# Sensitivity analysis and relaxation of the Static Traffic Assignment Problem with Capacity Constraints

Théophile Cabannes*†, Elizabeth Glista*, Kshama Dwarakanath*, Xu Rao*, Tanya Veeravalli*, Alexandre M.Bayen*
*Department of Electrical Engineering and Computer Science, University of California at Berkeley, CA
†LIX, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris

***Abstract*— This article introduces sensitivity analysis, reduction of the feasible set around the optimal solution, and LP and QP relaxations on convex, capacity-constrained network flow problems to evaluate the impact of a change in link capacity on the optimal flow allocation. This is done in the context of the static traffic assignment problem under capacity constraints [TAP-C], to understand the impact of traffic incidents on traffic flow in road networks, though the results also apply beyond transportation.**

**The dual formulation of the convex [TAP-C] using generalized travel costs is exploited to show that dual variables associated to the link capacity constraints can be used to understand sensitivity of link flows to perturbation. To avoid expensive computations of the [TAP-C] optimal flow solution, the impact of a perturbation of link capacities on the optimal flow allocation is shown to be lower-bounded by a dual sensitivity analysis, and upper-bounded by LP and QP relaxations of the [TAP-C] on a reduced set of feasible flow allocations around the optimal solution.**

**Simulations are conducted on a benchmark city network (the Sioux Falls network with 75 links) to assess the impact of an incident on the traffic flow allocation. When one road (2 links) is closed, the total vehicle hours of travel (VHT) doubles in the network. For this network, the computation of the solution through relaxation is shown to be 300 times faster than recomputing the [TAP-C].**

## I. Introduction

### A. Motivation

Traffic congestion has high costs for modern society: it results in lost efficiency, wasted fuel, and added stress. In 2018, the transportation data company INRIX attempted to quantify these losses in a study on the cost of congestion [1]. Taking into account direct costs like wasted fuel and time as well as indirect costs like price increases for goods shipped by trucks, INRIX calculated an annual loss of around $305 billion in the U.S.. This article aims to help improve our understanding of how traffic patterns respond to traffic incidents. Better predictions of the evolution of traffic patterns during traffic incidents will help to better allocate emergency resources to alleviate congestion.

To this end, this article uses techniques such as sensitivity analysis, reduction of the feasible set around the optimal solution, and linear programming (LP) and quadratic programming (QP) relaxations for the network flow control problem. The applications presented involve modeling traffic flow pattern response to traffic incidents. Other network flow applications involving perturbations in link capacity can be derived from this work. For instance, in packet-switched networks, available capacity on both the client and server sides is limited by a variable bottleneck link (as can be seen in [2]), prompting an optimal rerouting of flow.

### B. Approach and contributions

This article aims to improve dynamic control of network flow when incidents occur, by efficiently assessing the impact of link capacity perturbations on the optimal flow allocation. First, the network flow control problem – the traffic assignment problem with capacity constraints [TAP-C] [3] – is introduced in Section II. Since computing the optimal flow allocation through a gradient descent is computationally expensive (up to 3,000 gradient descent steps for a 75-link network, see [3], [4], [5]), the new optimal flow allocation after a perturbation of link capacities is computed from the old optimal one using sensitivity analysis and LP and QP relaxations (Section III). For the specific scenario of Section IV, current results show that computing the link flow changes due to perturbation of link capacities through relaxation is 300 times faster than recomputing the new convex problem.

The contributions of this article include:

1) Sensitivity analysis to obtain an approximation of the variation in optimal flow allocation when link capacities are perturbed in the [TAP-C].
2) Interpretation of dual variables of the [TAP-C] problem for understanding sensitivity of link flows with respect to link capacities.
3) Improvement in computation speed to model flow changes due to decreases in link capacities using LP and QP “relaxations” on a subset of feasible flow allocation around the optimal solution.
4) Demonstration of the impact of a link removal on the optimal flow allocation on a benchmark network (75 links).

Applications to traffic are presented in order to model the impact of a traffic incident on the road traffic flow allocation and to help better allocate resources (like price-directive traffic control schemes).

## II. The Static Traffic Assignment with Capacity constraints

The following section first introduces the static traffic assignment (TAP) framework and notation. Next, the static traffic assignment with capacity constraints (TAP-C) is presented to model the impact of a link capacity change on the network optimal flow allocation. Then, the dual formulation of the TAP-C is analyzed in order to provide meaningful information on the potential impact and recovery measures.

### A. The Static Traffic Assignment [4]

*1) Framework and notation:* The static traffic assignment is a network flow convex problem defined on a strongly connected directed graph $G$ (the network) with vertices in $\mathcal{V}$ and links in $\mathcal{A}$: $G = (\mathcal{V}, \mathcal{A})$. A proportion of flow travels from an origin $o \in \mathcal{V}$ to a destination $d \in \mathcal{V}$, contributing to the route demand of this $od$ pair. We use the following notations:

- $\mathcal{A}$ Set of all links (arcs) in the network $\mathcal{G}$.
- $\mathcal{K}$ Set of all origin-destination ($od$) pairs in the network.
- $\mathcal{R}_k$ Set of all routes for an $od$ pair $k \in \mathcal{K}$.
- $\mathcal{R} = \cup_{k\in\mathcal{K}}\mathcal{R}_k$ Set of all routes that can be used in the network.
- $f_a \in \mathbb{R}_+$ Flow on link $a \in \mathcal{A}$.
- $h_r \in \mathbb{R}_+$ Flow on route $r \in \mathcal{R}$.
- $d_k \in \mathbb{R}_{++}$ Demand from origin node to destination node in O-D pair $k \in \mathcal{K}$.
- $\Delta$ Link-Route incidence matrix $\Delta \in \mathbb{R}^{|\mathcal{A}|\times|\mathcal{R}|}$.
- Every vectorized form is noted in bold (like $\mathbf{f}$ and $\mathbf{h}$).

The link-route incidence matrix is defined as $\Delta^T = (\delta_{ra})_{r\in\mathcal{R}, a\in\mathcal{A}}$ where $\delta_{ra} := \begin{cases} 1, & \text{if route } r \text{ uses link } a \\ 0, & \text{otherwise} \end{cases}$.

*2) Feasible flow allocation and Wardrop conditions:* In the static traffic assignment, the demand is fixed. At the equilibrium, the route flows no longer evolve, implying that route flow is uniformly distributed among the links that compose the route. Then, for every link $a \in \mathcal{A}$ we have: $f_a = \sum_{k\in\mathcal{K}}\sum_{r\in R_k} \delta_{ra} h_r$. In vectorized form, we have $\mathbf{f} = \Delta\mathbf{h}$.

A flow allocation $\mathbf{h}$ is a feasible flow allocation if all the demand is distributed among the different routes:

$$\forall k \in \mathcal{K},\ d_k = \sum_{r\in\mathcal{R}_k} h_r$$

We define the cost on each link $a \in \mathcal{A}$, $t_a$ as a function of the link flow $f_a$: $t_a(f_a)$. In the traffic context, this cost will be taken as the travel time. Then, we define the route cost $c_r$ of a route $r$ as the sum of the cost of the links that compose the route. The cost function $t_a : \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be continuous and strictly increasing for each $a \in \mathcal{A}$.

We note:

- $t_a(\cdot)$ is the cost on link $a \in \mathcal{A}$ which is a function of the flow $f_a$ on link $a$.
- $c_r(\cdot)$ is the cost on route $r \in \mathcal{R}$, also a function of the link flow allocation $\mathbf{f} = \Delta\mathbf{h}$.
- $\tilde{\pi}_k(\cdot)$ Cost of the shortest route for an O-D pair $k \in \mathcal{K}$, also a function of the link flow allocation $\mathbf{f}$.

Wardrop's first condition [6] states that the equilibrium state is attained when: "the journey times on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route":

$$\forall k \in \mathcal{R},\ \forall r \in \mathcal{R}_k,\ h_r > 0 \implies c_r(\mathbf{f}) = \tilde{\pi}_k(\mathbf{f})$$

*3) Traffic assignment primal formulation:* In 1956, Beckmann et al. showed that the Wardrop's first condition leads to a convex formulation of the problem as follows [7]:

$$\text{[TAP]} \quad \min_{\mathbf{f},\mathbf{h}} \sum_{a\in\mathcal{A}} \int_0^{f_a} t_a(s)ds \tag{1}$$

subject to

$$\sum_{k\in\mathcal{K}}\sum_{r\in R_k} \delta_{ra} h_r = f_a \qquad \forall a \in \mathcal{A} \tag{2}$$

$$\sum_{r\in\mathcal{R}_k} h_r = d_k \qquad \forall k \in \mathcal{K} \tag{3}$$

$$h_r \geq 0 \qquad \forall r \in \mathcal{R}_k, \forall k \in \mathcal{K} \tag{4}$$

The Karush-Kuhn-Tucker (KKT) optimality conditions for the [TAP] correspond to the Wardrop equilibrium conditions.

### B. Primal form of [TAP-C] [3]

We model incidents on links by reducing the corresponding link capacity. The optimization formulation for [TAP-C] is derived from [TAP] as follows. Letting $u_a \in \mathbb{R}_+$ denote the capacity of link $a$ for all $a \in \mathcal{A}$, [TAP-C] is [TAP] with the additional link capacity constraints [3]:

$$\text{[TAP-C]} \quad \min_{\mathbf{f},\mathbf{h}} \sum_{a\in\mathcal{A}} \int_0^{f_a} t_a(s)ds$$

subject to

$$\sum_{k\in\mathcal{K}}\sum_{r\in R_k} \delta_{ra} h_r = f_a \qquad \forall a \in \mathcal{A}$$

$$\sum_{r\in\mathcal{R}_k} h_r = d_k \qquad \forall k \in \mathcal{K}$$

$$h_r \geq 0 \qquad \forall r \in \mathcal{R}_k, \forall k \in \mathcal{K}$$

$$f_a \leq u_a \qquad \forall a \in \mathcal{A}$$

The [TAP-C] can be unfeasible if capacity constraints $\mathbf{u}$ are incompatible with demand $\mathbf{d}$. In the remainder of this article, we study problems in which [TAP-C] is feasible.

When there is a traffic incident, we want to predict how it will affect the traffic flow. The [TAP-C] problem is a convex optimization problem with a strictly convex objective function with respect to link flows $\mathbf{f}$. However, because the size of $\mathcal{R}$ scales exponentially with the number of links in the network [8], [TAP-C] cannot be solved directly by solvers unless the network size is very small. Existing optimization algorithms, based on the projected gradient descent algorithm, are also computationally expensive [5]. These algorithms can take a long time to converge to the optimal solution [5]. Therefore, solving the new [TAP-C] problem may be time-consuming and may not provide near-instantaneous solutions.

### C. Dual form of [TAP-C]

We associate the dual constraints $\mu_a \in \mathbb{R}$ to $\sum_{k\in\mathcal{K}}\sum_{r\in R_k} \delta_{ra} h_r = f_a$, $\pi_k \in \mathbb{R}$ to $\sum_{r\in\mathcal{R}_k} h_r = d_k$, $\nu_a \in \mathbb{R}_+$ to $f_a \leq u_a$ and $\lambda_r \in \mathbb{R}_+$ to $h_r \geq 0$. We define $T_a(f_a) =$

$\int_0^{f_a} t_a(s)ds$ and $T(\mathbf{f}) = \sum\limits_{a\in\mathcal{A}} T_a(f_a)$. Then, the Lagrangian for [TAP-C] is:

$$\begin{aligned}L(\mathbf{f},\mathbf{h},\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\pi},\boldsymbol{\lambda}) = &\sum_{a\in\mathcal{A}}\Big(T_a(f_a) + (-\mu_a+\nu_a)f_a\Big)\\ &+\sum_{k\in\mathcal{K}}\sum_{r\in\mathcal{R}_k}\Big(\sum_{a\in\mathcal{A}}\mu_a\delta_{ra} - \lambda_r - \pi_k\Big)h_r\\ &+\sum_{k\in\mathcal{K}}\pi_k d_k - \sum_{a\in\mathcal{A}}\nu_a u_a\end{aligned}$$

The dual function is given by:

$$g(\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\pi},\boldsymbol{\lambda}) = \inf_{\mathbf{f},\mathbf{h}} L(\mathbf{f},\mathbf{h},\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\pi},\boldsymbol{\lambda})$$

If, $\forall k\in\mathcal{K}$ and $\forall r\in\mathcal{R}_k$, $\pi_k \le \sum_{a\in\mathcal{A}}\mu_a\delta_{ra} - \lambda_r$:

$$g(\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\pi},\boldsymbol{\lambda}) = -\sum_{a\in\mathcal{A}}\tilde{T}_a(\mu_a-\nu_a) + \sum_{k\in\mathcal{K}}\pi_k d_k - \sum_{a\in\mathcal{A}}\nu_a u_a$$

Else:

$$g(\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\pi},\boldsymbol{\lambda}) = -\infty$$

where $\tilde{T}_a$ is the Fenchel conjugate function of $T_a$ [9]. The dual problem can be expressed as

$$\begin{aligned}\max_{\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\pi},\boldsymbol{\lambda}}\; &-\sum_{a\in\mathcal{A}}\tilde{T}_a(\mu_a-\nu_a) + \sum_{k\in\mathcal{K}}\pi_k d_k - \sum_{a\in\mathcal{A}}\nu_a u_a\\ \text{s.t. } &\pi_k \le \sum_{a\in\mathcal{A}}\mu_a\delta_{ra} - \lambda_r,\ \ \forall r\in\mathcal{R}_k, \forall k\in\mathcal{K}\\ &\boldsymbol{\nu}\ge 0\\ &\boldsymbol{\lambda}\ge 0\end{aligned}$$

Solving the above maximization problem with respect to $\pi_k$, since $d_k > 0$ for all $k\in\mathcal{K}$, the optimal value for the dual variable $\pi_k$ is:

$$\pi_k^\star(\boldsymbol{\mu},\boldsymbol{\lambda}) = \min_{r\in\mathcal{R}_k}\left\{\sum_{a\in\mathcal{A}}\mu_a\delta_{ra} - \lambda_r\right\}$$

which can be interpreted as the minimum cost/travel time for the O-D pair $k$ by adding the link travel times $\boldsymbol{\mu}$. Therefore, the dual problem for static traffic assignment with capacity constraints can be written as follows [10]:

$$\begin{aligned}\text{[D-TAP-C]}\quad \max_{\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\lambda}}\; &-\sum_{a\in\mathcal{A}}\tilde{T}_a(\mu_a-\nu_a) + \sum_{k\in\mathcal{K}}\pi_k^\star(\boldsymbol{\mu},\boldsymbol{\lambda})d_k\\ &-\sum_{a\in\mathcal{A}}\nu_a u_a \quad \text{s.t. } \boldsymbol{\nu}\ge 0,\ \boldsymbol{\lambda}\ge 0\end{aligned}$$

### D. Karush–Kuhn–Tucker conditions of [TAP-C]

Since [TAP-C] is a convex optimization problem, the KKT conditions are necessary and sufficient conditions for the optimality of $\mathbf{f}$ for the primal problem and that of $(\boldsymbol{\mu},\boldsymbol{\nu},\boldsymbol{\pi})$ for the dual problem [9].

Defining the "generalized route travel cost" as in [3]:

$$\bar{c}_r(\mathbf{f}) := \sum_{a\in\mathcal{A}}\left(t_a(f_a)+\nu_a\right)\delta_{ra} \qquad \forall r\in\mathcal{R}_k, \forall k\in\mathcal{K}$$

KKT conditions gives that:

$$\begin{aligned}h_r > 0 \Rightarrow \bar{c}_r(\mathbf{f}) = \pi_k \qquad &\forall r\in\mathcal{R}_k, \forall k\in\mathcal{K}\\ h_r = 0 \Rightarrow \bar{c}_r(\mathbf{f}) \ge \pi_k \qquad &\forall r\in\mathcal{R}_k, \forall k\in\mathcal{K}\end{aligned}$$

i.e. for an O-D pair, all routes with nonzero flow have the same generalized travel cost, and all unused routes have a higher cost. Therefore, the solution with capacity constraints satisfies the Wardrop's first condition with respect to the generalized travel cost [3].

### E. Interpretation of the dual variables

Like for the [TAP] [4], dual variables can be interpreted:

1) $\nu_a$: Equilibrium queuing delay on saturated links [3]. The difference between the generalized travel cost and $t_a(f_a)$ (the actual cost on link $a$) can be interpreted as the queuing delay due to saturation. For unsaturated links we have $f_a < u_a$, giving $\nu_a = 0$, or zero queuing delay. For this traffic context, given the interpretation of the variable $\nu_a$, we can use it to construct price-directive traffic control schemes with link tolls to limit maximal traffic volume on some links (the higher $\nu_a$ is, the more expensive the road price or the closer the crisis resources should be). Additionally, it can be used to direct the allocation of emergency resources by indicating critical links for improving overall travel cost: the higher $\nu_a$ is, the more sensitive it is to any capacity change.
2) $\pi_k$: Minimum generalized travel cost which is the sum of the route cost and the queuing delay in each O-D pair.
3) The solution set (if non-empty) of [TAP-C] is equivalent to that of [TAP] with link travel cost $t_a(\cdot)+\nu_a$ for each link $a\in\mathcal{A}$ with $\boldsymbol{\nu}$ as the set of optimal Lagrange multipliers for the capacity constraints.

## III. Evaluation of the impact of capacity changes using dual variables and linear and quadratic relaxations

This section provides evaluations of the evolution of the flow assignment when there is a perturbation of the capacity constraints. These evaluations are computationally less expensive than computing the new [TAP-C] optimal solution (section IV). First, a lower bound resulting from a sensitivity analysis of the dual program is presented. Then, upper bounds using LP and QP relaxations inside a subset of feasible flows around the optimal solution are obtained.

### A. A lower bound - sensitivity analysis

For every link $a\in\mathcal{A}$ the capacity constraint on the link flow in the [TAP-C] is $f_a \le u_a$.

We model a change in link capacity, such as one due to a traffic accident or a worsening of road conditions, as a decrease in the corresponding capacity constraint. The perturbed [TAP-C] capacity constraint is

$$f_a \le u_a - \delta u_a \qquad\qquad \forall a\in\mathcal{A}$$

where $\delta u_a \geq 0$ is the magnitude of the reduction in the capacity of link $a$ and is assumed to be known.

Consider the Lagrangian for the [TAP-C] where only capacity constraints are dualized:

$$L(\mathbf{f}, \boldsymbol{\nu}, \mathbf{u}) = T(\mathbf{f}) + \boldsymbol{\nu}^T(\mathbf{f} - \mathbf{u})$$

Let $T^\star(\mathbf{u})$ be the optimal objective value of the original [TAP-C] problem (see Section II-B) and let $T^\star(\mathbf{u} - \boldsymbol{\delta}\mathbf{u})$ be the optimal objective value of the perturbed [TAP-C] problem (with $\boldsymbol{\delta}\mathbf{u} = (\delta u_a)_{a\in\mathcal{A}}$ given).

Denote by $\mathcal{F}$ the set of all feasible flows to the [TAP] without capacity constraints, so that the optimal primal and dual objective values can be given as

$$T^\star(\mathbf{u}) = \min_{\mathbf{f}\in\mathcal{F}} \max_{\boldsymbol{\nu}\geq 0} L(\mathbf{f}, \boldsymbol{\nu}, \mathbf{u})$$
$$d^\star(\mathbf{u}) = \max_{\boldsymbol{\nu}\geq 0} \min_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \boldsymbol{\nu}, \mathbf{u})$$

Since strong duality holds due to the convexity of the primal objective function,

$$T^\star(\mathbf{u}) = d^\star(\mathbf{u}) = \min_{\mathbf{f}\in\mathcal{F}} L(\mathbf{f}, \boldsymbol{\nu}(\mathbf{f}, \mathbf{u}), \mathbf{u})$$
$$\text{where } \boldsymbol{\nu}(\mathbf{f}, \mathbf{u}) = \arg\max_{\boldsymbol{\nu}\geq 0} L(\mathbf{f}, \boldsymbol{\nu}, \mathbf{u})$$

The optimal dual variables $\boldsymbol{\nu}^\star = \boldsymbol{\nu}(\mathbf{f}^\star, \mathbf{u})$ can be computed with the primal solution $\mathbf{f}^\star$ using the KKT conditions (Section II-D). The sensitivity analysis, as in [9], of the capacity perturbation provides a lower bound on the change in equilibrium cost when the capacity of links is decreased:

$$T^\star(\mathbf{u} - \boldsymbol{\delta}\mathbf{u}) \geq T^\star(\mathbf{u}) + {\boldsymbol{\nu}^\star}^T \boldsymbol{\delta}\mathbf{u}$$

We see that for a reduction of $\delta u_a$ in the capacity of the link $a$ in the traffic network, the optimal cost increases proportionally to the amount of capacity reduction $\delta u_a$ on saturated link $a$ (corresponding to $\nu_a^\star > 0$). In other words, the lower bound scales with the decrease in link capacity $\delta u_a$, as well as the optimal dual parameter $\nu_a^\star$, which corresponds to the equilibrium queuing delay on the saturated link $a$.

The equation $T^\star(\mathbf{u} - \boldsymbol{\delta}\mathbf{u}) \geq T^\star(\mathbf{u}) + {\boldsymbol{\nu}^\star}^T \boldsymbol{\delta}\mathbf{u}$ can be seen as a first order approximation of $T^\star(\mathbf{u} - \boldsymbol{\delta}\mathbf{u})$. We previously saw that an interpretation of the dual variable is the equilibrium queuing delay of a link. Another interpretation (in the traffic context) of $\nu_a$ is the time that one would save by choosing this saturated link $a$. Since the cost is the gradient of the objective function with respect to the link flow (Section II-A.3), the amount that one can save by using a saturated link is the gradient of the objective function with respect to the dual variable. We see this because rerouting a $\delta u_a$ amount of flow on an unsaturated link will increase the cost of this link; this flow could not change its cost on the link by less than ${\nu_a^*}^\top \delta u_a$. This explains why ${\boldsymbol{\nu}^*}^T \boldsymbol{\delta u}$ is a lower bound of the perturbation of the [TAP-C], which will be achieved if rerouting a $\boldsymbol{\delta u}$ amount of flow on unsaturated links will not change the cost on these links.

Another interpretation of the dual variables is that it is an evaluation of how saturated links are. The dual variables can help traffic engineers understand where a traffic incident will have the most impact (on links with highest values for dual variables). This can help to better allocate crisis resources (variable-message signs for example): the greater $\nu_a$ is, the more sensitive to perturbation the link $a$ is. The sensitivity analysis also provides an evaluation of the maximum efficiency gain that could be made by increasing the road capacity of a road.

### *B. An upper bound - feasibility analysis, LP and QP relaxations*

In order to find an upper bound on the change in the optimal travel time due to a reduction in capacity, we use the fact that the primal [TAP-C] problem is a minimization problem. Therefore, any feasible flow allocation provides an upper bound on the optimal objective value.

The optimal link flow allocation $\mathbf{f}^\star = \Delta\mathbf{h}^\star$ for the original [TAP-C] may be infeasible for the new [TAP-C]. Therefore, we approximate the optimal solution $\mathbf{f}^\star_{\boldsymbol{\delta u}}$ of the new [TAP-C] with LP and QP approximations of the objective function around the optimal solution $\mathbf{f}^\star$ of the original [TAP-C].

For each origin-destination pair $k \in \mathcal{K}$, we partition the set of all routes $r \in \mathcal{R}_k$ into those that contain saturated links – $a \in \mathcal{A}$ such that $f_a^\star > u_a - \delta u_a$ – (called saturated routes) and those that contain link flows which are all strictly feasible (called unsaturated routes).

We note $\bar{\Delta}$ the incidence matrix for the saturated routes: $\bar{\Delta} = (\bar{\delta}_{r,a})_{r\in\mathcal{R}, a\in\mathcal{A}}$ where:

$$\bar{\delta}_{ra} = \begin{cases} 1, & \text{if } \delta_{ra} = 1 \text{ and } f_a^\star > u_a - \delta u_a \\ 0, & \text{otherwise} \end{cases}$$

The total link flow due to saturated routes is then: $\bar{\mathbf{f}} = \bar{\Delta}\mathbf{h}^\star$. We just need to reallocate the flow from saturated routes to unsaturated routes in order to obtain a feasible flow allocation of the [TAP-C]. We denote $\bar{\mathbf{d}}$ the corresponding demand to reallocate:

$$\bar{d}_k = \sum_{r\in\mathcal{R}_k} h_r \cdot \max_{a\in\mathcal{A}} \bar{\delta}_{ra}$$

Then, we note $\bar{\mathcal{F}}_{\delta u}$ the feasible new flow allocation after reallocation:

$$\bar{\mathcal{F}}_{\delta u} = \left\{ (\mathbf{f}^\star - \bar{\mathbf{f}}) + \Delta\mathbf{h} \mid \forall \mathbf{h} \text{ s.t. } \forall k \in \mathcal{K}, \bar{d}_k = \sum_{r\in\mathcal{R}_k} h_r \right\}$$

The convexity of the travel time function gives that $\mathbf{f}^\star_{\delta u} \in \bar{\mathcal{F}}_{\delta u}$ [11]. We reduce the set of feasible flow allocations around $\mathbf{f}^\star_{\delta u}$ using our knowledge about the optimal flow allocation $\mathbf{f}^\star$ of the original [TAP-C]. The optimal solution $\mathbf{f}^\star_{\delta u}$ of the new [TAP-C] is computed faster using a projected gradient descent only inside $\bar{\mathcal{F}}_{\delta u}$ (Section IV).

To avoid doing a gradient descent algorithm, which might be computationally expensive, two possible approximations are mentioned below along with the corresponding optimal flows.

- The LP "relaxation" approximation: the following approximation of the objective function

$$t_a(s) \approx t_a(f_a^\star) \; \forall s \in \mathbb{R}_+$$

is equivalent to approximating the convex objective function by a linear function as

$$\hat{\mathbf{f}}_{\delta u} \in \arg\min_{\mathbf{f}\in\bar{\mathcal{F}}_{\delta u}} \sum_{a\in\mathcal{A}} \int_0^{f_a} t_a(f_a^\star)ds$$

$$\hat{\mathbf{f}}_{\delta u} \in \arg\min_{\mathbf{f}\in\bar{\mathcal{F}}_{\delta u}} \mathbf{f}^\top \mathbf{t}(\mathbf{f}^\star)$$

The idea for this relaxation stems from sensitivity analysis where we use a constant cost per unit flow. This "relaxation" is an LP and is easier to solve than a general convex program. We note that solving this LP is equivalent to finding all shortest paths between the origins and destinations of the saturated routes.

- The QP "relaxation" approximation: the following approximation of the objective function

$$t_a(s) \approx t_a(f_a^\star) + (s - f_a^\star)\cdot t_a'(f_a^\star)\ \forall s \in \mathbb{R}_+$$

is equivalent to approximating the convex objective function by a quadratic function.

$$\hat{\mathbf{f}}_{\delta u} = \arg\min_{\mathbf{f}\in\bar{\mathcal{F}}_{\delta u}} \mathbf{f}^\top c(\mathbf{f}^\star) + \tfrac{1}{2}\mathbf{f}^\top \nabla \mathbf{t}(\mathbf{f}^\star)\mathbf{f}$$

This "relaxation" is a QP and is also easier to solve than a general convex program.

The reason we do a sensitivity and feasibility analysis is that solving the [TAP-C] is expensive, so it may not be practical to perform these computations in real-time for small changes in link capacities. Since these optimization problems, involving computing the upper and lower bounds, are linear and/or quadratic programs, it is faster to solve these problems compared to solving a general convex optimization program. Therefore, it is easier to compute these bounds on the objective function so that we can guarantee a certain margin of change on the total travel time in the network.

## IV. Implementation and Results

In this section, we use a projected gradient algorithm with penalization to solve the [TAP-C]. We first solve the [TAP-C] on a small network (4 links). We present the upper bound and lower bounds evaluation on this network. Then, we solve the [TAP-C] on a larger network (75 links). We see that a decrease of the capacity constraint on one road can double the total vehicle hours of travel spent in the network.

### A. Algorithm

Convex solvers cannot directly solve the [TAP] because it would require enumerating all possible routes $\mathcal{R}$ in the network. Instead, traffic engineers currently use a modified version of the Frank-Wolfe algorithm to solve the [TAP] [12], [13], [5]. The Frank-Wolfe algorithm is a projected gradient descent method to find solutions to constrained convex optimization problems with polyhedral feasible sets. For [TAP], the Frank-Wolfe algorithm alternates between solving an LP sub-problem (finding an optimal all-or-nothing allocation) and a line search on the original problem. Solving the LP sub-problem consists of finding all shortest-paths for every origin-destination pair $k \in \mathcal{K}$. This is the bottleneck of the implementation, and there is much research on optimizing this part of the algorithm [5], [8]. We can modify this algorithm to incorporate capacity constraints by penalizing the objective function [3]. Research can be reproduced using the open-source code.[1]

Because the algorithm is a gradient descent, it does not find the optimal solution but an approximation of the solution. We stop the algorithm after 2,000 iterations of the gradient descent. In this particular set-up, the approximation of the [TAP-C] is 300 times faster than recomputing the new [TAP-C] (Table 1) for the Sioux Falls network (Section IV-C).

| | Mean (s) | Standard Deviation |
|---|---|---|
| TAP-C | 267.98 | ± 6.6 % |
| LP-Relaxation | 0.98 | ± 40 % |

Fig. 1. The LP "relaxation" is 300 times faster than recomputing [TAP-C].

### B. Results on a benchmark network

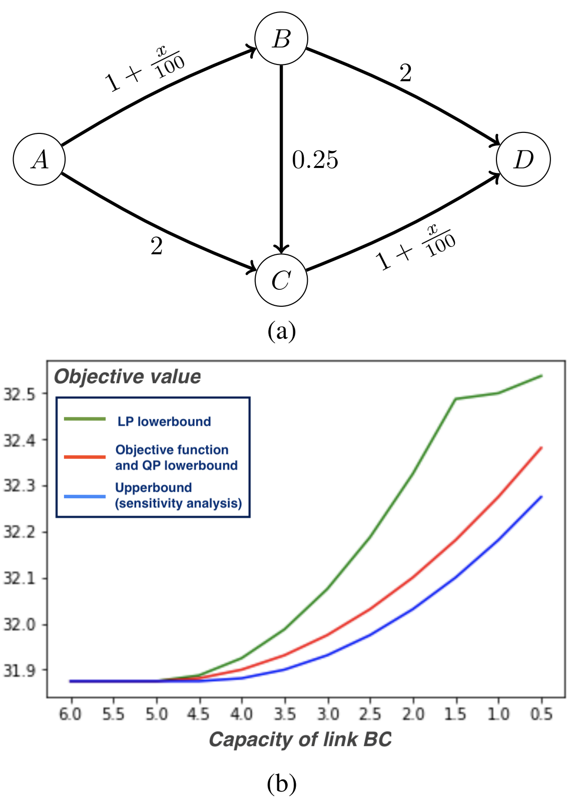


Fig. 2. Braess network considered – Cost on every link are given as a function of the link flow (for example: $t_{CD}(f_{CD}) = 1 + \frac{f_{CD}}{100}$) – and the value of the objective function of the TAP-C and the evaluations.

We first solve the [TAP-C] on a benchmark network (Figure 2) with 4 nodes ($A$, $B$, $C$ and $D$), 5 links and 3 paths. On this benchmark network we can enumerate all the possible paths and consequently use a convex optimization solver like CVX [14]. We fixed the demand to be 100

[1] https://github.com/TheoCabannes/Traffic_network_framework.

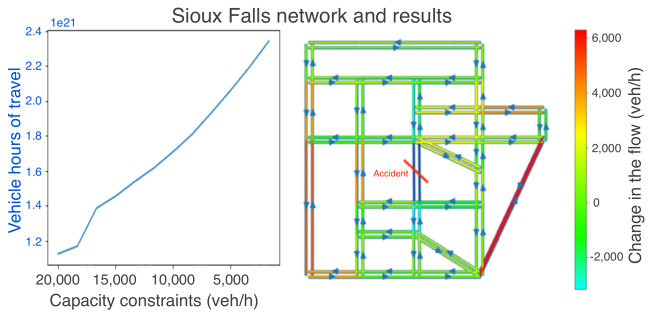


Fig. 3. Doubling of vehicle hours of travel as a function of the capacity reduction for the Sioux Falls network. The corresponding heat map shows flow changes of the traffic equilibrium solution to the [TAP-C] when the capacity of the two middle links decreases from $20,000$ veh/h to $0$ veh/h.

between $A$ to $D$, and we compare the solution when the capacity of link $BC$ decreases from 60 to 0. Figure 2 shows the exact solution and the approximation of the objective function. For this particular network, the QP "relaxation" gives the exact solution.

### C. Analysis on a larger network: the Sioux Falls network

We next perform simulations on a larger network: the Sioux Falls network (Figure 3). The network contains 24 nodes, 75 links and more than 2,500 paths without cycles. We obtained the network model, the traffic demand and the travel time of every link as a function of the link flow from an online repository.[2] The total demand is 360,600 trips/h. We reduce the capacity of the middle highway (for links show on Figure 3) from 20,000 veh/h to 0 veh/h. We do not consider capacity constraints for the other links.

We use the Frank-Wolfe algorithm with penalization [3], [13] to solve the [TAP-C] because we cannot enumerate all the paths in the network. By running 10 times the LP "relaxation" and the new [TAP-C], we observed that the LP "relaxation" is in average 300 times faster than recomputing the [TAP-C] when we observed a change in the capacity (Table 1). We plot the VHT as a function of the capacity reduction, observing that the total vehicle hours doubles when the capacity decreases to 2,500 veh/h (Figure 3). The difference of flow allocation for the 0 veh/h capacity with respect to the one for the 20,000 veh/h capacity is shown as a heatmap in Figure 3.

## V. Conclusion

The dual variables of the presented [TAP-C] formulation can be interpreted as link sensitivity with respect to a link traffic incident (i.e. a decrease of the link capacity). Therefore, the dual variables indicate where crisis management resources should be allocated and could improve dynamic decision making by traffic engineers. In section III, we develop an efficient method to evaluate the impact of a traffic incident on traffic flows using sensitivity analysis rather than recomputing the [TAP-C], which can be computationally expensive. We introduce an approximation of the [TAP-C] by first reducing the feasible flow allocation around the optimal solution and then performing an LP and a QP relaxation in this new set of feasible flow allocations. Results (section IV) show that approximating the [TAP-C] solution using the LP-relaxation reduces computational time by a factor of $\approx 300$ compared to recomputing the [TAP-C] on a network with 75 links.

Future research should aim to compute the [TAP-C] directly on the reduced set of feasible flow allocations. Reducing the set of feasible flow allocations using the convexity of the problem may significantly decrease the running time of computing the [TAP-C] with a projected gradient descent algorithm. Solving the [TAP-C] on this subset will give the optimal solution. One might also consider performing similar sensitivity analysis on different network parameters such as demand, different types of commodities (i.e. mode choice, etc.), and travel time to obtain a deeper insight into their impact on the traffic flow allocation.

ACKNOWLEDGEMENTS

We are grateful to Dr. Alexander Keimer and Pr. Laurent El-Ghaoui for their fruitful discussions.

## References

[1] *INRIX Global Traffic Scorecard*. INRIX Research, February 2018. [Online]. Available: http://inrix.com/scorecard/
[2] R. L. Carter and M. E. Crovella, "Measuring bottleneck link speed in packet-switched networks," *Performance evaluation*, vol. 27, pp. 297–318, 1996.
[3] T. Larsson and M. Patriksson, "An augmented lagrangean dual algorithm for link capacity side constrained traffic assignment problems," *Transportation Research Part B: Methodological*, vol. 29, no. 6, pp. 433–455, 1995.
[4] M. Patriksson, *The traffic assignment problem: models and methods*. Courier Dover Publications, 2015.
[5] H. Bast, D. Delling, A. Goldberg, M. Müller-Hanneman, T. Pajor, P. Sanders, D. Wagner, and R. F. Werneck, "Route planning in transportation networks," 2015, available via https://arxiv.org/pdf/1504.05140.pdf.
[6] J. G. Wardrop and J. I. Whitehead, "Correspondence. Some theoretical aspects of road traffic research," *ICE Proceedings: Engineering Divisions 1*, 1952.
[7] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation*. Yale University Press, 1956.
[8] B. V. Cherkassky, A. V. Goldberg, and T. Radzik, "Shortest paths algorithms: Theory and experimental evaluation," *Mathematical programming*, vol. 73, no. 2, pp. 129–174, 1996.
[9] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.
[10] T. Larsson, Z. Liu, and M. Patriksson, "A dual scheme for traffic assignment problems," *Optimization*, vol. 42, no. 4, pp. 323–358, 1997.
[11] M. A. Hall, "Properties of the equilibrium state in transportation networks," *Transportation Science*, vol. 12, no. 3, pp. 208–216, 1978. [Online]. Available: http://www.jstor.org/stable/25767914
[12] M. Frank and P. Wolfe, "An algorithm for quadratic programming," *Naval research logistics quarterly*, vol. 3, no. 1-2, pp. 95–110, 1956.
[13] M. Fukushima, "A modified frank-wolfe algorithm for solving the traffic assignment problem," *Transportation Research Part B: Methodological*, vol. 18, no. 2, pp. 169–177, 1984.
[14] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[2] https://github.com/bstabler/TransportationNetworks/blob/master/SiouxFalls/SiouxFalls_trips.tntp