Primal convergence from dual subgradient methods for convex optimization

(article starts on next page)

Primal convergence from dual subgradient methods for convex optimization

Emil Gustavsson · Michael Patriksson · Ann-Brith Strömberg

Abstract When solving a convex optimization problem through a Lagrangian dual reformulation subgradient optimization methods are favorably utilized, since they often find near-optimal dual solutions quickly. However, an optimal primal solution is generally not obtained directly through such a subgradient approach unless the Lagrangian dual function is differentiable at an optimal solution. We construct a sequence of convex combinations of primal subproblem solutions, a so called ergodic sequence, which is shown to converge to an optimal primal solution when the convexity weights are appropriately chosen. We generalize previous convergence results from linear to convex optimization and present a new set of rules for constructing the convexity weights that define the ergodic sequence of primal solutions. In contrast to previously proposed rules, they exploit more information from later subproblem solutions than from earlier ones. We evaluate the proposed rules on a set of nonlinear multicommodity flow problems and demonstrate that they clearly outperform the ones previously proposed.

E. Gustavsson (✉) · M. Patriksson · A.-B. Strömberg
Department of Mathematical Sciences, Chalmers University of Technology,
412 96 Göteborg, Sweden
e-mail: emilg@chalmers.se

E. Gustavsson · M. Patriksson · A.-B. Strömberg
Department of Mathematical Sciences, University of Gothenburg,
412 96 Göteborg, Sweden

1 Introduction and motivation

Lagrangian relaxation is a frequently utilized tool for solving large-scale convex minimization problems due to its simplicity and its property of systematically providing optimistic estimates on the optimal value. One popular tool for solving the dual problems of Lagrangian relaxation schemes is subgradient optimization. The advantage of subgradient methods is that they often find near-optimal dual solutions quickly, whilst a drawback is that near-optimal primal feasible solutions can not, in general, be obtained directly from the subgradient scheme. As the dual iterates in a subgradient scheme converge towards an optimal dual solution, primal convergence towards a near-optimal primal solution is not in general achieved by simply using the subproblem solutions as primal iterates. Even with a dual optimal solution at hand, an optimal primal solution can not easily be obtained. The reason for this inconvenience is that the dual objective function is typically nonsmooth at an optimal point, whence an optimal primal solution is a nontrivial convex combination of the extreme subproblem solutions.

This paper analyzes what is called *ergodic sequences* by Larsson et al. [30] or *recovering primal solutions* by Sherali and Choi [42]; we will use the notion ergodic sequences. To guarantee primal convergence for a linear program in a subgradient scheme, Shor [44, Chapter 4] and Larsson and Liu [27] (originally developed in [26]) utilize a strategy which, rather than using the subproblem solution as primal iterate, uses a convex combination of previously found subproblem solutions, denoted as an *ergodic sequence*. In [44, Chapter 4] the convex combinations are determined by the step lengths used in the subgradient scheme, while in [27, Theorem 3] the mean of the iterates previously found are used. These results are extended in [42] to a more general case of convex combinations and step lengths in the subgradient algorithm applied to linear programs. Larsson et al. [30] show that the convex combinations used in [44] and [27] yield primal convergence also for general convex optimization problems.

Several other methods for generating approximate primal solutions in a subgradient scheme have been studied. Barahona and Anbil [7] propose a method for approximating the solution to a linear program by utilizing a subgradient method in which a primal solution is created as a convex combination of the previous solution and the primal iterate obtained from the subgradient method. The method is denoted the *volume algorithm* and was revisited by Bahiense et al. [4] and Sherali and Lim [43], where they extended it to include more information in the dual scheme. Nesterov [34] analyzes a primal–dual subgradient method where a primal feasible approximation to the optimum is obtained by using control sequences in both the primal and dual space. Nedić and Ozdaglar [32,33] study methods which utilize the average of all previously found iterates as primal solutions. The latter algorithms employ a constant step length due to its simplicity and practical significance. For a more thorough overview of the history of strategies for the construction of primal iterates in dual subgradient schemes, see Anstreicher and Wolsey [1].

This paper generalizes the results in [42] to the class of convex programs, and extends the results in [30] to include more general convex combinations in the definition of the ergodic sequences. We present a new set of rules for constructing the convexity weights defining the ergodic sequence of primal iterates. In contrast to rules

previously utilized, they exploit more information from later subproblem solutions than from earlier ones. We evaluate the new rules on a set of nonlinear multicommodity flow problems (NMFPs) and show that they clearly outperform the previously utilized ones.

The remainder of this paper is organized as follows. In Sect. 2 we introduce some basic concepts regarding Lagrangian relaxation and subgradient methods. In Sect. 3, we describe the notion of primal ergodic sequences and present an important theorem regarding their convergence when considering general convex problems. Section 3 also includes a new set of rules for choosing convexity weights when defining the ergodic sequences. The final part of Sect. 3 includes a taxonomy of previous results and their connection to the results presented in this paper. In Sect. 4 we introduce the NMFP and describe a solution approach based on Lagrangian relaxation. Computational results for a set of NMFP test instances employing the new rules for choosing the convexity weights are presented in Sect. 5. Conclusions are then drawn in Sect. 6.

2 Background

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^n \to \mathbb{R}, i \in \mathcal{I} := \{1, \ldots, m\}$, be convex and (possibly) nonsmooth functions and the set $X \subset \mathbb{R}^n$ be convex and compact. Consider the program

$$f^* := \text{minimum} \quad f(\mathbf{x}), \tag{1a}$$

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

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

with solution set $X^* \subset \mathbb{R}^n$. We assume that the set X is simple and that the feasible set $\{\mathbf{x} \in X \mid h_i(\mathbf{x}) \leq 0, \ i \in \mathcal{I}\}$ is nonempty, implying that the solution set X^* is also nonempty. We define the Lagrange function $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ with respect to the relaxation of the constraints (1b) as $\mathcal{L}(\mathbf{x}, \mathbf{u}) := f(\mathbf{x}) + \mathbf{u}^T \mathbf{h}(\mathbf{x})$ for $(\mathbf{x}, \mathbf{u}) \in \mathbb{R}^n \times \mathbb{R}^m$. The dual objective function $\theta : \mathbb{R}^m \to \mathbb{R}$ is defined by

$$\theta(\mathbf{u}) := \min_{\mathbf{x} \in X} \left[f(\mathbf{x}) + \mathbf{u}^T \mathbf{h}(\mathbf{x}) \right], \quad \mathbf{u} \in \mathbb{R}^m. \tag{2}$$

The set X is compact which implies that θ is continuous [8, Theorem 6.3.1] on $\mathbb{R}^m$. The function θ is also concave on $\mathbb{R}^m$ and the nonempty and compact solution set for the subproblem in (2) at $\mathbf{u} \in \mathbb{R}^m$ is

$$X(\mathbf{u}) := \left\{ \mathbf{x} \in X \mid f(\mathbf{x}) + \mathbf{u}^T \mathbf{h}(\mathbf{x}) \leq \theta(\mathbf{u}) \right\}. \tag{3}$$

For $\mathbf{u} \in \mathbb{R}_+^m$, the set $X(\mathbf{u})$ is also convex. The Lagrangian dual problem is defined as

$$\theta^* := \underset{\mathbf{u} \in \mathbb{R}_+^m}{\text{supremum}} \ \theta(\mathbf{u}), \tag{4}$$

whose solution set $U^* \subseteq \mathbb{R}_+^m$ is convex. By weak duality, the inequality $\theta(\mathbf{u}) \leq f(\mathbf{x})$ holds for all $\mathbf{u} \in \mathbb{R}_+^m$ and all $\mathbf{x} \in X$ such that $\mathbf{h}(\mathbf{x}) \leq \mathbf{0}^m$ [8, Theorem 6.2.1].

Let S be a nonempty, closed, and convex set. We define the projection and distance operators by

$$\text{proj}(\mathbf{x}, S) := \underset{\mathbf{y} \in S}{\text{argmin}} \|\mathbf{y} - \mathbf{x}\|_2 \quad \text{and} \quad \text{dist}(\mathbf{x}, S) := \min_{\mathbf{y} \in S} \|\mathbf{y} - \mathbf{x}\|_2. \tag{5}$$

Note that the function $\text{dist}(\cdot, S)$ is convex and continuous. We denote by *closed map* a point-to-set map $X : \mathbb{R}^m \to 2^{\mathbb{R}^n}$ such that when $\{\mathbf{u}^t\} \subset \mathbb{R}^m$, $\{\mathbf{u}^t\} \to \mathbf{u}$, $\mathbf{x}^t \in X(\mathbf{u}^t)$ for all t, and $\{\mathbf{x}^t\} \to \mathbf{x}$, this implies $\mathbf{x} \in X(\mathbf{u})$. The following result states that the point-to-set map $X(\cdot)$ defined in (3) is a closed map.

Lemma 1 ($X(\cdot)$ is a closed map [30, Lemma 1]) *Let the sequence $\{\mathbf{u}^t\} \subset \mathbb{R}^m$, the map $X(\cdot) : \mathbb{R}^m \to 2^X$ be given by the definition* (3), *and the sequence $\{\mathbf{x}^t\}$ be given by the inclusion $\mathbf{x}^t \in X(\mathbf{u}^t)$. If $\{\mathbf{u}^t\} \to \mathbf{u} \in \mathbb{R}^m$, then $\text{dist}(\mathbf{x}^t, X(\mathbf{u})) \to 0$. If, in addition, $X(\mathbf{u}) = \{\mathbf{x}\}$, then $\{\mathbf{x}^t\} \to \mathbf{x}$.*

For each $\mathbf{u} \in \mathbb{R}^m$, we define the set of indices corresponding to strictly positive multiplier values as

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

Lemma 2 (affineness of the Lagrange function [30, Lemma 2]) *The functions f and h_i, $i \in \mathcal{I}(\mathbf{u})$, are affine on $X(\mathbf{u})$ for every $\mathbf{u} \in \mathbb{R}_+^m$. Further, if the function f is (the functions h_i, $i \in \mathcal{I}(\mathbf{u})$, are) differentiable, then ∇f is (∇h_i, $i \in \mathcal{I}(\mathbf{u})$ are) constant on $X(\mathbf{u})$.*

From Lemma 2 follows that for all $\mathbf{u} \in \mathbb{R}_+^m$ and every $i \in \mathcal{I}(\mathbf{u})$, ∂h_i is constant on rint $X(\mathbf{u})$; hence for every $\overline{\mathbf{x}} \in$ rint $X(\mathbf{u})$, each subgradient $\xi \in \partial h_i(\overline{\mathbf{x}})$ defines a hyperplane that supports the function h_i at every $\mathbf{x} \in$ rint $X(\mathbf{u})$. We define the *subdifferential* of θ at $\mathbf{u} \in \mathbb{R}^m$ as the set

$$\partial\theta(\mathbf{u}) := \left\{ \boldsymbol{\gamma} \in \mathbb{R}^m \mid \theta(\mathbf{v}) \leq \theta(\mathbf{u}) + \boldsymbol{\gamma}^T(\mathbf{v} - \mathbf{u}), \quad \mathbf{v} \in \mathbb{R}^m \right\}.$$

Proposition 1 (subdifferential to the dual function [30, Proposition 1]) *For each $\mathbf{u} \in \mathbb{R}^m$, it holds that $\partial\theta(\mathbf{u}) = \{\mathbf{h}(\mathbf{x}) \mid \mathbf{x} \in X(\mathbf{u})\}$. 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 all $\mathbf{x} \in X(\mathbf{u})$.*

To obtain primal–dual optimality relations, we assume *Slater's constraint qualification* as stated in Assumption 1.

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

Under Assumption 1, the solution set U^* is nonempty and compact and, by strong duality, the equality $\theta(\mathbf{u}^*) = f(\mathbf{x}^*)$ holds for some pair of primal–dual solutions $(\mathbf{x}^*, \mathbf{u}^*)$ fulfilling $\mathbf{u}^* \in \mathbb{R}_+^m$, $\mathbf{x}^* \in X$ and $\mathbf{h}(\mathbf{x}^*) \leq \mathbf{0}^m$ ([8, Theorem 6.2.5]).

Proposition 2 (optimality conditions, [8, Theorem 6.2.5]) *Let Assumption 1 hold. Then, $\mathbf{u} \in U^*$ and $\mathbf{x} \in X^*$ if and only if $\mathbf{u} \in \mathbb{R}_+^m$, $\mathbf{x} \in X(\mathbf{u})$, $\mathbf{h}(\mathbf{x}) \leq \mathbf{0}^m$ and $\mathbf{u}^T\mathbf{h}(\mathbf{x}) = 0$.*

2.1 Subgradient optimization

We consider solving the Lagrangian dual program by the *subgradient optimization method*. We start at some $\mathbf{u}^0 \in \mathbb{R}_+^m$ and compute iterates $\mathbf{u}^t$ according to

$$\mathbf{u}^{t+1} = \left[\mathbf{u}^t + \alpha_t \mathbf{h}(\mathbf{x}^t)\right]_+, \quad t = 0, 1, \ldots, \tag{6}$$

where $\mathbf{x}^t \in X(\mathbf{u}^t)$ solves the subproblem defined in (2) at $\mathbf{u}^t$, implying that $\mathbf{h}(\mathbf{x}^t) \in \partial\theta(\mathbf{u}^t)$, $\alpha_t > 0$ is the step length chosen at iteration t, and $[\cdot]_+$ denotes the Euclidian projection onto the nonnegative orthant $\mathbb{R}_+^m$. For some early development of the theory of the subgradient optimization method, see Shor [44, Chapter 2], Polyak [37,38], and Ermol'ev [15]. The convergence of the method (6) for the special case of a divergent series step length rule is established in the following proposition.

Proposition 3 (convergence of dual iterates, [1, Theorem 3]) *Suppose that Assumption 1 holds, and let the method* (6) *be applied to the program* (4)*, with the step lengths α_t fulfilling the conditions*

$$\alpha_t > 0, \ \forall t, \quad \lim_{t \to \infty} \sum_{s=0}^{t-1} \alpha_s = \infty, \quad \lim_{t \to \infty} \sum_{s=0}^{t-1} \alpha_s^2 < \infty. \tag{7}$$

Then $\{\mathbf{u}^t\} \to \mathbf{u}^\infty \in U^*$ and $\{\theta(\mathbf{u}^t)\} \to \theta^*$.

3 Ergodic primal convergence

In this section, we introduce the notion of an *ergodic sequence* and present two important results regarding the convergence of ergodic sequences depending on convexity weights and step lengths. Assume that the method (6) is applied to the problem (4). At each iteration t, an ergodic primal iterate $\bar{\mathbf{x}}^t$ is composed according to

$$\bar{\mathbf{x}}^t = \sum_{s=0}^{t-1} \mu_s^t \mathbf{x}^s, \quad \sum_{s=0}^{t-1} \mu_s^t = 1, \quad \mu_s^t \geq 0, \quad s = 0, \ldots, t-1, \tag{8}$$

where $\mathbf{x}^s$ is the primal solution found in iteration s, i.e., the solution to the subproblem defined in (2). The vector $\bar{\mathbf{x}}^t$ thus is a convex combination of all previous subproblem solutions. We define

$$\gamma_s^t = \mu_s^t / \alpha_s, \quad s = 0, \ldots, t-1, \quad t = 1, 2, \ldots, \tag{9a}$$

and

$$\Delta\gamma_{\max}^t = \max_{s \in \{1,\ldots,t-1\}} \{\gamma_s^t - \gamma_{s-1}^t\}, \quad t = 2, 3, \ldots. \tag{9b}$$

Assumption 2 (relations between convexity weights and step lengths) The step lengths α_t and the convexity weights μ_s^t are chosen such that the following conditions are satisfied:

A1: $\gamma_s^t \geq \gamma_{s-1}^t$, $s = 1, \ldots, t-1$, $t = 2, 3, \ldots$,

A2: $\Delta\gamma_{\max}^t \to 0$ as $t \to \infty$, and

A3: $\gamma_0^t \to 0$ as $t \to \infty$ and, for some $\Gamma > 0$, $\gamma_{t-1}^t \leq \Gamma$ for all t.

The condition A1 requires that $\mu_s^t/\mu_{s-1}^t \geq \alpha_s/\alpha_{s-1}$, $s = 1, \ldots, t-1, t = 1, 2, \ldots$. This can be interpreted as the requirement that whenever the step length at iteration s (α_s) is larger than the previous one at iteration $s-1$ (α_{s-1}), the corresponding convexity weight (μ_s^t) should be larger than the previous one (μ_{s-1}^t). By condition A2, the difference between each pair of subsequent convexity weights tends to zero as t increases, meaning that no primal iterate should be completely neglected. Condition A3 assures that, for decreasing step lengths, the convexity weights decrease at a rate not slower than that of the step lengths.

Remark 1 For any fixed value of $s \in \{0, \ldots, t-1\}$, it follows from Assumption 2 that $\gamma_s^t \leq \gamma_0^t + s\Delta\gamma_{\max}^t \to 0$ as $t \to \infty$. This implies that $\gamma_s^t = \mu_s^t/\alpha_s \to 0$, which yields that $\mu_s^t \to 0$ as $t \to \infty$, since $0 < \alpha_s < \infty$. □

One example of convexity weights and step lengths fulfilling Assumption 2 is when each ergodic iterate equals the average of all previously found subproblem solutions, i.e., $\mu_s^t = 1/t$, $s = 0, \ldots, t-1$, $t = 1, 2, \ldots$, and the step lengths are chosen according to a harmonic series, i.e., $\alpha_t = a/(b+ct)$, $t = 0, 1, \ldots$, where $a, b, c > 0$. Note that in [42, Theorem 1], Assumption 2 is included in the hypothesis.

We now present a special case of a result of Silverman and Toeplitz (proven in [25]) which will be utilized in the analysis to follow.

Lemma 3 (convergence of convex combinations, [25, p. 35]) *Assume that the sequence $\{\mu_s^t\} \subset \mathbb{R}$ fulfills the conditions*

$$\mu_s^t \geq 0, \ s = 0, \ldots, t-1, \ \sum_{s=0}^{t-1} \mu_s^t = 1, \ t = 1, 2, \ldots, \ and$$

$$\lim_{t \to \infty} \mu_s^t = 0, \ s = 0, 1, \ldots.$$

If the sequence $\{\mathbf{b}^s\} \subset \mathbb{R}^r$, $r \geq 1$, is such that $\lim_{s\to\infty} \mathbf{b}^s = \mathbf{b}$ holds, then it holds that $\lim_{t\to\infty}\left(\sum_{s=0}^{t-1}\mu_s^t\mathbf{b}^s\right) = \mathbf{b}$.

3.1 Feasibility in the limit

We here show that, assuming convergence towards a dual feasible point in the subgradient method (6), and that the step lengths, α_t, and convexity weights, μ_s^t, are chosen such that Assumption 2 is fulfilled, the ergodic sequence of iterates, $\bar{\mathbf{x}}^t$, converges to the set of primal feasible solutions.

Proposition 4 (feasibility of $\bar{\mathbf{x}}^t$ in the limit) *Suppose that the method (6) operated with a suitable step length rule yields $\{\mathbf{u}^t\} \to \mathbf{u}^\infty \in \mathbb{R}_+^m$. If the step lengths α_t and*

convexity weights μ_s^t fulfill Assumption 2, then the sequence $\{\bar{\mathbf{x}}^t\}$ generated according to (8) fulfills

$$\limsup_{t \to \infty} \mathbf{h}(\bar{\mathbf{x}}^t) \leq \mathbf{0}^m, \quad and \quad \bar{\mathbf{x}}^t \in X, \ t = 1, 2, \ldots.$$

Proof For all $t \geq 2$, we have that

$$\mathbf{h}(\bar{\mathbf{x}}^t) \leq \sum_{s=0}^{t-1} \mu_s^t \mathbf{h}(\mathbf{x}^s) \leq \sum_{s=0}^{t-1} \mu_s^t \frac{1}{\alpha_s} \left(\mathbf{u}^{s+1} - \mathbf{u}^s \right) = \sum_{s=0}^{t-1} \gamma_s^t \left(\mathbf{u}^{s+1} - \mathbf{u}^s \right) \tag{10a}$$

$$= -\gamma_0^t \mathbf{u}^0 - \sum_{s=1}^{t-1} \left(\gamma_s^t - \gamma_{s-1}^t \right) \mathbf{u}^s + \gamma_{t-1}^t \mathbf{u}^t \tag{10b}$$

$$= -\gamma_0^t \mathbf{u}^0 + \gamma_{t-1}^t \mathbf{u}^t - \mathbf{u}^\infty \sum_{s=1}^{t-1} \left(\gamma_s^t - \gamma_{s-1}^t \right) + \sum_{s=1}^{t-1} \left(\gamma_s^t - \gamma_{s-1}^t \right) \left(\mathbf{u}^\infty - \mathbf{u}^s \right) \tag{10c}$$

$$= \gamma_0^t \left(\mathbf{u}^\infty - \mathbf{u}^0 \right) + \gamma_{t-1}^t (\mathbf{u}^t - \mathbf{u}^\infty) + \sum_{s=1}^{t-1} (\gamma_s^t - \gamma_{s-1}^t) \left(\mathbf{u}^\infty - \mathbf{u}^s \right), \tag{10d}$$

where the inequalities in (10a) follow from the convexity of the function $\mathbf{h}$ and the iteration formula (6), respectively. By the condition A3 in Assumption 2, the first term in (10d) tends to $\mathbf{0}^m$ as $t \to \infty$. From the hypothesis, $\{\mathbf{u}^t\} \to \mathbf{u}^\infty$ and by condition A3 in Assumption 2, $\gamma_{t-1}^t \leq \Gamma$ holds, implying that the second term in (10d) tends to $\mathbf{0}^m$ as $t \to \infty$. Let $\sigma_t = \sum_{s=1}^{t-1} (\gamma_s^t - \gamma_{s-1}^t)(\mathbf{u}^\infty - \mathbf{u}^s)$. We need to show that $\sigma_t \to \mathbf{0}^m$ as $t \to \infty$. Given any $\varepsilon > 0$, let $\kappa \geq 1$ be large enough so that $||\mathbf{u}^\infty - \mathbf{u}^s|| \leq \varepsilon/2\Gamma$ holds for all $s \geq \kappa + 1$. Then, for $t \geq \kappa + 2$ and large enough so that $\Delta\gamma_{\max}^t \sum_{s=1}^{\kappa} ||\mathbf{u}^\infty - \mathbf{u}^s|| \leq \varepsilon/2$ holds, we have that

$$||\sigma_t|| \leq \sum_{s=1}^{t-1} \left(\gamma_s^t - \gamma_{s-1}^t \right) ||\mathbf{u}^\infty - \mathbf{u}^s|| \tag{11a}$$

$$\leq \Delta\gamma_{\max}^t \sum_{s=1}^{\kappa} ||\mathbf{u}^\infty - \mathbf{u}^s|| + \frac{\varepsilon}{2\Gamma} \sum_{s=\kappa+1}^{t-1} \left(\gamma_s^t - \gamma_{s-1}^t \right) \tag{11b}$$

$$\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2\Gamma} \left(\gamma_{t-1}^t - \gamma_\kappa^t \right) \tag{11c}$$

$$\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \tag{11d}$$

where the inequality (11a) follows from the triangle inequality and condition A1 in Assumption 2, and the inequality (11b) from condition A2 in Assumption 2 and the assumption that $\kappa \geq 1$ is large enough. The inequality (11c) follows from the assumption that $t \geq \kappa + 2$ is large enough, and the inequality (11d) follows from the

condition A3 in Assumption 2 and the fact that $\gamma_\kappa^t \geq 0$. Since $\varepsilon > 0$ is arbitrary, we deduce that $\boldsymbol{\sigma}_t \to \mathbf{0}^m$ as $t \to \infty$. It follows that

$$\limsup_{t\to\infty} h_i(\bar{\mathbf{x}}^t) \leq \mathbf{0}^m, \quad i \in \mathcal{I}.$$

Furthermore, since X is convex and $\bar{\mathbf{x}}^t$ is a convex combination of $\mathbf{x}^s \in X(\mathbf{u}^s) \subseteq X$, it holds that $\bar{\mathbf{x}}^t \in X$ for all t.   □

Proposition 4 states that as long as the sequence $\{\mathbf{u}^t\}$ of dual iterates converges to some feasible point in the Lagrangian dual problem (4), and the step lengths and convexity weights are appropriately chosen, the corresponding sequence of primal iterates defined by (8) will produce a primal feasible solution in the limit. If there is an accumulation point $\bar{\mathbf{x}}$ such that $\{\bar{\mathbf{x}}^t\} \to \bar{\mathbf{x}}$, then Proposition 4 states that $\bar{\mathbf{x}}$ is feasible in the original problem (1). If the functions f and h_i, $i \in \mathcal{I}$, are affine, and the set X is a polytope, then Proposition 4 reduces to [42, Theorem 1].

Note that the conditions A2 and A3 of Assumption 2 are fulfilled if condition A1 in Assumption 2 holds together with the condition that $\gamma_{t-1}^t \to 0$ as $t \to \infty$. Below, we present a result for strengthened assumptions on the convexity weights and step lengths, but where the sequence $\{\mathbf{u}^t\}$ is only assumed to be bounded.

Corollary 1 (bounded dual sequence) *Suppose that the sequence $\{\mathbf{u}^t\}$ generated by the formula* (6) *is bounded, and that condition* A1 *of Assumption* 2 *holds along with the condition that $\gamma_{t-1}^t \to 0$ as $t \to \infty$. Then, the sequence $\{\bar{\mathbf{x}}^t\}$ generated by* (8) *fulfills*

$$\limsup_{t\to\infty} \mathbf{h}(\bar{\mathbf{x}}^t) \leq \mathbf{0}^m, \quad and \quad \bar{\mathbf{x}}^t \in X, \quad t = 1, 2, \ldots.$$

Proof From the relations (10a)–(10b) and the condition A1 of Assumption 2 follows that $\mathbf{h}(\bar{\mathbf{x}}^t) \leq \gamma_{t-1}^t \mathbf{u}^t$, $t \geq 2$. Since $\gamma_{t-1}^t \to 0$ and $\{\mathbf{u}^t\}$ is bounded, $\limsup_{t\to\infty} \mathbf{h}(\bar{\mathbf{x}}^t) \leq \mathbf{0}^m$ holds.   □

Note that, under the assumptions of Corollary 1, any accumulation point $\bar{\mathbf{x}}$ to the sequence $\{\bar{\mathbf{x}}^t\}$ is feasible in (1).

3.2 Optimality in the limit

We next establish—assuming that Slater's constraint qualification (Assumption 1) is fulfilled—primal convergence to the set of optimal solutions X^* of the problem (1) as long as the step lengths, α_t, and the convexity weights, μ_s^t, are chosen to satisfy Assumption 2.

Theorem 1 (optimality of $\bar{\mathbf{x}}^t$ in the limit) *Suppose Assumption* 1 *holds and that the subgradient method* (6) *operated with a suitable step length rule attains dual convergence, i.e., $\{\mathbf{u}^t\} \to \mathbf{u}^\infty \in \mathbb{R}_+^m$, and let $\bar{\mathbf{x}}^t$ be generated according to* (8). *If the step lengths α_t and the convexity weights μ_s^t satisfy Assumption* 2, *then*

$$\mathbf{u}^\infty \in U^* \quad and \quad \mathrm{dist}(\bar{\mathbf{x}}^t, X^*) \to 0.$$

Proof From Proposition 4 follows that $\limsup_{t\to\infty} \mathbf{h}(\overline{\mathbf{x}}^t) \leq \mathbf{0}^m$ and $\overline{\mathbf{x}}^t \in X$, $t \geq 1$. In view of Proposition 2, it suffices to show that $\{\mathrm{dist}(\overline{\mathbf{x}}^t, X(\mathbf{u}^\infty))\} \to 0$ and that $\{\mathbf{h}(\overline{\mathbf{x}}^t)^T \mathbf{u}^\infty\} \to 0$ as $t \to \infty$.

By the convexity and nonnegativity of the function $\mathrm{dist}(\cdot, S)$, and the definition (8), the inequalities

$$0 \leq \mathrm{dist}\left(\overline{\mathbf{x}}^t, X(\mathbf{u}^\infty)\right) \leq \sum_{s=0}^{t-1} \mu_s^t \, \mathrm{dist}\left(\mathbf{x}^s, X(\mathbf{u}^\infty)\right), \quad t = 1, 2, \ldots,$$

hold. By Lemma 1, $\{\mathrm{dist}(\mathbf{x}^s, X(\mathbf{u}^\infty))\} \to 0$ as $s \to \infty$. Utilizing Remark 1 and Lemma 3 with $\mathbf{b}^s = \mathrm{dist}(\mathbf{x}^s, X(\mathbf{u}^\infty))$ and $\mathbf{b} = 0$, it follows that $\left\{\mathrm{dist}(\overline{\mathbf{x}}^t, X(\mathbf{u}^\infty))\right\} \to 0$ as $t \to \infty$.

Whenever $\mathcal{I}(\mathbf{u}^\infty) = \emptyset$, the equation $\mathbf{h}(\overline{\mathbf{x}}^t)^T \mathbf{u}^\infty = 0$ holds for all t, so by Proposition 2, $\mathbf{u}^\infty \in U^*$ and $\mathrm{dist}(\overline{\mathbf{x}}^t, X^*) \to 0$. Now, assume that $\mathcal{I}(\mathbf{u}^\infty) \neq \emptyset$, and consider an $i \in \mathcal{I}(\mathbf{u}^\infty)$. Since $\{\mathbf{u}^t\} \to \mathbf{u}^\infty$, it follows, for some fixed $\tau \geq 1$ that is large enough, that $u_i^t > 0$ for all $t \geq \tau$. Therefore, by the iteration formula (6), it holds that

$$h_i(\mathbf{x}^s) = \frac{u_i^{s+1} - u_i^s}{\alpha_s}, \quad s \geq \tau.$$

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

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

The function h_i is uniformly continuous over the compact set X, so for every $\delta > 0$ there exists an $\varepsilon > 0$ such that for any $\mathbf{x}$ with $\mathrm{dist}(\mathbf{x}, X(\mathbf{u}^\infty)) \leq \varepsilon$, the inequality

$$h_i(\mathbf{x}) \leq h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T (\mathbf{x} - \overline{\mathbf{x}}) + \frac{\delta}{3}$$

holds. If $\mathrm{rint}\, X(\mathbf{u}^\infty) = \emptyset$, i.e., the set $X(\mathbf{u}^\infty)$ is a singleton, the same reasoning holds when $\{\overline{\mathbf{x}}\} = X(\mathbf{u}^\infty)$. From Lemma 1, we know that $\{\mathrm{dist}(\mathbf{x}^s, X(\mathbf{u}^\infty))\} \to 0$ as $t \to \infty$, and hence, for some fixed $\kappa \geq \tau + 1$, the inequality $\mathrm{dist}(\mathbf{x}^s, X(\mathbf{u}^\infty)) \leq \varepsilon$ holds for all $s \geq \kappa$. Therefore, it holds that

$$h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T (\mathbf{x}^s - \overline{\mathbf{x}}) \geq \frac{u_i^{s+1} - u_i^s}{\alpha_s} - \frac{\delta}{3}, \quad s \geq \kappa. \tag{12}$$

Hence, for $t \geq \kappa + 1$, we have that

$$h_i(\overline{\mathbf{x}}^t) \geq h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T (\overline{\mathbf{x}}^t - \overline{\mathbf{x}}) = \sum_{s=0}^{t-1} \mu_s^t \left(h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T (\mathbf{x}^s - \overline{\mathbf{x}}) \right) \tag{13a}$$

$$\geq \sum_{s=0}^{\kappa-1} \mu_s^t \left(h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T (\mathbf{x}^s - \overline{\mathbf{x}}) \right) + \sum_{s=\kappa}^{t-1} \mu_s^t \left(\frac{1}{\alpha_s}(u_i^{s+1} - u_i^s) - \frac{\delta}{3} \right) \tag{13b}$$

$$\geq \sum_{s=0}^{\kappa-1} \mu_s^t \left(h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T (\mathbf{x}^s - \overline{\mathbf{x}}) \right) + \sum_{s=\kappa}^{t-1} \gamma_s^t (u_i^{s+1} - u_i^s) - \frac{\delta}{3}, \qquad (13c)$$

where the inequality in (13a) follows from the convexity of h_i, the inequality (13b) follows from (12), and the inequality (13c) from the fact that $\mu_s^t \leq 1$ for $s = 0, \ldots, t - 1$. For $t \geq \kappa$, under the conditions A1–A3 of Assumption 2, we have that

$$\sum_{s=0}^{\kappa-1} \gamma_s^t \leq \sum_{s=0}^{\kappa-1} \left(\gamma_0^t + (\kappa - 1) \Delta \gamma_{\max}^t \right) = \kappa \gamma_0^t + \kappa (\kappa - 1) \Delta \gamma_{\max}^t \to 0 \quad \text{as} \ \ t \to \infty,$$

which implies that $\gamma_s^t \to 0$ as $t \to \infty$ for $s = 0, \ldots, \kappa - 1$. Since $\alpha_s < \infty$ for $s = 0, \ldots, \kappa - 1$, it follows that $\mu_s^t \to 0$ as $t \to \infty$ for $s = 0, \ldots, \kappa - 1$. Therefore, for large enough values of $t \geq \kappa$, the inequality $\sum_{s=0}^{\kappa-1} \mu_s^t \left(h_i(\overline{\mathbf{x}}) + \boldsymbol{\xi}_i^T (\mathbf{x}^s - \overline{\mathbf{x}}) \right) \geq -\delta/3$ holds.

By the same reasoning as in the proof of Proposition 4 [the inequalities (11)], for large enough values of t, the inequality $\sum_{s=\kappa}^{t-1} \gamma_s^t (u_i^{s+1} - u_i^s) \geq -\delta/3$ holds. Hence,

$$h_i(\overline{\mathbf{x}}^t) \geq -\frac{\delta}{3} - \frac{\delta}{3} - \frac{\delta}{3} = -\delta$$

for large enough values of $t \geq \kappa + 1$. Therefore, $\liminf_{t \to \infty} h_i(\overline{\mathbf{x}}^t) \geq 0$ holds. From Proposition 4 follows that $\limsup_{t \to \infty} h_i(\overline{\mathbf{x}}^t) \leq 0$. We deduce 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 since $u_i^\infty = 0$ whenever $i \in \mathcal{I} \backslash \mathcal{I}(\mathbf{u}^\infty)$, it follows that

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

By Proposition 2, the theorem then follows. $\square$

For the case when (a) the functions f and h_i, $i \in \mathcal{I}$, are affine, and (b) the set X is a polytope, Theorem 1 reduces to the result of Sherali and Choi [42, Theorem 2].

3.3 A new rule for choosing the convexity weights when utilizing harmonic series step lengths

We now study the special case when the step lengths α_t are chosen according to a harmonic series, i.e.,

$$\alpha_t := \frac{a}{b + ct}, \quad t = 0, 1, \ldots, \quad \text{where } a > 0, \ b > 0, \ c > 0. \qquad (14)$$

This choice of step lengths was used by Larsson and Liu [27], and Larsson et al. [30] and guarantees, according to Proposition 3, convergence to a dual optimum. We

define

$$\Delta\mu_{\max}^t := \max_{s\in\{1,\ldots,t-1\}}\{\mu_s^t - \mu_{s-1}^t\}, \quad t = 2, 3, \ldots,$$

where μ_s^t are the convexity weights employed in (8).

Assumption 3 (convexity weights when employing harmonic step lengths) The convexity weights are chosen to satisfy

B1: $\mu_s^t \geq \mu_{s-1}^t$, $s = 1,\ldots,t-1$, $t = 2, 3, \ldots$,
B2: $t\Delta\mu_{\max}^t \to 0$, as $t \to \infty$, and
B3: $t\mu_{t-1}^t \leq M < \infty$, $t = 1, 2, \ldots$.

Condition B1 states that the convex combinations $\bar{\mathbf{x}}^t$, defined in (8), should put more weight on later observations (that is, primal subproblem solutions $\mathbf{x}^t$). Condition B2 states that no particular primal iterate should be favoured, meaning that the difference between the weights for two subsequent iterates should tend to zero. Condition B3 states that the convexity weights μ_{t-1}^t should decrease at a rate not lower than $1/t$ as $t \to \infty$.

Consider the following result.

Proposition 5 (convexity weights fulfilling Assumption 3 together with step lengths defined by (14) fulfill Assumption 2) *If the step lengths, α_t, fulfill (14) and the convexity weights, μ_s^t, satisfy Assumption 3, then Assumption 2 is fulfilled.*

Proof From (14) follows that the inequality $\alpha_s \leq \alpha_{s-1}$ holds for all $s \geq 1$, which implies that $\gamma_s^t - \gamma_{s-1}^t = \mu_s^t/\alpha_s - \mu_{s-1}^t/\alpha_{s-1} \geq (\mu_s^t - \mu_{s-1}^t)/\alpha_s \geq 0$. Hence, the condition A1 in Assumption 2 is fulfilled. Next, we have that

$$\gamma_s^t - \gamma_{s-1}^t = \frac{b+cs}{a}\mu_s^t - \frac{b+c(s-1)}{a}\mu_{s-1}^t$$
$$= \frac{1}{a}\left(b(\mu_s^t - \mu_{s-1}^t) + c\mu_{s-1}^t + cs(\mu_s^t - \mu_{s-1}^t)\right),$$

which implies that

$$\Delta\gamma_{\max}^t = \max_{s\in\{1,\ldots,t-1\}}\left\{\gamma_s^t - \gamma_{s-1}^t\right\} \tag{15a}$$

$$= \frac{1}{a}\max_{s\in\{1,\ldots,t-1\}}\left\{b\left(\mu_s^t - \mu_{s-1}^t\right) + c\mu_{s-1}^t + cs\left(\mu_s^t - \mu_{s-1}^t\right)\right\} \tag{15b}$$

$$\leq \frac{b}{a}\Delta\mu_{\max}^t + \frac{c}{a}\mu_{t-2}^t + \frac{c}{a}(t-1)\Delta\mu_{\max}^t \to 0, \quad \text{as } t \to \infty, \tag{15c}$$

where the inequality (15c) follows by maximizing each term in (15b) separately. The first term in (15c) tends to zero due to condition B2 in Assumption 3, the second term

converges to zero due to the conditions B1 and B3 in Assumption 3 and the third term converges to zero due to the condition B2 in Assumption 3. Hence, the condition A2 in Assumption 2 is fulfilled. We also have that

$$
\gamma_{t-1}^t = \frac{\mu_{t-1}^t}{\alpha_{t-1}} = \frac{(b + c(t-1))\mu_{t-1}^t}{a}
$$

$$
\leq \frac{\mu_{t-1}^t(b-c) + cM}{a} \leq \frac{M}{a}\left(\frac{|b-c|}{t} + c\right) < \infty,
$$

for any $t \geq 1$, which implies that the condition A3 in Assumption 2 is satisfied. □

Larsson et al. [30] show that by using the convexity weights $\mu_s^t = 1/t$, primal convergence can always be guaranteed for the harmonic series step lengths (14). We here refer to this *rule* for creating an ergodic sequence as the $1/t$-*rule*; it was first analyzed by Larsson and Liu [27], who prove convergence for the case when (1) is a linear program. Clearly, the $1/t$-rule fulfills the conditions of Corollary 5; hence the primal convergence of the $1/t$-rule is a special case of the analysis above.

One drawback of the $1/t$-rule is the fact that when creating the ergodic sequences of primal solutions, all previously found iterates are weighted equally. We expect that later subproblem solutions in the subgradient method are more likely to belong to the set of optimal solutions to the subproblem (2) at a dual optimal solution, $\mathbf{u}^* \in U^*$. We therefore propose a more general set of rules for creating ergodic sequences of primal iterates which allows for later primal iterates to receive larger convexity weights.

Definition 1 (the s^k-rule) Let $k \geq 0$. The s^k-rule creates an ergodic sequence by choosing convexity weights according to

$$
\mu_s^t = \frac{(s+1)^k}{\sum_{l=0}^{t-1}(l+1)^k}, \quad \text{for} \quad s = 0, \ldots, t-1, \quad t \geq 1.
$$

Note that the s^0-rule is equivalent to the $1/t$-rule. For $k > 0$, the s^k-rule results in an ergodic sequence in which the later iterates are assigned higher weights than the earlier ones. For larger values of k, the weights are shifted towards later iterates. In Fig. 1, the convexity weights μ_s^t are illustrated for $t = 10$ and $k \in \{0, 1, 2, 10\}$. The following proposition establishes primal convergence for the ergodic sequence created
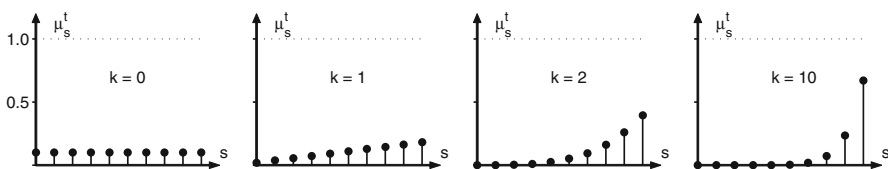

Fig. 1 Illustration of the convexity weights, μ_s^t, when $t = 10$, for the s^k-rule when $k = 0, 1, 2, 10$ respectively

by the s^k-rule, given that a harmonic series step length is utilized when solving the Lagrangian dual problem (4).

Proposition 6 (the s^k-rule satisfies Assumption 3) *The convexity weights μ_s^t, chosen according to Definition 1, fulfill Assumption 3.*

Proof The convexity weights μ_s^t clearly fulfill the condition B1 of Assumption 3. For any $t \geq 2$, it holds that

$$\Delta\mu_{\max}^t = \max_{s \in \{1,\dots,t-1\}} \{\mu_s^t - \mu_{s-1}^t\} = \max_{s \in \{1,\dots,t-1\}} \left\{ \frac{(s+1)^k - s^k}{\sum_{l=0}^{t-1}(l+1)^k} \right\},$$

where the maximum is obtained for $s = 1$ when $0 \leq k < 1$, and for $s = t - 1$ when $k \geq 1$. Noting that

$$\sum_{l=0}^{t-1}(l+1)^k = \sum_{l=0}^{t-1} \int_l^{l+1} \lceil x \rceil^k \, dx \geq \sum_{l=0}^{t-1} \int_l^{l+1} x^k \, dx = \int_0^t x^k \, dx = \frac{t^{k+1}}{k+1}, \quad (16)$$

it follows that, for $0 \leq k < 1$,

$$t \Delta\mu_{\max}^t = t \frac{1}{\sum_{l=0}^{t-1}(l+1)^k} \leq \frac{(k+1)}{t^k} \to 0, \quad \text{as} \quad t \to \infty.$$

For $k \geq 1$, it holds that

$$t \Delta\mu_{\max}^t = t \frac{t^k - (t-1)^k}{\sum_{l=0}^{t-1}(l+1)^k} \leq \frac{(k+1)t(t^k - (t-1)^k)}{t^{k+1}}$$

$$= (k+1)\left(1 - \left(\frac{t-1}{t}\right)^k\right) \to 0,$$

as $t \to \infty$. Hence, condition B2 of Assumption 3 holds. Condition B3 in Assumption 3 holds, due to the fact that

$$t\mu_{t-1}^t = \frac{t^{k+1}}{\sum_{l=0}^{t-1}(l+1)^k} \leq \frac{(k+1)t^{k+1}}{t^{k+1}} = k+1 < \infty, \quad t = 2, 3, \dots,$$

where the first inequality follows from (16). $\qquad\square$

One should note that when constructing the ergodic iterate, $\bar{\mathbf{x}}^t$, defined by the s^k-rule, not all previously found iterates, $\mathbf{x}^s$, $s = 0, \dots, t - 1$, have to be stored since the ergodic iterate can be updated by

$$\bar{\mathbf{x}}^0 = \mathbf{x}^0, \quad \bar{\mathbf{x}}^t = \frac{\sum_{s=0}^{t-2}(s+1)^k}{\sum_{s=0}^{t-1}(s+1)^k}\bar{\mathbf{x}}^{t-1} + \frac{t^k}{\sum_{s=0}^{t-1}(s+1)^k}\mathbf{x}^{t-1}, \quad t = 1, 2, \dots. \quad (17)$$

Hence, in iteration t, only the previous ergodic iterate, $\bar{\mathbf{x}}^{t-1}$, and the new primal iterate, $\mathbf{x}^t$, are required for constructing the new ergodic iterate, $\bar{\mathbf{x}}^t$.

We now summarize the results obtained in this section in the following theorem.

Theorem 2 (convergence of the s^k-rule) *Suppose Assumption 1 holds, that $\mathbf{u}^t$ is generated by the subgradient method (6) operated with harmonic series step lengths (14), and that $\bar{\mathbf{x}}^t$ is generated according to (8), where the convexity weights are chosen according to the s^k-rule (Definition 1). Then,*

$$\mathbf{u}^t \to \mathbf{u}^\infty \in U^* \quad and \quad \mathrm{dist}(\bar{\mathbf{x}}^t, X^*) \to 0.$$

Proof By Proposition 3, it follows that $\mathbf{u}^t \to \mathbf{u}^\infty \in U^*$. Using Propositions 5 and 6 yields that the assumptions of Theorem 1 hold, which completes the proof. $\square$

3.4 Connection with previous results

In this section, we present some previous proposals for choosing step lengths and convexity weights. For simplicity, we define

$$A_t = \sum_{s=0}^{t-1} \alpha_s, \quad B_t = \sum_{s=0}^{t-1} \alpha_s^2, \quad t = 1, 2, \ldots.$$

In the volume algorithm developed by Baharona and Anbil [7], each ergodic iterate is constructed as a convex combination of the previous ergodic iterate and the new primal iterate, i.e., $\bar{\mathbf{x}}^t = \beta \mathbf{x}^t + (1 - \beta)\bar{\mathbf{x}}^{t-1}$, where $0 < \beta < 1$. Translating this into the framework of our analysis, this is equivalent to letting

$$\mu_0^t = (1 - \beta)^{t-1}, \quad \mu_s^t = \beta(1 - \beta)^{t-s+1}, \quad s = 1, \ldots, t - 1. \tag{18}$$

In the taxonomy in Table 1, we present the following attributes of the previously developed algorithms which utilize the subgradient method to solve the dual problem:

Problem	Type of problem considered. For the case when problem (1) is a linear program, the assumptions are that f and h_i, $i \in \mathcal{I}$, are affine functions and that X is a nonempty and bounded polyhedron. This is denoted in the table as LP. The case of a general convex optimization problem is denoted CP
Step lengths	The step lengths α_t employed in the subgradient method (6). Step lengths defined according to (14) are denoted Harmonic series. If the step lengths fulfill $\alpha_t > 0$, $\lim_{t \to \infty} \alpha_t = 0$ and $\lim_{t \to \infty} A_t = \infty$, we denote this by Divergent series and if also $\lim_{t \to \infty} B_t < \infty$, we denote this by Divergent series, QB (quadratically bounded)
Conv. weights	The convexity weights, μ_s^t, defined in (8), defining the ergodic sequences of primal iterates
Theorem 1	Whether or not Theorem 1 guarantees the convergence of the method
Theorem 2	Whether or not Theorem 2 guarantees the convergence of the method

Table 1 Taxonomy of dual subgradient algorithms

	Problem	Step lengths	Conv. weights	Theorem 1	Theorem 2
Shor [44, Chapter 4]	LP	Divergent series	$\mu_s^t = \alpha_s/A_t$	Yes	No
Larsson and Liu [26]	LP	Harmonic series	$\mu_s^t = 1/t$	Yes	Yes
Sherali and Choi [42]	LP	$\alpha_t = (t+1)^{-\kappa}$	$\mu_s^t = 1/t$	Yes	No
Baharona and Anbil [7]	LP	Polyak step size	(18)	No	No
Larsson et al. [30]	CP	Divergent series, QB	$\mu_s^t = \alpha_s/A_t$	Yes	No
Larsson et al. [30]	CP	Harmonic series	$\mu_s^t = 1/t$	Yes	Yes
Nedić and Ozdaglar [32]	CP	Constant	$\mu_s^t = 1/t$	No	No
Gustavsson et al. (this art.)	CP	Harmonic series	s^k-rule (Definition 1)	Yes	Yes

Since the work presented in this paper utilizes the traditional subgradient method to solve the dual problem, we only include algorithms which employ this method for the dual problem in Table 1. More sophisticated methods for solving the dual problem include *deflected conditional subgradient methods* (d'Antonio and Frangioni [12], Burachik and Kaya [10]), *bundle methods* (Lemaréchal et al. [31], Kiwiel [20]), *augmented Lagrangian methods* (Rockafellar [41], Bertsekas [9]), and *ballstep subgradient methods* (Kiwiel et al. [22,23]). All of these methods utilize information from previously computed subgradients when updating the iterates in the subgradient scheme. In order to approximate the primal solutions, the convexity weights defining the primal iterates are then acquired from the information obtained in these dual schemes (e.g., Robinson [40]).

4 Applications to multicommodity network flows

We apply subgradient methods to a Lagrangian dual of the NMFP. Primal solutions are computed from ergodic sequences of subproblem solutions. For a more thorough description of multicommodity flow problems and solution methods for these, see [19, 35,36].

4.1 The nonlinear multicommodity network flow problem

Consider a graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ with a node set $\mathcal{N}$ and a set of directed arcs $\mathcal{A}$. There is a set $\mathcal{C} \subseteq \mathcal{N} \times \mathcal{N}$ of origin-destination pairs (OD-pairs). For each pair $k \in \mathcal{C}$, there is a flow demand, $d_k > 0$, associated with a specific commodity. We denote the set of simple routes from the origin to the destination of OD-pair k by $\mathcal{R}_k$, and the flow on route $r \in \mathcal{R}_k$ by h_{kr}. Let $[\delta_{kra}]_{r \in \mathcal{R}_k, k \in \mathcal{C}, a \in \mathcal{A}}$ 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 flow on arc $a \in \mathcal{A}$ is denoted by f_a and is defined by the route flows h_{kr} through

$$f_a = \sum_{k \in \mathcal{C}} \sum_{r \in \mathcal{R}_k} \delta_{kra} h_{kr}, \quad a \in \mathcal{A}. \tag{19}$$

To each arc $a \in \mathcal{A}$, we associate a convex cost function $g_a : \mathbb{R}_+ \to \mathbb{R}_+$ of its flow f_a. The NMFP then is the program

$$z^* = \text{minimum} \quad \sum_{a \in \mathcal{A}} g_a(f_a), \tag{20a}$$

$$\text{subject to} \quad \sum_{r \in \mathcal{R}_k} h_{kr} = d_k, \quad k \in \mathcal{C}, \tag{20b}$$

$$h_{kr} \geq 0, \quad r \in \mathcal{R}_k, \ k \in \mathcal{C}, \tag{20c}$$

$$\sum_{k \in \mathcal{C}} \sum_{r \in \mathcal{R}_k} \delta_{kra} h_{kr} = f_a, \quad a \in \mathcal{A}, \tag{20d}$$

$$f_a \geq 0, \quad a \in \mathcal{A}. \tag{20e}$$

One should note that the constraints (20e) are implied by (20c) and (20d), and do not have to be incorporated explicitly in the model. We will consider two definitions of the arc cost functions, $g_a(f_a)$, $a \in \mathcal{A}$. The first, *BPR (Bureau of Public Roads) nonlinear delay*, is used in the field of transportation (e.g., [6,29]) and is defined as

$$g_a(f_a) = r_a f_a \left(1 + \frac{p_a}{q_a + 1} \left(\frac{f_a}{c_a} \right)^{q_a} \right), \quad f_a \in \mathbb{R}_+, \ a \in \mathcal{A}, \tag{21}$$

where $r_a \geq 0$ is the *free-flow travel time* and $c_a > 0$ is the *practical capacity* of arc $a \in \mathcal{A}$. The parameters $p_a \geq 0$ and $q_a \geq 0$ are arc-specific. The second, *Kleinrock's average delays*, is used in the field of telecommunications (e.g., [24,35]) and is defined as

$$g_a(f_a) = \frac{f_a}{c_a - f_a}, \quad f_a \in [0, c_a), \ a \in \mathcal{A}, \tag{22}$$

where c_a, $a \in \mathcal{A}$, are the *arc capacities*.

4.2 A Lagrangian dual formulation

For the NMFP, i.e., the program (20), we utilize a Lagrangian dual approach in which the arc flow defining constraints (20d) are relaxed. For a more thorough explanation of the Lagrangian reformulation, see [28]. The resulting Lagrangian subproblem essentially consists of solving one shortest path problem for each commodity $k \in \mathcal{C}$.

Let $\mathbf{u} = [u_a]_{a \in \mathcal{A}}$ be the multipliers associated with the constraints (20d). We define the Lagrangian dual objective function by

$$\theta(\mathbf{u}) = \sum_{k \in \mathcal{C}} \phi_k(\mathbf{u}) + \sum_{a \in \mathcal{A}} \xi_a(u_a), \tag{23}$$

where, for each $k \in \mathcal{C}$ and any $\mathbf{u} \in \mathbb{R}^{|\mathcal{A}|}$, $\phi_k(\mathbf{u})$ is the optimal value of the shortest simple route subproblem, with arc costs u_a, $a \in \mathcal{A}$, given by

$$\phi_k(\mathbf{u}) = \text{minimum} \quad \sum_{r \in \mathcal{R}_k} \left(\sum_{a \in \mathcal{A}} u_a \delta_{kra} \right) h_{kr},$$

$$\text{subject to} \quad \sum_{r \in \mathcal{R}_k} h_{kr} = d_k, \tag{24}$$

$$h_{kr} \geq 0, \quad r \in \mathcal{R}_k,$$

and with solution set $H_k(\mathbf{u}) \subseteq \mathbb{R}_+^{|\mathcal{R}_k|}$. For each $a \in \mathcal{A}$ and any $u_a \in \mathbb{R}$, $\xi_a(u_a)$ is the optimal value of the single-arc subproblem

$$\xi_a(u_a) := \text{minimum}_{f_a \geq 0} \{g_a(f_a) - u_a f_a\}, \tag{25}$$

with solution $f_a(u_a) \subseteq \mathbb{R}_+$. For the cost functions (21) and (22), $f_a(u_a)$ can be expressed in closed form as

$$f_a(u_a) = \begin{cases} (g_a')^{-1}(u_a), & u_a \geq g_a'(0), \\ 0, & u_a < g_a'(0), \end{cases} \tag{26}$$

where $(g_a')^{-1}$ is the continuous inverse mapping of the derivative of g_a at u_a. The function $\theta : \mathbb{R}^{|\mathcal{A}|} \to \mathbb{R}$ is the sum of $|\mathcal{C}|$ concave and piecewise linear functions ϕ_k, $k \in \mathcal{C}$, and $|\mathcal{A}|$ concave and differentiable functions ξ_a, $a \in \mathcal{A}$. It is thus finite, continuous, concave, and subdifferentiable on $\mathbb{R}^{|\mathcal{A}|}$. Its subdifferential mapping at $\mathbf{u} \in \mathbb{R}^{|\mathcal{A}|}$ equals the bounded polyhedron

$$\partial\theta(\mathbf{u}) = \left\{ \left[\sum_{k \in \mathcal{C}} \sum_{r \in \mathcal{R}_k} \delta_{kra} h_{kr} - f_a(u_a) \right]_{a \in \mathcal{A}} \mid [h_{kr}]_{r \in \mathcal{R}_k} \in H_k(\mathbf{u}), \ k \in \mathcal{C} \right\}. \tag{27}$$

By weak duality, $\theta(\mathbf{u}) \leq z^*$ holds for all $\mathbf{u} \in \mathbb{R}^{|\mathcal{A}|}$. Consider an arbitrary $\mathbf{u} \in \mathbb{R}^{|\mathcal{A}|}$, and define $\widehat{u}_a = \max\{u_a, g_a'(0)\}$, $a \in \mathcal{A}$. Then, $f_a(\widehat{u}_a) = f_a(u_a)$, implying that $\xi_a(\widehat{u}_a) = \xi_a(u_a)$ $a \in \mathcal{A}$. Further, $\phi_k(\widehat{\mathbf{u}}) \geq \phi_k(\mathbf{u})$, $k \in \mathcal{C}$, since $\widehat{\mathbf{u}} \geq \mathbf{u}$, and it follows that $\theta(\widehat{\mathbf{u}}) \geq \theta(\mathbf{u})$. Since the Lagrangian dual objective function θ, defined in (23), is to be maximized on $\mathbb{R}^{|\mathcal{A}|}$, one can, without loss of generality, impose the restriction $u_a \geq g_a'(0)$, $a \in \mathcal{A}$. The Lagrangian dual can thus be stated as

$$\theta^* = \text{maximum} \ \theta(\mathbf{u}),$$
$$\text{subject to} \ u_a \geq g_a'(0), \quad a \in \mathcal{A}, \tag{28}$$

with solution set $U^* \subseteq \mathbb{R}^{|\mathcal{A}|}$.

Proposition 7 (primal–dual optimality, [30, Proposition 6]) *Let* $\mathbf{u}^* \in U^*$. *Then, strong duality holds, that is,* $\theta^* = \theta(\mathbf{u}^*) = z^*$. *Further,* $f_a^* = f_a(u_a^*)$, $a \in \mathcal{A}$, *and*

$$H_k^* = H_k(\mathbf{u}^*) \cap \left\{ [h_{kr}]_{r \in \mathcal{R}_k} \left| \sum_{l \in \mathcal{C}} \sum_{r \in \mathcal{R}_l} \delta_{lra} h_{lr} = f_a^*, \ a \in \mathcal{A} \right. \right\}, \quad k \in \mathcal{C}.$$

Proposition 7 states that the optimal arc flow $[f_a^*]_{a \in \mathcal{A}}$ is obtained from the solutions to the subproblems $\xi_a(u_a^*)$, $a \in \mathcal{A}$. However, an optimal route flow pattern $[h_{kr}^*]_{r \in \mathcal{R}_k} \in H_k^*$ is, in general, not directly available from the solution to the subproblem (24). This depends on the set $\prod_{k \in \mathcal{C}} H_k(\mathbf{u}^*)$ typically not being a singleton, since the functions ϕ_k, $k \in \mathcal{C}$, typically are nonsmooth at $\mathbf{u}^*$.

4.3 The algorithm

We solve the Lagrangian dual problem (28) by the subgradient method defined in (6). 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, defined by

$$y_a^t = \sum_{k \in \mathcal{C}} \sum_{r \in \mathcal{R}_k} \delta_{kra} h_{kr}, \quad a \in \mathcal{A}, \tag{29}$$

a subgradient of θ at $\mathbf{u}^t$ is given by the vector $[y_a^t - f_a(u_a^t)]_{a \in \mathcal{A}}$. The subgradient method (6) is then given by

$$u_a^{t+1} := \max \left\{ u_a^t + \alpha_t (y_a^t - f_a(u_a^t)), \ g_a'(0) \right\}, \quad a \in \mathcal{A}, \ t = 0, 1, \ldots,$$

where α_t is the step length used in iteration t. We create ergodic sequences of arc flows according to

$$\hat{f}_a^t := \sum_{s=0}^{t-1} \mu_s^t y_a^s, \quad a \in \mathcal{A}, \quad t = 1, 2, \ldots,$$

by choosing the convexity weights μ_s^t according to the s^k-rule (see Definition 1). Since all arc flows y_a^s, $a \in \mathcal{A}$, $s = 0, \ldots, t-1$, are feasible, the ergodic iterate $\hat{f}_a^t$ will also be feasible in (20) for $t \geq 1$. The ergodic iterates $\hat{f}_a^t$ are updated analogously to the update of $\bar{\mathbf{x}}^t$ in (17). In each iteration $t \geq 0$, we obtain a lower bound, $\underline{z}^t$, and an upper bound, $\bar{z}^t$, on the optimal objective value z^*, according to

$$\underline{z}^t := \max_{s \in \{0,1,\ldots,t\}} \{\theta(\mathbf{u}^s)\} \quad \text{and} \quad \bar{z}^t := \min_{s \in \{0,1,\ldots,t\}} \left\{ \sum_{a \in \mathcal{A}} g_a(\hat{f}_a^s) \right\}. \tag{30}$$

5 Numerical tests and results

We now utilize the subgradient approach described in Sect. 4.3 on a set of convex multicommodity flow problems to evaluate the performance of a number of different rules for choosing the convexity weights defining the ergodic sequences.

5.1 Implementation issues

The algorithm described in Sect. 4.3 has been implemented in Fortran95 on a Pentium Dual Core 2.50 GHz with 4 GB RAM under Linux Red Hat 2.16.0.

To solve the shortest-path subproblems defined in (24), we use Dijkstra's algorithm [13] as implemented in the subroutine L2QUE described in [18]. In every iteration, Dijkstra's algorithm is called $|\mathcal{S}|$ times, where $\mathcal{S} \subseteq \mathcal{N}$ is the union of all origin nodes of the OD set $\mathcal{C}$. No comparisons have been made between this implementation and other shortest-path solvers.

In the dual subgradient method (6), we adopt a harmonic series step length $\alpha_t = \widehat{\alpha}/(t+1)$, $t = 0, 1, \ldots$, where $\widehat{\alpha}$ is chosen for each specific problem instance. The subgradient algorithm is terminated when the relative optimality gap is below a specified limit, $\varepsilon_{\text{opt}} > 0$, i.e., when

$$\frac{\overline{z}^t - \underline{z}^t}{\max\{\underline{z}^t, 1\}} < \varepsilon_{\text{opt}}, \tag{31}$$

where $\overline{z}^t$ and $\underline{z}^t$ are the upper and lower bounds defined in (30).

5.2 Test problems

We evaluate our algorithm on three sets of test problems, which are also used in [2] and [21]. The first set, the *planar problems*,[1] consists of ten instances, in which nodes have been randomly chosen as points in the plane, and the arcs are such that the resulting graph is planar; the OD-pairs have been chosen at random. The *grid problems* (see footnote 1) collection contains 15 networks with a grid structure, meaning that each node has four incoming and four outgoing arcs; the OD-pairs have been chosen at random. The third set consists of three *telecommunication problems* of various sizes. The arc cost functions have been generated as in [2, Section 8.1] for all the test problems.

In Table 2, the characteristics of the problems are given, where $|\mathcal{N}|$ is the number of nodes, $|\mathcal{A}|$ is the number of arcs, $|\mathcal{C}|$ is the number of commodities and $|\mathcal{S}|$ is the number of calls to Dijkstra's algorithm needed in each iteration. Note that the characteristics are taken from [21] since some values in [2] are incorrect; see [3]. We also include in Table 2 some computational characteristics of the subgradient algorithm described in Sect. 4.3, CPU time and time spent on shortest path problems.

---

[1] Available at http://www.di.unipi.it/di/groups/optimize/Data/MMCF.html (Accessed 2013-12-10).

Table 2 Data for the test problems of Babonneau and Vial [2]

Problem ID	$|\mathcal{N}|$	$|\mathcal{A}|$	$|\mathcal{C}|$	$|\mathcal{S}|$	CPU (ms)	%SP
Planar problems						
planar30	30	150	92	29	0.09	83.5
planar50	50	250	267	50	0.19	85.8
planar80	80	440	543	80	0.75	92.4
planar100	100	532	1,085	100	0.92	89.8
planar150	150	850	2,239	150	2.79	92.9
planar300	300	1,680	3,584	300	9.55	94.1
planar500	500	2,842	3,525	500	50.06	94.3
planar800	800	4,388	12,756	800	185.88	95.1
planar1000	1,000	5,200	20,026	1,000	130.19	88.3
planar2500	2,500	12,990	81,430	2,500	1,382.11	91.8
Grid problems						
grid1	25	80	50	23	0.06	78.1
grid2	25	80	100	25	0.09	82.7
grid3	100	360	50	40	0.17	77.9
grid4	100	360	100	63	0.33	84.5
grid5	225	840	100	83	0.73	80.8
grid6	225	840	200	135	1.32	87.5
grid7	400	1,520	400	247	5.39	76.0
grid8	625	2,400	500	343	12.22	79.4
grid9	625	2,400	1,000	495	21.03	78.6
grid10	625	2,400	2,000	593	38.40	85.9
grid11	625	2,400	4,000	625	74.58	90.0
grid12	900	3,480	6,000	899	164.67	89.3
grid13	900	3,480	12,000	900	317.34	91.9
grid14	1,225	4,760	16,000	1,225	593.30	91.8
grid15	1,225	4,750	32,000	1,225	1,180.56	92.5
Telecommunication problems						
ndo22	14	22	23	5	0.00	30.5
ndo148	61	148	122	61	0.17	66.5
904	106	904	11,130	106	1.65	80.5

CPU (ms) denotes the CPU time in milliseconds per iteration of the subgradient algorithm, and %SP denotes the percentage of time spent on solving the shortest path problems defined in (24)

5.3 Convexity weight rules

We have chosen to analyze the $1/t$-rule [27,30], the volume algorithm (VA) [7] and the proposed s^k-rule for $k = 1, 2, 4$, and 10 on the problem instances listed in Table 2. For the VA, we update the ergodic iterates by $\overline{\mathbf{x}}^t = \beta \mathbf{x}^t + (1 - \beta)\overline{\mathbf{x}}^{t-1}$, where $\beta = 0.1$, as proposed in [7]. We decided not to include the rule described in [44, Chapter 4], since for most of the problem instances, it did not reach the optimality threshold chosen within 10,000 iterations.

5.4 Results

Tables 3 and 4 present our results when defining the arc cost functions as the BPR congestion function (21) and the Kleinrock function (22), respectively. In these tables,

- $\widehat{\alpha}$ represents the initial step length used in the subgradient algorithm (6) which was chosen as the integer power of 10 that yielded the best performance for each problem instance, and
- for the $1/t$-rule, the VA, and the s^k-rules, the number of subgradient iterations required to reach an optimality gap below the given threshold, ε_{opt}, are listed.

We impose a limit of 10,000 iterations, and denote by a dash ('–') when the optimality gap did not reach below ε_{opt} within this limit.

In Fig. 2, the performance profiles [14] for the methods are illustrated for the 56 test problems considered (28 using the BRP congestion function and 28 using the Kleinrock delay function). The graphs in the figure represent the portion of problems solved (that is, attained an optimality gap below the given threshold ε_{opt}) within a factor τ times the number of iterations needed by the method that reached the threshold within the least number of iterations.

The s^k-rule for $k = 1, 2, 4$, and 10 clearly outperforms both the $1/t$-rule and the VA for the test instances. The best performance was shown by the s^4-rule which reached the acquired relative optimality gap [defined in (31)] for 37 out of the 56 instances using the least number of iterations. For the problem instance where the s^4-rule performed the poorest it still solved the problem within a factor $\tau \approx 1.25$ times the number of iterations needed by the method which solved that instance within the least number of iterations. The VA ($1/t$-rule) failed to obtain the given optimality threshold within 10,000 iterations for ten (five) problem instances, while the s^k-rules failed on only four of the problem instances.

6 Conclusions and future research

We generalize the convergence results in [42] to convex optimization problems and extend the analysis in [30] to include more general convex combinations for creating the ergodic sequences.

The proposed s^k-rule for choosing convexity weights for the primal iterates allows putting more weight on later iterates in the subgradient scheme. Computational results for three sets of NMFPs demonstrate that the s^k-rule is convincing and shows a performance superior to that of previously proposed rules. Section 5 presents a comparison between different rules for choosing the convexity weights in the subgradient scheme, and should *not* be viewed as an attempt to provide a new, competitive solution method for the NMFP.

Since the convergence results are presented for general convex optimization problems, we have not analyzed the performance of the s^k-rule specifically for linear programs. Preliminary numerical tests indicate, however, a similar performance.

Future interesting research includes an analysis of the performance of the s^k-rule for other problems, for which subgradient schemes have proven to be successful. Examples

Table 3 Results for the BPR congestion function (21): the number of iterations until the relative optimality gap defined in (31) is below $\varepsilon_{\mathrm{opt}} = 10^{-4}$

Problem ID	$\widehat{\alpha}$	$1/t$-rule (s^0-rule)	VA	s^k-rule			
				$k = 1$	$k = 2$	$k = 4$	$k = 10$
Planar problems							
planar30	10^2	9,965	**4,650**	4,986	4,916	4,814	4,773
planar50	10^2	1,846	350	273	**252**	**252**	273
planar80	10^2	6,317	–	**1,353**	**1,353**	**1,353**	**1,353**
planar100	10^1	1,497	477	271	**266**	**266**	**266**
planar150	10^1	6,371	–	1,613	**1,568**	**1,568**	**1,568**
planar300	10^0	1,122	493	338	332	**314**	332
planar500	10^0	3,323	162	189	162	**142**	148
planar800	10^{-1}	1,458	510	524	468	428	**415**
planar1000	10^{-1}	–	–	–	–	–	–
planar2500	10^{-1}	–	–	–	–	–	–
Grid problems							
grid1	10^{-1}	434	164	**162**	**162**	**162**	163
grid2	10^{-1}	712	233	233	188	**167**	168
grid3	10^{-1}	648	143	143	136	**114**	136
grid4	10^{-1}	758	166	161	**157**	**157**	**157**
grid5	10^{-1}	755	156	155	**137**	**137**	140
grid6	10^{-1}	916	426	271	**242**	**242**	252
grid7	10^{-2}	277	138	138	126	**119**	132
grid8	10^{-2}	839	988	509	470	443	**432**
grid9	10^{-2}	400	232	159	**149**	**149**	175
grid10	10^{-2}	597	176	154	**128**	**128**	150
grid11	10^{-3}	720	534	470	436	413	**406**
grid12	10^{-3}	209	73	80	70	**69**	79
grid13	10^{-3}	374	77	96	78	**74**	78
grid14	10^{-3}	488	68	78	**56**	**56**	68
grid15	10^{-4}	338	210	214	195	**185**	**185**
Telecommunication problems							
ndo22	10^0	2,279	68	56	21	**15**	**15**
ndo148	10^0	341	80	80	80	**76**	**76**
904	10^1	7,698	2,322	**1,302**	**1,302**	**1,302**	**1,302**

For each problem ID, the bold entry indicates the rule(s) that achieved the given optimality gap within the least number of iterations

–: Failed to reach the required relative optimality gap within 10,000 iterations

are found within the fields of discrete optimization (e.g., Ceria et al. [11], Fisher [16]), network design (e.g., Balakrishnan et al. [5], Frangioni and Gendron [17]), and traffic assignment (e.g., Patriksson [36]).

Table 4 Results for the Kleinrock delay function (22): the number of iterations until the relative optimality gap defined in (31) is below $\varepsilon_{\text{opt}} = 10^{-2}$

Problem ID	$\widehat{\alpha}$	$1/t$-rule (s^0-rule)	VA	s^k-rule			
				$k = 1$	$k = 2$	$k = 4$	$k = 10$
Planar problems							
planar30	10^{-3}	116	49	54	**46**	**46**	62
planar50	10^{-3}	350	**113**	140	122	114	**113**
planar80	10^{-3}	360	132	146	132	**129**	131
planar100	10^{-4}	161	54	50	**45**	**45**	59
planar150	10^{-4}	1,732	7,365	736	599	564	**456**
planar300	10^{-5}	71	36	28	**26**	**26**	35
planar500	10^{-5}	112	40	36	28	**26**	27
planar800	10^{-6}	54	31	22	**18**	**18**	26
planar1000	10^{-6}	234	125	114	**103**	**103**	120
planar2500	10^{-7}	–	5,784	7,279	6,162	5,358	**4,600**
Grid problems							
grid1	10^{-4}	830	503	435	420	**418**	**418**
grid2	10^{-4}	–	–	–	–	–	–
grid3	10^{-4}	150	64	**63**	**63**	**63**	81
grid4	10^{-4}	348	157	144	136	**135**	143
grid5	10^{-4}	219	100	96	**85**	**85**	102
grid6	10^{-4}	884	793	515	488	465	**462**
grid7	10^{-5}	120	82	64	**60**	67	95
grid8	10^{-5}	697	3,004	448	423	**414**	431
grid9	10^{-5}	665	–	436	404	**397**	436
grid10	10^{-6}	5,683	–	5,226	5,191	5,177	**5,163**
grid11	10^{-7}	1,089	–	956	**948**	956	986
grid12	10^{-7}	147	229	101	**98**	106	142
grid13	10^{-8}	921	–	810	**807**	810	843
grid14	10^{-8}	103	144	**81**	**81**	89	121
grid15	10^{-9}	147	349	118	**114**	121	156
Telecommunication problems							
ndo22	10^{-1}	119	**98**	**98**	**98**	**98**	**98**
ndo148	10^{-2}	–	–	–	–	–	–
904	10^{-4}	627	456	451	**444**	**444**	**444**

For each problem ID, the bold entry indicates the rule(s) that achieved the given optimality gap within the least number of iterations

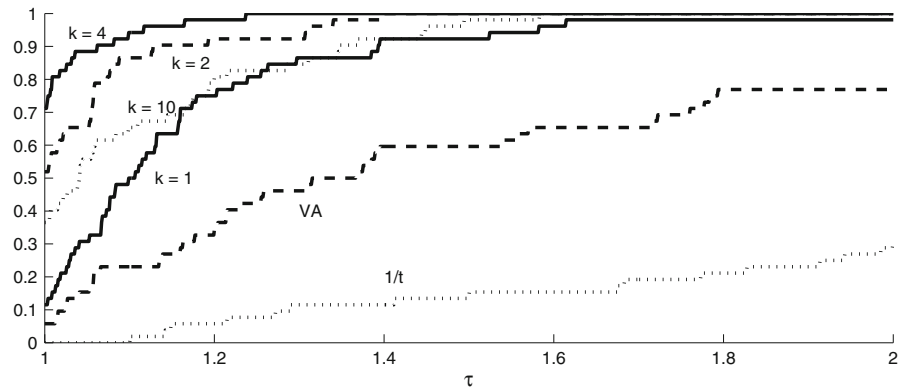–: Failed to reach the required relative optimality gap within 10,000 iterations

Fig. 2 Performance profiles of the methods for the 56 test instances (28 instances, each with the BPR congestion function (21) and the Kleinrock delay function (22), respectively). The graphs illustrate the proportion of the instances which each of the methods solved within τ times the number of iterations required by the method which solved each corresponding instance within the least number of iterations

The s^k-rule is utilized together with harmonic series step lengths, and future research should also investigate convergence results and the practical performance of the rule when utilizing other step lengths, for example Polyak step lengths [39, Chapter 5.3]. We also aim at analyzing the convergence rate of the ergodic sequences in terms of infeasibility and sub-optimality depending on the convexity weight rules utilized. Another extension of the results presented here would be to analyze the convergence of the ergodic sequences when allowing inexact solutions of the subproblems; such solutions would provide ε-subgradients of the dual objective function (e.g., d'Antonio and Frangioni [12]).

We are currently investigating the feasibility and computational potential of using the s^k-rule when employing other methods for solving the dual problem, for example augmented Lagrangian methods (e.g., Rockafellar [41], Bertsekas [9]), bundle methods (e.g., Lemaréchal et al. [31]) and ballstep subgradient methods (e.g., Kiwiel et al. [22,23]).

References

1. Antreicher, K.M., Wolsey, L.A.: Two "well-known" properties of subgradient optimization. Math. Program. **120**, 213–220 (2009)
2. Babonneau, F., Vial, J.-P.: ACCPM with a nonlinear constraint and an active set strategy to solve nonlinear multicommodity flow problems. Math. Program. **120**, 170–210 (2009)
3. Babonneau, F., Vial, J.-P.: ACCPM with a nonlinear constraint and an active set strategy to solve nonlinear multicommodity flow problems: a corrigendum. Math. Program. **120**, 211–212 (2009)
4. Bahiense, L., Maculan, N., Sagastizábal, C.: The volume algorithm revisited: relation with bundle methods. Math. Program. **94**, 41–69 (2002)

5. Balakrishnan, A., Magnanti, T.L., Wong, R.T.: A dual-ascent procedure for large-scale uncapacitated network design. Oper. Res. **37**, 716–740 (1989)
6. Bar-Gera, H.: Origin-based algorithm for the traffic assignment problem. Transp. Sci. **36**(4), 398–417 (2002)
7. Barahona, F., Anbil, R.: The volume algorithm: producing primal solutions with a subgradient method. Math. Program. **87**, 385–399 (2000)
8. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming Theory and Applications, 2nd edn. Wiley, New York (1993)
9. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, San Diego, CA (1982)
10. Burachik, R.S., Kaya, C.Y.: A deflected subgradient method using a general augmented Lagrangian duality with implications on penalty methods. In: Burachik, R.S., Yao, J.C. (eds.) Variational Analysis and Generalized Differentiation in Optimization and Control, Springer Optimization and Its Applications, vol. 47, pp. 109–132. Springer, New York (2010)
11. Ceria, S., Nobili, P., Sassano, A.: A Lagrangian-based heuristic for large-scale set covering problems. Math. Program. **81**, 215–228 (1998)
12. d'Antonio, G., Frangioni, A.: Convergence analysis of deflected conditional approximate subgradient methods. SIAM J. Optim. **20**, 357–386 (2009)
13. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numer. Math. **1**, 269–271 (1959)
14. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program. **91**(2), 201–213 (2002)
15. Ermol'ev, Y.M.: Methods of solution of nonlinear extremal problems. Cybernetics **2**, 1–14 (1966)
16. Fisher, M.L.: The Lagrangian relaxation method for solving integer programming problems. Manag. Sci. **27**, 626–642 (1991)
17. Frangioni, A., Gendron, B.: 0-1 reformulations of the multicommodity capacitated network design problem. Discret. Appl. Math. **157**, 1229–1241 (2009)
18. Gallo, G., Pallottino, S.: Shortest path algorithms. Ann. Oper. Res. **13**, 1–79 (1988)
19. Goffin, J.-L., Gondzio, J., Sarkissian, R., Vial, J.-P.: Solving nonlinear multicommodity flow problems by the analytic center cutting plane method. Math. Program. **76**, 131–154 (1996)
20. Kiwiel, K.C.: Proximity control in bundle methods for convex nondifferentiable minimization. Math. Program. **46**, 105–122 (1990)
21. Kiwiel, K.C.: An alternative linearization bundle method for convex optimization and nonlinear multicommodity flow problems. Math. Program. **130**, 59–84 (2011)
22. Kiwiel, K.C., Larsson, T., Lindberg, P.O.: The efficiency of ballstep subgradient level methods for convex optimization. Math. Oper. Res. **24**, 237–254 (1999)
23. Kiwiel, K.C., Larsson, T., Lindberg, P.O.: Lagrangian relaxation via ballstep subgradient methods. Math. Oper. Res. **32**, 669–686 (2007)
24. Kleinrock, L.: Communication Nets; Stochastic Message Flow and Delay. Dover, New York (1972)
25. Knopp, K.: Infinite Sequences and Series. Dover Publications, New York, NY (1956)
26. Larsson, T., Liu, Z.: A Primal Convergence Result for Dual Subgradient Optimization with Application to Multi-Commodity Network Flows. Technical Report. Department of Mathematics, Linköping Institute of Technology (1989)
27. Larsson, T., Liu, Z.: A Lagrangean relaxation scheme for structured linear programs with application to multicommodity network flows. Optimization **40**, 247–284 (1997)
28. Larsson, T., Liu, Z., Patriksson, M.: A dual scheme for traffic assignment problems. Optimization **42**, 323–358 (1997)
29. Larsson, T., Patriksson, M.: An augmented Lagrangean dual algorithm for link capacity side constrained traffic assignment problems. Transp. Res. Part B Methodol. **29**(6), 433–455 (1995)
30. Larsson, T., Patriksson, M., Strömberg, A.-B.: Ergodic, primal convergence in dual subgradient schemes for convex programming. Math. Program. **86**, 283–312 (1999)
31. Lemaréchal, C., Nemirovskii, A., Nesterov, Y.: New variants of bundle methods. Math. Program. **69**, 111–147 (1995)
32. Nedić, A., Ozdaglar, A.: Approximate primal solutions and rate analysis for dual subgradient methods. SIAM J. Optim. **19**, 1757–1780 (2009)
33. Nedić, A., Ozdaglar, A.: Subgradient methods for saddle-point problems. J. Optim. Theory Appl. **142**, 205–228 (2009)

34. Nesterov, Y.: Primal-dual subgradient methods for convex problems. Math. Program. Ser. B **120**, 221–259 (2009)
35. Ouorou, A., Mahey, P., Vial, J.-P.: A survey of algorithms for convex multicommodity flow problems. Manag. Sci. **46**, 126–147 (2000)
36. Patriksson, M.: The Traffic Assignment Problem: Models and Methods. Topics in Transportation series, VSP, Utrecht, The Netherlands (1994)
37. Polyak, B.T.: A general method of solving extremum problems. Sov. Math. Dokl. **8**, 593–597 (1967)
38. Polyak, B.T.: Minimization of unsmooth functionals. Comput. Math. Math. Phys. **9**, 14–29 (1969)
39. Polyak, B.T.: Introduction to Optimization. Optimization Software, Publications Division, NY (1987)
40. Robinson, S.M.: Bundle-based decomposition: conditions for convergence. In: International Institute for Applied Systems Analysis (1987)
41. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex optimization. Math. Oper. Res. **1**, 97–116 (1976)
42. Sherali, H.D., Choi, G.: Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs. Oper. Res. Lett. **19**, 105–113 (1996)
43. Sherali, H.D., Lim, C.: On embedding the volume algorithm in a variable target value method. Oper. Res. Lett. **32**, 455–462 (2004)
44. Shor, N.Z.: Minimization Methods for Non-Differentiable Functions. Springer, Berlin (1985)