CE 272
Traffic Network Equilibrium

Lecture 19

Overview of Other TAP Variants

Previously on Traffic Network Equilibrium...

- All travelers select paths at the same instant and we do not explicitly model the departure times and times at which travelers enter a link. Hence, this process of modeling route choice is also called **static traffic assignment**.
- In other words, we only analyze the spatial distribution of congestion and ignore the temporal dimension. There are advanced models which capture both aspects (dynamic traffic assignment).
- Travelers are fully rational and selfish and seek routes that minimize their travel times.
- **Travelers are fully aware of the network topology and its response to congestion (perhaps from experience).**

This simplified setting is representative of a "steady state" and allows us to study equilibrium analytically.

Previously on Traffic Network Equilibrium...

- The BPR functions are defined for flow values that exceed capacity. We'll ignore capacity constraints in this course but there are formulations and solution techniques which explicitly model capacity constraints.
- Alternately, we can define delay functions that exhibit a steep increase at flows close to the roadway capacity.
- Practitioners often use V/C ratios, i.e., x_{ij}/C_{ij}, to identify the links that are heavily congested.

Lecture Outline

1. Stochastic User Equilibrium
2. Day-to-Day Traffic Assignment
3. Capacitated Traffic Assignment
4. Equilibrium with Recourse

Lecture Outline

Stochastic User Equilibrium

Stochastic User Equilibrium
Introduction

So far, we assumed that travelers select routes with least travel time (or generalized cost).

In practice, travelers may

- Have other criteria for route selection. (Such as ?)
- Not know travel time conditions in the network (perception errors).

Let us, as usual, consider the path selection process with fixed travel times and then extended it to the equilibrium problem with flow-dependent link times.

Stochastic User Equilibrium

Choice Modeling

In such settings, we may assume that the utility of traveling on a path p is expressed as

$$u_p = -\theta\tau_p + \epsilon_p$$

where θ is a scale parameter (else the utilities depend on the units of time), that can be related to the level of perception errors.

Travelers choose from a finite number of paths (discrete choices) and they select a path p only if it's utility is greater than or equal to the utilities of all other paths.

$$u_p \geq u_q \,\forall q \in P_{rs}$$
$$\Rightarrow -\theta\tau_p + \epsilon_p \geq -\theta\tau_q + \epsilon_q \,\forall\, q \in P_{rs}$$

Stochastic User Equilibrium

Choice Modeling

Thus, the probability with which p is be chosen is given by

$$\mathbb{P}[-\theta\tau_p + \epsilon_p \geq -\theta\tau_q + \epsilon_q \, \forall \, q \in P_{rs}]$$

If the ϵ_p variables are identically distributed extreme value type I (Gumbel) distributed, then the above probability has the following closed form expression

$$\frac{\exp(-\theta\tau_p)}{\sum_{q \in P_{rs}} \exp(-\theta\tau_q)}$$

This framework is also called the logit model. The number of users on path p is therefore given by

$$d_{rs} \frac{\exp(-\theta\tau_p)}{\sum_{q \in P_{rs}} \exp(-\theta\tau_q)}$$

Stochastic User Equilibrium

Example

Suppose there are two parallel routes between O and D with travel times 10 and 20. If the demand between O and D is 100 and if $\theta = 0.1$. What are the flows on each route? What if $\theta = 1$?

- Hence, as $\theta \to \infty$, we get the deterministic user equilibrium model.
- For different assumptions on the distributions of the error terms, we get different types of SUE models.
- For logit-based SUE, Dial in 1971 suggested a network loading method that avoids path enumeration.

Stochastic User Equilibrium

Equilibrium Definitions

In practice, path travel times are a function of path flows, which in-turn determine choice probabilities and new path flows.

We assume that travelers in a stochastic user equilibrium setting select paths that minimize their perceived travel times.

Suppose $\boldsymbol{\pi}(\boldsymbol{\tau})$ represents a function that provides the path flows for a given vector of path travel times (e.g., the logit formula we saw earlier).

Then, $\mathbf{y}$ is a stochastic user equilibrium if and only if $\mathbf{y} = \boldsymbol{\pi}(\boldsymbol{\tau}(\mathbf{y}))$.

Stochastic User Equilibrium

Fisk's Optimization Formulation

Fisk in 1980 provided the following optimization model for solving the logit-based SUE.

$$\min \frac{1}{\theta} \sum_{p \in P} y_p \ln y_p + \sum_{(i,j) \in A} \int_0^{\sum_{p \in P} \delta_{ij}^p y_p} t_{ij}(\omega) d\omega$$

$$\text{s.t.} \quad \sum_{p \in P_{rs}} y_p = d_{rs} \ \forall \ (r,s) \in Z^2$$

$$y_p \geq 0 \ \forall \ p \in P$$

The KKT conditions of this program be shown to satisfy the logit probabilities.

Supplementary Reading

Dial, R. B. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. Transportation research, 5(2), 83-111.

Daganzo, C. F., & Sheffi, Y. (1977). On stochastic models of traffic assignment. Transportation science, 11(3), 253-274.

Lecture Outline

Day-to-Day Traffic Assignment

Day-to-Day Traffic Assignment

Discrete Time Version

In the discrete time version, travelers are typically assumed to be finite.

Each network user observes the network performance on the last m days and users that to construct a probability distribution across all the paths.

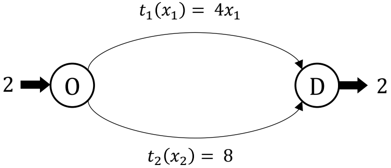
The evolution of traffic is represented a discrete time Markov chain. The state space consists of the number of travelers on different paths.

The transition probabilities for moving from one state to another is constructed using the individual choice probabilities (such as the logit choice equation).

Day-to-Day Traffic Assignment

Discrete Time Version

Consider the following network with three states $(2,0)$, $(0,2)$, and $(1,1)$, which we will refer to as 1, 2 and 3 respectively. State 1 is a NE and state 3 is SO.



Suppose both travelers use the logit choice model in which the probability of choosing the top paths is

$$\frac{\exp(-t_1(x_1))}{\exp(-t_1(x_1)) + \exp(-t_2(x_2))}$$

Where $t_1(x_1)$ and $t_2(x_2)$ represents the travel times as a function of the previous day's flow.

Day-to-Day Traffic Assignment

Discrete Time Version

The stochastic process is Markovian and the steady state probabilities of observing states 1, 2, and 3 are 0.5654, 0.1414, and 0.2932 respectively.

The expected TSTT is $16(0.5654) + 16(0.1414) + 12(0.2932) = 14.8272$.

In these models, the system constantly transitions from one state to another and the equilibrium is defined with respect to the long run probabilities of observing it in different states.

There are some interesting connections between these type of models, SUE, and the learning approaches we saw in Lecture 10.

Day-to-Day Traffic Assignment

Continuous Time Version

In the continuous time models, travelers are at a non-equilibrium state, aware of travel times on different routes, and switch from one route to another at a rate proportional to the difference in their travel times.

Using Lyapunov functions, it can be shown that such dynamical systems converge to the set of equilibira (under some conditions on the travel time functions).

Supplementary Reading

Cascetta, E. (1989). A stochastic process approach to the analysis of temporal dynamics in transportation networks. Transportation Research Part B: Methodological, 23(1), 1-17.

Smith, M. J. (1984). The stability of a dynamic model of traffic assignment - an application of a method of Lyapunov. Transportation Science, 18(3), 245-252.

Capacitated Traffic Assignment

Capacitated Traffic Assignment

Introduction

In all models discussed so far, we let flow exceed capacity. This can be avoided by imposing explicit constraints.

In reality, when flow reaches capacity traffic queues spill back to upstream links.

Capacity constraints are bundle constraints, i.e., we can no longer analyze each OD pair separately by repeatedly solving shortest path problems.

Instead, the subproblem is a multicommodity minimum cost flow problem, which is relatively harder to solve.

Capacitated Traffic Assignment

Formulation

Suppose C_{ij} is the capacity of link (i, j). The capacitated traffic assignment problem (CTAP) can be formulated as

$$\min \sum_{(i,j)\in A} \int_0^{\sum_{p\in P} \delta_{ij}^p y_p} t_{ij}(\omega)\, d\omega$$

$$\text{s.t.} \sum_{p\in P_{rs}} y_p = d_{rs} \; \forall \; (r, s) \in Z^2$$

$$\sum_{p\in P} \delta_{ij}^p y_p \leq C_{ij} \; \forall \; (i, j) \in A$$

$$y_p \geq 0 \; \forall \; p \in P$$

For strictly increasing delay functions, the problem is strictly convex and has unique link flow solutions.

Capacitated Traffic Assignment

Formulation

Does the new constraint alter the definition of Wardrop equilibria?

In terms of the path flows, the Lagrangian can be written as

$$\mathcal{L}(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) = \sum_{(i,j)\in A} \int_0^{\sum_{p\in P} \delta_{ij}^p y_p} t_{ij}(\omega)\, d\omega + \sum_{p\in P} \lambda_p(-y_p)$$
$$+ \sum_{(r,s)\in Z^2} \mu_{rs} \left(d_{rs} - \sum_{p\in P_{rs}} y_p \right) + \sum_{(i,j)\in A} \nu_{ij} \left(\sum_{p\in P} \delta_{ij}^p y_p - C_{ij} \right)$$

Path-based Formulation

KKT Conditions

Primal feasibility:

$$\sum_{p \in P_{rs}} y_p = d_{rs} \ \forall \ (r, s) \in Z^2$$

$$\sum_{p \in P} \delta_{ij}^p y_p \leq C_{ij} \ \forall \ (i, j) \in A$$

$$y_p \geq 0 \ \forall \ p \in P$$

Dual feasibility:

$$\lambda_p \geq 0 \, \forall \, p \in P$$

$$\nu_{ij} \geq 0 \, \forall \, (i, j) \in A$$

Complementary Slackness:

$$\lambda_p y_p = 0 \, \forall \, p \in P$$

$$\nu_{ij} \left(\sum_{p \in P} \delta_{ij}^p y_p - C_{ij} \right) = 0 \, \forall \, (i, j) \in A$$

Gradient of the Lagrangian vanishes:

$$\sum_{(i,j) \in A} \delta_{ij}^p t_{ij}(x_{ij}) + \sum_{(i,j) \in A} \delta_{ij}^p \nu_{ij} - \lambda_p - \mu_{rs} = 0 \, \forall \, (r, s) \in Z^2, p \in P_{rs}$$

Path-based Formulation

KKT Conditions

Replacing $\sum_{p \in P} \delta_{ij}^p y_p$ by x_{ij}, the complementary slackness condition can be written as

$$\nu_{ij}(x_{ij} - C_{ij}) = 0 \, \forall \, (i,j) \in A$$

Hence, if $\nu_{ij} > 0$, then $x_{ij} = C_{ij}$, i.e., the link is at its capacity. One can thus interpret ν_{ij} as the extra delay caused when a link is at capacity (due to queue spillback).

Path-based Formulation

KKT Conditions

Using the last equation and other complementary slackness conditions, we can write

$$\sum_{(i,j)\in A} \delta_{ij}^{p}\big(t_{ij}(x_{ij}) + \nu_{ij}\big) \geq \mu_{rs}$$

$$y_p \left(\sum_{(i,j)\in A} \delta_{ij}^{p}\big(t_{ij}(x_{ij}) + \nu_{ij}\big) - \mu_{rs} \right) = 0$$

Thus, users in the CTAP problem are assumed to minimize "generalized cost of travel", where the generalized cost on a link is defined as $t_{ij}(x_{ij})+\nu_{ij}$ (the delay on the link + delay caused when the link is at capacity).

Capacitated Traffic Assignment

Penalty Methods

Penalty methods ignore the capacity constraints but modify the objective function by adding an extra term which ensure that the flows remain bounded.

Two common penalty methods used are

- Inner Penalty Function (IPF) method
- Augmented Lagrange Multiplier (ALM) method

Capacitated Traffic Assignment

Inner Penalty Functions

The IPF method (also called Barrier method) imposes an asymptotic penalty function at the boundary of the feasible set.

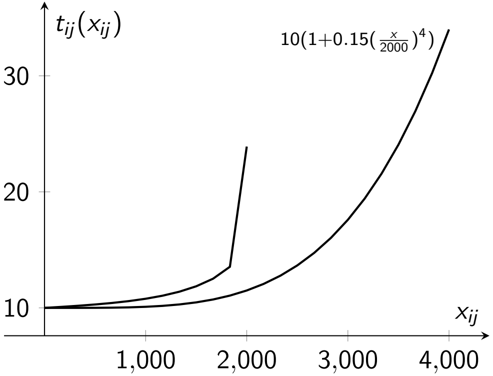
$$
\begin{aligned}
\min \sum_{(i,j)\in A} \int_0^{x_{ij}} t_{ij}(\omega)\, d\omega + \gamma \sum_{(i,j)\in A} \pi_{ij}(x_{ij}) \\
\text{s.t.} \sum_{p\in P_{rs}} y_p = d_{rs}\ \forall\ (r,s) \in Z^2 \\
y_p \geq 0\ \forall\ p \in P
\end{aligned}
$$

where the function $\pi_{ij}(x_{ij})$ is assumed to be positive, continuous, and tends to ∞ as $x_{ij} \to C_{ij}$.

It can be shown that the sequence of solutions of the above formulation converges to the optimum of CTAP as $\gamma \to 0$.

Capacitated Traffic Assignment

Inner Penalty Functions



The ALM method imposes a quadratic penalty to the objective. Consider a non-linear program with equality constraints,

$$\begin{aligned} &\min_{\mathbf{x}} f(\mathbf{x}) \\ \text{s.t. } &h_i(\mathbf{x}) = 0 \ \forall \ i = 1, 2, \ldots, m \end{aligned}$$

The augmented Lagrangian function is defined as

$$\mathbb{L}(\mathbf{x}, \boldsymbol{\mu}, \rho) = f(\mathbf{x}) + \sum_{i=1}^{m} \mu_i h_i(\mathbf{x}) + \frac{1}{2}\rho \sum_{i=1}^{m} h_i(\mathbf{x})^2$$

In iteration k, given a value of $\boldsymbol{\mu}^k$ and ρ^k, we solve an unconstrained problem $\min \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}^k, \rho^k)$ to get $\mathbf{x}^k$. The $\boldsymbol{\mu}$s and ρ are then updated as (κ is a constant)

$$\begin{aligned} \boldsymbol{\mu}^{k+1} &= \boldsymbol{\mu}^k + \rho^k \mathbf{h}(\mathbf{x}^k) \\ \rho^{k+1} &= \kappa \rho^k \end{aligned}$$

Capacitated Traffic Assignment

Augmented Lagrange Multiplier

For the CTAP, the capacity constraints are inequalities. To cast it as equalities, we can add slack variables as follows.

$$\begin{aligned} x_{ij} - C_{ij} + \xi_{ij} = 0 \ \forall\ (i,j) \in A \\ \xi_{ij} \geq 0 \ \forall\ (i,j) \in A \end{aligned}$$

Thus, the augmented Lagrangian takes the form,

$$\begin{aligned} \mathbb{L}(\mathbf{x}, \boldsymbol{\xi}, \boldsymbol{\mu}, \rho) = \sum_{(i,j)\in A} \int_0^{x_{ij}} t_{ij}(\omega)\, d\omega + \sum_{(i,j)\in A} \mu_i \left(x_{ij} - C_{ij} + \xi_{ij}\right) \\ + \frac{1}{2}\rho \sum_{(i,j)\in A} \left(x_{ij} - C_{ij} + \xi_{ij}\right)^2 \end{aligned}$$

Fixing $\boldsymbol{\mu}$ and ρ, one can minimize the above function to find $\mathbf{x}$ and $\boldsymbol{\xi}$ and use them to update the $\boldsymbol{\mu}$ and ρ values.

Supplementary Reading

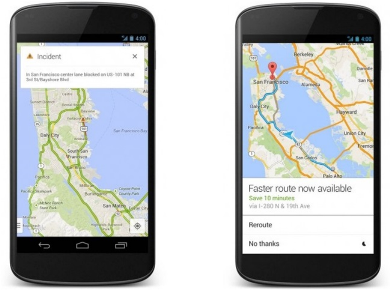
Larsson, T., & Patriksson, M. (1995). An augmented Lagrangean dual algorithm for link capacity side constrained traffic assignment problems. Transportation Research Part B: Methodological, 29(6), 433-455.

Nie, Y., Zhang, H. M., & Lee, D. H. (2004). Models and algorithms for the traffic assignment problem with link capacity constraints. Transportation Research Part B: Methodological, 38(4), 285-312.

Lecture Outline

Equilibrium with Recourse

Imagine you are traveling along route A. Say your app informs you of an incident downstream and suggests an alternate route B.



Suppose the accident reduces capacity for 15 min in the three-hour peak period, it is reasonable to assume that 1/12 of the travelers will observe the disrupted state and 11/12 of the travelers will not. (Or think about signals)

Equilibrium with Recourse

Online Routing

A traveler has to keep in mind that the new route is likely to get congested as travelers shift to it and that the incident will reduce capacity but will clear up after sometime.

Choices of travelers arriving at the fork

1. Always take A
2. Always take B
3. **Take A if there is an incident and B otherwise**
4. **Take B if there is an incident and A otherwise**

Equilibrium with Recourse

Assumptions

When network states change, drivers arriving at a node learn adjacent link-states and choose a downstream node to minimize their expected travel times.

Let each link in the network (i, j) exist in different states $s \in S_{ij}$ with link performance functions $t_{ij}^s(x_{ij}^s)$.

The probability with which a traveler will encounter a link (i, j) in state s is also assumed to be known.

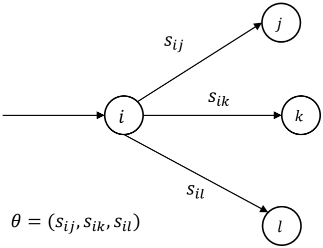
We also make a **full-reset assumption**, i.e., even if a traveler revisits a node, the downstream link-states are re-sampled from the assumed distributions.

Equilibrium with Recourse

Online Shortest Paths

Imagine that the travel times for each link-state are not flow dependent.

Upon arriving at a node i, a traveler observes a information vector $\theta \in \Theta_i = \times_{(i,j) \in A} S_{ij}$ with probability q^θ informing him or her of the state of each link leaving node i.

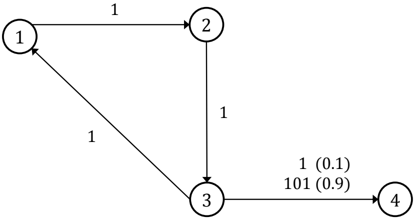


A policy $\pi(i, \theta)$ is a function that maps each node-information vector pair to a downstream node.

Equilibrium with Recourse

Online Shortest Paths

The problem can be formulated as a total cost Markov decision process, and in order to find the optimal policy, we can use techniques such as **value or policy iteration**.



Consider the policy: Take arc (3,4) only if its cost is 1, else return to node 3 via nodes 1 and 2. The expected cost of the policy is an arithmetico geometric sequence!

$$3(0.1) + 6(0.9)(0.1) + 9(0.9)^2(0.1) + \ldots = 3(0.1)\left[1(1) + 2(0.9) + 3(0.9)^2 + \ldots\right]$$
$$= 30$$

Equilibrium with Recourse

Online Shortest Paths

Since the optimal policy may have cycles, the all-or-nothing step is non-trivial.

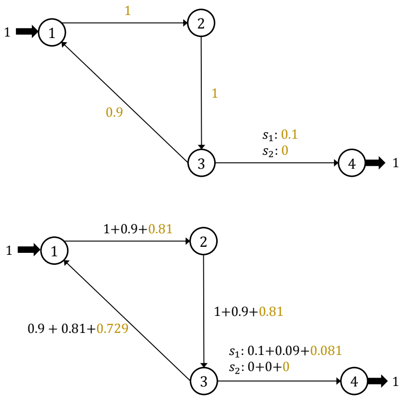
However, cycles are an artifact of the full-reset assumption and ideally we'd prefer loading travelers on acyclic policies.

There are some methods to handle the issue of cycles but for now we will assume that travelers can take cyclic policies.

Equilibrium with Recourse

All-or-Nothing Assignment

Suppose that the demand between nodes 1 and 4 is 1. Assume everyone follows the policy described earlier.



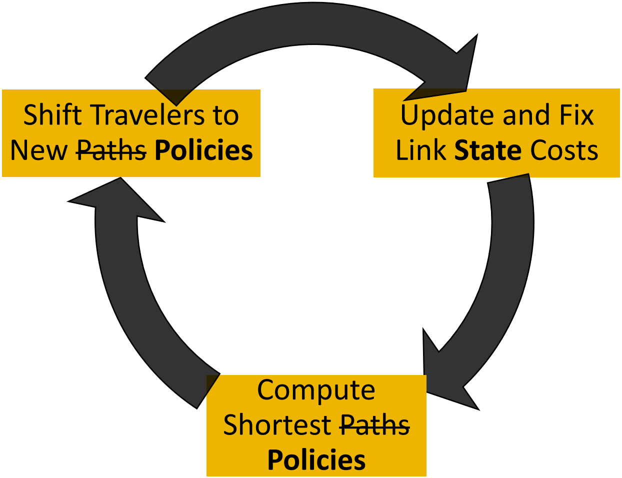
Equilibrium with Recourse

Equilibrium Framework

The Wardrop principle in this setting requires that travelers on all **used policies** must have equal and minimal **expected times**.

Equilibrium with Recourse

Equilibrium Framework



Supplementary Reading

Waller, S. T., & Ziliaskopoulos, A. K. (2002). On the online shortest path problem with limited arc cost dependencies. Networks, 40(4), 216-227.

Rambha, T., Boyles, S. D., Unnikrishnan, A., & Stone, P. (2017). Marginal cost pricing for system optimal traffic assignment with recourse under supply-side uncertainty. Transportation Research Part B: Methodological, 110, 104-121.

Your Moment of Zen