Torbjörn Larsson · Michael Patriksson · Ann-Brith Strömberg

Ergodic, primal convergence in dual subgradient schemes for convex programming

Abstract. Lagrangean dualization and subgradient optimization techniques are frequently used within the field of computational optimization for finding approximate solutions to large, structured optimization problems. The dual subgradient scheme does not automatically produce primal feasible solutions; there is an abundance of techniques for computing such solutions (via penalty functions, tangential approximation schemes, or the solution of auxiliary primal programs), all of which require a fair amount of computational effort.

We consider a subgradient optimization scheme applied to a Lagrangean dual formulation of a convex program, and construct, at minor cost, an ergodic sequence of subproblem solutions which converges to the primal solution set. Numerical experiments performed on a traffic equilibrium assignment problem under road pricing show that the computation of the ergodic sequence results in a considerable improvement in the quality of the primal solutions obtained, compared to those generated in the basic subgradient scheme.

Key words. convex programming – Lagrangean duality – Lagrangean relaxation – subgradient optimization – ergodic convergence – primal convergence – traffic equilibrium assignment – road pricing

1. Introduction

When solving large, structured optimization problems through the utilization of Lagrangean dual formulations, subgradient optimization methods are popular since they remarkably often are able to quickly identify near-optimal dual solutions. However, the subgradient schemes do not directly provide solutions to the primal problem. We present a simple means to construct such a solution by forming an ergodic sequence of Lagrangean subproblem solutions; our analysis generalizes those for *linear* programs by Shor [60] and Larsson and Liu [39] to the case of general *convex* programs.

Associated with a convex minimization program is a Lagrangean dual program. The corresponding Lagrange function is convex (concave) with respect to its primal (dual) variables, and the solutions to this primal–dual pair of programs are the saddle points to the Lagrange function. In a Lagrangean dual approach for solving a convex program, the Lagrange function is maximized with respect to its dual variables, and the corresponding primal solutions are derived from subproblems, which are usually considerably more easily solved than the original program. For some early and important developments of Lagrangean duality theory in nonlinear programming, see Uzawa [63], Everett [20], and Falk [21].

T. Larsson: Division of Optimization, Department of Mathematics, Linköping University, S-581 83 Linköping, Sweden

M. Patriksson, A.-B. Strömberg: Department of Mathematics, Chalmers University of Technology, S-412 96 Göteborg, Sweden, Fax: +46 31 16 19 73, e-mail: `anstr@math.chalmers.se`

Over the last few decades, Lagrangean dualization techniques (e.g., Geoffrion [25]) have become widely used within the field of computational optimization. Some examples are given in the following. An early application of Lagrangean dual techniques was the *Dantzig–Wolfe decomposition method* [13] for linear programming problems (although it was not described in these terms), which employs an inner approximation of the convex set defined by their constraints; this approximation can be interpreted as the dual of a tangential approximation of the objective function of the Lagrange dual with respect to these constraints (e.g., Lasdon [46, Section 8.6]). The *Lagrangean relaxation* technique has proven to be a powerful tool for designing efficient heuristic solution methods for many classes of structured large scale optimization problems, in particular within the field of discrete optimization (e.g., Geoffrion [26], Fisher [23], and Beasley [4]). In *augmented Lagrangean methods* (e.g., Rockafellar [56], and Bertsekas [7]) a quadratic penalty term is added to the objective function before the program is dualized; this type of combination of Lagrangean relaxation and penalty methods has been successful in many applications (e.g., Ruszczyński [58]). The application of Lagrangean dual techniques to network flow problems with (strictly) convex and separable cost functions some times gives rise to highly parallelizable algorithms (see Bertsekas and Tsitsiklis [8, Chapter 5]).

We next review a number of applications that demonstrate the strength of dual solution approaches to large scale structured optimization problems. In an algorithm for optimum structural design, Svanberg [62] solves an approximate design problem — with strictly convex objective and linear inequality constraints — using Lagrangean duality; the dual objective is smooth and it is maximized by steepest ascent (slightly modified to take care of the nonnegativity restrictions). Lamond and Stewart [38] show that many balancing methods used in transportation planning and in other fields are special cases of a method developed by Bregman [9] applied to a Lagrange dual of a gravity model. The minimization of a strictly convex, separable function subject to convexity constraints can be efficiently made by utilizing Lagrangean duality (e.g., Cottle et al. [11]); special cases of this program (e.g., the Euclidean projection of a point onto a simplex) arise as subproblems in many iterative schemes for structured large scale programming. Balakrishnan et al. [2] develop a dual ascent procedure to solve large scale uncapacitated network design problems; they report successful results from applications to models arising in freight transportation. Fisher [24] solves vehicle routing problems to optimality by a branch–and–bound algorithm, where lower bounds are generated using Lagrangean relaxation, the resulting subproblem solutions essentially being k-trees; the algorithm has produced proven optimal solutions for several difficult problems.

In the last few years, Lagrangean dual approaches have received a renewed interest for the solution of large scale *linear* programs. Goffin et al. [27,28] propose a new treatment of the master program in the Dantzig–Wolfe method; the resulting method performs well in applications to large scale structured linear programs. Hauer and Hoganson [29] solve large linear programs arising in forest management scheduling, using Lagrangean relaxation and subgradient optimization; here the Lagrangean dual approach is especially appropriate since the constraints are *soft*, that is, they need not be fulfilled exactly. Jones et al. [34] apply the Dantzig–Wolfe decomposition principle to different formulations of linear multicommodity network flow problems.

1.1. Primal convergence in Lagrangean dual schemes

What then are the main advantages of Lagrangean dual approaches? Many computationally demanding optimization problems can be interpreted as fairly easily solvable problems which are complicated by side constraints. An example of a problem which can be characterized as such is the travelling salesman polytope, which can be expressed by spanning tree constraints and node degree (side) constraints. In a Lagrangean relaxation formulation of a program, these side constraints are moved to the objective, where they are included, weighted by multipliers. The relaxed problem that is thereby created takes the side constraints into account implicitly (and the new program is more easily solved than the original one). The solution to the relaxed problem will, in general, not satisfy the relaxed constraints, but the violation can, in a certain sense, be minimized through the solution of a Lagrangean dual program, which is always convex. Moreover, the feasible solutions to the dual program yield bounds on the optimal objective value of the original program.

However, the Lagrangean duality concept has a few drawbacks, some of which we aim to remedy through this work. If the original program is nonconvex — if it is, for example, a discrete optimization problem — then there is usually a gap between the optimal primal and dual objective values; this property makes it difficult to construct proper termination criteria for algorithms based on Lagrangean relaxation formulations. Moreover, not even in the (unlikely) case that an exact dual solution is at hand is a primal optimal solution easily available. One reason for this inconvenience is that the Lagrangean function may not have a saddle point in the nonconvex case. Another reason is that the dual objective function (also in the convex case) is typically nonsmooth, especially at an optimal dual solution; then an optimal primal solution is (usually) a nontrivial convex combination of the extreme subproblem solutions. Within linear programming this property has been referred to as the *noncoordinability* phenomenon [16]. In the case that the original objective function is strictly convex, the Lagrangean dual objective function is differentiable, whence this phenomenon does not appear. A lot of effort has been put into inducing primal convergence in Lagrangean relaxation schemes for convex programming. We go on to an overview of the main types of approaches used for this purpose.

Approach 1 (Removing the nonsmoothness). In the methods of Jennergren [32] and Mangasarian [48] linear price functions and a quadratic perturbation of the linear objective, respectively, are employed. Closely related to these approaches are the augmented Lagrangean dual solution methods for convex programming (e.g., Rockafellar [56] and Bertsekas [7]); these methods can be interpreted as combinations of Lagrangean dualization and penalty approaches, where the nonlinear penalty term induces coordinability. Dem'yanov and Malozemov [14, p. 230] solve convex–concave saddle point problems; in order to receive *strict* convexity–concavity they add (subtract) a strictly convex quadratic term to (from) the convex (concave) component of the function. Feinberg [22] introduces coordinability in a Dantzig–Wolfe type scheme, by using strictly convex price functions.

In these methods, exact primal feasibility and complementarity are typically reached in the limit only. Although their memory requirements are fairly low, each of their

subproblems is computationally more demanding than that of the ordinary Lagrangean dual approach.

□

Approach 2 (Using ascent methods). For large block-angular linear programs, Ruszczyński [58] proposes an augmented Lagrangean method in which the multiplier steps form an ascent procedure with respect to the ordinary Lagrange dual problem; its convergence is finite. Ben-Tal and Bendsøe [5] reformulate a problem of truss topology design to an unconstrained minimization problem with a convex and piecewise quadratic objective function; it is solved by an ε-steepest descent algorithm. In a proximal bundle method applied to a Lagrange dual of a convex program, aggregated subproblem solutions asymptotically solve the primal program (Kiwiel [35]); the aggregation weights are computed through the solution of a quadratic program.

In these methods, feasibility and complementarity are typically reached in the limit only (for linear programs, sometimes finitely, however). For large scale problems, these quadratic programs are normally very large.

□

Approach 3 (Solving a master problem). In the Dantzig–Wolfe decomposition principle [13] a linear coordinating master program combines extreme Lagrangean subproblem solutions into an approximate solution to the original program (see also [46, Section 8.6]). In the stochastic decomposition method for two-stage linear programs by Higle and Sen [30], the objective function of the master program is a piece-wise linear approximation of the original (implicitly defined) objective function. For convex programs, Bazaraa et al. [3, p. 230] acquire feasibility and finite ε-optimality through the minimization of an inner approximation of the objective function over the convex hull of the subproblem solutions.

In these methods, feasibility is maintained through the iterative process, while complementarity is typically reached in the limit only. The master program is usually a large linear (or, convex) program.

□

Approach 4 (Utilizing ergodic sequences). The construction of sequences of (weighted) averages (that is, ergodic sequences) of solutions is a widely used technique for inducing convergence properties that an original sequence lacks. Examples of this are the method of successive averages by Powell and Sheffi [53], and the mean value cross decomposition method by Holmberg [31]. In the stochastic decomposition method by Higle and Sen [30], averages of approximate supporting hyperplanes are used to obtain statistically valid lower bounds, and Petersson and Patriksson [50] employ averaging techniques to solve saddle point problems arising from applications in the topology optimization of mechanical structures. Shor [60, pp. 116–118] uses ergodic sequences of Lagrangean subproblem solutions to generate optimal primal solutions in linear programming; his ideas are further investigated and developed, and computationally tested by Larsson et al. [40] for the traffic equilibrium assignment problem, and by Larsson and Liu [39] for structured linear programs. Sherali and Choi [59] extend the results of Shor and of Larsson and Liu to allow for more general choices of convexity weights and step lengths in the subgradient scheme.

In these methods, feasibility and complementarity are reached in the limit only. In some cases the convergence is slow, but no additional optimization problem has to be solved and the memory requirements are low.

$\square$

1.2. Motivation and outline

In this work, we continue and further develop the ideas of Shor and of Larsson and Liu, as described in Approach 4. Their results are generalized to convex programs with possibly nonsmooth objective and/or constraint functions.

While in the methods of Approaches 1–3 above, the sub- or master problems are computationally demanding, in our ergodic approach no auxiliary optimization problem needs to be solved to induce primal convergence. Further, our method requires a relatively small amount of additional memory, which is in contrast to the methods of Approaches 2 and 3. In the methods of Approach 3, primal feasibility is maintained throughout the iterative procedure. Our method, however, guarantees neither primal feasibility nor complementarity in finite time when applied to general convex programs. Therefore we also propose the use of heuristic projections of averaged solutions onto the primal feasible set. (In the application to traffic equilibrium assignment under road pricing, which is reported in Section 4, the special problem structure enables us, however, to obtain primal feasibility throughout the iterative procedure.) Our method is also motivated by applications with *soft* constraints, such as the large forest management scheduling problem solved by Hauer and Hoganson [29], and applications involving capacity expansion decisions, such as production and work force planning problems (e.g., Johnson and Montgomery [33, Example 4-14]). We believe that our analysis fills a gap between the analyses of linear programs and of strictly convex programs; in the latter case it is well known that primal convergence holds without the generation of ergodic sequences.

In Section 2 we briefly review Lagrangean duality theory for convex programming, together with a convergence result for conditional subgradient optimization applied to the dual program, under a general step length rule that extends the divergent series rule. The main contribution of this paper is contained in Section 3, where we present two schemes for generating ergodic sequences of subproblem solutions which induce convergence to the solution set of the primal program. We then show that a sequence of heuristic projections of the averaged solutions onto the original feasible set finitely reaches ε-optimality. In Section 4 we present results from an application to traffic equilibrium assignment under road pricing, and in Section 5 we draw conclusions and discuss briefly some opportunities for further research.

2. Preliminaries

Let the functions $f : \Re^n \mapsto \Re$ and $h_i : \Re^n \mapsto \Re$, $i \in \mathcal{I} = \{1, \dots, m\}$, be convex and (possibly) nonsmooth, the set $X \subset \Re^n$ be convex and compact, and consider the convex program

$$f^* = \min \ f(\mathbf{x}), \tag{1a}$$

$$\text{s.t. } h_i(\mathbf{x}) \leq 0, \quad i \in \mathcal{I}, \tag{1b}$$

$$\mathbf{x} \in X, \tag{1c}$$

with solution set X^*. We assume that the set X is simple and that the feasible set $\{\mathbf{x} \in X \mid h_i(\mathbf{x}) \leq 0, \quad i \in \mathcal{I}\}$ is nonempty.

The following definition is to be used in the sequel. Letting S be a nonempty, closed, and convex set, we denote by

$$\texttt{proj}(\mathbf{x}, S) = \arg \min_{\mathbf{y} \in S} \|\mathbf{y} - \mathbf{x}\|_2 \quad \text{and} \quad \texttt{dist}(\mathbf{x}, S) = \min_{\mathbf{y} \in S} \|\mathbf{y} - \mathbf{x}\|_2 \tag{2}$$

the Euclidean projection of the vector $\mathbf{x}$ onto the set S, and the Euclidean distance from the point $\mathbf{x}$ to $\texttt{proj}(\mathbf{x}, S)$, respectively. The function $\texttt{dist}(\cdot, S)$ is convex and continuous.

We first give a summary of the relevant Lagrangean duality theory for convex programs. Although this is a classical subject, with contributions dating back several decades (see, e.g., Kuhn and Tucker [37] for an early work), to assist the reader we have chosen the textbook by Bazaraa et al. [3] as our basic reference.

2.1. Lagrangean duality

The Lagrange function $\mathcal{L} : \Re^n \times \Re^m \mapsto \Re$ with respect to the relaxation of the constraints (1b) is $\mathcal{L}(\mathbf{x}, \mathbf{u}) = f(\mathbf{x}) + \mathbf{u}^T \mathbf{h}(\mathbf{x})$ for all $(\mathbf{x}, \mathbf{u}) \in \Re^n \times \Re^m$, where $\mathbf{h}(\mathbf{x}) = [h_i(\mathbf{x})]_{i \in \mathcal{I}}$ for all $\mathbf{x} \in \Re^n$ and $\mathbf{u} = [u_i]_{i \in \mathcal{I}}$. For any $\mathbf{u} \in \Re^m_+$, $\mathcal{L}(\cdot, \mathbf{u})$ is convex on $\Re^n$. The dual objective function $\theta : \Re^m \mapsto \Re$ is concave and continuous and is defined by the dual subproblem

$$\theta(\mathbf{u}) = \min_{\mathbf{x} \in X} f(\mathbf{x}) + \mathbf{u}^T \mathbf{h}(\mathbf{x}), \qquad \mathbf{u} \in \Re^m. \tag{3}$$

The nonempty, convex and compact solution set to this subproblem at $\mathbf{u} \in \Re^m$ is

$$X(\mathbf{u}) = \left\{ \mathbf{x} \in X \ \middle| \ f(\mathbf{x}) + \mathbf{u}^T \mathbf{h}(\mathbf{x}) \leq \theta(\mathbf{u}) \right\}. \tag{4}$$

The Lagrange dual to the program (1) then is

$$\begin{aligned} \theta^* = \sup \ & \theta(\mathbf{u}), \\ \text{s.t. } & \mathbf{u} \geq \mathbf{0}, \end{aligned} \tag{5}$$

with the convex solution set U^*. This dual program consists of maximizing a concave function over a convex set; it is thus a convex programming problem. Further, by *weak duality* for the primal-dual pair (1), (5), the inequality $\theta(\mathbf{u}) \leq f(\mathbf{x})$ holds whenever $\mathbf{u} \geq \mathbf{0}, \mathbf{x} \in X$, and $\mathbf{h}(\mathbf{x}) \leq \mathbf{0}$ (Bazaraa et al. [3, Theorem 6.2.1]).

The two following results are crucial in the development of our continued analysis. We utilize the following notion of a *closed map*: A point–to–set map $X : \Re^m \mapsto 2^{\Re^n}$ is closed if $\{\mathbf{u}^t\} \subset \Re^m, \{\mathbf{u}^t\} \to \mathbf{u}, \mathbf{x}^t \in X(\mathbf{u}^t)$ for all t, and $\{\mathbf{x}^t\} \to \mathbf{x}$ imply that $\mathbf{x} \in X(\mathbf{u})$.

Lemma 1 ($X(\cdot)$ **is a closed map**). *Let the sequence $\{\mathbf{u}^t\} \subset \Re^m$, the map $X(\cdot) : \Re^m \mapsto 2^X$ be given by the definition (4), and the sequence $\{\mathbf{x}^t\}$ by the inclusion $\mathbf{x}^t \in X(\mathbf{u}^t)$. If $\{\mathbf{u}^t\} \to \mathbf{u}$, then $\{\mathrm{dist}(\mathbf{x}^t, X(\mathbf{u}))\} \to 0$. If, in addition, $X(\mathbf{u}) = \{\mathbf{x}\}$, then $\{\mathbf{x}^t\} \to \mathbf{x}$.*

Proof. Since the functions f and $\mathbf{h}$ are continuous, the map $X(\cdot)$ is closed. Since the set X is compact, any sequence $\{\mathbf{x}^t\} \subseteq X$ is bounded, whence $\{\mathrm{dist}(\mathbf{x}^t, X(\mathbf{u}))\} \to 0$ follows. The second result is immediate.

$\square$

For each $\mathbf{u} \in \Re^m$, we define the set $\mathcal{I}(\mathbf{u})$ of indices corresponding to strictly positive multipliers, u_i, as

$$\mathcal{I}(\mathbf{u}) = \{i \in \mathcal{I} \mid u_i > 0\}. \tag{6}$$

Lemma 2 (The Lagrange function is affine on the subproblem solution set). *The functions f and h_i, $i \in \mathcal{I}(\mathbf{u})$, are affine on $X(\mathbf{u})$ for every $\mathbf{u} \geq \mathbf{0}$. Further, if the function f (the function h_i, $i \in \mathcal{I}(\mathbf{u})$) is differentiable, then ∇f (∇h_i, $i \in \mathcal{I}(\mathbf{u})$) is constant on $X(\mathbf{u})$.*

Proof. For any $\mathbf{u} \geq \mathbf{0}$, let $\mathbf{x}, \mathbf{y} \in X(\mathbf{u})$ be arbitrary and $\lambda \in [0, 1]$. By the convexity of the functions f and h_i, $i \in \mathcal{I}(\mathbf{u})$, and the definitions (3) and (4),

$$\theta(\mathbf{u}) \leq f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) + \sum_{i \in \mathcal{I}(\mathbf{u})} u_i h_i(\lambda \mathbf{x} + (1-\lambda)\mathbf{y})$$

$$\leq \lambda \left(f(\mathbf{x}) + \sum_{i \in \mathcal{I}(\mathbf{u})} u_i h_i(\mathbf{x}) \right) + (1-\lambda) \left(f(\mathbf{y}) + \sum_{i \in \mathcal{I}(\mathbf{u})} u_i h_i(\mathbf{y}) \right)$$

$$= \lambda \theta(\mathbf{u}) + (1-\lambda) \theta(\mathbf{u}).$$

Hence, the above inequalities must hold with equality. Especially, then, $\lambda \mathbf{x} + (1-\lambda)\mathbf{y} \in X(\mathbf{u})$. Since $\mathbf{x}$ and $\mathbf{y}$ are arbitrary in $X(\mathbf{u})$, this implies the first statement. The second statement is then immediate.

$\square$

A consequence of Lemma 2 is that, for all $\mathbf{u} \geq \mathbf{0}$ and every $i \in \mathcal{I}(\mathbf{u})$, ∂h_i is constant on $\mathrm{rint}\, X(\mathbf{u})$; hence, for example, for every $\overline{\mathbf{x}} \in \mathrm{rint}\, X(\mathbf{u})$, each subgradient $\boldsymbol{\xi}_i \in \partial h_i(\overline{\mathbf{x}})$ defines a hyperplane that supports the function h_i at every $\mathbf{x} \in X(\mathbf{u})$.

The *subdifferential* of the concave function θ at $\mathbf{u} \in \Re^m$ is (Bazaraa et al. [3, Definition 3.2.3])

$$\partial \theta(\mathbf{u}) = \left\{ \boldsymbol{\gamma} \in \Re^m \,\Big|\, \theta(\mathbf{v}) \leq \theta(\mathbf{u}) + \boldsymbol{\gamma}^T(\mathbf{v} - \mathbf{u}), \quad \mathbf{v} \in \Re^m \right\},$$

the elements of which are called *subgradients*. The next proposition follows from [3, Theorem 6.3.7], the convexity of the set X, and Theorem 11 in [45].

Proposition 1 (Subdifferential to the dual objective function). *For each $\mathbf{u} \in \Re^m$, $\partial \theta(\mathbf{u}) = \{\mathbf{h}(\mathbf{x}) \mid \mathbf{x} \in X(\mathbf{u})\}$. Further, θ is differentiable at $\mathbf{u}$ if and only if each h_i is constant on $X(\mathbf{u})$, in which case $\nabla \theta(\mathbf{u}) = \mathbf{h}(\mathbf{x})$ for any $\mathbf{x} \in X(\mathbf{u})$.*

$\square$

This result implies that the function θ is differentiable at $\mathbf{u} \in \Re^m$ if the set $X(\mathbf{u})$ is a singleton (Bazaraa et al. [3, Theorem 6.3.3]).

The *normal cone* to the set $\Re^m_+$ at $\mathbf{u} \in \Re^m_+$ is

$$N_{\Re^m_+}(\mathbf{u}) = \left\{ \boldsymbol{v} \in \Re^m_- \mid v_i u_i = 0, \quad i \in \mathcal{I} \right\}.$$

The optimality conditions for the dual program (5) are given by the following (e.g., [3, Theorem 3.4.3]).

Proposition 2 (Optimality conditions for the dual problem). *The point $\mathbf{u} \in U^*$ if and only if there exists a $\boldsymbol{\gamma} \in \partial\theta(\mathbf{u})$ such that $\boldsymbol{\gamma} \leq \mathbf{0}$ and $\mathbf{u}^T \boldsymbol{\gamma} = 0$, or, equivalently, $\partial\theta(\mathbf{u}) \cap N_{\Re^m_+}(\mathbf{u}) \neq \emptyset$ holds, that is, $-\partial\theta(\mathbf{u}) + N_{\Re^m_+}(\mathbf{u}) \ni \mathbf{0}$.*

□

To obtain primal–dual optimality relations, the primal feasible set must fulfil a *constraint qualification*.

Assumption 1 (Slater constraint qualification). *The set $\{\mathbf{x} \in X \mid \mathbf{h}(\mathbf{x}) < \mathbf{0}\}$ is non-empty.*

□

Under Assumption 1, the convex set U^* is nonempty and compact, and, by *strong duality*, $\theta(\overline{\mathbf{u}}) = f(\overline{\mathbf{x}})$ holds for some pair $(\overline{\mathbf{x}}, \overline{\mathbf{u}})$ such that the relations $\overline{\mathbf{u}} \geq \mathbf{0}$, $\overline{\mathbf{x}} \in X$, and $\mathbf{h}(\overline{\mathbf{x}}) \leq \mathbf{0}$ hold ([3, Theorem 6.2.4]). The next proposition states conditions under which a point $\mathbf{x}$ is optimal in (1), for the (unlikely) case that an optimal dual solution is at hand; it follows from [3, Theorem 6.2.5].

Proposition 3 (Primal–dual optimality conditions). *Let Assumption 1 hold and let $\mathbf{u} \in U^*$. Then, $\mathbf{x} \in X^*$ if and only if $\mathbf{x} \in X(\mathbf{u})$, $\mathbf{h}(\mathbf{x}) \leq \mathbf{0}$, and $\mathbf{u}^T \mathbf{h}(\mathbf{x}) = 0$.*

□

The composite mapping $(\partial\theta) \cap (N_{\Re^m_+})$ is constant on the solution set U^* (e.g., Larsson et al. [44, Proposition 2.1] and Burke and Ferris [10, Lemma 2]). Hence, under Assumption 1, the solution set to the primal program (1) may be expressed as

$$X^* = \left\{ \mathbf{x} \in X(\mathbf{u}) \,\middle|\, \mathbf{h}(\mathbf{x}) \leq \mathbf{0}, \quad \mathbf{u}^T \mathbf{h}(\mathbf{x}) = 0 \right\}, \tag{7}$$

irrespective of the choice of $\mathbf{u} \in U^*$, and the primal–dual optimality conditions may be expressed as

$$(\mathbf{x}, \mathbf{u}) \in X^* \times U^* \qquad \Longleftrightarrow \qquad \mathbf{h}(\mathbf{x}) \in \partial\theta(\mathbf{u}) \cap N_{\Re^m_+}(\mathbf{u}). \tag{8}$$

At a dual solution $\mathbf{u} \in U^*$, the subproblem solution set $X(\mathbf{u})$ is typically *not* a singleton; as a consequence, the dual objective function is nonsmooth on U^*, and a subgradient that can be used to verify the optimality of such a solution, according to Proposition 2, is not directly available.

2.2. Convergence of dual subgradient optimization

We consider solving the Lagrangean dual program (5) by the *conditional subgradient optimization* method [42], which is given by the following. Choose a starting solution $\mathbf{u}^0 \geq \mathbf{0}$ and compute iterates $\mathbf{u}^t$ according to the formula

$$\mathbf{u}^{t+\frac{1}{2}} = \mathbf{u}^t + \alpha_t \left(\mathbf{h}(\mathbf{x}^t) - \mathbf{v}^t \right), \qquad \mathbf{u}^{t+1} = \left[\mathbf{u}^{t+\frac{1}{2}} \right]_+, \qquad t = 0, 1, \dots, \qquad (9)$$

where $\mathbf{x}^t \in X(\mathbf{u}^t)$ solves the dual subproblem (4) at $\mathbf{u}^t$, so that $\mathbf{h}(\mathbf{x}^t) \in \partial\theta(\mathbf{u}^t)$ is a subgradient to θ at $\mathbf{u}^t$, $\mathbf{v}^t \in N_{\Re_+^m}(\mathbf{u}^t)$ is an element of the normal cone to $\Re_+^m$ at $\mathbf{u}^t \in \Re_+^m$, α_t is the step length chosen at iteration t, and $[\cdot]_+$ denotes the Euclidean projection onto the nonnegative orthant $\Re_+^m$.

If $\{\mathbf{v}^t\} = \{\mathbf{0}\}$, then the method (9) reduces to the traditional subgradient optimization method (e.g., Shor [60, Section 2]). Choosing $\mathbf{v}^t = \texttt{proj}(\mathbf{h}(\mathbf{x}^t), N_{\Re_+^m}(\mathbf{u}^t))$ results in

$$h_i(\mathbf{x}^t) - v_i^t = \begin{cases} 0, & \text{if } u_i^t = 0 \text{ and } h_i(\mathbf{x}^t) < 0, \\ h_i(\mathbf{x}^t), & \text{otherwise}, \end{cases} \qquad i = 1, \dots, m, \qquad (10)$$

(see Larsson et al. [42, Lemma 3.2]), defining the special case of the method (9) that is called the *subgradient projection method*. (Note that this name is sometimes used for traditional subgradient optimization although it there refers to the projection of $\mathbf{u}^{t+\frac{1}{2}}$ onto $\Re_+^m$.) The direction $\mathbf{h}(\mathbf{x}^t) - \mathbf{v}^t$ from $\mathbf{u}^t \in \Re_+^m$, as defined in (10), is feasible in the program (5).

The convergence of the method (9) is established in [42] for two different step length rules. For the case when $\{\mathbf{v}^t\} = \{\mathbf{0}\}$, Polyak [51,52] and Ermol'ev [19], among others, establish convergence using several different step length rules (see also Shor [60, Chapter 2]). In this work we utilize special cases of the *divergent series* step length rule [19].

The next convergence result specializes that established by Larsson et al. [42, Theorem 2.7] to the Lagrangean dual problem (5); note that the boundedness condition on the sequence $\{\mathbf{v}^t\}$ can always be fulfilled by construction.

Proposition 4. *Suppose that Assumption 1 holds, and let the method* (9) *be applied to the program* (5), *with the step lengths α_t fulfilling the conditions*

$$\alpha_t > 0, \ \forall t, \quad \lim_{t\to\infty} \alpha_t = 0, \quad \lim_{t\to\infty} \sum_{s=0}^{t-1} \alpha_s = \infty, \quad \text{and} \quad \lim_{t\to\infty} \sum_{s=0}^{t-1} \alpha_s^2 < \infty. \quad (11)$$

If the sequence $\{\mathbf{v}^t\}$ is bounded, then $\{\mathbf{u}^t\} \to \mathbf{u}^\infty \in U^*$ and $\{\theta(\mathbf{u}^t)\} \to \theta^*$.

$\square$

To induce convergence of the method (9) for *adaptive* step length selection rules, based on line searches or formulas involving estimates of the optimal value, the *almost complete relaxation* strategy of Dem'yanov and Vasil'ev [15, Section 3.4] can be used ([42, Corollary 2.8]). It works as follows. Define the sequences $\{\underline{\alpha}_t\}$ and $\{\overline{\alpha}_t\}$ with $\underline{\alpha}_t \leq \overline{\alpha}_t$

for all t, and both satisfying the conditions (11).[1] Suppose that at some iteration t, an adaptive rule has generated a (tentative) step length $\widetilde{\alpha}_t$. The step length α_t used is defined as the projection of $\widetilde{\alpha}_t$ onto the interval $[\underline{\alpha}_t, \overline{\alpha}_t]$. The sequence $\{\alpha_t\}$, resulting from this safeguarding strategy, satisfies the conditions (11).

From Propositions 2 and 4 it follows that the set $\partial\theta(\mathbf{u}^\infty) \cap N_{\mathfrak{R}_+^m}(\mathbf{u}^\infty)$ is nonempty. The next proposition, which is a special case of Theorem 3.9 in [44], establishes that the sequence $\{\mathbf{h}(\mathbf{x}^t)\}$ of subgradients to the dual objective function converges in an ergodic sense to an element that verifies optimality in terms of Proposition 2. We define the sequence $\{A_t\}$ of cumulative step lengths by

$$A_t = \sum_{s=0}^{t-1} \alpha_s, \qquad t = 1, 2, \ldots . \tag{12}$$

Proposition 5. *Let Assumption 1 hold and the method* (9)–(11) *be applied to the program* (5). *Further, let the sequences* $\{A_t\}$ *and* $\{\mathbf{g}^t\}$ *be defined by* (12) *and*

$$\mathbf{g}^t = A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \mathbf{h}(\mathbf{x}^s), \qquad t = 1, 2, \ldots ,$$

respectively, where $\mathbf{x}^s \in X(\mathbf{u}^s)$ *for all s. If the sequence* $\{\boldsymbol{v}^t\}$ *is bounded, then*

$$\left\{ \text{dist}\left(\mathbf{g}^t, \partial\theta(\mathbf{u}^\infty) \cap N_{\mathfrak{R}_+^m}(\mathbf{u}^\infty) \right) \right\} \to 0.$$

$\square$

In the next section we establish that the sequence $\{\mathbf{x}^t\}$ of subproblem solutions converges in an ergodic sense to the solution set X^* as expressed in (7).

3. Ergodic primal convergence

The application of the method (9)–(11) to the program (5) produces a sequence $\{\mathbf{x}^t\}$ of solutions to the subproblem (3). We propose two schemes for generating an ergodic sequence of subproblem solutions; each of these sequences is shown to converge to the solution set, X^*. Their generation is computationally cheap, and their storage requires a relatively small amount of memory. In the first scheme, the sequence is defined by (convexity) weights that are proportional to the step lengths, α_t. The second scheme presumes the use of step lengths that generalize a modified harmonic series to allow for the utilization of the almost complete relaxation strategy (and which also satisfies the conditions (11)); in this case, the ergodic sequence is defined by equal weights. We also present a heuristic projection procedure for the finite attainment of primal ε-optimality.

Henceforth, we make repeated use of the following lemma; it is a special case of a result of Silverman and Toeplitz, and a proof can be found in, e.g., Knopp [36, Theorem 2, p. 35].

---

[1] It may, for example, be appropriate to let the sequences be given by $\underline{\alpha}_t = \mu/(b+t)$ and $\overline{\alpha}_t = M/(b+t)$, $t = 0, 1, \ldots$, where $\mu > 0$ ($M > 0$) is a very small (very large) constant and $b > 0$.

Lemma 3. *Assume that the sequence $\{\beta_{ts}\} \subset \Re$ fulfils the conditions*

$$\beta_{ts} \geq 0, \quad s = 0, \dots, t-1, \quad \sum_{s=0}^{t-1} \beta_{ts} = 1, \quad t = 1, 2, \dots,$$

$$and \quad \lim_{t \to \infty} \beta_{ts} = 0, \quad s = 0, 1, \dots.$$

If the sequence $\{\mathbf{b}^s\} \subset \Re^r$ is such that $\lim_{s \to \infty} \mathbf{b}^s = \mathbf{b}$, then $\lim_{t \to \infty} \left(\sum_{s=0}^{t-1} \beta_{ts} \mathbf{b}^s \right) = \mathbf{b}$.

$\square$

3.1. Divergent series step lengths and proportional weights

The ergodic sequence $\{\bar{\mathbf{x}}^t\}$ of subproblem solutions that is computed within the method (9)–(11) applied to the program (5), is defined as the weighted averages

$$\bar{\mathbf{x}}^t = A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \mathbf{x}^s, \quad t = 1, 2, \dots, \tag{13}$$

where the sequence $\{A_t\}$ is defined in (12). Hence, each vector $\bar{\mathbf{x}}^t$ is a convex combination of the subproblem solutions found up to iteration t, and thus $\bar{\mathbf{x}}^t \in X$ holds for all t.

The convergence of the sequence $\{\bar{\mathbf{x}}^t\}$ to the set X^* is established in terms of the fulfilment of the optimality conditions of Proposition 3.

Theorem 1 ($\{\bar{\mathbf{x}}^t\}$ **converges to the solution set**). *Suppose that Assumption 1 holds, let the method (9)–(11) be applied to the program (5), the set X^* and the sequence $\{\bar{\mathbf{x}}^t\}$ be given by the definition (7) and (13), respectively, and suppose that the sequence $\{\mathbf{v}^t\}$ is bounded. Then,*

$$\{\mathrm{dist}\left(\bar{\mathbf{x}}^t, X^*\right)\} \to 0.$$

Proof. Letting $\mathbf{u}^\infty$ be the limit of the sequence $\{\mathbf{u}^t\}$, as given in Proposition 4, we first show that the sequence $\{\bar{\mathbf{x}}^t\}$ converges to the set of subproblem solutions at $\mathbf{u}^\infty$, then that it is feasible in the limit, and, finally, that it is complementary to $\mathbf{u}^\infty$ in the limit.

By the convexity and nonnegativity of the function $\mathrm{dist}(\cdot, S)$, and the definition (12), the inequalities

$$0 \leq \mathrm{dist}\left(\bar{\mathbf{x}}^t, X(\mathbf{u}^\infty)\right) \leq A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \mathrm{dist}\left(\mathbf{x}^s, X(\mathbf{u}^\infty)\right)$$

hold for all t. By Lemma 1 and Proposition 4,

$$\{\mathrm{dist}\left(\mathbf{x}^s, X(\mathbf{u}^\infty)\right)\} \to 0 \quad \text{as} \quad s \to \infty. \tag{14}$$

Utilizing Lemma 3, with $\beta_{ts} = A_t^{-1} \alpha_s$, $\mathbf{b}^s = \mathrm{dist}\{\mathbf{x}^s, X(\mathbf{u}^\infty)\}$ and $\mathbf{b} = 0$, it then follows that

$$\{\mathrm{dist}\left(\bar{\mathbf{x}}^t, X(\mathbf{u}^\infty)\right)\} \to 0 \quad \text{as} \quad t \to \infty. \tag{15}$$

By the convexity of the functions $h_i, i \in \mathcal{I}$, $\mathbf{h}(\overline{\mathbf{x}}^t) \leq A_t^{-1} \sum_{s=0}^{t-1} \alpha_s \mathbf{h}(\mathbf{x}^s)$ for all t, and from the iteration formula (9) it follows that $\mathbf{h}(\mathbf{x}^s) \leq (\mathbf{u}^{s+1} - \mathbf{u}^s)/\alpha_s$ for all s. Hence, $\mathbf{h}(\overline{\mathbf{x}}^t) \leq A_t^{-1}(\mathbf{u}^t - \mathbf{u}^0)$ for all t. Proposition 4 implies that the sequence $\{\mathbf{u}^t - \mathbf{u}^0\}$ is bounded and, therefore,

$$\limsup_{t \to \infty} h_i(\overline{\mathbf{x}}^t) \leq 0, \qquad \forall i \in \mathcal{I}. \tag{16}$$

Now, consider an $i \in \mathcal{I}(\mathbf{u}^\infty)$. From Proposition 4 it follows that, for some fixed τ that is large enough, $u_i^t > 0$ for all $t \geq \tau$, and, by the iteration formula (9), $v_i^t u_i^t = 0$ holds for all t. Therefore, by (9),

$$h_i(\mathbf{x}^t) = \frac{u_i^{t+1} - u_i^t}{\alpha_t}, \qquad \forall t \geq \tau. \tag{17}$$

Choosing $\overline{\mathbf{x}} \in \mathrm{rint}\, X(\mathbf{u}^\infty)$ and $\boldsymbol{\xi}_i \in \partial h_i(\overline{\mathbf{x}})$, Lemma 2 yields that

$$h_i(\mathbf{x}) = h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T (\mathbf{x} - \overline{\mathbf{x}}), \qquad \forall \mathbf{x} \in X(\mathbf{u}^\infty).$$

Then, by the continuity of the function h_i, for every $\delta > 0$ there exists an $\varepsilon > 0$ such that

$$h_i(\mathbf{x}) \leq h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T (\mathbf{x} - \overline{\mathbf{x}}) + \frac{\delta}{3}, \qquad \forall \mathbf{x} : \mathrm{dist}\left(\mathbf{x}, X(\mathbf{u}^\infty)\right) \leq \varepsilon.$$

From (14) follows that $\mathrm{dist}(\mathbf{x}^s, X(\mathbf{u}^\infty)) \leq \varepsilon$ for all $s \geq \kappa$, for some fixed $\kappa > \tau$. The relation (17) then yields

$$h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T \left(\mathbf{x}^s - \overline{\mathbf{x}}\right) \geq \frac{u_i^{s+1} - u_i^s}{\alpha_s} - \frac{\delta}{3}, \qquad \forall s \geq \kappa. \tag{18}$$

Using the definition (13), we have for all $t > \kappa$,

$$\begin{aligned}
h_i(\overline{\mathbf{x}}^t) &\geq h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T \left(\overline{\mathbf{x}}^t - \overline{\mathbf{x}}\right) \\
&\geq A_t^{-1} \sum_{s=0}^{\kappa-1} \alpha_s \left(h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T \left(\mathbf{x}^s - \overline{\mathbf{x}}\right)\right) + A_t^{-1} \sum_{s=\kappa}^{t-1} \left(u_i^{s+1} - u_i^s - \frac{\alpha_s \delta}{3}\right) \\
&= \frac{A_\kappa}{A_t} \left(h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T \left(\overline{\mathbf{x}}^\kappa - \overline{\mathbf{x}}\right)\right) + \frac{u_i^t - u_i^\kappa}{A_t} - \left(1 - \frac{A_\kappa}{A_t}\right) \frac{\delta}{3},
\end{aligned}$$

where the first inequality follows from the definition of $\boldsymbol{\xi}_i$ and the second is implied by the inequality (18). Since $\{A_t\} \to \infty$ and $\{u_i^t\} \to u_i^\infty$, then $A_t^{-1} A_\kappa(h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T(\overline{\mathbf{x}}^\kappa - \overline{\mathbf{x}})) \geq -\delta/3$ and $A_t^{-1}(u_i^t - u_i^\kappa) \geq -\delta/3$ for all $t > \kappa$ that are large enough. It follows that $h_i(\overline{\mathbf{x}}^t) \geq -\delta$ for all $t > \kappa$, that are large enough. Therefore, $\liminf_{t \to \infty} h_i(\overline{\mathbf{x}}^t) \geq 0$, and the inequalities (16) then yield that $\lim_{t \to \infty} h_i(\overline{\mathbf{x}}^t) = 0$. Since this result holds for all $i \in \mathcal{I}(\mathbf{u}^\infty)$, and, by the definition (6), $u_i^\infty = 0$ for all $i \in \mathcal{I} \setminus \mathcal{I}(\mathbf{u}^\infty)$, it follows that

$$\left\{(\mathbf{u}^\infty)^T \mathbf{h}(\overline{\mathbf{x}}^t)\right\} \to 0 \qquad \text{as} \qquad t \to \infty. \tag{19}$$

The theorem follows from the relations (15), (16) and (19), and Proposition 3.

$\square$

For the case when the functions f and $h_i, i \in \mathcal{I}$, are affine and the set X is a polytope (that is, when (1) is a linear program), and $\{\boldsymbol{v}^t\} = \{\mathbf{0}\}$ in the method (9)–(11) applied to the program (5), Theorem 1 reduces to a result of Shor [60, pp. 116–118].

The next result follows from Proposition 4, Theorem 1, and the relation (8).

Corollary 1 ($\overline{\mathbf{x}}^t$ **verifies optimality in the limit).** *Under the assumptions of Theorem 1,*

$$\left\{ \mathtt{dist}\left(\mathbf{h}(\overline{\mathbf{x}}^t), \partial\theta(\mathbf{u}^\infty) \cap N_{\mathfrak{R}^m_+}(\mathbf{u}^\infty) \right) \right\} \to 0.$$

$\square$

3.2. Generalization of modified harmonic series step lengths, and equal weights

Now, let the step lengths used in the conditional subgradient optimization method (9) be minorized and majorized, respectively, by the elements of two modified harmonic series, that is

$$\alpha_t \in \left[\frac{\mu}{b+t}, \frac{M}{b+t} \right], \qquad b > 0, \qquad 0 < \mu \leq M < \infty, \qquad t = 0, 1, \dots, \quad (20)$$

and let the sequence $\{\widehat{\mathbf{x}}^t\}$ of averaged subproblem solutions be defined by equal weights, as

$$\widehat{\mathbf{x}}^t = \frac{1}{t} \sum_{s=0}^{t-1} \mathbf{x}^s, \qquad t = 1, 2, \dots . \quad (21)$$

Analogously to the previous subsection, we will here derive the convergence of the sequence $\{\widehat{\mathbf{x}}^t\}$ to the solution set X^*, as expressed in the definition (7).

Theorem 2 ($\{\widehat{\mathbf{x}}^t\}$ **converges to the solution set).** *Suppose that Assumption 1 holds, let the method* (9), (20) *be applied to the program* (5), *the set X^* and the sequence $\{\widehat{\mathbf{x}}^t\}$ be given by the definitions* (7) *and* (21), *respectively, and suppose that the sequence $\{\boldsymbol{v}^t\}$ is bounded. Then,*

$$\left\{ \mathtt{dist}\left(\widehat{\mathbf{x}}^t, X^* \right) \right\} \to 0.$$

Proof. Using arguments analogous to those used in the derivation of the result (15), we have that

$$\left\{ \mathtt{dist}\left(\widehat{\mathbf{x}}^t, X(\mathbf{u}^\infty) \right) \right\} \to 0 \quad \text{as} \quad t \to \infty. \quad (22)$$

By the iteration formula (9) and the definition (20), $\mathbf{h}(\mathbf{x}^s) \leq \alpha_s^{-1}(\mathbf{u}^{s+1} - \mathbf{u}^s) \leq \mu^{-1}(b+s)(\mathbf{u}^{s+1} - \mathbf{u}^s)$, for all s. Hence, for all $t \geq 1$,

$$\frac{1}{t} \sum_{s=0}^{t-1} \mathbf{h}(\mathbf{x}^s) \leq \frac{b}{\mu t} \sum_{s=0}^{t-1} \left(\mathbf{u}^{s+1} - \mathbf{u}^s \right) + \frac{1}{\mu t} \sum_{s=0}^{t-1} s \left(\mathbf{u}^{s+1} - \mathbf{u}^s \right)$$

$$= \frac{b-1}{\mu t} \left(\mathbf{u}^t - \mathbf{u}^0 \right) + \frac{1}{\mu} \left(\mathbf{u}^t - \frac{1}{t} \sum_{s=0}^{t-1} \mathbf{u}^s \right). \quad (23)$$

Since $\{\mathbf{u}^t\} \to \mathbf{u}^\infty$, $\lim_{t\to\infty}(\mu t)^{-1}(b-1)(\mathbf{u}^t - \mathbf{u}^0) = 0$. Applying Lemma 3, with $\beta_{ts} = t^{-1}$, $\mathbf{b}^s = \mathbf{u}^s$, and $\mathbf{b} = \mathbf{u}^\infty$, yields that $\{t^{-1}\sum_{s=0}^{t-1}\mathbf{u}^s\} \to \mathbf{u}^\infty$. It follows that the right-hand side of (23) tends to the origin as $t \to \infty$. By the convexity of the functions h_i, $h_i(\widehat{\mathbf{x}}^t) \le t^{-1}\sum_{s=0}^{t-1} h_i(\mathbf{x}^s)$ holds for all $i \in \mathcal{I}$, and therefore

$$\limsup_{t\to\infty} h_i(\widehat{\mathbf{x}}^t) \le 0, \qquad \forall i \in \mathcal{I}. \tag{24}$$

Now, consider an $i \in \mathcal{I}(\mathbf{u}^\infty)$, and choose an $\overline{\mathbf{x}} \in \mathtt{rint}\, X(\mathbf{u}^\infty)$ and $\boldsymbol{\xi}_i \in \partial h_i(\overline{\mathbf{x}})$. Similarly to the derivation of the inequalities (18), utilizing the definition (20), we have that, for every $\delta > 0$ there exists a $\kappa > 0$ such that

$$h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T\left(\mathbf{x}^s - \overline{\mathbf{x}}\right) \ge \frac{b+s}{M}\left(u_i^{s+1} - u_i^s\right) - \frac{\delta}{4}, \qquad \forall s \ge \kappa. \tag{25}$$

Using the definition (21) we have, for all $t > \kappa$,

$$h_i(\widehat{\mathbf{x}}^t) \ge h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T\left(\widehat{\mathbf{x}}^t - \overline{\mathbf{x}}\right)$$

$$\ge \frac{1}{t}\sum_{s=0}^{\kappa-1}\left(h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T\left(\mathbf{x}^s - \overline{\mathbf{x}}\right)\right) + \frac{b}{Mt}\sum_{s=\kappa}^{t-1}\left(u_i^{s+1} - u_i^s\right)$$

$$+ \frac{1}{Mt}\sum_{s=\kappa}^{t-1} s\left(u_i^{s+1} - u_i^s\right) - \frac{t-\kappa}{t}\cdot\frac{\delta}{4}$$

$$= \frac{\kappa}{t}\left(h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T\left(\widehat{\mathbf{x}}^\kappa - \overline{\mathbf{x}}\right)\right) + \frac{b-1+\kappa}{Mt}\left(u_i^t - u_i^\kappa\right)$$

$$+ \frac{1}{M}\left(u_i^t - \left(1 - \frac{\kappa}{t}\right)\frac{1}{t-\kappa}\sum_{s=\kappa}^{t-1}u_i^s\right) - \left(1 - \frac{\kappa}{t}\right)\frac{\delta}{4},$$

where the first inequality follows from the definition of $\boldsymbol{\xi}_i$ and the second from the inequality (25). Since $\{u_i^t\} \to u_i^\infty$, for all $t > \kappa$ that are large enough, it follows that $t^{-1}\kappa(h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T(\widehat{\mathbf{x}}^\kappa - \overline{\mathbf{x}})) \ge -\delta/4$, $(Mt)^{-1}(b-1+\kappa)(u_i^t - u_i^\kappa) \ge -\delta/4$, and, utilizing Lemma 3, that $M^{-1}(u_i^t - (1 - t^{-1}\kappa)(t-\kappa)^{-1}\sum_{s=\kappa}^{t-1}u_i^s) \ge -\delta/4$. It follows that $h_i(\widehat{\mathbf{x}}^t) \ge -\delta$ for all $t > \kappa$ that are large enough. Analogous to the result (19), then

$$\left\{(\mathbf{u}^\infty)^T\mathbf{h}(\widehat{\mathbf{x}}^t)\right\} \to 0 \qquad \text{as} \qquad t \to \infty. \tag{26}$$

The theorem follows from the relations (22), (24) and (26), and Proposition 3.

$\square$

For the case when (1) is a linear program, and $\{\boldsymbol{v}^t\} = \{\mathbf{0}\}$ in the method (9), (20) applied to the program (5), our Theorem 2 reduces to Theorem 3 of Larsson and Liu [39].

The next result is a consequence of Proposition 4, Theorem 2, and the relation (8).

Corollary 2 ($\widehat{\mathbf{x}}^t$ **verifies optimality in the limit**). *Under the assumptions of Theorem 2,*

$$\left\{\mathtt{dist}\left(\mathbf{h}(\widehat{\mathbf{x}}^t), \partial\theta(\mathbf{u}^\infty) \cap N_{\mathfrak{R}_+^m}(\mathbf{u}^\infty)\right)\right\} \to 0.$$

$\square$

3.3. Finite primal feasibility and ε-optimality

The following discussion concerns the method of Section 3.1, but is applicable to that of Section 3.2 as well.

Theorem 1 establishes optimality in the limit for the sequence $\{\bar{\mathbf{x}}^t\}$ of primal solutions. While dual feasibility is maintained throughout the iterative procedure (9), neither primal feasibility nor complementarity will, in general, be finitely satisfied by the elements of the sequence $\{\bar{\mathbf{x}}^t\}$. However, eventually $\bar{\mathbf{x}}^t$ will be both near-feasible and near-complementary. For applications with soft constraints, such as the forest management scheduling problem presented by Hauer and Hoganson [29], for which primal feasibility is not crucial, the elements $\bar{\mathbf{x}}^t$ will thus eventually be sufficiently close to a feasible (and optimal) solution. Nevertheless, whenever primal feasibility is required finitely, it might be necessary to apply a procedure that converts any finitely generated primal solution $\bar{\mathbf{x}}^t$ into a feasible solution to the original convex program. One procedure for enforcing feasibility is the solution of the Euclidean projection problem

$$\texttt{proj}\left(\bar{\mathbf{x}}^t, Y\right) = \arg\min_{\mathbf{x} \in Y} \left\| \mathbf{x} - \bar{\mathbf{x}}^t \right\|_2, \qquad \text{where} \qquad Y = \{\, \mathbf{x} \in X \mid \mathbf{h}(\mathbf{x}) \leq \mathbf{0}\,\}. \tag{27}$$

Solving this program regularly may, however, be computationally too expensive. It is probably better to develop a heuristic procedure, which exploits the structure of the set Y when searching for a feasible and near-optimal solution to the projection problem (27).

We let $\texttt{proj}_H(\mathbf{x}, Y)$ denote a heuristic projection of a point $\mathbf{x} \in \Re^n$ onto the feasible set of the program (1) which enjoys the following property.

Assumption 2 (Property of heuristic projection). [2] *Let $\mathbf{x} \in X$ and the set Y be defined by (27). There exists a function $\delta : \Re_+ \mapsto \Re_+$ such that $\delta(\varepsilon) > 0$ for all $\varepsilon > 0$, $\lim_{\varepsilon \to 0^+} \delta(\varepsilon) = 0$, and*

$$\|\texttt{proj}_H(\mathbf{x}, Y) - \texttt{proj}(\mathbf{x}, Y)\|_2 \leq \delta(\varepsilon) \qquad \text{whenever} \qquad \texttt{dist}(\mathbf{x}, Y) \leq \varepsilon.$$

$\square$

Theorem 3 (Convergence by heuristic projection). *Let the set Y be defined by (27) and suppose that Assumption 1 holds. Let the method (9)–(11) be applied to the program (5), the sequence $\{\bar{\mathbf{x}}^t\}$ be given by the definition (13), and suppose that the sequence $\{v^t\}$ is bounded. Then, under Assumption 2,*

$$\left\{ \texttt{dist}\left(\texttt{proj}_H\left(\bar{\mathbf{x}}^t, Y\right), X^*\right) \right\} \to 0.$$

Proof. By the definition (2) and the triangle inequality,

$$\texttt{dist}\left(\texttt{proj}_H\left(\bar{\mathbf{x}}^t, Y\right), X^*\right) \leq \left\| \texttt{proj}_H\left(\bar{\mathbf{x}}^t, Y\right) - \texttt{proj}\left(\bar{\mathbf{x}}^t, Y\right)\right\|_2$$
$$+ \texttt{dist}\left(\bar{\mathbf{x}}^t, Y\right) + \texttt{dist}\left(\bar{\mathbf{x}}^t, X^*\right). \tag{28}$$

---

[2] This assumption expresses a continuity property on the difference between the heuristic projection and the exact Euclidean projection.

Since $X^* \subseteq Y$, it follows from Theorem 1 that for any $\varepsilon > 0$ there is a $\tau > 0$ such that $\mathrm{dist}\left(\overline{\mathbf{x}}^t, Y\right) \leq \mathrm{dist}\left(\overline{\mathbf{x}}^t, X^*\right) \leq \varepsilon$ for all $t \geq \tau$. Then, by Assumption 2, $\|\mathrm{proj}_H(\overline{\mathbf{x}}^t, Y) - \mathrm{proj}(\overline{\mathbf{x}}^t, Y)\|_2 \leq \delta(\varepsilon)$ for all $t \geq \tau$. From the inequality (28) it then follows that $\mathrm{dist}(\mathrm{proj}_H(\overline{\mathbf{x}}^t, Y), X^*) \leq \delta(\varepsilon) + 2\varepsilon$ for all $t \geq \tau$. The result follows by letting $\varepsilon \to 0^+$.

$\square$

We can now construct an algorithm that employs heuristic projections and yields convergence to the optimal value in the primal as well as in the dual procedure.

Corollary 3 (Finite termination at ε-optimality). *Let the assumptions of Theorem 3 hold. For every $\varepsilon > 0$ there is a $\tau > 0$ such that $f(\mathrm{proj}_H(\overline{\mathbf{x}}^t, Y)) - \theta(\mathbf{u}^t) \leq \varepsilon$ holds for all $t \geq \tau$.*

Proof. Choose an $\varepsilon > 0$. From Theorem 3 and the continuity of the function f, it follows that there is a $\kappa > 0$ such that $f(\mathrm{proj}_H(\overline{\mathbf{x}}^t, Y)) \leq f^* + \varepsilon/2$ for all $t \geq \kappa$. By Proposition 4, there is a $\tau \geq \kappa$ such that $\theta(\mathbf{u}^t) \geq \theta^* - \varepsilon/2$ holds for all $t \geq \tau$. By Proposition 3, $f^* = \theta^*$, and therefore, $f(\mathrm{proj}_H(\overline{\mathbf{x}}^t, Y)) - \theta(\mathbf{u}^t) \leq \varepsilon$ for all $t \geq \tau$.

$\square$

Since exact Euclidean projection is a special case of heuristic projection, the finite attainment of ε-optimality also occurs when an exact Euclidean projection is employed.

Remark 1 (Delayed start of averaging). When developing the results of this section we utilize the properties *in the limit* of the sequence $\{\overline{\mathbf{x}}^t\}$. Since the information from the initial iterations of a subgradient optimization scheme is usually of low quality, it is preferable to delay the initialization of the sequence $\{\overline{\mathbf{x}}^t\}$ until the iterates, $\mathbf{u}^t$, are near-optimal, in the sense that they are located on optimal segments of the dual objective function (implying that $\mathbf{h}(\mathbf{x}^t) \in \partial\theta(\mathbf{u}^\infty)$ if θ is polyhedral, and an indication of which is that certain vectors occur repeatedly as subproblem solutions), and that $u_i^t > 0$ holds for all i such that $u_i^\infty > 0$ (implying that $\mathbf{v}^t + \alpha_t^{-1}(\mathbf{u}^{t+\frac{1}{2}} - \mathbf{u}^{t+1}) \in N_{\mathfrak{R}_+^m}(\mathbf{u}^\infty)$). This behaviour is also apparent from the numerical experiments performed in the following section. In our theoretical analysis, we initialize the ergodic sequences at the first iteration. Since subgradient methods are memoryless, all of the results can, however, be modified to the initialization of these sequences at any iteration $t_0 \geq 1$, with the obvious modification of the convex combination formula (13).

$\square$

4. Application to traffic equilibrium assignment under road pricing

Traffic assignment deals with the estimation of route flows in each of the origin–destination relations of a road network, and the travel times resulting from this allocation. We consider the traffic equilibrium assignment model, which is based on the assumption that all road users have complete information about the current traffic conditions, and that they choose among the shortest routes available. An equilibrium state therefore has the property that the travel times on routes that are used are equal, that is, they are all

shortest with respect to the current traffic flow (Wardrop [64]). Since travellers use the shortest routes, they do not generally fulfil society's goal of efficient road usage (which may, for example, be a minimal total travel time). In order to fulfil this goal, during the last several years *road pricing* systems have been paid much attention (e.g., Small [61] and *The Economist* [18]). In such systems, each vehicle pays a prespecified toll for traversing certain arcs, possibly relative to the current traffic flow on these arcs.

4.1. Statement of the model

Consider a connected transportation network $\mathcal{G} = (\mathcal{N}, \mathcal{A})$, with node set $\mathcal{N}$ and a set $\mathcal{A}$ of directed arcs. Let $\mathcal{C} \subset \mathcal{N} \times \mathcal{N}$ be a set of origin–destination pairs. For each pair, $k \in \mathcal{C}$, there is a fixed positive demand d_k of flow, associated with a specific commodity. We denote the nonempty set of simple routes from the origin to the destination of the k:th pair by $\mathcal{R}_k$ and the flow on route $r \in \mathcal{R}_k$ by h_{kr}. Feasibility with respect to demands requires the route flows $[h_{kr}]_{r \in \mathcal{R}_k, k \in \mathcal{C}}$ to satisfy the conditions

$$\sum_{r \in \mathcal{R}_k} h_{kr} = d_k, \qquad k \in \mathcal{C}, \tag{29a}$$

$$h_{kr} \geq 0, \qquad r \in \mathcal{R}_k, \ k \in \mathcal{C}. \tag{29b}$$

Letting $[\delta_{kra}]_{r \in \mathcal{R}_k, k \in \mathcal{C}, a \in \mathcal{A}}$ be an arc-route incidence matrix for $\mathcal{G}$, with

$$\delta_{kra} = \begin{cases} 1, & \text{if route } r \in \mathcal{R}_k \text{ contains arc } a \in \mathcal{A}, \\ 0, & \text{otherwise,} \end{cases}$$

the arc flows, f_a, are defined by the route flows, h_{kr}, through

$$f_a = \sum_{k \in \mathcal{C}} \sum_{r \in \mathcal{R}_k} \delta_{kra} h_{kr}, \qquad a \in \mathcal{A}. \tag{29c}$$

With each arc $a \in \mathcal{A}$ is associated the travel time $c_a(f_a)$ for traversing it at arc flow $f_a \geq 0$. We presume that the free-flow travel time, $c_a(0)$, is positive, and that the functions $c_a : \Re_+ \mapsto \Re_+$ are continuous, strictly increasing, and weakly coercive, that is, they tend to infinity with the flow (these assumptions are quite natural because of the congestion effects).

The Wardrop equilibrium conditions for the model (29a)–(29c) with cost functions c_a, $a \in \mathcal{A}$, are equivalent to the first-order optimality conditions (e.g., Patriksson [49, Theorem 2.1]) for the program

$$\min \sum_{a \in \mathcal{A}} \int_0^{f_a} c_a(s) ds, \tag{30}$$
$$\text{s.t. (29a)–(29c),}$$

which has a convex and differentiable objective function and linear constraints.

For each arc $a \in \mathcal{A}$, let $\tau_a \geq 0$ be the toll (measured in time equivalents) charged each vehicle traversing the arc when the arc flow is greater than $\varphi_a \geq 0$. The toll mapping $\pi_a : \Re_+ \mapsto 2^{\Re_+}$ is defined by[3]

$$\pi_a(f_a) = \begin{cases} \{0\}, & 0 \leq f_a < \varphi_a, \\ [0, \tau_a], & f_a = \varphi_a, \qquad a \in \mathcal{A}. \\ \{\tau_a\}, & f_a > \varphi_a, \end{cases} \tag{31}$$

The mapping π_a is nonnegative, convex-valued,[4] and closed (see Section 2.1) on $\Re_+$ for all $a \in \mathcal{A}$. Figure 1 illustrates three different types of toll mappings.
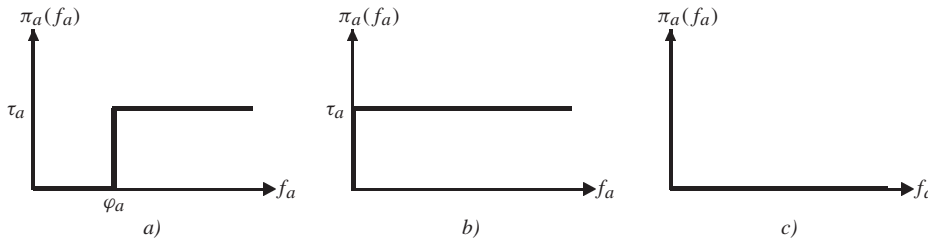
Fig. 1. The toll mapping π_a for arc $a \in \mathcal{A}$ with respect to the arc flow $f_a \geq 0$. a) $\varphi_a > 0$, $\pi_a > 0$. b) $\varphi_a = 0$, $\pi_a > 0$. c) $\pi_a = 0$, i.e., an untolled arc

Letting μ_k denote the travel cost (time plus toll) at flow f_a, $a \in \mathcal{A}$, on the shortest route from the origin to the destination of the k:th pair, the Wardrop user equilibrium conditions for the model (29) with cost mappings $c_a + \pi_a$ may be stated as

$$h_{kr} > 0 \implies \sum_{a \in \mathcal{A}} \delta_{kra} (c_a(f_a) + p_a(f_a)) = \mu_k, \qquad r \in \mathcal{R}_k, \tag{32a}$$

$$h_{kr} = 0 \implies \sum_{a \in \mathcal{A}} \delta_{kra} (c_a(f_a) + p_a(f_a)) \geq \mu_k, \qquad r \in \mathcal{R}_k, \tag{32b}$$

where

$$p_a(f_a) \in \pi_a(f_a), \qquad a \in \mathcal{A}, \tag{32c}$$

is the toll charged for traversing arc a at flow f_a. According to (31), if $\tau_a > 0$ and $f_a = \varphi_a$, then the toll level $p_a(f_a)$ has a certain degree of freedom; by allowing this degree of freedom, we can show that an equilibrium can be found by solving a (nonsmooth) convex optimization problem. Asmuth [1] shows that a user equilibrium always exists when the network is strongly connected and the cost mappings are positive,

---

[3] All the results to be derived may be generalized to several toll levels for each arc $a \in \mathcal{A}$. For simplicity of notation, however, we consider only one level (which may also be zero) for each arc.

[4] A point–to–set map $\pi : \Re_+ \mapsto 2^{\Re_+}$ is convex-valued if the set $\pi(f)$ is convex for all $f \in \Re_+$.

upper semicontinuous,[5] and convex-valued. Bernstein and Smith [6] consider single-valued, lower semicontinuous[6] cost functions; they show that under fairly mild regularity conditions a user equilibrium always exists. The theory of road pricing has also been discussed by Dafermos and Sparrow [12], among others. However, none of these papers describe equivalent optimization models.

Similarly to the formulation (30) it can be shown (see the proof of Proposition 6) that the conditions (29) and (32) are the first-order optimality conditions for the program

$$z^* = \min \sum_{a \in \mathcal{A}} \int_0^{f_a} (c_a(s) + \pi_a(s)) \, ds, \tag{33a}$$

$$\text{s.t.} \qquad \sum_{r \in \mathcal{R}_k} h_{kr} = d_k, \qquad k \in \mathcal{C}, \tag{33b}$$

$$h_{kr} \geq 0, \qquad r \in \mathcal{R}_k, \ \ k \in \mathcal{C}, \tag{33c}$$

$$\sum_{k \in \mathcal{C}} \sum_{r \in \mathcal{R}_k} \delta_{kra} h_{kr} = f_a, \qquad a \in \mathcal{A}, \tag{33d}$$

$$f_a \geq 0, \qquad a \in \mathcal{A}. \tag{33e}$$

(The constraints (33e) are implied by (33c)–(33d); they have been added to strengthen the Lagrangean dual formulation to be presented.) As a result of the assumptions made above, this program is a highly structured, convex optimization problem with a nonlinear and possibly nonsmooth objective function. A real world instance of this model may have several thousand nodes, arcs, and origin–destination pairs.

From the properties of the functions c_a and point–to–set mappings π_a, it follows that the objective function (33a) is strictly convex with respect to arc flows. Hence, the optimal arc flow, f_a^*, $a \in \mathcal{A}$, is unique. However, if the arc flow variables are eliminated from the problem, then the objective of the resulting equivalent problem in route flow variables is, in general, non-strictly convex, since an arc flow pattern may correspond to several route flow patterns. Hence, the sets of optimal route flows, H_k^*, $k \in \mathcal{C}$, are in general *not* singleton sets, but polytopes.

4.2. A Lagrangean dual formulation

For the untolled program (30) Larsson et al. [40] propose a Lagrangean dual approach in which the arc flow defining constraints (29c) are relaxed. The resulting solution method essentially consists of the repeated solution of shortest path problems; it is very simple, both from a conceptual and implementational point of view. Its merits are that, despite its dual character, it produces a feasible flow in each iteration, and that this is done without the solution of any additional optimization problem. In this paper, we generalize this method to the model (33).

---

[5] A point–to–set map $\pi : \Re_+ \mapsto 2^{\Re_+}$ is upper semicontinuous if $f_i \in \Re_+$ for all i, $\{f_i\} \to f$ and $t_i \in \pi(f_i)$, for all i, imply that $\{t_i\}_{i \in \mathcal{I}} \to t \in \pi(f)$ for some subsequence $\mathcal{I}$.

[6] A single-valued function $c : \Re \mapsto \Re$ is lower semicontinuous if $\liminf_{f \to \overline{f}} c(f) \geq c(\overline{f})$ for all $\overline{f} \in \Re$.

Letting $\mathbf{u} = [u_a]_{a \in \mathcal{A}}$ be multipliers associated with the constraints (33d), we define the Lagrangean dual objective function (cf., Definition 3) by

$$\theta(\mathbf{u}) = \sum_{k \in \mathcal{C}} \theta_k(\mathbf{u}) + \sum_{a \in \mathcal{A}} \theta_a(u_a).$$

For each $k \in \mathcal{C}$ and all $\mathbf{u} \in \Re^m$, $\theta_k(\mathbf{u})$ is the optimal value of the shortest simple route subproblem, with arc costs u_a, $a \in \mathcal{A}$, given by

$$
\begin{aligned}
\theta_k(\mathbf{u}) = \min \sum_{r \in \mathcal{R}_k} & \left(\sum_{a \in \mathcal{A}} u_a \delta_{kra} \right) h_{kr}, \\
\text{s.t.} \sum_{r \in \mathcal{R}_k} & h_{kr} = d_k, \\
& h_{kr} \geq 0, \quad r \in \mathcal{R}_k.
\end{aligned}
\tag{34}
$$

The solution set $H_k(\mathbf{u})$ to this program is a bounded polyhedron; it is not necessarily a singleton set (especially not at an optimal point, $\mathbf{u}^*$). We let $[h_{kr}(\mathbf{u})]_{r \in \mathcal{R}_k} \in H_k(\mathbf{u})$, $k \in \mathcal{C}$, denote an arbitrary set of solutions to the subproblems (34) at $\mathbf{u}$. Note that, by a result of Robinson [54, Lemma 3.5], $H_k(\mathbf{u}) \subseteq H_k(\mathbf{u}^*)$ holds for any $\mathbf{u}$ in a sufficiently small neighbourhood of $\mathbf{u}^*$; this result has a bearing to the algorithm devised in Section 4.3.

For each $a \in \mathcal{A}$ and all $u_a \in \Re$, $\theta_a(u_a)$ is the optimal value of the single–arc subproblem

$$\theta_a(u_a) = \min_{f_a \geq 0} \int_0^{f_a} (c_a(s) + \pi_a(s) - u_a)\, ds, \tag{35}$$

which has a strictly convex and nonsmooth objective function. Since each mapping $c_a + \pi_a$ is strictly increasing and weakly coercive, the program (35) is uniquely solved by

$$
f_a(u_a) = \begin{cases}
0, & u_a \leq c_a(0), \\
c_a^{-1}(u_a), & c_a(0) \leq u_a \leq c_a(\varphi_a), \\
\varphi_a, & c_a(\varphi_a) \leq u_a \leq c_a(\varphi_a) + \tau_a, \\
c_a^{-1}(u_a - \tau_a), & u_a \geq c_a(\varphi_a) + \tau_a,
\end{cases} \quad a \in \mathcal{A}, \tag{36}
$$

where c_a^{-1} is the continuous inverse mapping (e.g., Rudin [57, Theorem 4.17]) of the continuous one–to–one mapping c_a, $a \in \mathcal{A}$. One may note that c_a^{-1} is explicit for most travel time functions used and that c_a need not be differentiable. Figure 2 illustrates the function $u_a \mapsto f_a(u_a)$ for arc $a = (16, 18)$ in the Sioux Falls network (see Section 4.4), with $c_a(f_a) = 0.03 + 0.00000003 \cdot f_a^4$.

The function $\theta : \Re^{|\mathcal{A}|} \mapsto \Re$ is the sum of the $|\mathcal{C}|$ concave and piecewise linear functions θ_k, $k \in \mathcal{C}$, and the $|\mathcal{A}|$ concave and differentiable functions θ_a, $a \in \mathcal{A}$. It is thus finite, continuous, concave, and subdifferentiable on $\Re^{|\mathcal{A}|}$; its subdifferential mapping at $\mathbf{u} \in \Re^{|\mathcal{A}|}$ is the bounded polyhedron (cf., Proposition 1)

$$\partial\theta(\mathbf{u}) = \left\{ \left[\sum_{k \in \mathcal{C}} \sum_{r \in \mathcal{R}_k} \delta_{kra} h_{kr} - f_a(u_a) \right]_{a \in \mathcal{A}} \middle| [h_{kr}]_{r \in \mathcal{R}_k} \in H_k(\mathbf{u}), \quad k \in \mathcal{C} \right\}. \tag{37}$$
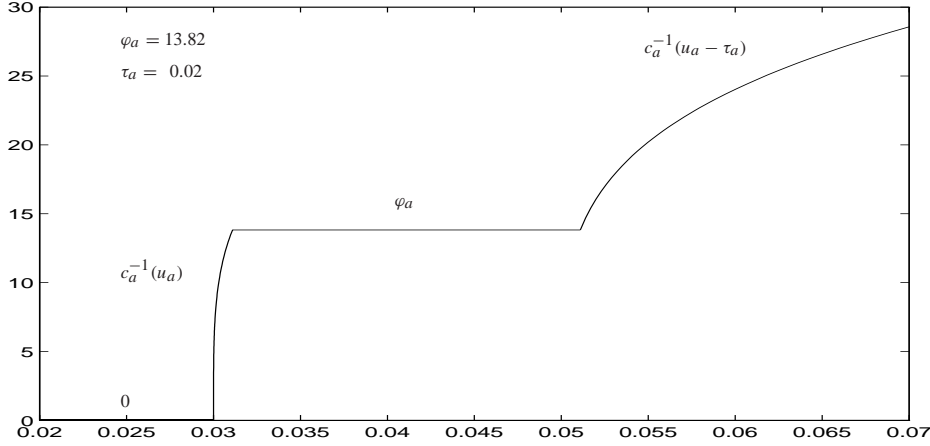
Fig. 2. The solution f_a to the single arc flow subproblem as a function of u_a

By weak duality, $\theta(\mathbf{u}) \leq z^*$ holds for all $\mathbf{u} \in \mathcal{R}^{|\mathcal{A}|}$. To formulate the Lagrangean dual program (cf., the program (5)) we consider an arbitrary point $\mathbf{u} \in \mathfrak{R}^{|\mathcal{A}|}$, and define

$$\widetilde{u}_a = \max\{u_a; \; c_a(0)\}, \qquad a \in \mathcal{A}.$$

Then, $f_a(\widetilde{u}_a) = f_a(u_a)$, so that $\theta_a(\widetilde{u}_a) = \theta_a(u_a)$. Further, $\theta_k(\widetilde{\mathbf{u}}) \geq \theta_k(\mathbf{u})$ since $\widetilde{\mathbf{u}} \geq \mathbf{u}$, and it follows that $\theta(\widetilde{\mathbf{u}}) \geq \theta(\mathbf{u})$. Since the dual objective is maximized on $\mathcal{R}^{|\mathcal{A}|}$, one can therefore, without loss of generality, impose the restrictions $u_a \geq c_a(0)$, $a \in \mathcal{A}$ (this is done by Larsson et al. [40] for the untolled program (30), with the same motivation). The Lagrange dual (5) may now be stated as

$$\begin{aligned} \max \quad & \theta(\mathbf{u}), \\ \text{s.t.} \quad & u_a \geq c_a(0), \qquad a \in \mathcal{A}, \end{aligned} \tag{38}$$

with solution set U^*. Due to the conclusions drawn in Section 2, this is a convex program. It has an interesting interpretation; whereas in the primal program (33) the equilibrium *arc flows* are sought, (38) is the problem of determining the equilibrium *arc travel times*.

The following proposition relates the primal and dual solutions.

Proposition 6 (Primal–dual optimality). *Let* $\mathbf{u}^* \in U^*$ *be arbitrary. Then, strong duality holds, that is,* $\theta(\mathbf{u}^*) = z^*$. *Further,* $f_a^* = f_a(u_a^*)$, $a \in \mathcal{A}$, *and*

$$H_k^* = H_k(\mathbf{u}^*) \bigcap \left\{ [h_{kr}]_{r \in \mathcal{R}_k} \; \middle| \; \sum_{\ell \in \mathcal{C}} \sum_{r \in \mathcal{R}_\ell} \delta_{\ell r a} h_{\ell r} = f_a^*, \, a \in \mathcal{A} \right\}, \quad k \in \mathcal{C}. \tag{39}$$

Proof. The strong duality follows from Theorem 6.2.4 in Bazaraa et al. [3]. (The application of that theorem requires that the program (33) has a feasible solution and that the inclusion $\mathbf{0} \in \mathrm{int}\{[f_a - \sum_{k \in \mathcal{K}} \sum_{r \in \mathcal{R}_k} \delta_{kra} h_{kr}]_{a \in \mathcal{A}} \mid ([f_a]_a, [h_{kr}]_{kr})$ satisfies (33b)–(33c)\}$ holds; the latter of these assumptions differs slightly from Assumption 1

and is clearly satisfied.) By the above and Theorem 28.3 in Rockafellar [55] (see also Patriksson [49, Theorem 2.1]), the conditions (29) and (32) are the first-order optimality conditions for the program (33). Further, the set of optimal solutions to (33) may be characterized as the set of Lagrangean subproblem solutions at $\mathbf{u}^*$ that also satisfy the constraints (33d) (Bazaraa et al. [3, Theorem 6.2.5]; cf. also the relation (7)). The uniqueness of $f_a(u_a)$, $u_a \in \mathfrak{R}$, yields that $f_a^* = f_a(u_a^*)$ for all $a \in \mathcal{A}$, and the expression (39) follows.

□

The proposition states that the optimal arc flow $[f_a^*]_{a \in \mathcal{A}}$ is obtained from the solutions to the subproblems (35) at $\mathbf{u}^* \in U^*$. However, an optimal route flow pattern $[h_{kr}^*]_{r \in \mathcal{R}_k} \in H_k^*$ is, in general, not directly available from the subproblem (34) even if an optimal dual solution is at hand. This is so because the set $\prod_{k \in \mathcal{C}} H_k(\mathbf{u}^*)$ is usually not a singleton, since the function θ_k is usually nonsmooth at $\mathbf{u}^*$. (For the case when $[f_a^*]_{a \in \mathcal{A}}$ is known, an algorithm for calculating a solution $[h_{kr}^*]_{r \in \mathcal{R}_k} \in H_k^*$ is given in Drissi-Kaïtouni [17].)

Proposition 7 (Equilibrium tolls). *Let* $\mathbf{u}^* \in U^*$ *be arbitrary and* $[f_a^*]_{a \in \mathcal{A}}$ *optimal in the program (33). Then, the equilibrium tolls satisfy*

$$p_a^* \begin{cases} = u_a^* - c_a(f_a^*), & \text{if } f_a^* > 0, \\ \in \pi_a(0), & \text{if } f_a^* = 0, \end{cases} \qquad a \in \mathcal{A}.$$

Proof. Since $[f_a^*]_{a \in \mathcal{A}}$ and $\mathbf{u}^*$ solves the primal–dual pair (33), (38), the generalized KKT conditions (Rockafellar [55, Theorem 28.3]) yield, for all $a \in \mathcal{A}$, that $p_a(f_a^*) \in \pi_a(f_a^*)$ holds, and that $u_a^* = c_a(f_a^*) + p_a(f_a^*)$ holds whenever $f_a^* > 0$.

□

4.3. The algorithm

The algorithm is based on the solution of the Lagrange dual (38) by the subgradient optimization method (9)–(11). A few comments regarding its implementation are needed. All calculations can be made in arc flows exclusively. By aggregating the feasible shortest route flow pattern $[h_{kr}(\mathbf{u}^t)]_{r \in \mathcal{R}_k, k \in \mathcal{C}}$ into a *feasible arc flow* solution

$$y_a^t = \sum_{k \in \mathcal{C}} \sum_{r \in \mathcal{R}_k} \delta_{kra} h_{kr}(\mathbf{u}^t), \qquad a \in \mathcal{A}, \tag{40}$$

a subgradient to θ at $\mathbf{u}^t$ (cf. the definition (37)) is defined by $y_a^t - f_a(u_a^t)$, $a \in \mathcal{A}$. The adaption of the standard subgradient algorithm (that is, the iteration formula (9) with $\mathbf{v}^t = \mathbf{0}$) to the program (38) is then given by

$$u_a^{t+\frac{1}{2}} = u_a^t + \alpha_t \left(y_a^t - f_a(u_a^t) \right), \quad u_a^{t+1} = \max \left\{ u_a^{t+\frac{1}{2}}; c_a(0) \right\}, \quad a \in \mathcal{A}, \ t = 0, 1, \ldots.$$

Remark 2 (Redundant subgradient projection). Every subgradient of θ at $\mathbf{u}^t$ defines a locally feasible direction in the program (38). This is so because if, for some $a \in \mathcal{A}$, $u_a^t = c_a(0)$, then the formula (36) yields $f_a(u_a^t) = 0$, implying that $y_a^t - f_a(u_a^t) \geq 0$. Therefore, when applied to the program (38), the subgradient projection (see Section 2.2) is equivalent to the standard subgradient method, in which $\{\mathbf{v}^t\} = \{\mathbf{0}\}$.

□

The evaluation of the objective function θ at $\mathbf{u}$ essentially requires the calculation of a shortest route pattern; in each iteration t, the value $\theta(\mathbf{u}^t)$ defines a *lower bound* on the optimal value z^* of (33).

Embedded in the subgradient scheme is the generation of feasible solutions to the primal problem (33) through the computation of convex combinations of shortest route flow patterns, according to the formula (21). Previous experience from applications of the method (9)–(11) has demonstrated that the ergodic convergence of the sequence of subproblem solutions defined by (20)–(21) is superior to that defined by (13) (see Larsson and Liu [39], Larsson et al. [40], and Petersson and Patriksson [50]). Therefore, we employ step lengths according to the formula (20) and define the ergodic sequence of subproblem solutions by (21). For illustration purposes, we however also show results from the application of the formula (13). As pointed out in Remark 1, the ergodic sequences may be initialized at any iteration $t_0 \geq 1$; previous experience also indicates that it is indeed preferable to choose a $t_0 \gg 1$ (see also Larsson et al. [43]).

When applying the formula (21), the ergodic sequences $\{\widehat{f}_a^t\}$, $a \in \mathcal{A}$, of feasible arc flow solutions are computed as[7]

$$\widehat{f}_a^{t_0} = y_a^{t_0-1}, \quad \widehat{f}_a^t = \frac{t-t_0}{t-t_0+1}\widehat{f}_a^{t-1} + \frac{1}{t-t_0+1}y_a^{t-1}, \quad t = t_0+1, t_0+2, \dots \quad (41)$$

(When applying (13), a corresponding formula is obtained.) It is thus not necessary to store all the route flow patterns $[h_{kr}(\mathbf{u}^t)]_{k,r,t}$. Denoting the objective function of the program (33) by $z(\mathbf{f})$, where $\mathbf{f} = [f_a]_{a \in \mathcal{A}}$, it follows that $z(\widehat{\mathbf{f}}^t)$, $t \geq t_0$, are upper bounds converging to z^*.

4.4. Numerical experiments

The proposed method was implemented in Fortran-77 on a Digital AlphaStation 2004/166 and tested on the Sioux Falls network (LeBlanc et al. [47]). This network has 24 nodes, 76 directed arcs, and 528 origin–destination pairs. We imposed positive tolls on 12 arcs in the network, namely $\tau_a = 0.02$ on arcs (16, 18) and (18, 16), $\tau_a = 0.05$ on arcs (4, 5), (5, 4), (11, 12), and (12, 11), $\tau_a = 0.06$ on arcs (9, 10), (10, 9), (15, 22), and (22, 15), and, finally, $\tau_a = 0.07$ on arcs (21, 24) and (24, 21). The flow breakpoint, φ_a, for each of these arcs was chosen as 90% of the flow on the arc at equilibrium for the original (untolled) problem (these values were computed by the DSD code of Larsson and Patriksson [41]).

After calibrations of the method, the following strategies and parameter values were chosen. The step lengths were generated according to the formula (20) with $\mu = M = 1/75$ and $b = 1$, that is, $\alpha_t = (75(t+1))^{-1}$. In order to receive the best primal convergence the sequence of step lengths used in the dual method should be slowly decreasing so that all of the optimal segments of the dual function θ are attained by the sequence $\{\mathbf{u}^t\}$ with appropriate frequencies. The proper choice of the sequence $\{\alpha_t\}$ is evidently problem-dependent. In the first test, the feasible arc flows were computed according to the formula (41) with $t_0 = 1$, that is, the averaging was initiated

---

[7] The reason for not defining the sequences $\{\widehat{f}_a^t\}$ by averages of the subproblem solutions $f_a(u_a^s)$, $s = t_0, \dots, t-1$, is that, whenever $\mathbf{u}^t \neq \mathbf{u}^*$, the subproblem solutions $[f_a(u_a^t)]_{a \in \mathcal{A}}$ typically do *not* define a feasible flow; the sequences $\{f_a(u_a^t)\}$ are, however, convergent to u^*.

at iteration 1. In the second test, the averaging was initiated after 49 iterations, that is, $t_0 = 50$. The shortest route calculations were made using a standard implementation of Dijkstra's algorithm.

For these two tests, lower and upper bounds were recorded for 100 and 50 iterations, respectively, and the bounds found up to iteration t, for $t = 1, \ldots, 100$, are shown in Figure 3. The upper bounds obtained when the averaging starts at iteration t_0 are defined by $\widehat{z}_t^{t_0} = \min_{s=t_0,\ldots,t}\{z(\widehat{\mathbf{f}}^s)\}$, $t \geq t_0$, where $\widehat{\mathbf{f}}^s$ is given by the formula (41). The lower bounds from the dual feasible solutions are defined by $\theta_t = \max_{s=0,\ldots,t}\{\theta(\mathbf{u}^s)\}$, $t \geq 0$. We define the relative difference between the upper and lower bounds after 100 iterations as $(\widehat{z}_{100}^{t_0} - \theta_{100})/\theta_{100}$. For the two values of t_0 (1 and 50) these relative differences were 0.656% and 0.171%, respectively. The final lower bound was $\theta_{100} = 42.7037$, and the final upper bounds were $\widehat{z}_{100}^1 = 42.9840$ and $\widehat{z}_{100}^{50} = 42.7766$, which exceed the optimal value $z^* = 42.7576 \pm 0.0006$ (this value was obtained by our code running for 10,000 iterations) with 0.529% and 0.0444%, respectively.
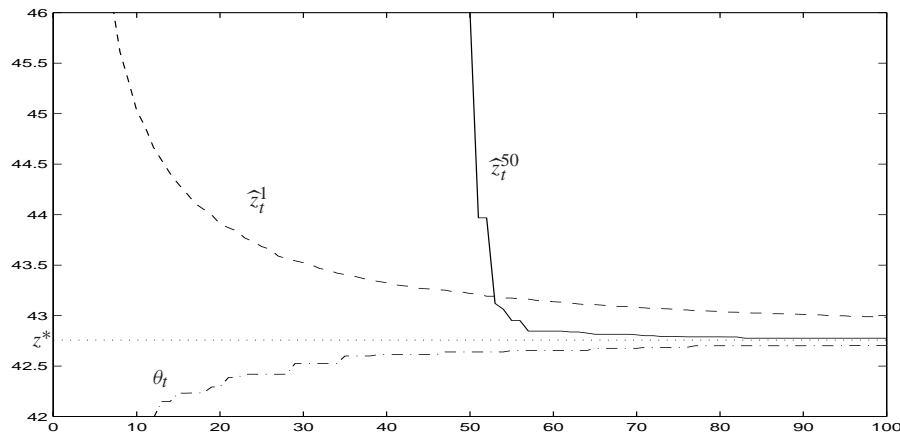

Fig. 3. The bounds on z^* found up to iteration t using the formula (21), for $t_0 = 1$ and 50, respectively

The curves in Figure 4 are generated similarly to those in Figure 3. The difference is that the convex combinations were here computed according to the formula (13), and that we made an additional test where the averaging scheme was initiated after 4 iterations, that is, $t_0 = 5$. The corresponding upper bounds are here denoted $\overline{z}_t^{t_0}$. For the three values of t_0 (that is, 1, 5 and 50) the relative differences between the bounds were 4.487%, 0.754% and 0.188%, respectively. The final upper bounds were $\overline{z}_{100}^1 = 44.6200$, $\overline{z}_{100}^5 = 43.0258$ and $\overline{z}_{100}^{50} = 42.7838$, which exceed the optimal value with 4.36%, 0.627%, and 0.0613%, respectively.

Concerning the difference in behaviour between the schemes using either the formula (13) or (21), consider the convergence of $\{\overline{z}_t^1\}$ and $\{\widehat{z}_t^1\}$, respectively. The initial all–or–nothing flow pattern $\mathbf{y}^0 = \overline{\mathbf{f}}^1 = \widehat{\mathbf{f}}^1$ (generated with link costs corresponding to zero flow) is composed of routes that are not used to such a large extent in the equilibrium solution, $\mathbf{f}^*$, since their travel costs are too high when the flow is large (in fact,
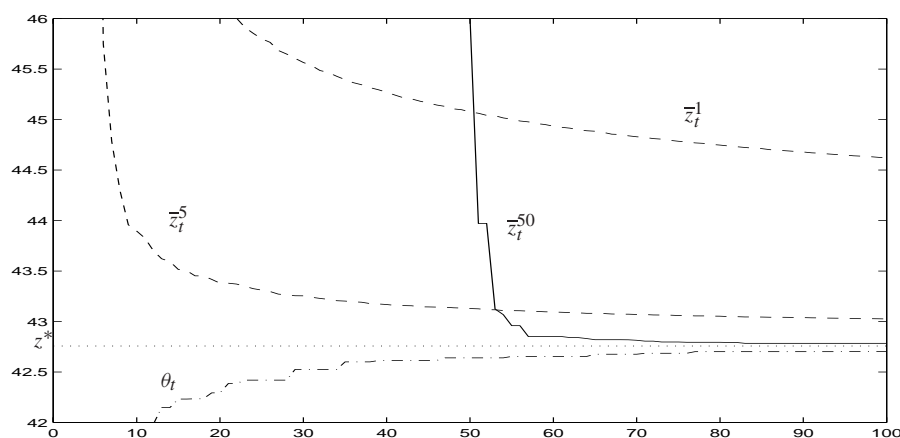
Fig. 4. The bounds on z^* found up to iteration t using the formula (13), for $t_0 = 1, 5$ and 50, respectively

$z(\mathbf{y}^0) = 160.29$). In the formula (13), however, this flow pattern receives a weight that is considerably larger than that of the subsequent patterns, and hence its weight in $\overline{\mathbf{f}}^t$ will be substantial even for large values of t. This explains the very poor convergence of $\{\overline{z}_t^1\}$, compared to $\{\widehat{z}_t^1\}$, where, according to the formula (21), the weight of $\mathbf{y}^0$ clearly becomes negligible rather quickly. From this discussion and the experiments shown in Figures 3 and 4 we conclude that the formula (21) combined with a relatively large value of t_0 yields the best results. Note that the delayed start of averaging is supported in theory by the result that $H_k(\mathbf{u}^t) \subseteq H_k(\mathbf{u}^\infty)$ holds for all t that are sufficiently large (cf. Proposition 4 and the discussion following the formulation (34)).

Defining $\mathbf{y}^t = [y_a^t]_{a \in \mathcal{A}}$, we compare the primal objective values obtained from the sequence $\{\mathbf{y}^t\}_{t=0}^{100}$ of feasible arc flows with those from the corresponding ergodic sequence $\{\widehat{\mathbf{f}}^t\}_{t=50}^{100}$. Figure 5 illustrates that the objective values from the ergodic sequence converges rapidly to the optimal value z^*, whereas the sequence of objective values from the all–or–nothing solutions $\mathbf{y}^t$ is clearly nonconvergent. Moreover, the time used for computing the ergodic sequence $\{\widehat{\mathbf{f}}^t\}$ is negligible (in fact, it amounts to less than one percent of the total computing time used by the procedure). Performing 100 iterations (including initializations and generation of output files) took 0.6 CPU-seconds on average.

We conclude that the proposed method, where the activation of the averaging procedure is delayed until the dual iterates are near-optimal, is a feasible approach to the traffic equilibrium assignment problem under road pricing. In particular, for $t_0 = 50$, the rate of convergence of the upper bound is very good once the averaging scheme has been activated.

The distance between a subproblem solution $(\mathbf{f}(\mathbf{u}^t), [h_{kr}(\mathbf{u}^t)]_{k,r})$ and the set defined by the constraints (33d) is proportional to the length of a subgradient $\boldsymbol{\gamma}^t = \mathbf{y}^t - \mathbf{f}(\mathbf{u}^t)$ of θ at $\mathbf{u}^t$. Analogously, $\mathbf{g}^t = \widehat{\mathbf{f}}^t - (t - 49)^{-1} \sum_{s=49}^{t-1} \mathbf{f}(\mathbf{u}^s)$ measures the distance between an averaged subproblem solution and the set defined by (33d). Figure 6 illustrates
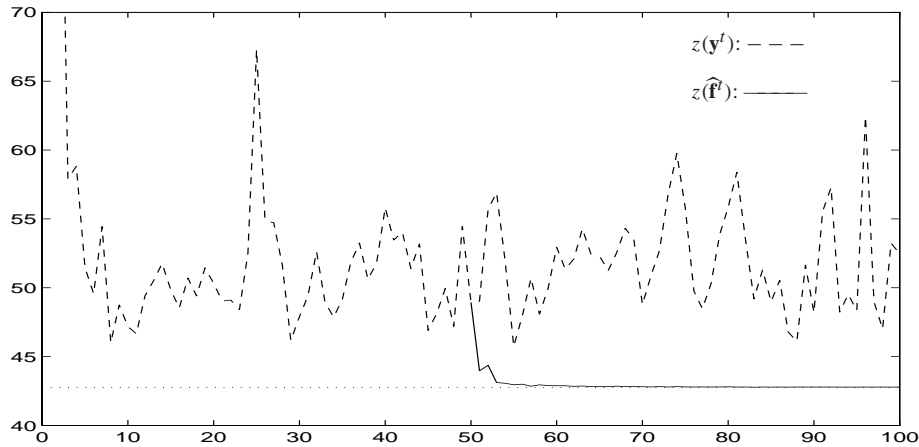
Fig. 5. Objective values with and without averaging the all–or–nothing flow patterns

that the sequence $\{\boldsymbol{\gamma}^t\}$ is nonconvergent, whereas the sequence $\{\mathbf{g}^t\}$ converges to zero (cf., Proposition 5), that is, the ergodic sequence of subproblem solutions converges to a feasible solution to (33).
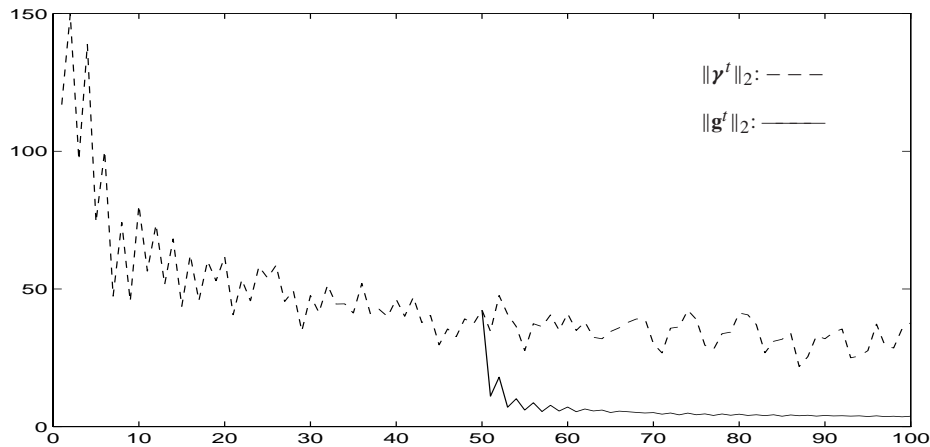

Fig. 6. A measure of the distance from subproblem solutions and ergodic subproblem solutions to the feasible set

We will now study two specific tolled arcs in the network; one with an equilibrium flow strictly greater than its flow breakpoint and one with a flow equal to its breakpoint, in order to study the convergence characteristics of the toll levels.

In Figure 7 we consider arc $a = (16, 18)$ with travel time function $c_a(f_a) = 0.03 + 0.00000003 \cdot f_a^4$, toll level $\tau_a = 0.02$, flow breakpoint $\varphi_a = 13.82$, and equilibrium flow $f_a^* \approx 14.39$ (that is, $f_a^* > \varphi_a$, and hence the equilibrium toll for this arc must be
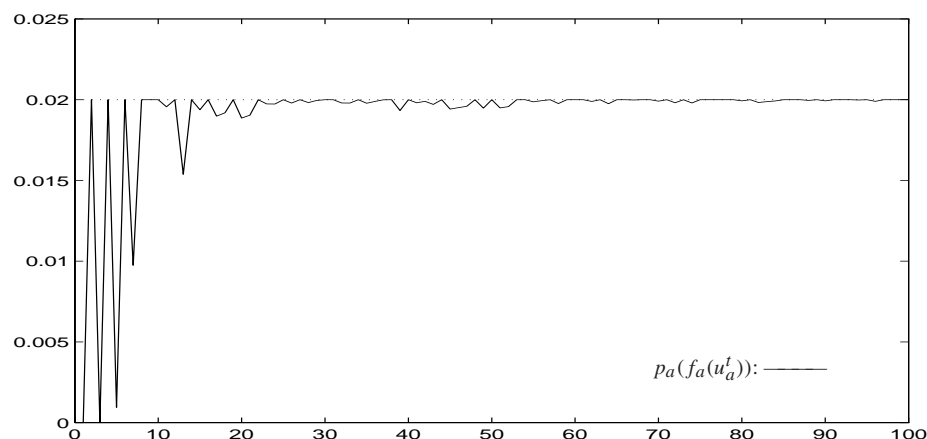
Fig. 7. The toll levels $u_a^t - c_a(f_a(u_a^t))$ for an arc with $f_a^* > \varphi_a$

$\pi_a(f_a^*) = \tau_a$; cf., the formula (31)). The graph shows the estimated tolls $p_a(f_a(u_a^t)) = u_a^t - c_a(f_a(u_a^t))$ for $t = 1, \ldots, 100$. The total arc cost at equilibrium is $u_a^* \approx 0.05129$; it is composed by the toll $p_a^* = 0.02$ and the travel time $c_a(f_a^*) \approx 0.03129$. Note that, by definition, $0.00 \leq p_a(f_a(u_a^t)) \leq 0.02$ for all t.
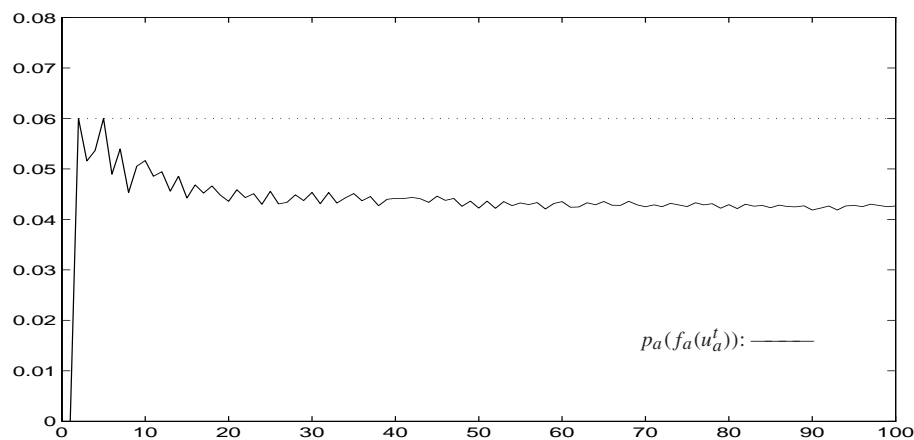

Fig. 8. The toll levels $u_a^t - c_a(f_a(u_a^t))$ for an arc with $f_a^* = \varphi_a$

The arc chosen for Figure 8 is $a = (9, 10)$ with travel time function $c_a(f_a) = 0.03 + 0.00000012 \cdot f_a^4$, toll level $\tau_a = 0.06$, flow breakpoint $\varphi_a = 19.53$, and equilibrium flow $f_a^* = \varphi_a$. The set of equilibrium tolls for this arc must clearly be a subinterval of $[0.00, 0.06]$, according to the formula (31); the actual value, $p_a(f_a(u_a^\infty))$, obtained in the limit from the application of the dual scheme depends on the initial dual solution and the step lengths chosen. The graph illustrates the estimated tolls $p_a(f_a(u_a^t)) =$

$u_a^t - c_a(f_a(u_a^t))$ for $t = 1, \ldots, 100$. Note that, by definition, $0.00 \le p_a(f_a(u_a^t)) \le 0.06$ for all t.

5. Conclusions and further research

The procedure analyzed in this paper generates an ergodic sequence of subproblem solutions within a Lagrangean dual subgradient scheme for the solution of a convex program. This ergodic sequence converges to the primal solution set, without requiring the solution of any coordinating master problem.

The method has been implemented for solving traffic equilibrium assignment problems under road pricing. The numerical experiments show that the solutions obtained from the ergodic sequence of subproblem solutions are of considerably better quality than those generated by the basic subgradient scheme.

A property of the dual subgradient method is that it can be initialized at any dual feasible solution. For the traffic assignment problem, an estimate of the equilibrium travel times may be used as a starting solution.

We remark that more general equilibrium models with pricing may be constructed and solved with the proposed methodology, along the lines presented in Larsson et al. [40] for the basic equilibrium model. In particular, the fixed demand model considered in Section 4.1 may be generalized to allow for elastic demands, that is, where each d_k is a (nonnegative and nonincreasing) function of the least route cost for origin–destination pair k.

Sherali and Choi [59] solve Lagrangean dual formulations of *linear programs* by subgradient methods, and establish the ergodic convergence of sequences of primal subproblem solutions; their analysis allows for more general choices of convexity weights and step lengths in the subgradient scheme than our analysis does. An interesting subject for further research is therefore the generalization of our results to these more general choices.

We have recently been able to extend the results of this paper to general convex–concave saddle point problems.

Further, it would be interesting to employ our method in applications such as, for example, the large linear programs with *soft* constraints studied in Hauer and Hoganson [29].

An extension of the results of this paper that would be of practical interest is to retain the ergodic convergence while allowing for an inexact solution of the subproblems; such solutions would provide ε_t-subgradients to the dual objective function. Convergence is indeed retained provided that $\Re_+ > \{\varepsilon_t\} \to 0$ and $\sum_{t=0}^{\infty} \alpha_t \varepsilon_t < \infty$.

We are currently investigating the application of the results of this paper in the field of discrete optimization. In such applications, the elements of an ergodic sequence in the limit solve a convexification of the original problem; this property can be exploited in various solution strategies.

References

1. Asmuth, R.L. (1978): Traffic Network Equilibria. PhD dissertation, Stanford University, Stanford, CA
2. Balakrishnan, A., Magnanti, T.L., Wong, R.T. (1989): A dual-ascent procedure for large-scale uncapacitated network design. Oper. Res. **37**, 716–740
3. Bazaraa, M.S., Sherali, H.D., Shetty, C.M. (1993): Nonlinear Programming. Theory and Algorithms, 2nd edition. Wiley, New York, NY
4. Beasley, J.E. (1993): Lagrangean relaxation. In: Reeves, C.R., ed., Modern Heuristic Techniques for Combinatorial Problems, pp. 243–303. Blackwell Scientific Publications, Oxford
5. Ben-Tal, A., Bendsøe, M.P. (1993): A new method for optimal truss topology design. SIAM J. Optim. **3**, 322–358
6. Bernstein, D., Smith, T.E. (1994): Equilibria for networks with lower semicontinuous costs: With an application to congestion pricing. Transp. Sci. **28**, 221–235
7. Bertsekas, D.P. (1982): Constrained Optimization and Lagrange Multiplier Methods. Academic Press, San Diego, CA
8. Bertsekas, D.P., Tsitsiklis, J.N. (1989): Parallel and Distributed Computation. Numerical Methods. Prentice–Hall, Englewood Cliffs, NJ
9. Bregman, L.M. (1967): The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Comp. Math. Math. Phys. **7**, 200–217
10. Burke, J.V., Ferris, M.C. (1991): Characterization of solution sets of convex programs. Oper. Res. Lett. **10**, 57–60
11. Cottle, R.W., Duvall, S.G., Zikan, K. (1986): A Lagrangean relaxation algorithm for the constrained matrix problem. Nav. Res. Logist. Quarterly **33**, 55–76
12. Dafermos, S., Sparrow, F.T. (1971): Optimal resource allocation and toll patterns in user-optimised transport networks. J. Transp. Econ. Policy **5**, 187–200
13. Dantzig, G.B., Wolfe, P. (1960): Decomposition principle for linear programs. Oper. Res. **8**, 101–111
14. Dem'yanov, V.F., Malozemov, V.N. (1974): Introduction to Minimax. Wiley, New York, NY
15. Dem'yanov, V.F., Vasil'ev, L.V. (1985): Nondifferentiable Optimization. Optimization Software, New York, NY
16. Dirickx, Y.M.I., Jennergren, L.P. (1979): Systems Analysis by Multilevel Methods. Wiley, Chichester
17. Drissi-Kaïtouni, O. (1990): An algorithm for the decomposition of arc flows into path flows for the general spatial price equilibrium problem. INFOR **28**, 403–411
18. Economist, The (1997): Jam today, road pricing tomorrow. December 6–12, pp. 21–24
19. Ermolev, Yu.M. (1966): Methods for solving nonlinear extremal problems. Cybernetics **2**(4), 1–14
20. Everett, H. (1963): Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. Oper. Res. **11**, 399–417
21. Falk, J.E. (1967): Lagrange multipliers and nonlinear programming. J. Math. Anal. Appl. **19**, 141–159
22. Feinberg, B. (1989): Coercion functions and decentralized linear programming. Math. Oper. Res. **14**, 177–187
23. Fisher, M.L. (1981): The Lagrangian relaxation method for solving integer programming problems. Manage. Sci. **27**, 1–18
24. Fisher, M.L. (1994): Optimal solution of vehicle routing problems using minimum k-trees. Oper. Res. **42**, 626–642
25. Geoffrion, A.M. (1971): Duality in nonlinear programming: A simplified applications-oriented development. SIAM Rev. **13**, 1–37
26. Geoffrion, A.M. (1974): Lagrangian relaxation for integer programming. Math. Program. Study **2**, 82–114
27. Goffin, J.L., Haurie, A., Vial, J.P. (1992): Decomposition and nondifferentiable optimization with the projective algorithm. Manage. Sci. **38**, 284–302
28. Goffin, J.-L., Haurie, A., Vial, J.-Ph., Zhu, D.L. (1993): Using central prices in the decomposition of linear programs. Eur. J. Oper. Res. **64**, 393–409
29. Hauer, G.K., Hoganson, H.M. (1996): Tailoring a decomposition method to a large forest management scheduling problem in northern Ontario. INFOR **34**, 209–231
30. Higle, J.L., Sen, S. (1991): Stochastic decomposition: An algorithm for two-stage linear programs with recourse. Math. Oper. Res. **16**, 650–669
31. Holmberg, K. (1992): Linear mean value cross decomposition: A generalization of the Kornai–Lipták method. Eur. J. Oper. Res. **62**, 55–73
32. Jennergren, P. (1973): A price schedules decomposition algorithm for linear programming problems. Econometrica **41**, 965–979
33. Johnson, L.A., Montgomery, D.S. (1974): Operations Research in Production Planning, Scheduling and Inventory Control. Wiley, New York, NY

34. Jones, K.L., Lustig, I.J., Farvolden, J.M., Powell, W.B. (1993): Multicommodity network flows: The impact of formulation on decomposition. Math. Program. **62**, 95–117
35. Kiwiel, K.C. (1995): Approximations in bundle methods and decomposition of convex programs. J. Optim. Theory Appl. **84**, 529–548
36. Knopp, K. (1956): Infinite Sequences and Series. Dover Publications, New York, NY
37. Kuhn, H.W., Tucker, A.W. (1951): Nonlinear programming. In: Neyman, J., ed., Proceedings of the second Berkeley symposium on mathematical statistics and probability, Department of Mathematics, University of California, July 31 – August 12, 1950, pp. 481–492. University of California Press, Berkeley, CA
38. Lamond, B., Stewart, N.F. (1981): Bregman's balancing method. Transp. Res. **15**B, 239–248
39. Larsson, T., Liu, Z. (1997): A Lagrangean relaxation scheme for structured linear programs with application to multicommodity network flows. Optimization **40**, 247–284
40. Larsson, T., Liu, Z., Patriksson, M. (1997): A dual scheme for traffic assignment problems. Optimization **42**, 323–358
41. Larsson, T., Patriksson, M. (1992): Simplicial decomposition with disaggregated representation for the traffic assignment problem. Transp. Sci. **26**, 4–17
42. Larsson, T., Patriksson, M., Strömberg, A.-B. (1996a): Conditional subgradient optimization – Theory and applications. Eur. J. Oper. Res. **88**, 382–403
43. Larsson, T., Patriksson, M., Strömberg, A.-B. (1996b): Ergodic results and bounds on the optimal value in subgradient optimization. In: Kleinschmidt, P. et al., eds., Operations Research Proceedings 1995, Selected papers of the Symposium on Operations Research, Passau, Germany, September, 1995, pp. 30–35. Springer, Berlin
44. Larsson, T., Patriksson, M., Strömberg, A.-B. (1998): Ergodic convergence in subgradient optimization. Optim. Methods Software **9**, 93–120
45. Lasdon, L.S. (1968): Duality and decomposition in mathematical programming. IEEE Trans. Systems Sci. Cybernet. **SSC4**, 86–100
46. Lasdon, L.S. (1970): Optimization Theory for Large Systems. Macmillan, London
47. LeBlanc, L.J., Morlok, E.K., Pierskalla, W.P (1975): An efficient approach to solving the road network equilibrium traffic assignment problem. Transp. Res. **9**, 309–318
48. Mangasarian, O.L. (1981): Iterative solution of linear programs. SIAM J. Numer. Anal. **18**, 606–614
49. Patriksson, M. (1994): The Traffic Assignment Problem: Models and Methods, Topics in Transportation series. VSP, Utrecht, The Netherlands
50. Petersson, J., Patriksson, M. (1997): Topology optimization of sheets in contact by a subgradient method. Int. J. Numer. Methods Eng. **40**, 1295–1321
51. Polyak, B.T. (1967): A general method of solving extremum problems. Sov. Math. Dokl. **8**(3), 593–597
52. Polyak, B.T. (1969): Minimization of unsmooth functionals. USSR Comp. Math. Math. Phys. **9**, 14–29
53. Powell, W.B., Sheffi, Y. (1982): The convergence of equilibrium algorithms with predetermined step sizes. Transp. Sci. **16**, 45–55
54. Robinson, S.M. (1984): Local structure of feasible sets in nonlinear programming, part II: Nondegeneracy. Math. Program. Study **22**, 217–230
55. Rockafellar, R.T. (1970): Convex Analysis. Princeton University Press, Princeton, NJ
56. Rockafellar, R.T. (1976): Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**, 97–116
57. Rudin, W. (1976): Principles of Mathematical Analysis, third edition. McGraw–Hill, Singapore
58. Ruszczyński, A. (1989): An augmented Lagrangean decomposition method for block diagonal linear programming problems. Oper. Res. Lett. **8**, 287–294
59. Sherali, H.D., Choi, G. (1996): Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs. Oper. Res. Lett. **19**, 105–113
60. Shor, N.Z. (1985): Minimization Methods for Non-Differentiable Functions. Springer, Berlin
61. Small, K.A. (1992): Using the revenues from congestion pricing. Transportation **19**, 359–381
62. Svanberg, K. (1982): An algorithm for optimum structural design using duality. Math. Program. Study **20**, 161–177
63. Uzawa, H. (1958): Iterative methods for concave programming. In: Arrow, K.J., Hurwicz, L., Uzawa, H., eds., Studies in Linear and Nonlinear Programming, pp. 154–165. Stanford University Press, Stanford, CA
64. Wardrop, J.G. (1952): Some theoretical aspects of road traffic research. In: Proceedings of the Institute of Civil Engineers, Part II, pp. 325–378