# MPO Datasets, Our Recent Work, and Future Directions

Jing Zhang
jzh@bu.edu, http://people.bu.edu/jzh/
https://github.com/jingzbu

Division of Systems Engineering,
and Center for Information and Systems Engineering
Boston University

May 5, 2017

# Road Map of Eastern Massachusetts

Road map of Eastern Massachusetts (EMA), over which we have access to a vast amount of actual traffic data

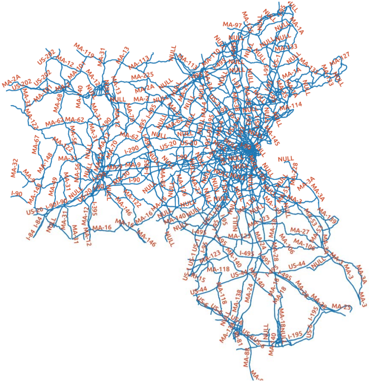


Figure: All available road segments in the road map of EMA

# Datasets

- spatial average speeds for more than 13,000 road segments in EMA; covers every minute of the year 2012; 50+ GB csv files
- flow capacity (# of vehicles per hour) for more than 100,000 road segments in EMA
- Confidential! – raw data available within the CODES/NOC labs only
- See Github repository InverseVIsTraffic for some of the processed data

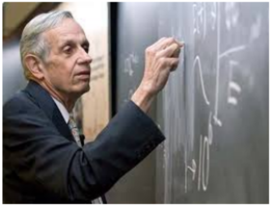
# Data Processing

- Preprocessing
  - Select road segments (QGIS)
  - Extract speeds from raw data (Python)
  - Filter capacity data (Python)
  - Handle missing data (Python)
    - set missing speeds as 0.1 mph and travel times as 0.01 hrs
    - set missing flow capacities as 2000 vehicles/hr
  - Convert speeds to flows (Python)
- OD demand estimation
  - GLS method – QP & QCP (Gurobi; Python)
- Cost function estimation
  - QP (Gurobi; Julia)
- PoA evaluation
  - NLP (IPOPT, JuMP; Julia)
- Data sharing among different programming languages: JSON

# Our Recent Work

- Evaluate/reduce Price of Anarchy induced by selfish driving
  - CDC16 (8 nodes, 24 links, 56 OD pairs)
    - arXiv:1606.02194
    - slides url: `http://people.bu.edu/jzh/cdc16_slides.pdf`
  - IFAC17 (22 nodes, 74 links, 462 OD pairs)
    - arXiv:1610.09580
  - IEEE18 (74 nodes, 34 zones, 258 links, 1122 valid OD pairs)
    - EMA highway benchmark network released (Github/Kaggle)
- Estimate cost functions in multi-class transportation networks
  - CDC17 (use other benchmark networks)
    - arXiv:1703.04010

# Inferring User Flows — Converting Speeds to Flows



(from Google images)

- Greenshield's model (Mathew (2014)):

$$x_a = 4m_a \left[ \frac{v_a}{v_a^0} - \left( \frac{v_a}{v_a^0} \right)^2 \right],$$

where $m_a$ is the flow capacity, $v_a$ the current average speed, and $v_a^0$ the free-flow speed

- Assume these inferred flow observations form an equilibrium (Wardrop (1952)) under a "user-centric" routing policy; $x_a^{\text{user}}$

# Estimating OD Demand Matrix

- Define $\mathbf{A}$ as link-route incidence matrix, $\mathbf{P}$ route choice probability matrix, $\mathbf{S}$ sample covariance matrix, and $\boldsymbol{\lambda}$ vectorized OD demand matrix
- Let $\mathscr{R}_i$ be the set of all feasible routes connecting OD pair $i$
- Estimate $\boldsymbol{\lambda}$ by using a Generalized Least Squares (GLS) method (Hazelton (2000)):

$$\begin{aligned} \min_{\mathbf{P}\geq 0,\, \boldsymbol{\lambda}\geq 0} \quad & \sum_{k=1}^{K} \left(\mathbf{x}^{(k)} - \mathbf{A}\mathbf{P}'\boldsymbol{\lambda}\right)' \mathbf{S}^{-1} \left(\mathbf{x}^{(k)} - \mathbf{A}\mathbf{P}'\boldsymbol{\lambda}\right) \\ \text{s.t.} \quad & p_{ir} = 0 \quad \forall (i,r) \in \{(i,r) : r \notin \mathscr{R}_i\}, \\ & \mathbf{P}\mathbf{1} = \mathbf{1}. \end{aligned}$$

  - Could be dependent on time-intervals (AM/MD/PM/NT) of day and days of week (weekday/weekend)
  - Estimate $\boldsymbol{\lambda}$ using data with different time stamps accordingly

# Estimating Cost Functions

- $t_a(x_a) = t_a^0 g\left(\frac{x_a}{m_a}\right)$
- Seek to find a cost function $g(\cdot)$ under which the observed user flows $x_a^{\text{user}}$ and the estimated OD demand $\boldsymbol{\lambda}$ are as consistent as possible (good data reconciling)
  - Could be dependent on time-intervals (AM/MD/PM/NT) of day and days of week (weekday/weekend)
  - Estimate $g(\cdot)$ using data with different time stamps accordingly
- Seek to find $g(\cdot)$ having strong predictive power (good generalization properties)
- Achieve these by solving an inverse optimization problem, which is reduced to a Quadratic Programming (QP) problem

# Estimating Cost Functions (cont.)

Given user flows $\{(x_a^k;\ a \in \mathscr{A}_k);\ k = 1, \ldots, K\}$. Let $\mathscr{H}$ be a Reproducing Kernel Hilbert Space (RKHS). Solve the following inverse optimization problem (role of $\gamma > 0$, regularization) (Bertsimas et al. (2014)):

$$
\begin{aligned}
\min_{g,\mathbf{y},\boldsymbol{\epsilon}} \quad & \|\boldsymbol{\epsilon}\| + \gamma \|g\|_{\mathscr{H}}^2 \\
\text{s.t.} \quad & \mathbf{e}_a' \mathbf{N}_k' \mathbf{y}^{\mathbf{w}} \le t_a^0 g\left(\frac{x_a}{m_a}\right), \qquad \forall \mathbf{w} \in \mathscr{W}_k,\ a \in \mathscr{A}_k,\ k = 1, \ldots, K, \\
& \sum_{a \in \mathscr{A}_k} t_a^0 x_a g\left(\frac{x_a}{m_a}\right) - \sum_{\mathbf{w} \in \mathscr{W}_k} \left(\mathbf{d}^{\mathbf{w}}\right)' \mathbf{y}^{\mathbf{w}} \le \epsilon_k, \qquad \forall k = 1, \ldots, K, \\
& \boldsymbol{\epsilon} \ge 0, \quad g \in \mathscr{H}, \\
& g\left(\frac{x_a}{m_a}\right) \le g\left(\frac{x_{\tilde{a}}}{m_{\tilde{a}}}\right), \qquad \forall a,\ \tilde{a} \in \bigcup_{k=1}^{K} \mathscr{A}_k \text{ s.t. } \frac{x_a}{m_a} \le \frac{x_{\tilde{a}}}{m_{\tilde{a}}}, \\
& g(0) = 1.
\end{aligned}
$$

# Estimating Cost Functions (cont.)

Take polynomial kernel $\phi(x,y) = (c+xy)^n$. Reformulate the inverse optimization problem as the following QP:

$$\min_{\boldsymbol{\beta},\mathbf{y},\boldsymbol{\epsilon}} \quad \|\boldsymbol{\epsilon}\| + \gamma \sum_{i=0}^{n} \frac{\beta_i^2}{\binom{n}{i} c^{n-i}}$$

$$\text{s.t. } \mathbf{e}_a' \mathbf{N}_k' \mathbf{y}^{\mathbf{w}} \le t_a^0 \sum_{i=0}^{n} \beta_i \left(\frac{x_a}{m_a}\right)^i, \ \forall \mathbf{w} \in \mathscr{W}_k, \ a \in \mathscr{A}_k, \ k = 1, \dots, K,$$

$$\sum_{a \in \mathscr{A}_k} t_a^0 x_a \sum_{i=0}^{n} \beta_i \left(\frac{x_a}{m_a}\right)^i - \sum_{\mathbf{w} \in \mathscr{W}_k} (\mathbf{d}^{\mathbf{w}})' \mathbf{y}^{\mathbf{w}} \le \epsilon_k, \ \forall k = 1, \dots, K,$$

$$\epsilon_k \ge 0, \ \forall k = 1, \dots, K,$$

$$\sum_{i=0}^{n} \beta_i \left(\frac{x_a}{m_a}\right)^i \le \sum_{i=0}^{n} \beta_i \left(\frac{x_{\tilde{a}}}{m_{\tilde{a}}}\right)^i, \ \forall a, \ \tilde{a} \in \bigcup_{k=1}^{K} \mathscr{A}_k \text{ s.t. } \frac{x_a}{m_a} \le \frac{x_{\tilde{a}}}{m_{\tilde{a}}},$$

$$\beta_0 = 1.$$

# Estimating Cost Functions (cont.)

Solving the QP gives an estimator $\hat{g}(\cdot)$ of $g(\cdot)$:

$$\hat{g}(x) = \sum_{i=0}^{n} \beta_i^* x^i = 1 + \sum_{i=1}^{n} \beta_i^* x^i.$$

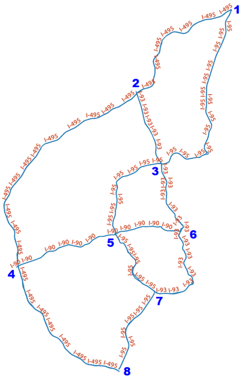
# Finding Social Flows

$$\text{PoA} = \frac{\sum_{a\in\mathscr{A}} x_a^{\text{user}} t_a(x_a^{\text{user}})}{\sum_{a\in\mathscr{A}} x_a^{\text{social}} t_a(x_a^{\text{social}})} \geq 1$$

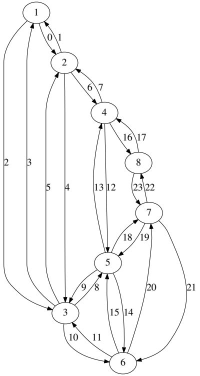
- Now ready to calculate the social flows
- Find the social flows $x_a^{\text{social}}$ by solving the following NLP (Patriksson (1994)):

$$\text{(socialOpt)} \quad \min_{\mathbf{x}\in\mathscr{F}} \sum_{a\in\mathscr{A}} x_a t_a(x_a)$$

# A Sub-Map of EMA



(a)



(b)

- 8 nodes
- 24 links
- $8 \times (8-1) = 56$ OD pairs

Figure: (a) An interstate highway sub-network; (b) The topology of the sub-network.

# Results for Cost Function Estimation

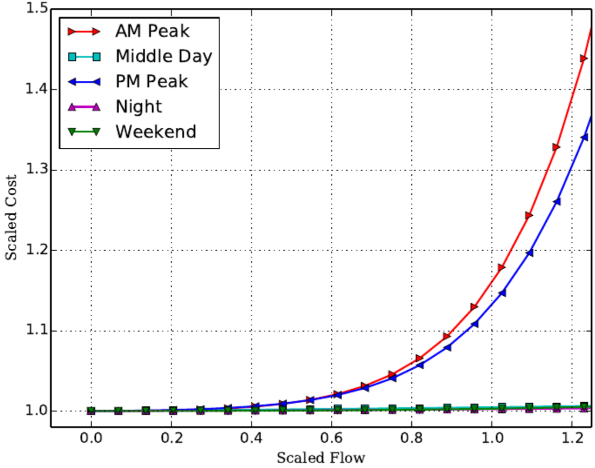
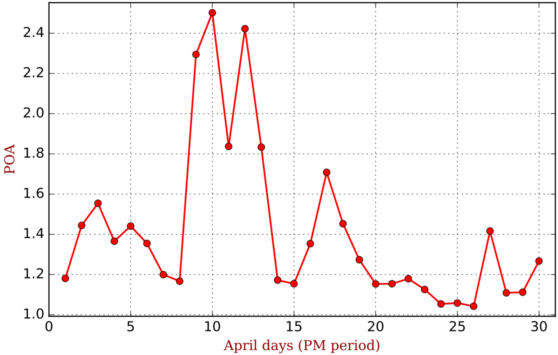



Figure: Estimated $g(\cdot)$ for different time periods (Apr. 2012).

# Results for PoA




Average PoA $\approx 1.5$, meaning we can improve the road network by about 50%; some PoA $> 2$, meaning we can gain more than 100% improvement!

# Ongoing/Future Directions

- Extend single-class model to multi-class model
  - cf. CDC17; estimate cost functions only
- Do the following jointly (finished):
  - estimate cost functions and adjust OD demands (assuming an initial "rough" OD demand matrix is at hand)
- Use more complicated model to convert speeds to flows
- Take all roads into account rather than highway roads only
- Evaluate PoA for special dates; July 4, Dec. 25, Jan. 1, etc.
- Develop alternative methods to estimate OD demand matrices
- Inverse problem (cost function estimation): consider flow observations with noises or missing data
  - other people have done some work in this regard
- Consider stochastic user equilibrium problem and its inverse

etc.

# Q & A

## Thank You!

**Questions:** jzh@bu.edu

**Further References:**

http://people.bu.edu/jzh/

https://github.com/jingzbu/InverseVIsTraffic (contains Github/Kaggle links to EMA highway benchmark network)