CALCULUS FOR THE MORNING COMMUTE

Prof. David Bernstein
Department of Computer Science
JAMES MADISON UNIVERSITY®

JMU Department of Mathematics and Statistics
Fall 2007 Colloquia

Motivation

- Some History:

  The shortest path problem is easy to solve

  25 years ago it became clear that we could develop route guidance systems if we only had network data

- The Breakthrough (in the U.S.):

  The Census Bureau digitized the entire country

- Today:

  A few companies maintain and market (very similar) street databases

  Many companies develop and market route guidance systems

- A Thought Experiment:

  What would happen if "everyone" used such a system to determine their route from home to work?

Notation

- The Network:

  $\mathcal{N}$ denotes the set of nodes

  $\mathcal{A}$ denotes the set of arcs (or links)

  $\mathcal{W} \subseteq \mathcal{N}^2$ denotes the set of origin-destinat pairs

  $\mathcal{R}_w$ denotes the set of routes connecting $w \in \mathcal{W}$ (with $R_w = |\mathcal{R}_w|$)

  $\mathcal{R} = \cup_{w \in \mathcal{W}} \mathcal{R}_w$ (with $R = \Sigma_{w \in \mathcal{W}} R_w$)

  $c : \mathbb{Z}_+^R \rightarrow \mathbb{R}_+^R$ denotes the vector of route cost functions

- Demand and Flows:

  $D = (D_w : w \in \mathcal{W}) \in \mathbb{Z}_{++}^R$ denotes the demand vector

  $h = (h_r : r \in \mathcal{R}) \in \mathbb{Z}_+^R$ denotes the route "flow" vector

  $H_D = \left\{ h \in \mathbb{Z}_+^R : \Sigma_{r \in \mathcal{R}_w} h_r = D_w, w \in \mathcal{W} \right\}$ denotes a feasible route flow pattern

Behavioral Model

- The Key Behavioral Assumption:

  Each commuter attempts to minimize her/his own travel cost, given the behavior of other commuters

- The Equilibrium Model:

  A flow pattern, $h \in H_D$ is said to be a (Nash) network equilibrium iff:

  $$h_r > 0 \Rightarrow c_r(h) \leq c_s(h + 1_s - 1_r) \forall s \in \mathcal{R}_w$$

  for all $r \in \mathcal{R}_w$ and $w \in \mathcal{W}$.

- An Interpretation:

  In equilibrium, no commuter has any incentive to change her/his route

The Important Questions

- Existence:

  When will such an equilibrium exist?

- Uniqueness:

  When will there be exactly one equilibrium?

- Stability:

  When will the equilibrium (or set of equilibria) be stable?

- Computation:

  Can we find equilibria efficiently (using a numerical algorithm)?

A Simplifying Assumption

- Something a Computer Scientist Should Never Do:

  Consider the limiting case

- Operationalizing this Simplification:

  The equilibrium condition:

  $$h_r > 0 \Rightarrow c_r(h) \leq c_s(h + 1_s - 1_r) \forall s \in \mathcal{R}_w$$

  becomes:

  $$h_r > 0 \Rightarrow c_r(h) \leq c_s(h + \epsilon 1_s - \epsilon 1_r) \forall s \in \mathcal{R}_w$$

  and we consider the limit as $\epsilon \rightarrow 0$

Starting Again

- Notation:

  Replace $\mathbb{Z}$ with $\mathbb{R}$ above

  Let $\underline{c}_w(h) = \min\{c_s(h) : s \in \mathcal{R}_w\}$

- The Equilibrium Model:

  A flow pattern, $h \in H_D$ is said to be a continuous network equilibrium iff:

  $$h_r > 0 \Rightarrow c_r(h) = \underline{c}_w(h)$$

  for all $r \in \mathcal{R}_w$ and $w \in \mathcal{W}$.

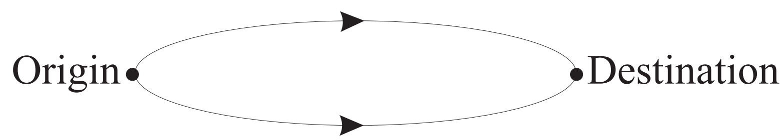
  The set of such flow patterns is denoted by CNE(c, D)

- Interpretation:

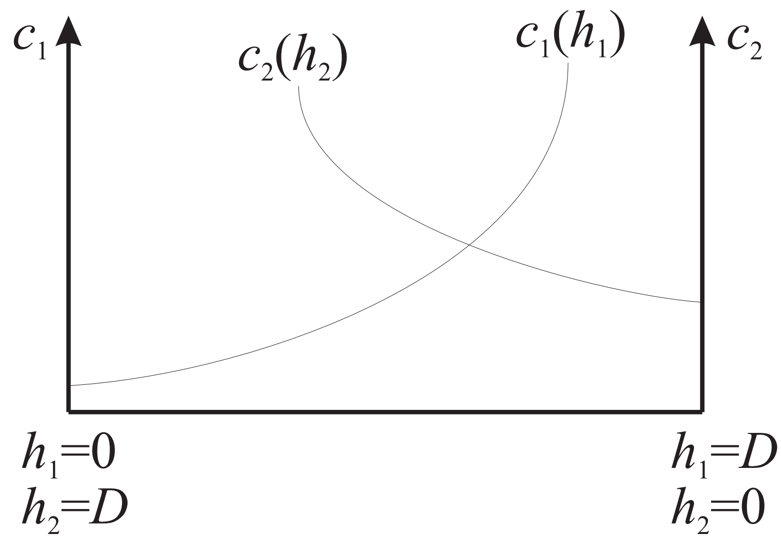
  In equilibrium, the cost on all used routes (connecting an OD-pair) must be minimal (and, hence, equal)

A Graphical Approach

- The Network:

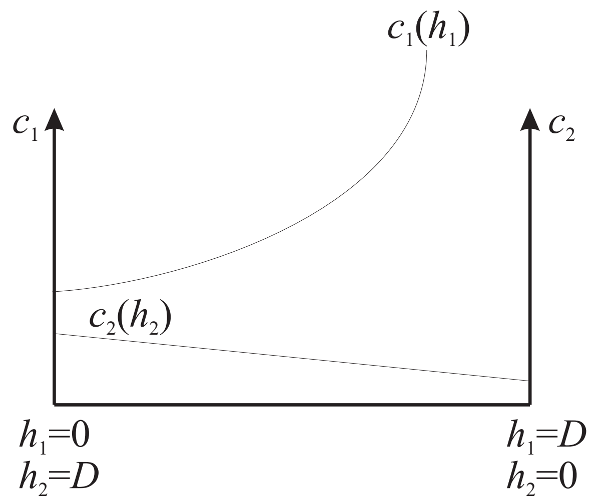


- A Simple Case:

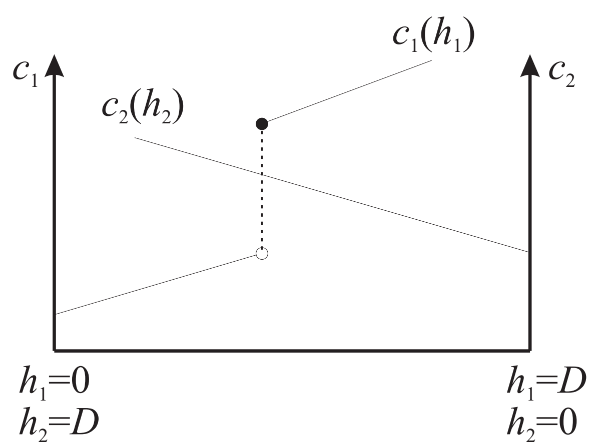


Some Interesting Cases

- No Intersection:

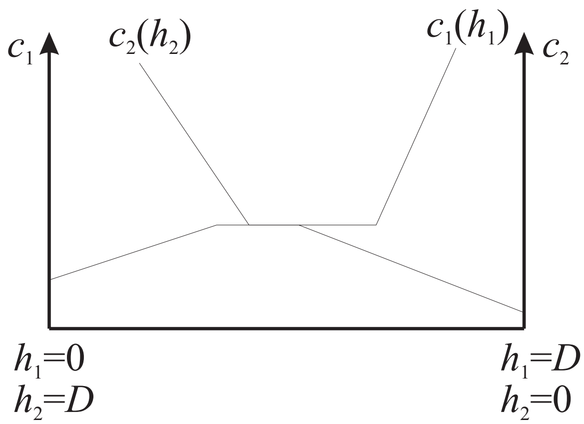


- Discontinuous Costs:

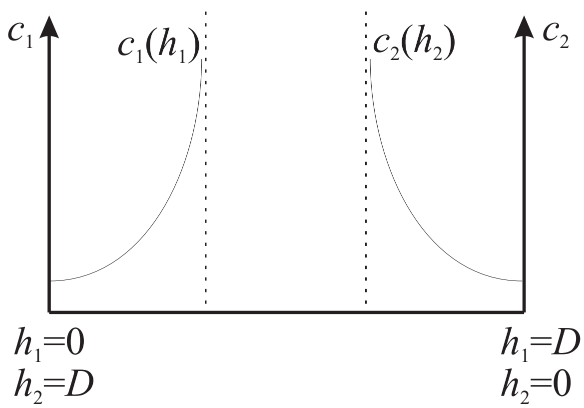


Some Interesting Cases (cont.)

- Many Points of Equal Cost:



- No Intersection:



Lagrange Multipliers: A Review

- The Theorem:

  Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $h : \mathbb{R}^m \longrightarrow \mathbb{R}$ be "appropriately continuous and differentiable" functions. Let $\overline{x} \in \mathbb{R}^n$ with $h(\overline{x}) = c$, and let S denote the level set for h with value c. If f restricted to S has an optimum at $\overline{x}$ then there exists a $\lambda \in \mathbb{R}$ such that:

  $$\nabla f(\overline{x}) - \lambda \nabla h(\overline{x}) = 0$$

- An Interpretation:

  Let

  $$\mathcal{L}(x) = f(x) - \lambda h(x)$$

  and consider the critical points of $\mathcal{L}$

Karush-Kuhn-Tucker Conditions

- Motivation:

  L involves problems of the form:

  $$\begin{array}{ll} \min_x & f(x) \\ \text{s.t.} & h_i(x) = 0 \text{ for } i = 1, \ldots, l \end{array}$$

  KKT deal with problems of the form:

  $$\begin{array}{ll} \min_x & f(x) \\ \text{s.t.} & h_i(x) = 0 \text{ for } i = 1, \ldots, l \\ & g_i(x) \leq 0 \text{ for } i = 1, \ldots, m \end{array}$$

- The Necessary Conditions:

  For "appropriately continuous and differentiable" functions, if $\overline{x}$ solves the problem above then there exist scalars μ_i and λ_i such that:

  $$\begin{aligned} \nabla f(\overline{x}) + \sum_{i=1}^{m} \mu_i \nabla g_i(\overline{x}) + \sum_{i=1}^{l} \lambda_i \nabla h_i(\overline{x}) &= 0 \\ \mu_i g_i(\overline{x}) &= 0 \text{ for } i = 1, \ldots, m \\ \mu_i &\geq 0 \text{ for } i = 1, \ldots, m \end{aligned}$$

- The Sufficient Conditions:

  Require appropriate convexity of f, g_i and h_i

Back to the Continuous Equilibrium Problem

- An Optimization Problem:

$$\begin{array}{lll} \min_h & V(h) & \\ \text{s.t.} & \Sigma_{r \in \mathcal{R}_w} h_r = D_w & w \in \mathcal{W} \\ & h_r \geq 0 & r \in \mathcal{R} \end{array}$$

- The KKT Conditions:

$$\nabla_r V(h) - \mu_w - \eta_r \geq 0 \text{ for all } r \in \mathcal{R}_w \text{ and } w \in \mathcal{W}$$

$$\eta_r h_r = 0 \text{ for all } r \in \mathcal{R}$$

$$\eta_r \geq 0 \text{ for all } r \in \mathcal{R}$$

Using the KKT Conditions

$$\nabla_r V(h) - \mu_w - \eta_r \geq 0 \text{ for all } r \in \mathcal{R}_w \text{ and } w \in \mathcal{W}$$
$$\eta_r h_r = 0 \text{ for all } r \in \mathcal{R}$$
$$\eta_r \geq 0 \text{ for all } r \in \mathcal{R}$$

- Some Simple Results:

  It follows from $\eta_r \geq 0$ and $\nabla_r V(h) - \mu_w - \eta_r \geq 0$ that $\nabla_r V(h) \geq \mu_w$

  It follows from $\eta_r h_r = 0$ that $h_r > 0 \Rightarrow \eta_r = 0$

  It follows that $h_r > 0 \Rightarrow \nabla_r V(h) = \mu_w$

- It Would Be Nice If:

  $\nabla_r(h) = c_r(h)$ for all $r \in \mathcal{R}$

  (i.e., $\nabla V(h) = c(h)$)

- Because Then:

  μ_w would be $\underline{c}_w(h)$

  A minimizer of $V(h)$ subject to $h \in H_D$ would be an equilibrium

In Order to Move Ahead

- An Important Question:

  Given a function, $c : \mathbb{R}_+^R \to \mathbb{R}_+^R$, under what conditions does there exist a function, $V \in \mathcal{C}^1(\mathbb{R}_+^R)$, with $\nabla V = c$?

- This Question Arises Elsewhere:

  In physics V is called a potential function for the gradient vector field, c

In Order to Move Ahead (cont.)

- One Important Sufficient Condition:

  Path Independence [Stoke's Theorem]

- A Potential Function:

$$V(h) = \oint_0^h c(\omega) d\omega$$

- A More Easily Verified Sufficient Condition:

  Lipschitz continuous costs (i.e., $||c(x) - c(y)|| \leq K||x - y||$ where $|| \cdot ||$ denotes the max norm)

  Symmetry (i.e., $\nabla_r c_s(h) = \nabla_s c_r(h)$ for almost all h where both gradients exist) [Frobenius Theorem]

- A Usable Formulation:

  The Rectilinear Path of Integration:

$$V(h) = \sum_{i=1}^{R} \int_0^{h_i} c_i(h_1, h_2, \ldots, h_{i-1}, x, 0, \ldots, 0) dx$$

How Does This Help?

- Existence:

  An equilibrium will exist whenever a minimum exists

- Uniqueness:

  The objective function is, in general, not strictly convex so equilibrium route flows are not unique

  However, if route costs are additive we can use the following function of link flows:

$$\sum_{i=1}^{L} \int_0^{f_i} t_i(f_1, f_2, \ldots, f_{i-1}, x, 0, \ldots, 0)dx$$

  which is strictly convex if the link cost function, t, is strictly monotone

How Does This Help? (cont.)

- Calculation:

  Consider feasible descent algorithms – we need to calculate an initial feasible solution and then we need to iteratively find feasible descent directions. Can we?

- Stability:

  We need a behaviorally meaningful adjustment process

A Behavioral Adjustment Mechanism

- Notation:

  $\dot{h}_{rs}$ denotes the *switching rate* from r to s (where $r, s \in \mathcal{R}_w$) and $\dot{h}_r \equiv \Sigma_{s \in \mathcal{R}_w - r}(\dot{h}_{sr} - \dot{h}_{rs})$.

  $\dot{h}_{rs}$ is assumed to be determined by some *route switching process* $a_{rs}(h) \geq 0$ with an associated *adjustment operator* $a_r(h) \equiv \Sigma_{s \in \mathcal{R}_w - r}(a_{sr} - a_{rs})$.

  A function $p : \mathbb{R}_+ \rightarrow \mathbb{R}^R$ satisfying $\dot{p}(t) = a[p(t)], t \in \mathbb{R}_+$ is called an *adjustment path* for a.

- Behavioral Assumptions:

  A c-adjustment operator, a, must satisfy:

  (Rationality) $a_{rs}(h) > 0 \Rightarrow c_r(h) > c_s(h)$
  (Feasibility) $p(0) \in H_D \Rightarrow p \subseteq H_D$
  (Persistency) $a(h) = 0 \Leftrightarrow h \in \text{CNE}(c, D)$

  for all D and $r, s \in \mathcal{R}_w$

Adjustments and Stability

- Notation:

  The *Hausdorff distance* from a point, $h \in \mathbb{R}^R$, to a nonempty set, $S \subseteq \mathbb{R}^R$, is given by $\rho(h, S) = \inf\{\|h - x\| : x \in S\}$.

  The *neighborhood* of S in H_D is given by $H_D(S, \epsilon) = \{h \in H_D : \rho(h, S) \leq \epsilon\}$.

  The set of *minima* is given by $\text{MIN}(V^c, S) = \{h \in S : V^c(h) = \min_{g \in S} V^c(g)\}$.

- Definitions:

  A set $S \subseteq H_D$ is *locally* V^c*-minimal* iff there exists some $\epsilon > 0$ such that $S = \text{MIN}[V^c, H_D(S, \epsilon)]$ and S is *isolated* iff $S = \text{KKT}(V^c, H_D) \cap H_D(S, \epsilon)$.

  A set, $S \subseteq H_D$ is *asymptotically a-stable* iff there exists some $\epsilon > 0$ such that

  $$p(0) \in H_D(S, \epsilon) \Rightarrow \lim_{t \to \infty} \rho[p(t), S] = 0$$

  and for each $\epsilon > 0$ there is some $\alpha_\epsilon \in (0, \epsilon]$ such that

  $$p(0) \in H_D(s, \alpha_\epsilon) \Rightarrow p(\mathbb{R}_+) \subseteq H_D(S, \epsilon)$$

A Stability Property

Theorem:

If c is a gradient cost structure with cost potential V^c, then for all c-adjustment processes a and demand patterns D each isolated, locally V^c-minimal set, $S \subseteq H_D$ is asymptotically a-stable.

Sketch of the Proof:

1. Show that V^c is a strict *Liapunov function* for every c-adjustment process a (i.e., that $h \in H_D \Rightarrow \dot{V}^c(h) \leq 0$ and that $\dot{V}^c(h) = 0 \Rightarrow a(h) = 0$).
2. Show that if there exists a Liapunov function V^c for a on H_D then every locally V^c-minimal set is a-stable.
3. Show that all c-adjustment processes, a, for gradient cost structures eventually converge to network equilibria.

An Interesting Adjustment Process

- Notation:

  For any $x \in \mathbb{R}$ let $x_+ = \max\{0, x\}$

  $\delta_0(0) = 1$ and $\delta_0(x) = 0$ for all $x > 0$

- The Process:

  $a^{\infty}_{rs}(h) = h_r[c_r(h) - c_s(h)]_+ \delta_0[c_s(h) - \underline{c}_w(h)]$

  $a^{\infty}_{r}(h) = \Sigma_{s \in \mathcal{R}_w - r}[a^{\infty}_{sr}(h_+) - a_{rs}(h_+)]$

- What's Interesting About It?

  It involves discontinuities at every point where the set of minimal cost routes in $\mathcal{R}_w$ changes (for any $w \in \mathcal{W}$)

A Solution Concept for Discontinuous Dif. Eqs.

- Notation:

  Given $x \in \mathbb{R}^n$ and a nonempty set $S \subseteq \mathbb{R}^n$, the Hausdorff distance from x to S is defined as $\rho(x, S) = \inf\{||x - y|| : y \in S\}$

  For any $\epsilon > 0$, closed set $X \subseteq \mathbb{R}^n$ and nonempty set $S \subseteq X$, the ϵ-neghborhood of S in X is defined as $X(S, \epsilon) = \{x \in X : \rho(x, S) \leq \epsilon\}$

  conv(S) denotes the convex hull of the set $S \in \mathbb{R}^n$

  cconv(S) denotes the closure of the convex hull of the set $S \in \mathbb{R}^n$

- Getting Started:

  If the image set of a^∞ is denoted by $a^\infty[H_D(h, \epsilon)] = \{a^\infty(g) : g \in H_D(h, \epsilon)\}$ then cconv$(a^\infty[H_D(h, \epsilon)])$ contains all of the convex combinations of vectors in $a^\infty[H_D(h, \epsilon)]$

  $\bigcap_{\epsilon>0}$ cconv$(a^\infty[H_D(h, \epsilon)])$ must contain the limits of such convex combinations as the size of the neighborhood goes to zero

A Solution Concept for Disc. Dif. Eqs. (cont.)

- A Reminder

  A function, $f : [a, b] \longrightarrow \mathbb{R}^n$ is said to be absolutely continuous on $[a, b]$ if given $\epsilon > 0$ there exists a $\delta > 0$ such that

  $$\sum_{i=1}^{n} ||f(y_i) - f(z_i)||_\infty < \epsilon$$

  for every finite collextion, $\{(y_i, z_i) : i = 1, \ldots, n\}$ of nonoverlapping intervals with $\Sigma_{i=1}^{n} |y_i - z_i| < \delta$

- Krasovkij Solutions to a^∞:

  An absolutely continuous function, $p : \mathbb{R}_+ \longrightarrow \mathbb{R}_+^R$ is a solution to a^∞ iff for almost all $t \in \mathbb{R}_+$

  $$\dot{p}(t) \in \bigcap_{\epsilon > 0} \text{cconv}(a^\infty[H_D(p(t), \epsilon)])$$

- An Observation:

  Over regions of H_D where a^∞ is continuous, this reduces to $\dot{p}(t) = a^\infty[p(t)]$ (i.e., the classical Carathéodory solution)

An Example

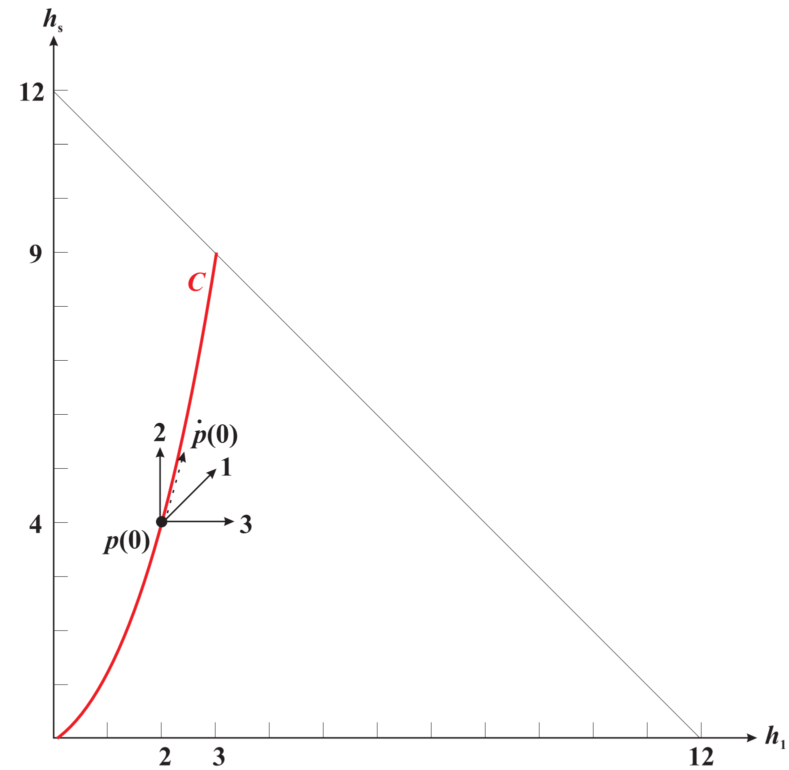
- One OD with three Routes:

$$D = 12$$
$$c_1(h) = 1 + h_1^2$$
$$c_2(h) = 1 + h_2$$
$$c_3(h) = 15 + h_3$$

- Solving:

The unique equilibrium is $h^* = (3, 9, 0)$

An Example



C, is the set of points with $c_1(h) = c_2(h)$

At $p(0)$ commuters on 3 will siwtch to 1 and 2 so p moves in direction 1

This movement collapses the set of minimum cost routes from $\{1, 2\}$ to $\{2\}$ so commuters instantly stop switching to 1, moving p in direction 2

This shifts the set of minimum cost routes to $\{1\}$ so that commuters instantly stop switching to 2 and start switching to 1, moving p in direction 3

An Example (cont.)

- Conclusion:

  No direction of movement can persist for more than an instant (called a *chattering regime*)

  Commuters are continuously switching to 1 or 2 (or both) so the p is moving up and to the right

- Interpretation:

  The cone of feasible direction vectors corresponds to $\cap_{\epsilon>0}\,\text{cconv}(a^\infty[H_D(p(0),\epsilon)])$ (i.e., the Krasovskij solutions)

- Observations:

  The only absoutely continuous adjustment path starting at a point on C is the path with trajectory along C

  $\dot{p}(t)$ must be almost everywhere tangent to C

  $a^\infty[p(t)]$ is always parallel to direction 1 for any point $p(t) \in C$

  So $\dot{p}(t)$ is *nowhere* equal to $a^\infty[p(t)]$

Lower-Semicontinuous Costs

- Definition of Equilibrium:

  The classical definition (Wardrop) doesn't work

  Other definitions (Dafermos, Heydecker) allow collaboration

  The most obvious alternative (Nash) doesn't work

- Modeling "Smallness":

  $\text{liminf}_{\epsilon \downarrow 0} c_r(h + \epsilon 1_r - \epsilon 1_s) =$
  $\lim_{\epsilon \downarrow 0} \inf\{c_r(h + \alpha 1_r - \alpha 1_s) : 0 < \alpha < \min(\epsilon, h_s)\}$

- New Definition:

  h is a *user equilibrium* iff:

  $$h_r > 0 \Rightarrow c_r(h) \leq \text{liminf}_{\epsilon \downarrow 0} c_s(h + \epsilon 1_s - \epsilon 1_r)$$

Lower-Semicontinuous Costs (cont.)

- One Result:

  If c is lower semicontinuous then every Wardrop eqm is a user eqm

  If c is upper semicontinuous then every user eqm is a Wardrop eqm

- Another Result:

  If c is lower semicontinuous and flow shifts between routes only create discontinuities on those links where the flow changes then a user equilibrium exists

SRD Equilibrium

- The Issue:

  Drivers simultaneously choose both a *path*, p, and a *departure time*, $t \in [0, T]$.

- Fluid Approximation:

  The *departure rate* on path p at time t is denoted by $h_p(t) \in \Re_+$

SRD Equilibrium (cont.)

- Notation:

  $$h(t) = (h_r(t) : r \in \mathcal{R})$$

  $H_D = \left\{h : \sum_{r \in \mathcal{R}_w} \int_0^T h_r(t) d\nu(t) = D_w, w \in W\right\}$ (where $\nu(t)$ is a Lebesgue measure on $[0, T]$).

- Assumptions:

  Each departure rate pattern, h, gives rise to a (time varying) *traffic pattern* , $x(t) = (x_a(t) : a \in A)$.

  The relationship between h and x is driven by the time needed to traverse arc a when entered at time t, $d_a(t)$.

  Link travel times are determined by the link occupancies at the time the link is entered (i.e., the number of vehicles ahead of you on a link when you enter).

SRD Equilibrium (cont.)

The time needed to traverse path r when departing from the origin at time t is given by:

$$d_p(t,h) = d_{a_1^r}[x_{a_1^r}(t)] \; + \; d_{a_2^r}[x_{a_2^r}(\tau_{a_1^r}(t))]$$

$$+ \; \cdots + d_{a_{m(r)}^r}[x_{a_{m(r)}^r}(\tau_{a_{m(r)-1}^r}(t))]$$

where

$$\tau_{a_1^r}(t) = t + d_{a_1^r}[x_{a_1^r}(t)] \quad \forall r \in \mathcal{R}$$

$$\tau_{a_i^r}(t) = \tau_{a_{i-1}^r}(t) + d_{a_i^r}[x_{a_i^r}(\tau_{a_{i-1}^r}(t))] \;\; \forall \; r \in \mathcal{R}, \; i \in [2, m(r)].$$

SRD Equilibrium (cont.)

- Travelers may arrive early or late but incur a *schedule cost*.

  α is the dollar penalty for early arrival.

  β is the dollar penalty for late arrival.

  $[T^* - \Delta, T^* + \Delta]$ is the set of "equally acceptable" arrival times.

- The schedule cost is given by

$$\Phi_r(t,h) = \left[\begin{array}{l} \alpha[(T^* - \Delta) - (t + d_r(t,h))] \quad \text{if} \quad (T^* - \Delta) > [t + d_r(t,h)] \\ \\ 0 \quad \text{if} \quad (T^* - \Delta) \leq [t + d_r(t,h)] \leq (T^* + \Delta) \\ \\ \beta[(t + d_r(t,h)) - (T^* + \Delta)] \quad \text{if} \quad (T^* + \Delta) < [t + d_r(t,h)], \end{array} \right.$$

- Letting γ denote the value of travel time, total cost is given by

$$c_r(t,h) = \gamma d_r(t,h) + \Phi_r(t,h) \quad \forall\, r \in \mathcal{R}$$

SRD Equilibrium (cont.)

Letting $\mu_r(h) = \text{ess inf}\{c_r(t,h) : t \in [0,T]\}$ the relevant lower bound on achievable costs for a w-commuter is given by $\mu_w(h) = \min\{\mu_r(h) : r \in \mathcal{R}_w\}$ we can define an equilibrium as follows:

Definition 1 (SRD Equilibrium) *A departure rate pattern, $h \in H_D$ is said to be a* simultaneous route and departure-time choice equilibrium *(SRD equilibrium) for D if and only if (iff) h satisfies the following condition for all $w \in W$, and $r \in \mathcal{R}_w$:*

$$h_r(t) > 0 \Longrightarrow c_r(t,h) = \mu_w(h)$$

for ν-almost all $t \in [0,T]$

SRD Equilibrium (cont.)

- It is possible to formulate the SRD equilibrium problem as an infinite dimensional variational inequality:

  Theorem 1 *A departure rate pattern, $\hat{h} \in H_D$ is an SRD equilibrium for c if and only if*

  $$\sum_{r \in \mathcal{R}} \int_0^T c_r(t, \hat{h})[h_r(t) - \hat{h}_r(t)]d\nu(t) \geq 0 \qquad (1)$$

  for all $h \in H_D$.

- This result is quite general, and does not depend on the form of d_a or c_r.

- This result allows us to consider questions of existence and solution algorithms.

The Big Open Questions

- SRD Equilibrium:

  Existence and Uniqueness

  Efficient Algorithms

- Information Provision:

  How will users respond to traffic forecasts?

  How do we incorporate their reponses into our forecasts?