A Unified Description of Iterative Algorithms for Traffic Equilibria *

Michael Patriksson †

LiTH-MAT-R-91-35
Revised April 1992

Abstract

The purpose of this paper is to provide a unified description of iterative algorithms for the solution of traffic equilibrium problems. We demonstrate that a large number of well known solution techniques can be described in a unified manner through the concept of partial linearization, and establish close relationships with other algorithmic classes for nonlinear programming and variational inequalities. In the case of nonseparable travel costs, the class of partial linearization algorithms are shown to yield new results in the theory of finite-dimensional variational inequalities. The possibility of applying truncated algorithms within the framework is also discussed.

Keywords: *Transportation,* *Networks,* *Mathematical Programming*

1 Introduction

Traffic planning and transportation analysis problems have motivated a large amount of academic research into mathematical modelling and the design of algorithms during the last 25 years. The analysis of mathematical models of traffic planning problems aid planners in predicting what effects changes in the traffic network topology will have on the network's performance, with respect to congestion, queueing, safety and so on, without having to perform these changes in the real traffic network.

In this paper we consider the static model of traffic assignment with and without link interactions, with emphasis on the problem resulting from the principle of user equilibrium, formulated by Wardrop [97]. Given link performance functions, modelled as strictly increasing functions of the flow on the links, the equilibrium conditions can be shown to be the first order optimality conditions of a convex mathematical program; hence, the traffic equilibrium problem may be analyzed by using standard techniques in nonlinear optimization. The special problem structure

*The author wishes to thank Drs. O. Drissi-Kaïtouni, T. Larsson, A. Migdalas and J.H. Wu for discussions leading to improvements of the paper. The research was partially supported by the Swedish Transport Research Board (TFB) Grant Dnr. 16/88-62.

†Department of Mathematics, Linköping Institute of Technology, S-581 83 Linköping, Sweden

and the size of real-world problems, however, make the problem a challenge for the development of specialized methods. A large amount of research has been made in the development of efficient algorithms for these problems and also to validate the use of the models in practice. When there are link interactions present in the model, a reformulation as a convex program is not possible to make in general. Through a reformulation of the equilibrium conditions as a finite-dimensional variational inequality, it is still possible to solve these more general problems efficiently. Recently, variational inequalities have shown to be equivalent to (generally) nonconvex, and also possibly nondifferentiable, optimization problems. Based on these optimization formulations new algorithms have been developed, and applied to traffic equilibrium.

Many of the feasible direction methods and successive approximation algorithms presented for the traffic assignment problem may be given a unified description. In this paper we introduce a class of algorithms that is based on iterative travel cost approximations, and show that a number of well known iterative algorithms for traffic assignment belong to this class. Examples of such methods are the Frank-Wolfe algorithm, Jacobi and Gauss-Seidel type methods and Newton methods. We will also establish a close relationship between this algorithmic class and the classes of algorithms defined by Dafermos [24], Migdalas [69] and Tseng [93].

The remainder of the paper is organized as follows. In Section 2 we introduce the traffic equilibrium problem, and discuss some of its important properties. In Section 3 we introduce the concept of partial linearization and provide some interpretations. We discuss the global convergence of the general algorithm, and validate truncated subproblem solutions within the framework. Well known instances of the algorithm, as well as possible extensions, are discussed in Section 4. Parallel (partially) asynchronous implementations of the partial linearization algorithm are also validated. In Section 5 we present the variational inequality formulation of the general traffic equilibrium problem. Optimization reformulations are also presented, and we establish a new equivalence result for variational inequality algorithms. We provide in Section 6 some well known instances of algorithms for traffic equilibria, and discuss possible extensions by the introduction of line searches with respect to merit functions. Finally, in Section 7, we draw some conclusions and present research opportunities.

2 The traffic equilibrium model

Consider a transportation network $\mathcal{G}=(\mathcal{N},\mathcal{A})$, where each directed arc $a \in \mathcal{A}$ is associated with a positive travel time, $t_a(\mathbf{f})$. This travel time, or transportation cost, measures the disutility of using the arc as a function of the network flow $\mathbf{f}$. The functions $t_a(\mathbf{f})$ are usually referred to as *arc performance functions*, and are monotone as a result of congestion. For certain pairs of origins and destinations, $(p,q) \in \mathcal{C}$, where $\mathcal{C} \subset \mathcal{N} \times \mathcal{N}$, there is a given positive flow demand d_{pq}. Each O-D pair (p,q) is associated with a specific commodity. We denote the commodity flow directed from node p to node q through arc a by f_{apq}, giving rise to the total arc flow equation $f_a = \sum_{(p,q)\in\mathcal{C}} f_{apq}$. To measure the commodity flow into and out of a specific node $i \in \mathcal{N}$, we define $\mathcal{W}_i$ and $\mathcal{V}_i$ to be the sets of arcs initiated at node i and terminating at node i, respectively. The problem of determining a network flow fulfilling the travel demands and a prescribed performance criterion is referred to as the *traffic assignment problem.*

In this paper, we consider static traffic assignment problems, modelling peak-hour urban traffic. Two main principles of optimality are usually considered. These are attributed to Wardrop [97]. The first optimality principle is based on the intuitive behaviour of traffic, i.e., each user of the traffic network seeks to minimize his/her own travel time, leading to a steady-state where all used routes between a given pair of origin and destination have equal travel times; it is therefore known as the principle of *user equilibrium*. With the assumptions that the function $\mathbf{t}(\mathbf{f})$ is integrable and that the Jacobian of $\mathbf{t}$ is positive semidefinite for all feasible flows, Wardrop's first conditions of optimality can be reformulated as a convex mathematical program (e.g. [22]). First to formulate this program are Beckmann *et al.* [7], in the case of separable cost functions (i.e., $t_a(\mathbf{f}) = t_a(f_a), \forall a \in \mathcal{A}$). The assumption that the travel cost function is integrable is in some applications too restrictive. A nonintegrable function $\mathbf{t}(\mathbf{f})$ corresponds to an asymmetric Jacobian of travel costs. The problem is therefore known as the *asymmetric traffic assignment problem*, and may be formulated for instance, as a variational inequality [90, 23], a nonlinear complementarity problem [2], a fixed point problem (e.g. [71]) or as a, generally nonconvex, mathematical program by the use of so called *gap functions* (e.g. [44]). If separable costs are considered, the corresponding mathematical program will have an objective with additive terms $g_a(f_a) = \int_0^{f_a} t_a(s)ds$.

The second optimality principle is known as the *system optimum* principle, and corresponds to the situation in which the whole transportation system's disutility is minimized. The flows corresponding to the system optimum must be imposed upon the users, thus giving a problem of *prescription*, as opposed to the problem of *description* in the case of user equilibrium. Under the assumption that each function $t_a(f_a)$ is positive, monotonically increasing and convex, this principle can be shown to be equivalent to a convex mathematical program, and the objective will have additive terms $g_a(f_a) = t_a(f_a)f_a$.

The fixed demand Traffic Assignment Problem may, in both cases of optimality principles, be stated as (e.g. [74])

$$
[\mathbf{TAP}] \qquad \begin{array}{lrcll}
\min & T(\mathbf{f}) = \displaystyle\sum_{a\in\mathcal{A}} g_a(f_a) & & & \\
\text{s.t.} & \mathbf{A}\mathbf{f}_{pq} & = & \mathbf{d}_{pq} & \forall (p,q) \in \mathcal{C} \\
 & \mathbf{f}_{pq} & \geq & \mathbf{0} & \forall (p,q) \in \mathcal{C} \\
 & \displaystyle\sum_{(p,q)\in\mathcal{C}} f_{apq} & = & f_a & \forall a \in \mathcal{A},
\end{array}
$$

where $\mathbf{A}$ is an $|\mathcal{N}| \times |\mathcal{A}|$ arc-node incidence matrix, $\mathbf{f}_{pq}$ is the vector of commodity flows f_{apq} and where $\mathbf{d}_{pq}$ denotes the demand vector for commodity (p, q). This formulation is referred to as the *arc-node* formulation.

It is also possible to formulate [**TAP**] in terms of route flow variables. Define the set of simple routes from p to q by $\mathcal{R}_{pq}$, and the flow on route $r \in \mathcal{R}_{pq}$ by h_{pqr}. If we define an arc-route incidence matrix $\boldsymbol{\Delta} = (\delta_{pqra})$ for $\mathcal{G}$, i.e., $\delta_{pqra} = 1$ if $r \in \mathcal{R}_{pq}$ contains arc a, and 0 otherwise, then arc flows are calculated as $f_a = \sum_{(p,q)\in\mathcal{C}} \sum_{r\in\mathcal{R}_{pq}} \delta_{pqra} h_{pqr}$, and the traffic assignment problem becomes (e.g. [39, 27])

$$
[\mathbf{TAP}] \qquad \begin{array}{lrclll}
\min & T(\mathbf{f}) & = & \displaystyle\sum_{a\in\mathcal{A}} g_a(f_a) & & \\
\text{s.t.} & \displaystyle\sum_{r\in\mathcal{R}_{pq}} h_{pqr} & = & d_{pq} & & \forall (p,q) \in \mathcal{C} \\
 & h_{pqr} & \geq & 0 & \forall r \in \mathcal{R}_{pq} & \forall (p,q) \in \mathcal{C} \\
 & \displaystyle\sum_{(p,q)\in\mathcal{C}} \sum_{r\in\mathcal{R}_{pq}} \delta_{pqra} h_{pqr} & = & f_a & \forall a \in \mathcal{A}. &
\end{array}
$$

We will refer to this formulation as the *arc-route* formulation. In the sequel we will concentrate on the case of user equilibrium.

The reader may note that an inherent property of [**TAP**] is that its objective function is nonseparable with respect to commodity flows, while the feasible set is a Cartesian product set. Furthermore, one can show that, for strictly increasing arc performance functions, the objective is strictly convex with respect to total arc flows f_a, but only convex with respect to commodity flows f_{apq} (e.g. [89]). Thus, the program [**TAP**] is convex.

The development of formalized methods for the solution of the traffic assignment problem arose in the middle of the 1960's. The most influential and well known algorithm is an adaptation of the method of Frank and Wolfe [37], first suggested for use in this field by Bruynooghe *et al.* [16] and implemented for a small city by LeBlanc *et al.* [54]. Based on a feasible flow $\mathbf{f}^{(l)}$, a first order Taylor approximation of the objective yields a linear subproblem, which in the case of user equilibrium is equivalent to $|\mathcal{C}|$ independent shortest route problems in the transportation network, based on the arc costs $\mathbf{t}\left(\mathbf{f}^{(l)}\right)$. A line search is then made in the direction $\mathbf{y}^{(l)} - \mathbf{f}^{(l)}$, where $\mathbf{y}^{(l)}$ is a shortest route pattern.

Another important class of algorithms is the class of *cyclic decomposition methods*, in which the problem is solved for subsets of the variables sequentially. Cyclic decomposition schemes extend block Gauss-Seidel methods known from the solution of systems of linear and nonlinear equations to constrained optimization [75]. In the traffic assignment context, the cyclic decomposition is made over the commodities (e.g. [73, 82]). The advantage of these algorithms, as compared to the Frank-Wolfe approach, is that the nonlinearity of the objective is preserved, thus better retaining the properties of the original problem. A disadvantage, however, is that the algorithm is inherently sequential, as opposed to the Frank-Wolfe algorithm.

The idea behind the original formulation of a *partial linearization method*[1] [49] was to combine the advantages of these two approaches, by making subproblems nonlinear and separable with respect to commodities. (This is accomplished by identifying the nonseparable part of the objective, and linearizing it.) It is a generalization of the Frank-Wolfe algorithm that avoids extreme point solutions to the subproblems by retaining them nonlinear. In this way the subproblems will better resemble the original problem, and therefore hopefully provide better search directions. By making the subproblems separable with respect to the commodities, the algorithm will avoid the sequential character of cyclic decomposition schemes, and hence, the algorithm may utilize parallel computing environments.

In the following, we will develop a class of algorithms, whose common feature is that only a part of the objective is linearized in each iteration, as opposed to the Frank-Wolfe approach. The algorithm extends the one given in [49]; by choosing the part to linearize in different ways, a number of well known methods for nonlinear programming and traffic assignment may be recognized.

[1] Originally, the term partial linearization was used by Evans [30] for an iterative algorithm for the combined distribution and assignment problem. This algorithm is, in fact, an instance of the partial linearization algorithm.

3 The partial linearization algorithm

3.1 Introduction

The basis for the discussion in this section is an article by the author [81], where the convergence of partial linearization methods is established for, not necessarily differentiable, convex programming and pseudo-convex differentiable programming. The algorithm is stated for a general problem over a compact and convex feasible set, where the objective is pseudo-convex and continuously differentiable.

[**P**] $$\min_{\mathbf{x} \in \mathbf{X}} T(\mathbf{x})$$

The general algorithm works as follows.

In a feasible point, $\mathbf{x}^{(l)}$, the objective function is expressed as

$$T(\mathbf{x}) = \Phi\left(\mathbf{x}, \mathbf{x}^{(l)}\right) + \left(T(\mathbf{x}) - \Phi\left(\mathbf{x}, \mathbf{x}^{(l)}\right)\right), \tag{3.1}$$

where $\Phi(\mathbf{x}, \mathbf{y})$ is assumed convex and continuously differentiable with respect to $\mathbf{x}$ and continuous with respect to $\mathbf{y}$. The second term in (3.1) can be seen as expressing the error when replacing the original objective, $T(\mathbf{x})$, by the function $\Phi\left(\mathbf{x}, \mathbf{x}^{(l)}\right)$. The idea then is to take this error into account by replacing it with a first order Taylor expansion around $\mathbf{x}^{(l)}$.

The convex subproblem solved in iteration l then is

[**PL-SUB**$^{(l)}$] $$\min_{\mathbf{x} \in \mathbf{X}} T^{(l)}\left(\mathbf{x}, \mathbf{x}^{(l)}\right), \quad \text{where}$$

$$T^{(l)}\left(\mathbf{x}, \mathbf{x}^{(l)}\right) = \Phi\left(\mathbf{x}, \mathbf{x}^{(l)}\right) + T\left(\mathbf{x}^{(l)}\right) - \Phi\left(\mathbf{x}^{(l)}, \mathbf{x}^{(l)}\right) + \left[\nabla T\left(\mathbf{x}^{(l)}\right) - \nabla_{\mathbf{x}} \Phi\left(\mathbf{x}^{(l)}, \mathbf{x}^{(l)}\right)\right]^{\mathrm{T}} \left(\mathbf{x} - \mathbf{x}^{(l)}\right),$$

and where $\nabla_{\mathbf{x}} \Phi(\mathbf{y}, \mathbf{y}) = \nabla_{\mathbf{x}} \Phi(\mathbf{x}, \mathbf{y})|_{\mathbf{x}=\mathbf{y}}$. We immediately see that the algorithm extends the Frank-Wolfe algorithm, which corresponds to letting $\Phi \equiv 0$.

The basic descent property is given by the following theorem.

Theorem 3.1 [81, Theorem 2.1] *Assume that* $\mathbf{x}^{(l)} \in \mathbf{X}$, *and that* $\overline{\mathbf{x}}^{(l)}$ *is any optimal solution to* [**PL-SUB**$^{(l)}$]. *If* $\mathbf{x}^{(l)}$ *solves* [**PL-SUB**$^{(l)}$], *then* $\mathbf{x}^{(l)}$ *solves* [**P**]. *Otherwise, the direction* $\mathbf{d}^{(l)} = \overline{\mathbf{x}}^{(l)} - \mathbf{x}^{(l)}$ *is a feasible descent direction with respect to* T.

If convergence is not detected after having solved [**PL-SUB**$^{(l)}$], the algorithm proceeds by performing a line search in the direction $\mathbf{d}^{(l)}$ with respect to T, defining the new point $\mathbf{x}^{(l+1)}$. The basic convergence property of this algorithm is given below.

Theorem 3.2 [81, Theorem 2.2] *The partial linearization algorithm either terminates in a finite number of iterations or it generates an infinite sequence* $\left\{\mathbf{x}^{(l)}\right\}$ *such that any accumulation point solves* [**P**].

Under the additional assumption that $T(\mathbf{x})$ has a *Lipschitz continuous* gradient, i.e., there exists a constant $K \geq 0$ such that

$$\left\| \nabla T\left(\mathbf{x}^1\right) - \nabla T\left(\mathbf{x}^2\right) \right\| \leq K \left\| \mathbf{x}^1 - \mathbf{x}^2 \right\|, \forall \mathbf{x}^1, \mathbf{x}^2 \in \mathbf{X},$$

then it is possible to show that the partial linearization algorithm is globally convergent even when line searches are made only approximately. See [81] for further details.

Later, we will develop a variant of the partial linearization algorithm where the subproblems are solved inaccurately, i.e., where [$\mathbf{PL\text{-}SUB}^{(l)}$] is solved with a *truncated* algorithm. Global convergence holds for any descent algorithm applied to [$\mathbf{PL\text{-}SUB}^{(l)}$], and for almost arbitrary inaccuracy, provided that the algorithmic map of the descent algorithm is closed. (See [99] for a general discussion on closedness of algorithmic maps.) The Frank-Wolfe algorithm is an interesting special case, since such a *truncated partial linearization algorithm* may be easily implemented by slightly modifying an existing Frank-Wolfe code.

Some differences between the general algorithm and the Frank-Wolfe method is noted below.

A nice property of the Frank-Wolfe algorithm is that for convex problems a termination criterion is available through the lower bound on the optimal objective value, obtained from the linear subproblem. In the partial linearization algorithm this termination criterion is not valid since the error function $T(\mathbf{x}) - \Phi(\mathbf{x}, \mathbf{y})$ is not convex in general. It is, however, easy to show that $T\left(\mathbf{x}^{(l)}\right) - T^{(l)}\left(\overline{\mathbf{x}}^{(l)}, \mathbf{x}^{(l)}\right) > 0$ whenever $\mathbf{x}^{(l)}$ is not a solution to [$\mathbf{P}$], and that this difference tends towards zero. The reader may note that an optimal solution is obtained also from the sequence of solutions to the subproblems, if $\Phi(\mathbf{x}, \mathbf{y})$ is strictly convex with respect to $\mathbf{x}$, i.e., $\lim_{l\to\infty}\left(\mathbf{x}^{(l)} - \overline{\mathbf{x}}^{(l)}\right) = \mathbf{0}$. This is not the case in the Frank-Wolfe algorithm.

Interpretations of the subproblem in the partial linearization algorithm are made below.

If the function $\Phi(\mathbf{x}, \mathbf{y})$ is chosen so that $\nabla_{\mathbf{x}}\Phi(\mathbf{x}, \mathbf{x}) = \mathbf{0}$, then [$\mathbf{PL\text{-}SUB}^{(l)}$] reduces to

$$\min_{\mathbf{x}\in\mathbf{X}} \ \nabla T\left(\mathbf{x}^{(l)}\right)^{\mathrm{T}}\left(\mathbf{x} - \mathbf{x}^{(l)}\right) + \Phi\left(\mathbf{x}, \mathbf{x}^{(l)}\right). \tag{3.2}$$

Algorithms of this type is studied by Migdalas [69], who refers to them as *regularized Frank-Wolfe algorithms*, and Tseng [93], by the name *nonlinear proximal descent* (NPD) methods. (As will be shown later, these algorithms are, in fact, equivalent to the partial linearization algorithm.) The expression (3.2) provides a nice characterization of the method class as *partial* linearization methods as opposed to *complete* linearization methods, such as the Frank-Wolfe algorithm. In the Frank-Wolfe algorithm, the first order Taylor expansion, which is valid only locally around $\mathbf{x}^{(l)}$, is used globally in the subproblem phase. The subproblem solved in a partial linearization method introduces a regularization term in the objective function of the Frank-Wolfe subproblem, restricting the distance between the current point $\mathbf{x}^{(l)}$ and the subproblem solution $\overline{\mathbf{x}}^{(l)}$. By retaining the nonlinearity of the original objective function, partial linearization methods will therefore avoid the *tailing-off* phenomena inherent in the Frank-Wolfe method, caused by the generation of extreme point subproblem solutions (e.g. [59]).

Apart from avoiding tailing-off phenomena the partial linearization method may be used to regularize not strictly convex problems by adding a strictly convex term to the objective. Assume that the objective function $T(\mathbf{x})$ is convex but *not* strictly convex everywhere. Assume also that the function $\Phi(\mathbf{x}, \mathbf{y})$ is strictly convex with respect to $\mathbf{x}$. Let $\overline{\Phi}(\mathbf{x}, \mathbf{y}) = T(\mathbf{x}) + \Phi(\mathbf{x}, \mathbf{y})$, which then is strictly convex with respect to $\mathbf{x}$. Rewrite the objective as in (3.1), i.e., let

$$\begin{aligned} T(\mathbf{x}) &= \overline{\Phi}(\mathbf{x},\mathbf{y}) + \left(T(\mathbf{x}) - \overline{\Phi}(\mathbf{x},\mathbf{y})\right) \\ &= T(\mathbf{x}) + \Phi(\mathbf{x},\mathbf{y}) + (-\Phi(\mathbf{x},\mathbf{y})). \end{aligned}$$

The subproblem solved in the partial linearization method then is to minimize

$$T^{(l)}(\mathbf{x},\mathbf{x}^{(l)}) = T(\mathbf{x}) + \Phi\left(\mathbf{x},\mathbf{x}^{(l)}\right) - \nabla_{\mathbf{x}}\Phi\left(\mathbf{x}^{(l)},\mathbf{x}^{(l)}\right)^{\mathrm{T}}\left(\mathbf{x}-\mathbf{x}^{(l)}\right),$$

i.e., the subproblem in the partial linearization algorithm may also be seen as a *strengthening* of a not strictly convex problem by adding a strictly convex term, Φ, to the objective. As an example, consider the function $\Phi(\mathbf{x},\mathbf{y}) = \frac{c}{2}\|\mathbf{x}-\mathbf{y}\|_2^2$, where $c > 0$. The subproblem then is to minimize $T(\mathbf{x}) + \dfrac{c}{2}\left\|\mathbf{x}-\mathbf{x}^{(l)}\right\|_2^2$, which is equivalent to the subproblem of the proximal point algorithm [86].

The possibility of obtaining strictly convex subproblems by choosing Φ strictly convex also facilitates the use of duality based methods for [**PL-SUB**$^{(l)}$]. Efficient dual ascent methods have been developed for strictly convex problems in single-commodity networks, matrix balancing, quadratic programming and entropy maximization, see e.g. [11, 14, 15, 17, 18, 57, 87, 92, 95, 96, 100]. Through the use of partial linearization such methods are applicable to [**TAP**] since the choice of Φ as a separable strictly convex function with respect to the commodities induces both strict convexity and separability of [**PL-SUB**$^{(l)}$] (see Section 4.1). Since these methods are amenable to parallel computations efficient parallel partial linearization methods are defined (see Section 4.2).

The subproblem [**PL-SUB**$^{(l)}$] may also be given a more general formulation. Assume that [**P**] is given by its variational inequality characterization (see Section 5.1)

[**VIP**] find $\mathbf{x}^* \in \mathbf{X}$ such that $\nabla T(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{x}-\mathbf{x}^*) \geq 0, \forall \mathbf{x} \in \mathbf{X}$.

Assume also that $\varphi : \mathbf{X}\times\mathbf{X} \mapsto \Re^n$ is monotone with respect to its first argument and continuous with respect to its second, and that $\varphi(\cdot,\mathbf{y})$ is *not* necessarily the gradient of a convex function $\Phi(\cdot,\mathbf{y})$. The subproblem of the partial linearization method may then be described as follows. In iteration l, the cost vector ∇T is approximated by $\varphi\left(\cdot,\mathbf{x}^{(l)}\right)$, and the error made when replacing ∇T with φ is taken into account by adding the term $\nabla T\left(\mathbf{x}^{(l)}\right) - \varphi\left(\mathbf{x}^{(l)},\mathbf{x}^{(l)}\right)$ to the cost. The subproblem is then the following variational inequality.

[**PL-SUB-VIP**$^{(l)}$] Find $\overline{\mathbf{x}}^{(l)} \in \mathbf{X}$ such that

$$\left[\varphi\left(\overline{\mathbf{x}}^{(l)},\mathbf{x}^{(l)}\right) + \nabla T\left(\mathbf{x}^{(l)}\right) - \varphi\left(\mathbf{x}^{(l)},\mathbf{x}^{(l)}\right)\right]^{\mathrm{T}}\left(\mathbf{x}-\overline{\mathbf{x}}^{(l)}\right) \geq 0, \forall \mathbf{x} \in \mathbf{X}$$

(By associating the map $\varphi(\cdot,\mathbf{y})$ with the gradient of $\Phi(\cdot,\mathbf{y})$, using its variational inequality characterization the subproblem [**PL-SUB**$^{(l)}$] can be shown to be equivalent to [**PL-SUB-VIP**$^{(l)}$].)

This more general description of the subproblem of the partial linearization algorithm will be used later to establish that cyclic decomposition methods are instances of the general algorithm.

3.2 Truncated partial linearization algorithms

From a practical point of view subproblems can not be solved exactly, and there is an obvious trade-off between the amount of work spent on solving [**PL-SUB**$^{(l)}$] and obtaining sufficiently steep descent directions. We conclude this section by establishing the global convergence of a version of the partial linearization algorithm, where each subproblem is solved with a truncated descent algorithm with closed algorithmic map.

The idea behind the truncated partial linearization algorithm is to bound the work performed on [**PL-SUB**$^{(l)}$], by bounding from above the number of iterations performed with a finite integer, $k^{(l)}$. These numbers can either be determined *a priori*, or be viewed as being the consequence of the algorithm and stopping criteria chosen for [**PL-SUB**$^{(l)}$]. It will be shown that the sequence $\left\{k^{(l)}\right\}$ may be chosen arbitrary, with $k^{(l)} \geq 1, \forall l$, and convergence will still be ensured under the condition that the method used for solving [**PL-SUB**$^{(l)}$] has a closed algorithmic map.

The following lemma and theorem establish the global convergence of the *truncated partial linearization algorithm.*

Lemma 3.1 *If* $T^{(l)}\left(\mathbf{x}, \mathbf{x}^{(l)}\right) < T^{(l)}\left(\mathbf{x}^{(l)}, \mathbf{x}^{(l)}\right)$ *holds for some* $\mathbf{x} \in \mathbf{X}$*, then* $\mathbf{d}^{(l)} = \mathbf{x} - \mathbf{x}^{(l)}$ *is a feasible descent direction of descent with respect to* T*.*

Proof The fact that $\nabla_{\mathbf{x}} T^{(l)}(\mathbf{x}, \mathbf{x}) = \nabla T(\mathbf{x}), \forall \mathbf{x}$, and the convexity of $T^{(l)}$ yields

$$\begin{aligned} \nabla T\left(\mathbf{x}^{(l)}\right)^{\mathrm{T}}\left(\mathbf{x} - \mathbf{x}^{(l)}\right) &= \nabla T^{(l)}\left(\mathbf{x}^{(l)}, \mathbf{x}^{(l)}\right)^{\mathrm{T}}\left(\mathbf{x} - \mathbf{x}^{(l)}\right) \\ &\leq T^{(l)}\left(\mathbf{x}, \mathbf{x}^{(l)}\right) - T^{(l)}\left(\mathbf{x}^{(l)}, \mathbf{x}^{(l)}\right) \\ &< 0. \end{aligned}$$

□

The implication of this lemma is that is suffices with a single iteration of a descent algorithm on [**PL-SUB**$^{(l)}$] to obtain a feasible descent direction with respect to T.

Theorem 3.3 *Assume that* $T(\mathbf{x})$ *is pseudo-convex and continuously differentiable, and that* $\Phi(\mathbf{x}, \mathbf{y})$ *is convex and continuously differentiable with respect to* $\mathbf{x}$*, and continuous with respect to* $\mathbf{y}$*. Furthermore, assume that the algorithm used for solving* [**PL-SUB**$^{(l)}$] *is a descent algorithm with closed algorithmic map, and that the termination criteria chosen for* [**PL-SUB**$^{(l)}$] *are such that* $1 \leq k^{(l)} < \infty, \forall l$*. Then any accumulation point of the sequence* $\left\{\mathbf{x}^{(l)}\right\}$ *solves* [**P**].

Proof By Lemma 3.1, the sequence $\left\{\mathbf{d}^{(l)}\right\}$ is a sequence of descent directions. Following the proof of [81, Theorem 2.2] we conclude that there exists a subsequence $\hat{L}$ so that

$$\lim_{l \in \hat{L}} \nabla T\left(\mathbf{x}^{(l)}\right)^{\mathrm{T}}\left(\overline{\mathbf{x}}^{(l)} - \mathbf{x}^{(l)}\right) = \nabla T\left(\mathbf{x}^*\right)^{\mathrm{T}}\left(\overline{\mathbf{x}} - \mathbf{x}^*\right) = 0, \tag{3.3}$$

where $\mathbf{x}^* = \lim_{l \in \hat{L}} \mathbf{x}^{(l)}$ and $\overline{\mathbf{x}} = \lim_{l \in \hat{L}} \overline{\mathbf{x}}^{(l)}$, and where $\overline{\mathbf{x}}^{(l)}$ is the result of the truncated solution of [**PL-SUB**$^{(l)}$].

Taking the limit of $\hat{L}$, we have by Lemma 3.1 that

$$\lim_{l \in \hat{L}} T^{(l)}\left(\overline{\mathbf{x}}^{(l)}, \mathbf{x}^{(l)}\right) \stackrel{\text{def}}{=} T^*\left(\overline{\mathbf{x}}, \mathbf{x}^*\right) \leq T^*\left(\mathbf{x}^*, \mathbf{x}^*\right).$$

The convexity of $T^{(l)}, \forall l$ implies that

$$T^*\left(\overline{\mathbf{x}}, \mathbf{x}^*\right) \geq T^*\left(\mathbf{x}^*, \mathbf{x}^*\right) + \nabla T^*\left(\mathbf{x}^*, \mathbf{x}^*\right)^{\mathrm{T}}\left(\overline{\mathbf{x}} - \mathbf{x}^*\right).$$

But the fact that $\nabla T^{(l)}(\mathbf{x}, \mathbf{x}) = \nabla T(\mathbf{x}), \forall \mathbf{x}$ and using (3.3) we obtain $T^*\left(\overline{\mathbf{x}}, \mathbf{x}^*\right) \geq T^*\left(\mathbf{x}^*, \mathbf{x}^*\right)$, and hence we have that

$$T^*\left(\overline{\mathbf{x}}, \mathbf{x}^*\right) = T^*\left(\mathbf{x}^*, \mathbf{x}^*\right). \tag{3.4}$$

Assume that, for any positive constant M, there exists a finite integer i such that $k^{(l)} \geq M, \forall l \geq i$. (We say that $\lim_{l \to \infty} k^{(l)} = \infty$.) Then the subproblem is solved accurately in the limit, and convergence is ensured by Theorem 3.2.

Let us now show that the subproblem is solved accurately in the limit also when $k^{(l)} \not\to \infty$.

Since $k^{(l)} \not\to \infty$, there must be an integer k^* that occurs in the sequence $\left\{k^{(l)}\right\}$, $l \in \hat{L}$, an infinite number of times. Choose the subsequence of $\hat{L}$ corresponding to these indices. The subproblem [**PL-SUB**$^{(l)}$] is in the limit of this sequence solved with $k^* \geq 1$ iterations of a descent algorithm with closed algorithmic map. We next show that $\overline{\mathbf{x}}$ is a solution to this problem.

Define $\overline{\mathbf{x}}^{(l)} \in \mathbf{M}^{(l)}\left(\mathbf{x}^{(l)}\right)$ to be the algorithmic map defining the solution of [**PL-SUB**$^{(l)}$] obtained after $k^{(l)}$ iterations. It suffices to show that $\overline{\mathbf{x}} \in \overline{\mathbf{M}}(\mathbf{x}^*)$, where $\overline{\mathbf{M}}$ is the composite map of k^* consecutive direction finding problems (the map $\mathbf{C}$) and line searches (the map $\mathbf{B}$), i.e., $\overline{\mathbf{M}} = \underbrace{\mathbf{BCBC} \times \ldots \times \mathbf{BC}}_{k^* \text{ times } \mathbf{BC}}$.

But since the mapping $\mathbf{C}$ is closed by assumption, and also the mapping $\mathbf{B}$ [6, Theorem 8.3.1], the composite mapping $\overline{\mathbf{M}}$ is closed [6, Theorem 7.3.2]. Hence, $\overline{\mathbf{x}} \in \overline{\mathbf{M}}(\mathbf{x}^*)$ so that $\overline{\mathbf{x}}$ is the solution to the truncated subproblem in the limit.

By using (3.4), Lemma 3.1 and the fact that $k^* \geq 1$, we conclude that $\mathbf{x}^*$ must solve the subproblem in the limit, i.e., that $\mathbf{x}^* \in \arg\min_{\mathbf{x} \in \mathbf{X}} T^*\left(\mathbf{x}, \mathbf{x}^*\right)$. Theorem 3.1 then implies that $\mathbf{x}^*$ solves [**P**]. □

Related convergence results have been obtained earlier for certain conceptual algorithms, see e.g. [83, Section A.2].

Note that, for the interesting special case of using Frank-Wolfe for solving [**PL-SUB**$^{(l)}$], whenever $k^{(l)} = 1$, iteration l in the truncated partial linearization algorithm reduces to a Frank-Wolfe iteration. Note also that by the proof of Theorem 3.3, the limit point of the truncated subproblem solutions, $\overline{\mathbf{x}}$, in fact solves the subproblem in the limit *exactly*, despite the fact that each subproblem is solved inaccurately.

Migdalas [69] presents a truncated Frank-Wolfe algorithm for the subproblem of the regularization algorithm. There, the subproblems are truncated when the relative error is smaller than an *a priori* determined value. To ensure convergence, the sequence of values are assumed to have limit zero, so that, in essence, the subproblems are solved with a higher and higher accuracy.

It is interesting to note that, by the discussion made above, this is accomplished automatically regardless of the number of iterations performed on each [$\mathbf{PL\text{-}SUB}^{(l)}$].

The feasible set in the arc-node formulation of the traffic equilibrium problem is, in general, not bounded due to the presence of cycles in the network. When applying the Frank-Wolfe algorithm to [$\mathbf{PL\text{-}SUB}^{(l)}$], it is possible that negative arc costs can occur, and that, as a consequence, the linear subproblem may have an infinite solution. (This can, however, not happen during the first iteration of the subproblem solution process, since the arc costs then are defined by $\mathbf{t}\left(\mathbf{f}^{(l)}\right)$, positive by assumption.) In a practical implementation the subproblem may be truncated prematurely if a negative cycle is detected. The global convergence of such an algorithm is still ensured by Theorem 3.3.

In the next section we will outline the use of partial linearization methods for separable traffic equilibrium problems. The presentation is divided into two parts; we first discuss Jacobi/Gauss-Seidel type algorithms and then Newton type approaches. Algorithms incorporating truncated algorithms and simplicial decomposition/column generation are also discussed.

4 Partial linearization algorithms for separable traffic equilibrium problems

4.1 Jacobi/Gauss-Seidel type algorithms

The Jacobi and Gauss-Seidel methods are well known for the solution of systems of linear and nonlinear equations [75], and optimization problems over Cartesian product sets [17]. Assume that the feasible set $\mathbf{X} \subseteq \Re^n$ in [$\mathbf{P}$] is a Cartesian product, i.e., that $\mathbf{X} = \prod_{i \in \mathcal{C}} \mathbf{X}_i$, where $\mathbf{X}_i \subset \Re^{n_i}$ and $\sum_{i \in \mathcal{C}} n_i = n$. Given a point $\mathbf{x}^{(l)} \in \mathbf{X}$, the Jacobi subproblem then consists of the following $|\mathcal{C}|$ independent problems

[$\mathbf{J}_i^{(l)}$] $$\mathbf{x}_i^{(l+1)} \in \arg \min_{\mathbf{x}_i \in \mathbf{X}_i} \; T\left(\mathbf{x}_1^{(l)}, \mathbf{x}_2^{(l)}, \ldots, \mathbf{x}_{i-1}^{(l)}, \mathbf{x}_i, \mathbf{x}_{i+1}^{(l)}, \ldots, \mathbf{x}_{|\mathcal{C}|}^{(l)}\right).$$

It is easily verified that Jacobi is a partial linearization algorithm, through the subproblem defining function $\Phi(\mathbf{x}, \mathbf{y}) = \sum_{i \in \mathcal{C}} T\left(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{i-1}, \mathbf{x}_i, \mathbf{y}_{i+1}, \ldots, \mathbf{y}_{|\mathcal{C}|}\right)$. As applied to traffic assignment, the subproblem solved for commodity $i \in \mathcal{C}$ in a Jacobi method, given a feasible flow $\mathbf{f}^{(l)}$, then is

[$\mathbf{TAP}_i^{(l)}$] $$\begin{aligned} \min \quad & \sum_{a \in \mathcal{A}} \int_0^{f_{ai} + \sum_{j \neq i} f_{aj}^{(l)}} t_a(s)\,ds \\ \text{s.t.} \quad & \mathbf{A}\mathbf{f}_i = \mathbf{d}_i \\ & \mathbf{f}_i \geq \mathbf{0}. \end{aligned}$$

The reader should note here that a decomposition of the original problem into independent subproblems is accomplished by any function Φ of the form

$$\Phi(\mathbf{x}, \mathbf{y}) = \sum_{i \in \mathcal{C}} \Phi_i(\mathbf{x}_i, \mathbf{y}), \tag{4.1}$$

where each function Φ_i is convex with respect to the variables $\mathbf{x}_i$.

An alternative to the parallel solution of the subproblems is to instead use the new information obtained from one subproblem when solving the next. Although this strategy is inherently sequential, the better utilization of information may result in a better practical convergence rate. This is the Gauss-Seidel approach, and for the variables $\mathbf{x}_i$ the subproblem is

$$[\mathbf{GS}_i^{(l)}] \qquad \mathbf{x}_i^{(l+1)} \in \arg \min_{\mathbf{x}_i \in \mathbf{X}_i} T\left(\mathbf{x}_1^{(l+1)}, \mathbf{x}_2^{(l+1)}, \ldots, \mathbf{x}_{i-1}^{(l+1)}, \mathbf{x}_i, \mathbf{x}_{i+1}^{(l)}, \ldots, \mathbf{x}_{|\mathcal{C}|}^{(l)}\right).$$

The Gauss-Seidel method is an instance of the partial linearization algorithm. The subproblem, defining one main iteration, may be described by the variational inequality

$$[\mathbf{GS}^{(l)}] \qquad \text{find } \mathbf{x}^{(l+1)} \in \mathbf{X} \text{ such that}$$

$$\varphi\left(\mathbf{x}^{(l+1)}, \mathbf{x}^{(l)}\right)^{\mathrm{T}}\left(\mathbf{x} - \mathbf{x}^{(l+1)}\right) \geq 0, \forall \mathbf{x} \in \mathbf{X},$$

where the ith block component of φ is $\varphi_i(\mathbf{x}, \mathbf{y}) = \nabla_i T\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{y}_{i+1}, \ldots, \mathbf{y}_{|\mathcal{C}|}\right)$. Since $\varphi(\mathbf{x}, \mathbf{x}) = \nabla T(\mathbf{x})$, Gauss-Seidel is a special case of partial linearization (compare [$\mathbf{GS}^{(l)}$] with [$\mathbf{PL\text{-}SUB\text{-}VIP}^{(l)}$]). Applying Gauss-Seidel to traffic equilibrium results in a subproblem for commodity i equal to [$\mathbf{TAP}_i^{(l)}$], with the objective replaced by $\int_0^{f_{ai} + \sum_{j<i} f_{aj}^{(l+1)} + \sum_{j>i} f_{aj}^{(l)}} t_a(s)ds$.

A large number of articles have been devoted to Jacobi and Gauss-Seidel approaches. Cyclic decomposition methods for [**TAP**] are presented in e.g. [20, 21, 22, 16, 27, 73, 82, 88]. In combination with simplicial decomposition/column generation (see [47, 51]), such algorithms have also been presented in [39, 55, 72, 74, 36, 34]. Jacobi/Gauss-Seidel methods have also been applied to the problem of optimal routing in computer communication networks (a problem with the same structure as [**TAP**]), often in combination with Newton type approaches. We would like to stress that most applications of Jacobi or Gauss-Seidel algorithms for traffic assignment is of the *successive approximation* type, i.e., the solution to the separable subproblem is taken as the new iterate point, while, in the partial linearization scheme a line search is included between the current iterate and the subproblem solution point. Not only may the global convergence criteria be weakened (it is well known that the convergence of Jacobi and Gauss-Seidel methods require, typically, strict convexity of either the whole objective or with respect to each independent variable block component, see e.g. [15, Section 3.3.5]), but also the practical convergence is expected to be much better by including line searches [31].

The partial linearization algorithm defines extensions of the Gauss-Seidel approach. A general Gauss-Seidel type algorithm is defined by an ordering $\{i_l\}_{l=1}^{\infty}$ of indices to be chosen, and a mapping φ such that, for the variable block component $\mathbf{x}_i$, the following variational inequality is solved.

$$[\mathbf{GS\text{-}PL}_i^{(l)}] \qquad \text{Find } \overline{\mathbf{x}}_i^{(l)} \in \mathbf{X}_i \text{ such that}$$

$$\left[\varphi_i\left(\overline{\mathbf{x}}_i^{(l)}, \mathbf{x}^{(l)}\right) + \nabla_i T\left(\mathbf{x}^{(l)}\right) - \varphi_i\left(\mathbf{x}_i^{(l)}, \mathbf{x}^{(l)}\right)\right]^{\mathrm{T}}\left(\mathbf{x}_i - \overline{\mathbf{x}}_i^{(l)}\right) \geq 0, \forall \mathbf{x}_i \in \mathbf{X}_i,$$

which reduces to a convex nonlinear program if φ_i is a gradient.

By letting the mapping φ_i be such that $\varphi_i(\mathbf{x}_i, \mathbf{x}) = \mathbf{0}$, we obtain the generic subproblem of Tseng [93]:

$[\mathbf{GS\text{-}NPD}_i^{(l)}]$ find $\overline{\mathbf{x}}_i^{(l)} \in \mathbf{X}_i$ such that

$$\left[\varphi_i\left(\overline{\mathbf{x}}_i^{(l)}, \mathbf{x}^{(l)}\right) + \nabla_i T\left(\mathbf{x}^{(l)}\right)\right]^{\mathrm{T}}\left(\mathbf{x}_i - \overline{\mathbf{x}}_i^{(l)}\right) \geq 0, \forall \mathbf{x}_i \in \mathbf{X}_i.$$

Tseng extends the Gauss-Seidel type approach by including an inexact line search with respect to the original cost in the direction $\overline{\mathbf{x}}_i^{(l)} - \mathbf{x}_i^{(l)}$, after each subproblem. Furthermore, the sequence of indices i_l chosen is not restricted to be the standard cyclic order ($i_l = 1 + l \bmod |\mathcal{C}|$). Rather, in order to establish convergence, the sequence of indices $\{i_l\}$ must satisfy the condition that there exists a constant $B \geq |\mathcal{C}|$ such that every index is chosen at least once every B successive iterations (essentially cyclic rule).

Based on the results obtained in this paper we may extend Tseng's Gauss-Seidel type algorithm in several directions. In view of the results of Section 3.2, the subproblems need not be solved exactly in order to establish convergence (cf. [15, Prop. 3.3.8] and [68]). Furthermore, other orders of iteration are possible, such as variants of the Gauss-Southwell (remotest) order [17], in which the index chosen corresponds to variables being, in some measure, farthest from the optimum. In the traffic assignment context an algorithm based on such an ordering is established convergent in [82]. In iteration l the commodity to be considered is chosen based on the error in the Wardrop conditions, using dual prices obtained from the Frank-Wolfe subproblem. A simpler way of using the Frank-Wolfe algorithm in the choice of commodity in a Gauss-Seidel type scheme is to instead use the value of the primal gap function (see Section 5.3), which is directly available from the computation of shortest routes. Evaluating, for each commodity, the current total cost of transportation minus the total cost of transportation along a shortest route provides a measure of the distance (gap) from the optimum for each commodity. The commodity chosen in iteration l then is the one with the largest (relative) gap. The convergence of such an algorithm is established analogously to [82].

Other Jacobi type methods for [**TAP**] have been presented in [19, 50]. These are briefly discussed below, and shown to be instances of the general algorithm.

In [19] a *shifted function* $h(\mathbf{b}) = T(\mathbf{b} + \mathbf{f}) - T(\mathbf{f})$ is defined for feasible flows $\mathbf{f}$. Then [**TAP**] is equivalent to minimizing $h(\mathbf{b})$ subject to $\mathbf{b} : \mathbf{b} + \mathbf{f}$ is feasible. By defining h_a to be the term of h corresponding to arc a, they then define a *scaled separable function* ($\sigma > 0$ is a constant)

$$\begin{aligned} h^s(\mathbf{b}, \sigma) &= \sum_a \sum_i \frac{1}{\sigma} h_a(\sigma b_{ai}) = \\ &= \sum_a \sum_i \frac{1}{\sigma}\left(\int_0^{\sigma b_{ai} + \sum_i f_{ai}^{(l)}} t_a(s) ds - \int_0^{\sum_i f_{ai}^{(l)}} t_a(s) ds\right), \end{aligned}$$

which is minimized with respect to $\mathbf{b}$. Associating the scaled separable function h^s with the function Φ we may conclude that the algorithm is a partial linearization algorithm. It should be noted that for the special case of $\sigma = 1$, the algorithm reduces to the Jacobi algorithm.

In [50] an algorithm is defined, which also is highly related to the Jacobi method with line searches. Here, the separable approximation defining Φ is $\sum_{i \in \mathcal{C}} \sum_{a \in \mathcal{A}} \int_0^{f_{ai}} t_a(s) ds$.

How the choice of separable functions Φ, i.e., of the form (4.1), induces parallel partial linearization methods is shown below.

4.2 Parallel asynchronous implementations of partial linearization methods

Partial linearization algorithms, when applied to problems over Cartesian product sets, may also be implemented in a (partially) asynchronous manner. Assume that we have access to a parallel computer, where we can allocate each of the $|\mathcal{C}|$ single-commodity problems to one independent processor. In a *synchronized* implementation each processor must halt at predetermined points to exchange information (in this context commodity flows). In an *asynchronous* implementation local processors do not wait to receive the latest information available. The advantage of such an approach is a minimal communication delay and a potential speed-up over synchronized algorithms. The drawback is that the conditions of convergence are more severe than for the synchronized version, and that the termination criteria are not as easily detected. In a *partially asynchronous* implementation there is a bound on the communication delays, i.e., the information that any processor holds is not arbitrarily old. We follow the definition given in [15, p. 483], and introduce a set T^i of times at which $\mathbf{x}_i$ is updated by processor i, variables $\tau_j^i(t)$ for each $i, j \in \mathcal{C}$ and t, measuring the amount by which the information used in an update of $\mathbf{x}_i$ is outdated, with $0 \leq \tau_j^i(t) \leq t$. By partial asynchronism we mean that there exists a positive constant B such that for each index i and for each $t \geq 0$ at least one element in the set $\{t, t+1, \ldots, t+B-1\}$ belongs to T^i, $t - B < \tau_j^i(t) \leq t$ for each i, j and $t \in T^i$, and $\tau_i^i(t) = t$ for all i and $t \in T^i$. (Note the similarities between the definitions of partial asynchronism and the essentially cyclic rule.)

To establish convergence of a (partially) asynchronous implementation of the partial linearization algorithm, we make the additional assumptions that ∇T is Lipschitz continuous with modulus K, and that the function Φ is of the form (4.1), where each function $\Phi_i(\mathbf{x}_i, \mathbf{y})$ is *strongly convex* with modulus m_{Φ_i}, i.e.,

$$\left[\nabla\Phi_i\left(\mathbf{x}_i^1, \mathbf{y}\right) - \nabla\Phi_i\left(\mathbf{x}_i^2, \mathbf{y}\right)\right]^{\mathrm{T}}\left(\mathbf{x}_i^1 - \mathbf{x}_i^2\right) \geq m_{\Phi_i}\left\|\mathbf{x}_i^1 - \mathbf{x}_i^2\right\|_2^2, \forall \mathbf{x}_i^1, \mathbf{x}_i^2 \in \mathbf{X}_i.$$

We also define $m_\Phi = \min_i m_{\Phi_i}$. In order to carry out the computations in an asynchronous fashion the line search is replaced by a predetermined step length, $\alpha > 0$. The following convergence result extends the ones obtained in [15, Props. 3.3.4 and 7.5.3] for a general algorithmic scheme in unconstrained optimization and for gradient projection type methods, respectively. (The fact that gradient projection is an instance of the partial linearization algorithm is shown in Section 4.3.) Since the theorem is easily established from these propositions, the proof is omitted.

Theorem 4.1 *Assume that T is pseudo-convex and has a Lipschitz continuous gradient, and that the functions Φ_i are strongly convex. Then the synchronized partial linearization algorithm using fixed step lengths α converges globally if $0 < \alpha < \frac{2m_\Phi}{K}$. In the partially asynchronous version, we obtain the same result if steps are chosen in the interval $0 < \alpha < \frac{m_\Phi}{K}\frac{1}{\frac{1}{2}+(n+1)B}$.*

For a discussion on the roles of the different parameters in the step length restriction, see [15, Section 7.5.2]. In particular, if the problem is *weakly coupled* with respect to the commodities, i.e., the interaction between commodity flows in the network is modest, then the second result obtained in Theorem 4.1 may be improved in the sense that a longer step α is allowed. In the special case of gradient projection type methods [56, 8], which is also the algorithm considered in [15, Section 7.5], Tseng [94] establishes a linear convergence rate for the partially asynchronous version. (This result extends the linear convergence result obtained in e.g. [58], for the basic method.)

We conclude this section by studying Newton type approaches for separable traffic assignment problems.

4.3 Newton-type methods

We first show that the constrained Newton method is an instance of the partial linearization algorithm.

Given $\mathbf{y} \in \mathbf{X}$, let the function Φ be given by $\Phi(\mathbf{x}, \mathbf{y}) = \frac{1}{2}(\mathbf{x}-\mathbf{y})^{\mathrm{T}} \nabla^2 T(\mathbf{y})(\mathbf{x}-\mathbf{y})$. The subproblem of the partial linearization algorithm corresponding to this function then is to minimize $\nabla T(\mathbf{y})^{\mathrm{T}}(\mathbf{x}-\mathbf{y}) + \frac{1}{2}(\mathbf{x}-\mathbf{y})^{\mathrm{T}} \nabla^2 T(\mathbf{y})(\mathbf{x}-\mathbf{y})$ over $\mathbf{x} \in \mathbf{X}$, which is equivalent to the constrained Newton subproblem.

The function $\Phi(\mathbf{x}, \mathbf{y})$ defined above is convex with respect to $\mathbf{x}$ for all values of $\mathbf{y}$ if the Hessian matrix $\nabla^2 T(\mathbf{y})$ is positive semidefinite for all feasible points, which holds for convex functions T. The convergence of the constrained Newton algorithm is then guaranteed by the convergence of the partial linearization method. Furthermore, if T is strongly convex, this convergence is known to be superlinear [85, Theorem III.3.6]. If the Hessian is also Lipschitz continuous everywhere on $\mathbf{X}$, the convergence is quadratic [85, Theorem III.3.5]. Hence, the class of partial linearization methods define methods with convergence rates ranging from sublinear (the Frank-Wolfe method, by the choice of $\Phi \equiv 0$) to quadratic (the constrained Newton method). It is also possible to show that for a convex problem [**P**] with a *sharp minimum* [84], i.e., if there exists an $\alpha > 0$ such that, for all $\mathbf{x} \in \mathbf{X}$,

$$T(\mathbf{x}) - T(\mathbf{x}^*) \geq \alpha \|\mathbf{x} - \mathbf{x}^*\|,$$

then the partial linearization algorithm is *finite*, whenever ∇T is Lipschitz continuous, and the function Φ is chosen convex with a Lipschitz continuous gradient $\nabla_{\mathbf{x}} \Phi(\cdot, \mathbf{y})$. (This result generalizes those obtained for the proximal point algorithm [86], gradient projection [84], constrained Newton [84] and Frank-Wolfe [84].)

From the above discussion on regularizations of the Frank-Wolfe algorithm we may interpret the Newton method as a special case, where the regularizing term is given by a particular matrix norm, see [69].

Approximate Newton methods are defined through approximations $\mathbf{B}^{(l)}$ of the Hessian matrix, for instance as block diagonal matrices, making the corresponding subproblems separable with respect to the commodities, when applied to traffic equilibrium. In [40, 56] the matrix is chosen as the identity matrix, while in [8] it is chosen positive diagonal. (These methods are often referred to as *(scaled) gradient projection algorithms.*) Such algorithms are seen to be instances of the partial linearization methods, by replacing $\nabla^2 T\left(\mathbf{x}^{(l)}\right)$ with $\mathbf{B}^{(l)}$ in the above derivation of Newton's method.

Newton algorithms of the types mentioned above have been applied to routing and traffic equilibrium (e.g. [9, 28]), and incorporating simplicial decomposition/column generation in [10, 13, 79, 45, 46, 51]. Dembo and Tulowitzki [28] present a Newton method for the arc-node formulation of [**TAP**], based on the use of truncated Frank-Wolfe iterations. Using a subproblem termination criterion based on decreasing relative errors, the algorithm is shown to converge with a superlinear rate. They do not address the problem with negative arc costs (cf. Section 3.2), which could imply that in practice this is not an issue. It is interesting to note their conclusion regarding the number of Frank-Wolfe iterations that should be considered to maximize the efficiency of the algorithm. Their conclusion is that not more than four Frank-Wolfe steps should be considered in any iteration.

5 Partial linearization for variational inequalities

In this section we consider the variational inequality formulation of the general traffic equilibrium problem, arising from the first Wardrop conditions when the cost interactions between user classes, or between the flows on different links, are not symmetric. The possibility of formulating merit functions for asymmetric variational inequalities are also discussed. The partial linearization algorithm will then be extended to variational inequality problems, and some algorithmic equivalence results will be given. Based on these results we will, in the next section, give a unified description of some well known traffic equilibrium methods. The extension to descent methods through the use of merit functions will also be addressed.

It is shown in e.g. [22] that the Wardrop conditions of user equilibrium may be reformulated as a convex mathematical program under the assumption that the travel time function $\mathbf{t}$ has a symmetric Jacobian matrix for all feasible flows, i.e., $\mathbf{t}'(\mathbf{f})$ is symmetric for all feasible $\mathbf{f}$. When, in the general case, arc costs are asymmetric, an equivalent mathematical program corresponding to user equilibrium can not be derived by using the same arguments as used when formulating [**TAP**]. The Wardrop conditions can instead be reformulated into other well known problem types, such as variational inequality, nonlinear complementarity or fixed point problems. In this paper we consider the variational inequality formulation. Through the variational inequality concept equivalent mathematical programs have been presented. Algorithms based on merit functions are typically based on the solution of convex subproblems or affine variational inequalities, and nontrivial line searches with respect to these functions.

Before formulating the variational inequality problem, equivalent to the Wardrop conditions for user equilibrium, we will discuss some important properties of variational inequalities in finite-dimensional space.

5.1 Variational inequalities

Consider a continuous mapping $\mathbf{F}(\mathbf{x}) : \mathbf{X} \mapsto \Re^n$. Throughout this exposition $\mathbf{X}$ will be assumed nonempty, closed and convex.

The Variational Inequality Problem is defined as

[**VIP**] find $\mathbf{x}^* \in \mathbf{X}$ such that $\mathbf{F}(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^*) \geq 0, \forall \mathbf{x} \in \mathbf{X}$.

A solution to [**VIP**] exists under certain regularity assumptions on $\mathbf{X}$ and $\mathbf{F}$, and a solution $\mathbf{x}^*$ to [**VIP**] is unique under the condition that $\mathbf{F}$ is *strictly monotone*, i.e., that

$$[\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})]^{\mathrm{T}} (\mathbf{x} - \mathbf{y}) > 0, \forall \mathbf{x}, \mathbf{y} \in \mathbf{X} \text{ and } \mathbf{x} \neq \mathbf{y}.$$

The problem [**VIP**] is related to nonlinear programming by the following.

Theorem 5.1 [6, Theorem 4.2] *Let T be continuously differentiable with* $\mathbf{F}(\mathbf{x}) = \nabla T(\mathbf{x})$. *If* $\mathbf{x}^*$ *solves* $\min_{\mathbf{x}\in\mathbf{X}} T(\mathbf{x})$, *then* $\mathbf{x}^*$ *solves* [**VIP**]. *The converse holds whenever T is convex.*

$\mathbf{F}(\mathbf{x})$ is a gradient of a function $T(\mathbf{x})$ under the condition that the Jacobian, $\mathbf{F}'(\mathbf{x})$, is symmetric for all $\mathbf{x} \in \mathbf{X}$ [75, Theorem 4.1.6], and the problem [**VIP**] can hence be put into an equivalent mathematical programming problem as follows:

$$[\mathbf{P}] \qquad \min_{\mathbf{x}\in\mathbf{X}} T(\mathbf{x}) \Leftrightarrow \min_{\mathbf{x}\in\mathbf{X}} \oint_{\mathbf{0}}^{\mathbf{x}} \mathbf{F}(\mathbf{s})d\mathbf{s}.$$

The following two theorems provide basic tools in algorithmic procedures for solving [**VIP**].

Theorem 5.2 [75, Theorem 5.1.3] *Let* $\mathbf{X}$ *be closed in* $\Re^n$ *and let* $\mathbf{T} : \mathbf{X} \mapsto \mathbf{X}$ *be a* contraction mapping with modulus K, *i.e., that* $\|\mathbf{T}(\mathbf{x}) - \mathbf{T}(\mathbf{y})\| \leq K\|\mathbf{x} - \mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in \mathbf{X}$, *where* $K \in [0, 1)$ *and* $\|\cdot\|$ *is some norm. Then* $\mathbf{T}$ *has a unique fixed point, and the sequence* $\mathbf{x}^{(l+1)} = \mathbf{T}(\mathbf{x}^{(l)}), l = 0, 1, \ldots$, *converges to the fixed point geometrically.*

Theorem 5.3 [29] *Let* γ *be a positive constant and* $\mathbf{G}$ *a symmetric and positive definite matrix. Then* $\mathbf{x}^*$ *solves* [**VIP**] *if and only if* $\mathbf{x}^* = \mathbf{P}_{\mathbf{X}}^{\mathbf{G}}\left[\mathbf{x}^* - \gamma\mathbf{G}^{-1}\mathbf{F}(\mathbf{x}^*)\right]$, *where* $\mathbf{P}_{\mathbf{X}}^{\mathbf{G}}[\cdot]$ *is the projection on* $\mathbf{X}$ *with respect to the norm* $\|\mathbf{x}\|_{\mathbf{G}} = (\mathbf{x}^{\mathrm{T}}\mathbf{G}\mathbf{x})^{1/2}$.

Theorem 5.3 provides a fixed point characterization of a solution to [**VIP**]. Furthermore, since $\mathbf{T}(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{P}_{\mathbf{X}}^{\mathbf{G}}\left[\mathbf{x} - \gamma\mathbf{G}^{-1}\mathbf{F}(\mathbf{x})\right]$ is a contraction mapping if $\mathbf{F}$ is strongly monotone and Lipschitz continuous, and if γ is sufficiently small, a convergent algorithm may be defined by employing the iterations

$$\mathbf{x}^{(l+1)} = \mathbf{T}\left(\mathbf{x}^{(l)}\right), l = 0, 1, \ldots. \tag{5.1}$$

This methodology is referred to as the *projection algorithm* (e.g. [23]).

Another fixed point characterization of solutions to variational inequalities is given below.

Theorem 5.4 *Let the problem* [**VIP**] *be given as above. Define the point-to-set mapping*

$$W(\mathbf{x}) = \arg\max_{\mathbf{y}\in\mathbf{X}} \mathbf{F}(\mathbf{x})^{\mathrm{T}}(\mathbf{x} - \mathbf{y}). \tag{5.2}$$

Then $\mathbf{x}^*$ *solves* [**VIP**] *if and only if* $\mathbf{x}^*$ *is a fixed point of the mapping* $W(\mathbf{x})$, *i.e.,* $\mathbf{x}^* \in W(\mathbf{x}^*)$.

The theorem is developed from the text given in Zuhovickiĭ *et al.* [102] (see also [43, 62]). The latter fixed point formulation does not directly define an algorithm as the projection characterization does, since the mapping W cannot be shown to be contractive. (In the context of traffic equilibrium, the algorithm is equivalent to the heuristic, iterated all-or-nothing assignment algorithm. However, it provides a nice interpretation of the Wardrop conditions of equilibria, see [80, Section 5.4].) Both above fixed point characterizations will be shown to be special cases of a general fixed point characterization, given by a generalization of Theorem 3.1.

For further discussions on variational inequalities we refer to the book by Kinderlehrer and Stampacchia [48], and the recent survey by Harker and Pang [42].

5.2 Traffic equilibria via variational inequality and mathematical programming reformulations

We begin by stating the Wardrop conditions of user equilibrium, and state the equivalent variational inequality formulation.

[**W**]
$$
\begin{array}{rcllr}
h^*_{pqr}\left(c_{pqr}(\mathbf{h}^*) - \pi_{pq}\right) & = & 0 & \forall r \in \mathcal{R}_{pq} \quad \forall (p,q) \in \mathcal{C} & (1)\\
c_{pqr}(\mathbf{h}^*) - \pi_{pq} & \geq & 0 & \forall r \in \mathcal{R}_{pq} \quad \forall (p,q) \in \mathcal{C} & (2)\\
\sum_{r \in \mathcal{R}_{pq}} h^*_{pqr} & = & d_{pq} & \qquad\qquad \forall (p,q) \in \mathcal{C} & (3)\\
h^*_{pqr} & \geq & 0 & \forall r \in \mathcal{R}_{pq} \quad \forall (p,q) \in \mathcal{C} & (4)
\end{array}
$$

where π_{pq} denotes the equilibrium travel time on the utilized routes for commodity (p,q), d_{pq} the demand for commodity (p,q) and $\mathcal{R}_{pq}$ the set of simple routes between p and q, and c_{pqr} the travel time on route r resulting from the route flow $\mathbf{h}^*$.

The Wardrop conditions of user equilibrium can be shown (see e.g. [90]) to be equivalent to the following variational inequality formulation, defining the general Traffic Equilibrium Problem:

[**TEP-VIP**] find $\mathbf{h}^* \in \mathbf{X}$ such that $\mathbf{c}(\mathbf{h}^*)^{\mathrm{T}}(\mathbf{h} - \mathbf{h}^*) \geq 0, \forall \mathbf{h} \in \mathbf{X}$,

where $\mathbf{X}$ is defined by the demand feasibility constraints, i.e., the constraints (3) and (4) of [**W**]. Using the arc-route flow relationship, the problem of traffic assignment with fixed demands and additive costs can be cast in terms of arc flows:

[**TEP-VIP**] find $\mathbf{f}^* \in \mathbf{X}$ such that $\mathbf{t}(\mathbf{f}^*)^{\mathrm{T}}(\mathbf{f} - \mathbf{f}^*) \geq 0, \forall \mathbf{f} \in \mathbf{X}$,

where $\mathbf{X}$ is defined by the commodity flow conservation constraints and nonnegativity restrictions. The first to formulate this problem as a variational inequality are Smith [90] and Dafermos [23]. The existence and uniqueness properties of traffic equilibria are analyzed by studying the properties of the corresponding formulation as a variational inequality, nonlinear complementarity or fixed point problem.

We now turn our attention to the possible reformulations of variational inequalities as mathematical programs, without imposing a symmetry restriction on the Jacobian of the mapping $\mathbf{F}$.

5.3 A class of generalized gap functions

When the Wardrop conditions for traffic equilibrium problems with asymmetric travel costs were equivalently formulated as variational inequalities, and other problem types, the lack of equivalent mathematical programming formulation was taken almost as self-evident. However, such formulations had, in saddle point problems and game theory, been known for more than ten years among Soviet researchers, see e.g. [102]. Because of the fact that variational inequalities are special cases of games, mathematical programming formulations existed even before models involving asymmetry had become a research topic. For relations between saddle point problems, variational inequalities and games, in general terms as well as for traffic equilibria, the reader is referred to [80].

The reformulation of asymmetric variational inequalities as mathematical programs involves the use of so called *gap functions* as valid merit functions. As defined in [52], a gap function is a function from $\mathbf{X}$ to $\Re$ that is nonnegative for all feasible arguments, and equals zero exactly at solutions to [**VIP**]. The consequence of this definition is that by minimizing a gap function, a solution to the variational inequality is obtained. Even though these functions are nonconvex in general, under appropriate assumptions stationary points to the minimization problem can be shown to coincide with solutions to the variational inequality.

Gap functions are highly related to the partial linearization algorithm, as will be shown below. The following is developed from two recent papers on gap functions and variational inequalities, Auchmuty [5] and Larsson and Patriksson [52].

If ∇T is replaced by $\mathbf{F}$ in [$\mathbf{PL\text{-}SUB}^{(l)}$], the subproblem is equivalent to the convex problem of calculating

$$\tilde{G}\left(\mathbf{x}^{(l)}\right) = \sup_{\mathbf{x}\in\mathbf{X}} \ \tilde{L}\left(\mathbf{x}, \mathbf{x}^{(l)}\right),$$

where

$$\tilde{L}(\mathbf{x},\mathbf{y}) = \Phi(\mathbf{y},\mathbf{y}) - \Phi(\mathbf{x},\mathbf{y}) + [\mathbf{F}(\mathbf{y}) - \nabla_{\mathbf{x}}\Phi(\mathbf{y},\mathbf{y})]^{\mathrm{T}}(\mathbf{y}-\mathbf{x}).$$

We will show that the class of functions $\tilde{G}$ define a class of gap functions.

Theorem 5.5 *Let $\Phi(\mathbf{x},\mathbf{y})$ be convex and continuously differentiable with respect to $\mathbf{x}$, and continuous with respect to $\mathbf{y}$. Then $\tilde{G}$ is a gap function.*

Proof For any $\mathbf{x}\in\mathbf{X}$, $\tilde{G}(\mathbf{x}) \geq 0$, since $\tilde{L}(\mathbf{x},\mathbf{x}) = 0$.

Assume that $\mathbf{x}^*$ solves [**VIP**]. Then

$$\begin{aligned}
\tilde{L}(\mathbf{x},\mathbf{x}^*) &= \Phi(\mathbf{x}^*,\mathbf{x}^*) - \Phi(\mathbf{x},\mathbf{x}^*) + [\mathbf{F}(\mathbf{x}^*) - \nabla_{\mathbf{x}}\Phi(\mathbf{x}^*,\mathbf{x}^*)]^{\mathrm{T}}(\mathbf{x}^*-\mathbf{x}) \\
&\leq \mathbf{F}(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{x}^*-\mathbf{x}) \\
&\leq 0, \forall \mathbf{x}\in\mathbf{X},
\end{aligned}$$

due to the convexity of $\Phi(\mathbf{x},\mathbf{y})$ with respect to $\mathbf{x}$, and the fact that $\mathbf{x}^*$ solves [**VIP**], respectively. Since $\tilde{L}(\mathbf{x}^*,\mathbf{x}^*) = 0$, we have that $\tilde{G}(\mathbf{x}^*) = 0$.

Conversely, assume that $\tilde{G}(\mathbf{x}^*) = 0$. Then $\tilde{L}(\mathbf{x},\mathbf{x}^*) \leq 0, \forall\mathbf{x}\in\mathbf{X}$. But since $\tilde{L}(\mathbf{x}^*,\mathbf{x}^*) = 0$, $\mathbf{x}^*$ must be a solution to the inner problem defined in the calculation of $\tilde{G}(\mathbf{x}^*)$. The variational inequality that is defined by this inner problem then yields that $\mathbf{x}^*$ solves [**VIP**]. $\square$

Special cases of this class of gap functions are recovered for $\Phi \equiv 0$ [102, 43, 44, 61, 62, 66] (cf. Theorem 5.4), $\Phi(\mathbf{x},\mathbf{y}) = \frac{1}{2}\|\mathbf{x}\|^2_{\mathbf{G}}$ [38] (cf. Theorem 5.3) and $\Phi(\mathbf{x},\mathbf{y}) \equiv \Phi(\mathbf{x})$ [5, 52], where $\mathbf{G}$ is a positive definite and symmetric matrix. In [98] the algorithmic class of [69, 93] is extended to variational inequalities, and the gap functions developed in these references are therefore equivalent to the ones given here in the sense of Theorem 5.7 below.

The following theorem provides a new fixed point characterization of solutions to [**VIP**]. (The characterization extends the ones given in Theorems 5.3 and 5.4, and is a generalization of the first part of Theorem 3.1.) The proof follows immediately from Theorem 5.5.

Theorem 5.6 $\mathbf{x}^*$ solves [**VIP**] $\Leftrightarrow \mathbf{x}^* \in \ \arg\max_{\mathbf{x}\in\mathbf{X}} \ \tilde{L}(\mathbf{x},\mathbf{x}^*)$.

The basis for the convergence results in gap minimization methods is the fact that a local minimum to $\min_{\mathbf{x}\in\mathbf{X}} \tilde{G}(\mathbf{x})$ is a global one under suitable assumptions on the function Φ defining $\tilde{G}$ [52]. The convergence analysis made in [52] thus suggests the possibility of extending algorithms for variational inequalities by including line searches with respect to a gap function. The class of gap functions given here are given through the subproblem of the partial linearization algorithm. Gap decreasing algorithms can therefore be seen as a natural extension of partial linearization methods to asymmetric variational inequalities.

The most practical cases of gap functions are, seemingly, those described by functions $\Phi(\mathbf{x})$ rather than by functions $\Phi(\mathbf{x}, \mathbf{y})$. This is because the convergence conditions in the former case can be formulated in terms of the original data [52]. This is not the case for functions $\Phi(\mathbf{x}, \mathbf{y})$, where convergence conditions are rather difficult to verify (see [98]). Note here that, for a Cartesian product feasible set, by choosing the function Φ as a separable function with respect to this set, i.e., a function of the form (4.1), the calculation of $\tilde{G}$ can be done in parallel. An algorithm based on the minimization of such a gap function is therefore highly parallelizable, also in the line search phase.

5.4 The partial linearization algorithm extended to variational inequalities

In this section we extend the partial linearization algorithm to asymmetric variational inequalities. We will also establish some algorithmic equivalence results, in that we show that the generic subproblems of Dafermos [24] and Tseng [93] are equivalent to the subproblem of the partial linearization algorithm. These three algorithmic classes are thus equivalent apart from the inclusion of line searches with respect to a merit function.

Let the function $\varphi : \mathbf{X} \times \mathbf{X} \mapsto \Re^n$ be continuously differentiable and monotone with respect to its first argument and continuous with respect to its second. In iteration l rewrite

$$\mathbf{F}(\mathbf{x}) = \varphi\left(\mathbf{x}, \mathbf{x}^{(l)}\right) + \left(\mathbf{F}(\mathbf{x}) - \varphi\left(\mathbf{x}, \mathbf{x}^{(l)}\right)\right).$$

The error in approximating the original mapping $\mathbf{F}$ by φ is taken into account by fixing the values of $\mathbf{x}$ to $\mathbf{x}^{(l)}$ within the parenthesis. (This approximation is equivalent to the one made in Section 3.1, with $\mathbf{F}$ being the gradient of a function T). A choice of $\varphi(\mathbf{x}, \mathbf{y})$ as a gradient mapping with respect to $\mathbf{x}$, i.e., with $\varphi(\mathbf{x}, \mathbf{y}) = \nabla_{\mathbf{x}}\Phi(\mathbf{x}, \mathbf{y})$ for some function $\Phi : \mathbf{X} \times \mathbf{X} \mapsto \Re$, convex and continuously differentiable with respect to $\mathbf{x}$ and continuous with respect to $\mathbf{y}$, results in a subproblem equivalent to the calculation of a gap function $\tilde{G}$.

The subproblem solved in iteration l is the following variational inequality.

[PL-SUB-VIP$^{(l)}$] Find $\overline{\mathbf{x}}^{(l)} \in \mathbf{X}$ such that

$$\left[\varphi\left(\overline{\mathbf{x}}^{(l)}, \mathbf{x}^{(l)}\right) + \mathbf{F}\left(\mathbf{x}^{(l)}\right) - \varphi\left(\mathbf{x}^{(l)}, \mathbf{x}^{(l)}\right)\right]^{\mathrm{T}} \left(\mathbf{x} - \overline{\mathbf{x}}^{(l)}\right) \geq 0, \forall \mathbf{x} \in \mathbf{X}$$

Dafermos [24] defines a smooth function $\mathbf{g} : \mathbf{X}\times\mathbf{X} \mapsto \Re^n$, with the properties (i) $\mathbf{g}(\mathbf{x}, \mathbf{x}) = \mathbf{F}(\mathbf{x})$ for all $\mathbf{x} \in \mathbf{X}$, and (ii) $\nabla_{\mathbf{x}}\mathbf{g}(\mathbf{x}, \mathbf{y})$ is symmetric and positive definite for any fixed $\mathbf{x}, \mathbf{y} \in \mathbf{X}$. She then defines a successive approximation algorithm through the following variational inequality:

[DAF$^{(l)}$] find $\mathbf{x}^{(l+1)} \in \mathbf{X}$ such that

$$\mathbf{g}\left(\mathbf{x}^{(l+1)}, \mathbf{x}^{(l)}\right)^{\mathrm{T}} \left(\mathbf{x} - \mathbf{x}^{(l+1)}\right) \geq 0, \forall \mathbf{x} \in \mathbf{X}.$$

The sequence of points generated by the algorithm is proven convergent towards the solution to [**VIP**] under the assumption that

$$\left\| \left| \nabla_{\mathbf{x}} \mathbf{g}^{-1/2}\left(\mathbf{x}^1, \mathbf{y}^1\right) \cdot \nabla_{\mathbf{y}} \mathbf{g}\left(\mathbf{x}^2, \mathbf{y}^2\right) \cdot \nabla_{\mathbf{x}} \mathbf{g}^{-1/2}\left(\mathbf{x}^3, \mathbf{y}^3\right) \right| \right\| < 1, \forall \mathbf{x}^i, \mathbf{y}^i \in \mathbf{X}, \tag{5.3}$$

for $i = 1, 2, 3$, and where $|||\cdot|||$ denotes the operator norm on $\mathcal{L}(\Re^n, \Re^n)$ induced by the Euclidean norm $\|\cdot\|_2$ on $\Re^n$.

It will be shown below that this class of methods is, in a particular sense, equivalent to the partial linearization methods and also to the algorithms defined by Tseng [93], extended to variational inequalities.

For easier reference we state again the generic subproblem of Tseng [93], for the case of a single block component, and applied to variational inequalities (cf. [$\mathbf{GS\text{-}NPD}_i^{(l)}$] and the variational inequality characterization of (3.2)).

[$\mathbf{NPD}^{(l)}$] Find $\overline{\mathbf{x}}^{(l)} \in \mathbf{X}$ such that

$$\left[\varphi\left(\overline{\mathbf{x}}^{(l)}, \mathbf{x}^{(l)}\right) + \mathbf{F}\left(\mathbf{x}^{(l)}\right)\right]^{\mathrm{T}} \left(\mathbf{x} - \overline{\mathbf{x}}^{(l)}\right) \geq 0, \forall \mathbf{x} \in \mathbf{X}$$

Theorem 5.7 *The generic subproblems of Dafermos [24] and Tseng [93] are equivalent to that of the partial linearization algorithm.*

Proof We will establish the equivalence through three inclusions.

1. ([$\mathbf{NPD}^{(l)}$] $\subseteq$ [$\mathbf{PL\text{-}SUB\text{-}VIP}^{(l)}$]) If the mapping $\varphi(\mathbf{x}, \mathbf{y})$ is chosen so that $\varphi(\mathbf{x}, \mathbf{x}) = \mathbf{0}$, [$\mathbf{PL\text{-}SUB\text{-}VIP}^{(l)}$] reduces to [$\mathbf{NPD}^{(l)}$].
2. ([$\mathbf{PL\text{-}SUB\text{-}VIP}^{(l)}$] $\subseteq$ [$\mathbf{DAF}^{(l)}$]) For a given choice of mapping $\varphi(\mathbf{x}, \mathbf{y})$, choose the mapping $\mathbf{g}(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x}, \mathbf{y}) + \mathbf{F}(\mathbf{y}) - \varphi(\mathbf{y}, \mathbf{y})$. We then observe that $\mathbf{g}(\mathbf{x}, \mathbf{x}) = \mathbf{F}(\mathbf{x})$, so that we recover a subproblem of the type [$\mathbf{DAF}^{(l)}$].
3. ([$\mathbf{DAF}^{(l)}$] $\subseteq$ [$\mathbf{NPD}^{(l)}$]) For a given choice of mapping $\mathbf{g}(\mathbf{x}, \mathbf{y})$, choose the mapping $\varphi(\mathbf{x}, \mathbf{y}) = \mathbf{g}(\mathbf{x}, \mathbf{y}) - \mathbf{F}(\mathbf{y})$. From the property ($i$) of $\mathbf{g}$ above, $\varphi(\mathbf{x}, \mathbf{x}) = \mathbf{0}$, so that a subproblem of the type [$\mathbf{NPD}^{(l)}$] is defined. □

There are several implications of this algorithmic equivalence result. A variety of algorithms for variational inequalities have been analyzed in the framework of [24], especially different equilibrium problems in traffic networks, spatial oligopolistic markets and Walrasian price and general economic equilibrium problems (see e.g. [26, 101] and the references cited therein). The convergence condition (5.3) may, in some cases, be too restrictive (see e.g. [76]), and are, in general, difficult to establish to hold in a practical application. The discussions made in Section 5.3 imply the possibility of including line searches in successive approximation methods, and thereby obtain a potential speed-up as well as both less restrictive and more easily examined convergence conditions. (An example of such an algorithm is Newton's algorithm, for which line search rules have been proposed in e.g. [62, 63, 64, 65, 66, 77, 91].)

The fixed point characterization given in Theorem 5.6 carries over to the more general algorithm treated here. The proof follows also easily from Theorem 5.5.

Theorem 5.8 $\mathbf{x}^*$ *solves* [**VIP**] $\Leftrightarrow \mathbf{x}^*$ *solves* [**PL-SUB-VIP**$^{(l)}$] *defined at* $\mathbf{x}^*$.

As an example of the partial linearization algorithms for variational inequalities, we show that the proximal minimization algorithm [67] is an instance of this class of methods. The algorithm proceeds as follows.

Find $\mathbf{x}^{(l+1)} \in \mathbf{X}$ such that

$$\left[\mathbf{F}\left(\mathbf{x}^{(l+1)}\right) + \left(\mathbf{x}^{(l)} - \mathbf{x}^{(l+1)}\right)\right]^{\mathrm{T}} \left(\mathbf{x} - \mathbf{x}^{(l+1)}\right) \geq 0, \forall \mathbf{x} \in \mathbf{X}.$$

Let $\varphi(\mathbf{x}, \mathbf{y}) = \mathbf{F}(\mathbf{x}) + (\mathbf{y} - \mathbf{x})$. Then [**PL-SUB-VIP**$^{(l)}$] is equivalent to the above variational inequality problem, with $\overline{\mathbf{x}}^{(l)} = \mathbf{x}^{(l+1)}$, so that the proximal minimization algorithm is a special case of partial linearization methods for variational inequalities.

We next show that a number of well known successive approximation algorithms applied to [**TEP-VIP**] may be recognized as partial linearization algorithms, excluding line searches.

6 Partial linearization for asymmetric traffic equilibria

6.1 Diagonalization/Relaxation methods

The class of Jacobi/Gauss-Seidel methods discussed earlier have also been extended to general traffic equilibria. The methods are referred to as *diagonalization methods* or *relaxation methods*, since the approximation of the mapping $\mathbf{F}$ in Jacobi/Gauss-Seidel methods define mappings with block diagonal Jacobians, and since the interaction between commodities, user classes or arc flows are relaxed. The corresponding variational inequality then is equivalent to a convex mathematical program or a sequence of such programs, solvable by the same methods as applied to [**TAP**]. Replacing ∇T by $\mathbf{F}$ in the analysis made in Section 4.1, we conclude that these methods may be seen as instances of the partial linearization methods for variational inequalities. This result may be applied to traffic equilibrium where the index i may be associated with different user classes, commodities or arcs.

Jacobi methods for asymmetric traffic equilibrium problems are discussed in [4, 35, 89, 41, 60]. Sheffi [89] and Mahmassani and Mouskos [60] suggest using truncated Frank-Wolfe steps on the convex subproblems, i.e., these are instances of the truncated partial linearization algorithm for variational inequalities. Sheffi suggests using only one iteration of the Frank-Wolfe algorithm in his *streamlined* approach. In their tests, Mahmassani and Mouskos conclude that not more than four steps of the Frank-Wolfe algorithm should be considered; this result complies with that of Dembo and Tulowitzki [28] for the separable case (see Section 4.3).

Linearized Jacobi are considered e.g. in [79, 12, 53], in combination with simplicial decomposition/column generation. As in the separable case, this methodology has also been applied in a Gauss-Seidel framework. Examples are found in [25, 70].

6.2 Linearization methods

The class of linearization methods defines special cases of partial linearization. As described by e.g. Pang and Chan [78], the class of linearization methods is based on an affine approximation of the mapping $\mathbf{F}$ around the iteration point $\mathbf{x}^{(l)}$:

$$\mathbf{F}^{(l)}(\mathbf{x}) = \mathbf{F}\left(\mathbf{x}^{(l)}\right) + \mathbf{A}\left(\mathbf{x}^{(l)}\right)\left(\mathbf{x} - \mathbf{x}^{(l)}\right),$$

and a linearization method is defined by the choice of map $\mathbf{A}$, and the variational inequality

$$\text{find } \mathbf{x}^{(l+1)} \in \mathbf{X} \text{ such that } \mathbf{F}^{(l)}\left(\mathbf{x}^{(l+1)}\right)^{\mathrm{T}}\left(\mathbf{x} - \mathbf{x}^{(l+1)}\right) \geq 0, \forall \mathbf{x} \in \mathbf{X}.$$

Let $\varphi(\mathbf{x}, \mathbf{y}) = \mathbf{A}(\mathbf{y})(\mathbf{x} - \mathbf{y})$. Then [$\mathbf{PL\text{-}SUB\text{-}VIP}^{(l)}$] is equivalent to the above problem, by letting $\mathbf{x}^{(l+1)} = \overline{\mathbf{x}}^{(l)}$. Hence, the generic subproblem of linearization methods is a special case of the subproblem of the partial linearization methods. Based on the exposition of Pang and Chan, we list some well known linearization methods below.

Let the Jacobian of $\mathbf{F}$ at $\mathbf{x} \in \mathbf{X}$ be written as $\mathbf{F}'(\mathbf{x}) = \mathbf{L}(\mathbf{x}) + \mathbf{D}(\mathbf{x}) + \mathbf{U}(\mathbf{x})$, where $\mathbf{D}(\mathbf{x})$ is the diagonal part of $\mathbf{F}'(\mathbf{x})$, and $\mathbf{L}(\mathbf{x})$ and $\mathbf{U}(\mathbf{x})$ the lower and upper triangular parts, respectively. Also, let $\gamma > 0$, $0 < \omega < 2$, and $\mathbf{G}$ a positive definite symmetric matrix. The following well known algorithms are instances of the linearization method:

$$\begin{array}{rcll}
\mathbf{A}(\mathbf{x}) & = & \mathbf{F}'(\mathbf{x}) & \text{Newton's method} \\
 & \approx & \mathbf{F}'(\mathbf{x}) & \text{Quasi-Newton methods} \\
 & = & \left\{ \begin{array}{l} \mathbf{L}(\mathbf{x}) + \mathbf{D}(\mathbf{x})/\omega \text{ or} \\ \mathbf{U}(\mathbf{x}) + \mathbf{D}(\mathbf{x})/\omega \end{array} \right. & \text{SOR; Linearized Gauss-Seidel } (\omega = 1) \\
 & = & \mathbf{D}(\mathbf{x}) & \text{Linearized Jacobi} \\
 & = & \frac{1}{\gamma}\mathbf{G} & \text{Projection method}
\end{array}$$

The projection algorithm was one of the first methods applied to general traffic equilibria, and may be viewed as the successive approximation algorithm based on the fixed point Theorem 5.3. References include [23, 33, 41].

In the case where the map $\mathbf{A}$ is symmetric, the affine variational inequality subproblem reduces to a quadratic program. Various choices of symmetric approximations of the Jacobian matrix $\mathbf{F}'$ define the so called quasi-Newton methods. One example of approximations is the *symmetrized Newton* method, where $\mathbf{A}(\mathbf{x}) = \frac{1}{2}\left(\mathbf{F}'(\mathbf{x}) + \mathbf{F}'(\mathbf{x})^{\mathrm{T}}\right)$, i.e., the symmetric part of the Jacobian matrix. In the traffic equilibrium context, where the subproblem is a quadratic multicommodity flow problem, such a method is applied in e.g. [23, 33].

Newton algorithms in combination with simplicial decomposition/column generation are presented in e.g. [1, 3, 53, 46]. A gap-decreasing algorithm (for $\Phi \equiv 0$, see Section 5.3) based on Newton's method is given in [66].

7 Conclusions and further research

In this paper we have shown that a number of algorithms frequently applied to symmetric and asymmetric traffic equilibrium problems may be described in a unified manner as instances of the partial linearization algorithm. Examples include the standard Frank-Wolfe algorithm, Newton-type methods, projection methods and Jacobi/Gauss-Seidel methods. For the separable model we also validated truncated solutions of the subproblems and asynchronous implementations of the partial linearization algorithm. An equivalence result was also obtained, relating the frameworks of Dafermos and Tseng to the class of partial linearization methods. Through the framework of partial linearization the theoretical (as well as the practical) convergence rate obtained for different properties of $T(\mathbf{x})$ and $\Phi(\mathbf{x}, \mathbf{y})$, and the effects of different truncation strategies, would be of interest to investigate.

Historically, methods used for asymmetric traffic equilibria have almost all been based on successive approximation. The use of merit functions in descent algorithms opens up possibilities for speeding up the practical convergence rate of these algorithms as well as weakening the convergence conditions. The line searches can also ensure global convergence where else only local convergence can be ensured. (An example is Newton's method, which is not globally convergent in its successive approximation version, but globally convergent under an additional line search with respect to a gap function [66].) The possibility of solving subproblems in gap-decreasing algorithms only approximately [52, 98] and still ensuring global convergence makes the algorithms even more interesting for practical use. It would also be interesting to study the convergence properties of methods based on line searches on the variational inequality as is done in the symmetric case; this has been done to a very limited extent [41] but showing promising results.

References

[1] Aashtiani, H.Z. (1979), *The multi-modal traffic assignment problem.* Ph.D. dissertation, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.

[2] Aashtiani, H.Z. and Magnanti, T.L. (1981), *Equilibria on a congested transportation network.* SIAM Journal on Algebraic and Discrete Methods **2**, pp. 213-226.

[3] Aashtiani, H.Z. and Magnanti, T.L. (1982), *A linearization and decomposition algorithm for computing urban traffic equilibria.* In: Proceedings of the IEEE Large Scale Systems Symposium, pp. 8-19.

[4] Abdulaal, M. and LeBlanc, L.J. (1979), *Methods for combining modal split and equilibrium assignment models.* Transportation Science **13**, pp. 292-314.

[5] Auchmuty, G. (1989), *Variational principles for variational inequalities.* Numerical Functional Analysis and Optimization **10**, pp. 863-874.

[6] Bazaraa, M.S. and Shetty, C.M. (1979). Nonlinear Programming. Theory and Algorithms. John Wiley & Sons, New York, NY.

[7] Beckmann, M. McGuire, C. and Winsten, C.B. (1956). Studies in the Economics of Transportation. Yale University Press, New Haven, CT.

[8] Bertsekas, D.P. (1976), *On the Goldstein-Levitin-Polyak gradient projection method.* IEEE Transactions on Automatic Control **AC-21**, pp. 174-184.

[9] Bertsekas, D.P. (1979), *Algorithms for nonlinear multicommodity network flow problems.* In: Proceedings of the International Symposium on Systems Optimization and Analysis, Bensoussan and Lions (eds.), Springer-Verlag, N.Y., pp. 210-224.

[10] Bertsekas, D.P. (1982), *Optimal routing and flow control methods for communication networks.* In: Analysis and Optimization of Systems, Proceedings of the International Symposium on Systems Optimization and Analysis, Versailles, A. Bensoussan and J.L. Lions (eds.), Springer-Verlag, New York, NY, pp. 615-643.

[11] Bertsekas, D.P. and El Baz, D. (1987), *Distributed asynchronous relaxation methods for convex network flow problems.* SIAM Journal on Control and Optimization **25**, 74-85.

[12] Bertsekas, D.P. and Gafni, E.M. (1982), *Projection methods for variational inequalities with application to the traffic assignment problem.* Mathematical Programming Study **17**, pp. 139-159.

[13] Bertsekas, D.P. and Gafni, E.M. (1983), *Projected Newton methods and optimization of multicommodity flows.* IEEE Transactions on Automatic Control **AC-28**, pp. 1090-1096.

[14] Bertsekas, D.P., Hosein, P.A. and Tseng, P. (1987), *Relaxation methods for network flow problems with convex arc costs.* SIAM Journal on Control and Optimization **25**, pp. 1219-1243.

[15] Bertsekas, D.P. and Tsitsiklis, J.N. (1989). Parallel and Distributed Computation. Numerical Methods. Prentice-Hall, London.

[16] Bruynooghe, M., Gibert, A. and Sakarovitch, M. (1969), *Une méthode d'affectation du trafic.* In: Fourth International Symposium on the Theory of Traffic Flow, Karlsruhe, 1968, W. Lentzback and P. Baron (eds.), Beiträge zur Theorie des Verkehrsflusses Strassenbau und Strassenverkehrstechnik Heft 86, Herausgeben von Bundesminister fur Verkehr, Abteilung Strassenbau, Bonn, pp. 198-204.

[17] Censor, Y. (1981), *Row-action methods for huge and sparse systems and their applications.* SIAM Review **23**, pp. 444-466.

[18] Censor, Y. (1988), *Parallel application of block-iterative methods in medical imaging and radiation therapy.* Mathematical Programming **42**, pp. 307-325.

[19] Chen, R.J. and Meyer, M.M. (1988), *Parallel optimization for traffic assignment.* Mathematical Programming **42**, pp. 327-345.

[20] Dafermos, S.-S.C. (1968), *Traffic assignment and resource allocation in transportation networks.* Ph.D. dissertation, The Johns Hopkins University, Baltimore, MA.

[21] Dafermos, S.C. (1971), *An extended traffic assignment model with applications to two-way traffic.* Transportation Science **5**, pp. 366-389.

[22] Dafermos, S.C. (1972), *The traffic assignment problem for multiclass-user transportation networks.* Transportation Science **6**, pp. 73-87.

[23] Dafermos, S. (1980), *Traffic equilibrium and variational inequalities.* Transportation Science **14**, pp. 42-54.

[24] Dafermos, S. (1983), *An iterative scheme for variational inequalities.* Mathematical Programming **26**, pp. 40-47.

[25] Dafermos, S. (1983), *Convergence of a network decomposition algorithm for the traffic equilibrium model.* In: Proceedings of the 8th International Symposium on Transportation and Traffic Theory, Toronto, Canada, 1981, V.F. Hurdle *et al.* (eds.), University of Toronto Press, Toronto, pp. 143-156.

[26] Dafermos, S. (1990), *Exchange price equilibria and variational inequalities.* Mathematical Programming **46**, pp. 391-402.

[27] Dafermos, S.C. and Sparrow, F.T. (1969), *The traffic assignment problem for a general network.* Journal of Research of the National Bureau of Standards **73B**, pp. 91-118.

[28] Dembo, R.S. and Tulowitzki, U. (1988), *Computing equilibria on large multicommodity networks: an application of truncated quadratic programming algorithms.* Networks **18**, pp. 273-284.

[29] Eaves, B.C. (1971), *On the basic theorem of complementarity.* Mathematical Programming **1**, pp. 68-75.

[30] Evans, S.P. (1976), *Derivation and analysis of some models for combining trip distribution and assignment.* Transportation Research **10**, pp. 37-57.

[31] Feijoo, B. and Meyer, R.R. (1988), *Piecewise-linear approximation methods for nonseparable convex optimization.* Management Science **34**, pp. 411-419.

[32] Ferris, M.C. (1991), *Finite termination of the proximal point algorithm.* Mathematical Programming **50**, pp. 359-366.

[33] Fisk, C. and Nguyen, S. (1982), *Solution algorithms for network equilibrium models with asymmetric user costs.* Transportation Science **16**, pp. 361-381.

[34] Florian, M. (1986), *Nonlinear cost network models in transportation analysis.* Mathematical Programming Study **26**, pp. 167-196.

[35] Florian, M., Chapleau, R., Nguyen, S., Achim, C., James-Lefevbre, L., Galarneau, S., Lefevbre, J. and Fisk, C. (1979), *Validation and application of EMME: an equilibrium based two-mode urban transportation planning method.* Transportation Research Record **728**, pp. 14-23.

[36] Florian, M. and Nguyen, S. (1974), *A method for computing network equilibrium with elastic demands.* Transportation Science **8**, pp. 321-332.

[37] Frank, M. and Wolfe, P. (1956), *An algorithm for quadratic programming.* Naval Research Logistics Quarterly **3**, pp. 95-110.

[38] Fukushima, M. (1992), *Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems.* Mathematical programming **53**, pp. 99-110.

[39] Gibert, A. (1968), *A method for the traffic assignment problem.* Report LBS-TNT-95, Transportation Network Theory Unit, London Business School, London.

[40] Goldstein, A.A. (1964), *Convex programming in Hilbert space.* Bulletin of the American Mathematical Society **70**, pp. 709-710.

[41] Harker, P.T. (1988), *Accelerating the convergence of the diagonalization and projection algorithms for finite-dimensional variational inequalities.* Mathematical Programming **41**, pp. 29-59.

[42] Harker, P.T. and Pang, J.-S. (1990), *Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications.* Mathematical Programming **48**, pp. 161-220.

[43] Hearn, D.W. (1982), *The gap function of a convex program.* Operations Research Letters **1**, pp. 67-71.

[44] Hearn, D.W., Lawphongpanich, S. and Nguyen, S. (1984), *Convex programming formulations of the asymmetric traffic assignment problem.* Transportation Research **18B**, pp. 357-365.

[45] Hearn, D.W., Lawphongpanich, S. and Ventura, J.A. (1987), *Restricted simplicial decomposition: computation and extensions.* Mathematical Programming Study **31**, pp. 99-118.

[46] Hearn, D.W., Lawphongpanich, S. and Ventura, J.A. (1987), *Optimization algorithms for congested network models.* In: Flow Control of Congested Networks, Proceedings of the NATO Advanced Research Workshop on Flow Control of Congested Networks, Capri, Italy, 1986, A.R. Odoni *et al.* (eds.), NATO ASI Series **F 38**, Springer-Verlag, New York, NY.

[47] von Hohenbalken, B. (1977), *Simplicial decomposition in nonlinear programming algorithms.* Mathematical Programming **13**, pp. 49-68.

[48] Kinderlehrer, D. and Stampacchia (1980). An Introduction to Variational Inequalities and Their Applications. Academic Press, New York, NY.

[49] Larsson, T. and Migdalas, A. (1990), *An algorithm for nonlinear programs over Cartesian product sets.* Optimization **21**, pp. 535-542.

[50] Larsson, T., Migdalas, A. and Patriksson, M. (1989), *A partial linearization method for the traffic assignment problem.* Report LiTH-MAT-R-89-06, Linköping Institute of Technology, Department of Mathematics, Linköping, Sweden.

[51] Larsson, T. and Patriksson, M. (1992), *Simplicial decomposition with disaggregated representation for the traffic assignment problem.* Transportation Science **26**, pp. 4-17.

[52] Larsson, T. and Patriksson, M. (1990), *A generalized gap function for variational inequalities.* Report LiTH-MAT-R-90-18, Linköping Institute of Technology, Department of Mathematics, Linköping, Sweden. Revised Spring 1992.

[53] Lawphongpanich, S. and Hearn, D.W. (1984), *Simplicial decomposition of the asymmetric traffic assignment problem.* Transportation Research **18B**, pp. 123-133.

[54] LeBlanc, L.J., Morlok, E. and Pierskalla, W.P. (1975), *An efficient approach to solving the road network equilibrium traffic assignment problem.* Transportation Research, **9**, pp. 308-318.

[55] Leventhal, T., Nemhauser, G. and Trotter, jr., L. (1973), *A column generation algorithm for optimal traffic assignment.* Transportation Science **7**, pp. 168-176.

[56] Levitin, E.S. and Polyak, B.T. (1966), *Constrained minimization methods.* USSR Computational Mathematics and Mathematical Physics **6**, pp. 1-50.

[57] Lin, Y.Y. and Pang, J.-S. (1987), *Iterative methods for large scale convex quadratic programs: a survey.* SIAM Journal on Control and Optimization **25**, pp. 383-411.

[58] Luo, Z.-Q. and Tseng, P. (1992), *On the linear convergence of descent methods for convex essentially smooth minimization.* SIAM Journal on Control and Optimization **30**, pp. 408-425.

[59] Lupi, M. (1986), *Convergence of the Frank-Wolfe algorithm in transportation networks.* Civil Engineering Systems **3**, pp. 7-15.

[60] Mahmassani, H.S. and Mouskos, K.C. (1989), *Vectorization of transportation network equilibrium assignment codes.* Conference Paper, presented at the ORSA Conference, Williamsburg, Virginia. Also as Report, Department of Civil Engineering, University of Texas, Austin.

[61] Marcotte, P. (1985), *A new algorithm for solving variational inequalities with application to the traffic assignment problem.* Mathematical Programming **33**, pp. 339-351.

[62] Marcotte, P. (1986), *Gap-decreasing algorithms for monotone variational inequalities.* Conference Paper, presented at the ORSA/TIMS Joint National Meeting, Miami Beach, FL.

[63] Marcotte, P. and Dussault, J.-P. (1985), *A modified Newton method for solving variational inequalities.* In: Proceedings of the 24th IEEE Conference on Decision and Control, Fort Lauderdale, FL, pp. 1433-1436.

[64] Marcotte, P. and Dussault, J.-P. (1987), *A note on a globally convergent Newton method for solving monotone variational inequalities.* Operations Research Letters **6**, pp. 35-42.

[65] Marcotte, P. and Dussault, J.-P. (1989), *A sequential linear programming algorithm for solving monotone variational inequalities.* SIAM Journal on Control and Optimization **27**, pp. 1260-1278.

[66] Marcotte, P. and Guélat, J. (1988), *Adaptation of a modified Newton method for solving the asymmetric traffic equilibrium problem.* Transportation Science **22**, pp. 112-124.

[67] Martinet, B. (1970), *Regularisation d'inequations variationnelles par approximations successives.* Revue francaise d'informatique et de recherche opérationnelle **4**, pp. 154-158.

[68] Migdalas, A. (1990), *Cyclic linearization vs. Frank-Wolfe decomposition for nonlinear problems over Cartesian product sets.* Unpublished note, Linköping Institute of Technology, Department of Mathematics, Linköping, Sweden.

[69] Migdalas, A. (1990), *A regularization of the Frank-Wolfe algorithm.* Report LiTH-MAT-R-90-10, Linköping Institute of Technology, Department of Mathematics, Linköping, Sweden.

[70] Nagurney, A. (1988), *An equilibration scheme for the traffic assignment problem with elastic demands.* Transportation Research **22B**, pp. 73-79.

[71] Netter, M. and Sender, J.G. (1970), *Equilibre offre-demande et tarification sur un réseau de transport.* Institut de Recherche des Transport, Arcueil, France.

[72] Nguyen, S. (1973), *A mathematical programming approach to equilibrium methods of traffic assignment with fixed demands.* Publication no. 138, Département d'informatique et de recherche opérationelle, Université de Montréal.

[73] Nguyen, S. (1974), *An algorithm for the traffic assignment problem.* Transportation Science **8**, pp. 203-216.

[74] Nguyen, S. (1976), *A unified approach to equilibrium methods for traffic assignment.* In: Traffic Equilibrium Methods, Proceedings of the International Symposium in Montreal, M.A. Florian (ed.), Lecture Notes in Economics and Mathematical Systems **118**, Springer-Verlag, New York, NY, pp. 148-182.

[75] Ortega, J.M. and Rheinboldt, W.C. (1970). Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York, NY.

[76] Pang, J.-S. (1985), *Asymmetric variational inequality problems over Cartesian product sets: applications and iterative methods.* Mathematical Programming **31**, pp. 206-219.

[77] Pang, J.-S. (1990), *Newton's method for B-differentiable equations.* Mathematics of Operations Research **15**, pp. 311-341.

[78] Pang, J.-S. and Chan, D. (1982), *Iterative methods for variational and complementarity problems.* Mathematical Programming **24**, pp. 284-313.

[79] Pang, J.-S. and Yu, C.-S. (1984), *Linearized simplicial decomposition methods for computing traffic equilibria on networks.* Networks **14**, pp. 427-438.

[80] Patriksson, M. (1990), *The traffic assignment problem - theory and algorithms.* Report LiTH-MAT-R-90-29, Linköping Institute of Technology, Department of Mathematics, Linköping, Sweden.

[81] Patriksson, M. (1991), *Partial linearization methods in nonlinear programming.* Report LiTH-MAT-R-91-11, Linköping Institute of Technology, Department of Mathematics, Linköping, Sweden. Forthcoming in Journal of Optimization Theory and Applications.

[82] Petersen, E.R. (1975), *A primal-dual traffic assignment algorithm.* Management Science **22**, pp. 87-95.

[83] Polak, E. (1971). Computational Methods in Optimization. A Unified Approach. Academic Press, New York, NY.

[84] Polyak, B.T. (1987). Introduction to Optimization. Optimization Software, New York, NY.

[85] Pshenichny, B.N. and Danilin, Yu. M. (1978). Numerical Methods in Extremal Problems. MIR Publishers, Moscow.

[86] Rockafellar, R.T. (1976), *Augmented Lagrangians and applications of the proximal point algorithm in convex programming.* Mathematics of Operations Research **1**, pp. 97-116.

[87] Schneider, M.H. and Zenios, S.A. (1990), *A comparative study of algorithms for matrix balancing.* Operations Research **38**, pp. 439-455.

[88] Serra, P. and Weintraub, A. (1981), *Convergence of decomposition algorithms for the traffic assignment problem.* In: Studies on Graphs and Discrete Programming, P. Hansen (ed.), North-Holland Publishing Company, pp. 313-318.

[89] Sheffi, Y. (1985). Urban Transportation Networks. Equilibrium Analysis with Mathematical Methods. Prentice-Hall, Englewood Cliffs, NJ.

[90] Smith, M.J. (1979), *The existence, uniqueness and stability of traffic equilibria.* Transportation Research **13B**, pp. 295-304.

[91] Taji, K., Fukushima, M. and Ibaraki, T. (1991), *A globally convergent Newton method for solving strongly monotone variational inequalities.* Forthcoming in Mathematical Programming.

[92] Tseng, P. (1990), *Dual ascent methods for strictly convex costs and linear constraints.* SIAM Journal on Control and Optimization **28**, pp. 214-242.

[93] Tseng, P. (1991), *Decomposition algorithm for convex differentiable minimization.* Journal of Optimization Theory and Applications **70**, pp. 109-135.

[94] Tseng, P. (1991), *On the rate of convergence of a partially asynchronous gradient projection algorithm.* SIAM Journal on Optimization **1**, pp. 603-619.

[95] Tseng, P. and Bertsekas, D.P. (1991), *Relaxation methods for problems with strictly convex costs and linear constraints.* Mathematics of Operations Research **16**, pp. 462-481.

[96] Tseng, P., Bertsekas, D.P. and Tsitsiklis, J.N. (1990), *Partially asynchronous, parallel algorithms for network flow and other problems.* SIAM Journal on Control and Optimization **28**, pp. 678-710.

[97] Wardrop, J.G. (1952), *Some theoretical aspects of road traffic research.* In: Proceedings of the Institute of Civil Engineers, part II, pp. 325-378.

[98] Wu, J.H., Florian, M. and Marcotte, P. (1990), *A general descent framework for the monotone variational inequality problem.* Publication no. 723, Centre de recherche sur les transports, Université de Montréal.

[99] Zangwill, W.I. (1969). Nonlinear Programming: A Unified Approach. Prentice-Hall, Englewood Cliffs, NJ.

[100] Zenios, S.A. (1990), *Matrix balancing on a massively parallel connection machine.* ORSA Journal on Computing 2, pp. 112-125.

[101] Zhao, L. and Dafermos, S. (1991), *General economic equilibrium and variational inequalities.* Operations Research Letters **10**, pp. 369-376.

[102] Zuhovickiĭ, S.I., Poljak, R.A. and Primak, M.E. (1969), *Two methods of search for equilibrium points of n-person concave games.* Soviet Mathematics Doklady **10**, pp. 279-282.