AN INFORMAL OVERVIEW OF VARIATIONAL THEORY

Carlos F. Daganzo

Prepared for the Graduate Research School on Transport, Infrastructure and Logistics, Delft, The Netherlands

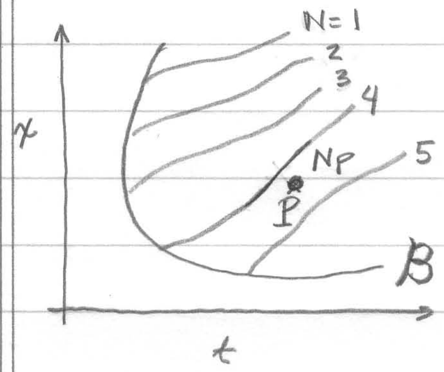
July 15, 2013

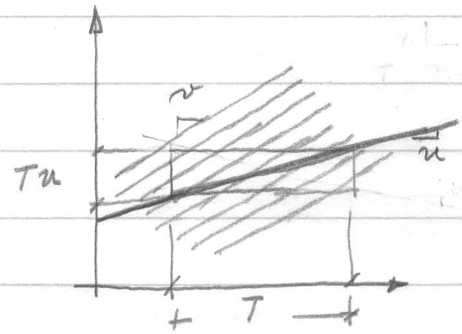
Prepared by:
Date:

CONTENTS

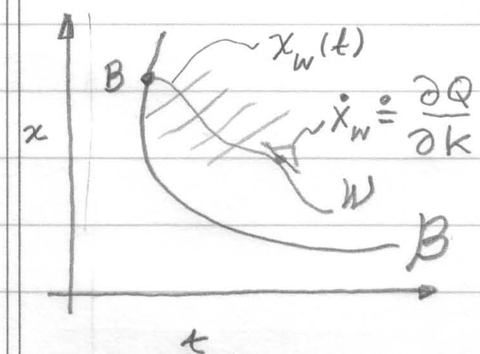
PLATE 1: KINEMATIC WAVE THEORY

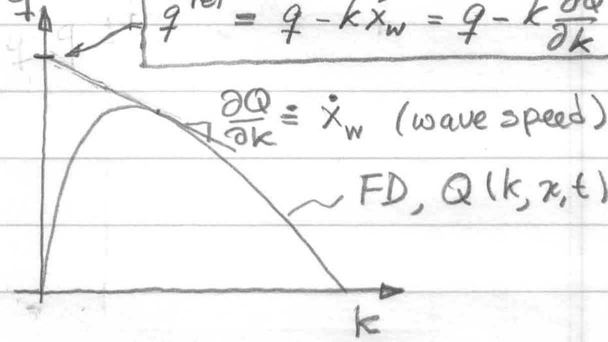
Prepared by:	
Date:	





$$\boxed{q^{rel} = q - ku}$$





$$\boxed{q^{rel} = q - k\dot{x}_w = q - k\frac{\partial Q}{\partial k}}$$

Postulates: (i) N is continuous; $\frac{\partial N}{\partial x \partial t} = \frac{\partial N}{\partial t \partial x} \rightarrow \frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = 0$, except at shocks.

(ii) FD: $q = Q(k, x, t)$.

(iii) stability: $|k - k'|$ declines with t.

PDE:

$$\text{(i), (ii)} \longrightarrow \frac{\partial Q}{\partial k}\frac{\partial k}{\partial x} + \frac{\partial k}{\partial t} = -\frac{\partial Q}{\partial x}$$ (a PDE for $k(t, x)$)

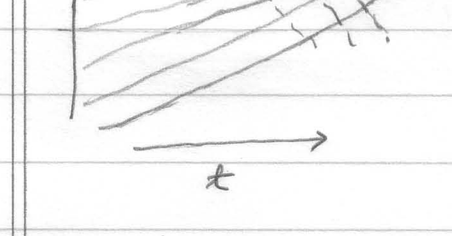
ODE along waves (characteristics):

$$\dot{x}_w(t) \doteq \frac{\partial Q}{\partial k} \longrightarrow \dot{k}_w(t) = -\frac{\partial Q}{\partial x}$$

Solution for homogeneous IVP with concave Q: [5]

It exists and is unique. Waves are truncated, and each point P is reached by only 1 wave.

Shocks are the loci of all truncation points



Of note:

$$N_P(W) = N_B + \int_{t_B}^{t_P} \underbrace{\left[Q(k_w(t), x, t) - k_w(t)\,\dot{x}_w(t)\right]}_{q_w^{rel}(t)} dt \qquad (1)$$

PLATE 2: VARIATIONAL THEORY – A SIMPLIFICATION OF KWT

P2

Prepared by:

Date:

Newell's minimum principle:

$$N_P^{KW} = \inf_{\mathcal{W} \in \mathbf{W}_P} \{ N_P(\mathcal{W}) \}$$

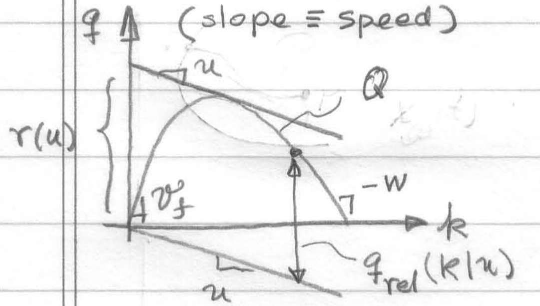
$N_P(\mathcal{W})$ – solve ODE

$\mathbf{W}_P$ – tough to identify

For homogeneous highways: $\frac{\partial Q}{\partial x} = 0$
- → waves are straight lines and
- → density & q^{rel} are constant
- → $N_{\mathcal{W}}(t)$ increases linearly
- → $N(t,x)$ is LE of ruled surface

Still tough to deal with moving bottlenecks, general boundary conditions and inhomogeneous highways

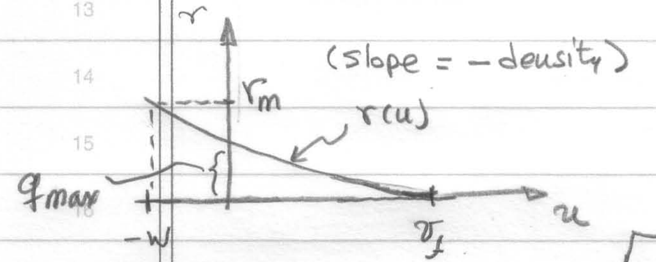
Variational Theory (VT):

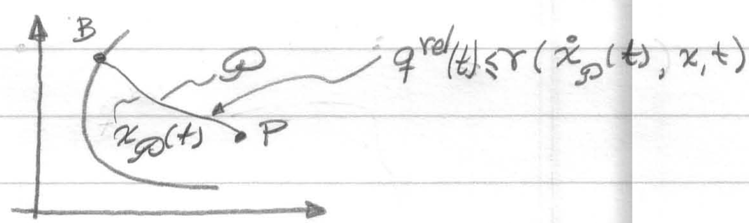


$q_{rel}(k|u) \le r(u)$ (relative capacity function)

curve "r" is LT of "curve "−Q"

$Q \begin{pmatrix} \text{concave} \\ \text{linear} \end{pmatrix} \rightarrow r \begin{pmatrix} \text{convex} \\ \text{linear} \end{pmatrix}$





- Path Capacity for $\mathcal{P} \in \mathbf{V}_P$ (valid paths):

$$\boxed{\begin{aligned} N_P^{KW} &= N_B + \int_{t_B}^{t_P} q^{rel}(t)\,dt \le N_B + \int_{t_B}^{t_P} r(\dot{x}_{\mathcal{P}}(t), x, t)\,dt \\ &\doteq N_B + \Delta_P(\mathcal{P}) \end{aligned}}$$

- Definition of VT: $N_P^{VT} \doteq \inf_{\mathcal{P} \in \mathbf{V}_P} \{ N_{B(\mathcal{P})} + \Delta_P(\mathcal{P}) \} \ge N_P^{KW}$ (2)

- As per plate 1: $\exists\, \mathcal{W} \in \mathbf{V}$; & according to (1) $q_{\mathcal{W}}^{rel}(t) = r(\dot{x}_{\mathcal{W}}(t), x, t)$

$\therefore N_P^{KW} = N_{B(\mathcal{W})} + \Delta_P(\mathcal{W}) \ge \inf_{\mathcal{P} \in \mathbf{V}} \{ N_B + \Delta_P(\mathcal{P}) \} \doteq N_P^{VT}$

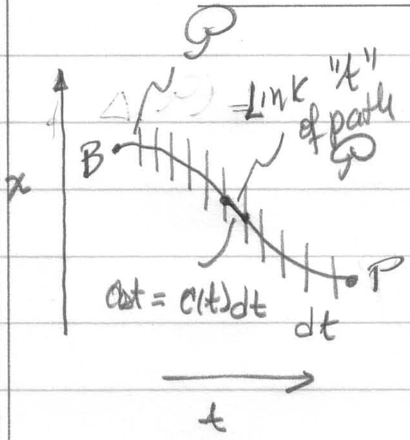
Hence $\boxed{N_P^{VT} = N_P^{KW}}$

- So we can do away with waves and ODE's! Solve (2) instead

PLATE 3 : VT INTERPRETATION & APPLICATION

Prepared by:

Date:



Suppose path is broken into links of ~~lea~~ duration dt and "cost" $c_{\mathcal{P}}(t)dt$, ~~where~~ The path cost is then: $\int_{t_B}^{t_P} c_{\mathcal{P}}(t)\,dt$.

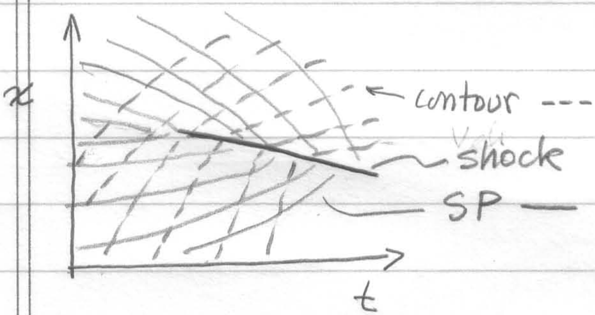
Assign costs: $c_{\mathcal{P}}(t) \doteq r(\dot{x}_{\mathcal{P}}(t), x, t)$

Then $\Delta_P(\mathcal{P})$ is the path cost.

> Eq. (2) of VT defines a SP problem w/ link costs on the t, x plane defined by the cost rate function $r(\dot{x}, x, t)$ of slope, time and location.

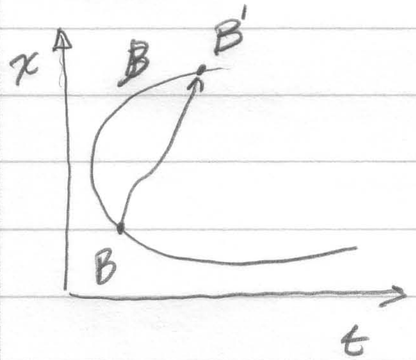
Waves = Shortest Paths ; Shocks = SP ends

Vehicle trajectories = Iso-cost contours



A simple & general framework for defining problems w/ complex boundary conditions including MB's

Caveat.



Problem data must include no contradictions; i.e. problem should be "well-posed"

If $B' \in \mathcal{B}$ can be reached by a valid path and $N_{B'}$ is given $N_{B'}^{given}$ then:

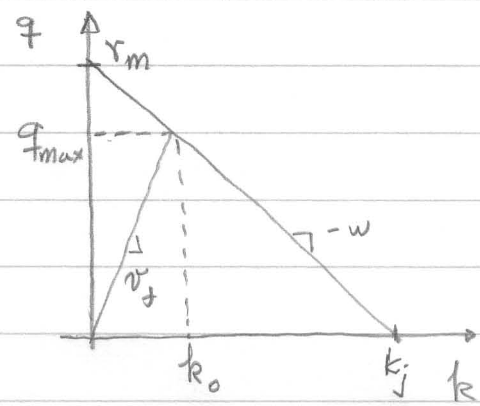
$$N_{B'}^{VT} = N_{B}^{given}$$

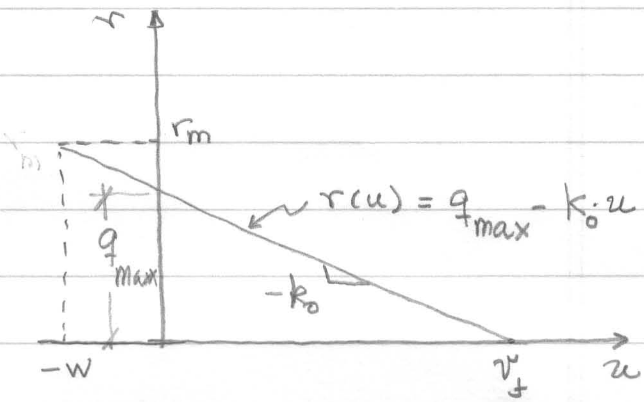
For more information on well-posed problems see [6]

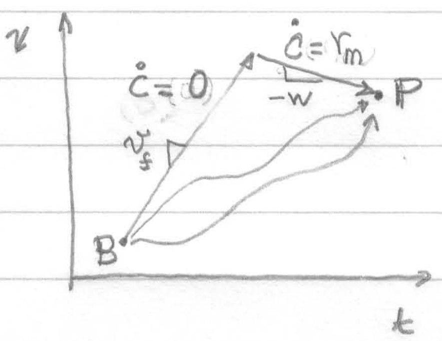
PLATE 4: THE TRIANGULAR FD SPECIAL CASE

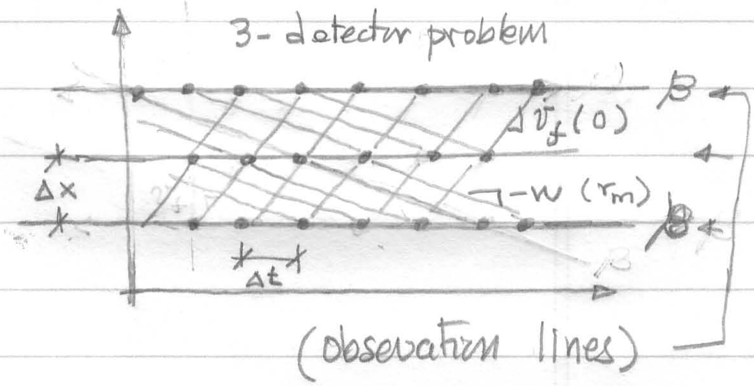
Prepared by:

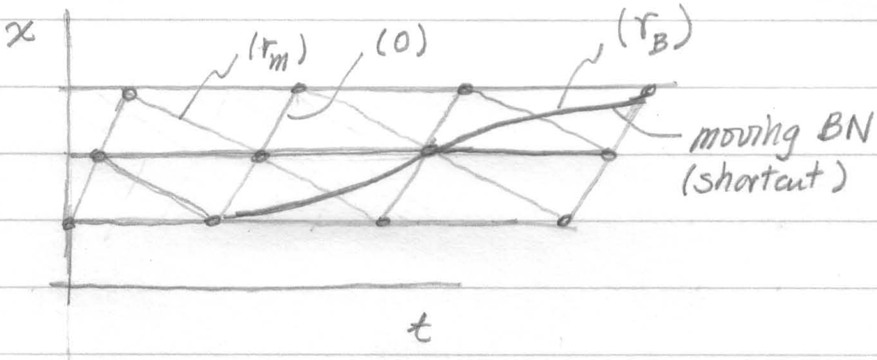
Date:

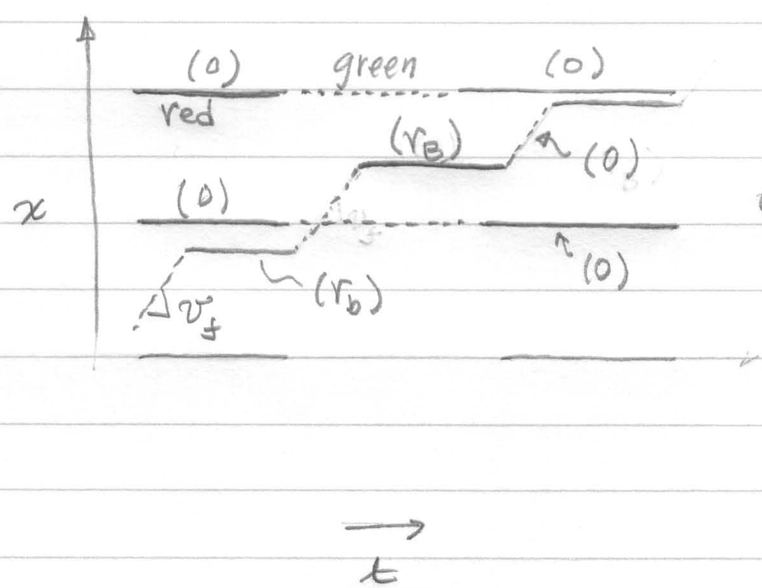








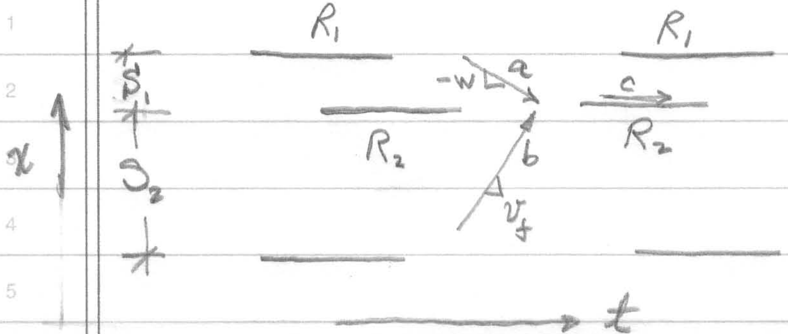




solid lines (shortcuts)
cost rates in parentheses

PLATE 5: ARTERIALS

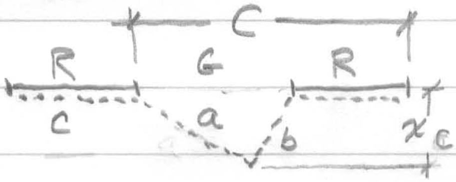
Prepared by:
Date:

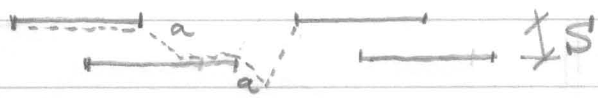


- ∞-SP w/ SLOPE = 0
- 3 types of links
- Only "a" has cost > 0

GOAL: Use a-type links the least time possible

capacity = avg cost $\bar{r}$ per unit time



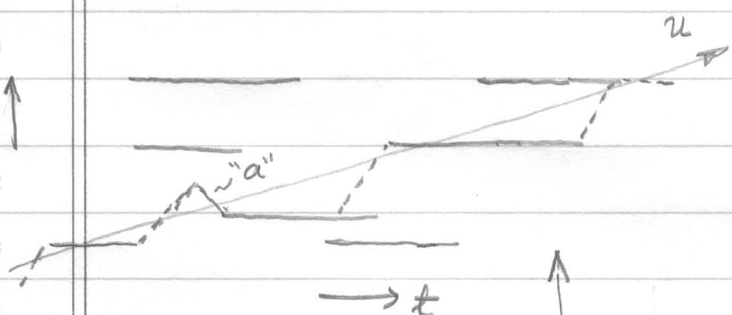


single intersection:
capacity = $q_{max} G$
$x_c = q_{max} G / k_j$

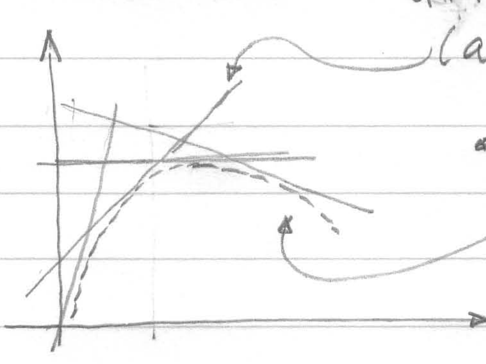
two intersections:
capacity < $q_{max} G$
only if $s' < x_c = q_{max} G / k_j$
and offset $\neq 0$

multiple intersections: capacity can be less than that of the worst pair of intersections

MFD:



- ∞-SP w/ SLOPE u
- Find least avg. cost rate $\bar{r}(u)$
- $\bar{Q}(k|u) = \bar{r}(u) + ku$ (a cut)

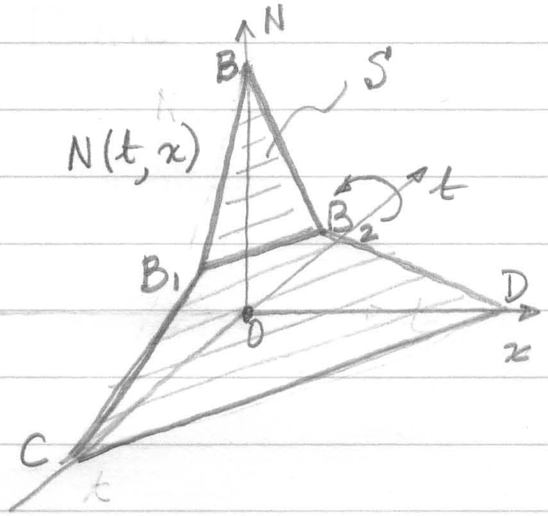


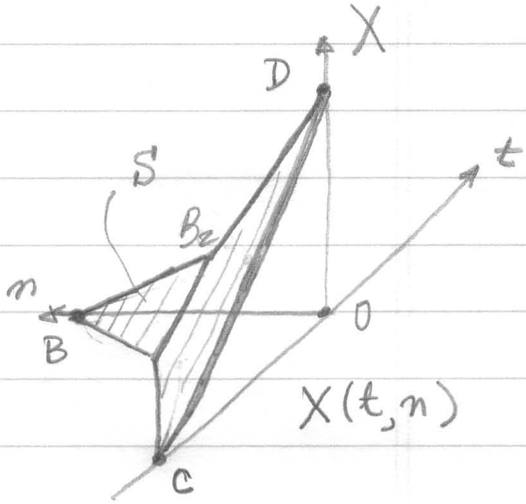
- $Q^{VT}(k) = \min_{u} \{\bar{Q}(k|u)\}$

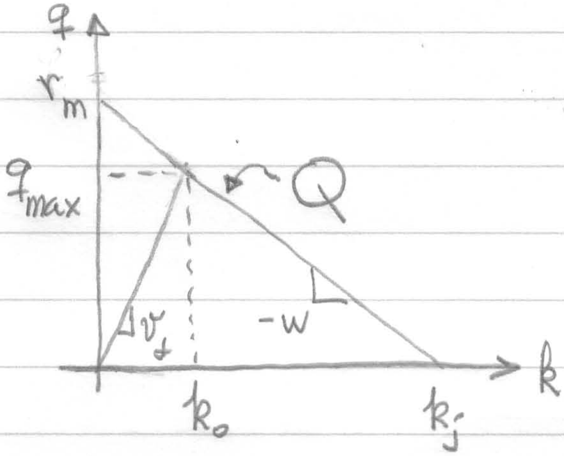
PLATE 6 : DUALITY

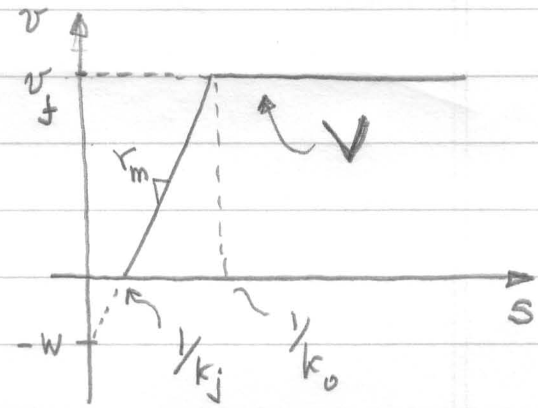
Prepared by:

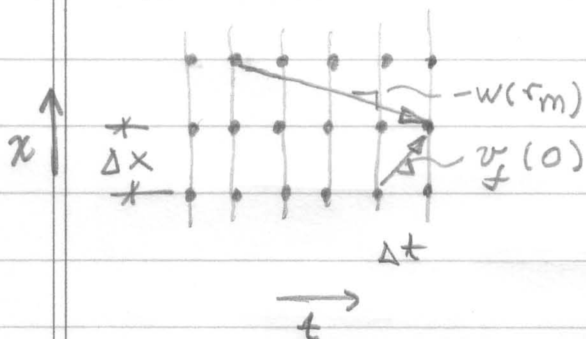
Date:

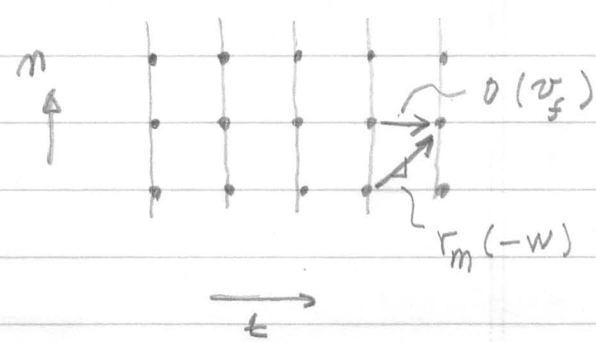












PRIMAL

$\Delta x = v_f \, \Delta t$

v_f = multiple of w

DUAL

$\Delta n = r_m \, \Delta t$

OBJECTIVE

To present the key ideas of VT in a single place as intuitively as possible

How it came about

• How it came about...

In the 1990's the great Gordon Newell invented this simplified theory of kinematic waves.[1] Much earlier (in the 1960's) he had shown how shocks came about in non-linear car-following models, connecting in this way KW theory and CF theory.[2] He believed that in both theories drivers always acted to be as far along the road as possible, which is intuitive because that's how most people drive. We talked about this in the late 90's. After he passed away the though gnawed at me & I wondered if there was a way in which traffic flow theory could be formulated as an optimization problem.

Inspiration struck me in the early 2000's.[3] Newell had shown in his KW work of the 90's that the vehicle found at any time and space point (t, x) along a road is the smallest of the vehicle numbers predicted by following the kinematic waves that reach the point in question. This is useful but tedious to do for problems with time- and space-dependent condition because the wave paths can be complicated and it is hard to tell a priori how many waves pass through a point. (Newell's theory works very well for simple problems with triangular fundamental diagrams when waves are straight lines and at most 2 waves can pass through each point (t, x).)

Given the difficulty dealing with waves... the inspiration was asking whether we could deal instead with <u>all</u> the spacetime paths passing through our point (t, x). This is of course a much larger set but one that did not require doing anything special to identify the members of the set.

If it turned out that one really should minimize a number associated with each path, then calculus of variations could be used to solve the problem! But... which number should be associated with each path for a minimization approach to make sense? And if something was found that made sense, could it be shown to produce the same answers as KWT?

• Description of VT

The answer to the first question, for a path $\mathcal{P}$ leading from the boundary $\mathcal{B}$ to a point P in the solution domain, was the largest possible vehicle number that could be observed at point P.

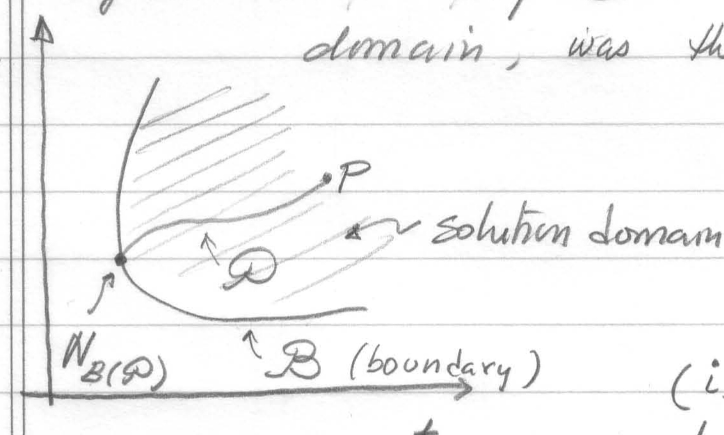


Figure 1

This quantity is the sum of two numbers: (i) The vehicle number observed at the boundary, at the point $B(\mathcal{P})$ where the path begins; and (ii) the maximum number of vehicles that could pass an observer traveling alongside the road on path $\mathcal{P}$, $\Delta(\mathcal{P})$. If we use the convention that $N_{\circ}$ denotes the vehicle number at point "$\circ$", the number associated with a path $\mathcal{P}$ is:

$$N(\mathcal{P}) \doteq N_{B(\mathcal{P})} + \Delta(\mathcal{P}). \qquad (1)$$

Definition (1) clearly implies that the actual number at P, N_P, cannot exceed the RHS; i.e.:

$$N_P \leq N(\mathcal{P}) \doteq N_{B(\mathcal{P})} + \Delta(\mathcal{P}). \quad (2a)$$

and

$$N_P \leq \inf_{\forall \mathcal{P}} \{N(\mathcal{P})\}. \quad (2b)$$

The leap of faith is conjecturing that N_P is the largest LHS number satisfying (2b); i.e. that the VT prediction is the largest possible value satisfying the capacity constraints.

$$N_P^{VT} = \inf_{\forall \mathcal{P}} \{N(\mathcal{P})\} \quad (3)$$

This equation satisfies our desire for reducing the traffic flow theory problem to an optimization problem.†

This "smells" right but we still need to show how to calculate $\Delta(\mathcal{P})$ and then prove that (3) actually matches the KWT prediction.

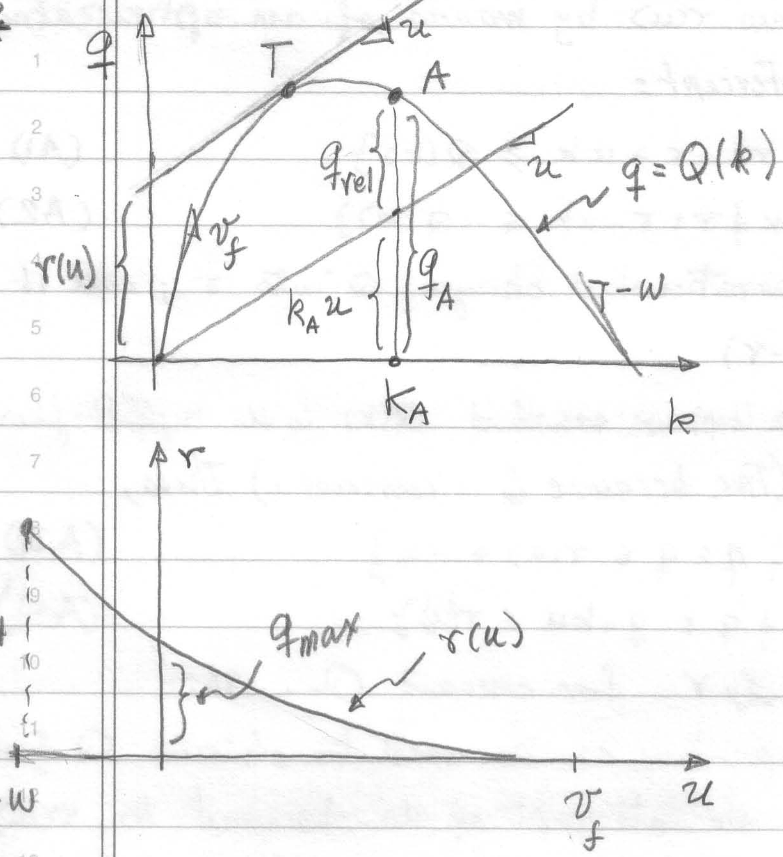
<u>How to calculate $\Delta(\mathcal{P})$:</u> Assume road is homogeneous.

Given is a FD (see Figure). We know from traffic flow theory that if traffic is in state A with flow q_A

---

† Given N_0 along a boundary curve $\mathcal{B}$, recipe (3) identifies N_P for every point P in the solution domain; i.e., a surface $N(t, x)$ whose partial derivatives (if they exist) are the flow $q(t, x)$ and negative of the density $k(t, x)$. This assumes that traffic flows in the direction of increasing x and vehicles are numbered in the direction of increasing time. Then: $\frac{\partial N}{\partial t} = q$; $\frac{\partial N}{\partial x} = -k$.

Figure 2



then the relative flow seen by an observer moving with speed u (i.e. the rate at which the observer is being passed) is $q_A - k_A u$, as shown.

The maximum passing rate is at "T" and is the intercept of the tangent line, denoted $r(u)$.

I call this quantity the "relative capacity" of the road because it is the maximum flow that can be observed in a frame of reference moving with the observer.

The figure on the bottom shows how $r(u)$ changes as u changes from $-w$ to v_f (the minimum and maximum slopes -- wave speeds -- of the FD.).
The transformation from $Q(k)$ to $r(u)$ is called the Legendre transform. You can verify that:

Q concave $\Rightarrow$ r convex (as shown)

Q triangular $\Rightarrow$ r linear

(See Aside at the end for more properties.)

Note: when $u=0$, $r(u) = q_{max}$ & when $u = v_f$, $r(u)=0$.
This should be intuitive.

Variational theory uses $r(u)$ instead of the MFD.

We can now write the formula for $\Delta(\mathcal{P})$. We consider only "valid paths" with trajectories $x_{\mathcal{P}}(t)$ such that the observer speed, $\dot{x}_{\mathcal{P}}(t) \doteq dx_{\mathcal{P}}(t)/dt$ is always in the $[-w, v_f]$ range.

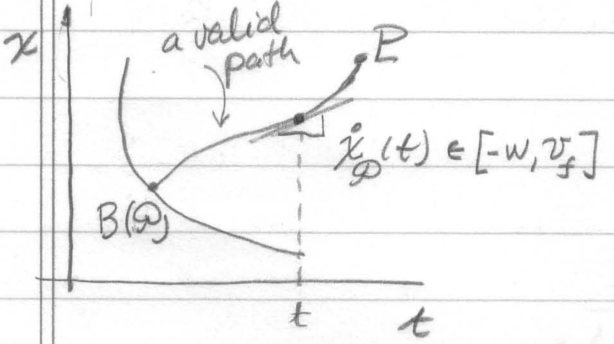


Figure 3

Then we have:

$$\Delta(\mathcal{P}) = \int_{t_{B(\mathcal{P})}}^{t_P} r(\dot{x}_{\mathcal{P}}(t))\,dt. \quad (4)$$

Thus, the capacity constraint (2b) becomes:

$$N_P \leq \inf_{\mathcal{P} \in \mathcal{V}} \left\{ N_{B(\mathcal{P})} + \int_{t_{B(\mathcal{P})}}^{t_P} r(\dot{x}_{\mathcal{P}}(t))\,dt \right\} \quad (5)$$

where $\mathcal{V}$ is the set of valid paths. And the prediction of VT is:

$$N_P^{VT} = \inf_{\mathcal{P} \in \mathcal{V}} \left\{ N_{B(\mathcal{P})} + \int_{t_{B(\mathcal{P})}}^{t_P} r(\dot{x}_{\mathcal{P}}(t))\,dt \right\} \quad (6)$$

This is an ordinary calculus of variations problem for the curve $\mathcal{P}$. In other words it is a shortest path problem in the (t, x) continuum with $r(\dot{x})$ as the "cost" of moving a unit time with slope $\dot{x}$, and where the cost of every point $B \in \mathcal{B}$ is given, N_B. For this reason I often call $r(\cdot)$ the "cost function"... SIMPLE!

Note that in the solution of this shortest path problem: (1) the vehicle trajectories are the iso-cost contours in the (t, x) plane; (2) the set of determining waves that reach every point is the set of shortest paths emanating from the boundary; and (3) shock waves are where the shortest paths end... BEAUTIFUL!

Simplicity and beauty are usually a good thing because they make solutions easy to get and interpret, and they suggest correctness... but this needs to be verified

I verified correctness in three ways: (1) by building on Newell's method [4]; (2) by calculus of variations arguments [4]; and (3) by building on Lax's KW work [3].

The simplest proof is (3) and therefore I can show it now:

It is well known from Lax's work [5] that if any "initial value problem" of KW theory in which Q is concave and the data satisfy some mild smoothness conditions then through every point P in the solution domain there passes a wave W_P with trajectory $x_w(t)$, such such that the traffic state, S, at time t is a point on the FD where the slope of Q is $dQ/dk = \dot{x}_w(t)$. See Figure 4.

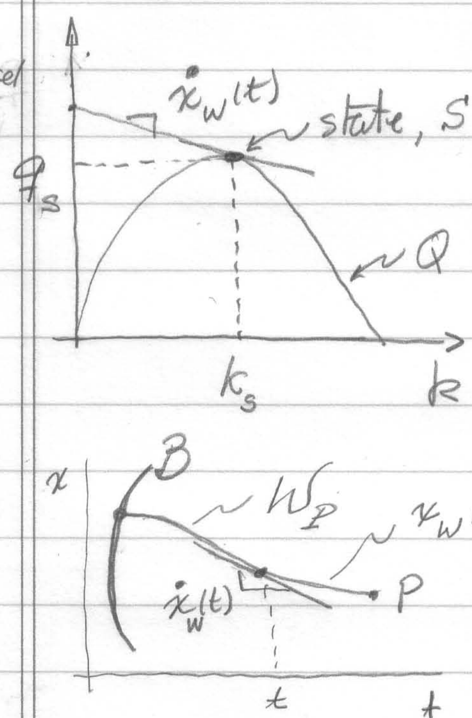


Figure 4

The rate at which traffic passes the observer traveling with the wave in the KW solution is q_S^{rel}. Note q_S^{rel} satisfies:

$$q_S^{rel} = r(\dot{x}_w(t)).$$

(Compare Figures 4 and 2.)

Hence the KW solution assigns a vehicle number N_P^{KW} to point P with the following formula (compare with (5) and (6)):

$$N_P^{KW} = N_{B(W_P)} + \int_{t_B}^{t_P} r(\dot{x}_w(t))\, dt = N(W_P). \quad (7)$$

Note that W_P is a valid path and therefore a value in the optimization (5). Hence, $N_P^{KW} \geq N_P^{VT}$.

We saw earlier that N_p^{VT} is the largest possible value for N_p of any theory satisfying the capacity constraints — such as KWT; i.e. that $N_p^{VT} \geq N_p^{KW}$.

Hence, since $N_p^{VT} \geq N_p^{KW}$ and $N_p^{VT} \leq N_p^{KW}$, it follows that

$$N_p^{VT} = N_p^{KW}.$$

Both theories yield the same result for the initial value problem studied by Lax, and are therefore identical.

References [4, 6] show that the ideas extend to more general problems, as they arise in practical transportation contexts, including complex boundary conditions. In essence, it is shown that the two theories are equivalent if the boundary data contains no contradiction. That is, if a valid path can go to a point on the boundary, $P \in B$, then the pre-specified value (or set of values) for N_p must be consistent with the theory's prediction:

$$N_p = N_p^{VT}.$$

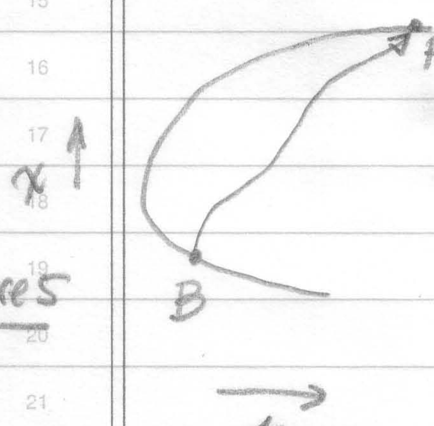


Figure 5

In the initial value problem paths cannot reach the boundary, so this issue does not arise.

- **Applications: the case of triangular FD's (based on [8]).**

If the FD is triangular the cost function is linear.

In this case traffic forecasting becomes a shortest path problem on planar networks, solvable with dynamic programming.

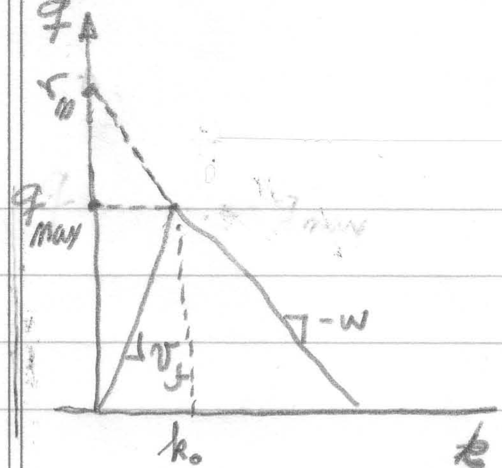
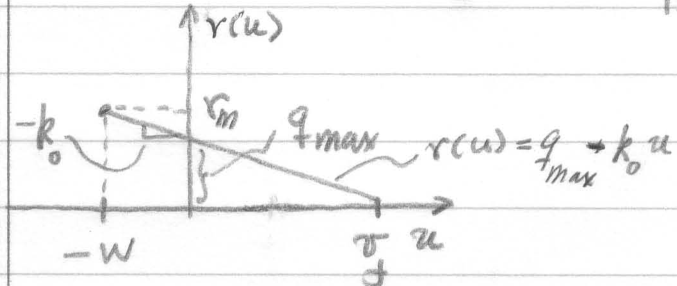
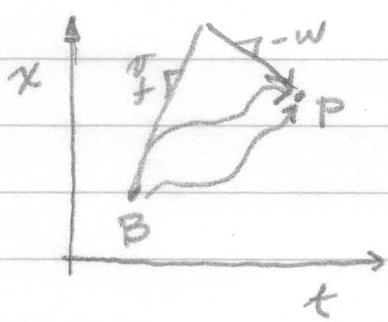


Figure 6





In the linear case shown in Fig. 6, all the valid paths between two points (B and P) have the same cost:

$$\Delta(\mathcal{P}) = \int_{t_B}^{t_P} r(\dot{x}(t))\,dt =$$

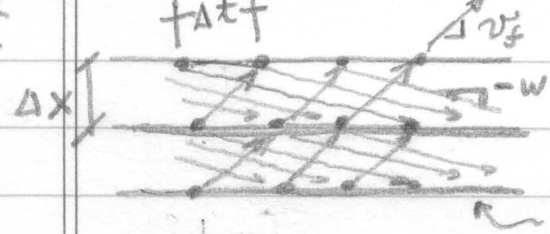
$$= q_{max}(t_P - t_B) - k_0(x_P - x_B).$$

Thus, to find a solution it suffices to consider "extremal paths" in which slopes are allowed to be either v_f or $-w$ as shown by the upper path of Fig. 6.

Computation

We can therefore construct "sufficient" networks made up of links with such slopes that suffice to evaluate the least possible costs between all their nodes. See Fig. 7.

Figure 7



on the forward links $r = 0$

on the backward links $r = r_m$

To solve a problem, place the nodes of such a network on the boundary and solve a shortest path problem for as many nodes of the network as possible*

* i.e. for nodes whose backward paths in the network intersect the boundary.

This can be done with shortest path algorithms, requiring only a few computations per node. The solution is exact even if the network steps Δt & Δx are long. Thus, we can place the observation lines at the beginning and the end of links without bothering about intermediate points. Newell's simplified KW solution is a special case of this framework.

Moving bottlenecks

In VT moving bottlenecks are easily modeled. Each bottleneck consists of its trajectory $x_B(t)$ and a maximum passing rate $r_B(t)$. (Presumably $r_B(t) < r(\dot{x}_B(t))$)

Since bottlenecks reduce the passing rate, they reduce the "cost rate" along their paths. They act as "shortcuts" in spacetime for the shortest path problem that determines N_p. Examples are presented in [9].

To solve a problem, simply overlay the moving bottleneck on the sufficient network; represent it by links with capacities determined from $r_B(t)$; and solve the shortest path on the expanded network. The level of complexity does not increase.

Figure 8

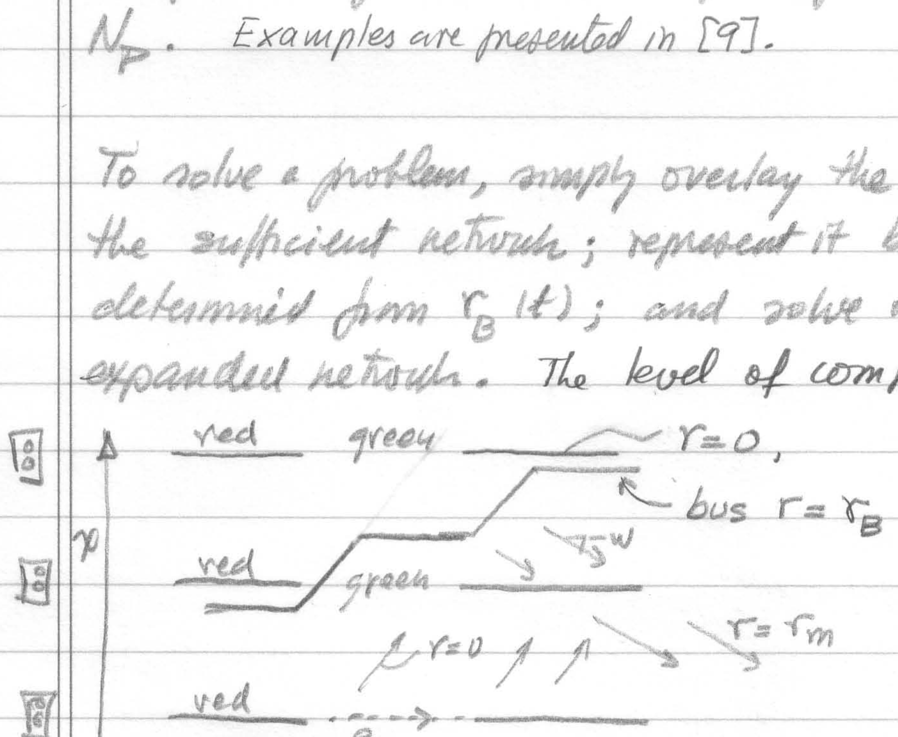


Figure 8 shows by means of solid lines the shortcuts that would be added for a bus that passes through a link containing 3 signals and one bus stop

Moving bottlenecks affect the traffic stream. But the traffic stream can also affect the moving bottlenecks. If the latter rules are understood from observation and theory, then it is possible to build models where the behavior of the bottlenecks is endogenous; e.g. models of mixed traffic with cars and buses; or models of multilane traffic where lane changers act as MB's, as in [10].

Capacity of arterials

These principles can be used to find the capacity of arterial streets with few turning movements as a function of the traffic offsets, cycles, green phase durations, number of lanes and intersection spacings -- even if these parameters change along the arterial.

The problem is a shortest path problem involving only 3 types of links (a, b, c). Links a and b have slopes $-w$ and v_f and cost rates r_m and 0. They can connect neighboring intersections at a cost of $r_m s_i$; which turns out to be the # of vehicles that can be stored in between intersections i and $i+1$ at jam density, K_i. Links c follow the red periods (as shown) and have zero cost rate.

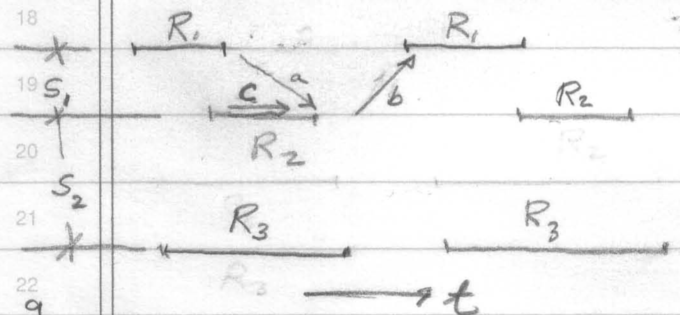
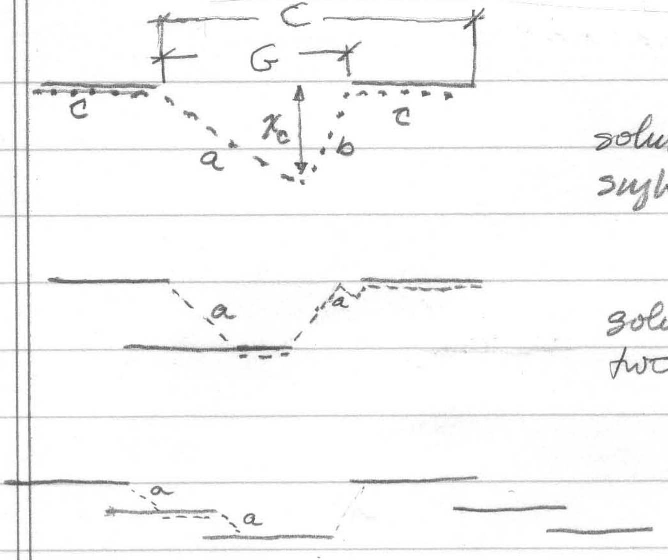


Figure 9

According to VT the arterial capacity is the smallest possible average cost rate of an ∞-ly long path with average slope zero. The path can be found numerically or graphically. For the latter, the game is finding a path that uses links "a" for the least time possible.

Figure 10



solution for single intersection (capacity $= q_{max}\, G/C$)

solution for two intersection (system capacity is less than the capacity of the individual intersection)

solution for 3 intersection (capacity reduced further)

Above we show solutions for 3 different problems. You can derive general expressions for the cases with 1 and 2 intersections and you will find that they match the expressions given in the literature. For example, you can readily see from the top that if intersections are separated by more than x_c (where $x_c(1/w + 1/v_f) = G \Rightarrow x_c k_j/q_m = G \Rightarrow x_c = Gq_m/k_0$) then there is no capacity reduction due to the 2nd intersection's proximity. (The formula $x_c = Gq_m/k_0$ is well known.)

So we see that VT allows us to derive the capacity of arterials in a convenient way.

MFD estimation

It can also help us define the steady state flows that can be observed when the total number of vehicles in the arterial is fixed (as if the arterial was a loop); i.e. the arterial's "macroscopic fundamental diagram", or MFD, linking its average flow to its density. And this can be done for any pattern of signal timings so that the MFD arising from engineering decisions can be predicted.

According to VT the MFD, $\bar{Q}(k)$, on such an arterial would exist.

To find it one would find the least cost path with average slope u, to find the long run maximum passing rate $\bar{r}(u)$. Thus we know that $\bar{Q}(k) \leq \bar{r}(u) + ku$, and call this inequality a "cut". (The earlier examples used the cut with $u=0$ to get the system capacity.)

If we now consider all the cuts we will have determined the function $\bar{r}(u)$. And we know from VT that flow is maximal across all possible paths with all possible slopes so:

$$\bar{Q}(k) = \max_{\forall u} \{ q : q \leq \bar{r}(u) + k u \}$$

(Compare with (A3).) An approximation from above can be obtained by evaluating just a few cuts. This is the approach used in [11] to find an approximate formula for the MFD for urban areas.

— · — · —

Duality

Let us now discuss a duality principle that can be used together with VT to unify several models of traffic flow.

In VT we solve for the 3-D surface described by the function $N(t,x)$. This function increases with t and declines with x. Therefore it admits an inverse $X(t,n)$ that increases with t and decreases with vehicle number, n. Both functions are equally valid descriptors of the 3-D surface. Could we solve for $X(t,n)$ instead? i.e. for the position of every vehicle (n) at every time (t)?

Two Different ways of casting the same problem

The answer is yes and here is how. For a more detailed description see reference [6] which introduces the ideas.

First note that the KW solution $k(t,x)$ that matches the VT solution is uniquely characterized by the properties:

(i) $q = Q(k)$ where the density is continuous

(ii) density can be discontinuous at curves (or "shocks")

(iii) $k(t_B, x_B) = k_B$ if B is a point on the boundary

(iv) solution is stable; i.e if k and k' are solutions with boundary data $\{k_B\}$ and $\{k'_B = k_B + \varepsilon\}$ where ε is a vehicle-conserving perturbation to the data, then the two solutions k and k' become equal as $t \to \infty$: $|k - k'| \to 0$ $\forall x$ as $t \to \infty$.

The stability condition means that small perturbations to the initial data are smoothed out as time goes on.

These properties can be rewritten in terms of the VT solution function $N(t,x)$, by remembering that the partial derivatives of N, N_x and N_t are $-k$ and q respectively:

(i) $N_t = Q(-N_x)$ where N_x and N_t are continuous

(ii) N is continuous but derivatives can be discontinuous

(iii) $N(t_B, x_B) = N_B$

(iv) $(N - N') \to 0$ as $t \to \infty$ for any vehicle conserving perturbation.

Property (i) is a non-linear PDE. Property (iii) are boundary conditions. Properties (ii) and (iv) rule out non-sensical solutions and ensure uniqueness of the shock placements. (This is all well known.)

Now we look for the equivalent conditions for function X. Note that

$$N_x = -k = -\frac{1}{\text{spacing}} = 1/X_n$$

$$N_t = q = kv = k X_t = -X_t/X_n$$

Thus, we have (regarding condition (i)):

$$(i) \quad -X_t / X_n = Q(-1/X_n)$$

$$X_t = -X_n\, Q(-1/X_n) \doteq V(-X_n)$$

Note that the last equality defines a new function "V" from the FD "Q". This transformation preserves concavity. It also preserves bilinearity as Fig. 11 shows. The transformation, D, is its own inverse so that

$$D \cdot Q = V \quad \text{and} \quad D \cdot V = Q$$

as is easy to verify. Thus, we can think of Q and V as primal and dual versions of the same thing.

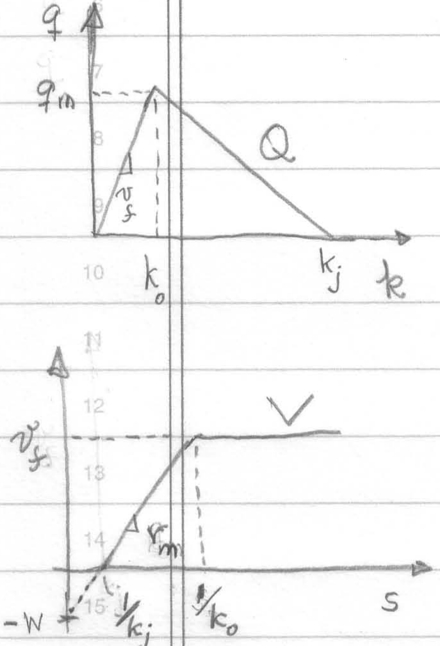


Figure 11

Now, since (ii)-(iv) are properties of the 3-D surface which is described by X as well as by N, it follows that X also satisfies (ii)-(iv).

Thus, X satisfies (i)-(iv) with Q replaced by V and the boundary conditions expressed as: $x_B = X(t_B, n_B)$;

This dual problem is of the same (Hamilton-Jacobi) type as the primal. It is also a KW problem like the primal, where the unknown "flow" and "density" have been replaced by "speed" and "spacing". The new dual problem is a continuum "car-following" problem of the exact same type as the original. The iso-cost contours for a given X are the N-curves of cumulative count at X on the queuing diagram. And the problem can be solved with VT.

We now apply this idea to the bilinear case. We will construct sulhrien networks that can be solved with dynamic programming. The DP formulae turn out to be special cases of different types of existing models. This shows that these existing models are intimately related.

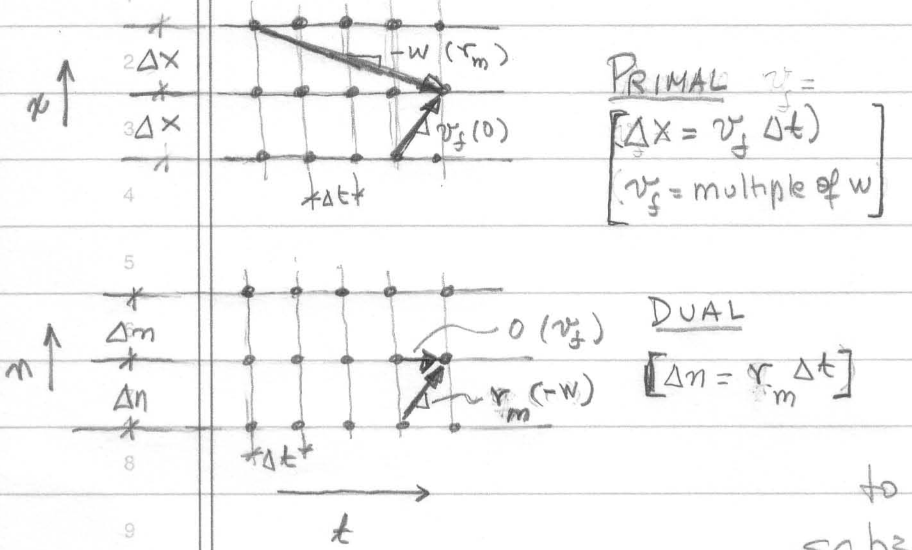


Figure 12

Figure 12 shows the "stencils" of links pointing to a generic link in the primal and dual problems corresponding to the FD of Figure 11. For the shown networks to be sufficient they must satisfy the conditions expressed in brackets, (so that links with extremal slopes start and end at nodes of the network.)

Equivalence of several models

The quantities by each link are the slopes of the link and (in parenthesis) its cost rate. The dynamic programming recursions are:

PRIMAL: $N(t,x) = \min\left\{N(t-\Delta t, x-\Delta x);\ N\left(t-\frac{v_f}{w}\Delta t, x+\Delta x\right) + r_m \frac{v_f}{w}\Delta t\right\}$

(This is Newell's formula in his simplified theory of KW.)

DUAL: $X(t,n) = \min\{X(t-\Delta t, n-\Delta n) - w\Delta t,\ X(t-\Delta t, n) + v_f\Delta t\}$

(This is the continuum car-following model of KW, CF(L), where cars can be handled in groups Δn as large as desired.[12] If we choose Δt so that $\Delta n = 1$ (i.e. $\Delta t = 1/r_m$) then $w\Delta t = w/r_m = 1/k_j = s_j$, the jam spacing, and we get Newell's "lower order" car-following model.[13]

Now, for the primal choose Δt so that $\kappa_m v_f/w \, \Delta t = 1$ and define $v_f/w \doteq m$ (an integer). Also, index the lattice with consecutive integers (j, i). Then, the primal relation becomes:

$$\text{PRIMAL:} \quad N_{ji} = \min\{N_{j-1,\, i-1} \, ; \, N_{j-m,\, i+1} + 1\}$$

Note that if all the data in the RHS are integer, the LHS is integer. Thus, if the initial data are integer the solution is integer. The recursion says that the vehicle on point $(j-1, i-1)$ can jump to point (j, i) (i.e advance one lattice spacing in one clock tick) iff the vehicle in front has been observed at least m clock ticks earlier at lattice point $i+1$ (i.e. so that it has cleared off the space at "i" with time to spare.). This is the cellular automata model CA(M) proposed in [12].

Finally, for the dual choose Δt so that $w\Delta t = 1$ and $v_f \Delta t = m$. And index the lattice with consecutive indices (j, k). Then we have:

$$\text{DUAL:} \quad X_{jk} = \min\{X_{j-1,\, k-1} - 1, \; X_{j-1,\, k} + m\}$$

Again this is integer arithmetic. So vehicles stay on a lattice. This states that vehicle k advances m lattice positions in one clock tick from $j-1$ to j iff the vehicle immediately before has vacated the target position before the jump. This is a special case of the cellular automata model in [14], called the CA(L) model in [12].

The above shows that several seemingly unrelated models are actually the same thing if one takes care of choosing their lattice spacing and their parameters in certain ways; and that all these models are then exact solutions of the ~~same~~ KW/VT problem.

— ∘ — • —

Closure

In the explanation so far we assumed that the FD was time- and space-independent. The connection between VT and KWT is preserved, however, if Q depends on time and space so that $q = Q(k, t, x)$. The only thing that changes is that the cost function of VT: $r(u, t, x) = \mathcal{L} \cdot Q(k, t, x)$ (i.e. the Legendre transform of Q with respects to k) is also time and space dependent. Therefore, parallel links of equal length will no longer necessarily have the same cost. The solution, however, is still a shortest path problem which is no more difficult to solve. Reference [9] shows how this is done, uses examples, and shows how VT improves on other methods to solve this type of problem. This is an advantage of VT.

There is one caveat, however. One must be careful when formulating a problem not to stipulate conditions that lead to impossibilities; e.g. stipulating that the road's jam density drops to zero if the road is not empty of vehicles. Conditions to ensure that a VT problem is well posed have been laid out in [4] and [6]. In practice, however, it suffices to verify that the statement of a problem "makes physical sense".

* Aside

Properties of LT

Note we can redefine $r(u)$ by means of an optimization for the lowest possible intercept:

$$r(u) = \min\{r : r + uk \geq Q(k)\} \qquad (A1)$$

or

$$-r(u) = \max\{r : r - uk \leq -Q(k)\} \qquad (A2)$$

Equation (A2) is an operation that changes Q into r; call it $\mathcal{L}$ and write $\mathcal{L} \bullet (-Q) = (-r)$

From (A1) and Fig. 2 (top) we see that $Q(k)$ is the highest flow beneath all the tangents. (True because Q is concave.) Thus,

$$Q(k) = \max\{q : q \leq r(u) + ku\} \qquad (A3)$$

$$= \max\{q : q - ku \leq r(u)\} \qquad (A4)$$

In other words $Q = \mathcal{L} \cdot r$ for concave Q. This shows that a similar operation can be used to obtain Q from r. For each k, $Q(k)$ is the intercept of the tangent to $r(u)$ with slope $-k$.

Interpretation of (A3): Since $r(u)$ is the maximum passing rate of an observer, and the term "ku" converts this rate into "absolute" flow (i.e. relative to the road), the RHS of (A3) is an upper bound on flow allowed by an observer traveling at speed u.

Eq. (A3) merely states that the flow of the FD is the highest possible that does not violate any of the capacity constraints imposed by observers traveling with constant speed. Graphically, the FD is the lower envelope of the family of lines $\{q = r(u) + ku\}$, as implied by (A3).

APPENDIX B

Background: A review of kinematic wave theory* (B1

The problem

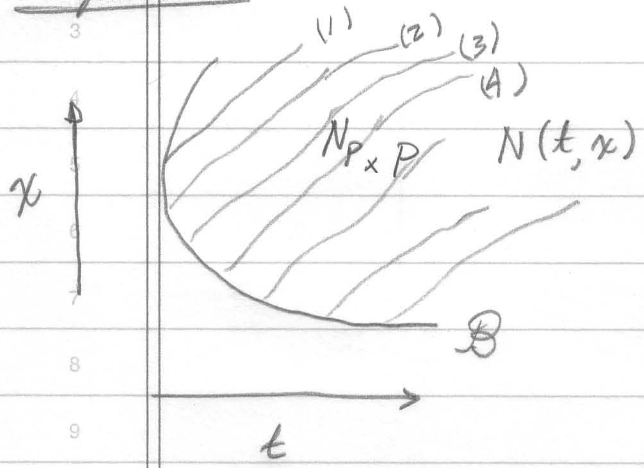


Figure B1.

The traffic forecasting problem consists in determining all the vehicle trajectories in space-time, giving the vehicle positions along some boundary curve (or curves) $\mathcal{B}$. See Figure B1.

In other words determining the vehicle number N_P observed at every point P in the solution domain. The collection of N_P's should define a continuous surface $N(t,x)$, if we treat vehicles like a fluid. The contours of this surface are the vehicle trajectories. The slopes of the surface are related to the vehicle flow and density. For vehicles going in the direction of increasing x (as shown in Figure B1) the relation is:

$$q(t,x) = \frac{\partial N}{\partial t} \; ; \; k(t,x) = -\frac{\partial N}{\partial x}. \qquad (B1)$$

The KWT formulation

In kinematic wave theory we look for $k(t,x)$ instead of $N(t,x)$. (This is the same for practical purposes since one can get N from k -- up to a constant -- by integrating.) So given is the density k_B along all points of the boundary, $B \in \mathcal{B}$. The theory has 3 postulates:

1. Vehicle conservation (N is continuous) expressed as:

$$\partial^2 N/_{\partial x \partial t} = \partial^2 N/_{\partial t \partial x} \Leftrightarrow \frac{\partial q}{\partial x} = -\frac{\partial k}{\partial t} \Leftrightarrow \frac{\partial q}{\partial x} + \frac{\partial k}{\partial t} = 0$$

where derivatives exist. (B2)

* Also known as LWR theory after proposers: Lighthill and Whitham; and Richards.

Prepared by:

Date:

(B2

2. Equation of state (Fundamental Diagram, FD):

At every point flow is uniquely determined from the density, the location x and the time t.

$$q(t,x) = Q(k(t,x), x, t). \qquad (B3)$$

If the highway is homogeneous and conditions time-independent then flow only depends on density.

$$q = Q(k). \qquad \text{(the Fundamental Diagram)}$$

Solutions consistent with 1 and 2 and with the boundary data usually require that we introduce curves called "shocks" where the derivatives of N (i.e., k and q) are discontinuous -- although N itself must remain continuous -- so that along these curves (B2) and (B3) are not meaningful and not required to hold. Because there usually are multiple ways in which shocks can be introduced to produce a valid solution, the theory includes a third condition that rules out undesirable solutions.

3. Stability (decay of small perturbations with time.)

If you take two solutions k and k' with similar but not identical initial data, then the maximum difference between the solutions at any given time, declines with time:

$$\max_x |k(t,x) - k'(t,x)| \quad \text{declines with time.}$$

It is known [5] that conditions 1, 2, 3 produce unique solutions to the initial value problem (where $\mathcal{B}$ is the line $t=0$) if the initial density satisfies some mild regularity conditions that are always met for practical problems.

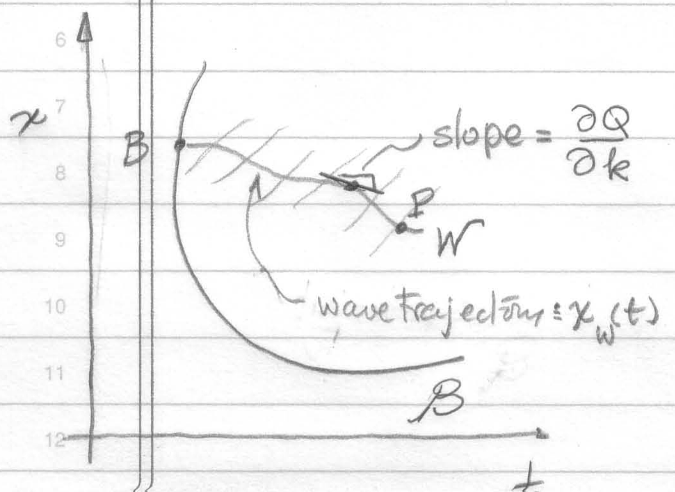
Prepared by:
Date:

(B3)

Solution with waves

If you replace q in (P2) by the RHS of (P3) (to eliminate q as an unknown) we find that:

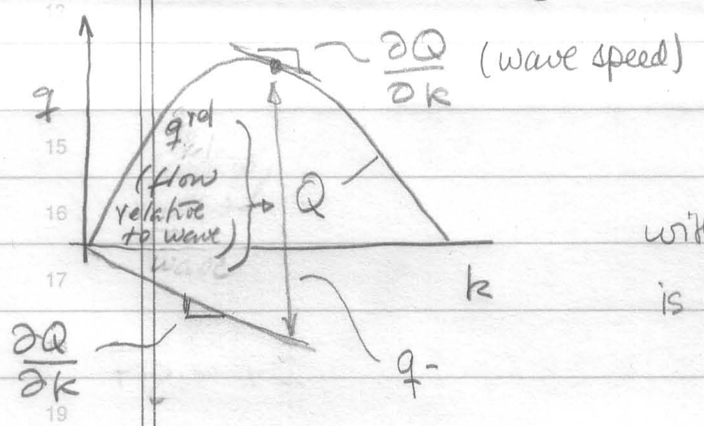
$$\left(\frac{\partial Q}{\partial k}\right)\cdot\frac{\partial k}{\partial x}+\frac{\partial k}{\partial t}=-\left(\frac{\partial Q}{\partial x}\right). \qquad (B4)$$



Consider a curve W (called a "wave" or "characteristic") emanating from $B \in \mathcal{B}$ as shown in the figure with the property that everywhere along the curve:

$$\frac{dx_w(t)}{dt} \doteq \dot{x}_w(t)=\frac{\partial Q}{\partial k}. \qquad (B5)$$

The slope of the curve matches the slope of the FD for the prevailing density. The reason for defining waves like this is that then the rate at which k increases with time along the wave (denoted dk_w/dt) is the LHS of (B4):



$$\frac{dk_w(t)}{dt}=-\frac{\partial Q}{\partial x} \qquad (B6)$$

Equations (B5) and (B6) are a pair of ODE's that can be solved by stepping through time, starting from any point on the boundary such as "B". Knowledge of $x_w(t)$ and $k_w(t)$ serves to identify N_P from N_B. The recipe uses the formula for flow seen by a moving observer:

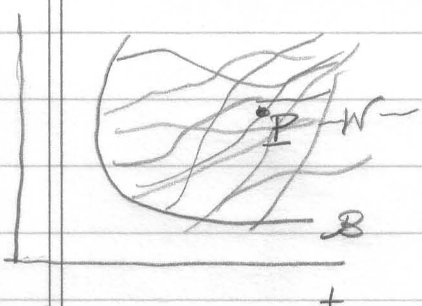
$$N_P \doteq N_B+\begin{bmatrix}\#\text{vehs.}\\ \text{passing}\\ \text{wave}\end{bmatrix}=N_B+\int_{t_B}^{t_P} q^{rel}(t)\,dt = N_B+\int_{t_B}^{t_P}\left[Q(k_w,x_w,t)-\dot{x}_w k_w\right]dt \qquad (B7)$$

(The subscripts of "t" refer to the points to which they apply.)

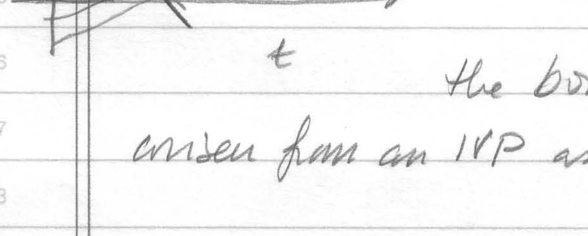
Prepared by:
Date:

(B4)

To obtain a complete solution we must evaluate (B5-B7) for every point such as B along the boundary. If you think of B as a scalp and W as a hair growing from the scalp, one must evaluate a complete hair mane. In real problems this mane will have hair tangles (see Figure) and at tangled locations where



there are multiple hairs there is a multiple choice problem for N_p. Likewise there could be locations without hairs and at these locations a prediction cannot be made. Fortunately it turns out (see [5]) that if Q is concave at every x (and independent of t) and if the initial density satisfies some mild conditions, then every IVP has a unique solution that satisfies the 3 KWT postulates and this solution can be obtained by giving our "mane" a unique haircut that (i) will remove all tangles (i.e all wave crossings) and (ii) fill the solution domain. In other words waves are truncated and every point that is not at the end of a wave is visited by only 1 (truncated) wave. Shocks are the loci containing the ends of all the waves. See figure.



This property also holds for well-posed problems with more general boundary conditions; i.e. with curvy boundaries (as in the figure) or multiple boundaries, provided the data along the boundaries is well-posed (i.e. could have arisen from an IVP as illustrated by boundary B in the figure.)

Prepared by:
Date:

(B5

Although it is in principle possible to find the "mane", finding the relevant haircut is difficult. For homogeneous highways, where $\frac{\partial Q}{\partial x} = 0$, (B6) implies that k_w is constant along each wave. And since k is constant, then (B5) implies that waves are straight lines. Thus, for homogeneous highways growing the "mane" is easy. Clipping was still difficult.

Newell simplified theory [1]

In the early 1990's, Newell [1] proposed that the relevant solution was obtained by using a minimum rule. View the waves as values of N_w and x for different t's, i.e. as curves in 3-D space of N, x and t : $\{N_w(t), x_w(t)\}$. (The family of waves is a ruled surface for the homogeneous problem since all the 3-D waves are straight lines.) Surfaces of this type can intersect and have multiple layers. Newell proposed that one always had to choose the layer with the lowest N. In other words, where a point (t, x) is reached by multiple waves, the relevant wave is the one with the lowest N.

Newell's method was based on physical and geometric intuition. It produces the solution we desire with continuous $N(t,x)$ and non-crossing waves. Newell showed [1] how the procedure can be streamlined very elegantly to solve problems with piece-wise homogeneous freeways.

However, the procedure is still cumbersome for inhomogeneous problems, and for problems involving moving bottlenecks.

R1

REFERENCES

1. Newell, G.F. (1993) A simplified theory of kinematic waves: (i) general theory; (ii) queuing at freeway bottlenecks; (iii) multi-destination flows. *Transportation Research Part B* 37, 281-313.

2. Newell, G. F. (1961). Nonlinear Effects in the Dynamics of Car Following. *Operations Research* 9(2), 209-229.

3. Daganzo, C. F. (2003). A variational formulation for class of first order PDE's. Research Report: UCB-ITS-RR-2003-4, Univ. of California, Berkeley.

4. Daganzo, C.F. (2005). A variational formulation of kinematic waves: Basic theory and complex boundary conditions. *Transportation Research Part B* 39(2), 187-196.

5. Lax, P.D. (1973) *Hyperbolic systems of conservation laws and the mathematical theory of shockwaves.* SIAM Regional Conference Series in Applied Mathematics, J.W. Arrowsmith Ltd., Bristol, U.K.

6. Daganzo, C.F. (2006). On the variational theory of traffic flow: well-posedness, duality and applications. Working Paper: UCB-ITS-VWP-2006-2, Univ. of California, Berkeley; reprinted in *Networks and Heterogeneous Media* 1(4) 601-619 (2006)

7. Daganzo, C. F. (2003). A variational formulation of kinematic waves: solution methods: Research Report: UCB-ITS-RR-2003-7, Univ. of California, Berkeley.

8. Daganzo, C. F. (2003). A variational formulation of kinematic waves: solution methods. Research Report: UCB-ITS-RR-2003-7, Univ. of California, Berkeley; reprinted in *Transportation Research Part B* 39(10), 934-950 (2005).

9. Daganzo, C. F. and Menendez, M. (2005). A variational formulation of kinematic waves: bottleneck properties and examples. in *Transportation and Traffic Theory*, Proc. 16th Int. Symposium on Transportation and Traffic Theory (H. Mahmassani, editor) pp.345-364, Elsevier, New York, N.Y., July 2005

10. Laval, J. and Daganzo, C.F. (2006). Lane changing in traffic streams. *Transportation Research Part B* 40(3), 251-264.

11. Daganzo, C.F. and Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B* 42(9), 771-781

12. Daganzo, C.F. (2004) In traffic flow, cellular automata = kinematic waves. Research Report: UCB-ITS-RR-2004-5, Univ. of California, Berkeley; reprinted in *Transportation Research Part B* 40 (2006) 396–403.

13. Newell, G.F. (2002) A simplified car-following theory: a lower order model. *Transportation Research Part B* 36, 195-205.

14. Nagel, K. and Schreckenberg, M. (1992). A cellular automata model for freeway traffic. *J. de Physique I* 2, 2221-22229.