43/2006

Raport Badawczy

Research Report

RB/57/2006

Lagrangian relaxation via ballstep subgradient methods

K. C. Kiwiel

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 8373578

fax: (+48) (22) 8372772

Kierownik Pracowni zgłaszający pracę:
Prof. dr hab. inż. Krzysztof C. Kiwiel

Warszawa 2006

Lagrangian Relaxation via Ballstep Subgradient Methods

Krzysztof C. Kiwiel
Systems Research Institute, Newelska 6, 01–447 Warsaw, Poland
email: kiwiel@ibspan.waw.pl

Torbjörn Larsson
Linköping University, S-58183 Linköping, Sweden
email: tolar@math.liu.se

P. O. Lindberg
Linköping University, S-58183 Linköping, Sweden
email: polin@math.liu.se

We exhibit useful properties of ballstep subgradient methods for convex optimization that use level controls for estimating the optimal value. Augmented with simple averaging schemes, they asymptotically find objective and constraint subgradients involved in optimality conditions. When applied to Lagrangian relaxation of convex programs, they find both primal and dual solutions, and have practicable stopping criteria. Up till now, similar results have only been known for proximal bundle methods, and for subgradient methods with divergent series stepsizes, whose convergence can be slow. Encouraging numerical results are presented for large-scale nonlinear multicommodity network flow problems.

Key words: convex programming; nondifferentiable optimization; subgradient optimization, Lagrangian relaxation, level projection methods

MSC2000 Subject Classification: Primary: 65K05; Secondary: 90C25

OR/MS subject classification: Primary: programming; Secondary: convex

---

1. Introduction. We consider subgradient methods for the convex minimization problem

$$f_* := \min\{\, f(x) : x \in S \,\} \tag{1}$$

under the following assumptions. S is a nonempty closed convex set in $\mathbb{R}^n$, the objective function $f : \mathbb{R}^n \to \mathbb{R}$ is convex, for each $x \in S$ we can find the value $f(x)$ and a subgradient $g_f(x) \in \partial f(x)$ of f at x, and for each $x \in \mathbb{R}^n$ we can find $P_S x := \arg\min_S |x - \cdot|$, its projection on S in the Euclidean norm $|\cdot|$. Finally, we assume that the *optimal solution set* $S_* := \operatorname{Arg}\min_S f$ of problem (1) is nonempty.

This setting covers many applications, but we are mostly interested in Lagrangian relaxation (see, e.g., Hiriart-Urruty and Lemaréchal [19, Chap. XII]) in the framework given below.

EXAMPLE 1.1 Consider the following *primal* convex optimization problem:

$$\psi_0^{\max} := \max\ \psi_0(z) \quad \text{s.t.} \quad \psi_j(z) \ge 0,\ j = 1{:}n,\ z \in Z, \tag{2}$$

where the set $\emptyset \ne Z \subset \mathbb{R}^{\bar m}$ is compact and convex, and each function ψ_j is concave, proper and closed (upper semicontinuous) with $\operatorname{dom}\psi_j \supset Z$. The Lagrangian of (2) has the form $\psi_0(z) + \langle x, \psi(z)\rangle$, where $\psi := (\psi_1, \dots, \psi_n)$ and x is a multiplier. Suppose that, at each multiplier x in the *dual feasible set* $\breve S := \mathbb{R}^n_+$, the *dual function*

$$f(x) := \max\{\, \psi_0(z) + \langle x, \psi(z)\rangle : z \in Z \,\} \tag{3}$$

can be evaluated by finding a *partial Lagrangian solution*

$$z(x) \in Z(x) := \operatorname{Arg}\max\{\, \psi_0(z) + \langle x, \psi(z)\rangle : z \in Z \,\}. \tag{4}$$

Thus f is finite convex and has a subgradient mapping $g_f(\cdot) := \psi(z(\cdot))$ on $\breve S$. For algorithmic purposes, suppose that this mapping g_f is locally bounded on $\breve S$ (e.g., f is the restriction to $\breve S$ of a convex function finite on an open neighborhood of $\breve S$, or $\inf_Z \min_{j=1}^n \psi_j > -\infty$, or ψ is continuous on Z). Finally, assume that the *dual optimal set* $\breve S_* := \operatorname{Arg}\min_{\breve S} f$ is nonempty; e.g., if Slater's condition holds ($\psi(\check z) > 0$ for some $\check z \in Z$), then $\breve S_*$ is both nonempty and bounded. For $S := \breve S$, problem (1) is the standard dual of (2). However, if we know strict upper bounds on a dual solution in the form of a point $x^{\rm up}$ such that $x^{\rm up} > \bar x$ for some $\bar x \in \breve S_*$, then it may be more efficient to take $S := \{x : 0 \le x \le x^{\rm up}\}$.

This paper shows that in the Lagrangian relaxation setting of Example 1.1, the ballstep subgradient method of Kiwiel et al. [26] applied to the dual problem (1) can provide a solution of the primal problem (2) at no extra cost. In its simplest form, this method proceeds like standard subgradient methods, except for a special choice of stepsizes. At iteration $k \geq 1$, for the current iterate $x^k \in S$ and the *target level* $f_{\text{lev}}^k < f(x^k)$ that estimates the optimal value f_* of (1), it uses the *subgradient linearization* of f

$$f_k(\cdot) := f(x^k) + \langle g_f^k, \cdot - x^k \rangle \leq f(\cdot) \quad \text{with} \quad g_f^k := g_f(x^k) \in \partial f(x^k) \tag{5}$$

and its halfspace

$$H_k := \{x : f_k(x) \leq f_{\text{lev}}^k\} \tag{6}$$

as an outer approximation to the f_{lev}^k*-level set* of f:

$$\mathcal{L}_f(f_{\text{lev}}^k) := \{x : f(x) \leq f_{\text{lev}}^k\} \subset H_k = \mathcal{L}_{f_k}(f_{\text{lev}}^k). \tag{7}$$

Then, as in the algorithm of Polyak [37], successive projections onto H_k and S give the next iterate

$$x^{k+1} := P_S(x^k + t_k[P_{H_k}x^k - x^k]) = P_S(x^k - t_k[f_k(x^k) - f_{\text{lev}}^k]g_f^k/|g_f^k|^2), \tag{8}$$

where the second equality is due to $f_k(x^k) = f(x^k) > f_{\text{lev}}^k$, and t_k is a *relaxation* factor satisfying

$$t_k \in T := [\,t_{\min}, t_{\max}\,] \quad \text{for some fixed} \quad 0 < t_{\min} \leq t_{\max} < 2. \tag{9}$$

The targets are chosen via a ballstep strategy that works in *groups* of iterations (because a single subgradient iteration does not provide enough information for changing the current target). Within each group, the target f_{lev}^k is fixed, and the method attempts to minimize f over a certain ball around the best point found so far. Two outcomes may arise. Either the objective f decreases sufficiently relative to the target, in which case the ball is shifted to the best iterate and the target is lowered, or it is discovered that the target is too low, in which case the ball is shrinked and the target is increased. For discovering whether the target is unattainable, we may use the two level schemes of Kiwiel et al. [26, §§2 and 5]; both schemes ensure that $\inf_k f(x^k) = f_*$ and provide efficiency estimates when the optimal set S_* is bounded.

For comparisons with other approaches, we note that although our iteration (8) with the *stepsizes*

$$\nu_k := t_k[f_k(x^k) - f_{\text{lev}}^k]/|g_f^k|^2 > 0 \tag{10}$$

conforms with the standard subgradient iteration

$$x^{k+1} := P_S(x^k - \nu_k g_f^k) \quad \text{with } \nu_k > 0, \tag{11}$$

our stepsizes do not have to obey the popular *divergent series* condition

$$\sum_{k=1}^{\infty} \nu_k = \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \nu_k^2 < \infty, \tag{12}$$

or other conditions typically required for convergence of subgradient methods; see Kiwiel [24].

In this paper we augment the ballstep method with simple *averaging* schemes, using the *convex weights*

$$\nu_j^k := \nu_j/\tilde{\nu}_f^k \quad \text{for} \quad j = k(l){:}\,k \quad \text{with} \quad \tilde{\nu}_f^k := \sum_{j=k(l)}^{k} \nu_j, \tag{13}$$

where $k(l)$ is the iteration number at which the current lth group started. These convex weights lead to aggregate versions of various quantities related to our method. For instance, by combining the oracle linearizations of (5), we obtain the *aggregate linearization* $\tilde{f}_k := \sum_{j=k(l)}^k \nu_j^k f_j$, which is an affine minorant of f. We show that its gradient $\nabla \tilde{f}_k$ can be used for finding asymptotically objective and constraint subgradients involved in optimality conditions for problem (1). Similarly, in Lagrangian relaxation, we may combine the partial Lagrangian solutions $z(x^j)$ of (4) to produce the *aggregate primal solution* $\tilde{z}^k := \sum_{j=k(l)}^k \nu_j^k z(x^j)$. We show that these aggregate solutions $\tilde{z}^k$ converge subsequentially to the set of optimal solutions to the primal problem (2). Further, we provide practicable stopping criteria, which allow the method to terminate when $\tilde{z}^k$ is an ϵ-solution of (2) for a given $\epsilon > 0$. To sum up, in Lagrangian relaxation, our method finds both primal and dual solutions. Up till now, for subgradient methods similar results have only been known for the iteration (11) with stepsizes obeying (12) and weights given by (13) with $k(l) = 1$, whose convergence can be slow; see Zhurbenko [44], Shor [42, §4.4], Anstreicher and Wolsey [1], Larsson and Liu [28], Larson et al. [31, 32], and Sherali and Choi [41].

Our results parallel ones given by Feltenmark and Kiwiel [12] for the proximal bundle method of Hiriart-Urruty and Lemaréchal [19, §XV.3] and Kiwiel [20]. At first sight, this method has little in common with our simple subgradient algorithm, since it accumulates many linearizations for its QP subproblems, and uses the QP multipliers for averaging. But in fact there are more similarities than differences. Our key observation is that, from the convergence viewpoint, a *group* of iterations of the ballstep method is similar to *one* iteration of the bundle method. Thus, once suitable estimates for a group of ballstep iterations are established, the remainder of our convergence analysis is almost identical to that of Feltenmark and Kiwiel [12]. Also the efficiency analysis of both methods is quite similar; see Kiwiel [23] and Kiwiel et al. [26]. Up till now, the literature has only contrasted simple subgradient methods with more advanced bundle methods, whereas our paper highlights their similarities.

Good reviews of the subgradient algorithm may be found in Bertsekas [9], Polyak [38] and Shor [42], and more recent variants in Ben-Tal et al. [7], Kiwiel [24], Kiwiel and Lindberg [27], Nedić and Bertsekas [34], Nedić et al. [35]. It is widely used, mainly due to its simplicity and good performance, especially in Lagrangian relaxation. In many applications it solves the dual of an LP relaxation of the original problem; then even quite approximate primal solutions delivered by our averaging schemes could be useful, e.g., in primal heuristics, variable fixing, etc.; see Balas and Cerna [3], Barahona and Chudak [6], Bahiense et al. [2], and Ceria et al. [11].

Also the recent volume algorithm of Barahona and Anbil [4] performs well in practice; see Barahona and Anbil [5] and Bahiense et al. [2]. Its averaging is similar to that of a version of our method that employs past aggregate subgradients to avoid zigzags (cf. (45)). However, in contrast with our method, the volume algorithm has no proof of convergence; see Bahiense et al. [2]. We hope, therefore, that our results may stimulate research on the development of simple subgradient methods that are both theoretically convergent and practically effective.

As a partial justification of our hope, we give preliminary numerical results for the traffic assignment and message routing problems (see, e.g., Bertsekas [8]) on apparently the largest instances reported in the literature. For modest solution accuracy (typical in such applications) our implementation seems to be competitive with the methods reviewed in the recent survey of Ouorou et al. [36].

The paper is organized as follows. In §2 we review briefly the simplest ballstep method of Kiwiel et al. [26] and its convergence properties. In §3 we show how averaging may produce affine minorants of f and the indicator function i_S of S, and a useful optimality estimate. Their uses for indentifying subgradients of f and i_S involved in optimality conditions for $\min_S f$ are discussed in §4. Applications to Lagrangian relaxation are studied in §5. Extensions to the accelerations of Kiwiel et al. [26, §7] are discussed in §6. Applications to multicommodity network flows are reported in §7.

Our notation is fairly standard. $B(x,r) := \{y : |y-x| \le r\}$ is the ball with center x and radius r. $d_C(\cdot) := \inf_{y\in C} |\cdot - y|$ is the distance function of a set $C \subset \mathbb{R}^n$ ($d_C = \infty$ if $C = \emptyset$).

2. The ballstep level algorithm. The simplest version of the ballstep subgradient method of Kiwiel et al. [26] stated below employs the following notation. At iteration k, $x^k_{\rm rec}$ is the *record* point with the best objective value $f^k_{\rm rec} := \min_{j=1}^k f(x^j)$ obtained so far. The iterations are split into *groups*

$$K_l := \{k(l) : k(l+1) - 1\}, \quad l \ge 1. \tag{14}$$

In group l, starting from the point $x^{k(l)}_{\rm rec}$, the method attempts to reach the *frozen* target level $f^k_{\rm lev} := f^{k(l)}_{\rm rec} - \delta_l$ within the ball of a certain *radius* R_l centered at $x^{k(l)}_{\rm rec}$, where the *level gap* $\delta_l > 0$ controls the stepsizes (10). If *sufficient descent* $f(x^k) \le f^{k(l)}_{\rm rec} - \frac{1}{2}\delta_l$ occurs for some $k > k(l)$ (i.e., at least half of the desired objective reduction δ_l is achieved), the next group $l+1$ starts with the same gap $\delta_{l+1} := \delta_l$ and radius $R_{l+1} := R_l$. Otherwise, the method eventually discovers that the target is infeasible in the sense that

$$f^k_{\rm lev} := f^{k(l)}_{\rm rec} - \delta_l < \min\{\, f(x) : x \in B(x^{k(l)}, R_l) \cap S \,\}. \tag{15}$$

Our test for detecting (15) (see (17) below) was derived in Kiwiel et al. [26] via fairly complicated geometric arguments; we only sketch the main idea because a much simpler validation of this test will be given in §3. Suppose (15) does not hold: $f(x) \le f^k_{\rm lev}$ for some $x \in B(x^{k(l)}, R_l) \cap S$. Let $t_k \equiv 1$. Viewing the iteration (8) as a subgradient step $x^{k+1/2} := P_{H_k} x^k$ followed by a projection step $x^{k+1} := P_S x^{k+1/2}$, simple estimates show that the sum of squares of these steps $\rho_{k+1} := \sum_{j=k(l)}^k (|x^{j+1/2} - x^j|^2 + |x^{j+1} - x^{j+1/2}|^2)$

satisfies $\rho_{k+1} \le |x^{k(l)} - x|^2 - |x^{k+1} - x|^2 \le R_l^2$, because by (7), the sets on which projections occur have a common point x. Thus, the inequality $\rho_{k+1} > R_l^2$ implies (15). Intuitively, if (15) holds, then oscillations in successive projections eventually produce $\rho_{k+1} > R_l^2$; the weaker test (17) below may detect (15) even sooner. Then the next group $l+1$ starts with a contracted gap $\delta_{l+1} := \frac{1}{2}\delta_l$ and a shrinked radius $R_{l+1} := R_l/2^\beta$, where $\beta \in [0,1)$ is a parameter (typically $\beta = \frac{1}{2}$).

We now state a detailed description of our method. Further comments on its rules are given below and in §3; also see Kiwiel et al. [26] for additional motivations.

ALGORITHM 2.1 (ballstep level method).

STEP 0 (*Initialization*). Select an initial point $x^1 \in S$, a level gap $\delta_1 > 0$, ballstep parameters $R > 0$, $\beta \in [0,1)$, and relaxation bounds $t_{\min}$, $t_{\max}$ (cf. (9)). Set $f_{\text{rec}}^0 := \infty$, $\rho_1 := 0$. Set the counters $k := l := k(1) := 1$ ($k(l)$ is the iteration number of the lth change of f_{lev}^k).

STEP 1 (*Objective evaluation*). Calculate $f(x^k)$ and $g_f(x^k)$. If $f(x^k) < f_{\text{rec}}^{k-1}$, set $f_{\text{rec}}^k := f(x^k)$ and $x_{\text{rec}}^k := x^k$, else set $f_{\text{rec}}^k := f_{\text{rec}}^{k-1}$ and $x_{\text{rec}}^k := x_{\text{rec}}^{k-1}$ (so that $f(x_{\text{rec}}^k) = \min_{j=1}^k f(x^j)$).

STEP 2 (*Stopping criterion*). If $g_f^k := g_f(x^k) = 0$, terminate ($x^k \in S_*$).

STEP 3 (*Sufficient descent detection*). If $f(x^k) \le f_{\text{rec}}^{k(l)} - \frac{1}{2}\delta_l$, start the next group: set $k(l+1) := k$, $\delta_{l+1} := \delta_l$, $\rho_k := 0$ and increase the group counter l by 1.

STEP 4 (*Projections*). Set the level $f_{\text{lev}}^k := f_{\text{rec}}^{k(l)} - \delta_l$. Choose the relaxation factor $t_k \in T$ (cf. (9)). Set

$$x^{k+1/2} := x^k + t_k(P_{H_k}x^k - x^k), \quad \check\rho_k := t_k(2 - t_k)d_{H_k}^2(x^k), \quad \rho_{k+1/2} := \rho_k + \check\rho_k, \tag{16a}$$

$$x^{k+1} := P_S x^{k+1/2}, \quad \check\rho_{k+1/2} := |x^{k+1} - x^{k+1/2}|^2, \quad \rho_{k+1} := \rho_{k+1/2} + \check\rho_{k+1/2}. \tag{16b}$$

STEP 5 (*Target infeasibility detection*). Set the ball radius $R_l := R(\delta_l/\delta_1)^\beta$. If

$$(R_l - |x^{k+1} - x^{k(l)}|)^2 > R_l^2 - \rho_{k+1}, \tag{17}$$

i.e., the target level is too low, then go to Step 6; otherwise, increase k by 1 and go to Step 1.

STEP 6 (*Level increase*). Start the next group: set $k(l+1) := k$, $\delta_{l+1} := \frac{1}{2}\delta_l$, $\rho_k := 0$, replace x^k by x_{rec}^k and g_f^k by $g_f(x_{\text{rec}}^k)$, increase the group counter l by 1 and go to Step 4.

Assuming the method doesn't terminate, we now recall some results of Kiwiel et al. [26, §2–3].

REMARKS 2.1 (i) If group $l+1$ starts at Step 3, then $f_{\text{rec}}^{k(l+1)} \le f_{\text{rec}}^{k(l)} - \frac{1}{2}\delta_l$ and $x^{k(l+1)} = x_{\text{rec}}^{k(l+1)}$ (since $f(x^j) > f_{\text{rec}}^{k(l)} - \frac{1}{2}\delta_l$ for $j < k$). Thus, by the rules Step 6, at Step 4 we have $x^{k(l)} = x_{\text{rec}}^{k(l)} \in S$ and $f_{\text{rec}}^{k(l)} = f(x^{k(l)})$ for all l.

(ii) At Step 4, in view of (5) and (6) with $f_k(x^k) = f(x^k) > f_{\text{lev}}^k$, we have $x^{k+1/2} = x^k - \nu_k g_f^k$ by (10), and $d_{H_k}(x^k) = [f_k(x^k) - f_{\text{lev}}^k]/|g_f^k|$. Hence the *Fejér* quantities $\check\rho_k$, $\rho_{k+1/2}$ and ρ_{k+1} are positive (because ρ_k is set to zero at Steps 0, 3 and 6). The rôle of these quantities will be explained in §3.

(iii) At Step 5, the *ball radius* $R_l := R(\delta_l/\delta_1)^\beta \le R$ is nonincreasing. Ideally, R_l should be of order $d_{S_*}(x^{k(l)})$, and hence shrink as the ball center $x^{k(l)}$ approaches the optimal set S_*. As shown by Kiwiel et al. [26, Rem. 3.9(i)], for convergence it suffices to choose R_l so that $\delta_l/R_l \to 0$; our results will additionally require boundedness of the sequence $\{R_l\}$. This makes room for other choices of R_l.

(iv) By Kiwiel et al. [26, Lem. 3.1(v)] or Lemma 3.1(iv,v) below, the *Fejér test* (17) discovers that the target is infeasible in the sense of (15). Then the gap δ_l is halved at Step 6, the target f_{lev}^k is increased at Step 4 and the candidate point x^{k+1} is recomputed. Note that the group counter l increases at Step 6, but the iteration counter k does not, so relations like $f_{\text{lev}}^k := f_{\text{rec}}^{k(l)} - \delta_l$ always involve the *current values* of k and l at Step 4.

(v) Notice that if $|x^{k+1} - x^{k(l)}| > 2R_l$, then the Fejér test (17) is passed. It follows that at Step 1 we have the basic *local boundedness* property: $\{x^k\}_{k=k(l)}^{k(l+1)} \subset B(x^{k(l)}, 2R_l)$.

We shall need the following convergence properties of Algorithm 2.1, which follow from the analysis of Kiwiel et al. [26, §3] and our standing assumption that the optimal set S_* of problem (1) is nonempty.

THEOREM 2.1 *We have $f(x^{k(l)}) \downarrow f_*$, $\delta_l \downarrow 0$, and each cluster point of the sequence $\{x^{k(l)}\}$ (if any) lies in the optimal set S_* of problem* (1). *Moreover, the sequence $\{x^{k(l)}\}$ is bounded if the optimal set S_* is bounded. These results require only finiteness of the objective f and local boundedness of the subgradient mapping g_f on the feasible set S (in which case f is continuous on S).*

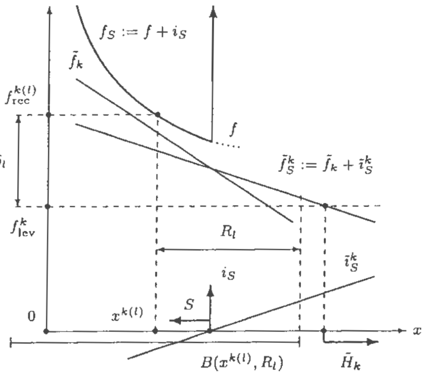

Figure 1: Target infeasibility $f_{\rm lev}^k < \min_{B(x^{k(l)},R_l)} f_S$ if $d_{\bar H_k}(x^{k(l)}) > R_l$.

PROOF. The first assertion follows from the results of Kiwiel et al. [26, Lemma 3.6 and Theorem 3.7], the second one from [26, Corollary 3.8], and the third one from [26, Remark 3.9(ii)]. □

3. Dual subgradient interpretations. For theoretical purposes, it is convenient to regard our constrained problem $f_* := \min_S f$ of (1) as the unconstrained problem $f_* = \min f_S$ with the *essential objective*

$$f_S := f + i_S, \tag{18}$$

where i_S is the *indicator function* of the feasible set S ($i_S(x) = 0$ if $x \in S$, ∞ if $x \notin S$). Clearly, the objective f_S is convex. Let $\mathcal{N}_S := \partial i_S$ denote the *normal cone* operator of the feasible set S.

We now outline our main results. At each iteration, Step 1 delivers the linearization f_k (cf. (5)) of the objective f, whereas at Step 4, the projection $x^{k+1} := P_S x^{k+1/2}$ gives rise to a subgradient linearization of the constraint function i_S at x^{k+1}. At iteration k, we construct affine minorants $\bar f_k$ and $\bar\imath_S^k$ of the functions f and i_S by combining their past subgradient linearizations with suitable weights. Then the function $\bar f_S^k := \bar f_k + \bar\imath_S^k$ is an affine minorant of $f_S := f + i_S$, and hence its halfspace $\bar H_k := \mathcal{L}_{\bar f_S^k}(f_{\rm lev}^k)$ contains the level set $\mathcal{L}_{f_S}(f_{\rm lev}^k)$. Now, in terms of the minimum ball value $f_*^l := \min_{B(x^{k(l)},R_l)} f_S$, condition (15) reads $f_{\rm lev}^k < f_*^l$. It follows that $f_{\rm lev}^k < f_*^l$ if $B(x^{k(l)}, R_l) \cap \bar H_k = \emptyset$ (see Figure 1); the latter condition is shown to be equivalent to the Fejér test (17) by fairly simple algebra. Next, when this condition holds, we get the inclusion $\nabla \bar f_S^k \in \partial_{\delta_l} f_S(x^{k(l)})$ and the bound $|\nabla \bar f_S^k| \le \delta_l/R_l$ as in Figure 1; since $\delta_l \to 0$ and $\delta_l/R_l \to 0$, these relations ensure asymptotic optimality and suggest practical stopping criteria.

3.1 Aggregate linearizations. We first derive a dual interpretation of the Fejér test (17) by identifying below affine minorants $\bar f_k$, $\bar\imath_S^k$, $\bar f_S^k$ of the functions f, i_S, f_S, respectively. As mentioned earlier, $\bar f_k$ is obtained by combining the subgradient linearizations f_j of (5) with the convex weights ν_j^k of (13), i.e., the stepsizes ν_j of (10) divided by the *cumulative stepsize* $\bar\nu_f^k := \sum_{j=k(l)}^k \nu_j$ so that $\sum_{j=k(l)}^k \nu_j^k = 1$. For aggregating constraint information, we shall use the fact that at Step 4, the vector

$$g_S^k := x^{k+1/2} - x^{k+1} \tag{19}$$

is a subgradient of i_S at x^{k+1} stemming from the construction of $x^{k+1} := P_S x^{k+1/2}$. Accordingly, we shall employ the following *aggregate linearizations* of f, i_S and f_S (cf. (18)):

$$\bar f_k(\cdot) := \sum_{j=k(l)}^k \nu_j^k f_j(\cdot), \quad \bar\imath_S^k(\cdot) := \sum_{j=k(l)}^k \langle g_S^j, \cdot - x^{j+1}\rangle/\bar\nu_f^k, \quad \bar f_S^k(\cdot) := \bar f_k(\cdot) + \bar\imath_S^k(\cdot), \tag{20}$$

and the corresponding *aggregate halfspace* $\bar H_k$ of $\bar f_S^k$ and the *aggregate level* $\bar f_{\rm lev}^k$ given by

$$\bar H_k := \mathcal{L}_{\bar f_S^k}(\bar f_{\rm lev}^k) = \{\, x : \bar f_S^k(x) \le \bar f_{\rm lev}^k \,\} \quad \text{with} \quad \bar f_{\rm lev}^k := \sum_{j=k(l)}^k \nu_j^k f_{\rm lev}^j. \tag{21}$$

The following technical result lists their basic properties, which are commented upon below.

LEMMA 3.1 (i) *At Step 4, the point* x^{k+1} *and the Fejér sum* ρ_{k+1} *satisfy*

$$x^{k+1} - x^{k(l)} = -\sum_{j=k(l)}^{k} (\nu_j g_f^j + g_S^j), \tag{22}$$

$$L_k := -\tfrac{1}{2}\Big|\sum_{j=k(l)}^{k} \nu_j g_f^j + g_S^j\Big|^2 + \sum_{j=k(l)}^{k} \Big\{\nu_j[f_j(x^{k(l)}) - f_{\rm lev}^j] + \langle g_S^j, x^{k(l)} - x^{j+1}\rangle\Big\} = \tfrac{1}{2}\rho_{k+1}. \tag{23}$$

(ii) *The aggregate linearizations satisfy* $\tilde f_k \le f$, $\tilde\imath_S^k \le i_S$, $\tilde f_S^k \le f_S$. *Further,* $\tilde\nu_f^k \nabla \tilde f_S^k = x^{k(l)} - x^{k+1}$,

$$2\tilde\nu_f^k[\tilde f_S^k(x^{k(l)}) - \tilde f_{\rm lev}^k] = |x^{k+1} - x^{k(l)}|^2 + \rho_{k+1}. \tag{24}$$

(iii) *We have* $\tilde f_S^k(x^{k(l)}) > \tilde f_{\rm lev}^k$, *and the distance from the point* $x^{k(l)}$ *to the halfspace of* (21) *satisfies*

$$d_{\tilde H_k}(x^{k(l)}) = [\tilde f_S^k(x^{k(l)}) - \tilde f_{\rm lev}^k]/|\nabla \tilde f_S^k| \ge \rho_{k+1}^{1/2}. \tag{25}$$

(iv) *For the minimum ball value* $f_*^l := \min_{B(x^{k(l)},R_l)} f_S$, *we have the following. If* $\tilde f_{\rm lev}^k \ge f_*^l$, *then* $d_{\tilde H_k}(x^{k(l)}) \le R_l$. *Consequently,* $\tilde f_{\rm lev}^k < f_*^l$ *if* $d_{\tilde H_k}(x^{k(l)}) > R_l$.

(v) $(R_l - |x^{k+1} - x^{k(l)}|)^2 > R_l^2 - \rho_{k+1}$ *(i.e., the Fejér test* (17) *is true) iff* $d_{\tilde H_k}(x^{k(l)}) > R_l$.

PROOF. (i) Since $x^{k+1/2} - x^k = -\nu_k g_f^k$ by Remark 2.1(ii), and $x^{k+1} - x^{k+1/2} = -g_S^k$ by (19), summing gives (22). Let $\Delta L_k := L_k - L_{k-1}$. Since by (22), $x^k - x^{k(l)} = -\sum_{j=k(l)}^{k-1}(\nu_j g_f^j + g_S^j)$ in (23), we have

$$\begin{aligned}
\Delta L_k &= -\tfrac{1}{2}|\nu_k g_f^k + g_S^k|^2 + \langle \nu_k g_f^k + g_S^k, x^k - x^{k(l)}\rangle + \nu_k[f_k(x^{k(l)}) - f_{\rm lev}^k] + \langle g_S^k, x^{k(l)} - x^{k+1}\rangle \\
&= -\tfrac{1}{2}|\nu_k g_f^k|^2 + \nu_k[f_k(x^{k(l)}) + \langle g_f^k, x^k - x^{k(l)}\rangle - f_{\rm lev}^k] + \langle g_S^k, x^k - x^{k+1} - \nu_k g_f^k - \tfrac{1}{2} g_S^k\rangle \\
&= -\tfrac{1}{2}|\nu_k g_f^k|^2 + \nu_k[f_k(x^k) - f_{\rm lev}^k] + \langle g_S^k, x^{k+1/2} - x^{k+1} - \tfrac{1}{2} g_S^k\rangle \\
&= (-\tfrac{1}{2}t_k^2 + t_k)\{[f_k(x^k) - f_{\rm lev}^k]/|g_f^k|\}^2 \\
&\quad + \langle x^{k+1/2} - x^{k+1}, x^{k+1/2} - x^{k+1} - \tfrac{1}{2}(x^{k+1/2} - x^{k+1})\rangle \\
&= \tfrac{1}{2}\{t_k(2-t_k)d_{H_k}^2(x^k) + |x^{k+1} - x^{k+1/2}|^2\} = \tfrac{1}{2}(\check\rho_k + \check\rho_{k+1/2}) = \tfrac{1}{2}(\rho_{k+1} - \rho_k),
\end{aligned}$$

where the first equality follows from expansion of L_k, the third one from the definition (5) of f_k and the fact that $x^{k+1/2} = x^k - \nu_k g_f^k$, the fonrth one from the definitions (10) of ν_k and (19) of g_S^k, the fifth one from the fact that $d_{H_k}(x^k) = [f_k(x^k) - f_{\rm lev}^k]/|g_f^k|$, and the final two ones from (16). Consequently, (23) can be obtained by induction, starting from $L_{k(l)-1} := \rho_{k(l)} := 0$ (cf. Steps 0, 3 and 6).

(ii) Combining the subgradient inequalities $f_j \le f$ of (5) in (20) gives $\tilde f_k \le f$. Next, since $g_S^j := x^{j+1/2} - x^{j+1}$ by (19) and $x^{j+1} := P_S x^{j+1/2}$ by Step 4, using the well-known projection property

$$\langle g_S^j, x - x^{j+1}\rangle = \langle x^{j+1/2} - P_S x^{j+1/2}, x - P_S x^{j+1/2}\rangle \le 0 \quad \forall x \in S$$

gives $\tilde\imath_S^k \le i_S$ in (20) by summing, and hence $\tilde f_S^k := \tilde f_k + \tilde\imath_S^k \le f + i_S =: f_S$. Now, using the definitions (13) and (20) yields $\tilde\nu_f^k \nabla \tilde f_S^k = \sum_{j=k(l)}^{k}(\nu_j g_f^j + g_S^j) = x^{k(l)} - x^{k+1}$ by (22), as well as, by (21),

$$\tilde\nu_f^k[\tilde f_S^k(x^{k(l)}) - \tilde f_{\rm lev}^k] = \sum_{j=k(l)}^{k} \Big\{\nu_j[f_j(x^{k(l)}) - f_{\rm lev}^j] + \langle g_S^j, x^{k(l)} - x^{j+1}\rangle\Big\}.$$

These two expressions allow us to rewrite (23) in the following useful form

$$L_k = -\tfrac{1}{2}|\tilde\nu_f^k \nabla \tilde f_S^k|^2 + \tilde\nu_f^k[\tilde f_S^k(x^{k(l)}) - \tilde f_{\rm lev}^k] = \tfrac{1}{2}\rho_{k+1} > 0, \tag{26}$$

where $\rho_{k+1} > 0$ by Remark 2.1(ii); then (24) follows from (26), where $\tilde\nu_f^k \nabla \tilde f_S^k = x^{k(l)} - x^{k+1}$.

(iii) By (26), $L_k = -\tfrac{1}{2}a^2 + b = \tfrac{1}{2}c^2$ with $a := |\tilde\nu_f^k \nabla \tilde f_S^k|$, $b := \tilde\nu_f^k[\tilde f_S^k(x^{k(l)}) - \tilde f_{\rm lev}^k]$, $c := \rho_{k+1}^{1/2} > 0$. Then $b = \tfrac{1}{2}(a^2 + c^2) \ge |ac|$, so that by the definition of $\tilde H_k$ in (21), $d_{\tilde H_k}(x^{k(l)}) = b/a \ge c$ implies (25).

(iv) Consider any point $x \in \operatorname{Arg\,min}_{B(x^{k(l)},R_l)} f_S$. If $f_*^l \le \tilde f_{\text{lev}}^k$, then $x \in \tilde H_k$ by (21), because $f_S(x) = f_*^l$ and $\tilde f_S^k \le f_S$ by statement (ii). Together with $x \in B(x^{k(l)}, R_l)$, this implies that $d_{\tilde H_k}(x^{k(l)}) \le R_l$.

(v) $(R_l - |x^{k+1} - x^{k(l)}|)^2 > R_l^2 - \rho_{k+1} \Leftrightarrow |x^{k+1} - x^{k(l)}|^2 + \rho_{k+1} > 2R_l|x^{k+1} - x^{k(l)}| \Leftrightarrow 2\tilde\nu_f^k[\tilde f_S^k(x^{k(l)}) - \tilde f_{\text{lev}}^k] > 2R_l\tilde\nu_f^k|\nabla \tilde f_S^k| \Leftrightarrow [\tilde f_S^k(x^{k(l)}) - \tilde f_{\text{lev}}^k]/|\nabla \tilde f_S^k| > R_l \Leftrightarrow d_{\tilde H_k}(x^{k(l)}) > R_l$, where we have used (24), the fact that $|x^{k+1} - x^{k(l)}| = \tilde\nu_f^k|\nabla f_S^k|$ by statement (ii), and (25). □

REMARKS 3.1 (i) By Lemma 3.1(v), the Fejér test (17) is *equivalent* to the *distance test*

$$d_{\tilde H_k}(x^{k(l)}) > R_l. \tag{27}$$

The fact that the Fejér test (17) implies $f_{\text{lev}}^k < f_*^l$ (cf. (15)) was derived in Kiwiel et al. [26, Lem. 3.1(v)] from Fejér estimates via analytic arguments, which are quite difficult to interpret. In contrast, the distance test (27) has a straightforward interpretation: with $\tilde f_{\text{lev}}^k = f_{\text{lev}}^k$ in (21), (27) means that the minimum of the linearization $\tilde f_S^k$ over the ball $B(x^{k(l)}, R_l)$, and hence also that of f_S (since $\tilde f_S^k$ underestimates f_S), is greater than f_{lev}^k, i.e., $f_{\text{lev}}^k < \min_{B(x^{k(l)},R_l)} \tilde f_S^k \le \min_{B(x^{k(l)},R_l)} f_S =: f_*^l$ (cf. Fig. 1).

(ii) To cover the modifications of Kiwiel et al. [26, §6], which need not use constant levels $f_{\text{lev}}^j = f_{\text{lev}}^k$ for $j = k(l){:}\,k$, note that the proof of Lemma 3.1 holds if at Step 4, for all k, we only have

$$f_{\text{rec}}^{k(l)} - \delta_l \le f_{\text{lev}}^k < \min\{f_{\text{rec}}^{k(l)}, f(x^k)\}. \tag{28}$$

In general, since $\tilde f_{\text{lev}}^k \ge \min_{j=k(l)}^k f_{\text{lev}}^j$ by (21) and (13), if we have $\min_{j=k(l)}^k f_{\text{lev}}^j \ge f_{\text{rec}}^{k(l)} - \delta_l$, then (27) yields $f_{\text{rec}}^{k(l)} - \delta_l < f_*^l$. It follows that Lemma 3.1(iv,v) subsumes the corresponding result of Kiwiel et al. [26, Lem. 3.1(v)], and hence that the level condition (28) suffices for our convergence results.

(iii) Suppose momentarily that $S = \mathbb{R}^n$, so that $g_S^k \equiv 0$. It is instructive to observe that our algorithm acts like a dual coordinate ascent method for the QP subproblem

$$\min\left\{\tfrac12|x - x^{k(l)}|^2 : f_j(x) \equiv f_j(x^{k(l)}) + \langle g_f^j, x - x^{k(l)}\rangle \le f_{\text{lev}}^j,\ j = k(l){:}\,k\right\}. \tag{29}$$

Indeed, the Lagrangian of (29) with multipliers ν_j is minimized by the point x^{k+1} (cf. (22)) to give the dual function value L_k of (23), and $\nu_k = t_k\check\nu_k$ by (10), where $\check\nu_k := [f_k(x^k) - f_{\text{lev}}^k]/|g_f^k|^2$ maximizes $\Delta L_k = -\frac12|\nu_k g_f^k|^2 + \nu_k[f_k(x^k) - f_{\text{lev}}^k]$ (see the proof of Lemma 3.1(i)). Thus our algorithm may be regarded as a poor man's version of the proximal level methods of Kiwiel [21] and Lemaréchal et al. [33], which employ subproblem (29) with $f_{\text{lev}}^j = f_{\text{lev}}^k$ for all j.

3.2 An optimality estimate. We now derive an optimality estimate from the aggregate linearizations $\tilde f_k$, $\tilde\imath_S^k$ and $\tilde f_S^k$ defined in (20). These linearizations are described by their constant gradients, as well as their *linearization errors* at the current ball center $x^{k(l)}$ (cf. Fig. 1):

$$\tilde\epsilon_f^k := f(x^{k(l)}) - \tilde f_k(x^{k(l)}),\quad \tilde\epsilon_S^k := -\tilde\imath_S^k(x^{k(l)}),\quad \tilde\epsilon_k := f(x^{k(l)}) - \tilde f_S^k(x^{k(l)}); \tag{30}$$

note that $i_S(x^{k(l)}) = 0$ and $f_S(x^{k(l)}) = f(x^{k(l)})$ from $x^{k(l)} \in S$. In view of Remark 3.1(ii), from now on we assume only that the level condition (28) holds at Step 4 for all k.

LEMMA 3.2 *The linearization errors of* (30) *are nonnegative, with $\tilde\epsilon_k = \tilde\epsilon_f^k + \tilde\epsilon_S^k$, and we have*

$$\nabla\tilde f_k \in \partial_{\tilde\epsilon_f^k} f(x^{k(l)}),\quad \nabla\tilde\imath_S^k \in \partial_{\tilde\epsilon_S^k} i_S(x^{k(l)}),\quad \nabla\tilde f_S^k \in \partial_{\tilde\epsilon_k} f_S(x^{k(l)}).$$

Further,

$$f_S(\cdot) \ge \tilde f_S^k(\cdot) = f(x^{k(l)}) - \tilde\epsilon_k + \langle\nabla\tilde f_S^k, \cdot - x^{k(l)}\rangle, \tag{31}$$

where

$$\tilde\epsilon_k := f(x^{k(l)}) - \tilde f_S^k(x^{k(l)}) < f_{\text{rec}}^{k(l)} - \tilde f_{\text{lev}}^k \le \delta_l, \tag{32}$$

$$|\nabla\tilde f_S^k| = [\tilde f_S^k(x^{k(l)}) - \tilde f_{\text{lev}}^k]/d_{\tilde H_k}(x^{k(l)}) \le \delta_l/d_{\tilde H_k}(x^{k(l)}). \tag{33}$$

PROOF. By Lemma 3.1(ii), $\tilde f_k$ is an affine minorant of f; thus, by (30), the inequality

$$f(\cdot) \ge \tilde f_k(\cdot) = \tilde f_k(x^{k(l)}) + \langle\nabla\tilde f_k, \cdot - x^{k(l)}\rangle = f(x^{k(l)}) - \tilde\epsilon_f^k + \langle\nabla\tilde f_k, \cdot - x^{k(l)}\rangle$$

means that $\nabla\tilde f_k \in \partial_{\tilde\epsilon_f^k} f(x^{k(l)})$ with $\tilde\epsilon_f^k \ge 0$. Arguing similarly for $\tilde\imath_S^k$ and $\tilde f_S^k$ yields the first assertion and (31). The inequalities in (32) stem from the facts that $f(x^{k(l)}) = f_{\text{rec}}^{k(l)}$ by Remark 2.1(i), $\tilde f_S^k(x^{k(l)}) > \tilde f_{\text{lev}}^k$ by Lemma 3.1(iii), $\tilde f_{\text{lev}}^k \ge \min_{j=k(l)}^k f_{\text{lev}}^j$ by (21) and (13), and $\min_{j=k(l)}^k f_{\text{lev}}^j \ge f_{\text{rec}}^{k(l)} - \delta_l$ by condition (28) used at iterations $j = k(l){:}\,k$. Then the equality in (33) follows from (25), and the inequality from the fact that $\tilde f_S^k(x^{k(l)}) \le f_S(x^{k(l)}) = f_{\text{rec}}^{k(l)}$ (by Remark 2.1(i)) and the last inequality of (32). □

3.3 Ballstep modifications. We now consider two more efficient modifications of Kiwiel et al. [26].

To detect that $\min_{j=k(l)}^k f_{\text{lev}}^j < f_*^l$ more quickly, Step 5 may use the additional test

$$(R_l - |x^{k+1/2} - x^{k(l)}|)^2 > R_l^2 - \rho_{k+1/2}, \tag{34}$$

replacing (17) by "(34) or (17)". In view of the results of Kiwiel et al. [26, §3], Step 4 may set $x^{k+1} := x^{k+1/2}$ if condition (34) holds, so that $\rho_{k+1} = \rho_{k+1/2}$ and (17) holds; then all the preceding and subsequent results remain valid. Further, we may replace $x^{k+1/2}$ and $\rho_{k+1/2}$ in (34) by $P_{H_k}x^k$ and $\rho_k + d_{H_k}^2(x^k)$, as if $t_k = 1$; see Kiwiel et al. [26, Rem. 3.2(ii)].

Similarly, our preceding and subsequent results hold for the "true" ballstep version of Kiwiel et al. [26, Lem. 3.10], which additionally projects the point x^{k+1} on the ball $B(x^{k(l)}, R_l)$ to ensure that $\{x^k\}_{k=k(l)}^{k(l+1)} \subset B(x^{k(l)}, R_l)$ (instead of $\{x^k\}_{k=k(l)}^{k(l+1)} \subset B(x^{k(l)}, 2R_l)$ as before). Since this only needs more complicated notation, we refer the interested readers to Kiwiel et al. [25, Lem. 3.10].

4. Optimal objective and constraint subgradients. Our asymptotic convergence results will deal *exclusively* with relations holding at Step 6, using groups and iterations in the sets

$$L := \{\, l : \delta_{l+1} = \tfrac{1}{2}\delta_l \,\} \quad \text{and} \quad K := \{\, k(l+1) : l \in L \,\}. \tag{35}$$

The set L indexes groups l terminating at Step 6 when the distance test (27) ($\equiv$(17) by Remark 3.1(i)) holds at Step 5 for the current iteration $k = k(l+1)$ in the set of "interesting" iterations K. Of course, it would be nice to have results for the remaining iterations as well, but our estimate (33) involves the quantity $\delta_l / d_{\tilde{H}_k}(x^{k(l)})$, which in general converges to 0 only for $k = k(l+1) \in K$, as will be seen below.

We now begin our study of asymptotic properties of the aggregate linearizations $\tilde{f}_k$, $\tilde{\imath}_S^k$, $\tilde{f}_S^k$ of (20). First, we show that their errors $\tilde{\epsilon}_f^k$, $\tilde{\epsilon}_S^k$, $\tilde{\epsilon}_k$ (cf. (30)), as well as the gradient of $\tilde{f}_S^k$, vanish asymptotically for $k \in K$. Our further results will require local boundedness of the gradient of $\tilde{f}_k$. Since this gradient $\nabla \tilde{f}_k$ is a convex combination of the past subgradients $\{g_f^j\}_{j=k(l)}^k$ (cf. (20), (13) and (5)), its local boundedness will follow from the local boundedness of the subgradient mapping g_f.

Lemma 4.1 (i) *In the notation of* (30), (20) *and* (35), *we have*

$$\tilde{\epsilon}_f^k \to 0, \quad \tilde{\epsilon}_S^k \to 0, \quad \tilde{\epsilon}_k = \tilde{\epsilon}_f^k + \tilde{\epsilon}_S^k \to 0 \quad \textit{and} \quad \nabla \tilde{f}_S^k = \nabla \tilde{f}_k + \nabla \tilde{\imath}_S^k \xrightarrow{K} 0.$$

(ii) *Suppose the sequence $\{x^{k(l)}\}_{l \in L}$ has a cluster point x^∞. Let $L' \subset L$ be such that $x^{k(l)} \xrightarrow{L'} x^\infty$, and let $K' := \{k(l+1) : l \in L'\}$ (cf. (35)). Then $x^\infty \in S_*$ and $f(x^{k(l)}) \downarrow f_* = f(x^\infty)$. Moreover, the sequences $\{x^k\}_{k \in K'_l, l \in L'}$ and $\{g_f^k\}_{k \in K'_l, l \in L'}$ are bounded, where $K'_l := \{k(l) : k(l+1)\}$.*

Proof. (i) We have $0 \le \tilde{\epsilon}_f^k, \tilde{\epsilon}_S^k, \tilde{\epsilon}_k \le \delta_l$ by Lemma 3.2 (cf. (32)), where $\delta_l \downarrow 0$ by Theorem 2.1. Next, we have $|\nabla \tilde{f}_S^k| \le \delta_l / d_{\tilde{H}_k}(x^{k(l)})$ by (33) with $d_{\tilde{H}_k}(x^{k(l)}) > R_l$ for $k \in K$ (see below (35)), $R_l := R(\delta_l/\delta_1)^\beta$ by Step 5 and $\beta \in [0,1)$ by Step 0; consequently, we obtain that $\delta_l / R_l \to 0$ and hence $\nabla \tilde{f}_S^k \xrightarrow{K} 0$.

(ii) Of course, $x^\infty \in S_*$ by Theorem 2.1, but the estimate (31) combined with statement (i) and the fact that the sequence $\{x^k\}$ lies in the closed set S on which f is continuous provide an independent verification: $f_S(\cdot) \ge f_S(x^\infty)$. The final assertion follows from the inclusion $\{x^k\}_{k=k(l)}^{k(l+1)} \subset B(x^{k(l)}, 2R_l)$ of Remark 2.1(v), since $g_f^k := g_f(x^k)$ for all k and the mapping g_f is locally bounded on the set S. □

In the asymptotic setting of Lemma 4.1, let x^∞ be an arbitrary cluster point of the sequence $\{x^{k(l)}\}_{l \in L}$ corresponding to groups L' and iterations K' such that (cf. (35))

$$x^{k(l)} \xrightarrow{L'} x^\infty \quad \text{with} \quad L' \subset L := \{\, l : \delta_{l+1} = \tfrac{1}{2}\delta_l \,\},\ K' := \{k(l+1) : l \in L'\} \subset K; \tag{36}$$

note that $x^\infty \in S_*$ by Theorem 2.1. We now show that the corresponding subsequence of the *aggregate subgradients* $\nabla \tilde{f}_k$ converges to the *optimal subgradient set* of our problem $\min_S f$:

$$\mathcal{G} := \partial f(x^\infty) \cap -\mathcal{N}_S(x^\infty). \tag{37}$$

This set *does not depend* on the point x^∞, as long as $x^\infty \in S_*$: $\mathcal{G} = \partial f(x) \cap -\mathcal{N}_S(x)\ \forall x \in S_*$ by Burke and Ferris [10, Lem. 2], and it is closed convex (such are the sets $\partial f(x^\infty)$ and $\mathcal{N}_S(x^\infty) := \partial i_S(x^\infty)$).

THEOREM 4.1 *Suppose the sequence $\{x^{k(l)}\}_{l\in L}$ has a cluster point x^∞. Let $L' \subset L$ be such that $x^{k(l)} \xrightarrow{L'} x^\infty$, and let $K' := \{k(l+1) : l \in L'\}$ (cf. (35)). Then we have the following statements.*

(i) *The sequence $\{\nabla\tilde{f}_k\}_{k\in K'}$ is bounded and its cluster points lie in the subdifferential $\partial f(x^\infty)$.*

(ii) *Every cluster point of the sequence $\{\nabla\tilde{f}_k\}_{k\in K'}$ lies in the optimal subgradient set $\mathcal{G}$ of (37).*

(iii) $d_{\mathcal{G}}(\nabla\tilde{f}_k) \xrightarrow{K'} 0$, *i.e., the sequence $\{\nabla\tilde{f}_k\}_{k\in K'}$ converges to the optimal subgradient set $\mathcal{G}$.*

PROOF. (i) Since $\nabla\tilde{f}_k \in \operatorname{co}\{g_f^j\}_{j=k(l)}^k$ by (13) and (20), the sequence $\{\nabla\tilde{f}_k\}_{k\in K'}$ is bounded by Lemma 4.1(ii). Next, since $\nabla\tilde{f}_k \in \partial_{\tilde{\epsilon}_f^k} f(x^{k(l)})$ by Lemma 3.2, where $x^{k(l)} \xrightarrow{L'} x^\infty$ and $\tilde{\epsilon}_f^k \xrightarrow{K'} 0$ by Lemma 4.1(i), we see that each cluster point of the sequence $\{\nabla\tilde{f}_k\}_{k\in K'}$ lies in $\partial f(x^\infty)$, since the mapping $(x,\epsilon) \mapsto \partial_\epsilon f(x)$ is closed on $S \times \mathbb{R}_+$; see, e.g., Hiriart-Urruty and Lemaréchal [19, §XI.4.1].

(ii) Let $K'' \subset K'$ be such that the sequence $\{\nabla\tilde{f}_k\}_{k\in K''}$ has a limit $\nabla\tilde{f}_\infty$. By statement (i), $\nabla\tilde{f}_\infty \in \partial f(x^\infty)$. Since $\nabla\tilde{f}_S^k - \nabla\tilde{f}_k = \nabla\tilde{i}_S^k \in \partial_{\tilde{\epsilon}_S^k} i_S(x^{k(l)})$ (by (20) and Lemma 3.2) with $\nabla\tilde{f}_S^k \xrightarrow{K} 0$ and $\tilde{\epsilon}_S^k \to 0$ by Lemma 4.1(i), we see that $\nabla\tilde{i}_S^k \xrightarrow{K''} -\nabla\tilde{f}_\infty \in \partial i_S(x^\infty)$ by the closedness of $\partial_\epsilon i_S(x)$ as above.

(iii) This follows from statements (i), (ii) and the continuity of the distance function $d_{\mathcal{G}}$: pick $K'' \subset K'$ such that $d_{\mathcal{G}}(\nabla\tilde{f}_k) \xrightarrow{K''} \overline{\lim}_{k\in K'} d_{\mathcal{G}}(\nabla\tilde{f}_k)$ and $\nabla\tilde{f}_k \xrightarrow{K''} \nabla\tilde{f}_\infty \in \mathcal{G}$ to get $d_{\mathcal{G}}(\nabla\tilde{f}_k) \xrightarrow{K''} 0$. □

COROLLARY 4.1 *If the sequence $\{x^{k(l)}\}$ is bounded (e.g., the optimal set S_* is bounded), then the sequence $\{\nabla\tilde{f}_k\}_{k\in K}$ is bounded (cf. (35)), its cluster points lie in the optimal subgradient set $\mathcal{G}$ defined by (37) (for any point $x^\infty \in S_*$), and it converges to this set $\mathcal{G}$, i.e., $d_{\mathcal{G}}(\nabla\tilde{f}_k) \xrightarrow{K} 0$.*

PROOF. This follows from Theorem 2.1 and Theorem 4.1. □

Concerning Corollary 4.1, note that the sequence $\{x^{k(l)}\}$ is bounded if such is the feasible set S; also having S bounded is useful for stopping criteria; see Kiwiel et al. [25, Rem. 3.8]. As observed in Feltenmark and Kiwiel [12, §3], in some applications one wants to find the minimum $\min_{\breve{S}} f$ for an unbounded set $\breve{S}$, but one can find a bounded set $\bar{S}$ that intersects the optimal set $\operatorname{Arg\,min}_{\breve{S}} f$. Then it is natural to solve, instead of the original problem $\min_{\breve{S}} f$, its *restricted* version $\min_S f$ with a bounded feasible set $S = \breve{S} \cap \bar{S}$. Both problems have the same optimal subgradient set $\mathcal{G}$ if the "bounding" set $\bar{S}$ is "large enough", as explained in the following result of Feltenmark and Kiwiel [12, Lem. 3.7].

FACT 4.1 *Suppose $\min_S f$ is a restriction of the original problem $\min_{\breve{S}} f$ in the sense that $S = \breve{S} \cap \bar{S}$ for two convex sets $\breve{S}$ and $\bar{S}$. Let $\breve{S}_* := \operatorname{Arg\,min}_{\breve{S}} f$. Suppose $\breve{S}_* \cap \operatorname{int} \bar{S} \neq \emptyset$. Then $\emptyset \neq S_* \subset \breve{S}_*$, and we have both $\mathcal{G} = \partial f(x) \cap -\mathcal{N}_S(x)$ for every x in S_*, and $\mathcal{G} = \partial f(x) \cap -\mathcal{N}_{\breve{S}}(x)$ for every x in $\breve{S}_*$.*

REMARK 4.1 Under the assumptions of Fact 4.1, $\mathcal{N}_{\breve{S}}$ may replace $\mathcal{N}_S$ in Theorem 4.1; then $\mathcal{G} := \partial f(x^\infty) \cap -\mathcal{N}_{\breve{S}}(x^\infty)$ characterizes "optimal" subgradients for both $\min_S f$ *and* $\min_{\breve{S}} f$, also in Corollary 4.1. In general, if $\breve{S}_* \neq \emptyset$, then it suffices to choose $\bar{S}$ "large enough" but compact to have S bounded as well.

Following Feltenmark and Kiwiel [12, §4], the results of this section can be specialized as in Kiwiel et al. [25, §5] to the cases where we have explicit representations of f as a finite-max-type function, and of S as the solution set of finitely many nonlinear inequalities and linear equalities. The resulting schemes for identifying multipliers of objective pieces and constraints work under more general conditions than those in Anstreicher and Wolsey [1] and Larsson et al. [31]; see Kiwiel et al. [25, Rem. 5.15].

5. Lagrangian relaxation. For Lagrangian relaxation, in the general setting of Example 1.1, we consider the following two choices of the dual feasible set S:

$$S := \breve{S} := \mathbb{R}^n_+ \quad \text{or} \quad S := \{x : 0 \le x \le x^{\text{up}}\} \text{ with } x^{\text{up}} > \bar{x} \text{ for some } \bar{x} \in \breve{S}_*. \tag{38}$$

For the second choice, our problem $\min_S f$ is a restricted version of the classical dual problem $\min_{\breve{S}} f$ in the sense of Fact 4.1.

In this setting, our method employs the partial Lagrangian solutions and their constraint values

$$z^k := z(x^k) \quad \text{and} \quad g_f^k := \psi(z^k) \quad \text{for all } k; \tag{39}$$

note that, by (3)–(5),

$$f_k(\cdot) = \psi_0(z^k) + \langle \cdot, \psi(z^k)\rangle. \tag{40}$$

Using the convex weights $\{\nu_j^k\}_{j=k(l)}^k$ of (13), we define the kth *aggregate primal solution*

$$\tilde{z}^k := \sum_{j=k(l)}^{k} \nu_j^k z^j. \tag{41}$$

This construction is related to the aggregate linearization $\tilde{f}_k := \sum_{j=k(l)}^k \nu_j^k f_j$ of (20). By expressing each linearization f_j as in (40), we now derive bounds on the primal function values $\psi_0(\tilde{z}^k)$ and $\psi(\tilde{z}^k)$ that are useful for both asymptotic analysis and practical stopping criteria.

LEMMA 5.1 *The kth aggregate primal solution defined by (41) satisfies $\tilde{z}^k \in Z$,*

$$\psi_0(\tilde{z}^k) \ge \tilde{f}_k(0) \ge f(x^{k(l)}) - \tilde{\epsilon}_k - \langle \nabla \tilde{f}_S^k, x^{k(l)}\rangle \quad \text{and} \quad \psi(\tilde{z}^k) \ge \nabla \tilde{f}_k,$$

where $\nabla \tilde{f}_k \ge \nabla \tilde{f}_S^k$ if $S = \mathbb{R}_+^n$, and $\nabla \tilde{f}_k = \psi(\tilde{z}^k)$ if the primal constraint function ψ is affine.

PROOF. In view of (13) and (41), we have $\tilde{z}^k \in \operatorname{co}\{z^j\}_{j=k(l)}^k \subset Z$, $\psi_0(\tilde{z}^k) \ge \sum_j \nu_j^k \psi_0(z^j)$ and $\psi(\tilde{z}^k) \ge \sum_j \nu_j^k \psi(z^j)$ by convexity of Z and concavity of ψ_0, ψ. Next, using (20) and (40), we get

$$\tilde{f}_k(\cdot) := \sum\nolimits_j \nu_j^k f_j(\cdot) = \sum\nolimits_j \nu_j^k \left[\psi_0(z^j) + \langle \cdot, \psi(z^j)\rangle\right] = \sum\nolimits_j \nu_j^k \psi_0(z^j) + \langle \nabla \tilde{f}_k, \cdot\rangle$$

with $\nabla \tilde{f}_k = \sum_j \nu_j^k \psi(z^j)$. The above equality combined with the facts that $\tilde{f}_S^k := \tilde{f}_k + \tilde{\imath}_S^k$ by (20) and $\tilde{\imath}_S^k(0) \le i_S(0) = 0$ by Lemma 3.1(ii) and (38), and the representation of $\tilde{f}_S^k$ in (31) imply that

$$\sum\nolimits_j \nu_j^k \psi_0(z^j) = \tilde{f}_k(0) = \tilde{f}_S^k(0) - \tilde{\imath}_S^k(0) \ge \tilde{f}_S^k(0) = f(x^{k(l)}) - \tilde{\epsilon}_k - \langle \nabla \tilde{f}_S^k, x^{k(l)}\rangle.$$

Finally, if $S = \mathbb{R}_+^n$, then the minorization $\tilde{\imath}_S^k \le i_S$ of Lemma 3.1(ii) gives $\nabla \tilde{\imath}_S^k \le 0$, and hence that $\nabla \tilde{f}_k = \nabla \tilde{f}_S^k - \nabla \tilde{\imath}_S^k \ge \nabla \tilde{f}_S^k$. Combining the preceding relations yields the conclusion. □

Let Z_* denote the *primal solution set* of problem (2). We now show in the setting of (36) that the aggregate primal solutions $\{\tilde{z}^k\}_{k\in K'}$, generated via (41), converge to the primal solution set Z_*.

THEOREM 5.1 *Suppose the sequence $\{x^{k(l)}\}_{l\in L}$ has a cluster point x^∞. Let $L' \subset L$ be such that $x^{k(l)} \xrightarrow{L'} x^\infty$, and let $K' := \{k(l+1) : l \in L'\}$ (cf. (35)). Then we have the following statements.*

(i) *The sequence $\{\tilde{z}^k\}_{k\in K'}$ is bounded and all its cluster points lie in the set Z.*

(ii) $f(x^{k(l)}) \downarrow f_* = f(x^\infty)$, $\tilde{\epsilon}_k + \langle \nabla \tilde{f}_S^k, x^{k(l)}\rangle \xrightarrow{K'} 0$, *and* $\underline{\lim}_{k\in K'} \min_{i=1}^n (\nabla \tilde{f}_k)_i \ge 0$.

(iii) *Let $\tilde{z}^\infty$ be a cluster point of the sequence $\{\tilde{z}^k\}_{k\in K'}$. Then $\tilde{z}^\infty$ lies in the primal solution set Z_* and in the set $Z(x^\infty)$ of (4). Moreover, the optimal primal and dual values satisfy $\psi_0^{\max} = f_*$ (i.e., there is no duality gap). Finally, we have $\psi_0(\tilde{z}^k) \xrightarrow{K'} \psi_0^{\max}$ and $\underline{\lim}_{k\in K'} \psi_j(\tilde{z}^k) \ge 0$ for $j = 1{:}n$.*

(iv) $d_{Z_*}(\tilde{z}^k) \xrightarrow{K'} 0$, *i.e., the sequence $\{\tilde{z}^k\}_{k\in K'}$ converges to the primal solution set Z_*.*

PROOF. (i) By Lemma 5.1, each $\tilde{z}^k$ lies in the set Z, which is compact by our assumption.

(ii) The first two relations follow from Lemma 4.1. By Theorem 4.1(i,ii), (38) and Remark 4.1, the sequence $\{\nabla \tilde{f}_k\}_{k\in K'}$ is bounded and its cluster points lie in the set $\mathcal{G} \subset -\mathcal{N}_{\check{S}}(x^\infty)$; since $\mathcal{N}_{\check{S}}(x^\infty) \subset -\mathbb{R}_+^n$ (see, e.g., Hiriart-Urruty and Lemaréchal [19, Ex. III.5.2.6(b)]), the third relation follows.

(iii) By statement (i), $\tilde{z}^\infty \in Z$. Pick $K'' \subset K'$ such that $\tilde{z}^k \xrightarrow{K''} \tilde{z}^\infty$. Using statement (ii) in Lemma 5.1 together with the closedness (upper semicontinuity) of ψ_0 and ψ on Z gives

$$\psi_0(\tilde{z}^\infty) \ge \overline{\lim_{k\in K''}}\, \psi_0(\tilde{z}^k) \ge \underline{\lim_{k\in K''}}\, \psi_0(\tilde{z}^k) \ge f(x^\infty) = f_*, \tag{42a}$$

$$\psi_j(\tilde{z}^\infty) \ge \overline{\lim_{k\in K''}}\, \psi_j(\tilde{z}^k) \ge \underline{\lim_{k\in K''}}\, \psi_j(\tilde{z}^k) \ge 0, \quad j = 1{:}n. \tag{42b}$$

Thus the point $\tilde{z}^\infty$ is primal feasible. Since $\psi_0(\tilde{z}^\infty) \le \psi_0^{\max} \le f(x^\infty)$ by weak duality, (42a) yields that $\psi_0(\tilde{z}^\infty) = \psi_0^{\max} = f(x^\infty)$ and hence $\tilde{z}^\infty \in Z_*$. Then the inequalities $\psi(\tilde{z}^\infty) \ge 0$ and $x^\infty \ge 0$ (due to

$x^\infty \in S$) give $\psi_0(\tilde{z}^\infty) + \langle x^\infty, \psi(\tilde{z}^\infty)\rangle \geq f(x^\infty)$, so that $\tilde{z}^\infty \in Z(x^\infty)$ by (3)–(4). Next, since (42a) with $\psi_0(\tilde{z}^\infty) = f(x^\infty)$ yields $\psi_0(\tilde{z}^k) \xrightarrow{K''} \psi_0^{\max}$, whereas the sequence $\{\tilde{z}^k\}_{k\in K'}$ is bounded by statement (i), the final assertion may be obtained by considering convergent subsequences and using (42).

(iv) This follows from statements (i), (iii) and the continuity of the distance funtion d_{Z_*}. □

COROLLARY 5.1 *Suppose that the sequence $\{x^{k(l)}\}$ is bounded; e.g., the optimal dual set S_* is bounded (see Example 1.1 for a sufficient condition). Then the optimal primal and dual values satisfy $\psi_0^{\max} = f_*$, the sequence $\{\tilde{z}^k\}_{k\in K}$ is bounded and all its cluster points lie in the primal solution set Z_*, $d_{Z_*}(\tilde{z}^k) \xrightarrow{K} 0$, $f(x^{k(l)}) \downarrow \psi_0^{\max}$, $\psi_0(\tilde{z}^k) \xrightarrow{K} \psi_0^{\max}$ and $\underline{\lim}_{k\in K} \psi_j(\tilde{z}^k) \geq 0$ for $j = 1{:}n$.*

PROOF. Consider suitable convergent subsequences of $\{x^{k(l)}\}_{l\in L}$ and $\{\tilde{z}^k\}_{k\in K}$ in Theorem 5.1. □

REMARKS 5.1 (i) Given an accuracy tolerance $\epsilon > 0$, the method may stop if

$$\psi_0(\tilde{z}^k) \geq f(x^{k(l)}) - \epsilon \quad \text{and} \quad \psi_j(\tilde{z}^k) \geq -\epsilon, \quad j = 1{:}n.$$

Then $\psi_0(\tilde{z}^k) \geq \psi_0^{\max} - \epsilon$ from $f(x^{k(l)}) \geq \psi_0^{\max}$ (weak duality); in other words, the point $\tilde{z}^k \in Z$ is an *ϵ-solution* of the primal problem (2). By Lemma 5.1 and Theorem 5.1(ii), this stopping criterion will be satisfied for some k in at least two cases: if $S := \mathbb{R}^n_+$ and $|x^{k(l)}| \not\to \infty$ (e.g., if the dual optimal set $\check{S}_*$ is bounded; cf. Theorem 2.1), or if $S := \{x : 0 \leq x \leq x^{\text{up}}\}$ for the point x^{up} chosen as in (38).

(ii) If $\psi(\check{z}) > 0$ for some $\check{z} \in Z$, then for any points $\bar{x} \in \check{S}_* := \operatorname{Arg\,min}_{\mathbb{R}^n_+} f$ and $x \geq 0$, we have

$$\bar{x}_j \leq [f(x) - \psi_0(\check{z})]/\psi_j(\check{z}), \quad j = 1{:}n$$

(since $\psi_0(\check{z}) + \langle \bar{x}, \psi(\check{z})\rangle \leq f(\bar{x}) \leq f(x)$ by (3)). Such bounds may be used for choosing $x^{\text{up}} > \bar{x}$ in (38).

(iii) Our results may mitigate common critiques of subgradient optimization (see, e.g., Sen and Sherali [40]), which claim that such methods need heuristic stepsizes, lack effective stopping criteria and are not dual adequate (cf. (i) above).

(iv) For the standard subgradient iteration (11)–(12), the results in Larsson et al. [32] and Sherali and Choi [41] (where each function ψ_j is affine and the condition $\sum_k \nu_k^2 < \infty$ is replaced by the assumption that $x^k \to \bar{x} \in S_*$) correspond to replacing the set K by $\{1, 2, \ldots\}$ in Corollary 5.1, and $k(l)$ by 1 in (41). Hence our estimates may be expected to converge faster, since information from early steps is explicitly discarded. Further, Sherali and Choi [41] give partial results only for deflected subgradient approaches, which are easily handled in our framework; cf. §6.

We now indicate briefly two useful extensions of the framework of Example 1.1.

REMARKS 5.2 (i) Consider the equality constrained version of the primal problem (2):

$$\psi_0^{\max} := \max\ \psi_0(z) \quad \text{s.t.} \quad \psi(z) := Az - b = 0,\ z \in Z, \tag{43}$$

where $A \in \mathbb{R}^{n\times \bar{m}}$, $b \in \mathbb{R}^n$. Modifying (38), we may take either $S := \check{S} := \mathbb{R}^n$ or $S := \{x : x^{\text{low}} \leq x \leq x^{\text{up}}\}$ for bounding vectors that satisfy $x^{\text{low}} < \bar{x} < x^{\text{up}}$ for some dual solution $\bar{x} \in \check{S}_*$. Then Lemma 5.1 holds with $\psi(\tilde{z}^k) = \nabla \bar{f}_k$ (where $\nabla \bar{f}_k = \nabla \bar{f}_S^k$ if $S = \mathbb{R}^n$), and Theorem 5.1 holds with $\psi(\tilde{z}^k) = \nabla \bar{f}_k \xrightarrow{K'} 0$ in statement (ii) (using $\mathcal{N}_{\check{S}}(x^\infty) = \{0\}$), and hence $\psi(\tilde{z}^\infty) = 0$ in statement (iii).

(ii) Instead of assuming that the set Z is compact, suppose Z is closed and the mapping $z(\cdot)$ of (4) is locally bounded on the dual feasible set S. The preceding results of this section are not affected, since statement (i) of Theorem 5.1 follows from (39), (41) and Lemma 4.1(ii). This observation can also be exploited in the bundle framework of Feltenmark and Kiwiel [12, §5].

6. Accelerations.

As shown by Kiwiel et al. [26, §7], we may accelerate Algorithm 2.1 by replacing the subgradient linearization f_k with a more accurate model ϕ_k of f_S; this means that Step 4 sets

$$x^{k+1/2} := x^k + t_k(P_{\mathcal{L}_k} x^k - x^k), \quad \check{\rho}_k := t_k(2 - t_k) d^2_{\mathcal{L}_k}(x^k) \quad \text{with} \quad \mathcal{L}_k := \mathcal{L}_{\phi_k}(f^k_{\text{lev}}). \tag{44}$$

In other words, the halfspace H_k is replaced by the (hopefully tighter) approximation $\mathcal{L}_k$ of the objective level set $\mathcal{L}_{f_S}(f^k_{\text{lev}})$. The main idea is that the model ϕ_k should accumulate information from the past linearizations in order to prevent zigzags. Even fairly simple models yield faster convergence in practice.

Yet, for aggregation, we need to know the weights of past linearizations f_j in the current model, and the necessary notation becomes quite complex. To save space, we provide below formulas for several popular models, referring the interested readers to Kiwiel et al. [25, §6] for their justifications.

For the model choices specified below, ϕ_k is an affine minorant of f_S such that $\phi_k(x^k) > f_{\text{lev}}^k$. Therefore, if its gradient $g_\phi^k := \nabla\phi_k$ is nonzero, then (44) implies that we have $d_{\mathcal{L}_k}(x^k) = [\phi_k(x^k) - f_{\text{lev}}^k]/|g_\phi^k|$ and $x^{k+1/2} = x^k - \bar{\nu}_k g_\phi^k$ for the stepsize $\bar{\nu}_k := [\phi_k(x^k) - f_{\text{lev}}^k]/|g_\phi^k|^2$, which replaces ν_k in (10) and (13); hence the cumulative stepsize $\bar{\nu}_f^k := \sum_{j=k(l)}^k \bar{\nu}_j$ is updated by setting $\bar{\nu}_f^k := \bar{\nu}_f^{k-1} + \bar{\nu}_k$ if $k > k(l)$, $\bar{\nu}_f^k := \bar{\nu}_k$ otherwise. When $g_\phi^k = 0$, we have $d_{\mathcal{L}_k}(x^k) = \infty$, and we may set $x^{k+1/2} := x^k$ and $\bar{\nu}_f^k := \bar{\nu}_k := 1$.

Our implementation tested in §7.5 generates ϕ_k by combining the current linearization f_k with a past linearization $\bar{\phi}_{k-1}$ of f; to account for constraints, they are turned into linearizations $\breve{f}_k$ and $\breve{\phi}_{k-1}$ of f_S by using the subgradient reduction technique of Kiwiel [22, §7]. Specifically, we use the following formulas

$$\phi_k := (1-\alpha_k)\breve{f}_k + \alpha_k\breve{\phi}_{k-1} \quad \text{with} \quad \alpha_k \in [0,1],$$

$$\breve{f}_k(\cdot) := f_k(x^k) + \langle \breve{g}^k, \cdot - x^k\rangle, \quad \breve{\phi}_{k-1}(\cdot) := \bar{\phi}_f^{k-1}(x^k) + \langle \breve{g}_\phi^{k-1}, \cdot - x^k\rangle,$$

where $\breve{g}^k := g_f^k + P_{\mathcal{N}_S(x^k)}(-g_f^k)$ and $\breve{g}_\phi^{k-1} := \bar{g}_{\phi_f}^{k-1} + P_{\mathcal{N}_S(x^k)}(-\bar{g}_{\phi_f}^{k-1})$ are reduced subgradients, for $\bar{g}_{\phi_f}^{k-1} := \nabla\bar{\phi}_f^{k-1}$, updating

$$\bar{\phi}_f^k := (1-\alpha_k)f_k + \alpha_k\bar{\phi}_f^{k-1}, \quad \bar{\phi}_f^0 := f_1.$$

The choices of the weight α_k above, given by Kiwiel et al. [26, Ex. 7.4(v) and Rem. 7.6], include:

(i) the *ordinary subgradient strategy* (OSS): $\alpha_k := 0$;

(ii) the *conjugate subgradient strategy* (CSS):

$$\alpha_k := \begin{cases} \dfrac{\langle \breve{g}^k, \breve{g}_\phi^{k-1}\rangle}{\langle \breve{g}^k, \breve{g}_\phi^{k-1}\rangle - |\breve{g}_\phi^{k-1}|^2} & \text{if } \langle \breve{g}^k, \breve{g}_\phi^{k-1}\rangle < 0 \text{ and } \breve{\phi}_{k-1}(x^k) \ge f_{\text{lev}}^k, \\ 0 & \text{otherwise;} \end{cases}$$

(iii) the *average direction strategy* (ADS):

$$\alpha_k := \begin{cases} \alpha_k := \dfrac{|\breve{g}^k|}{|\breve{g}^k| + |\breve{g}_\phi^{k-1}|} & \text{if } \breve{g}_\phi^{k-1} \ne 0 \text{ and } \breve{\phi}_{k-1}(x^k) \ge f_{\text{lev}}^k, \\ 0 & \text{otherwise;} \end{cases}$$

(iv) the *aggregate subgradient strategy* (ASS): α_k is such that the projection of the point x^k on the f_{lev}^k-level set of ϕ_k coincides with its projection on the f_{lev}^k-level set of $\max\{\breve{f}_k, \breve{\phi}_{k-1}\}$ if the latter set is nonempty, otherwise α_k is such that the former set is empty; see Kiwiel [22, Rem. 4.1].

For OSS and ASS, if the Fejér tests (34) and (17) are false and $\max\{f_k(x^{k+1}), \bar{\phi}_f^k(x^{k+1})\} > f_{\text{rec}}^{k(l)} - \frac{3}{4}\delta_l$, then Step 4 is repeated with x^k and $\bar{\phi}_f^{k-1}$ replaced by x^{k+1} and $\bar{\phi}_f^k$. Such *repeated projections* are justified by Kiwiel et al. [26, Rem. 7.11] (but not for CSS and ADS). They provide an inexact implementation of the "best" single projection of x^k on the set $\mathcal{L}_{\max\{f_k, \bar{\phi}_f^{k-1}\}}(f_{\text{lev}}^k) \cap S$, which may be too expensive.

For primal aggregation (cf. (41)), we use the following updates (where $\bar{z}^0 := z_\phi^0 := z^1$):

$$\bar{z}^k := (\bar{\nu}_k/\bar{\nu}_f^k)z_\phi^k + (1 - \bar{\nu}_k/\bar{\nu}_f^k)\bar{z}^{k-1} \quad \text{with} \quad z_\phi^k := (1-\alpha_k)z^k + \alpha_k z_\phi^{k-1}. \tag{45}$$

Here one point should be noted. If we set $\alpha_{k(l)} := 0$ when a group starts, these constructions produce $(\bar{f}_k, \bar{z}^k) \in \text{co}\{(f_j, z^j)\}_{j=k(l)}^k$ and $(\bar{\phi}_f^k, z_\phi^k) \in \text{co}\{(f_j, z^j)\}_{j=k(l)}^k$; otherwise 1 replaces $k(l)$ in these inclusions. However, we may allow $\alpha_{k(l)} \ne 0$ in at least two cases. First, suppose the subgradient mapping g_f is bounded on the set S (e.g., ψ is continuous in Example 1.1); then the sequence $\{\nabla\bar{f}_k\}$ is bounded, as required for Theorem 4.1(i). Second, suppose the optimal set S_* is bounded. Then, by Theorem 2.1 and Remark 2.1(v), the sequences $\{x^k\}$ and $\{g_f^k\}$ are bounded, so that again the sequence $\{\nabla\bar{f}_k\}$ is bounded.

7. Application to multicommodity network flows. In this section we discuss an application of our method to the traffic assignment and message routing problems, which are important instances of nonlinear multicommodity network flow problems; see, e.g., Bertsekas [8, Chap. 8] for a textbook introduction, Ouorou et al. [36] for a recent survey, Fukushima [13, 14] for the pioneering dual developments, and Goffin et al. [17], Goffin et al. [18], Larsson et al. [29], and Larsson et al. [32] for recent comparable approaches. In particular, in §7.4 we relax the standard assumption of strictly convex arc costs, because our real-life instances include linear costs. Incidentally, our theoretical developments also lay ground for the application of the proximal bundle method in Feltenmark and Kiwiel [12, §5] to such problems.

7.1 The nonlinear multicommodity flow problem. Let $(\mathcal{N}, \mathcal{A})$ be a directed graph with N nodes and n arcs. Let $E \in \mathbb{R}^{N\times n}$ be its node-arc incidence matrix. There are m commodities to be routed through the network. For each commodity i there is a *required flow* $r_i > 0$ from its *source* node o_i to its *sink* node d_i. Let s_i be the *supply* N-vector of commodity i, having components $s_{io_i} = r_i$, $s_{id_i} = -r_i$, $s_{il} = 0$ if $l \neq o_i, d_i$. Our *convex separable multicommodity flow* problem is stated as follows:

$$\min \quad \breve{\psi}_0(z_0) := \sum_{j=1}^{n} \breve{\psi}_{0j}(z_{0j}) \tag{46a}$$

$$\text{s.t.} \quad \psi_j(z) := z_{0j} - \sum_{i=1}^{m} z_{ij} = 0, \quad j = 1{:}\,n, \tag{46b}$$

$$z := (z_0, z_1, \ldots, z_m) \in Z := Z_0 \times Z_1 \times \cdots \times Z_m, \tag{46c}$$

$$Z_0 := \mathbb{R}^n, \quad Z_i := \{\, z_i : Ez_i = s_i, 0 \le z_i \le \bar{z}_i \,\}, \quad i = 1{:}\,m, \tag{46d}$$

where z_i is the *flow vector of commodity* $i \in \{1{:}\,m\}$, $z_0 = \sum_{i=1}^{m} z_i$ is the *total flow vector*, and $\bar{z}_i$ is a fixed positive vector of *flow bounds* for each i. We assume that each *arc cost* function $\breve{\psi}_{0j}$ is closed proper strictly convex and increasing on its effective domain that equals $[0, \kappa_j)$ or $[0, \kappa_j]$ for a constant κ_j, and either $0 < \kappa_j < \infty$ or $\kappa_j = \infty$ and $\lim_{t\to\infty} \breve{\psi}'_{0j}(t) = \infty$, where $\breve{\psi}'_{0j}$ denotes the right derivative of $\breve{\psi}_{0j}$. (Here and in what follows, we assume basic familiarity with convex univariate functions; see, e.g., Bertsekas [8, §9.1], Rockafellar [39, pp. 227–230].) Finally, we suppose that

$$\check{z}_0 \in [0, \kappa_1) \times \cdots \times [0, \kappa_n) \quad \text{for some } \check{z} \in Z \text{ with } \psi(\check{z}) = 0. \tag{47}$$

7.2 Dual approach. In the framework of Remarks 5.2, letting $\psi_0(z) := -\breve{\psi}_0(z_0)$ and $\breve{S} := \mathbb{R}^n$, we may view problem (46) as an instance of the primal problem (43). Then, for each multiplier x, the dual function value of (3) and the partial Lagrangian solution of (4) can be written as $f(x) = \sum_{i=0}^{m} f^i(x)$ and $z(x) = (z_0(x), \ldots, z_m(x))$, where $f^0(x) := \sum_{j=1}^{n} f_j^0(x_j)$,

$$f_j^0(x_j) := \max_t \{\, x_j t - \breve{\psi}_{0j}(t) \,\} = \breve{\psi}_{0j}^*(x_j), \quad j = 1{:}\,n, \tag{48a}$$

$$z_{0j}(x) := \arg\min_t \{\, \breve{\psi}_{0j}(t) - x_j t \,\} = \nabla \breve{\psi}_{0j}^*(x_j) = \nabla f_j^0(x_j), \quad j = 1{:}\,n, \tag{48b}$$

and

$$f^i(x) := \max\{\, -\langle x, z_i\rangle : Ez_i = s_i, 0 \le z_i \le \bar{z}_i \,\}, \quad i = 1{:}\,m, \tag{49a}$$

$$z_i(x) \in \operatorname{Arg\,min}\{\, \langle x, z_i\rangle : Ez_i = s_i, 0 \le z_i \le \bar{z}_i \,\} = -\partial f^i(x), \quad i = 1{:}\,m. \tag{49b}$$

Concerning (48), note that, since each cost function $\breve{\psi}_{0j}$ is strictly convex, its conjugate function $\breve{\psi}_{0j}^*$ is continuously differentiable; hence the mapping $z_0(\cdot)$ is locally bounded. In turn, the mappings $z_i(\cdot)$ produced by (49b) are bounded by $0 \le z_i(\cdot) \le \bar{z}_i$. Consequently, the mappings $z(\cdot)$ and $g_f(\cdot) := \psi(z(\cdot))$ are locally bounded (as stipulated in Example 1.1 and Remark 5.2(ii)).

As for practical aspects, in typical applications the conjugate functions $\breve{\psi}_{0j}^*$ are available in closed form, and the computations involved in (48) are easy. In contrast, (49b) involves solving, for each i, a shortest path problem with some negative arc lengths if $x \not\geq 0$, and side constraints imposed by $\bar{z}_i$. Suppose momentarily that $x \ge 0$. Then this problem becomes much easier to solve. Further, consider the case where the required flow r_i and the flow bound $\bar{z}_i$ satisfy $r_i \le \bar{z}_{ij}$ for all j. Then, ignoring $\bar{z}_i$ in (49b), we may find $z_i(x)$ by solving a shortest path problem with nonnegative arc lengths and no side constraints (since this solution satisfies $z_{ij}(x) \le r_i$ for all j); this problem is easy; see, e.g., Gallo and Pallotino [15]. In particular, this means that we can handle problems where the flow bounds $\bar{z}_i$ are omitted in (46d) and (49b) (as happens in many applications), since the algorithm will proceed as if we had flow bounds satisfying $\bar{z}_{ij} \ge r_i$ for all i and j (i.e., we may pick such bounds for *theoretical* purposes only).

To sum up, the work in solving subproblems (49b) would reduce significantly if we took $S = \mathbb{R}^n_+$ as the dual feasible set for our method; a better choice due to Fukushima [13] is validated below.

THEOREM 7.1 *Under the assumptions of* §7.1, *we have the following statements.*

(i) *Problem* (46) *has a solution, and it is equivalent to the following inequality constrained problem:*

$$\breve{\psi}_0^{\min} := \min\ \breve{\psi}_0(z_0) \quad s.t. \quad \psi(z) \geq 0,\ z \in Z. \tag{50}$$

(ii) *The set* $\tilde{S}_* := \operatorname{Arg\,min}_{\mathbb{R}^n_+} f$ *of Lagrange multipliers of problem* (50) *is nonempty and bounded, and it is contained in the set* $\breve{S}_* := \operatorname{Arg\,min} f$ *of Lagrange multipliers of problem* (46).

(iii) *For the* restricted dual feasible set S *and the lower bounding vector* x^{low} *defined by*

$$S := \{\, x : x \geq x^{\text{low}} \,\} \quad with \quad x_j^{\text{low}} := \breve{\psi}_{0j}'(0) \geq 0 \ for\ j = 1{:}\,n, \tag{51}$$

the dual optimal set $S_* := \operatorname{Arg\,min}_S f$ *is nonempty and lies in the bounded set* $\tilde{S}_*$ *of statement* (ii).

(iv) *The* primal solution set *of problem* (46) (*and of the equivalent problem* (50)) *has the form*

$$Z_* = \{z_0^*\} \times Z_*^f \quad with \quad Z_*^f := \left\{ (z_1, \ldots, z_m) \in Z_1 \times \cdots \times Z_m : z_0^* = \sum\nolimits_{i=1}^m z_i \right\}, \tag{52}$$

where z_0^* *is the unique optimal total flow.*

(v) *If each arc cost function* $\breve{\psi}_{0j}$ *is finite and differentiable on the segment* $(0, \infty)$, *then the dual function* f *is strictly convex on the set* S *of* (51), *and hence the dual optimal set* S_* *is a singleton.*

PROOF. (i) Both problems have solutions: their feasible sets are closed, whereas their objective $\breve{\psi}_0$ is closed, has a finite value at the point $\check{z}$ of (47), and satisfies $\breve{\psi}_0(z_0) \to \infty$ if $|z_0| \to \infty$. The equivalence follows from the following observation: if a point z is feasible in (50) and $\breve{\psi}_0(z_0) < \infty$, but $\psi_j(z) > 0$ for some j, then, since the function $\breve{\psi}_{0j}$ in (46) is increasing on its effective domain, we can reduce $\breve{\psi}_0(z_0)$ by decreasing z_{0j} to $z_{0j} = \sum_{i=1}^m z_{ij}$, thus obtaining a better feasible point with $\psi_j(z) = 0$.

(ii) Since (47) is equivalent to Slater's condition for (50) ($\psi(z) > 0$ for some $z \in Z$ with $z_0 \in \operatorname{dom} \breve{\psi}_0$), the first assertion follows from Rockafellar's [39, Cors. 28.2.1, 28.4.1 and 29.1.5]; in particular, $\min_{\mathbb{R}^n_+} f = -\breve{\psi}_0^{\min}$. Since $\breve{\psi}_0^{\min}$ is also the optimal value of (46) by statement (i), and thus $-\breve{\psi}_0^{\min} \leq \min_{\mathbb{R}^n} f$ by weak duality, we have $-\breve{\psi}_0^{\min} = \min_{\mathbb{R}^n} f = \min_{\mathbb{R}^n_+} f$ (no duality gap), and the second assertion follows.

(iii) Since each $\breve{\psi}_{0j}$ is nondecreasing on its domain, we have $x_j^{\text{low}} \geq 0$, whereas for $0 \leq x_j \leq x_j^{\text{low}}$, $f_j^0(x_j)$ is constant in (48b), and each $f^i(x)$ is nonincreasing in (49a), so that $f(x)$ is nonicreasing. Hence $\min_{\mathbb{R}^n_+} f = \min_S f$, and it follows from the definitions that S_* is the projection of $\tilde{S}_*$ onto S.

(iv) This follows from the strict convexity of the objective $\breve{\psi}_0$ and the structure of the feasible set.

(v) Fix $j \in \{1{:}\,n\}$. Since $\breve{\psi}_{0j}$ is strictly convex, $\nabla \breve{\psi}_{0j}$ is increasing on $(0, \infty)$. Then, by (48b), $\nabla f_j^0(x_j) = \nabla \breve{\psi}_{0j}^*(x_j)$ is increasing for $x_j > \breve{\psi}_{0j}'(0)$ (since $\nabla \breve{\psi}_{0j}^*(x_j) = (\nabla \breve{\psi}_{0j})^{-1}(x_j)$), and thus $f_j^0(x_j)$ is strictly convex for $x_j > \breve{\psi}_{0j}'(0)$. In effect, f^0 and f are strictly convex on S. □

7.3 Algorithmic constructions and convergence. We now consider the application of our method in the setting of §7.2, using the mappings $z(\cdot)$ and $g_f(\cdot) := \psi(z(\cdot))$ defined via (48)–(49) at points in the feasible set S given by (51). Recall that these mappings are locally bounded. The local boundedness of g_f suffices for Theorem 2.1 and the convergence results of §4, with the optimal dual set S_* being bounded by Theorem 7.1(iii). On the other hand, the local boundedness of the mapping $z(\cdot)$ is crucial for extending the results of §5 as follows.

Here we view the inequality constrained problem (50) as an instance of the general problem (2) with the "flipped" objective $\psi_0(z) := -\breve{\psi}_0(z_0)$, so that their optimal values satisfy $\breve{\psi}_0^{\min} = -\psi_0^{\max}$. By Theorem 7.1(i), these two problems and our original problem (46) have a common solution set Z_*, and $\breve{\psi}_0^{\min}$ is the optimal value of (46). Now, in view of the local boundedness of $z(\cdot)$ and Remark 5.2(ii), the results of §5 would hold if we replaced S by $\mathbb{R}^n_+$ (cf. (38)); fortunately, this replacement is not needed. Namely, Theorem 5.1 is true: in the proof of statement (ii), we have $\mathcal{G} \subset -\mathcal{N}_S(x^\infty)$ and $\mathcal{N}_S(x^\infty) \subset -\mathbb{R}^n_+$ by (51), which also gives $x^\infty \geq 0$ in the proof of statement (iii). We conclude that *all* the results §5 still hold. In particular, the conclusions of Corollary 5.1 hold, since the optimal dual set S_* is bounded.

It follows that for any tolerance $\epsilon > 0$, the stopping criterion of Remark 5.1(i) will be met for some k. We now derive an alternative stopping criterion that is more efficient in practice. Basically, it involves turning the aggregate solution $\tilde{z}^k$ into another primal-feasible point $\breve{z}^k \in Z$ such that $\psi(\breve{z}^k) = 0$.

To this end, we first note that Remark 2.1(i) and Corollary 5.1 yield $f(x^k_{\rm rec}) \downarrow \psi_0^{\max} = -\breve{\psi}_0^{\min}$. Next, we observe that although the aggregate $\tilde{z}^k$ need not be feasible in the primal problem (46), it lies in the set Z by Lemma 5.1. Hence we may use its commodity components $\tilde{z}^k_i$, $i = 1{:}\,m$, to produce the *aggregate total flow*

$$\breve{z}^k_0 := \sum_{i=1}^m \tilde{z}^k_i \tag{53}$$

and the *primal feasible aggregate*

$$\breve{z}^k := (\breve{z}^k_0, \tilde{z}^k_1, \ldots, \tilde{z}^k_m) \in Z \quad \text{with } \psi(\breve{z}^k) = 0. \tag{54}$$

Note that

$$0 \le \breve{\psi}_0(\breve{z}^k_0) - \breve{\psi}_0^{\min} \le \breve{\psi}_0(\breve{z}^k_0) + f(x^k_{\rm rec}), \tag{55}$$

since $\breve{\psi}_0^{\min}$ is the optimal value of problem (46), and $-\breve{\psi}_0^{\min} = \psi_0^{\max} \le f(x^k_{\rm rec})$ as shown above. Therefore, the method may stop when $\breve{\psi}_0(\breve{z}^k_0) + f(x^k_{\rm rec}) \le \epsilon$ for a given tolerance $\epsilon > 0$, in which case $\breve{z}^k$ is a *feasible ϵ-solution* of problem (46). Among other things, the following result implies that this stopping criterion will be met for some k if the effective domain of each cost function $\breve{\psi}_{0j}$ has the form $[0, \kappa_j)$.

PROPOSITION 7.1 (i) $\psi_0^{\max} = -\breve{\psi}_0^{\min} = f_*$, $\breve{\psi}_0(\tilde{z}^k_0) \xrightarrow{K} \breve{\psi}_0^{\min}$ *and* $\psi(\tilde{z}^k) \xrightarrow{K} 0$.

(ii) $\tilde{z}^k_0 - \breve{z}^k_0 = \psi(\tilde{z}^k) \xrightarrow{K} 0$, $|\tilde{z}^k - \breve{z}^k| = |\tilde{z}^k_0 - \breve{z}^k_0| \xrightarrow{K} 0$, $d_{Z_*}(\tilde{z}^k) \xrightarrow{K} 0$ *and* $d_{Z_*}(\breve{z}^k) \xrightarrow{K} 0$.

(iii) $\tilde{z}^k_0 \xrightarrow{K} z^*_0$, $\breve{z}^k_0 \xrightarrow{K} z^*_0$, *and* $d_{Z^f_*}((\tilde{z}^k_1, \ldots, \tilde{z}^k_m)) \xrightarrow{K} 0$, *where z^*_0 is the unique optimal total flow, and the set Z^f_* of optimal commodity flows is given by* (52).

(iv) *If the optimal flow satisfies $z^*_0 \in \prod_{j=1}^n [0, \kappa_j)$, then* $\breve{\psi}_0(\breve{z}^k_0) \xrightarrow{K} \breve{\psi}_0^{\min}$ *and* $\breve{\psi}_0(\breve{z}^k_0) + f(x^k_{\rm rec}) \xrightarrow{K} 0$.

PROOF. (i) The optimal dual set S_* is bounded and $\psi_0(z) := -\breve{\psi}_0(z_0)$, so the first two relations follow from Corollary 5.1, which also yields that all cluster points of the bounded sequence $\{\tilde{z}^k\}_{k\in K}$ lie in Z_*; since Z_* is the solution set of our equality constrained problem (46), it follows that $\psi(\tilde{z}^k) \xrightarrow{K} 0$.

(ii) We have $\psi(\tilde{z}^k) = \tilde{z}^k_0 - \breve{z}^k_0$ and $|\tilde{z}^k - \breve{z}^k| = |\tilde{z}^k_0 - \breve{z}^k_0|$ by (46b) and (54); therefore, the first two relations follow from statement (i). Next, since $d_{Z_*}(\tilde{z}^k) \xrightarrow{K} 0$ by Corollary 5.1, the fourth relation is a consequence of the second one and the fact that the distance function d_{Z_*} is Lipschitz continuous.

(iii) Recalling the form (52) of the primal solution set Z_*, use the final two relations of statement (ii).

(iv) By statement (iii), $\breve{z}^k_0 \xrightarrow{K} z^*_0$ with $\breve{z}^k_0 \ge 0$ by (53), (41), (39) and (49b). Since each function $\breve{\psi}_{0j}$ in (46a) is continuous on $[0, \kappa_j)$, we have $\breve{\psi}_0(\breve{z}^k_0) \xrightarrow{K} \breve{\psi}_0^{\min}$, whereas $f(x^k_{\rm rec}) \downarrow -\breve{\psi}_0^{\min}$ as shown above. ☐

7.4 Extension to linear costs. We now present an extension to the case where some of the cost functions are linear. Thus, retaining the remaining assumptions of §7.1, suppose that for a fixed integer $0 \le \breve{n} < n$ and each index j such that $\breve{n} < j \le n$, the cost function $\breve{\psi}_{0j}$ is *linear* on its effective domain:

$$\breve{\psi}_{0j}(t) = \begin{cases} \breve{\psi}'_{0j}(0)t & \text{if } t \ge 0, \\ \infty & \text{otherwise,} \end{cases}$$

with $\breve{\psi}'_{0j}(0) > 0$. Then, by (48) and (51), $f^0_j(x_j) = 0$ and $z_{0j}(x) = 0$ if $x_j < x^{\rm low}_j$, $f^0_j(x_j) = \infty$ and $z_{0j}(x)$ is undefined if $x_j > x^{\rm low}_j$, but for $x_j = x^{\rm low}_j$, $f^0_j(x_j) = 0$ and $z_{0j}(x)$ could be arbitrary in $\mathbb{R}_+$. Exploiting this freedom, we may restrict attention to the following subset of the dual feasible set S of (51):

$$\hat{S} := \{\, x : x_j \ge x^{\rm low}_j \text{ for } j \le \breve{n},\ x_j = x^{\rm low}_j \text{ for } j > \breve{n} \,\}, \tag{56}$$

letting

$$z_{0j}(x) := \sum_{i=1}^m z_{ij}(x) \quad \text{if} \quad x \in \hat{S},\ j > \breve{n}. \tag{57}$$

This gives $[g_f(x)]_j := \psi_j(z(x)) = 0$ if $x \in \hat{S}$, $j > \breve{n}$. Hence, assuming that we choose an initial point $x^1 \in \hat{S}$, by induction on (8) we shall always have $x^k \in \hat{S}$ and $[g^k_f]_j := \psi_j(z^k) = 0$ for $j > \breve{n}$. In view of

(46b), this implies that $\psi_j(\tilde{z}^k) = 0$ by (41) and $\check{z}^k_{0j} = \tilde{z}^k_{0j}$ by (53) for $j > \check{n}$. In other words, for arcs with linear costs, the multipliers are fixed at their optimal values, and the aggregate flows are primal feasible. Clearly, the mappings $z(\cdot)$ and $g_f(\cdot) := \psi(z(\cdot))$ are locally bounded on the set $\hat{S}$ (such are $z_i(\cdot)$ for $i \ge 1$ and $z_{0j}(\cdot)$ for $j \le \check{n}$ as before, since $\hat{S} \subset S$, and then, by (57), also $z_{0j}(\cdot)$ for $j > \check{n}$).

The above observations suffice for proving the first two parts of Proposition 7.1 as before. The remaining two parts are modified as follows. In part (iii), since now the representation (52) of the primal solution set Z_* is replaced by $Z_* = \{(z^*_{01}, \ldots, z^*_{0\check{n}})\} \times \hat{Z}_*$ for a suitably chosen set $\hat{Z}_*$, we have

$$\tilde{z}^k_{0j} \xrightarrow{K} z^*_{0j}, \quad \check{z}^k_{0j} \xrightarrow{K} z^*_{0j} \quad \text{for } j \le \check{n}, \quad d_{\hat{Z}_*}((\tilde{z}^k_{0,\check{n}+1}, \ldots, \tilde{z}^k_{0n}, \tilde{z}^k_1, \ldots, \tilde{z}^k_m)) \xrightarrow{K} 0.$$

As for the proof of part (iv), we have $\psi_0(\tilde{z}^k) \xrightarrow{K} \psi_0^{\max}$ by Corollary 5.1 as before; in other words, $\check{\psi}_0(\tilde{z}^k_0) \xrightarrow{K} \check{\psi}_0^{\min}$. Now, since $\check{z}^k_{0j} = \tilde{z}^k_{0j}$ for $j > \check{n}$ (see below (57)), we have, by (46a),

$$\check{\psi}_0(\check{z}^k_0) = \check{\psi}_0(\tilde{z}^k_0) + \sum_{j \le \check{n}} [\check{\psi}_{0j}(\check{z}^k_{0j}) - \check{\psi}_{0j}(\tilde{z}^k_{0j})],$$

where $\check{\psi}_{0j}(\tilde{z}^k_{0j}), \check{\psi}_{0j}(\check{z}^k_{0j}) \xrightarrow{K} \check{\psi}_{0j}(z^*_{0j})$, since $0 \le \tilde{z}^k_{0j}, \check{z}^k_{0j} \xrightarrow{K} z^*_{0j} \in [0, \kappa_j)$ and the functions $\check{\psi}_{0j}$ are continuous on $[0, \kappa_j)$ for $j \le \check{n}$. Therefore, $\check{\psi}_0(\tilde{z}^k_0) \xrightarrow{K} \check{\psi}_0^{\min}$ yields $\check{\psi}_0(\check{z}^k_0) \xrightarrow{K} \check{\psi}_0^{\min}$, as desired.

7.5 Numerical results. Our method was programmed in Fortran 77 and run on a notebook PC (Pentium 4M 2 GHz, 768 MB RAM). We used the parameters $\beta = \frac{1}{2}$, $\delta_1 = \frac{1}{2}\delta_0$ and $R_l := R(\delta_l/\delta_0)^\beta$ with $\delta_0 = R|\breve{g}^1|$ for consistency with Kiwiel et al. [26, §8], $t_k \equiv 1$, the third projection of §3.3 and the aggregate subgradient strategy of §6, updating the total flows (cf. (45), (53))

$$\check{z}^k_0 = (\bar{v}_k/\bar{v}^k_f)\check{z}^k_{\phi,0} + (1 - \bar{v}_k/\bar{v}^k_f)\check{z}^{k-1}_0 \quad \text{with} \quad \check{z}^k_{\phi,0} := \sum_{i=1}^m z^k_{\phi,i} = (1-\alpha_k)\sum_{i=1}^m z^k_i + \alpha_k \check{z}^{k-1}_{\phi,0},$$

where $\check{z}^0_0 := \check{z}^0_{\phi,0} := \sum_{i=1}^m z^1_i$. We also computed record flows $\check{z}^k_{\text{rec}}$ as follows. Letting $\check{z}^1_{\text{rec}} := \check{z}^1$, every tenth iteration or when the loop counter l increased at Steps 3 or 6, we set $\check{z}^k_{\text{rec}} := \check{z}^k$ if $\check{\psi}_0(\check{z}^k_0) < \check{\psi}_0(\check{z}^k_{\text{rec},0})$, $\check{z}^k_{\text{rec}} := \check{z}^{k-1}_{\text{rec}}$ otherwise (we did not update $\check{z}^k_{\text{rec}}$ at every iteration to save time). In view of the optimality estimate (55), we employed the following stopping criterion

$$\check{\psi}_0(\check{z}^k_{\text{rec},0}) + f(x^k_{\text{rec}}) \le \epsilon_{\text{opt}}[1 + |\check{\psi}_0(\check{z}^k_{\text{rec},0})|], \tag{58}$$

which ensured a relative objective accuracy of ϵ_{opt}; we used $\epsilon_{\text{opt}} = 10^{-i/2}$ for $i = 4, 5, 6$.

We first give results for the CNET collection of Ouorou et al. [36], which describes message routing problems in a real-life telecommunication network with 106 nodes and 904 arcs. The instances have $m = 4452, 6678, 8904$ or 11130 commodities, and five load factors (1, 1.5, 2, 2.5, 3) that scale up the standard required flows r_i. The costs are Kleinrock's *average delays*

$$\check{\psi}_{0j}(t) := \begin{cases} t/(\kappa_j - t) & \text{if } t \in [0, \kappa_j), \\ \infty & \text{otherwise.} \end{cases}$$

The starting point had components $x^1_j := \kappa_j^{-1}(1-\rho_*)^{-2}$ for all j, with $\rho_* := \frac{1}{4}$ estimating the maximum traffic intensity $\max_j z^*_{0j}/\kappa_j$ as in Goffin [16] (this intensity sometimes exceeded $\frac{1}{2}$). By trying $R = 1, 10, 100$ on the first instance, we picked $R = 10$ as a reasonable value for all instances. Our results are given in Table 1, where Delay $:= \check{\psi}_0(\check{z}^k_{\text{rec},0})$ is the best primal value obtained until the final iteration k, times are given in seconds, and the optimal delays (communicated to us by A. Ouorou) are rounded to six digits. The accuracy attained was usually higher than that guaranteed by our stopping criterion (58); e.g., for $\epsilon_{\text{opt}} = 10^{-3}$, we had $[\check{\psi}_0(\check{z}^k_{\text{rec},0}) - \check{\psi}_0^{\min}]/\check{\psi}_0^{\min} < 10^{-4}$ for the unit load instances, where $\check{\psi}_0^{\min}$ is the optimal delay. Since each instance had 106 common sources, most work per iteration went into solving 106 shortest path subproblems via subroutine L2QUE of Gallo and Pallotino [15]. Our machine is about thirteen times faster than the one employed in Ouorou et al. [36]. Hence Table 1 suggests that our method is highly competitive with all the methods tested in Ouorou et al. [36, Tables 2 and 3], at least for modest accuracy requirements that are typical for such applications.

We next give results for five real-life traffic assignment problems described in Table 2. These problems

Table 1: Results for the CNET instances

		$\epsilon_{opt} = 10^{-2}$			$\epsilon_{opt} = 10^{-2.5}$			$\epsilon_{opt} = 10^{-3}$			Optimal
m	Load	Delay	k	Time	Delay	k	Time	Delay	k	Time	Delay
4452	1.0	12.6131	110	.421	12.5881	180	.601	12.5856	590	1.59	12.5847
	1.5	19.1949	150	.431	19.1831	350	.932	19.1815	600	1.52	19.1799
	2.0	25.9955	210	.581	25.9824	267	.721	25.9784	500	1.29	25.9755
	2.5	33.0326	200	.550	33.0017	330	.881	32.9838	1350	3.35	32.9809
	3.0	40.2486	230	.631	40.2173	480	1.25	40.2125	1421	3.34	40.2072
6678	1.0	19.6691	170	.591	19.6512	370	1.10	19.6494	720	1.94	19.6481
	1.5	30.2016	240	.671	30.1821	630	1.63	30.1806	900	2.30	30.1776
	2.0	41.2893	160	.471	41.2149	430	1.14	41.2106	1030	2.52	41.2066
	2.5	52.9117	220	.601	52.7989	350	.932	52.7842	950	2.31	52.7790
	3.0	64.9875	540	1.39	64.9573	900	2.23	64.9513	1851	4.42	64.9460
8904	1.0	26.4872	230	.741	26.4872	238	.761	26.4746	1050	2.71	26.4730
	1.5	41.0286	190	.541	40.9820	427	1.13	40.9772	900	2.26	40.9742
	2.0	56.4689	390	1.07	56.4301	630	1.67	56.4260	2032	4.96	56.4233
	2.5	73.0758	350	.961	72.9578	526	1.39	72.9454	944	2.37	72.9392
	3.0	90.7997	418	1.11	90.7069	580	1.51	90.6720	860	2.17	90.6620
11130	1.0	33.5348	190	.671	33.4978	440	1.33	33.4955	860	2.38	33.4931
	1.5	52.4137	200	.591	52.2819	710	1.92	52.2709	1217	3.17	52.2677
	2.0	72.6894	480	1.31	72.6634	780	2.06	72.6462	1500	3.82	72.6434
	2.5	95.0557	325	.921	94.9118	710	1.88	94.8916	1490	3.79	94.8838
	3.0	119.406	1250	3.23	119.321	1580	4.04	119.320	1830	4.65	119.306

Table 2: Traffic assignment problems and their best known primal values

Problem	Nodes	Arcs	OD pairs	Sources	Linear costs	Best delay
Barcelona	930	2522	7922	97	565	1.26846e+6
Linköping	335	882	12372	118	0	4.05602e+8
Winnipeg	1040	2836	4344	135	1176	8.85327e+5
Chicago	2552	7850	137417	445	0	4.03799e+6
Skåne	7722	18344	712466	1057	2262	7.63642e+7

have nonlinear *BPR delays*

$$\check\psi_{0j}(t) := \begin{cases} \alpha_j t + \beta_j t^{\gamma_j} & \text{if } t \geq 0, \\ \infty & \text{otherwise,} \end{cases}$$

with parameters $\alpha_j \geq 0$, $\beta_j > 0$, $\gamma_j > 1$, as well as linear costs

$$\check\psi_{0j}(t) := \begin{cases} \alpha_j t & \text{if } t \geq 0, \\ \infty & \text{otherwise,} \end{cases}$$

with $\alpha_j > 0$; column 6 of Table 2 gives their numbers. The first three medium-sized problems were used in Larsson et al. [30]. The Chicago problem of Tatineni et al. [43] is much bigger than the largest (random) problems considered in Goffin et al. [18] and Ouorou et al. [36]. The Skåne problem (not reported so far) is really huge. We used the starting points $x^1 = x^{\text{low}}$ and the ball parameters $R = 100$, except that we took $R = 10^4$ for the Linköping problem. Concerning these choices, we add that our earlier experience indicated that $R = 100$ should be reasonable when the BPR cost coefficients α_j are of order one. This turned out to be true for our current problems, except for Linköping, which has α_j of order 100, so we multiplied the "usual" R by 100. Our results are reported in Table 3. We add that again for the tolerance

Table 3: Results for the traffic assignment problems

	$\epsilon_{opt} = 10^{-2}$			$\epsilon_{opt} = 10^{-2.5}$			$\epsilon_{opt} = 10^{-3}$		
Problem	Delay	k	Time	Delay	k	Time	Delay	k	Time
Barcelona	1.27322e+6	120	3.00	1.26937e+6	310	7.58	1.26862e+6	790	19.2
Linköping	4.06050e+8	120	1.10	4.05774e+8	150	1.35	4.05716e+8	720	6.27
Winnipeg	8.89731e+5	56	1.67	8.86426e+5	116	3.31	8.85735e+5	220	6.18
Chicago	4.06493e+6	80	19.8	4.04446e+6	130	32.3	4.03941e+6	350	87.1
Skåne	7.64631e+7	20	37.9	7.63957e+7	44	82.6	7.63712e+7	80	150

$\epsilon_{\rm opt} = 10^{-3}$ in the stopping criterion (58), the final accuracy was quite high: 1.3e-4 for Barcelona, 2.8e-4 for Linköping, 4.6e-4 for Winnipeg, 3.5e-4 for Chicago, 9.2e-5 for Skåne.

Acknowledgments. We thank the reviewers and editors for comments that improved the presentation of the paper. We also thank D. Boyce for the Chicago data, H. Edwards for the Skåne data, and A. Ouorou and J.-P. Vial for the CNET data.

References

[1] K. M. Anstreicher and L. Wolsey, *On dual solutions in subgradient optimization*, Tech. report, Dept. of Management Sciences, Univ. of Iowa, Iowa City, IA, September 1993.

[2] L. Bahiense, N. Maculan, and C. A. Sagastizábal, *The volume algorithm revisited: Relation with bundle methods*, Math. Programming **94** (20002), 41–69.

[3] E. Balas and M. C. Carrera, *A dynamic subgradient-based branch-and-bound procedure for set covering*, Oper. Res. **44** (1996), 875–890.

[4] F. Barahona and R. Anbil, *The volume algorithm: Producing primal solutions with a subgradient method*, Math. Programming **87** (2000), 385–399.

[5] ______, *On some difficult linear programs coming from set partitioning*, Discrete Appl. Math. **118** (2002), 3–11.

[6] F. Barahona and D. Chudak, *Solving large scale uncapacitated facility location problems*, Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems (P. M. Pardalos, ed.), Kluwer, Dordrecht, 2000, pp. 48–62.

[7] A. Ben-Tal, T. Margalit, and A. Nemirovski, *The ordered subsets mirror descent optimization method with applications to tomography*, SIAM J. Optim. **12** (2001), 79–108.

[8] D. P. Bertsekas, *Network optimization: Continuous and discrete models*, Athena Scientific, Belmont, MA, 1998.

[9] ______, *Nonlinear programming*, second ed., Athena Scientific, Belmont, MA, 1999.

[10] J. V. Burke and M. C. Ferris, *Characterization of solutions sets of convex programs*, Oper. Res. Lett. **10** (1991), 57–60.

[11] S. Ceria, P. Nobili, and A. Sassano, *A Lagrangian-based heuristic for large-scale set covering problems*, Math. Programming **81** (1998), 215–228.

[12] S. Feltenmark and K. C. Kiwiel, *Dual applications of proximal bundle methods, including Lagrangian relaxation of nonconvex problems*, SIAM J. Optim. **10** (2000), 697–721.

[13] M. Fukushima, *A nonsmooth optimization approach to nonlinear multicommodity network flow problems*, J. Oper. Res. Soc. Japan **27** (1984), 151–176.

[14] ______, *On a dual approach to the traffic assignment problem*, Transportation Res. **18** (1984), 235–245.

[15] G. Gallo and S. Pallotino, *Shortest path algorithms*, Ann. Oper. Res. **13** (1988), 3–79.

[16] J.-L. Goffin, *The ellipsoid method and its predecessors*, Recent Advances in System Modelling and Optimization (L. Contesse, R. Correa, and A. Weintraub, eds.), Lecture Notes in Control and Information Sciences 87, Springer-Verlag, Berlin, 1987, pp. 127–141.

[17] J.-L. Goffin, J. Gondzio, R. Sarkissian, and J.-Ph. Vial, *Solving nonlinear multicommodity flow problems by the analytic center cutting plane method*, Math. Programming **76** (1996), 131–154.

[18] J. Gondzio, R. Sarkissian, and J.-Ph. Vial, *Using an interior point method for the master problem in a decomposition approach*, European J. Oper. Res. **101** (1997), 577–587.

[19] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, Springer, Berlin, 1993.

[20] K. C. Kiwiel, *Proximity control in bundle methods for convex nondifferentiable minimization*, Math. Programming **46** (1990), 105–122.

[21] ______, *Proximal level bundle methods for convex nondifferentiable optimization, saddle-point problems and variational inequalities*, Math. Programming **69** (1995), 89–109.

[22] ______, *The efficiency of subgradient projection methods for convex optimization, part II: Implementations and extensions*, SIAM J. Control Optim. **34** (1996), 677–697.

[23] ______, *Efficiency of proximal bundle methods*, J. Optim. Theory Appl. **104** (2000), 589–603.

[24] ______, *Convergence of approximate and incremental subgradient methods for convex optimization*, SIAM J. Optim. **14** (2004), 807–840.

[25] K. C. Kiwiel, T. Larsson, and P. O. Lindberg, *Dual properties of ballstep subgradient methods, with applications to Lagrangian relaxation*, Tech. Report LiTH-MAT-R-1999-24, Dept. of Mathematics, Linköping Univ., Linköping, Sweden, 1999, Revised October 2002, and December 2005.

[26] ______, *The efficiency of ballstep subgradient level methods for convex optimization*, Math. Oper. Res. **24** (1999), 237–254.

[27] K. C. Kiwiel and P. O. Lindberg, *Parallel subgradient methods for convex optimization*, Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications (D. Butnariu, Y. Censor, and S. Reich, eds.), Studies in Computational Mathematics 8, Elsevier, Amsterdam, 2001, pp. 335–344.

[28] T. Larsson and Z. Liu, *A Lagrangean relaxation scheme for structured linear programs with application to multicommodity network flows*, Optimization **40** (1997), 247–284.

[29] T. Larsson, Z. Liu, and M. Patriksson, *A dual scheme for traffic assignment problems*, Optimization **42** (1997), 323–358.

[30] T. Larsson, M. Patriksson, and C. Rydergren, *Applications of simplicial decomposition with nonlinear column generation to nonlinear network flows*, Network Optimization (P. M. Pardalos, D. W. Hearn, and W. W. Hager, eds.), Lecture Notes in Economics and Mathematical Systems 450, Springer-Verlag, Berlin, 1997, pp. 346–373.

[31] T. Larsson, M. Patriksson, and A.-B. Strömberg, *Ergodic convergence in subgradient optimization*, Optim. Methods Softw. **9** (1998), 93–120.

[32] ______, *Ergodic, primal convergence in dual subgradient schemes for convex programming*, Math. Programming **86** (1999), 283–312.

[33] C. Lemaréchal, A. S. Nemirovskii, and Yu. E. Nesterov, *New variants of bundle methods*, Math. Programming **69** (1995), 111–147.

[34] A. Nedić and D. P. Bertsekas, *Incremental subgradient methods for nondifferentiable optimization*, SIAM J. Optim. **12** (2001), 109–138.

[35] A. Nedić, D. P. Bertsekas, and V. S. Borkar, *Distributed asynchronous incremental subgradient methods*, Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications (D. Butnariu, Y. Censor, and S. Reich, eds.), Studies in Computational Mathematics 8, Elsevier, Amsterdam, 2001, pp. 381–407.

[36] A. Ouorou, P. Mahey, and J.-Ph. Vial, *A survey of algorithms for convex multicommodity flow problems*, Management Sci. **46** (2000), 126–147.

[37] B. T. Polyak, *Minimization of unsmooth functionals*, Zh. Vychisl. Mat. i Mat. Fiz. **9** (1969), 509–521 (Russian), English transl. in U.S.S.R. Comput. Math. and Math. Phys. 9 (1969) 14–29.

[38] ______, *Introduction to optimization*, Nauka, Moscow, 1983, English transl., Optimization Software Inc., New York, 1987.

[39] R. T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, NJ, 1970.

[40] S. Sen and H. D. Sherali, *A class of convergent primal-dual subgradient algorithms for decomposable convex programs*, Math. Programming **35** (1986), 279–297.

[41] H. D. Sherali and G. Choi, *Recovery of primal solutions when using subgradient optimization methods to solve Lagrangian duals of linear programs*, Oper. Res. Lett. **19** (1996), 105–113.

[42] N. Z. Shor, *Minimization methods for non-differentiable functions*, Naukova Dumka, Kiev, 1979 (Russian), English transl., Springer-Verlag, Berlin, 1985.

[43] M. Tatineni, H. Edwards, and D. Boyce, *Comparison of disaggregate simpicial decomposition and Frank-Wolfe algorithms for user-optimal route choice*, Transp. Res. Record **1617** (1998), 157–162.

[44] N. G. Zhurbenko, *Investigation of a class of algorithms for minimizing nonsmooth functions and their application to the solution of large-scale problems*, Dissertation, Cybernetics Institute, Academy of Sciences of the Ukrainian SSR, Kiev, 1977.