Markovian Traffic Equilibrium

Università di Pisa
Teoria e Metodi dell'Ottimizzazione

Nikita Deniskin

12 Ottobre 2021

Indice

1 Introduzione

In questo documento analizziamo un modello per lo studio di reti trafficate (*Traffic Networks*), proposto nell'articolo [1] da Baillon e Cominetti, e successivamente rielaborato per un libro [2, capitolo 3] da Cominetti.

Le reti trafficate vengono utilizzate come modelli per analizzare lo studio di reti presenti nel mondo reale, per esempio una rete stradale, di trasporti, oppure di collegamento tra computer. È quindi importante che il modello e i risultati ottenuti siano il più possibile verosimili.

Nel capitolo 2 vedremo le prime definizioni, e il modello classico per le reti trafficate, che è stato proposto da Wardrop nel 1952. Questo modello presenta dei difetti: la scelta del percorso viene effettuata tra tutti i percorsi possibili, e questo comporta un grandissimo dispendio computazionale.

Nel capitolo 3, vedremo il modello proposto nell'articolo: equilibrio di traffico Markoviano (MTE). A differenza di quello classico, questo è un modello non deterministico. Ogni arco ha un tempo di percorrenza non fisso, ma rappresentato da una variabile aleatoria; in questo modo possiamo incorporare piccole fluttuazioni, o percezioni personali degli utenti per il costo di un tragitto. Inoltre, la scelta del percorso avviene a livello locale: gli utenti non scelgono il percorso tra tutti i percorsi, ma ad ogni nodo seguono l'arco più conveniente. La scelta non è univocamente determinata (come i modelli "All-or-Nothing"), ma ad ogni arco uscente viene assegnata la probabilità che seguendo quell'arco, si arrivi col percorso più veloce. Gli utenti dunque si muoveranno seguendo una catena di Markov.

Nel capitolo 4 vedremo condizioni per assicurarsi coerenza del nostro modello, e garantire l'esistenza e unicità delle soluzioni del MTE.

Infine nel capitolo 5 vedremo l'approccio computazionale implementato nell'algoritmo, con opportuna scelta delle funzioni parametriche e della distribuzione delle variabili aleatorie. Gli algoritmi usati sono sostanzialmente dei metodi del gradiente, leggermente modificati per aumentare la velocità di convergenza (dal punto di vista sperimentale). Per questo algoritmo modificato, MSA, vedremo una dimostrazione teorica della convergenza nella sezione 5.5.
In molte delle proposizioni che dimostreremo, tra parentesi quadre è indicato il riferimento e il nome di tale proposizione o teorema come esposta in [2].

2 Rete trafficata

2.1 Definizioni

Una rete trafficata è definita da un grafo diretto $\mathcal{G} = (V, A)$, che ha $n = |V|$ vertici e $m = |A|$ archi. Ci sono degli utenti che si muovono lungo gli archi di questa rete; ciascuno ha un vertice di origine e uno di destinazione. La quantità di utenti dal vertice i al vertice d è rappresentata da g_i^d.

La rete può essere un modello per molti fenomeni del mondo reale, per cui tutti i flussi presenti (quelli attraverso un arco, oppure entranti o uscenti in un vertice) sono rappresentati da un numero reale non negativo. Per fissare la terminologia, supponiamo che il grafo rappresenti una rete stradale.

Se sull'arco a c'è un flusso w_a di utenti, si ha traffico e serve più tempo per percorrerlo. Per questa è introdotta la funzione di *congestione* $s_a : \mathbb{R} \to \mathbb{R}^+$, ovvero $t_a = s_a(w_a)$ è il tempo necessario per percorrere l'arco a se su questo c'è un traffico di w_a utenti. Richiediamo che s_a sia crescente, coerentemente con l'intuizione fisica: più traffico c'è, più si muove lentamente, dunque aumenta il tempo di percorrenza. È presente il pedice s_a perché il risultato può dipendere da parametri del collegamento (per esempio, quanto è larga la strada), ma il modello fisico è naturalmente lo stesso per tutti gli archi. Osserviamo che il traffico su ciascun arco è ≥ 0, quindi sembra che basterebbe definire la funzione con dominio $[0, \infty)$. È cosi, però per la trattazione che faremo ci è comodo supporre che sia definita su tutto $\mathbb{R}$ (ma non importa come sia su $(-\infty, 0)$, basta che sia crescente).

Definiamo A_i^+ come l'insieme degli archi uscenti da i e A_i^- quelli entranti. Se $a \in A_i^+$, allora $a = (i, *)$ e se $a \in A_i^-$, allora $a = (*, i)$.
Un'ipotesi che assumeremo sempre in questa trattazione è che il grafo sia fortemente connesso, ovvero esista un cammino tra qualsiasi coppia di vertici i e d.

Definizione 1. Dato un cammino $r = (a_1, a_2, \ldots, a_k)$, con gli archi concatenati, il suo costo è:

$$c_r = \sum_{a \in r} t_a = \sum_{a \in r} s_a(w_a)$$

Sia h_r il numero di utenti che segue il cammino r. Per un arco a passano tanti cammini; il flusso totale su a è la somma dei flussi di tutti i cammini che passano per esso:

$$w_a = \sum_{r \ni a} h_r$$

Supponiamo ora di fissare i flussi w_a e h_r su ogni arco e cammino. Preso un vertice i, l numero di persone che arriva in quel vertice è uguale al numero di persone che hanno destinazione i sommato al numero di persone che proseguono su altri archi, tolte le persone che hanno partono da i. La relazione è la seguente:

$$\sum_{a \in A_i^-} w_a = \sum_{k \in V} g_k^i + \sum_{a \in A_i^+} w_a - \sum_{j \in V} g_i^j \tag{1}$$

Si definisca v_a^d il flusso di utenti attraverso l'arco $a = (i, *)$ che hanno destinazione finale d, allora il flusso totale sull'arco è:

$$w_a = \sum_{d \neq i} v_a^d$$

L'equazione di conservazione di flusso 1, scritta nel vertice i per la destinazione $d \neq i$, diventa:

$$g_i^d + \sum_{a \in A_i^-} v_a^d = \sum_{a \in A_i^+} v_a^d, \qquad d \neq i \tag{2}$$

Sempre fissando tutti i flussi, definiamo il tempo minimo di viaggio τ_i^d, il costo del cammino ottimale tra i due vertici.

$$\tau_i^d = \min_{r \in R_i^d} c_r$$

Ci sono quindi dei cammini r da i a d che hanno costo τ_i^d, e data una generica distribuzione di flusso, è molto probabile che ci siano altri cammini con costo maggiore. Ogni utente che segue una strada più costosa si rende conto di viaggiare in modo poco ottimale, per cui cercherà di usare il percorso a costo minore. Tuttavia, cosi facendo diminuisce il flusso attraverso il suo vecchio cammino, e aumenta il flusso attraverso il cammino più conveniente. Per cui un costo diminuisce, e un altro aumenta.
Visto che lo fanno tutti gli utenti, che hanno origine e destinazioni diverse, un algoritmo "greedy" di rerouting potrebbe benissimo non convergere all'equilibrio, o comunque non migliorare la situazione.

Osserviamo inoltre che dai h_r è possibile stabilire i w_a (con la formula di sopra, è la somma). Tuttavia sapendo i w_a, gli h_r non sono univocamente determinati. Però è poco importante, perché il costo dipende solo da ciascun $t_a = s_a(w_a)$.
Un'altra osservazione è che i cammini r sono potenzialmente infiniti, ma per come è formulato il problema possiamo escludere quelli che passano per più di due volte da uno stesso vertice (sicuramente non sono ottimali); la ricerca sarà quindi limitata a un numero finito di cammini.

2.2 Modello di Wardrop

Definizione 2 (Equilibrio di Wardrop)**.** Fissati i flussi sugli archi w e sui cammini h, la coppia (w, h) si dice un equilibrio di Wardrop se per ogni cammino r che sia effettivamente percorso, ovvero $h_r > 0$, esso è un cammino minimo:

$$h_r > 0 \Longrightarrow c_r = \tau_i^d = \min_{\tilde{r} \in R_i^d} c_{\tilde{r}}$$

Definizione 3. Sja H l'insieme delle coppie $(h, w) = (h, w(h))$ che soddisfano le seguenti condizioni:

1. Conservazione del flusso $g_i^d = \sum_{r \in R_i^d} h_r$, con $h_r \geq 0$.
2. Traffico sull'arco $w_a = \sum_{r \ni a} h_r$.

Proposizione 1 (Theorem 1.2)**.** Una coppia $(w, h) \in H$ è un equilibrio di Wardrop se e solo se è soluzione del seguente problema di minimizzazione:

$$F(h) = F(h, w(h)) = \sum_{a \in A} \int_0^{w_a} s_a(z)\ \mathrm{d}z$$

$$\min_{(w,h) \in H} F(h) = \min_{(w,h) \in H} \sum_{a \in A} \int_0^{w_a(h)} s_a(z)\ \mathrm{d}z \tag{3}$$

Dimostrazione. Calcoliamo le derivate parziali di $F(h)$:

$$\frac{\partial}{\partial h_r} F(h) = \sum_{a \in A} \frac{\partial}{\partial h_r} \int_0^{w_a(h)} s_a(z)\, \mathrm{d}z = \sum_{a \in A} \frac{\partial w_a}{\partial h_r}(h)\, s_a(w_a(h))$$

Ma $w_a = \sum_{r \ni a} h_r$, per cui

$$\frac{\partial w_a}{\partial h_r}(h) = \begin{cases} 1 & \text{se } r \ni a \\ 0 & \text{se } r \not\ni a \end{cases} \tag{4}$$

Per cui

$$\frac{\partial}{\partial h_r} F(h) = \sum_{a \in r} s_a(w_a(h))$$

$F(h)$ **è una funzione convessa.** Chiamando $F_a(h) = \int_0^{w_a(h)} s_a(z)\, \mathrm{d}z$, vogliamo dimostrare che F_a è convessa. Questo implica che $F = \sum F_a$ è convessa.
$F_a(h)$ è costante sui piani dove è costante $\sum_{r \ni a} h_r$, perché $w_a(h) = \sum_{r \ni a} h_r$. Quindi attraverso una rotazione dello spazio (e cambio di variabili), si può scrivere $F_a(h) = F_a(x, y_1, \ldots y_{L-1}) = F_a(x)$, dove $x = \sum_{r \ni a} h_r$ rappresenta la coordinata sulla retta determinata dalle equazioni $\{h_1 = h_2 = \ldots = h_L\}$, mentre le altre coordiante y_i sono un qualsiasi complemento ortogonale.
Se F_a è convessa nella sola variabile x, lo è automaticamente nelle variabili $(x, y_1, \ldots, y_{L-1})$. Ma $F_a(x) = \int_0^x s_a(z)\, \mathrm{d}z$, $F_a'(x) = s_a(x)$ è una funzione crescente, dunque $F_a(x)$ è convessa e quindi anche $F(h)$.

Supponiamo che la coppia (w, h) sia il punto di minimo del problema. Scrivendo la relazione del sottogradiente, ovvero $F(h') - F(h) \geq \nabla F(h) \cdot (h' - h)$, si ha:

$$\nabla F(h) \cdot (h' - h) = \sum_r \frac{\partial F}{\partial h_r}(h)\, (h_r' - h_r) = \sum_r \left(\sum_{a \in r} s_a(w_a(h))\, (h_r' - h_r) \right) =$$

$$= \sum_r (h_r' - h_r) \left(\sum_{a \in r} s_a(w_a(h)) \right) = \sum_r (h_r' - h_r)\, c_r =$$

$$= \sum_{(i,d)} \left(\sum_{r \in R_i^d} c_r (h_r' - h_r) \right) = \sum_{(i,d)} \left(\sum_{r \in R_i^d} (c_r - \tau_i^d)(h_r' - h_r) \right) + \left(\sum_{r \in R_i^d} \tau_i^d (h_r' - h_r) \right)$$

Nel secondo termine si porta τ_i^d fuori e resta $\sum_{r \in R_i^d} h_r' = g_i^d = \sum_{r \in R_i^d} h_r$ per la conservazione del flusso, per cui segue che $\tau_i^d \left(\sum_{r \in R_i^d} (h_r' - h_r) \right) = 0$.

Per il primo termine, se si ha un r con $h_r > 0$, se si avesse $c_r > \tau_i^d$ allora si potrebbe scegliere $h_r' < h_r$ (e gli altri h aumentati in modo da mantenere costante la somma), che porta ad avere una direzione di decrescita per $F(h)$. Visto che sarebbe decrescente al primo ordine, in un intorno abbastanza piccolo si avrebbe un punto con $F(h') < F(h)$, assurdo. $\square$

Proposizione 2 (Corollary 1.3)**.** Esiste un equilbrio di Wardrop (w, h). Inoltre w è univocamente determinato: tutti gli equilibri, hanno sempre lo stesso w.

Dimostrazione L'insieme H è compatto. Infatti la condizione 1 implica la limitatezza di h, e di conseguenza anche di w per la condizione 2. Inoltre, entrambe le condizioni sono chiuse, da cui segue la compattezza di H. Dunque il problema di minimo (eq 3) ha una soluzione.
Guardiamo $F(w, h) = F(w)$ come funzione solo di w, per opportuni h; mostriamo che questa è strettamente convessa. Come visto nella proposizione 1, ogni $F_a(w)$ è strettamente convessa nella sola variabile w_a, ed è indipendente dalle altre w_c per $c \neq a$. Presi due punti distinti w^1 e w^2 e una loro combinazione convessa $w^\gamma = \gamma w^1 + (1-\gamma)w^2$, dall'equazione $F(w^1) = \sum_a F_a(w_a^1)$ e le sue analoghe, si ha:

$$\gamma F(w^1) + (1-\gamma)F(w^2) = \gamma \sum_a F_a(w_a^1) + (1-\gamma) \sum_a F_a(w_a^2) =$$

$$= \sum_a \gamma F_a(w_a^1) + (1-\gamma)F_a(w_a^2) \geq \sum_a F_a(w_a^\gamma)$$

Siccome w^1 e w^2 sono distinti, esiste un arco b tale che $w_b^1 \neq w_b^2$. Dunque $\gamma F_b(w_b^1) + (1-\gamma)F_b(w_b^2) > F_b(w_b^\gamma)$ e la disuguaglianza sopra è stretta. Questo prova che $F(w)$ è strettamente convessa, dunque il w che realizza il minimo è unico. $\square$

Denotiamo con w^* la soluzione trovata. Osserviamo però che h non è univocamente determinato a partire da w^*, quindi la coppia di equilibrio (w, h) potrebbe essere non unica. Questo non è un problema, perché per il traffico conta quanto traffico passa su ciascun arco, non come ciascun singolo utente scelga che strada percorrere.

La formulazione di Wardrop presenta delle scomodità dal punto di vista computazionale: il numero di cammini h_r cresce molto rapidamente all'aumentare della lunghezza di r. Diventa complicato calcolarli tutti, e il numero di variabili per $F(h)$ diventa molto presto insostenibile.
Per questo motivo, una prima modifica che si fa è di usare i flussi v_a^d, ovvero il numero di utenti attraverso a con destinazione d. Il numero di variabili è ora limitato superiormente da mn, e il minimo si può cercare al variare di (w, v) anziché (w, h).
Un'altro passo che si può fare è quello di passare al problema duale. Le variabili duali sono i tempi di percorrenza degli archi t_a e i tempi per arrivare a destinazione τ_i^d. Enunciamo, senza dimostrazione, la formulazione duale del problema di minimizzazione (cioé dell'equazione 3):

Proposizione 3 (Theorem 1.4)**.** Sia T l'insieme delle coppie (t, τ) che soddisfa

$$\tau_i^d \leq t_a + \tau_j^d, \qquad \forall d, \forall a \in A \text{ con } a = (i, j).$$

Sia $t_a^0 = s_a(0)$. Allora il seguente problema di minimizzazione ammette una soluzione:

$$\min_{(t,\tau)\in T} \sum_{a\in A} \int_{t_a^0}^{t_a} s_a^{-1}(y)\ \mathrm{d}y - \sum_{d\neq i} g_i^d \tau_i^d$$

Ogni soluzione (t^*, τ^*), soddisfa $t_a^* = s_a(w_a^*)$ dove w^* rappresenta i flussi trovati risolvendo il problema di minimizzazione originale.

Più avanti vedremo una generalizzazione di questo risultato, in cui i tempi di percorrenza degli archi (e di conseguenza tutte le altre quantità) sono variabili aleatorie. Nel caso in cui tutte le variabili abbiano varianza nulla, si ottiene la proposizione 3.

3 Markovian Traffic Equlibrium

3.1 Modello non deterministico

L'approccio di Wardrop presente alcuni difetti. Il numero di variabili è il numero di percorsi possibili, che cresce esponenzialmente con la dimensione del grafo. Diventa difficile gestire cosi tante variabili, e soprattutto ottimizzare e fare ricerca di minimi. Inoltre si è visto che modelli deterministici spesso danno risultati molto lontani dalla distribuzione di traffico del mondo reale; se si inserisce una componente stocastica, si hanno risultati migliori. Il ruolo della parte aleatoria è sia di incorporare piccole fluttuazioni sui tempi di percorrenza, a causa di svariati fattori, sia di distinguere preferenze personali e valutazioni diverse dei costi a seconda dell'utente che viaggia.

Per questi motivi, formuliamo un modello Markoviano di traffico. In questo ogni utente non sceglie già dall'inizio il percorso tra tutti i possibili, ma ad ogni nodo sceglie la strada che probabilmente sarà la più corta. Dunque per ogni destinazione formeremo una catena di Markov, seconda la quale gli utenti si muoveranno. Le probabilità di transizione, cosi come i flussi sugli archi e i tempi totali saranno tutti legati tra di loro, e formuleremo il problema come la ricerca di un minimo di un'opportuna funzine convessa.

Supponiamo che i tempi di percorrenza degli archi siano variabili aleatorie $\tilde{t}_a = t_a + \nu_a$, con $t_a = \mathbb{E}[\tilde{t_a}]$. Dato un cammino r, il suo costo è anch'esso una variabile aleatoria $\tilde{c}_r$, dato dalla somma dei costi degli archi che lo compongono.

$$\tilde{c}_r = \sum_{a\in r} \tilde{t}_a$$

Si definiscano le variabili aleatorie τ_i^d per ogni i, d che indicano il tempo per arrivare a d partendo da i.

$$\tilde{\tau}_i^d = \min_{r\in R_i^d} \tilde{c}_r$$

Da queste si ottengono le variabili $\tilde{z}_a^d = \tilde{t}_a + \tilde{\tau}_j^d$ per l'arco $a = (i, j)$. Per un utente che si trova nel vertice i, la variabile $\tilde{z}_a^d$ indica il tempo per raggiungere la destinazione se sceglie di percorrere l'arco a.

Fissata la destinazione d, per ogni vertice i guardo le $\tilde{z}_a^d$ uscenti da a. Queste sono tutte variabili aleatorie, quindi il loro costo può variare. Per ogni arco $a = (i, j)$ uscente da i, guardo la probabilità che $\tilde{z}_a^d$ sia più piccola rispetto a tutti gli altri archi b uscenti da a. Chiamo questa probabilità

$$P_{ij}^d = \begin{cases} \mathbb{P}(\tilde{z}_a^d \leq \tilde{z}_b^d \text{ per ogni arco } b \text{ uscente da } i\ ) & \text{se esiste l'arco } a = (i, j) \\ 0 & \text{se non esiste l'arco tra } i \text{ e } j \end{cases}$$

Inoltre, se $i = d$, vuol dire che l'utente ha già raggiunto la destinazione e non ha bisogno di muoversi oltre.

$$P_{dj}^d = \begin{cases} 1 & \text{se } j = d \\ 0 & \text{se } j \neq d \end{cases}$$

Creiamo una catena di Markov sul grafo: ogni utente, trovandosi sul nodo i e avente destinazione finale d, decide con probabilità P_{ij}^d di andare sull'arco $a = (i, j)$. Visto che abbiamo un grande flusso di utenti, possiamo assumere che il flusso entrante in i (ovvero x_i^d) si divida negli archi uscenti con proporzione P_{ij}^d.
Dunque in generale ci saranno n catene di Markov, una per ciascuna destinazione d. Per ognuna, d è uno stato assorbente. Le probabilità di transizione saranno generalmente molto diverse per ciascuna catena, per cui ci saranno n matrici di transizione P^d.

3.2 Minimo di variabili aleatorie

Vediamo ora come calcolare le probabilità P_{ij}^d. Per farlo, risolviamo il seguente problema più generale.

Problema. Supponiamo di avere una persona che vuole acquistare un oggetto tra n possibili scelte. Il costo di ciascun oggetto è rappresentata da una variabile aleatoria $\tilde{x}_i$, con $1 \leq i \leq n$. Per semplicità, si scomponga la variabile aleatoria nel seguente modo: $\tilde{x}_i = x_i + \varepsilon_i$, dove $x_i = \mathbb{E}[\tilde{x}_i]$ è la media (dunque un numero reale), mentre ε_i è il termine aleatorio. La persona osserva i costi e, per in ciascun possibile evento, sceglie l'alternativa con il costo minore. Il prezzo pagato p è dunque una variabile aleatoria

$$p = \min\{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n\} = \{x_1 + \varepsilon_1, x_2 + \varepsilon_2, \ldots x_n + \varepsilon_n\}$$

Supponiamo che di fissare le variabili aleatorie ε_i e di far variare le x_i. Il prezzo $p = p(x)$ è una variabile aleatoria che dipende dai x_i.

Definizione 4. Usando le quantità introdotte sopra, definiamo la funzione $\varphi(x) : \mathbb{R}^n \to \mathbb{R}$ come il valore atteso del prezzo, per ciascuna scelta di x

$$\varphi(x) = \mathbb{E}[\min\{x_1 + \varepsilon_1, \ldots, x_n + \varepsilon_n\}\,]$$

Per semplicità, sia $\mathcal{E}$ la classe di funzioni $\varphi(x)$, per tutte le n-uple di variabiali aleatorie ε_i che abbiano distribuzione continua e media nulla.
Vale il seguente risultato:

Proposizione 4 (Proposition A.1.)**.** Sia $\varphi \in \mathcal{E}$. Allora $\varphi(x)$ è una funzione concava, $\varphi(x) \leq \min\{x_1, \dots, x_n\}$. $\varphi(x)$ è monotona debolmente crescente in ciascuna variabile separatamente. Inoltre $\varphi(x)$ è di classe $\mathbb{C}^1$ e le sue derivate si possono esprimere come:

$$\frac{\partial \varphi}{\partial x_i}(x) = \mathbb{P}(\tilde{x}_i \leq \tilde{x}_j, \forall j) \tag{5}$$

Dimostrazione. Sia $m(x) = \min\{x_1, \dots, x_n\}$; $m(x)$ è una funzione concava.
Fissato un certo j, in generale vale:

$$\min\{x_1 + \varepsilon_1, \dots, x_n + \varepsilon_n\} \leq x_j + \varepsilon_j$$

Prendendo il valore atteso di ambo i membri, si ottiene:

$$\varphi(x) = \mathbb{E}[\min\{x_1 + \varepsilon_1, \dots, x_n + \varepsilon_n\}] \leq \mathbb{E}[x_j + \varepsilon_j] = x_j$$

Quindi per ogni j vale $\varphi(x) \leq x_j$, dunque $\varphi(x) \leq \min_j x_j = m(x)$.
Sia $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ e $F(\varepsilon)$ la funzione di ripartizione di ε. Allora vale

$$\varphi(x) = \int_{\mathbb{R}^n} m(x + \varepsilon) \, \mathrm{d}F(\varepsilon)$$

Approssimando l'integrale con sommatorie parziali, esse sono combinazioni lineari (con pesi tra 0 e 1, in particolare positivi) di funzioni $m(x + \varepsilon)$, che sono concave. Dunque le sommatorie parziali sono concave, e lo è anche il limite $\varphi(x)$.

Presi $x = (x_1, x_2, \dots, x_n)$ e $y = (y_1, x_2, \dots, x_n)$, allora vale la disuguaglianza puntualmente tra variabili aleatorie:

$$\min\{x_1 + \varepsilon_1, \dots, x_n + \varepsilon_n\} \leq \min\{y_1 + \varepsilon_1, \dots, x_n + \varepsilon_n\}$$

Dunque $\varphi(x) \leq \varphi(y)$. Questo mostra la monotonia separatamente su ogni componente.
Calcoliamo ora la derivata $\frac{\partial \varphi}{\partial x_i}(x)$ usando il rapporto incrementale:

$$\frac{\varphi(x + te_i) - \varphi(x)}{t} = \int_{\mathbb{R}^n} q_t(\varepsilon) \, \mathrm{d}F(\varepsilon)$$

dove $q_t(\varepsilon) = \frac{1}{t}\left(m(x + \varepsilon + te_i) - m(x + \varepsilon)\right)$. Si definiscano gli insiemi

$$A = \{\varepsilon \in \mathbb{R}^n : x_i + \varepsilon_i < x_j + \varepsilon_j, \ \forall j \neq i\}$$

$$B = \{\varepsilon \in \mathbb{R}^n : x_i + \varepsilon_i \leq x_j + \varepsilon_j, \ \forall j \neq i\}$$

Si vede che $\lim_{t \to 0^+} q_t(\varepsilon) = 1_A(\varepsilon)$ e $\lim_{t \to 0^-} q_t(\varepsilon) = 1_B(\varepsilon)$. Entrambe le successioni sono monotone, per cui per il teorema di convergenza monotona possiamo calcolare le derivate destra e sinistra di $\varphi(x)$:

$$D_i^+ \varphi(x) = \int_{\mathbb{R}^n} 1_A(\varepsilon) \, \mathrm{d}F(\varepsilon) = \mathbb{P}(A)$$

$$D_i^- \varphi(x) = \int_{\mathbb{R}^n} 1_B(\varepsilon) \, \mathrm{d}F(\varepsilon) = \mathbb{P}(B)$$

Osserviamo che ε ha distribuzione continua, dunque le probabilità di sopra sono uguali:

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(\tilde{x}_i \leq \tilde{x}_j, \forall j)$$

ed è uguale alla derivata parziale $\frac{\partial \varphi}{\partial x_i}(x)$. $\square$

3.3 Tempo necessario a raggiungere la destinazione

Definizione 5. Per ogni ogni coppia origine-destinazione (i, d), definisco la funzione φ_i^d dipendente da z_a^d, al variare di a tra gli archi uscenti da i:

$$\varphi_i^d(z^d) = \mathbb{E}\left[\ \min_{a \in A_i^+} \{\tilde{z}_a^d\}\ \right]$$

$\tilde{z}^d = (\tilde{z}_a^d)_{a \in A}$ è il vettore di tutti i tempi, ma nel minimo di sopra ci interessano solo per gli archi a uscenti da i.
Ricordiamo che le $\tilde{z}_a^d$ sono variabili aleatorie, che $\tilde{z}_a^d = z_a^d + \varepsilon_a^d$ dove z_a^d è la media e ε_a^d è il termine aleatorio. Dunque la funzione φ_i^d prende come argomento dei numeri reali z_a^d, ma il risultato è un valore atteso, che dipende dalle variabili aleatorie ε_a^d.
Inoltre, visto che al variare di i è diverso il numero di archi uscenti, ciascuna φ_i^d cerca il minimo tra un numero diverso di variabili.
Per ora i ragionamenti sono molto generali, applicabili a qualsiasi scelta di variabili aleatorie $\widetilde{t_a} = t_a + \nu_a$ e $\widetilde{z_a^d} = z_a^d + \varepsilon_a^d$. In particolare, ε_a^d dipende in maniera molto complicata dai ν_a: non è semplicemente la somma relativa agli archi del percorso più "breve", ma è una composizione iterata di minimi. Nella sezione 5.2 vedremo come scegliere le variabili aleatorie, e le conseguenze dai punti di vista computazionali e di realisticità del modello.

Usando i risultati della sezione 3.2, possiamo calcolare le probabilità di transizione. La funzione presa in considerazione è φ_i^d, con variabili aleatorie $\tilde{z}_a^d$. Dunque la probabilità che $\tilde{z}_a^d$ sia minore di $\tilde{z}_b^d$ per ogni b è data dalla derivata parziale di φ_i^d. Nella catena di Markov, partendo da $i \neq d$, la probabilità di andare in j è:

$$P_{ij}^d = \mathbb{P}(\tilde{z}_a^d \leq \tilde{z}_b^d \text{ per ogni arco } b \text{ uscente da } i\ ) = \frac{\partial \varphi_i^d}{\partial z_a^d}(z^d) \tag{6}$$

Vediamo come calcolare le $\tilde{z}_a^d$ e $\tilde{\tau}_i^d$ a partire dai tempi di percorrenza di ciascun arco $\tilde{(t)}_a$. Si hanno le seguenti espressioni, dove $a = (i, j_a)$:

$$\begin{cases} \tilde{z}_a^d = \tilde{t}_a + \tilde{\tau}_{j_a}^d \\ \tilde{\tau}_i^d = \min\limits_{a \in A_i^+} \tilde{z}_a^d \end{cases} \tag{7}$$

Prendendo il valore atteso, si ha:

$$\begin{cases} z_a^d = \mathbb{E}[\tilde{z}_a^d] = \mathbb{E}[\tilde{t}_a] + \mathbb{E}[\tilde{\tau}_{j_a}^d] = t_a + \tau_{j_a}^d \\ \tau_i^d = \mathbb{E}[\tilde{\tau}_i^d] = \mathbb{E}[\min\limits_{a \in A_i^+} \tilde{z}_a^d] = \varphi_i^d(\{z_a^d\}_{a \in A_i^+}) \end{cases}$$

Mettendo insieme le due espressioni, si ottiene:

$$\tau_i^d = \varphi_i^d(\{t_a + \tau_{j_a}^d\}_{a \in A_i^+}) \tag{8}$$

Questo sistema di equazioni, al variare di i, d, permette di trovare i tempi medi τ_i^d (che sono numeri reali). La parte stocastica è racchiusa nella funzione φ_i^d, che implicitamente contiene la variabile aleatoria $\varepsilon = \tilde{\tau}_j^d - \tau_j^d$.
Osserviamo che la funzione $\varphi_i^d(z^d)$ può essere pensata come approssimante continua e derivabile di $\overline{\varphi}_i^d(z^d) = \min\limits_{a \in A_i^+} z_a^d$. La funzione $\overline{\varphi}_i^d$ permette di stabilire il tempo minimo del percorso da i a d nel modello di Wardrop originario, ovvero senza la componente stocastica.

3.4 Matrici P,Q e flussi

Fissata la destinazione d, gli utenti si muovono secondo una catena di Markov, con matrice di transione P^d descritta in precedenza.
Indichiamo con v_a^d il flusso attraverso l'arco a con destinazione d, e x_i^d il flusso passante per il vertice i. Per conservazione del flusso, x_i^d è uguale a tutto il flusso entrante in i, e a tutto quello uscente.

$$x_i^d = \sum_{a \in A_i^-} v_a^d = g_i^d + \sum_{a \in A_i^+} v_a^d \tag{9}$$

Osserviamo che una parte del flusso entrante $\sum\limits_{a \in A_i^-} v_a^d$ è composta da utenti che hanno come destinazione i, ovvero $\sum\limits_k g_k^i$.

Per come abbiamo costruito la catena di Markov, ogni utente in i decide di seguire l'arco $a = (i, j)$ con probabilità $P_{ij}^d(z^d)$. Il flusso x_i^d attraverso i viene diviso in proporzione ai P_{ij}^d per ciascun arco:

$$v_a^d = x_i^d \cdot P_{ij}^d(z^d) = x_i^d \cdot \mathbb{P}(\tilde{z}_a^d \leq \tilde{z}_b^d, \forall b \in A_i^+) = x_i^d \frac{\partial \varphi_i^d}{\partial z_a^d}(z^d)$$

Esprimiamo questa relazione in forma vettoriale. v^d per tutti i valori di a, è un vettore con m componenti, mentre x^d al variare di i ne ha n.
Definiamo $\widehat{Q}^d$ una matrice $(n-1) \times m$, con entrate

$$\widehat{Q}_{ia}^d = \begin{cases} \dfrac{\partial \varphi_i^d}{\partial z_a^d}(z^d) & \text{se } a = (i, *), \text{ con } i \neq d \\ 0 & \text{se } a \neq (i, *), \text{ con } i \neq d \end{cases} \tag{10}$$

Si è tolto l'indice del vertice "assorbente" d. Similmente sia $\widehat{x}^d$ un vettore con $n-1$ componenti, in cui manca x_d^d. La relazione vettoriale è:

$$v^d = \widehat{Q}^d(z^d)^\top \, \widehat{x}^d \tag{11}$$

Vediamo come calcolare $\widehat{x}^d$. Il flusso uscente da i è dato dal flusso generato nel vertice, g_i^d, e da quello che ci arriva da altri. Questo porta alla relazione:

$$x_i^d = g_i^d + \sum_{k \neq d} P_{ki}^d(z^d)\, x_k^d \tag{12}$$

In notazione vettoriale, con $\widehat{P}^d$ la matrice $(n-1)\times(n-1)$ con la d-esima riga e colonna rimosse, si ottiene:

$$\widehat{x}^d = \widehat{g}^d + \widehat{P}^d(z^d)^\top \, \widehat{x}^d \tag{13}$$

$$\widehat{x}^d = (I - \widehat{P}^d(z^d)^\top)^{-1} \, \widehat{g}^d \tag{14}$$

Osserviamo che il trasposto nelle equazioni 11 e 13 è presente perché usiamo vettori colonna (e non riga, come è solito con le matrici di transizione).
Al posto delle matrici ridotte $\widehat{Q}^d$ e $\widehat{P}^d$, potevamo usare quelle contenenti la d-esima riga (e colonna), tanto le relative entrate (anche di x_d^d) sono nulle. Usiamo quelle più piccole per ridurre la dimensione, e per evitare problemi di singolarità nel calcolo dell'equazione 14 (infatti la matrice $I - P^d$ è singolare).

4 Equilibrio MTE e soluzioni delle equazioni

4.1 Esistenza ed unicità dei tempi di arrivo

Una volta che abbiamo a disposizione i tempi τ_i^d, è possibile ottenere i z_a^d e di conseguenza i flussi su ciascun arco e vertice. Tuttavia, per ora non sappiamo se l'equazione 8 ammetta soluzioni, o se ci sia unicità. Ricordiamo l'equazione in questione:

$$\tau_i^d = \varphi_i^d(\{t_a + \tau_{j_a}^d\}_{a \in A_i^+})$$

Inoltre, per ora non sappiamo nulla sulle Catene di Markov che abbiamo definito per ogni destinazione: a priori, potrebbero essere non irriducibili, o ergodiche; non è neanche garantita l'esistenza di un cammino da un qualsiasi vertice al (unico) stato assorbente d.
Per garantire l'esistenza di una soluzione e avere un modello funzionante, introduciamo un'ipotesi tecnica. Definiamo il seguente insieme di ammissibilità:

$$\mathcal{C} = \left\{ t \in \mathbb{R}^m : \; \exists \widehat{\tau}_i^d < \varphi_i^d \left(\{t_a + \widehat{\tau}_{j_a}^d\}_{a \in A_i^+} \right) \; \forall i \neq d \right\}$$

Osservazione 1. Siccome le φ_i^d sono continue, concave e debolmente monotone in ogni componente, segue che $\mathcal{C}$ è aperto e convesso. Inoltre, $\mathcal{C}$ è un cono.

Osserviamo che $\mathcal{C}$ non è vuoto: basta prendere t abbastanza grande in modo che $\varphi_i^d(t) > 0$ per ogni i, e poi scegliere $\widehat{\tau}^d = 0$.

Proposizione 5 (Lemma 2.1)**.** Dato un $t \in \mathcal{C}$, supponiamo che τ^d risolva l'equazione 8. Per questo t, denotiamo con $\widehat{\tau}^d$ la corrispondente sotto-soluzione. Sia $\delta_i^d = \tau_i^d - \widehat{\tau}_i^d$ per ogni $i \neq d$. Allora esiste un vertice j tale che $P_{ij}^d > 0$ e $\delta_j^d < \delta_i^d$.

Dimostrazione. Siano $\widehat{z}_a^d = t_a + \widehat{\tau}_{j_a}^d$ e $z_a^d = t_a + \tau_{j_a}^d$ per ogni $a \in A$.
La funzione φ_i^d è concava e derivabile, per cui sta sotto il grafico della tangente nel punto $\widehat{z}$:

$$\begin{aligned}\varphi_i^d(\widehat{z}^d) &\leq \varphi_i^d(z) + \sum_{a\in A} \frac{\partial \phi_i^d}{\partial z_a^d}(z^d)\,(\widehat{z}_a^d - z_a^d) = \\ &= \varphi_i^d(z^d) + \sum_{a\in A_i^+} P_{ij_a}^d\,(\widehat{\tau}_{j_a}^d - \tau_{j_a}^d) = \\ &= \varphi_i^d(z^d) - \sum_{(i,j)\in A_i^+} P_{ij}^d \delta_j^d\end{aligned}$$

Riarrangiando:

$$\sum_{(i,j)\in A_i^+} P_{ij}^d \delta_j^d \leq \varphi_i^d(z^d) - \varphi_i^d(\widehat{z}^d)$$

Visto che $\widehat{\tau}_i^d < \varphi_i^d(\widehat{z}^d)$ e $\tau_i^d = \varphi_i^d(z^d)$, segue che

$$\sum_{(i,j)\in A_i^+} P_{ij}^d \delta_j^d \leq \varphi_i^d(z^d) - \varphi_i^d(\widehat{z}^d) \leq \tau_i^d - \widehat{\tau}_i^d = \delta_i^d$$

Preso l tale che δ_l^d sia il minimo tra tutti i δ_j^d con $j \in A_i^+$, segue che

$$\delta_i^d > \sum_{(i,j)\in A_i^+} P_{ij}^d \delta_j^d > \sum_{(i,j)\in A_i^+} P_{ij}^d \delta_l^d = \left(\sum_{(i,j)\in A_i^+} P_{ij}^d\right) \delta_l^d = \delta_l^d$$

Dunque $j = l$ è quello cercato.□

Lo scopo dei δ_i^d è di avere un invariante strettamente decrescente, ogni volta che si segue un cammino. Questo permette di formare cammini garantendo di non entrare in un ciclo.

Proposizione 6 (Proposition 2.2)**.** Dato un $t \in \mathcal{C}$, supponiamo che τ^d risolva l'equazione 8. Sia $z_a^d = t_a + \tau_{j_a}^d$ e si costruiscano di conseguenza le matrici $\widehat{P}^d(z^d)$ e $\widehat{Q}^d(z^d)$. Allora lo stato assorbente d della relativa catena di Markov è raggiungibile da qualsiasi vertice di partenza. Inoltre la matrice $I - \widehat{P}^d(z^d)$ è invertibile, dunque sono ben definiti i vettori $\widehat{x}^d = (I - \widehat{P}^d(z^d)^\top)^{-1}\,\widehat{g}^d$ e $v^d = \widehat{Q}^d(z^d)^\top\,\widehat{x}^d$.

Dimostrazione. Usando ripetutamente la proposizione 5, a partire da un vertice i si può ottenere una sequeanza di vertici $i_0 = i, i_1, i_2, \ldots$ tale che $\widehat{P}_{i_l\, i_{l+1}}^d > 0$ e $\delta_{i_{l+1}}^d < \delta_{i_{l+1}}^d$. I δ decrescono fino a quando non si raggiunge $i_k = d$ (infatti nelle ipotesi della

proposizione 5 c'è $i \neq d$); visto che sono in numero finito, vuol dire che esiste un percorso da i a d, percorribile dalla catena di Markov con probabilità positiva.
Questo implica che la matrice ridotta $\widehat{P}^d$ (con le d-esime riga e colonna rimosse) ha autovalore dominante minore di 1, dunque $I - \widehat{P}^d(z^d)$ è invertibile e quindi sono ben definiti i vettori $v^d = \widehat{Q}^d(z^d)^\top \,\widehat{x}^d$ e $\widehat{x}^d = [I - \widehat{P}^d(z^d)^\top]^{-1}\, g^d$. $\square$

Dunque, conoscendo τ^d, (che determina P e Q), i flussi sono univocamente determinati. Resta da dimostrare la seguente:

Proposizione 7 (Proposition 2.3)**.** Per $t \in \mathcal{C}$, fissata una destinazione d, le equazioni 8 hanno un'unica soluzione $\tau^d(t)$ in funzione di t.
Siano $\overline{\tau}_i^d(t)$ le soluzioni alla variante deterministica del sistema 8:

$$\overline{\tau}_i^d = \min_{a \in A_i^+} \{t_a + \overline{\tau}_{j_a}^d\} = \overline{\varphi}_i^d \left(\{t_a + \overline{\tau}_{j_a}^d\}_{a \in A_i^+} \right)$$

Le mappe $t \longmapsto \tau_i^d(t)$ sono lisce, concave, debolmente monotone in ogni componente e vale $\tau_i^d(t) \leq \overline{\tau}_i^d(t)$ per ogni $i \neq d$.

Dimostrazione.
Esistenza di una soluzione τ. Per ipotesi $t \in \mathcal{C}$, quindi esiste $\widehat{\tau}$ tale che per ogni i:

$$\widehat{\tau}_i^d < \varphi_i^d(\{t_a + \widehat{\tau}_{j_a}^d\}_{a \in A_i^+})$$

Mostreremo che esiste una soluzione τ_i^d tale che $\widehat{\tau}_i^d \leq \tau_i^d \leq \overline{\tau}_i^d$. Qui $\overline{\tau}$ e $\widehat{\tau}$ sono già dati (e dipendono da t).
Consideriamo la sequenza:

$$\begin{cases} \tau^{(0),d} = \widehat{\tau}^d \\ \tau_i^{(k+1),d} = \varphi_i^d(\, \{t_a + \tau_{j_a}^{(k),d}\}_{a \in A_i^+}) \qquad \forall i \end{cases} \tag{15}$$

Mostriamo che questa sequenza è crescente (in ogni componente). Per ipotesi vale $\widehat{\tau} < \varphi(t + \widehat{\tau})$, dunque:

$$\tau^{(0),d} = \widehat{\tau} < \varphi(\{t_a + \widehat{\tau}_{j_a}^{(0),d}\}_{a \in A_i^+}) = \tau^{(1),d}$$

Le funzioni φ sono monotone debolmente crescenti in ogni componente, dunque

$$\tau_i^{(k),d} = \varphi(\{t_a + \widehat{\tau}_{j_a}^{(k-1),d}\}_{a \in A_i^+}) \leq \varphi(\{t_a + \widehat{\tau}_{j_a}^{(k),d}\}_{a \in A_i^+}) = \tau_i^{(k+1),d}$$

Ovvero $\tau^{(0),d} \leq \tau^{(1),d} \leq \ldots \leq \tau^{(k),d} \leq \tau^{(k+1),d} \leq \ldots$.
Sempre per induzione, mostriamo che $\tau^{(k),d} \leq \overline{\tau}^d$. Il passo base è $\tau^{(0),d} = \widehat{\tau}^d \leq \overline{\tau}^d$. Per il passo induttivo, supponendo che $\tau^{(k),d} \leq \overline{\tau}^d$, si ha:

$$\begin{aligned} \tau_i^{(k+1),d} &= \varphi_i^d(\{t_a + \widehat{\tau}_{j_a}^{(k),d}\}_{a \in A_i^+}) \leq \\ &\leq \varphi_i^d(\{t_a + \overline{\tau}_{j_a}^d\}_{a \in A_i^+}) \leq \\ &\leq \overline{\varphi}_i^d(\{t_a + \overline{\tau}_{j_a}^d\}_{a \in A_i^+}) = \overline{\tau}_i^d \end{aligned}$$

La sequenza $\tau^{(k),d}$ è crescente (in ogni componente) e limitata dall'alto da $\overline{\tau}^d$, perciò ha un limite τ^d che soddisfa $\widehat{\tau}^d \leq \tau^d \leq \overline{\tau}^d$ e $\tau_i^d = \varphi(\{t_a + \tau_{j_a}^d\}_{a\in A_i^+})$.

Unicità della soluzione τ. Siano τ^1, τ^2 due soluzioni, e sia $\alpha = \max_i \left\{\tau_i^{1,d} - \tau_i^{2,d}\right\}$. Sia V' l'insieme dei nodi i dove si ha il massimo, ovvero $i \in V' \iff \tau_i^{1,d} - \tau_i^{2,d} = \alpha$. In particolare, per qualsiasi nodo j si ha $\tau_j^{1,d} \leq \tau_j^{2,d} + \alpha$. Preso $i \in V'$, si ha per la monotonia di φ:

$$\tau_i^1 = \varphi_i^d(\{t_a + \tau_{j_a}^{1,d}\}_{a\in A_i^+}) \leq \varphi_i^d(\{t_a + \tau_{j_a}^{2,d} + \alpha\}_{a\in A_i^+}) =$$

$$= \varphi_i^d(\{t_a + \tau_{j_a}^{2,d}\}_{a\in A_i^+}) + \alpha = \tau_i^2 + \alpha = \tau_i^1$$

La derivata parziale di φ_i^d lungo la direzione $\frac{\partial}{\partial z_a^d}$, ovvero l'arco $a = (i,j)$, è $\frac{\partial \varphi_i^d}{\partial z_a^d}(\{t+\tau\}) = P_{ij}^d$. In particolare, se $P_{ij}^d > 0$, allora φ_i^d è strettamente crescente lungo la direzione z_a^d; visto che sopra si è ottenuta un'uguaglianza, questo vuol dire che $\tau_j^{1,d} = \tau_j^{2,d} + \alpha$.
Dunque ogni volta che $i \in V'$ e per un j si ha $P_{ij}^d > 0$, allora anche $j \in V'$. Per la proposizione 5, esiste un cammino che congiunge i a d fatto da archi percorsi con probabilità positiva. Dunque anche $d \in V'$, cioé $\tau_d^{1,d} = \tau_d^{2,d} + \alpha$. Ma $\tau_d^{1,d} = \tau_d^{2,d} = 0$, dunque $\alpha = 0$; questo implica che $\tau^{1,d} \leq \tau^{2,d}$. Scambiando i ruoli di $\tau^{1,d}$ e $\tau^{2,d}$ si ottiene la disuguaglianza opposta, da cui $\tau^{1,d} = \tau^{2,d}$.

Concavità. Dati $t^1, t^2 \in \mathcal{C}$, siano $\tau^1 = \tau^d(t^1)$ e $\tau^2 = \tau^d(t^2)$. Sia $t^\alpha = \alpha t^1 + (1-\alpha)t^2$ una combinazione convessa, e sia $\tau^3 = \alpha\tau^1 + (1-\alpha)\tau^2$. Le funzioni φ_i^d sono concave, per cui

$$\tau_i^d(t^\alpha) = \varphi_i^d(\{t_a + \tau_{j_a}^3\}_{a\in A_i^+}) \geq \alpha\varphi_i^d(\{t_a + \tau_{j_a}^1\}_{a\in A_i^+}) + (1-\alpha)\varphi_i^d(\{t_a + \tau_{j_a}^2\}_{a\in A_i^+}) =$$

$$= \alpha\tau_i^1 + (1-\alpha)\tau_i^2 = \tau_i^{3,d}$$

Quindi $\tau_i^{3,d} \leq \varphi(t + \tau^3)$, ovvero τ^3 soddisfa le ipotesi richieste su $\widehat{\tau}$ della dimostrazione di esistenza. Questa implica che $\tau_i^{3,d} \leq \tau_i^d(t^\alpha)$, dove il membro a destra è il limite della sequenza iterata. Tuttavia questa relazione è proprio quella di concavità:

$$\tau_i^d(t^\alpha) \geq \tau_i^{3,d} = \alpha\tau^1 + (1-\alpha)\tau^2$$

Monotonia. Siano $t^1 \leq t^2 \in \mathcal{C}$. Chiamato $\widehat{\tau} = \tau(t^1)$, per la monotonia delle φ_i^d si ha:

$$\widehat{\tau}_i^d = \varphi_i^d\{t_a^1 + \widehat{\tau}_{j_a}^d\}_{a\in A_i^+}) \leq \varphi_i^d\{t_a^2 + \widehat{\tau}_{j_a}^d\}_{a\in A_i^+}$$

Sempre per la dimostrazione di esistenza, questo implica che $\tau_i^d(t^2)$, ottenuto come limite della successione, è maggiore di $\widehat{\tau}_i^d$. Dunque:

$$\tau_i^d(t^1) = \widehat{\tau}_i^d \leq \tau_i^d(t^2)$$

$\square$

4.2 Equilibrio MTE

Fino a questo momento abbiamo trattato separatamente i flussi per ciascuna destinazione. Tuttavia, il traffico su un arco è dato da tutti gli utenti che ci passano, qualsiasi sia la loro destinazione, e dunque i tempi di percorrenza dipendono dal congestionamento totale.

Definizione 6. Definiamo il flusso aggregato $w_a = \sum_{d \in V} v_a^d$ per ciascun arco $a \in A$.

Definizione 7. Un vettore $w \in \mathbb{R}^m$ è un *equilibrio Markoviano del traffico* (MTE), se valgono le seguenti relazioni tra le variabili v_a^d, x_i^d, z_a^d, t_a, τ_i^d:

1. $w_a = \sum_{d \in V} v_a^d$ per ogni arco a.
2. $z_a^d = t_a + \tau_{j_a}^d$ dove $a = (*, j_a)$.
3. $\tau_i^d = \varphi_i^d(\{t_a + \tau_{j_a}^d\}_{a \in A_i^+})$
4. $v^d = \widehat{Q}^d(z^d)^\top \, \widehat{x}^d$
5. $\widehat{x}^d = (I - \widehat{P}^d(z^d)^\top)^{-1} \, \widehat{g}^d$
6. $t_a = s_a(w_a)$

Le prime 5 sono riepilogo di condizioni che abbiamo già visto sopra, mentre quella nuova è l'ultima: essa garantisce che i tempi di percorrenza di ciascun arco, sono esattamente quelli ottenuti dal traffico w_a tramite la funzione di congestione $s_a(\cdot)$.
Partendo da t_a possiamo calcolare tutte le altre quantità, però non è detto che il w che otterremo soddisferà l'equazione. Per questo dal punto di vista computazionale si userà un metodo iterativo per il calcolo di w e t, fino a quando non si ottiene convergenza.
La soluzione w trovata può essere caratterizzata in un modo simile al caso non-stocastico:

Proposizione 8 (Theorem 2.5)**.** Sia $t_a^0 = s_a(0)$ e assumiamo che il vettore $t^0 \in \mathcal{C}$. Allora esiste un unico MTE, che chiamiamo $w_a^* = s_a^{-1}(t_a)$ dove t^* è l'unica soluzione al seguente problema di minimo (strettamente convesso):

$$\min_{t \in \mathcal{C}} \Phi(t) \quad \text{dove} \quad \Phi(t) = \sum_{a \in A} \int_{t_a^0}^{t_a} s_a^{-1}(y) \, \mathrm{d}y - \sum_{d} \sum_{i \neq d} g_i^d \tau_i^d(t) \tag{16}$$

Dimostrazione:
Localizzazione del minimo di Φ. Sia $f_a(t_a) = \int_{t_a^0}^{t_a} s_a^{-1}(y) \, \mathrm{d}y$. $f_a(\cdot)$ è crescente per $t_a > t_a^0$ e decrescente per $t_a < t_a^0$ (e dunque ha minimo assoluto in $t_a = t_a^0$). Siccome le funzioni $\tau_i^d(t)$ sono debolmente monotone crescenti in ogni componente, allora la sommatoria $-\sum g_i^d \tau_i^d(t)$ è debolmente monotona decrescente in ogni componente.
Consideriamo la regione $\mathcal{A} = \{t \geq t^0\} \subseteq \mathcal{C}$. Preso $t \in \mathcal{C} \setminus \mathcal{A}$, possiamo trovare $t^1 \in \mathcal{A}$ con $t^1 \geq t$, aumentando opportune componenti. Per quanto detto prima, $\Phi(t^1) \geq \Phi(t)$, per cui (l'eventuale) minimo di $\Phi(\cdot)$ è da cercare nella regione $\{t \geq t^0\} \subseteq \mathcal{C}$

$\Phi(t)$ **è convessa**. Basta verificare che ogni termine della sommatoria è convessa. Per la proposizione 4 si ha che $\tau_i^d(t)$ è concava, dunque $-g_i^d \tau_i^d(t)$ è convessa per ogni i, d. Per il termine con l'integrale, notiamo che $s_a^{-1}(t_a)$ è crescente e con $\lim\limits_{t_a \to \infty} s_a^{-1}(t_a) = \infty$, dunque $f_a(t_a) = \int_{t_a^0}^{t_a} s_a^{-1}(y)\ \mathrm{d}y$ è convessa e coerciva. Inoltre, è strettamente convessa per $t_a \geq t_a^0$.
Possiamo concludere per linearità che $\Phi(t)$ è convessa e coerciva. Inoltre, nella regione $\{t \geq t^0\}$ è strettamente convessa: presi due punti distinti $t^1, t^2 \geq t^0$, esiste una componente a tale che $t_a^1 \neq t_a^2$. Chiamata $t^3 = \beta t^1 + (1-\beta)t^2$ una combinazione convessa, si ha

$$f_a(t_a^3) = f_a(\beta t_a^1 + (1-\beta)t_a^2) < \beta f_a(t_a^1) + (1-\beta) f_a(t_a^2)$$

per la stretta convessità di f_a. Sommando tutte le componenti, si ottiene $\Phi(t^3) < \beta\Phi(t^1) + (1-\beta)\Phi(t^2)$, ovvero che Φ è strettamente convessa su $\{t \geq t^0\}$ e dunque ha un unico minimo.
Caratterizzazione del minimo. Mostreremo che il minimo t^* è il punto in cui $\nabla\Phi(t^*) = 0$. Calcoliamo $\nabla\Phi(t)$.
Dall'equazione 8 ovvero $\tau_i^d = \varphi_i^d(\{t_a + \tau_{j_a}^d\}_{a \in A_i^+})$, facendo la derivata parziale in t_a, con $a = (i, j_a)$, si ottiene:

$$\frac{\partial \tau_i^d}{\partial t_a} = \sum_{b \in A_i^+} \frac{\partial \varphi_i^d}{\partial z_b}(\{z_c\}_{c \in A_i^+}) \cdot \left(1 + \frac{\partial \tau_{j_b}^d}{\partial t_a}\right) = \sum_{b \in A_i^+} P_{i\,j_b}^d + P_{i\,j_b}^d \frac{\partial \tau_{j_b}^d}{\partial t_a}$$

Ricoridamo che $P_{i\,j_b}^d = Q_{i\,b}^d$. Si riscriva l'equazione di sopra in forma vettoriale. Usiamo $\widehat{Q}^d$, $\widehat{P}^d$ e $\widehat{\tau}^d$ per ricordare che è stata tolta la componente relativa al vertice d (tra l'altro si ha $\tau_d^d = 0$).

$$\frac{\partial \widehat{\tau}^d}{\partial t_a} = \widehat{Q}_{:b}^d + P^d \frac{\partial \widehat{\tau}^d}{\partial t_a}$$

Qui $\widehat{Q}_{:b}^d$ è da intendersi in notazione di Matlab, ovvero la b-esima colonna di $\widehat{Q}^d$.

$$(I - \widehat{P}^d) \frac{\partial \widehat{\tau}^d}{\partial t_a} = \widehat{Q}_{:b}^d$$

$$\frac{\partial \widehat{\tau}^d}{\partial t_a} = (I - \widehat{P}^d)^{-1} \widehat{Q}_{:b}^d$$

Calcoliamo ora $\frac{\partial \Phi}{\partial t_a}(t)$, separatamente sul termine con l'integrale e sulla sommatoria. Per l'integrale, si ha semplicemente:

$$\frac{\partial}{\partial t_a}\left(\sum_{b \in A} \int_{t_b^0}^{t_b} s_b^{-1}(y)\ \mathrm{d}y\right) = \frac{\partial}{\partial t_a} \int_{t_a^0}^{t_a} s_a^{-1}(y)\ \mathrm{d}y = s_a^{-1}(t_a)$$

Si denoti $\psi^d(t) = \sum\limits_{i \neq d} g_i^d(t) \tau_i^d(t)$. Si ha:

$$\begin{aligned}\frac{\partial \psi^d}{\partial t_a}(t) &= \sum_i \left(\frac{\partial \tau_i^d}{\partial t_a}(t)\right) \cdot g_i^d = \left(\frac{\partial \widehat{\tau}^d}{\partial t_a}(t)\right)^\top \widehat{g}^d = \widehat{Q}_a^d(t)^\top \, (I - \widehat{P}^d(t)^\top)^{-1} \, \widehat{g}^d = \\ &= \widehat{Q}_a^d(t)^\top \, \widehat{x}^d(t) = v_a^d(t)\end{aligned}$$

Dunque si ha

$$\frac{\partial \Phi}{\partial t_a}(t) = s_a^{-1}(t_a) - \sum_d v_a^d(t)$$

La soluzione del problema di minimo (w^*, t^*) è tale che $s_a^{-1}(t^*) = w^*$. D'altronde, $\sum_d v_a^d(t) = w$ per definizione di w. Dunque $\frac{\partial \Phi}{\partial t_a}(t) = 0$ per ogni componente a se e solo se la coppia (w^*, t^*) è l'(unica) soluzione.

5 Approccio computazionale

5.1 Scelta delle funzioni parametriche

La funzione di congestione $s_a : \mathbb{R} \to \mathbb{R}^+$ per ora è sempre stata generica, con la sola richiesta che fosse crescente. Nelle prove sperimentali, si assume che la funzione abbia la seguente forma:

$$s_a(w_a) \;=\; t_a^0 \left(1 + b_a \left(\frac{w_a}{c_a}\right)^{p_a}\right) \tag{17}$$

dove

- t_a^0 è il tempo di transito sull'arco quando non c'è traffico, ovvero $s_a(0)$.
- c_a indica la capacità dell'arco. Per $w_a \leq c_a$, il traffico riesce a scorrere bene, mentre per $w_a > c_a$ inizia a crearsi congestione.
- Per evidenziare la formazione di traffico nel caso ci siano più utenti di quanto quell'arco riesca a sopportare, è introdotto un esponente p_a. Nel nostro caso, è sempre preso uguale a 4.
- b_a è un fattore di scala.

5.2 Modello logit

Nel modello MTE, abbiamo assunto che i tempi di percorrenza di ogni arco siano variabili aleatorie $\widetilde{t}_a = t_a + \nu_a$. Da queste ricaviamo $\widetilde{\tau}_i^d$ e $\widetilde{z}_a^d$, che indicano il valore atteso del tempo di percorrenza del percorso più corto da i fino alla destinazione. Dunque, la distribuzione di $\widetilde{z}_a^d$ ha un'espressione molto complicata in funzione delle ν_a: è un minimo, che dipende dagli altri $\widetilde{z}_a^d$ a cui sono sommati gli ν_a.
Scriviamo $\widetilde{z}_a^d = z_a^d + \varepsilon_a^d$, con $z_a^d = \mathbb{E}[\widetilde{z}_a^d]$ la media, e ε_a^d comprendente la parte stocastica.
L'approccio che useremo è il seguente: supponiamo che i tempi per raggiungere la destinazione siano variabili di Gumbel, con media z_a^d (e con un certo parametro di scala β_i^d). Ovverosia, consideriamo ogni possibile percorso da i a d separatamente, e i tempi di percorrenza di ciascuno sono variabili aleatorie ben stabilite, che possiamo calcolare facilmente.
Questo approccio facilita notevolmente i conti, perché è possibile calcolare esplicitamente le funzioni φ_i^d. Il rovescio della medaglia è che le ν_a non si possono ottenere dalle ε_a^d; anzi, un modello comodo dovrebbe funzionare al contrario: si potrebbero modificare

i parametri dei tempi di ciascun singolo arco e vedere come questo influenza la distribuzione di traffico. Il problema più grosso, però, è dal punto di vista probabilistico: assumiamo che le ε_a^d abbiano una certa distribuzione e soprattutto siano indipendenti tra di loro. Visto che rappresentano tanti cammini, questo è generalmente falso, perché ci saranno molti tratti in comune tra diversi cammini. L'idea di questo approccio risale a [3].
In ogni caso, usiamo questo modello per via della facile espressione esplicita. Una variabile di Gumbel, con media μ e parametro β è:

$$\begin{aligned} F(x) &= e^{-e^{-\frac{x-\mu}{\beta}}} && \text{funzione di ripartizione} \\ f(x) &= \frac{1}{\beta}\, e^{-e^{-\frac{x-\mu}{\beta}}-\frac{x-\mu}{\beta}} && \text{funzione densità} \end{aligned} \tag{18}$$

Fissiamo l'origine i, la destinazione d e un parametro β_i^d. Al variare di a tra gli archi uscenti da i, scegliamo $-\varepsilon_a^d$ come variabili di Gumbel a media nulla e parametro β_i^d. Allora $\widetilde{z}_a^d = z_a^d + \varepsilon_a^d$ è una variabile di Gumbell "specchiata", con media z_a^d e parametro β_i^d. In questo modo:

$$\mathbb{P}(\, \widetilde{z}_a^d \geq y \,) = F(y) = e^{-e^{-\frac{y-z_a^d}{\beta_i^d}}}$$

Calcoliamo $\mathbb{P}(\, \widetilde{z}_a^d \leq \widetilde{z}_b^d \quad \forall b \in A_i^+)$, che ricordiamo sono uguali alle probabilità di transizione $\frac{\partial \varphi_i^d}{\partial z_a^d}(z^d) = P_{ij}^d$ per l'arco $a = (i,j)$ della catena di Markov.

$$\begin{aligned} P_{ij}^d =& \mathbb{P}(\, \widetilde{z}_a^d \leq \widetilde{z}_b^d \;\; \forall b \in A_i^+ \,) = \\ =& \mathbb{P}(\, z_a^d + \varepsilon_a^d \leq z_b^d + \varepsilon_b^d \;\; \forall b \in A_i^+ \,) = \\ =& \mathbb{P}(\, \varepsilon_b^d \geq \varepsilon_a^d + z_a^d - z_b^d \;\; \forall b \in A_i^+ \,) \end{aligned}$$

Le ε_b^d sono indipendenti per ipotesi, quindi la probabilità composta è il prodotto delle probabilità separate:

$$\begin{aligned} \frac{\partial \varphi_i^d}{\partial z_a^d}(z^d) =& \mathbb{P}(\varepsilon_b^d \geq \varepsilon_a^d + z_a^d - z_b^d \;\; \forall b \in A_i^+) = \\ =& \prod_{b \in A_i^+} \mathbb{P}(\varepsilon_b^d \geq \varepsilon_a^d + z_a^d - z_b^d) = \\ =& \int \mathrm{d}y \prod_{b \in A_i^+,\, b \neq a} \mathbb{P}(\varepsilon_b^d \geq y + z_a^d - z_b^d) \cdot \mathbb{P}(\varepsilon_a^d = y) = \\ =& \int \mathrm{d}y \prod_{b \in A_i^+,\, b \neq a} e^{-e^{-\frac{y-z_a^d}{\beta_i^d}}} \cdot \frac{1}{\beta_i^d}\, e^{-e^{-\frac{y-z_a^d}{\beta_i^d}}-\frac{y-z_a^d}{\beta_i^d}} = \\ =& \frac{1}{\beta_i^d} \int \mathrm{d}y \prod_{b \in A_i^+,} e^{-e^{-\frac{y-z_a^d}{\beta_i^d}}} \cdot e^{\frac{y-z_a^d}{\beta_i^d}} = \\ =& \frac{e^{-\beta_i^d z_a^d}}{\sum_{b \in A_i^+} e^{-\beta_i^d z_b^d}} \end{aligned}$$

Abbiamo ottenuto:

$$P^d_{ij} = \frac{\partial \varphi^d_i}{\partial z^d_a}(z^d) = \mathbb{P}(\widetilde{z}^d_a \leq \widetilde{z}^d_b \;\; \forall b \in A^+_i) = \frac{e^{-\beta^d_i z^d_a}}{\sum\limits_{b \in A^+ - i} e^{-\beta^d_i z^d_b}}$$

Integrando, otteniamo anche un'espressione esplicita per φ^d_i:

$$\varphi^d_i(z^d) = -\frac{1}{\beta^d_i} \ln\left(\sum_{a \in A^+_i} e^{-\beta^d_i z^d_a}\right)$$

Questa scelta viene chiamata *modello Logit*, proprio perché le derivate parziali hanno la stessa espressione della funzione logistica multinomiale.
Se al posto di scegliere variabili Gumbel, si scegliesse una distribuzione normale, le espressioni di φ e delle sue derivate parziali sarebbero secondo il modello Probit. Questo, tuttavia, non ha un'espressione esplicita, per cui viene calcolato tramite Monte Carlo; questo diventa troppo costoso per reti di medio-grandi dimensioni.

5.3 Algoritmo

Per risolvere il problema 8, vogliamo trovare il minimo della funzione $\Phi(t)$, che chiamiamo t^* (con relativo w^*). L'algoritmo è il seguente:

- Si inizializzano i tempi di percorrenza degli archi $t_a = t^{(0)}_a$ con valori casuali.
- Per ogni destinazione $d = 1, 2, \ldots, n$:
  - Viene calcolato il punto fisso $\tau^d_i = \varphi^d_i(\{t_a + \tau^d_{j_a}\}_{a \in A^+_i})$. Questo viene fatto tramite un procedimento iterativo su τ^d.
  - Si calcola $\frac{\partial \varphi^d_i}{\partial z^d_a}(z^d)$.
  - Si costruiscono le matrici P^d e Q^d.
  - Vengono calcolati i flussi $x^d = (I - (P^d)^\top)^{-1} g^d$ e $v^d = (Q^d)^\top x^d$.
- Si aggregano i flussi per ogni destinazione $\widehat{w} = \sum_d v^d$
- Si usa un metodo del gradiente per calcolare i nuovi tempi di percorrenza $t^{(k+1)}$ a partire da $t^{(k)}$.
- Si itera fino a raggiungere la convergenza.

5.4 Metodi di discesa

Per il penultimo passo, ricordiamo che per $\Phi(t)$, le sue derivate parziali sono date da

$$\frac{\partial \Phi}{\partial t_a}(t) = s_a^{-1}(t_a) - \widehat{w}_a(t)$$

Un'implementazione del metodo del gradiente è la seguente:

$$t_a^{(k+1)} = t_a^{(k)} - \alpha_{(k)} \left(s_a^{-1}(t^{(k)})_a - \widehat{w}_a^{(k)} \right) \tag{19}$$

Dove $\alpha_{(k)}$ viene scelto opportunamente, per esempio

$$\alpha_{(k)} = \frac{\lambda_k}{|s_a^{-1}(t^{(k)})_a - \widehat{w}_a^{(k)}|}$$

con $\lambda_k \to 0$ e $\sum_k \lambda_k = \infty$. Secondo quanto riportato nell'articolo, il metodo converge (sia teoricamente che sperimentalmente), ma la convergenza è lenta.

Un'altra possibilità consiste nel fare un cambio di variabile, e lavorare con $\Psi(w) = \Phi(s(w)) = \Phi(t)$. Ψ, a differenza di Φ, non è più convessa; tuttavia, un minimo di Ψ corrisponde a un minimo di Φ (e viceversa). Dunque, un minimo di Ψ è necessariamente un punto stazionario, dunque ci basta cercare i punti stazionari di Ψ con $\nabla \Psi = 0$. Le derivate parziali di Ψ sono:

$$\frac{\partial \Psi}{\partial t_a}(w) = (w_a - \widehat{w}_a(s(w))) \cdot s'_a(w_a)$$

Per cui il metodo del gradiente diventa:

$$w_a^{(k+1)} = w_a^{(k)} - \alpha_{(k)} \left(w_a^{(k)} - \widehat{w}_a^{(k)} \right) s'_a(w_a) \tag{20}$$

Gli autori dell'articolo hanno osservato convergenza anche per questo metodo, però molto lenta. Hanno proposto la seguente modifica, che tralascia il termine della chain rule nel calcolo della derivata

$$w_a^{(k+1)} = w_a^{(k)} - \alpha_{(k)} \left(w_a^{(k)} - \widehat{w}_a^{(k)} \right) \tag{21}$$

Questo metodo iterativo viene denominato *MSA, Method of Successive Averages*. Infatti i pesi alla $k+1$ esima iterazione si ottengono pesando opportunamente i pesi della k-esima, $w^{(k)}$, con i pesi $\widehat{w}^{(k)}$ ottenuti imponendo l'equilibrio del sistema.
Il metodo MSA è quello che ha dato loro risultati migliori, e di cui riportano i grafici. Gli esperimenti sono stati fatti sulla rete stradale di Sioux Falls (24 vertici) e di Chicago (933 vertici), prese da [4].

Una variante dell'algoritmo MSA presentato combina la parte lineare, con un metodo di Newton una volta raggiunto il livello di precisione del 10%, ovvero $\|\widehat{w}^k - w^k\| \leq 0.1\|w^k\|$. Questo accorgimento velocizza l'esecuzione: per la rete di Sioux Falls, si passa da 3.7 secondi a 0.7 secondi, mentre per Chicago sono 29 minuti con MSA, e 11 per MSA con Newton.

5.5 Convergenza del MSA

Mostriamo che l'algoritmo MSA converge alla soluzione cercata.

Proposizione 9 (Theorem 2.8)**.** Sia $\alpha_{(k)}$ una sequenza di reali positivi convergenti a 0, che soddisfano $\sum\limits_k \alpha_{(k)} = \infty$ e $\sum\limits_k \alpha^2_{(k)} < \infty$. Supponiamo inoltre che le funzioni s_a siano di classe $\mathcal{C}^1$. Sia $\{\, w^{(k)} \,\}_{k\in\mathbb{N}}$ una sequenza di pesi generata dall'iterazione MSA, ovvero

$$w^{(k+1)} = (1-\alpha_{(k)})\, w^{(k)} + \alpha_{(k)} \widehat{w}^k$$

Assumiamo che $w^{(k)}$ sia limitata. Allora la sequenza converge alla soluzione MTE w^*.

Dimostrazione. Chiamati $t^{(k)} = s_a(w^{(k)})$, si ha che:

$$\frac{w^{(k+1)} - w^{(k)}}{\alpha_{(k)}} = -w^{(k)} + \widehat{w}^k = -\nabla\Phi(t^{(k)}) \tag{22}$$

Essendo Φ convessa, scriviamo la relazione di sottogradiente nel punto $t^{(k)}$ e usando come secondo punto t^*:

$$\begin{gathered}
\Phi(t^{(k)}) + \left\langle \frac{w^{(k+1)} - w^{(k)}}{\alpha_{(k)}}, t^k - t^* \right\rangle \leq \Phi(t^*) \\
\big(\Phi(t^{(k)}) - \Phi(t^*)\big)\ \alpha_{(k)} + \left\langle w^{(k+1)} - w^{(k)}, t^k - t^* \right\rangle \leq 0
\end{gathered}$$

Definiamo $h(w) = \sum\limits_{a\in A} \int\limits_{w_a^*}^{w_a} (s_a(z) - s_a(w_a^*))\ \mathrm{d}z$. Allora $\frac{\partial h}{\partial w_a} h(w^{(k)}) = s_a(w_a^{(k)}) - s_a(w_a^*) = t_a^{(k)} - t_a^*$; scritto in forma vettoriale è $\nabla_w h(w^{(k)}) = t^{(k)} - t^*$.
Si ottiene:

$$\big(\Phi(t^{(k)}) - \Phi(t^*)\big)\ \alpha_{(k)} + \left\langle \nabla_w h(w^{(k)}), t^k - t^* \right\rangle \leq 0 \tag{23}$$

Usando il resto dell'espansione di Taylor, esiste $\xi^{(k)} \in [w^{(k)}, w^{(k+1)}]$ tale che

$$\begin{aligned}
h(w^{(k+1)}) = {}& h(w^{(k)}) + \left\langle \nabla_w h(w^{(k)}), w^{(k+1)} - w^{(k)} \right\rangle + \\
&+ \frac{1}{2} \left\langle \nabla^2_w h(\xi^{(k)})(w^{(k+1)} - w^{(k)}),\ w^{(k+1)} - w^{(k)} \right\rangle
\end{aligned}$$

Osserviamo che $\frac{\partial h}{\partial w_a}$ dipende solo da w_a, per cui $\nabla^2_w h$ è diagonale, con entrata a-esima $\frac{\partial^2 h}{\partial w_a^2} = s'_a(w_a)$. Siccome i $w^{(k)}$ sono limitati, anche $\nabla^2_w h$ è limitata, dunque esiste una costante $\gamma \geq \|\nabla^2_w h\|$ tale che:

$$h(w^{(k+1)}) \leq h(w^{(k)}) + \left\langle \nabla_w h(w^{(k)}),\ w^{(k+1)} - w^{(k)} \right\rangle + \gamma |w^{(k+1)} - w^{(k)}|^2 \tag{24}$$

Mettendo insieme le disequazioni 23 e 24 si ha:

$$\begin{aligned}
&\big(\Phi(t^{(k)}) - \Phi(t^*)\big)\ \alpha_{(k)} + \big(h(w^{(k+1)}) - h(w^{(k)})\big) \leq \\
\leq &\big(\Phi(t^{(k)}) - \Phi(t^*)\big)\ \alpha_{(k)} + \left\langle \nabla_w h(w^{(k)}),\ w^{(k+1)} - w^{(k)} \right\rangle + \gamma |w^{(k+1)} - w^{(k)}|^2 \\
\leq &0 + \gamma |w^{(k+1)} - w^{(k)}|^2
\end{aligned} \tag{25}$$

Siccome $t_a^{(k)} = s_a(w_a^{(k)})$ è limitato, allora $|\nabla\Phi(t^{(k)})|$ è limitato superiormente da una certa costante β (infatti Φ è convessa e coerciva). Allora si ha dall'equazione 22:

$$|w^{(k+1)} - w^{(k)}| = \alpha_{(k)} |\nabla\Phi(t^{(k)})| \leq \beta\alpha_{(k)}$$

La stima diventa:

$$\big(\Phi(t^{(k)}) - \Phi(t^*)\big) \alpha_{(k)} + \big(h(w^{(k+1)}) - h(w^{(k)})\big) \leq \gamma\, |w^{(k+1)} - w^{(k)}|^2 \leq \gamma\, \beta^2\, \alpha_{(k)}^2 \tag{26}$$

Il primo termine è positivo in quanto $\Phi(t^{(k)}) \geq \Phi(t^*)$, dunque si ha:

$$0 \leq h(w^{(k+1)}) \leq h(w^{(k)}) + \gamma\beta^2\alpha_{(k)}^2$$

Siccome $\alpha_{(k)} \to 0$, si ha che $h(w^{(k)}) \to 0$. Invece sommando le disuguaglianze al variare di k, si ottiene una somma telescopia che cancella gli $h(w^{(k)})$ e resta:

$$\sum_{k=0}^{\infty} \big(\Phi(t^{(k)}) - \Phi(t^*)\big) \alpha_{(k)} \leq h(w^{(0)}) + \gamma\beta^2 \sum_{k=0}^{\infty} \alpha_{(k)}^2 < \infty$$

Se fosse $\liminf \Phi(t^{(k)}) - \Phi(t^*) = d > 0$, allora la somma $\sum_{k=0}^{\infty} \big(\Phi(t^{(k)} - \Phi(t^*)\big) \alpha_{(k)}$ sarebbe definitivamente maggiore di $d \cdot \sum_k \alpha_{(k)}$, che è un assurdo perché $\sum_k \alpha_{(k)}$ diverge.
Quindi $\liminf \Phi(t^{(k)}) = \Phi(t^*)$ (in quanto t^* è il minimo assoluto della funzione). Per convessità e coercività di Φ, i suoi sottolivelli sono limitati, quindi esiste una sottosuccessione $t^{(k_j)} \to t^*$. Dunque $w^{(k_j)} \to w^*$ e quindi

$$\lim_k h(w^{(k)}) = \lim_j h(w^{(k_j)}) = 0$$

Tuttavia dalla definizione di h, questo è possibile solo se i w^k tendono all'unico punto in cui $h(w) = 0$, ovvero w^*. La tesi è dimostrata.

Proposizione 10 (Theorem 2.10)**.** Supponiamo che $s_a(\cdot)$ e $s'_a(\cdot)$ siano limitate superioremente. Allora valgono le conclusioni della proposizione 9, senza l'assunzione che $w^{(k)}$ sia limitata.

Nella dimostrazione, si è usata la limitatezza di $w^{(k)}$ per dimostrare che t^k e $\nabla^2 h(\xi^{(k)})$ fossero limitati.
Se $s_a(\cdot)$ è limitata, allora anche $t_a^{(k)} = s_a(w_a^{(k)})$ lo è. Inoltre $\nabla^2 h(\xi^{(k)})$ contiene termini del tipo $s'_a(\xi^{(k)})$, dunque la limitatezza di $s'_a(\cdot)$ implica che $\nabla^2 h(\xi^{(k)})$ siano limitati.

Riferimenti bibliografici

[1] Baillon J.-B., Cominetti R., *Markovian traffic equilibrium*, Mathematical Programming 111(1-2), Ser. B (2008), 33–56.

[2] Cominetti R., Facchinei F., Lasserre J.B., *Modern Optimization Modelling Techniques*, Birkhäuser Basel (2012).

[3] Dial R.B., *A probabilistic multipath traffic assignment model which obviates path enumeration*, Transport. Res. 5 (1971), 83–111.

[4] Bargera H. et al. Transportation Networks repository, https://github.com/bstabler/TransportationNetworks/