}
&|\beta^t - \beta^\star| \\
&= \left| \int_0^1 \frac{\partial \hat{\beta}}{\partial \theta^\pm}(\theta^+ + \sigma u_+, \theta^- + \sigma u_-)^\top \, dt \right| \\
&= \left| \int_0^1 \left[\frac{\partial g}{\partial \beta}(\beta, \theta^\star - \sigma u_+, \theta^\star + \sigma u_-)\right]^{-1} \right. \\
&\qquad \cdot \frac{\partial g}{\partial \theta^\pm}(\beta, \theta^+ + \sigma u_-, \theta^- + \sigma u_+) \\
&\qquad \left. \cdot (\theta^+ - \theta^\star, \theta^- - \theta^\star) \, dt \right| \\
&\leq \frac{g_o}{\Delta_z} \cdot \int_0^1 \left| \frac{\partial g}{\partial \theta^\pm}(\beta, \theta^+ + \sigma u_-, \theta^- + \sigma u_+) \right. \\
&\qquad \left. \cdot (\theta^+ - \theta^\star, \theta^- - \theta^\star) \right| dt \\
&= \frac{g_o}{\Delta_z} \cdot \int_0^1 \left| \sum_{a\in A} \frac{\partial g}{\partial \theta^+_a}(\beta, \theta^+ + \sigma u_-, \theta^- + \sigma u_+) \right. \\
&\qquad \cdot (\theta^{t,+}_a - \theta^\star) \\
&\qquad + \sum_{a\in A} \frac{\partial g}{\partial \theta^-_a}(\beta, \theta^+ + \sigma u_-, \theta^- + \sigma u_+) \\
&\qquad \left. \cdot (\theta^{t,-}_a - \theta^\star) \right| dt,
\end{aligned}$$

where the inequality follows from (23). Next, let $A(i^\star)$ denote the set of all arcs along routes from the node $i^\star$ to the destination node d. Now, observe that, for any $a \in A$, $\beta > 0$ and $\theta^+, \theta^- \in \mathbb{R}^{|A|}$:

$$\begin{aligned}
\frac{\partial g}{\partial \theta^+_a}(\beta, \theta^+, \theta^-) &= -\beta w^t_a \cdot \mathbf{1}\{a \in A(i^\star)\}, \\
\frac{\partial g}{\partial \theta^-_a}(\beta, \theta^+, \theta^-) &= \frac{\exp(-\beta \cdot z_{\bar{a}}(\theta^{t,-}, w^t, p^t))}{\sum_{a'\in A^+_{i^\star}} \exp(-\beta \cdot z_{a'}(\theta^{t,-}, w^t, p^t))} \\
&\qquad \cdot \beta w^t_a \cdot \mathbf{1}\{a \in A(i^\star)\}.
\end{aligned}$$

Substituting into the above upper bound for $|\beta^t - \beta^\star|$, we obtain:

$$\begin{aligned}
&|\beta^t - \beta^\star| \\
&\leq \frac{g_o}{\Delta_z} \cdot \int_0^1 \sum_{a\in A(i^\star)} \left| \frac{\partial g}{\partial \theta^+_a}(\beta, \theta^+ + \sigma u_-, \theta^- + \sigma u_+) \right| \\
&\qquad \cdot (\theta^{t,+}_a - \theta^\star) \\
&\qquad + \sum_{a\in A(i^\star)} \left| \frac{\partial g}{\partial \theta^-_a}(\beta, \theta^+ + \sigma u_-, \theta^- + \sigma u_+) \right| \\
&\qquad \cdot (\theta^\star - \theta^{t,-}_a) \, dt \\
&\leq \frac{\beta^\star g_o}{\Delta_z} \cdot \sum_{a\in A(i^\star)} (\theta^{t,+}_a - \theta^{t,-}_a) w^t_a,
\end{aligned}$$

as desired. ∎

Notation: Throughout the appendix, the notation $x \lesssim y$ denotes that there exists some constant $K(\lambda, \Delta_z, c_\theta, C_\theta, c_\beta, \beta^\star)$, such that $x \leq Ky$.

D. Preliminary Lemmas

This subsection presents preliminary lemmas that will facilitate the proof of Theorem 1. We begin with a result derived from the Fundamental Theorem of Calculus.

Lemma 5: If $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable, then, for each $x_1, x_2 \in \mathbb{R}^n$:

$$\begin{aligned}
&\|f(x_2) - f(x_1)\|_2 \\
&\leq \max_{t\in[0,1]} \left\| \frac{\partial f}{\partial x}\big(x_1 + t(x_2 - x_1)\big) \right\|_2 \cdot \|x_2 - x_1\|_2.
\end{aligned}$$

Proof: Fix $x_1, x_2 \in \mathbb{R}^n$. For each $i \in [m]$, let $f_i : \mathbb{R}^n \to \mathbb{R}$ denote the i-th component of the map f. Define $g_i : \mathbb{R} \to \mathbb{R}$ by:

$$g_i(t) := f\big(x_1 + t(x_2 - x_1)\big).$$

Then, for each $x_1, x_2 \in \mathbb{R}^n$ and each $i \in [m]$:

$$\begin{aligned} & f_i(x_2) - f_i(x_1) \\ &= g(1) - g(0) \\ &= \int_0^1 \frac{dg_i}{dt}(t)\, dt \\ &= \int_0^1 \frac{\partial f_i}{\partial x}\big(x_1 + t(x_2 - x_1)\big)\, dt \cdot (x_2 - x_1). \end{aligned}$$

Concatenating the above equality across $i \in [m]$, we obtain:

$$f(x_2) - f(x_1) = \int_0^1 \frac{\partial f}{\partial x}\big(x_1 + t(x_2 - x_1)\big)\, dt \cdot (x_2 - x_1).$$

Finally, we apply the Cauchy-Schwarz inequality to obtain:

$$\begin{aligned} & \|f(x_2) - f(x_1)\|_2 \\ &\le \int_0^1 \left\| \frac{\partial f}{\partial x}\big(x_1 + t(x_2 - x_1)\big) \right\|_2 dt \cdot \|x_2 - x_1\|_2 \\ &\le \max_{t \in [0,1]} \left\| \frac{\partial f}{\partial x}\big(x_1 + t(x_2 - x_1)\big) \right\|_2 \cdot \|x_2 - x_1\|_2, \end{aligned}$$

as desired. ∎

Below, we establish a collection of upper bounds that will be used repeatedly throughout the remainder of the proofs (Lemmas 6 and 7).

Lemma 6: For any $a \in A$ and $t \in [T]$:

$$\gamma_a^t \lesssim \sqrt{\ln(Tg_o)}.$$

Proof: Recall the definition of γ_a^t in (7). After taking $\lambda_a = 1$, we have, for any $t \ge 2$:

$$\begin{aligned} \gamma_a^t &= \sqrt{\lambda_a} C_\theta + \sqrt{2\ln T + 2\ln\left(\frac{V_a^{t-1}}{\lambda_a}\right)} \\ &= C_\theta + \sqrt{2\ln T + 2\ln\left(1 + \sum_{t=1}^{t-1} \lfloor w_a^t \rfloor (w_a^t)^2\right)} \\ &\le C_\theta + \sqrt{2\ln T + 2\ln\big(1 + (t-1)g_o^3\big)} \\ &\lesssim \sqrt{\ln(Tg_o)}. \end{aligned}$$

This result can be straightforwardly extended to the $t = 1$ case by ensuring that the constant encapsulated in the "$\lesssim$" is selected to be large enough. ∎

Lemma 7: For any $a \in A$:

$$\sum_{t=1}^{T} \min\left\{1, \frac{\lfloor w_a^t \rfloor (w_a^t)^2}{V_a^{t-1}}\right\} \lesssim \ln(Tg_o).$$

Proof: First, observe that $\min\{1, x\} \le \frac{1}{\ln 2} \cdot \ln(1 + x)$ for each $x \ge 0$. Thus:

$$\begin{aligned} & \sum_{t=1}^{T} \min\left\{1, \frac{\lfloor w_a^t \rfloor (w_a^t)^2}{V_a^{t-1}}\right\} \\ &\le \frac{1}{\ln 2} \cdot \sum_{t=1}^{T} \ln\left(1 + \frac{\lfloor w_a^t \rfloor (w_a^t)^2}{V_a^{t-1}}\right) \\ &= \frac{1}{\ln 2} \cdot \sum_{t=1}^{T} \ln\left(\frac{V_a^{t-1} + \lfloor w_a^t \rfloor (w_a^t)^2}{V_a^{t-1}}\right) \\ &= \frac{1}{\ln 2} \cdot \sum_{t=1}^{T} \ln\left(\frac{V_a^{t-1} + \lfloor w_a^t \rfloor (w_a^t)^2}{V_a^{t-1}}\right) \\ &\le \frac{1}{\ln 2} \cdot \sum_{t=1}^{T} \ln\left(\frac{V_a^{t}}{V_a^{t-1}}\right) \\ &= \frac{1}{\ln 2} \cdot \ln V_a^T \\ &\le \frac{1}{\ln 2} \cdot \ln\big(1 + Tg_o^3\big) \\ &\lesssim \ln(Tg_o), \end{aligned}$$

as desired. ∎

Next, we bound the weighted sums of the magnitudes of the latency function parameter errors $\theta^{t,-} - \theta^\star$ and entropy parameter $\beta^t - \beta^\star$ across iterations $t \in [T]$. First, we require the following lemma.

Lemma 8: Under the good event E, for any $p > 0$:

$$\sum_{t=1}^{T} \sum_{a \in A} |\theta_a^{t,-} - \theta_a^\star| (w_a^t)^p \lesssim g_o^p |A| \sqrt{T} \ln(Tg_o). \tag{24}$$

Proof: The desired result follows by taking $p = 2$ in Lemma 8. ∎

Lemma 9: Recall that B denotes the number of arcs along routes from $i^\star$ to d, which are used to construct an estimate of $\beta^\star$ at each iteration t. Under the good event E:

$$\sum_{t=1}^{T} |\beta^t - \beta^\star| \lesssim g_o B \sqrt{T} \ln(Tg_o).$$

Proof: Let $A(i^\star)$ denote the set of all arcs on routes from $i^\star$ to d. By Lemma 3, under the good event E, we have $\beta^t \in [c_\beta, \beta^\star]$, so $|\beta^t - \beta^\star| \le \beta^\star - c_\beta$. Moreover, from (10), we have:

$$|\beta^t - \beta^\star| \lesssim g_o \cdot \sum_{a \in A(i^\star)} (\theta_a^{t,+} - \theta_a^{t,-}) w_a^t$$

We then have:

$$\begin{aligned} & \sum_{t=1}^{T} |\beta^t - \beta^\star| \\ &\lesssim g_o \cdot \sum_{t=1}^{T} \left| \min\left\{ \beta^\star - c_\beta, \sum_{a \in A(i^\star)} (\theta_a^{t,+} - \theta_a^{t,-}) w_a^t \right\} \right| \\ &\lesssim g_o \cdot \sum_{t=1}^{T} \min\left\{ 1, \sum_{a \in A(i^\star)} (\theta_a^{t,+} - \theta_a^{t,-}) w_a^t \right\}. \end{aligned}$$

Take $\tilde{a} \in \max_{a \in A(i^\star)} \big\{ \sum_{t=1}^{T} (\theta_a^{t,+} - \theta_a^{t,-}) w_a^t \big\}$. Then:

$$\sum_{t=1}^{T} |\beta^t - \beta^\star|$$

$$
\begin{aligned}
&\lesssim g_o \cdot \sum_{t=1}^{T} \min\left\{1, B \cdot \frac{2\gamma_{\tilde{a}}^t}{\sqrt{V_{\tilde{a}}^{t-1}}} w_a^t\right\} \\
&\leq 4 g_o B \gamma_{\tilde{a}}^T \cdot \sqrt{T} \cdot \sqrt{\sum_{t=1}^{T} \min\left\{1, \frac{1}{V_{\tilde{a}}^{t-1}}(w_a^t)^2\right\}} \\
&\leq 4 g_o B \gamma_{\tilde{a}}^T \cdot \sqrt{T} \cdot \sqrt{\sum_{t=1}^{T} \min\left\{1, \frac{\lfloor w_a^t \rfloor (w_a^t)^2}{V_{\tilde{a}}^{t-1}}\right\}} \\
&\lesssim g_o B \sqrt{T} \ln(T g_o)
\end{aligned}
$$

where we have used the fact that $\lfloor w_a^t \rfloor \geq 1$. ∎

E. Upper Bound for R_1

Lemma 10: Under the good event E:

$$
\begin{aligned}
R_1 &:= \sum_{t=1}^{T} \sum_{a \in A} (\theta_a^\star - \theta_a^{t,-})(w_a^t)^2 \qquad (25) \\
&\lesssim g_o^2 |A| \sqrt{T} \ln(T g_o).
\end{aligned}
$$

Proof: Take $\tilde{a} \in \arg\max_{a \in A} \left\{ \sum_{t=1}^{T} (\theta_a^\star - \theta_a^{t,-})(w_a^t)^2 \right\}$. Then, under the good event E:

$$
\begin{aligned}
R_1 &\leq |A| \cdot \sum_{t=1}^{T} (\theta_{\tilde{a}}^\star - \theta_{\tilde{a}}^{t,-})(w_a^t)^2 \\
&\leq |A| \sqrt{g_o} \cdot \sum_{t=1}^{T} (\theta_{\tilde{a}}^\star - \theta_{\tilde{a}}^{t,-})(w_{\tilde{a}}^t)^{3/2} \\
&\leq |A| \sqrt{g_o} \cdot \sum_{t=1}^{T} \min\left\{ C_\theta g_o^{3/2}, \frac{2\gamma_{\tilde{a}}^t}{\sqrt{V_{\tilde{a}}^{t-1}}} (w_{\tilde{a}}^t)^{3/2} \right\} \\
&\leq 2\sqrt{2} |A| \sqrt{g_o} \\
&\qquad \cdot \sum_{t=1}^{T} \min\left\{ C_\theta g_o^{3/2}, \frac{\gamma_{\tilde{a}}^t}{\sqrt{V_{\tilde{a}}^{t-1}}} \cdot \sqrt{\lfloor w_{\tilde{a}}^t \rfloor} \cdot w_{\tilde{a}}^t \right\}.
\end{aligned}
$$

where in the final inequality, we have used the fact that, since $w_a^t \geq 1$ by assumption, we have $w_a^t \leq 2\lfloor w_a^t \rfloor$. Thus, the Cauchy-Schwarz inequality gives:

$$
\begin{aligned}
R_1 &\leq 2\sqrt{2} C_\theta |A| g_o^2 \gamma_{\tilde{a}}^T \\
&\qquad \cdot \sum_{t=1}^{T} \min\left\{ 1, \frac{1}{\sqrt{V_{\tilde{a}}^{t-1}}} \cdot \sqrt{\lfloor w_{\tilde{a}}^t \rfloor} \cdot w_{\tilde{a}}^t \right\} \\
&\lesssim |A| g_o^2 \sqrt{\ln(T g_o)} \cdot \sqrt{T} \cdot \sqrt{\sum_{t=1}^{T} \min\left\{ 1, \frac{\lfloor w_{\tilde{a}}^t \rfloor (w_{\tilde{a}}^t)^2}{V_{\tilde{a}}^{t-1}} \right\}} \\
&\lesssim g_o^2 |A| \sqrt{T} \ln(T g_o),
\end{aligned}
$$

where the final inequality follows from (7). ∎

F. Upper Bound for R_2

Recall that in (11), we defined the entropy term $\chi : \mathcal{W} \to \mathbb{R}$ as follows:

$$
\begin{aligned}
&\chi(w) \\
&:= \sum_{i \in I \setminus \{d\}} \left[\sum_{a \in A_i^+} w_a \ln w_a - \left(\sum_{a \in A_i^+} w_a \right) \ln \left(\sum_{a \in A_i^+} w_a \right) \right]
\end{aligned}
$$

Lemma 11: For any $w \in \mathcal{W}$, we have:

$$
|\chi(w)| \leq g_o \cdot (|I| - 1) \ln\left(\frac{|A|}{|I| - 1} \right)
$$

Proof: First, fix $D > 0$ arbitrarily, and consider the following constrained optimization problem on $\mathbb{R}^d$:

$$
\begin{aligned}
\min_{x \in \mathbb{R}^d} \quad & \sum_{i=1}^{d} x_i \ln x_i - \left(\sum_{i=1}^{d} x_i \right) \ln \left(\sum_{i=1}^{d} x_i \right) \\
\text{s.t.} \quad & \sum_{i=1}^{d} x_i = D.
\end{aligned}
$$

The Lagrangian of the above problem is given by:

$$
\begin{aligned}
\mathcal{L}(x, \lambda, \mu) = &\sum_{i=1}^{d} x_i \ln x_i - \left(\sum_{i=1}^{d} x_i \right) \ln \left(\sum_{i=1}^{d} x_i \right) \\
&+ \lambda \left(\sum_{i=1}^{d} x_i - D \right) + \sum_{i=1}^{d} \mu_i x_i.
\end{aligned}
$$

The corresponding KKT conditions are therefore:

$$
\begin{aligned}
0 = \frac{\partial \mathcal{L}}{\partial x_i} &= \ln x_i + 1 - \ln \left(\sum_{j=1}^{d} x_j \right) - 1 + \lambda + \mu_i \\
&= \ln \left(\frac{x_i}{\sum_{j=1}^{d} x_j} \right) + \lambda + \mu_i, \qquad \forall\, i \in [d], \\
0 &= \mu_i x_i, \qquad \forall\, i \in [d],
\end{aligned}
$$

and $\sum_{i=1}^{d} x_i = D$. The optimal solution is thus $x^\star = \frac{D}{d}(1, \cdots, 1)$, with corresponding minimum value:

$$
\begin{aligned}
&\sum_{i=1}^{d} x_i^\star \ln x_i^\star - \left(\sum_{i=1}^{d} x_i^\star \right) \ln \left(\sum_{i=1}^{d} x_i^\star \right) \\
&= d \cdot \frac{D}{d} \ln\left(\frac{D}{d} \right) - D \ln D \\
&= -D \ln d.
\end{aligned}
$$

This implies that:

$$
\begin{aligned}
&\left| \sum_{a \in A_i^+} w_a \ln w_a - \left(\sum_{a \in A_i^+} w_a \right) \ln \left(\sum_{a \in A_i^+} w_a \right) \right| \\
&\leq \sum_{a \in A_i^+} w_i \cdot \ln |A_i^+|.
\end{aligned}
$$

Summing over all non-destination nodes, we obtain:

$$
\begin{aligned}
&\left| \sum_{i \in I \setminus \{d\}} \left[\sum_{a \in A_i^+} w_a \ln w_a - \left(\sum_{a \in A_i^+} w_a \right) \ln \left(\sum_{a \in A_i^+} w_a \right) \right] \right| \\
&\leq \sum_{i \in I \setminus \{d\}} \left(\sum_{a \in A_i^+} w_a \right) \ln |A_i^+|
\end{aligned}
$$

$$\leq g_o \cdot \sum_{i\in I\setminus\{d\}} \ln|A_i^+|$$
$$\leq g_o \cdot |I\setminus\{d\}| \ln\left(\prod_{i\in I\setminus\{d\}} \ln|A_i^+|^{1/|I\setminus\{d\}|}\right)$$
$$\leq g_o \cdot |I\setminus\{d\}| \ln\left(\frac{1}{|I\setminus\{d\}|}\sum_{i\in I\setminus\{d\}} |A_i^+|\right)$$
$$= g_o \cdot (|I|-1)\ln\left(\frac{|A|}{|I|-1}\right),$$

where the final inequality follows from the arithmetic-geometric inequality. ■

Lemma 12: Under the good event E:

$$R_2 := \sum_{t=1}^{T}\left(\frac{1}{\beta^t}-\frac{1}{\beta^\star}\right)\cdot\chi(w^t) \tag{26}$$
$$\lesssim g_o^2 \cdot B(|I|-1)\ln\left(\frac{|A|}{|I|-1}\right)\cdot\sqrt{T}\ln(Tg_o).$$

Proof: From Lemma 9:

$$\sum_{t=1}^{T}|\beta^t-\beta^\star| \lesssim g_o B\sqrt{T}\ln(Tg_o).$$

This bound, together with the upper bound on χ provided by Lemma 11, completes the proof. ■

G. Upper Bound for R_3

Lemma 13: Under the good event E:

$$R_3 := \sum_{t=1}^{T}\big|L\big(\bar{w}^{\theta^\star,\beta^\star}(p^t),\theta^{t,-},\beta^t\big) \tag{27}$$
$$- L\big(\bar{w}^{\theta^{t,-},\beta^t}(p^t),\theta^{t,-},\beta^t\big)\big| \tag{28}$$
$$\lesssim g_o^2\ln^2(g_o)|A|\sqrt{T}\ln(Tg_o)\cdot\max\left\{|I|\ln\left(\frac{|A|}{|I|}\right),B\right\}.$$

Proof: Define the map $\tilde{w}:\mathbb{R}^{|A|}\times\mathbb{R}\times\mathbb{R}^{|A|}\to\mathbb{R}^{|A|}$ by $\tilde{w}(\theta,\beta,p) := \bar{w}^{\theta,\beta}(p)$. Observe that $L(\cdot,\theta,\beta)$ is continuously differentiable on $\mathcal{W}$, for any fixed $\theta\in\mathbb{R}^{|A|}$, $\beta>0$; later, we will establish that $\tilde{w}$ is continuously differentiable as well. Then, from the Fundamental Theorem of Calculus to the maps L and $\tilde{w}$, we obtain:

$$L\big(\bar{w}^{\theta^\star,\beta^\star}(p^t),\theta^{t,-},\beta^t\big) - L\big(\bar{w}^{\theta^{t,-},\beta^t}(p^t),\theta^{t,-},\beta^t\big)$$
$$= \big[L\big(\bar{w}^{\theta^\star,\beta^\star}(p^t),\theta^{t,-},\beta^t\big) - L\big(\bar{w}^{\theta^{t,-},\beta^\star}(p^t),\theta^{t,-},\beta^t\big)\big]$$
$$+ \big[L\big(\bar{w}^{\theta^{t,-},\beta^\star}(p^t),\theta^{t,-},\beta^t\big)$$
$$- L\big(\bar{w}^{\theta^{t,-},\beta^t}(p^t),\theta^{t,-},\beta^t\big)\big] \tag{29}$$
$$= \int_0^1 \frac{\partial L}{\partial w}\Big(\bar{w}^{\theta^{t,-}+u(\theta^\star-\theta^{t,-}),\beta^t}(p^t),\theta^{t,-},\beta^t\Big) \tag{30}$$
$$\cdot \frac{\partial\tilde{w}}{\partial\theta}\big(\theta^{t,-}+u(\theta^\star-\theta^{t,-}),\beta^t,p^t\big)\,du$$
$$\cdot(\theta^\star-\theta^{t,-}) \tag{31}$$
$$+ \int_0^1 \frac{\partial L}{\partial w}\Big(\bar{w}^{\theta^{t,-},\beta^t+u(\beta^\star-\beta^t)}(p^t),\theta^{t,-},\beta^t\Big)$$
$$\cdot \frac{\partial\tilde{w}}{\partial\theta}\big(\theta^{t,-},\beta^t+u(\beta^\star-\beta^t),p^t\big)\,du$$
$$\cdot(\beta^\star-\beta^{t,-}). \tag{32}$$

For convenience, define:

$$S_{w,\theta} := \Big\{\bar{w}^{\theta^{t,-}+u(\theta^\star-\theta^{t,-}),\beta^t}(p^t) : u\in[0,1]\Big\},$$
$$S_{w,\beta} := \Big\{\bar{w}^{\theta^{t,-},\beta^t+u(\beta^\star-\beta^t)}(p^t) : u\in[0,1]\Big\},$$
$$S := S_{w,\theta}\cup S_{w,\beta},$$
$$S_\theta := \Big\{\theta^{t,-}+u(\theta^\star-\theta^{t,-}) : u\in[0,1]\Big\}$$
$$S_\beta := \Big\{\beta^t+u(\beta^\star-\beta^t) : u\in[0,1]\Big\}.$$

Then, by applying the Cauchy-Schwarz inequality to (5), we obtain:

$$L\big(\tilde{w}^{\theta^\star,\beta^\star}(p^t),\theta^{t,-},\beta^t\big) - L\big(\tilde{w}^{\theta^{t,-},\beta^t}(p^t),\theta^{t,-},\beta^t\big)$$
$$\leq \max_{w\in S_w}\left\|\frac{\partial L}{\partial w}(w,\theta^{t,-},\beta^t)\right\|_2 \tag{33}$$
$$\cdot\left[\max_{\theta\in S_\theta}\left\|\frac{\partial\tilde{w}}{\partial\theta}(\theta,\beta^\star,p^t)\cdot(\theta^\star-\theta^{t,-})\right\|_2\right.$$
$$\left.+\max_{\beta\in S_\beta}\left\|\frac{\partial\tilde{w}}{\partial\beta}(\theta^\star,\beta,p^t)\right\|_2\cdot|\beta^\star-\beta^t|\right]$$

We bound each of the max terms in (33) below.

1) Bounding $\max_{w\in S_w}\left\|\frac{\partial L}{\partial w}(w,\theta^{t,-},\beta^t)\right\|_2$:

For each $a\in A$, and any $w\in\mathcal{W}$, $\theta\in\mathbb{R}^{|A|}$, and $\beta>0$:

$$\frac{\partial L}{\partial w_a}(w,\theta,\beta) = 2\theta_a w_a + \frac{1}{\beta}\ln\left(\frac{w_a}{\sum_{a'\in A_{i_a}^+} w_{a'}}\right).$$

Note that $|\theta_a^{t,-}|\leq C_\theta$ for each $a\in A$, and that for any $w\in\mathcal{W}$, we have $\|w\|_2\leq\sum_{a\in A}w_a\leq m(G)g_o$. Moreover, by Lemma 11, and the assumption that $w_a\geq 1$ for each $a\in A$ (note that the set $\{w\in\mathbb{R}^{|A|} : w_a\geq 1,\forall\, a\in A\}$ is convex), we have for each $w\in\mathcal{W}$:

$$\sum_{a\in A}\left|\ln\left(\frac{w_a}{\sum_{a'\in A_{i_a}^+}w_{a'}}\right)\right|$$
$$= -\sum_{a\in A}\ln\left(\frac{w_a}{\sum_{a'\in A_{i_a}^+}w_{a'}}\right)$$
$$\leq -\sum_{a\in A} w_a\ln\left(\frac{w_a}{\sum_{a'\in A_{i_a}^+}w_{a'}}\right)$$
$$= |\chi(w)|$$
$$\leq g_o\cdot(|I|-1)\ln\left(\frac{|A|}{|I|-1}\right) \tag{34}$$

Meanwhile:

$$\sum_{a\in A}\left|\ln\left(\frac{w_a}{\sum_{a'\in A_{i_a}^+}w_{a'}}\right)\right|^2 \leq \ln^2(g_o)|A| \tag{35}$$

Thus, we obtain that, for any $w\in S_w$:

$$\left\|\frac{\partial L}{\partial w_a}(w,\theta^{t,-},\beta^t)\right\|_2$$
$$\leq 2C_\theta m(G)g_o \tag{36}$$

$$+\frac{1}{c_\beta}\min\left\{\ln(g_o)\sqrt{|A|}, g_o(|I|-1)\ln\left(\frac{|A|}{|I|-1}\right)\right\}. \tag{37}$$

2) Bounding $\max_{\theta\in S_\theta}\left\|\frac{\partial\tilde{w}}{\partial\theta}(\theta,\beta^\star,p^t)\cdot(\theta^\star-\theta^{t,-})\right\|_2$:

First, we verify that $\tilde{w}$ is indeed continuously differentiable, and compute the Jacobians $\frac{\partial L}{\partial w}$, $\frac{\partial\tilde{w}}{\partial\theta}$, and $\frac{\partial\tilde{w}}{\partial\beta}$. This requires the results of [8], Lemma 1, which we summarize below. Define $F:\mathcal{W}\times\mathbb{R}^{|A|}\times\mathbb{R}^{|A|}\times\mathbb{R}\times\mathbb{R}^{|A|}\to\mathbb{R}$ as follows—For each:

$$\begin{aligned}&F(w,\theta,\beta,p)\\&=\sum_{[a]\in A_O}\int_0^{w_a}\left[\theta_a z+p_a\right]dz\\&+\frac{1}{\beta}\sum_{i\neq d}\left[\sum_{a\in A_i^+}w_a\ln w_a-\left(\sum_{a\in A_i^+}w_a\right)\ln\left(\sum_{a\in A_i^+}w_a\right)\right]\end{aligned}$$

Note that $F(\cdot,\theta,\beta,p)$ is strongly convex, with parameter at least c_θ.

Next, observe that $\mathcal{W}$ is a compact subset of a strict affine subspace in $\mathbb{R}^{|A|}$. Let d be the dimension of the smallest affine subspace containing $\mathcal{W}$. Then, there exist $M\in\mathbb{R}^{|A|\times|I\setminus\{d\}|}$ with orthonormal columns, and $b\in\mathbb{R}^{|I\setminus\{d\}|}$ such that:

$$\mathcal{W}=\{w\in\mathbb{R}^{|A|}: M^\top w+b=0, w_a\geq 0, \forall\, a\in A\}.$$

Let $B\in\mathbb{R}^{|A|\times(|A|-|I\setminus\{d\}|)}$ consist of orthonormal columns orthogonal to the columns of M. We then use the theory of constrained optimization to completely characterize $\tilde{w}(\theta,\beta,p)=\bar{w}^{\theta,\beta}(p)$. In particular, $w=\tilde{w}(\theta,\beta,p)$ if and only if the following implicit equation, characterized by the map $J:\mathbb{R}^{|A|}\times\mathbb{R}^{|A|}\to\mathbb{R}^{|A|}$ defined below, is satisfied:

$$J(w,\theta,\beta,p):=\begin{bmatrix}M^\top w+b\\B^\top\nabla_w F(w,\theta,\beta,p)\end{bmatrix}=0.$$

Moreover, the proof of [8], Lemma 1 establishes that, for any fixed $\theta\in\mathbb{R}^{|A|}$, $\beta>0$, $p\in\mathbb{R}^{|A|}$:

$$\frac{\partial J}{\partial w}(\theta,\beta,p)=\begin{bmatrix}M^\top\\B^\top\nabla_w^2 F(w,\theta,\beta,p)\end{bmatrix}\in\mathbb{R}^{|A|\times|A|}$$

is non-singular. By the Implicit Function Theorem, this establishes the continuous differentiability of $\tilde{w}$. We can then compute $\frac{\partial\tilde{w}}{\partial\theta}\in\mathbb{R}^{|A|\times|A|}$ at any $(\theta,\beta,p)\in\mathbb{R}^{|A|}\times\mathbb{R}\times\mathbb{R}^{|A|}$ as:

$$\begin{aligned}&\frac{\partial\tilde{w}}{\partial\theta}(\theta,\beta,p)\\&=\left[\frac{\partial J}{\partial w}(\theta,\beta,p)\right]^{-1}\frac{\partial J}{\partial\theta}(\theta,\beta,p)\\&=\begin{bmatrix}M^\top\\B^\top\nabla_w^2F(w,\theta,\beta,p)\end{bmatrix}^{-1}\begin{bmatrix}0\\B^\top\frac{\partial}{\partial\theta}\nabla_w F(w,\theta,\beta,p)\end{bmatrix}\\&=B(B^\top\nabla_w^2F(w,\theta,\beta,p)B)^{-1}B^\top\\&\qquad\cdot\frac{\partial}{\partial\theta}\nabla_w F(w,\theta,\beta,p),\end{aligned}$$

where we have used the fact that by construction, $\begin{bmatrix}M & B\end{bmatrix}$ is an orthogonal matrix (see [8], Appendix A).

Now, observe that the (a,a')-entry of $\frac{\partial}{\partial\beta}\nabla_w F(w,\theta,\beta,p)\in\mathbb{R}^{|A|\times|A|}$ is given by:

$$\frac{\partial^2}{\partial\theta_{a'}\partial w_a}F(w,\theta,\beta,p)=2w_a\cdot\mathbf{1}\{a'=a\},\qquad\forall\, a\in A,$$

Substituting back into (33) and applying the Cauchy-Schwarz inequality, we obtain that, for each $\theta\in S_\theta$:

$$\begin{aligned}&\frac{\partial\tilde{w}}{\partial\theta}(\theta,\beta^\star,p^t)\cdot(\theta^\star-\theta^{t,-})\\&=B(B^\top\nabla_w^2F(w,\theta,\beta,p)B)^{-1}B^\top\\&\qquad\cdot\left((\theta^\star-\theta^{t,-})w_a^t\right)_{a\in A}.\end{aligned}$$

Applying the Cauchy-Schwarz inequality, we obtain:

$$\begin{aligned}&\max_{\theta\in S_\theta}\left\|\frac{\partial\tilde{w}}{\partial\theta}(\theta,\beta^\star,p^t)\cdot(\theta^\star-\theta^{t,-})\right\|_2\\&\leq\|B(B^\top\nabla_w^2F(w,\theta,\beta,p)B)^{-1}B^\top\|_2\\&\qquad\cdot\|\left((\theta^\star-\theta^{t,-})w_a^t\right)_{a\in A}\|_2.\end{aligned}$$

Since the columns of B are orthonormal, we have $\|B(B^\top\nabla_w^2F(w,\theta,\beta,p)B)^{-1}B^\top\|_2\leq\|\nabla_w^2F(w,\theta,\beta,p)\|_2\leq 1/c_\theta$. Moreover, we can upper bound $\|\left((\theta^\star-\theta^{t,-})w_a^t\right)_{a\in A}\|_2\leq\|\left((\theta^\star-\theta^{t,-})w_a^t\right)_{a\in A}\|_1=\sum_{a\in A}(\theta^\star-\theta^{t,-})w_a^t$. We thus obtain:

$$\begin{aligned}&\max_{\theta\in S_\theta}\left\|\frac{\partial\tilde{w}}{\partial\theta}(\theta,\beta^\star,p^t)\cdot(\theta^\star-\theta^{t,-})\right\|_2\\&\leq\frac{1}{c_\theta}\cdot\sum_{a\in A}(\theta^\star-\theta^{t,-})w_a^t.\end{aligned}\tag{38}$$

3) Bounding $\max_{\beta\in S_\beta}\left\|\frac{\partial\tilde{w}}{\partial\beta}(\theta^\star,\beta,p^t)\right\|_2\cdot|\beta^\star-\beta^t|$:

In the same manner that we used to compute $\frac{\partial\tilde{w}}{\partial\theta}$ above, we can compute $\frac{\partial\tilde{w}}{\partial\beta}\in\mathbb{R}^{|A|}$ at any $(\theta,\beta,p)\in\mathbb{R}^{|A|}\times\mathbb{R}\times\mathbb{R}^{|A|}$ as:

$$\begin{aligned}&\frac{\partial\tilde{w}}{\partial\beta}(\theta,\beta,p)\\&=\left[\frac{\partial J}{\partial w}(\theta,\beta,p)\right]^{-1}\frac{\partial J}{\partial\theta}(\theta,\beta,p),\\&=\begin{bmatrix}M^\top\\B^\top\nabla_w^2F(w,\theta,\beta,p)\end{bmatrix}^{-1}\begin{bmatrix}0\\B^\top\frac{\partial}{\partial\beta}\nabla_w F(w,\theta,\beta,p)\end{bmatrix}\\&=B(B^\top\nabla_w^2F(w,\theta,\beta,p)B)^{-1}B^\top\\&\qquad\cdot\frac{\partial}{\partial\beta}\nabla_w F(w,\theta,\beta,p),\end{aligned}$$

Now, observe that the a-th entry of $\frac{\partial}{\partial\beta}\nabla_w F(w,\theta,\beta,p)\in\mathbb{R}^{|A|}$ is:

$$\frac{\partial^2}{\partial\beta\partial w_a}F(w,\theta,\beta,p)=-\frac{1}{\beta^2}\ln\left(\frac{w_a}{\sum_{a'\in A_{i_a}^+}w_{a'}}\right).$$

Using (34) and (35), we obtain:

$$\begin{aligned}&\left\|\frac{\partial}{\partial\beta}\nabla_w F(w,\theta,\beta,p)\right\|_2\\ &\leq \frac{1}{\beta^2}\cdot\min\left\{\ln(g_o)\sqrt{|A|}, g_o(|I|-1)\ln\left(\frac{|A|}{|I|-1}\right)\right\}\end{aligned}$$

Finally, we conclude that:

$$\begin{aligned}&\max_{\beta\in S_\beta}\left\|\frac{\partial\tilde{w}}{\partial\beta}(\theta^\star,\beta,p^t)\right\|_2\cdot|\beta^\star-\beta^t|\\ &\leq\|B(B^\top\nabla_w^2 F(w,\theta,\beta,p)B)^{-1}B^\top\|_2\\ &\quad\cdot\left\|\frac{\partial}{\partial\beta}\nabla_w F(w,\theta,\beta,p)\right\|_2\cdot|\beta^\star-\beta^t|\\ &\leq\frac{1}{c_\theta\beta^2}\cdot\min\left\{\ln(g_o)\sqrt{|A|}, g_o(|I|-1)\ln\left(\frac{|A|}{|I|-1}\right)\right\}\\ &\quad\cdot|\beta^\star-\beta^t|.\end{aligned}\tag{39}$$

Substituting (36), (38), (39) back into (33), we obtain that:

$$\begin{aligned}R_3&=\sum_{t=1}^{T}\big|L\big(\bar{w}^{\theta^\star,\beta^\star}(p^t),\theta^{t,-},\beta^t\big)\\ &\qquad-L\big(\bar{w}^{\theta^{t,-},\beta^t}(p^t),\theta^{t,-},\beta^t\big)\big|\\ &\lesssim\left(m(G)g_o+\frac{1}{c_\beta}\cdot g_o\cdot(|I|-1)\ln\left(\frac{|A|}{|I|-1}\right)\right)\\ &\quad\cdot\left[\sum_{t=1}^{T}\sum_{a\in A}(\theta^\star-\theta^{t,-})w_a^t\right.\\ &\qquad+\min\left\{\ln^2(g_o)\cdot|A|, g_o|I|\ln\left(\frac{|A|}{|I|}\right)\right\}\\ &\qquad\left.\cdot\sum_{t=1}^{T}|\beta^\star-\beta^t|\right]\end{aligned}\tag{40}$$

Applying Lemmas 8 (with $p=1$) and 9, we obtain:

$$\begin{aligned}R_3&\lesssim\left(m(G)g_o+\min\left\{\ln(g_o)\sqrt{|A|}, g_o|I|\ln\left(\frac{|A|}{|I|}\right)\right\}\right)\\ &\quad\cdot\left[g_o|A|\sqrt{T}\ln(Tg_o)\right.\\ &\qquad+\min\left\{\ln(g_o)\sqrt{|A|}, g_o|I|\ln\left(\frac{|A|}{|I|}\right)\right\}\\ &\qquad\left.\cdot g_oB\sqrt{T}\ln(Tg_o)\right]\\ &\lesssim g_o^2m(G)|A|\sqrt{T}\ln(Tg_o)\\ &\quad+g_o^2\ln(g_o)m(G)\sqrt{|A|}B\sqrt{T}\ln(Tg_o)\\ &\quad+g_o^2|A||I|\ln\left(\frac{|A|}{|I|}\right)\sqrt{T}\ln(Tg_o)\\ &\quad+g_o\ln^2(g_o)|A|B\sqrt{T}\ln(Tg_o)\\ &\lesssim g_o^2\ln^2(g_o)|A|\sqrt{T}\ln(Tg_o)\\ &\quad\cdot\max\left\{|I|\ln\left(\frac{|A|}{|I|}\right),B\right\}.\end{aligned}$$

Note that we have used the fact that $m(G)\leq|I|$. ■

H. Upper Bound for R

Below, we combine the results of Lemmas 10, 12, and 13 in the above sections to conclude our proof of Theorem 1.

Proof: [**Proof of Theorem 1**] From Lemmas 10, 12, and 13, we have:

$$\begin{aligned}R_1&\lesssim g_o^2|A|\sqrt{T}\ln(Tg_o),\\ R_2&\lesssim g_o^2\cdot B|I|\ln\left(\frac{|A|}{|I|}\right)\cdot\sqrt{T}\ln(Tg_o),\\ R_3&\lesssim g_o^2\ln^2(g_o)|A|\sqrt{T}\ln(Tg_o)\cdot\max\left\{|I|\ln\left(\frac{|A|}{|I|}\right),B\right\}.\end{aligned}$$

Note that $R_1\lesssim R_3$ and $R_2\lesssim R_3$. We thus conclude that:

$$\begin{aligned}R&=R_1+R_2+R_3\\ &\lesssim g_o^2\ln^2(g_o)|A|\sqrt{T}\ln(Tg_o)\cdot\max\left\{|I|\ln\left(\frac{|A|}{|I|}\right),B\right\}.\end{aligned}$$

■