Raport Badawczy
Research Report

RB/76/2002

Lagrangian Relaxation via Ballstep Subgradient Methods

K.C. Kiwiel, T. Larsson,
P.O. Lindberg

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 2002

Lagrangian relaxation via ballstep subgradient methods*

Krzysztof C. Kiwiel† Torbjörn Larsson‡ P. O. Lindberg§

December 30, 2002

Abstract

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.

1 Introduction

We consider subgradient methods for the convex constrained minimization problem

$$f_* := \min\{ f(x) : x \in S \} \tag{1.1}$$

under the following assumptions. S is a nonempty closed convex set in $\mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, we can find the value $f(x)$ and a subgradient $g_f(x) \in \partial f(x)$ of f at any $x \in S$, 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 $|\cdot|$. We assume that the *optimal set* $S_* := \operatorname{Arg\,min}_S f$ is nonempty.

The ballstep subgradient method [KLL99b] finds f_* as follows. Given the kth iterate x^k in the feasible set S and a *target level* $f_{\rm lev}^k$ that estimates f_*, it uses the *linearization*

$$f_k(\cdot) := f(x^k) + \left\langle g_f^k, \cdot - x^k \right\rangle \le f(\cdot) \quad \text{with} \quad g_f^k := g_f(x^k) \in \partial f(x^k) \tag{1.2}$$

*Research supported by the Polish State Committee for Scientific Research under Grant 4T11A00622.
†Systems Research Institute, Newelska 6, 01–447 Warsaw, Poland (kiwiel@ibspan.waw.pl)
‡Linköping University, S-58183 Linköping, Sweden (tolar@math.liu.se)
§Linköping University, S-58183 Linköping, Sweden (polin@math.liu.se)

and its halfspace

$$H_k := \left\{x : f_k(x) \le f_{\rm lev}^k\right\} \tag{1.3}$$

to approximate the *level set* of f

$$\mathcal{L}_f(f_{\rm lev}^k) := \{x : f(x) \le f_{\rm lev}^k\} \subset H_k = \mathcal{L}_{f_k}(f_{\rm lev}^k). \tag{1.4}$$

Then, following the original algorithm of [Pol69], it generates the next iterate

$$x^{k+1} := P_S(x^k + t_k[P_{H_k}x^k - x^k]) = P_S(x^k - t_k[f(x^k) - f_{\rm lev}^k]g_f^k/|g_f^k|^2), \tag{1.5}$$

where

$$t_k \in T := [\,t_{\min}, t_{\max}\,] \quad \text{for some fixed} \quad 0 < t_{\min} \le t_{\max} < 2. \tag{1.6}$$

The targets are chosen via a ballstep strategy that works in *groups* of iterations. Within each group, the target $f_{\rm lev}^k$ is fixed, and the method attempts to minimize f over a ball around the best point found so far, shifting the ball and lowering the target when sufficient progress occurs, or shrinking the ball and increasing the target upon discovering that it is too low. The two level schemes of [KLL99b, §§2 and 5] ensure $\inf_k f(x^k) = f_*$ and provide efficiency estimates when the optimal set S_* is bounded. Although (1.5) with the *stepsizes* $\nu_k := t_k[f(x^k) - f_{\rm lev}^k]/|g_f^k|^2$ conforms with the standard subgradient iteration

$$x^{k+1} := P_S(x^k - \nu_k g_f^k) \quad \text{with } \nu_k > 0, \tag{1.7a}$$

such stepsizes needn't obey the popular *divergent series* condition

$$\sum_{k=1}^{\infty} \nu_k = \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \nu_k^2 < \infty \tag{1.7b}$$

or other conditions typically required for convergence of subgradient methods [Kiw03].

In this paper we augment the ballstep method with simple *averaging* schemes that asymptotically find objective and constraint subgradients involved in optimality conditions for problem (1.1). When applied to Lagrangian relaxation of convex programs, they find both primal and dual solutions, and provide practicable stopping criteria. Up till now, for subgradient methods similar results have only been known [Zhu77], [Sho79, §4.4], [AnW93, LaL97, LPS98, LPS99, ShC96] for the iteration (1.7), whose convergence can be slow.

Our results parallel ones in [FeK00] obtained recently for the proximal bundle method [HUL93, §XV.3], [Kiw90]. 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 proximal 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 [FeK00]; to stress the analogies, we quote freely from [FeK00]. Also the efficiency analysis of both methods is quite similar [Kiw00, KLL99b]. Up till now, the literature has only contrasted simple subgradient methods with more advanced proximal bundle methods, whereas our paper highlights their similarities.

Good reviews of the subgradient algorithm may be found in [Ber99, BSS93, Min86, Pol83, Sho79]. It is widely used, mainly due to its simplicity and good performance, especially in Lagrangian relaxation [Bea93]. 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. [BBP00, BaC96, BaC00a, BaC00b, BaL01, CFT96, CNS98].

Also the recent volume algorithm [BaA00] performs well in practice [BaA98]. Its averaging is similar to that of a version of our method that employs past aggregate subgradients to avoid zigzags (cf. (6.2)–(6.3) and Ex. 6.5). However, in contrast with our method, the volume algorithm has no proof of convergence [BMS01]. 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 numerical results for the traffic assignment and message routing problems [Ber98] 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 [OMV00].

The paper is organized as follows. In §2 we review briefly the simplest ballstep method of [KLL99b] and its convergence properties. In §3 we show how averaging may produce affine minorants of f and $\imath_S$ (the indicator of S), and corresponding optimality estimates and stopping criteria. Their uses for indentifying subgradients of f and $\imath_S$ involved in optimality conditions for $\min_S f$ are discussed in §4. Applications to Lagrangian decomposition of convex programs are studied in §5. Extensions to the accelerations of [KLL99b, §7] are given in §6. Applications to multicommodity network flows are reported in §7. The Appendix contains proofs of certain technical results.

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 $C \subset \mathbb{R}^n$.

2 The ballstep level algorithm

In the simplest version of the ballstep subgradient method [KLL99b] stated below, x^k_{rec} is the *record* point with the best objective value f^k_{rec} encountered till iteration k. The iterations are split into *groups* $K_l := \{k(l) : k(l+1)-1\}$, $l \ge 1$. In group l, starting from the point $x^{k(l)}_{\text{rec}}$, the method attempts to reach the *frozen* target level $f^k_{\text{lev}} := f^{k(l)}_{\text{rec}} - \delta_l$ within the ball of *radius* R_l centered at $x^{k(l)}_{\text{rec}}$, where the *level gap* $\delta_l > 0$ controls the stepsize. If *sufficient descent* $f(x^k) \le f^{k(l)}_{\text{rec}} - \frac{1}{2}\delta_l$ occurs, group $l+1$ starts with $\delta_{l+1} := \delta_l$ and $R_{l+1} := R_l$. Otherwise, target infeasibility is eventually discovered when the accumulated sum ρ_{k+1} of squares of subgradient and projection steps grows to about R_l^2 due to oscillations; then group $l+1$ starts with contracted $\delta_{l+1} := \frac{1}{2}\delta_l$ and $R_{l+1} := R_l/2^\beta$, where $\beta \in [0,1)$. Further comments on the rules of the method are given below and in §3; also see [KLL99b].

Algorithm 2.1.
Step 0 (*Initiation*). Select an initial point $x^1 \in S$, a level gap $\delta_1 > 0$, ballstep parameters $R > 0$, $\beta \in [0,1)$, and stepsize bounds $t_{\min}$, $t_{\max}$ (cf. (1.6)). Set $f^0_{\text{rec}} := \infty$, $\rho_1 := 0$. Set the counters $k := l := k(1) := 1$ ($k(l)$ is the iteration number of the lth change of f^k_{lev}).

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

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

Step 3 (*Sufficient descent detection*). If $f(x^k) \le f_{\rm rec}^{k(l)} - \frac{1}{2}\delta_l$, 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 $f_{\rm lev}^k := f_{\rm rec}^{k(l)} - \delta_l$. Choose the relaxation factor $t_k \in T$ (cf. (1.6)). Set $x^{k+1/2} := x^k + t_k(P_{H_k}x^k - x^k)$, $\check\rho_k := t_k(2-t_k)d_{H_k}^2(x^k)$, $\rho_{k+1/2} := \rho_k + \check\rho_k$, $x^{k+1} := P_S x^{k+1/2}$, $\check\rho_{k+1/2} := |x^{k+1} - x^{k+1/2}|^2$, $\rho_{k+1} := \rho_{k+1/2} + \check\rho_{k+1/2}$.

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

$$(R_l - |x^{k+1} - x^{k(l)}|)^2 > R_l^2 - \rho_{k+1}, \tag{2.1}$$

i.e., the target level is too low (see below), then go to Step 6; otherwise, go to Step 7.

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 $x_{\rm rec}^k$ and g_f^k by $g_f(x_{\rm rec}^k)$, increase l by 1 and go to Step 4.

Step 7. Increase k by 1 and go to Step 1.

Assuming the method doesn't terminate, we now recall some results of [KLL99b, §2–3].

Remarks 2.2. (i) If at least half of the *desired objective reduction* δ_l is achieved at Step 3, group $l+1$ starts with the same $\delta_{l+1} := \delta_l$, but $f_{\rm rec}^{k(l+1)} \le f_{\rm rec}^{k(l)} - \frac{1}{2}\delta_l$ with $x^{k(l+1)} = x_{\rm rec}^{k(l+1)}$ (since $f(x^k) > f_{\rm rec}^{k(l)} - \frac{1}{2}\delta_l \ \forall k \in K_l$). Thus by Step 6, we have $\delta_{l+1} \le \delta_l$, $x^{k(l)} = x_{\rm rec}^{k(l)} \in S$ and $f_{\rm rec}^{k(l)} = f(x^{k(l)})$ for all l. Hence $\inf_l f(x^{k(l)}) \ge f_* > -\infty$ gives $\delta_l \downarrow 0$ [KLL99b, Lem. 3.6] (otherwise we would have $f(x^{k(l)}) \downarrow -\infty$); in particular, the target infeasibility test (2.1) is met for infinitely many l such that $\delta_{l+1} := \frac{1}{2}\delta_l$ at Step 6.

(ii) At Step 4, $x^{k+1/2} = x^k - t_k[f(x^k) - f_{\rm lev}^k]g_f^k/|g_f^k|^2$ and $d_{H_k}(x^k) = [f(x^k) - f_{\rm lev}^k]/|g_f^k|$ with $f(x^k) > f_{\rm lev}^k$, so the *Fejér* quantities $\check\rho_k$, $\rho_{k+1/2}$ and ρ_{k+1} are positive, since ρ_k may only decrease to 0 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; $R_l \equiv R$ if $\beta = 0$. Ideally, R_l should be of order $d_{S_*}(x^{k(l)})$, and hence shrink as $x^{k(l)}$ approaches S_*. For convergence it suffices to choose R_l so that $\delta_l/R_l \to 0$ [KLL99b, Rem. 3.9(i)].

(iv) Algorithm 2.1 is a *ballstep* method, which in group l attempts to minimize f approximately over the intersection of the ball $B(x^{k(l)}, R_l)$ with the feasible set S, shifting the ball when sufficient progress occurs, or increasing the target level otherwise. By [KLL99b, Lem. 3.1(v)] or Lem. 3.1(iv,v) in §3, the target infeasibility test (2.1) implies

$$f_{\rm lev}^k := f_{\rm rec}^{k(l)} - \delta_l < \min\{\, f(x) : x \in B(x^{k(l)}, R_l) \cap S\,\}, \tag{2.2}$$

i.e., the target is too low, in which case δ_l is halved at Step 6, $f_{\rm lev}^k$ is increased at Step 4 and x^{k+1} is recomputed. Note that l increases at Step 6, but k does not, so relations like $f_{\rm lev}^k := f_{\rm rec}^{k(l)} - \delta_l$ always involve the *current values* of the counters k and l at Step 4.

(v) Since $|x^{k+1} - x^{k(l)}| > 2R_l$ suffices for passing the infeasibility test (2.1), this test also ensures at Step 1 the basic *local boundedness* property: $\{x^k\}_{k\in K_l} \subset B(x^{k(l)}, 2R_l)$.

(vi) By [KLL99b, Thm 3.7 and Cor. 3.8], $f(x^{k(l)}) \downarrow f_*$ and each cluster point of $\{x^{k(l)}\}$ lies in the optimal set S_*; moreover, $\{x^{k(l)}\}$ is bounded if S_* is bounded. These results only require finiteness of f on S and local boundedness of g_f on S [KLL99b, Rem. 3.9(ii)].

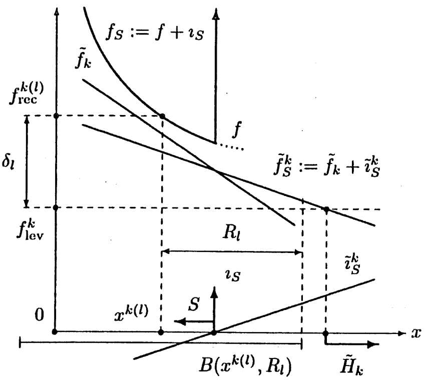

Figure 3.1: Target infeasibility $f_{\rm lev}^k < \min_{B(x^{k(l)},R_l)} f_S$ if $d_{\tilde{H}_k}(x^{k(l)}) > R_l$.

3 Dual subgradient interpretations

For theoretical purposes, it is convenient to regard our constrained problem $f_* := \min_S f$ (cf. (1.1)) as the unconstrained problem $f_* = \min f_S$ with the *essential objective*

$$f_S := f + \imath_S, \tag{3.1}$$

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

We now outline our main results. Suppose iteration k detects target infeasibility $f_{\rm lev}^k < f_*^l := \min_{B(x^{k(l)},R_l)} f_S$ (i.e., (2.2)) via the *Fejér* test (2.1). First, we construct affine minorants $\tilde{f}_k$ and $\tilde{\imath}_S^k$ of f and $\imath_S$ by combining their past subgradient linearizations with suitable weights. Then $\tilde{f}_S^k := \tilde{f}_k + \tilde{\imath}_S^k$ is an affine minorant of $f_S := f + \imath_S$ and hence $\mathcal{L}_{f_S}(f_{\rm lev}^k) \subset \tilde{H}_k := \mathcal{L}_{\tilde{f}_S^k}(f_{\rm lev}^k)$, so that $f_{\rm lev}^k < f_*^l$ if $B(x^{k(l)}, R_l) \cap \tilde{H}_k = \emptyset$ (see Fig. 3.1); the latter condition is shown to be equivalent to (2.1) by fairly simple algebra. Next, we get $\nabla \tilde{f}_S^k \in \partial_{\delta_l} f_S(x^{k(l)})$ with $|\nabla \tilde{f}_S^k| \le \delta_l / R_l$ as in Fig. 3.1; since $\delta_l \to 0$ and $\delta_l/R_l \to 0$, this ensures asymptotic optimality and suggests practical stopping criteria.

3.1 Aggregate linearizations

We first derive a dual subgradient interpretation of the test (2.1) by identifying below *linearizations* (affine minorants) $\tilde{f}_k$, $\tilde{\imath}_S^k$, $\tilde{f}_S^k$ of f, $\imath_S$, f_S, respectively. At Step 4, let

$$\nu_k := t_k[f_k(x^k) - f_{\rm lev}^k]/|g_f^k|^2, \tag{3.2}$$

$$g_S^k := x^{k+1/2} - x^{k+1}, \tag{3.3}$$

$$\nu_j^k := \nu_j / \tilde{\nu}_f^k \quad \text{for} \quad j = k(l){:}\,k \quad \text{with} \quad \tilde{\nu}_f^k := \sum_{j=k(l)}^{k} \nu_j. \tag{3.4}$$

Here ν_k is the *subgradient stepsize* such that $x^{k+1} = P_S(x^k - \nu_k g_f^k)$ (cf. Rem. 2.2(ii)), g_S^k is the *constraint subgradient* of $\imath_S$ at x^{k+1} stemming from $x^{k+1} := P_S x^{k+1/2}$, $\tilde{\nu}_f^k$ is the *cumulative stepsize*, whereas $\{\nu_j^k\}_{j=k(l)}^k$ are positive *convex weights* summing to 1. We shall employ the following *aggregate linearizations* of f, $\imath_S$ and f_S (cf. (3.1)):

$$\tilde{f}_k(\cdot) := \sum_{j=k(l)}^{k} \nu_j^k f_j(\cdot), \quad \tilde{\imath}_S^k(\cdot) := \sum_{j=k(l)}^{k} \left\langle g_S^j, \cdot - x^{j+1} \right\rangle / \tilde{\nu}_f^k, \quad \tilde{f}_S^k(\cdot) := \tilde{f}_k(\cdot) + \tilde{\imath}_S^k(\cdot), \tag{3.5}$$

and the corresponding *aggregate halfspace* and the *aggregate level*

$$\tilde{H}_k := \mathcal{L}_{\tilde{f}_S^k}(\tilde{f}_{\text{lev}}^k) = \{\, x : \tilde{f}_S^k(x) \le \tilde{f}_{\text{lev}}^k \,\} \quad \text{with} \quad \tilde{f}_{\text{lev}}^k := \sum_{j=k(l)}^{k} \nu_j^k f_{\text{lev}}^j. \tag{3.6}$$

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

Lemma 3.1. (i) *At Step* 4,

$$x^{k+1} - x^{k(l)} = - \sum_{j=k(l)}^{k} (\nu_j g_f^j + g_S^j), \tag{3.7}$$

$$L_k := -\tfrac{1}{2}\left| \sum_{j=k(l)}^{k} \nu_j g_f^j + g_S^j \right|^2 + \sum_{j=k(l)}^{k} \left\{ \nu_j [f_j(x^{k(l)}) - f_{\text{lev}}^j] + \left\langle g_S^j, x^{k(l)} - x^{j+1} \right\rangle \right\} = \tfrac{1}{2}\rho_{k+1}. \tag{3.8}$$

(ii) $\tilde{f}_k \le f$, $\tilde{\imath}_S^k \le \imath_S$, $\tilde{f}_S^k \le f_S$. *Further,* $\tilde{\nu}_f^k \nabla \tilde{f}_S^k = x^{k(l)} - x^{k+1}$,

$$2\tilde{\nu}_f^k [\tilde{f}_S^k(x^{k(l)}) - \tilde{f}_{\text{lev}}^k] = |x^{k+1} - x^{k(l)}|^2 + \rho_{k+1}. \tag{3.9}$$

(iii) $\tilde{f}_S^k(x^{k(l)}) > \tilde{f}_{\text{lev}}^k$,

$$d_{\tilde{H}_k}(x^{k(l)}) = [\tilde{f}_S^k(x^{k(l)}) - \tilde{f}_{\text{lev}}^k] / |\nabla \tilde{f}_S^k| \ge \rho_{k+1}^{1/2}. \tag{3.10}$$

(iv) *Let* $f_*^l := \min_{B(x^{k(l)}, R_l)} f_S$. *If* $\tilde{f}_{\text{lev}}^k \ge f_*^l$, *then* $d_{\tilde{H}_k}(x^{k(l)}) \le R_l$. *Consequently,* $\tilde{f}_{\text{lev}}^k < f_*^l$ *if* $d_{\tilde{H}_k}(x^{k(l)}) > R_l$.

(v) $d_{\tilde{H}_k}(x^{k(l)}) > R_l$ *iff* $(R_l - |x^{k+1} - x^{k(l)}|)^2 > R_l^2 - \rho_{k+1}$.

Proof. (i) Since at Step 4 (cf. Rem. 2.2(ii), (3.2)–(3.3)) $x^{k+1/2} - x^k = -\nu_k g_f^k$ and $x^{k+1} - x^{k+1/2} = -g_S^k$, (3.7) follows by induction. Let $\Delta L_k := L_k - L_{k-1}$. Since $x^k - x^{k(l)} = -\sum_{j=k(l)}^{k-1} (\nu_j g_f^j + g_S^j)$ in (3.8), using (1.2)–(1.3), $x^{k+1/2} - x^k = -\nu_k g_f^k$ and (3.2), we get

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

(cf. Step 4), so (3.8) follows by induction, using $L_{k(l)-1} := \rho_{k(l)} := 0$ (cf. Step 6).

(ii) At Step 4 (cf. (3.2), (1.2)) $\nu_k > 0$, so $\nu_j^k > 0$, $\sum_{j=k(l)}^k \nu_j^k = 1$ in (3.4). Use (cf. (1.2)) $f_j \le f$, (3.3) with (cf. Step 4) $x^{j+1} := P_S x^{j+1/2}$ and the well-known projection property

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

to get (cf. (3.5)) $\tilde f_k \le f$, $\tilde\imath_S^k \le \imath_S$ via positive combinations, and hence $\tilde f_S^k := \tilde f_k + \tilde\imath_S^k \le f + \imath_S =: f_S$. Since (cf. (3.4), (3.5), (3.7)) $\tilde\nu_f^k \nabla \tilde f_S^k = \sum_{j=k(l)}^k \nu_j g_f^j + g_S^j = x^{k(l)} - x^{k+1}$ and

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

(cf. (3.4)–(3.6)), (3.9) follows from (cf. (3.8) and Rem. 2.2(ii))

$$L_k = -\tfrac12|\tilde\nu_f^k \nabla \tilde f_S^k|^2 + \tilde\nu_f^k[\tilde f_S^k(x^{k(l)}) - \tilde f_{\rm lev}^k] = \tfrac12 \rho_{k+1} > 0. \tag{3.11}$$

(iii) By (3.11), $L_k = -\frac12 a^2 + b = \frac12 c^2$ 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 $b = \frac12(a^2 + c^2) \ge |ac|$, so (cf. (3.6)) $d_{\tilde H_k}(x^{k(l)}) = b/a \ge c > 0$.

(iv) If $f_*^l \le \tilde f_{\rm lev}^k$, then $\operatorname{Arg\,min}_{B(x^{k(l)}, R_l)} f_S \subset \tilde H_k$ from (3.6) and $\tilde f_S^k \le f_S$ (cf. (ii)).

(v) $(R_l - |x^{k+1} - x^{k(l)}|)^2 > R_l^2 - \rho_{k+1} \Leftrightarrow |x^{k+1} - x^{k(l)}|^2 + \rho_{k+1} > 2R_l|x^{k+1} - x^{k(l)}| \Leftrightarrow 2\tilde\nu_f^k[\tilde f_S^k(x^{k(l)}) - \tilde f_{\rm lev}^k] > 2R_l \tilde\nu_f^k|\nabla \tilde f_S^k| \Leftrightarrow [\tilde f_S^k(x^{k(l)}) - \tilde f_{\rm lev}^k]/|\nabla \tilde f_S^k| > R_l \Leftrightarrow d_{\tilde H_k}(x^{k(l)}) > R_l$, where we have used (3.9), $|x^{k+1} - x^{k(l)}| = \tilde\nu_f^k|\nabla f_S^k|$ (cf. (ii)) and (3.10). □

Remarks 3.2. (i) By Lem. 3.1(v), the Fejér test (2.1) is *equivalent* to the *distance* test

$$d_{\tilde H_k}(x^{k(l)}) > R_l. \tag{3.12}$$

The fact that the Fejér test (2.1) implies $f_{\rm lev}^k < f_*^l$ (cf. (2.2)) was derived in [KLL99b, Lem. 3.1(v)] from Fejér estimates via analytic arguments, which are quite difficult to interpret. In contrast, the distance test (3.12) has a straightforward interpretation: with $\tilde f_{\rm lev}^k = f_{\rm lev}^k$ in (3.6), (3.12) means that the minimum of $\tilde f_S^k$ over $B(x^{k(l)}, R_l)$, and hence also that of f_S (since $\tilde f_S^k$ underestimates f_S), is greater than $f_{\rm lev}^k$, i.e., $f_{\rm lev}^k < f_*^l$.

(ii) To cover the modifications of [KLL99b, §6], which need not use $f_{\rm lev}^j = f_{\rm lev}^k$ for $j = k(l){:}\,k$ (cf. §5), note that the proof of Lem. 3.1 holds if at Step 4 we only have

$$f_{\rm rec}^{k(l)} - \delta_l \le f_{\rm lev}^k < \min\{f_{\rm rec}^{k(l)}, f(x^k)\}. \tag{3.13}$$

In general, since (cf. (3.6), (3.4)) $\tilde f_{\rm lev}^k \ge \min_{j=k(l)}^k f_{\rm lev}^j$, if $\min_{j=k(l)}^k f_{\rm lev}^j \ge f_{\rm rec}^{k(l)} - \delta_l$ then (3.12) yields $f_{\rm rec}^{k(l)} - \delta_l < f_*^l$; thus Lem. 3.1(iv,v) subsumes [KLL99b, Lem. 3.1(v)].

(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

$$\min\left\{\tfrac12|x - x^{k(l)}|^2 : f_j(x) \equiv f_j(x^{k(l)}) + \left\langle g_f^j, x - x^{k(l)}\right\rangle \le f_{\rm lev}^j,\ j = k(l){:}\,k\right\}. \tag{3.14}$$

Indeed, the Lagrangian of (3.14) with multipliers ν_j is minimized by x^{k+1} (cf. (3.7)) to give the dual function value L_k (cf. (3.8)), and $\nu_k = t_k \check\nu_k$ (cf. (3.2)), where $\check\nu_k := [f_k(x^k) - f_{\rm lev}^k]/|g_f^k|^2$ maximizes $\Delta L_k = -\frac12|\nu_k g_f^k|^2 + \nu_k[f_k(x^k) - f_{\rm lev}^k]$ (cf. the proof of Lem. 3.1(i)). Thus our algorithm may be regarded as a poor man's version of proximal level methods [Kiw95, LNN95] that employ subproblem (3.14) with $f_{\rm lev}^j = f_{\rm lev}^k$.

3.2 Global optimality estimates

We now derive some global optimality estimates from the aggregate linearizations $\tilde f_k$, $\tilde\imath_S^k$ and $\tilde f_S^k$ (cf. (3.5)). The latter are described by their constant gradients, as well as their *linearization errors* at $x^{k(l)}$ (cf. Fig. 3.1):

$$\tilde\epsilon_f^k := f(x^{k(l)}) - \tilde f_k(x^{k(l)}), \quad \tilde\epsilon_S^k := -\tilde\imath_S^k(x^{k(l)}), \quad \tilde\epsilon_k := f(x^{k(l)}) - \tilde f_S^k(x^{k(l)}); \tag{3.15}$$

note that $\imath_S(x^{k(l)}) = 0$ and $f_S(x^{k(l)}) = f(x^{k(l)})$ from $x^{k(l)} \in S$. Suppose that (3.13) holds.

Lemma 3.3. *We have* $\nabla\tilde f_k \in \partial_{\tilde\epsilon_f^k} f(x^{k(l)})$, $\nabla\tilde\imath_S^k \in \partial_{\tilde\epsilon_S^k}\imath_S(x^{k(l)})$, $\nabla\tilde f_S^k \in \partial_{\tilde\epsilon_k} f_S(x^{k(l)})$, *with* $\tilde\epsilon_f^k \ge 0$, $\tilde\epsilon_S^k \ge 0$, $\tilde\epsilon_k = \tilde\epsilon_f^k + \tilde\epsilon_S^k \ge 0$. *Further,* $\tilde f_S^k(x^{k(l)}) > f_{\rm lev}^k \ge \min_{j=k(l)}^k f_{\rm lev}^j \ge f_{\rm rec}^{k(l)} - \delta_l$,

$$f_S(x) \ge \tilde f_S^k(x) = f(x^{k(l)}) - \tilde\epsilon_k + \left\langle \nabla\tilde f_S^k, x - x^{k(l)} \right\rangle \qquad \forall x, \tag{3.16}$$

$$\tilde\epsilon_k := f(x^{k(l)}) - \tilde f_S^k(x^{k(l)}) < f_{\rm rec}^{k(l)} - f_{\rm lev}^k \le \delta_l, \tag{3.17}$$

$$|\nabla\tilde f_S^k| = [\tilde f_S^k(x^{k(l)}) - f_{\rm lev}^k]/d_{\tilde H_k}(x^{k(l)}) \le \delta_l/d_{\tilde H_k}(x^{k(l)}), \tag{3.18}$$

$$f(x) \ge \tilde f_k(x) \ge \tilde f_S^k(x) \qquad \forall x \in S, \tag{3.19}$$

$$f(x^{k(l)}) - f_S(x) \le \delta_l \max\{|x - x^{k(l)}|/d_{\tilde H_k}(x^{k(l)}), 1\} \qquad \forall x. \tag{3.20}$$

Proof. By (3.5) and Lem. 3.1(ii), $\tilde f_k$ is an affine minorant of f; thus, by (3.15),

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

means $\nabla\tilde f_k \in \partial_{\tilde\epsilon_f^k} f(x^{k(l)})$ with $\tilde\epsilon_f^k \ge 0$. Arguing similarly for $\tilde\imath_S^k$ and $\tilde f_S^k$ yields the first assertion, (3.16), and (3.19) (since $\tilde\imath_S^k(x) \le \imath_S(x) = 0$ $\forall x \in S$). The inequalities in (3.17) stem from $f(x^{k(l)}) = f_{\rm rec}^{k(l)}$ (Rem. 2.2(i)), $\tilde f_S^k(x^{k(l)}) > f_{\rm lev}^k$ (Lem. 3.1(iii)) and (cf. (3.6), (3.4), (3.13)) $f_{\rm lev}^k \ge \min_{j=k(l)}^k f_{\rm lev}^j \ge f_{\rm rec}^{k(l)} - \delta_l$. Then (3.18) follows from (3.10), $\tilde f_S^k(x^{k(l)}) \le f_S(x^{k(l)}) = f_{\rm rec}^{k(l)}$ (Lem. 3.1(ii) and Rem. 2.2(i)) and the final inequality of (3.17). To prove (3.20), use (3.16)–(3.18) and $\epsilon := \tilde\epsilon_k/(f_{\rm rec}^{k(l)} - f_{\rm lev}^k) \in [0,1]$ to develop

$$\begin{aligned} f(x^{k(l)}) - f_S(x) &\le \tilde\epsilon_k + \left\langle \nabla\tilde f_S^k, x - x^{k(l)} \right\rangle \le \tilde\epsilon_k + |\nabla\tilde f_S^k||x - x^{k(l)}| \\ &= (f_{\rm rec}^{k(l)} - f_{\rm lev}^k)[\epsilon 1 + (1-\epsilon)|x - x^{k(l)}|/d_{\tilde H_k}(x^{k(l)})] \\ &\le \delta_l \max\{1, |x - x^{k(l)}|/d_{\tilde H_k}(x^{k(l)})\}. \qquad \square \end{aligned}$$

Remark 3.4. At Step 6, (3.12) and (3.20) yield the estimate of [KLL99b, Lem. 3.3]:

$$f(x^{k(l)}) - f_S(x) \le \delta_l \max\{|x - x^{k(l)}|/R_l, 1\} = \max\{|x - x^{k(l)}|\delta_l^{1-\beta}\delta_1^\beta/R, \delta_l\} \qquad \forall x.$$

3.3 First asymptotic results

Our asymptotic results will concentrate on Step 6, using the groups and iterations

$$L := \{\, l : \delta_{l+1} = \tfrac{1}{2}\delta_l \,\} \quad \text{and} \quad K := \{\, k(l+1) : l \in L \,\}. \tag{3.21}$$

Thus L comprises groups l terminating at Step 6 when the distance test (3.12) ($\equiv$(2.1) by Rem. 3.2(i)) holds at Step 5 with $k = k(l+1)$ in the set of "interesting" iterations K. Of course, it would be nice to have results for the remaining iterations as well, but our estimates (3.18) and (3.20) involve the quantities $\delta_l/d_{\tilde H_k}(x^{k(l)})$, which in general converge to 0 only for $k \in K$, 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 (3.5). First, we show that their errors $\tilde\epsilon_f^k$, $\tilde\epsilon_S^k$, $\tilde\epsilon_k$ (cf. (3.15)), as well as the gradient of $\tilde f_S^k$, vanish for $k \in K$. Our further results will require local boundedness of the gradient of $\tilde f_k$. Since $\nabla \tilde f_k$ is a convex combination of the past subgradients $\{g_f^j\}_{j=k(l)}^k$ (cf. (3.4)–(3.5)), its local boundedness will follow from the local boundedness of g_f^k.

Lemma 3.5. (i) *In the notation of* (3.15), (3.5) *and* (3.21), *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* $\{x^{k(l)}\}_{l\in L}$ *has a cluster point* x^∞ *such that* $x^{k(l)} \xrightarrow{L'} x^\infty$ *with* $L' \subset L$, *and let* $K' := \{k(l+1) : l \in L'\}$ *(cf.* (3.21)*). Then* $x^\infty \in S_*$, $f(x^{k(l)}) \downarrow f_* = f(x^\infty)$, *and both* $\{x^k\}_{k\in K_l, l\in L'}$ *and* $\{g_f^k\}_{k\in K_l, l\in L'}$ *are bounded, where* $K_l := \{k(l){:}\, k(l+1)-1\}$.

Proof. (i) By Lem. 3.3 and (3.17), $0 \le \tilde\epsilon_f^k, \tilde\epsilon_S^k, \tilde\epsilon_k \le \delta_l \downarrow 0$ (cf. Rem. 2.2(i)). Then $|\nabla \tilde f_S^k| \le \delta_l/d_{\tilde H_k}(x^{k(l)})$ (cf. (3.18)) with $d_{\tilde H_k}(x^{k(l)}) > R_l$ for $k = k(l+1)$ (cf. (3.12)), $R_l := R(\delta_l/\delta_1)^\beta$ (cf. Step 5) and $\beta \in [0,1)$ (cf. Step 0) give $\delta_l/R_l \to 0$ and hence $\nabla \tilde f_S^k \xrightarrow{K} 0$.

(ii) Of course, $x^\infty \in S_*$ by Rem. 2.2(vi), but (3.16) combined with (i) and the fact that $\{x^k\}$ 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 $\{x^k\}_{k\in K_l} \subset B(x^{k(l)}, 2R_l)$ (Rem. 2.2(v)), since $g_f^k := g_f(x^k)$ and g_f is locally bounded on S. □

Remark 3.6. To relate our preceding results with those of [FeK00], let $p_f^k := \nabla \tilde f_k$, $p_S^k := \nabla \tilde\imath_S^k$, $p^k := \nabla \tilde f_S^k$. Our $x^{k(l)}$, x^k and the index set $k(l){:}\,k$ usually correspond to x^k, y^k and J^k in [FeK00]. In this notation, Lem. 3.3 corresponds to [FeK00, Lem. 3.2]. However, our Lem. 3.5 says less than [FeK00, Lem. 3.3]. First, we only have $\nabla \tilde f_S^k \xrightarrow{K} 0$, instead of $p^k \to 0$. Second, in [FeK00, Lem. 3.3(ii)], $x^k \to x^\infty$ and $y^k \to x^\infty$ with $g_f(y^k)$ bounded, i.e., "everything converges" and "everything is bounded", whereas Lem. 3.5(ii) only speaks about suitable subsequences. Hence using the analysis of [FeK00] one may derive "subsequential" versions of the remaining results of [FeK00], as will be seen below.

3.4 Stopping criteria

The usual stopping criterion $\delta_l \le \epsilon_{\rm opt}(1+|f_{\rm rec}^k|)$ with $\epsilon_{\rm opt} > 0$ [KLL99b, Rem. 3.4(iii)] tends to work quite well, but it does not *guarantee* that $f(x_{\rm rec}^k) - f_* \le \epsilon_{\rm opt}(1+|f_{\rm rec}^k|)$ upon termination. The following result may be used for developing alternative stopping criteria when the feasible set S is bounded, as happens in many applications.

Lemma 3.7. *Suppose the feasible set S is bounded. Let $\tilde f_{\min}^k := \min_S \tilde f_S^k$ for all $k \ge 1$. Then $\tilde f_{\min}^k \le \min_S \tilde f_k \le f_*$ for all k, and $\tilde f_{\min}^k \xrightarrow{K} f_*$, where K is given by* (3.21).

Proof. Use Rem. 3.6 and the proof of [FeK00, Lem. 3.5], or see the Appendix. □

Remark 3.8. When S is bounded and simple enough, we may compute the lower bounds

$$\tilde f_{\rm low}^k := \max\left\{\min_S \tilde f_k, \tilde f_{\rm low}^{k-1}\right\} \quad \text{for} \quad k \ge 1, \quad \text{with} \quad \tilde f_{\rm low}^0 := -\infty. \tag{3.22}$$

Since $\tilde f_{\rm low}^k \uparrow f_*$ (cf. Lem. 3.7), whereas $f_{\rm rec}^k \downarrow f_*$ (Rem. 2.2(vi)), for any $\epsilon > 0$ there is k such that $f_{\rm rec}^k - \tilde f_{\rm low}^k \le \epsilon$ (implying $f(x_{\rm rec}^k) \le f_* + \epsilon$). This validates a stopping criterion of the form $f_{\rm rec}^k - \tilde f_{\rm low}^k \le \epsilon$. Note that it is better to use $\tilde f_k$ instead of $\tilde f_S^k$ in (3.22), since $\tilde f_k \ge \tilde f_S^k$ on S (cf. (3.19)). If the computation of $\min_S \tilde f_k$ is difficult, but it is easier to find $\min_{\tilde S} \tilde f_S^k$ for some "simpler" bounded set $\tilde S \supset S$, then $\min_{\tilde S} \tilde f_S^k$ may replace $\min_S \tilde f_k$ in (3.22) (since $\min_{\tilde S} \tilde f_S^k \le f_*$ and $\min_{\tilde S} \tilde f_S^k \xrightarrow{K} f_*$ by the proof of Lem. 3.7 with S replaced by $\tilde S$); in fact it may be more efficient to use $\tilde f_{\rm low}^k := \max\{\min_{\tilde S} \tilde f_S^k, \min_{\tilde S} \tilde f_k, \tilde f_{\rm low}^{k-1}\}$.

3.5 Ballstep modifications

We now consider two more efficient modifications of [KLL99b].

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

$$(R_l - |x^{k+1/2} - x^{k(l)}|)^2 > R_l^2 - \rho_{k+1/2}, \tag{3.23}$$

replacing (2.1) by "(3.23) or (2.1)". In view of the results of [KLL99b, §3], Step 4 may set $x^{k+1} := x^{k+1/2}$ if (3.23) holds, so that $\rho_{k+1} = \rho_{k+1/2}$ and (2.1) holds; then all the preceding and subsequent results remain valid. Further, we may replace $x^{k+1/2}$ and $\rho_{k+1/2}$ in (3.23) by $P_{H_k}x^k$ and $\rho_k + d_{H_k}^2(x^k)$, as if $t_k = 1$ [KLL99b, Rem. 3.2(ii)].

Similarly, our preceding and subsequent results hold for the "true" ballstep version of [KLL99b, Lem. 3.10], which additionally projects x^{k+1} on $B(x^{k(l)}, R_l)$ to ensure $\{x^k\}_{k\in K_l} \subset B(x^{k(l)}, R_l)$ (instead of $\{x^k\}_{k\in K_l} \subset B(x^{k(l)}, 2R_l)$ as before). Since this only needs more complicated notation, we refer the interested readers to [KLL99a, Lem. 3.10].

4 Optimal objective and constraint subgradients

Returning to the asymptotic setting of Lem. 3.5, let x^∞ be an arbitrary cluster point of $\{x^{k(l)}\}_{l\in L}$ corresponding to groups L' and iterations K' such that (cf. (3.21))

$$x^{k(l)} \xrightarrow{L'} x^\infty \quad \text{with} \quad L' \subset L := \{\, l : \delta_{l+1} = \tfrac{1}{2}\delta_l \,\},\ K' := \{k(l+1) : l \in L'\} \subset K. \tag{4.1}$$

We now show that the corresponding subsequences of the *aggregate subgradients* $\nabla\tilde{f}_k$ and $-\nabla\tilde{\imath}_S^k$ converge to the *optimal subgradient set* of our problem $\min_S f$:

$$\mathcal{G} := \partial f(x^\infty) \cap -\mathcal{N}_S(x^\infty); \tag{4.2}$$

this set does not depend on x^∞ ($\mathcal{G} = \partial f(x) \cap -\mathcal{N}_S(x)\ \forall x \in S_*$ [BuF91, Lem. 2]) and is closed convex (so are $\partial f(x^\infty)$ and $\partial\imath_S(x^\infty)$). We also show that $\tilde{f}_k$ and $\tilde{\imath}_S^k$ converge to the corresponding set of "optimal" linearizations of f and $\imath_S$ at x^∞. Similarly to [FeK00, Thm 3.4], this fairly abstract result will form the basis of the more concrete results of §5.

Theorem 4.1. *Suppose $\{x^{k(l)}\}_{l\in L}$ has a cluster point x^∞ such that $x^{k(l)} \xrightarrow{L'} x^\infty$ with $L' \subset L$, and let $K' := \{k(l+1) : l \in L'\}$. Then:*

(i) *$\{\nabla\tilde{f}_k\}_{k\in K'}$ is bounded and each cluster point of $\{\nabla\tilde{f}_k\}_{k\in K'}$ lies in $\partial f(x^\infty)$.*

(ii) *Let $\nabla\tilde{f}_\infty$ be a cluster point of $\{\nabla\tilde{f}_k\}_{k\in K'}$. Let $K'' \subset K'$ be such that $\nabla\tilde{f}_k \xrightarrow{K''} \nabla\tilde{f}_\infty$. Then $\nabla\tilde{f}_\infty \in \mathcal{G}$. Moreover,*

$$\nabla\tilde{\imath}_S^k \xrightarrow{K''} \nabla\tilde{\imath}_S^\infty, \quad \tilde{f}_k(\cdot) \xrightarrow{K''} \tilde{f}_\infty(\cdot) \quad \textit{and} \quad \tilde{\imath}_S^k(\cdot) \xrightarrow{K''} \tilde{\imath}_S^\infty(\cdot),$$

where

$$\nabla\tilde{\imath}_S^\infty := -\nabla\tilde{f}_\infty \in \mathcal{N}_S(x^\infty), \quad \tilde{f}_\infty(\cdot) := f(x^\infty) + \left\langle\nabla\tilde{f}_\infty, \cdot - x^\infty\right\rangle, \quad \tilde{\imath}_S^\infty(\cdot) := \langle\nabla\tilde{\imath}_S^\infty, \cdot - x^\infty\rangle.$$

(iii) *$\{\nabla\tilde{\imath}_S^k\}_{k\in K'}$ is bounded and each cluster point of $\{\nabla\tilde{\imath}_S^k\}_{k\in K'}$ lies in $\mathcal{N}_S(x^\infty)$.*

(iv) *$d_\mathcal{G}(\nabla\tilde{f}_k) \xrightarrow{K'} 0$ and $d_\mathcal{G}(-\nabla\tilde{\imath}_S^k) \xrightarrow{K'} 0$.*

Proof. Use Rem. 3.6 and the proof of [FeK00, Thm 3.4], or see the Appendix. □

Corollary 4.2. *If $\{x^{k(l)}\}$ is bounded (e.g., so is S_*), then $\{\nabla\tilde{f}_k\}_{k\in K}$ and $\{-\nabla\tilde{\imath}_S^k\}_{k\in K}$ are bounded, their cluster points lie in $\mathcal{G}$, $d_\mathcal{G}(\nabla\tilde{f}_k) \xrightarrow{K} 0$ and $d_\mathcal{G}(-\nabla\tilde{\imath}_S^k) \xrightarrow{K} 0$.*

Proof. This follows from Rem. 2.2(vi) and Thm 4.1. □

Concerning Thm 4.1 and Cor. 4.2, note that $\{x^{k(l)}\}$ is bounded if so is the feasible set S; also having S bounded is useful for stopping criteria (cf. Rem. 3.8). As observed in [FeK00, §3], in some applications (cf. Rem. 5.4(ii)), one wants to find $\min_{\breve{S}} f$ for an unbounded set $\breve{S}$, but one can find a bounded set $\bar{S}$ that intersects $\operatorname{Arg\,min}_{\breve{S}} f$. Then it is natural to solve, instead of the original problem $\min_{\breve{S}} f$, its restricted version $\min_S f$ with $S = \breve{S} \cap \bar{S}$ bounded. Both problems have the same optimal subgradient set $\mathcal{G}$ if the "bounding" set $\bar{S}$ is "large enough", as explained below.

Fact 4.3 ([FeK00, Lem. 3.7]). *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 4.4. Under the assumptions of Fact 4.3, $\mathcal{N}_{\breve{S}}$ may replace $\mathcal{N}_S$ in Thm 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 Cor. 4.2. 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 [FeK00, §4], the results of this section can be specialized [KLL99a, §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 [AnW93, LPS98]; cf. [KLL99a, Rem. 5.15].

5 Lagrangian relaxation

Following [FeK00, §5], we now consider the special case where problem (1.1) (i.e., $\min_S f$) is the Lagrangian dual problem of the following *primal* convex optimization problem:

$$\psi_0^{\max} := \max\ \psi_0(z) \quad \text{s.t.} \quad \psi_j(z) \geq 0,\ j = 1{:}n,\ z \in Z, \tag{5.1}$$

where $\emptyset \neq Z \subset \mathbb{R}^{\bar{m}}$ is compact and convex, and each ψ_j is closed (upper semicontinuous) proper and concave with $\operatorname{dom} \psi_j \supset Z$. The Lagrangian of (5.1) has the form $\psi_0(z) + \langle x, \psi(z) \rangle$, where $\psi := (\psi_1, \ldots, \psi_n)$ and x is a multiplier. Suppose that, at each multiplier x in the *dual feasible* set $\breve{S} := \mathbb{R}^n_+$, the *dual function*

$$f(x) := \max\{\, \psi_0(z) + \langle x, \psi(z) \rangle : z \in Z \,\} \tag{5.2}$$

can be evaluated by finding a partial Lagrangian solution

$$z(x) \in Z(x) := \operatorname{Arg\,max}\{\, \psi_0(z) + \langle x, \psi(z) \rangle : z \in Z \,\}. \tag{5.3}$$

Thus f is finite convex and has a subgradient mapping $g_f(\cdot) := \psi(z(\cdot))$ on $\breve{S}$. In view of Rem. 2.2(vi), we suppose that $\psi(z(\cdot))$ 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). Assuming nonemptiness of the *dual optimal* set $\breve{S}_* := \operatorname{Arg\,min}_{\breve{S}} f$ (e.g., Slater's condition $\psi(\breve{z}) > 0$ for some $\breve{z} \in Z$), we consider the following two choices:

$$S := \breve{S} := \mathbb{R}^n_+ \quad \text{or} \quad S := \{x : 0 \leq x \leq x^{\text{up}}\} \text{ with } x^{\text{up}} > \bar{x} \text{ for some } \bar{x} \in \breve{S}_*. \tag{5.4}$$

For the second choice, $\min_S f$ is a restricted version of the classical dual problem $\min_{\breve{S}} f$ in the sense of Fact 4.3.

We shall employ the partial Lagrangian solutions and their constraint values

$$z^k := z(x^k) \quad \text{and} \quad g_f^k := \psi(z^k) \tag{5.5}$$

for generating and analyzing the following estimates of solutions to the primal problem (5.1). Using the weights $\{\nu_j^k\}_{j=k(l)}^k$ (cf. (3.4)), we define the kth *aggregate primal solution*

$$\breve{z}^k := \sum_{j=k(l)}^{k} \nu_j^k z^j. \tag{5.6}$$

This construction is related to the aggregate linearization $\tilde{f}_k := \sum_{j=k(l)}^{k} \nu_j^k f_j$ (cf. (3.5)). By expressing each f_j in terms of $\psi_0(z^j)$ and $\psi(z^j)$, below we derive bounds on $\psi_0(\tilde{z}^k)$ and $\psi(\tilde{z}^k)$ that are useful for both asymptotic analysis and stopping criteria.

Lemma 5.1. (i) *For each* k, $f_k(\cdot) = \psi_0(z^k) + \left\langle \cdot, \psi(z^k) \right\rangle$.

(ii) $\tilde{z}^k \in Z$, $\psi_0(\tilde{z}^k) \geq \tilde{f}_k(0) \geq f(x^{k(l)}) - \tilde{\epsilon}_k - \left\langle \nabla \tilde{f}_S^k, x^{k(l)} \right\rangle$, $\psi(\tilde{z}^k) \geq \nabla \tilde{f}_k$, *where* $\nabla \tilde{f}_k \geq \nabla \tilde{f}_S^k$ *if* $S = \mathbb{R}_+^n$.

Proof. (i) Use (cf. (1.2)) $f_k(\cdot) = f(x^k) + \left\langle g_f^k, \cdot - x^k \right\rangle$, (5.2), (5.3) and (5.5).

(ii) We have (cf. (3.4)) $\sum_{j=k(l)}^{k} \nu_j^k = 1$ with $\nu_j^k > 0$. Hence $\tilde{z}^k \in \text{co}\{z^j\}_{j=k(l)}^k \subset Z$, $\psi_0(\tilde{z}^k) \geq \sum_j \nu_j^k \psi_0(z^j)$, $\psi(\tilde{z}^k) \geq \sum_j \nu_j^k \psi(z^j)$ by convexity of Z and concavity of ψ_0, ψ. Next, using (3.5) and (i), we get

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

with $\nabla \tilde{f}_k = \sum_j \nu_j^k \psi(z^j)$. The above equality, $\tilde{f}_S^k := \tilde{f}_k + \tilde{\imath}_S^k$ (cf. (3.5)), $\tilde{\imath}_S^k(0) \leq \imath_S(0) = 0$ (cf. Lem. 3.1(ii) and (5.4)) and (3.16) imply

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

Finally, if $S = \mathbb{R}_+^n$ then (cf. Lem. 3.1(ii)) $\tilde{\imath}_S^k \leq \imath_S$ gives $\nabla \tilde{\imath}_S^k \leq 0$, and hence $\nabla \tilde{f}_k = \nabla \tilde{f}_S^k - \nabla \tilde{\imath}_S^k \geq \nabla \tilde{f}_S^k$. Combining the preceding relations gives the conclusion. □

Let Z_* denote the solution set of the primal problem (5.1). We now show in the setting of (4.1) that the aggregate primal solutions $\{\tilde{z}^k\}_{k \in K'}$, generated via (5.6), converge to Z_*.

Theorem 5.2. *Suppose* $\{x^{k(l)}\}_{l \in L}$ *has a cluster point* x^∞ *such that* $x^{k(l)} \xrightarrow{L'} x^\infty$ *with* $L' \subset L$, *and let* $K' := \{k(l+1) : l \in L'\}$. *Then:*

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

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

(iii) *Let* $\tilde{z}^\infty$ *be a cluster point of* $\{\tilde{z}^k\}_{k \in K'}$. *Then* $\tilde{z}^\infty \in Z_*$. *Further,* $\psi_0^{\max} = f(x^\infty)$ *and* $\tilde{z}^\infty \in Z(x^\infty)$ *(cf.* (5.3)*).*

(iv) $d_{Z_*}(\tilde{z}^k) \xrightarrow{K'} 0$, *and* $f(x^{k(l)}) \downarrow \psi_0^{\max}$ *as* $k \to \infty$.

Proof. (i) By Lem. 5.1(ii), $\{\tilde{z}^k\}$ lies in Z, which is compact by our assumption.

(ii) By Lem. 3.5, $f(x^{k(l)}) \downarrow f(x^\infty)$, $\tilde{\epsilon}_k + \left\langle \nabla \tilde{f}_S^k, x^{k(l)} \right\rangle \xrightarrow{K'} 0$. By Thm 4.1(i,ii), (5.4) and Rem. 4.4, $\{\nabla \tilde{f}_k\}_{k \in K'}$ is bounded and its cluster points lie in $\mathcal{G} \subset -\mathcal{N}_S(x^\infty) \subset \mathbb{R}_+^n$.

(iii) By (i), $\tilde{z}^\infty \in Z$. Using (ii) in Lem. 5.1(ii) gives $\psi_0(\tilde{z}^\infty) \geq f(x^\infty)$, $\psi(\tilde{z}^\infty) \geq 0$ by closedness of ψ_0, ψ. Since $\psi_0(\tilde{z}^\infty) \leq \psi_0^{\max} \leq f(x^\infty)$ by weak duality, $\tilde{z}^\infty$ must solve (5.1) and $\psi_0(\tilde{z}^\infty) = \psi_0^{\max} = f(x^\infty)$. Further, $\psi(\tilde{z}^\infty) \geq 0$ and $x^\infty \geq 0$ yield $\psi_0(\tilde{z}^\infty) + \langle x^\infty, \psi(\tilde{z}^\infty) \rangle \geq f(x^\infty)$, so $\tilde{z}^\infty \in Z(x^\infty)$ by (5.2)–(5.3), using $\tilde{z}^\infty \in Z$.

(iv) This follows from (i–iii) and the continuity of d_{Z_*}. □

Corollary 5.3. *If $\{x^{k(l)}\}$ is bounded (e.g., so is S_*), then $\{\check{z}^k\}_{k\in K}$ is bounded, all its cluster points lie in Z_*, $d_{Z_*}(\check{z}^k) \xrightarrow{K} 0$, and $f(x^{k(l)}) \downarrow \psi_0^{\max}$ as $k \to \infty$.*

Remarks 5.4. (i) Given an (absolute) accuracy tolerance $\epsilon > 0$, the method may stop if

$$\psi_0(\check{z}^k) \geq f(x^{k(l)}) - \epsilon \quad \text{and} \quad \psi_i(\check{z}^k) \geq -\epsilon, \quad i = 1\colon n.$$

Then $\psi_0(\check{z}^k) \geq \psi_0^{\max} - \epsilon$ from $f(x^{k(l)}) \geq \psi_0^{\max}$ (weak duality), so $\check{z}^k \in Z$ is an ϵ-solution of (5.1). This stopping criterion will be satisfied for some k if $S := \mathbb{R}^n_+$ and $|x^{k(l)}| \not\to \infty$, e.g., if $\check{S}_*$ is bounded (cf. Rem. 2.2(vi)), or $S := \{x : 0 \leq x \leq x^{\text{up}}\}$ with $x^{\text{up}} > \bar{x}$ for some $\bar{x} \in \check{S}_*$ (cf. Lem. 5.1(ii) and Thm 5.2(ii)).

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

$$\bar{x}_i \leq [f(x) - \psi_0(\check{z})]/\psi_i(\check{z}), \quad i = 1\colon n$$

(since $\psi_0(\check{z}) + \langle \bar{x}, \psi(\check{z})\rangle \leq f(\bar{x}) \leq f(x)$ by (5.2)). Such bounds may be used for choosing $x^{\text{up}} > \bar{x}$ in (5.4).

(iii) Our results may mitigate common critiques of subgradient optimization (see, e.g., [SeS86]), 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 (1.7), the results in [LPS99] and [ShC96] (where each ψ_j is affine and $\sum_k \nu_k^2 < \infty$ is replaced by the assumption that $x^k \to \bar{x} \in S_*$) correspond to $\{1, 2, \ldots\}$ replacing K in Cor. 5.3, with 1 replacing $k(l)$ in (5.6). Hence our estimates may be expected to converge faster, since information from early steps is explicitly discarded. Further, [ShC96] gives partial results only for deflected subgradient approaches, which are easily handled in our framework; cf. Rem. 6.4(ii).

In some applications [Bea93], using the current multiplier x^k one may find a primal feasible point $\underline{z}^k \in Z_\psi := \{z \in Z : \psi(z) \geq 0\}$; then $\psi_0(\underline{z}^k) \leq f_*$. Such lower bounds may be exploited in the following modification of Algorithm 2.1 (cf. [KLL99b, §6]). At Step 0, set $f_{\text{low}}^0 := -\infty$, $\underline{L} := \emptyset$. At Step 1, find $\underline{z}^k \in Z_\psi$, and set $f_{\text{low}}^k := \psi_0(\underline{z}^k)$ and $z_{\text{low}}^k := \underline{z}^k$ if $\psi_0(\underline{z}^k) > f_{\text{low}}^{k-1}$, $f_{\text{low}}^k := f_{\text{low}}^{k-1}$ and $z_{\text{low}}^k := z_{\text{low}}^{k-1}$ otherwise. At Step 2, stop if $f_{\text{rec}}^k = f_{\text{low}}^k$ (since then $x_{\text{rec}}^k \in S_*$). Step 3 is replaced by

Step 3'. (i) If $f_{\text{low}}^k \geq f_{\text{rec}}^k - \frac{3}{4}\delta_l$, set $k(l+1) := k$, $\underline{L} \leftarrow \underline{L} \cup \{l\}$, $\check{z}^k := z_{\text{low}}^k$, $\rho_k := 0$, $\delta_{l+1} := f_{\text{rec}}^k - f_{\text{low}}^k$, replace x^k by x_{rec}^k and g_f^k by $g_f(x_{\text{rec}}^k)$, increase l by 1 and go to Step 4.

(ii) If $f(x^k) \leq f_{\text{rec}}^{k(l)} - \frac{1}{2}\delta_l$, set $k(l+1) := k$, $\rho_k := 0$, $\delta_{l+1} := \delta_l$ and increase l by 1.

At Step 4, set $f_{\text{lev}}^k := \max\{f_{\text{rec}}^{k(l)} - \delta_l, f_{\text{low}}^k\}$. At Step 6, set $\delta_{l+1} := \min\{\frac{1}{2}\delta_l, f_{\text{rec}}^k - f_{\text{low}}^k\}$.

This *lower bounding scheme* is analyzed in the following remarks.

Remarks 5.5. (i) The current lower bound $f_{\text{low}}^k = \psi_0(z_{\text{low}}^k) = \max_{j=1}^k \psi_0(\underline{z}^j) \leq f_*$ is used for adjusting δ_l and f_{lev}^k. As shown in [KLL99b, §6], the convergence results of Rem. 2.2(vi) remain valid, with $\delta_l \downarrow 0$ as $l \to \infty$ due to our assumption $f_* > -\infty$.

(ii) If $f_{\text{low}}^k \geq f_{\text{rec}}^{k(l)} - \frac{3}{4}\delta_l$ at Step 3'(i) then $\delta_{l+1} := f_{\text{rec}}^k - f_{\text{low}}^k \leq \frac{3}{4}\delta_l$. Hence if $\underline{K} := \{k(l+1) : l \in \underline{L}\}$ is infinite, then $\psi_0(\check{z}^k) \xrightarrow{\underline{K}} f_*$ from $\delta_l \downarrow 0$, $\psi_0(\check{z}^k) = f_{\text{low}}^k \leq f_* \leq f_{\text{rec}}^k$. This yields the following result. *If* $\#\underline{K}=\infty$, *then* $\{\tilde{z}^k\}_{k\in\underline{K}}$ *is bounded, all its cluster points lie in* Z_* *and* $d_{Z_*}(\tilde{z}^k)\xrightarrow{\underline{K}}0$. Indeed, $\{\tilde{z}^k\}_{k\in\underline{K}}\subset Z_\psi$, where Z_ψ is compact (so is Z and ψ is closed), and if $\tilde{z}^k\xrightarrow{K'}\tilde{z}^\infty$ with $\underline{K}'\subset\underline{K}$, then $\tilde{z}^\infty\in Z_\psi$ and $\psi_0(\tilde{z}^\infty)\geq f_*$ (ψ_0 is closed) give $\tilde{z}^\infty\in Z_*$ by weak duality, whereas d_{Z_*} is continuous.

(iii) As before, let L index groups l that terminate at Step 6. Note that $f_{\rm rec}^{k(l)}-\delta_l\leq f_{\rm lev}^k<\min\{f_{\rm rec}^{k(l)},f(x^k)\}$, as required in (3.13). Hence, in view of (i,ii), we have $\#L=\infty$ if $\#\underline{L}<\infty$, in which case Lem. 3.5 remains valid. In effect, Thm 5.2 remains true, whereas in Cor. 5.3, L and K are replaced by $L\cup\underline{L}$ and $K\cup\underline{K}$, respectively.

We now comment briefly on possible extensions.

Remarks 5.6. (i) Consider the equality constrained version of the primal problem (5.1)

$$\psi_0^{\max}:=\max\ \psi_0(z)\quad\text{s.t.}\quad\psi(z):=Az-b=0,\ z\in Z,\tag{5.7}$$

where $A\in\mathbb{R}^{n\times\tilde{m}}$, $b\in\mathbb{R}^n$. Then $\breve{S}:=\mathbb{R}^n$ and $S:=\breve{S}$ or $S:=\{x:x^{\rm low}\leq x\leq x^{\rm up}\}$ with $x^{\rm low}<\bar{x}<x^{\rm up}$ for some $\bar{x}\in\breve{S}_*$ (cf. (5.4)). Clearly, Lem. 5.1 holds with $\psi(\tilde{z}^k)=\nabla\tilde{f}_k$ (where $\nabla\tilde{f}_k=\nabla\tilde{f}_S^k$ if $S=\mathbb{R}^n$), and Thm 5.2 holds with $\psi(\tilde{z}^k)=\nabla\tilde{f}_k\xrightarrow{K'}0$ in (ii) (use $\mathcal{N}_{\breve{S}}(x^\infty)=\{0\}$) and hence $\psi(\tilde{z}^\infty)=0$ in (iii).

(ii) Instead of assuming Z compact, suppose Z is closed and (cf. (5.3)) $z(\cdot)$ is locally bounded on Z. The preceding results are not affected, since Thm 5.2(i) follows from (5.5)–(5.6) and Lem. 3.5(ii). This observation can also be used in [FeK00, §5].

6 Accelerations

As shown in [KLL99b, §9], we may accelerate Algorithm 2.1 by replacing the subgradient linearization f_k with a more accurate model ϕ_k of f_S from the family Φ_μ^k defined below.

Definition 6.1. Given $\mu\in(0,1]$, let $\Phi_\mu^k:=\{\phi\in\Phi:d_{\mathcal{L}(\phi,f_{\rm lev}^k)}(x^k)\geq\mu d_{H_k}(x^k)\}$, where $\Phi:=\{\phi:\mathbb{R}^n\to(-\infty,\infty]:\phi\text{ is closed proper convex and }\phi\leq f_S\}$, $\mathcal{L}(\phi,\cdot):=\mathcal{L}_\phi(\cdot)$.

Examples 6.2. (i) If $\phi_k\in\Phi$ and $\phi_k\geq f_k$ then $\phi_k\in\Phi_1^k$ (cf. (1.4)).

(ii) Let $\hat{f}^k:=\max_{j\in J^k}f_j$, where $k\in J^k\subset\{1\colon k\}$. Then $\hat{f}^k\in\Phi_1^k$.

(iii) Note that $\phi_k\in\Phi$ if ϕ_k is the maximum of several accumulated linearizations $\{f_j\}_{j=1}^k$, or their convex combinations, possibly augmented with $\imath_S$ or its affine minorants.

Fixing $\mu\in(0,1]$, suppose at Step 4 of Algorithm 2.1 we choose $\phi_k\in\Phi_\mu^k$ and set

$$x^{k+1/2}:=x^k+t_k(P_{\mathcal{L}_k}x^k-x^k),\ \check{\rho}_k:=t_k(2-t_k)d_{\mathcal{L}_k}^2(x^k)\quad\text{with}\quad\mathcal{L}_k:=\mathcal{L}_{\phi_k}(f_{\rm lev}^k),\tag{6.1}$$

i.e., $\mathcal{L}_k$ replaces H_k. Since the convergence analysis of [KLL99b, §7] covers this extension, we only need to exhibit suitable expressions of $\tilde{f}_k$ and $\tilde{\imath}_S^k$ in terms of the linearizations of f and $\imath_S$ that contribute to ϕ_k. Our fairly complex technical developments hinge on a simple idea. Namely, we may *a posteriori* replace ϕ_k by its linearization $\bar{\phi}_k$ that uncovers the weights with which $x^{k+1/2}$ in (6.1) is influenced by the past linearizations of f and $\imath_S$.

Lemma 6.3. (i) *Suppose* ϕ_k *is polyhedral and* $\mathcal{L}_k \neq \emptyset$ *in* (6.1). *Let* $y^k := P_{\mathcal{L}_k} x^k$. *Then there exist a subgradient* $g_\phi^k \in \partial\phi_k(y^k)$ *and a stepsize* $\check{\nu}_k > 0$ *s.t.* $y^k - x^k = -\check{\nu}_k g_\phi^k$. *Let*

$$\bar{\phi}_k(\cdot) := \phi_k(y^k) + \left\langle g_\phi^k, \cdot - y^k \right\rangle, \quad \bar{\nu}_k := t_k[\bar{\phi}_k(x^k) - f_{\text{lev}}^k]/|g_\phi^k|^2, \quad \bar{H}_k := \mathcal{L}_{\bar{\phi}_k}(f_{\text{lev}}^k).$$

Then $y^k = P_{\bar{H}_k} x^k$, $d_{\mathcal{L}_k}(x^k) = d_{\bar{H}_k}(x^k) = [\bar{\phi}_k(x^k) - f_{\text{lev}}^k]/|g_\phi^k|$, $x^{k+1/2} - x^k = -\bar{\nu}_k g_\phi^k$, $\bar{\nu}_k = t_k \check{\nu}_k > 0$, $\bar{\phi}_k \leq \phi_k$. *Further, if* $\phi_k = \phi_f^k + \phi_S^k$ *and* $g_\phi^k = g_{\phi_f}^k + g_{\phi_S}^k$ *with* $\phi_f^k \leq f$, $\phi_S^k \leq \imath_S$, $g_{\phi_f}^k \in \partial\phi_f^k(y^k)$, $g_{\phi_S}^k \in \partial\phi_S^k(y^k)$, *then* $\bar{\phi}_k = \bar{\phi}_f^k + \bar{\phi}_S^k$ *with*

$$\bar{\phi}_f^k(\cdot) := \phi_f^k(y^k) + \left\langle g_{\phi_f}^k, \cdot - y^k \right\rangle \leq f(\cdot), \quad \bar{\phi}_S^k(\cdot) := \phi_S^k(y^k) + \left\langle g_{\phi_S}^k, \cdot - y^k \right\rangle \leq \imath_S(\cdot).$$

(ii) *Suppose* $\mathcal{L}_j \neq \emptyset$, $\bar{\nu}_j$ *and* $\phi_j = \phi_f^j + \phi_S^j$ *are used as in* (i) *for iterations* $j = k(l){:}\,k$. *Then Lem.* 3.1 *holds with* (3.4)–(3.8) *modified by*

$$\tilde{f}_k(\cdot) := \sum_{j=k(l)}^{k} (\bar{\nu}_j/\tilde{\nu}_f^k)\bar{\phi}_f^j(\cdot),\ \tilde{\imath}_S^k(\cdot) := \sum_{j=k(l)}^{k} [\left\langle g_S^j, \cdot - x^{j+1}\right\rangle + \bar{\nu}_j \bar{\phi}_S^j(\cdot)]/\tilde{\nu}_f^k,\ \tilde{f}_S^k(\cdot) := \tilde{f}_k(\cdot) + \tilde{\imath}_S^k(\cdot),$$

$$\tilde{\nu}_f^k := \sum_{j=k(l)}^{k} \bar{\nu}_j, \quad \tilde{f}_{\text{lev}}^k := \sum_{j=k(l)}^{k} (\bar{\nu}_j/\tilde{\nu}_f^k) f_{\text{lev}}^j \geq \min_{j=k(l)}^{k} f_{\text{lev}}^j,$$

$$x^{k+1} - x^{k(l)} = -\tilde{\nu}_f^k \nabla \tilde{f}_S^k, \quad L_k := -\tfrac{1}{2}|\tilde{\nu}_f^k \nabla \tilde{f}_S^k|^2 + \tilde{\nu}_f^k[\tilde{f}_S^k(x^{k(l)}) - \tilde{f}_{\text{lev}}^k] = \tfrac{1}{2}\rho_{k+1}.$$

(iii) *Suppose in* (ii), $\bar{\phi}_f^k = \sum_{j=k(l)}^{k} \hat{\nu}_j^k f_j$ *with* $\hat{\nu}_j^k \geq 0$, $\sum_j \hat{\nu}_j^k = 1$, *and if* $k > k(l)$ *then*

$$\tilde{f}_{k-1} = \sum_{j=k(l)}^{k-1} (\tilde{\nu}_j^{k-1}/\tilde{\nu}_f^{k-1}) f_j \quad \textit{with} \quad \tilde{\nu}_f^{k-1} = \sum_{j=k(l)}^{k-1} \tilde{\nu}_j^{k-1}, \quad \tilde{\nu}_j^{k-1} \geq 0.$$

Then

$$\tilde{f}_k = \sum_{j=k(l)}^{k} (\tilde{\nu}_j^k/\tilde{\nu}_f^k) f_j \quad \textit{and} \quad \tilde{\nu}_f^k = \sum_{j=k(l)}^{k} \tilde{\nu}_j^k,$$

where

$$\tilde{\nu}_k^k := \bar{\nu}_k \hat{\nu}_k^k, \quad \tilde{\nu}_j^k := \bar{\nu}_k \hat{\nu}_j^k + \tilde{\nu}_j^{k-1} \geq 0 \textit{ for } j = k(l){:}\,k-1 \textit{ if } k > k(l).$$

Hence (3.3)–(3.6) *and Lem.* 3.1 *hold with* ν_j, g_S^j, $\left\langle g_S^j, \cdot - x^{j+1}\right\rangle$ *replaced by* $\tilde{\nu}_j^k$, $g_S^j + \bar{\nu}_j g_{\phi_S}^j$, $\left\langle g_S^j, \cdot - x^{j+1}\right\rangle + \bar{\nu}_j \bar{\phi}_S^j(\cdot)$, *respectively, except that now* $\tilde{f}_{\text{lev}}^k := \sum_{j=k(l)}^{k} (\bar{\nu}_j/\tilde{\nu}_f^k) f_{\text{lev}}^j$.

Proof. (i) Use the KKT conditions for $y^k = \arg\min\{\frac{1}{2}|x - x^k|^2 : \phi_k(x) \leq f_{\text{lev}}^k\}$ and note that $y^k \neq x^k$ (since $d_{\mathcal{L}_k}(x^k) \geq \mu d_{H_k}(x^k) > 0$) and $\bar{\phi}_k \leq \phi_k$ by convexity.

(ii) Use (i) and the proof of Lem. 3.1, replacing f_j by $\bar{\phi}_j$.

(iii) Develop $\tilde{f}_k = (\bar{\nu}_k \bar{\phi}_f^k + \tilde{\nu}_f^{k-1} \tilde{f}_{k-1})/\tilde{\nu}_f^k = (\bar{\nu}_k \sum_{j=k(l)}^{k} \hat{\nu}_j^k f_j + \sum_{j=k(l)}^{k-1} \tilde{\nu}_j^{k-1} f_j)/\tilde{\nu}_f^k$. □

The general constructions of Lem. 6.3 may be specialized as follows.

Remarks 6.4. (i) If $\phi_k = \max_{j\in J^k} f_j$, where $J^k \subset \{k(l): k\}$, then in Lem. 6.3(iii) we may use $\hat{\nu}_j^k \geq 0$ s.t. $g_\phi^k = \sum_j \hat{\nu}_j^k g_f^j$, $\sum_j \hat{\nu}_j^k = 1$, $\hat{\nu}_j^k[\phi_k(y^k) - f_j(y^k)] = 0$. In fact, such $\hat{\nu}_j^k$ are scaled Lagrange multipliers of the QP subproblem $\min\{\frac{1}{2}|x - x^k|^2 : f_j(x) \leq f_{\rm lev}^k, j \in J^k\}$.

(ii) If $\bar{\phi}_f^k = (1-\alpha_k)f_k + \alpha_k\bar{\phi}_f^{k-1}$ with $\alpha_k \in [0,1]$ and $\bar{\phi}_f^{k-1} = \sum_{j=k(l)}^{k-1} \hat{\nu}_j^{k-1} f_j$, where $\hat{\nu}_j^{k-1} \geq 0$ and $\sum_j \hat{\nu}_j^{k-1} = 1$, then in Lem. 6.3(iii) $\bar{\phi}_f^k = \sum_{j=k(l)}^k \hat{\nu}_j^k f_j$ with $\hat{\nu}_k^k := 1 - \alpha_k$, and $\hat{\nu}_j^k := \alpha_k\hat{\nu}_j^{k-1}$ for $j < k$. The examples of [KLL99b, Rem. 7.6] use $\alpha_k = 0$ if $k = k(l)$.

(iii) In view of Lem. 6.3(iii), the weights $\tilde{\nu}_j^k/\tilde{\nu}_f^k$ may replace ν_j^k in (5.6), so that

$$\tilde{z}^k := \sum_{j=k(l)}^k (\tilde{\nu}_j^k/\tilde{\nu}_f^k)z^j = (\bar{\nu}_k/\tilde{\nu}_f^k)z_\phi^k + (1 - \bar{\nu}_k/\tilde{\nu}_f^k)\tilde{z}^{k-1} \quad \text{with} \quad z_\phi^k := \sum_{j=k(l)}^k \hat{\nu}_j^k z^j, \tag{6.2}$$

where the multiplier $\bar{\nu}_k$ is given in Lem. 6.3(i); note that in (ii) above we may update

$$z_\phi^k = (1-\alpha_k)z^k + \alpha_k z_\phi^{k-1} \quad (z_\phi^0 := z^1). \tag{6.3}$$

Indeed,

$$\tilde{z}^k = (\bar{\nu}_k\hat{\nu}_k^k/\tilde{\nu}_f^k)z^k + \sum_{j=k(l)}^{k-1} (\bar{\nu}_k\hat{\nu}_j^k + \tilde{\nu}_j^{k-1})z^j/\tilde{\nu}_f^k = (\bar{\nu}_k/\tilde{\nu}_f^k)\sum_{j=k(l)}^k \hat{\nu}_j^k z^j + (\tilde{\nu}_f^{k-1}/\tilde{\nu}_f^k)\tilde{z}^{k-1}$$

with $\tilde{\nu}_f^k = \bar{\nu}_k + \tilde{\nu}_f^{k-1}$, and $z_\phi^k = (1-\alpha_k)z^k + \alpha_k\sum_{j=k(l)}^{k-1}\hat{\nu}_j^{k-1}z^j$ in (ii).

(iv) If we allow $\alpha_{k(l)} \neq 0$, then 1 replaces $k(l)$ in (ii,iii). Then for the proof of Thm 4.1(i), one may assume that g_f is bounded on S (e.g., S is bounded, $\psi(z(\cdot))$ is bounded on S or ψ is continuous in §5). Similar modifications may handle $J^k \not\subset \{k(l): k\}$ in (i).

(v) If ϕ_k is polyhedral and $\mathcal{L}_k = \emptyset$, then $\min\phi_k > f_{\rm lev}^k$ and $g_\phi^k := 0 \in \partial\phi_k(y^k)$ for any $y^k \in \operatorname{Arg\,min}\phi_k$. Hence if $\bar{\phi}_k$, ϕ_f^k, ϕ_S^k, $\bar{\phi}_f^k$ and $\bar{\phi}_S^k$ are as in as in Lem. 6.3(i), then $\tilde{f}_k := \bar{\phi}_f^k \leq f$, $\tilde{\imath}_S^k := \bar{\phi}_S^k \leq \imath_S$, $\tilde{f}_S^k := \tilde{f}_k + \tilde{\imath}_S^k = \bar{\phi}_k \leq f_S$, with $\nabla\tilde{f}_S^k = 0$, $\tilde{f}_S^k(\cdot) = \phi_k(y^k) > f_{\rm lev}^k$. Letting $\tilde{f}_{\rm lev}^k := f_{\rm lev}^k$ in (3.6), we have $\tilde{H}_k = \emptyset$, $d_{\tilde{H}_k}(x^{k(l)}) = \infty$. Clearly, Lem. 3.3 remains valid and the tests (2.1), (3.12) and (3.23) hold with $x^{k+1} = x^{k+1/2} = x^k$ and $\rho_{k+1} = \rho_{k+1/2} = \infty$, since $\mathcal{L}_k = \emptyset$ in (6.1). Hence if $\bar{\phi}_f^k = \sum_{j=k(l)}^k \hat{\nu}_j^k f_j$ as in Lem. 6.3(iii) (and (i,ii,iii) above), then $\hat{\nu}_j^k$ replaces ν_j^k in (5.6), i.e., $\tilde{z}^k = z_\phi^k$ as if $\bar{\nu}_k = \tilde{\nu}_f^k$ in (6.2).

(vi) The simplest ϕ_k employing some constraint information is $\phi_k = f_k + \langle \hat{a}^k, \cdot - x^k\rangle$ with $\hat{a}^k \in \mathcal{N}_S(x^k)$ (cf. [KLL99b, Rem. 7.8]); e.g., the "optimal" $\hat{a}^k = P_{\mathcal{N}_S(x^k)}(-g_f^k)$ maximizes $d_{\mathcal{L}_k}(x^k)$ [Kiw96, §7]. Then $\tilde{\imath}_S^k(\cdot) = \sum_{j=k(l)}^k[\bar{\nu}_j\langle\hat{a}^j, \cdot - x^j\rangle + \langle g_S^j, \cdot - x^{j+1}\rangle]/\tilde{\nu}_f^k$ if $\mathcal{L}_k \neq \emptyset$ (cf. Lem. 6.3(ii)), $\tilde{\imath}_S^k(\cdot) = \langle \hat{a}^k, \cdot - x^k\rangle$ otherwise by (v).

(vii) Suppose $\phi_k = f_k + \imath_S$ for all k. Then $g_\phi^k = g_f^k + \hat{a}^k$, $\hat{a}^k \in \mathcal{N}_S(y^k)$ in Lem. 6.3(i), so we may replace ϕ_k by $\bar{\phi}_k$ and x^k by y^k in (vi). Similar arguments apply to $\phi_k = \check{f}^k + \imath_S$ or $\phi_k = \max_{j\in J^k} f_j + \langle P_{\mathcal{N}_S(x^k)}(-g_f^j), \cdot - x^k\rangle$ [KLL99b, Rem. 7.8], since (cf. (i)) suitable $\tilde{f}_k$ and $\tilde{\imath}_S^k$ may be defined as in Lems. 6.3(ii,iii) and (v) above.

Example 6.5. For simple bounds $S = \{x : x^{\rm low} \leq x \leq x^{\rm up}\}$, our preliminary implementation employs

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

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

where $\check{g}^k := g_f^k + P_{\mathcal{N}_S(x^k)}(-g_f^k)$ and $\check{g}_\phi^{k-1} := g_{\phi_f}^{k-1} + P_{\mathcal{N}_S(x^k)}(-g_{\phi_f}^{k-1})$ are *reduced subgradients* (cf. Rem. 6.4(vi)), updating

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

Our choices of α_k [KLL99b, Ex. 7.4(v) and Rem. 7.6] include:

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

(ii) the *conjugate subgradient strategy* (CSS): $\alpha_k := \frac{\langle \check{g}^k, \check{g}_\phi^{k-1} \rangle}{\langle \check{g}^k, \check{g}_\phi^{k-1} \rangle - |\check{g}_\phi^{k-1}|^2}$ if $\langle \check{g}^k, \check{g}_\phi^{k-1} \rangle < 0$ and $\check{\phi}_{k-1}(x^k) \geq f_{\text{lev}}^k$, $\alpha_k := 0$ otherwise (cf. [KLL99b, Rem. 7.6]);

(iii) the *average direction strategy* (ADS): $\alpha_k := \frac{|\check{g}^k|}{|\check{g}^k| + |\check{g}_\phi^{k-1}|}$ if $\check{g}_\phi^{k-1} \neq 0$ and $\check{\phi}_{k-1}(x^k) \geq f_{\text{lev}}^k$, $\alpha_k := 0$ otherwise (cf. [KLL99b, Rem. 7.6]);

(iv) the *aggregate subgradient strategy* (ASS): α_k is s.t. $P_{\mathcal{L}_k} x^k = P_{\mathcal{L}(\max\{\check{f}_k, \check{\phi}_{k-1}\}, f_{\text{lev}}^k)} x^k$ if $\mathcal{L}_{\max\{\check{f}_k, \check{\phi}_{k-1}\}}(f_{\text{lev}}^k) \neq \emptyset$, α_k is s.t. $\mathcal{L}_k = \emptyset$ otherwise (cf. [Kiw96, Rem. 4.1]).

For OSS and ASS, if $\max\{f_k(x^{k+1}), \bar{\phi}_f^k(x^{k+1})\} > f_{\text{rec}}^{k(l)} - \frac{3}{4}\delta_l$ at Step 6, 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 [KLL99b, Rem. 7.11] (but not for CSS and ADS). They provide an inexact implementation of the "best" single projection $P_{\mathcal{L}(\max\{f_k, \bar{\phi}_f^{k-1}\}, f_{\text{lev}}^k) \cap S} x^k$, which would require more sophisticated QP.

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., [Ber98, Chap. 8] for a textbook introduction, [OMV00] for a recent survey, [Fuk84a, Fuk84b] for the pioneering dual developments, and [GGSV96, GSV97, LLP97, LPS99] 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 [FeK00, §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, with $s_{io_i} = r_i$, $s_{id_i} = -r_i$, $s_{il} = 0$ if $l \neq o_i, d_i$. Our *multicommodity flow* problem is stated as follows:

$$\min \quad \check{\psi}_0(z_0) := \sum_{j=1}^{n} \check{\psi}_{0j}(z_{0j}) \tag{7.1a}$$

$$\text{s.t.} \quad \psi_j(z) := z_{0j} - \sum_{i=1}^{m} z_{ij} = 0, \ j = 1\colon n, \tag{7.1b}$$

$$z := (z_0, z_1, \ldots, z_m) \in Z := Z_0 \times Z_1 \times \cdots \times Z_m, \tag{7.1c}$$

$$Z_0 := \mathbb{R}^n, \ Z_i := \{\, z_i : E z_i = s_i, 0 \le z_i \le \bar{z}_i \,\}, \ i = 1\colon m, \tag{7.1d}$$

where z_i is the *flow of commodity* $i \in \{1\colon m\}$, $z_0 := \sum_{i=1}^m z_i$ is the *total flow*, and $\bar{z}_i$ are fixed positive vectors of *flow bounds*. 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 $[0, \kappa_j]$, where either $\kappa_j > 0$ is finite 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 [Ber98, §9.1], [Roc70, pp. 227–230].) Finally, we suppose that

$$\check{z}_0 \in \textstyle\prod_{j=1}^{n} [0, \kappa_j) \quad \text{for some } \check{z} \in Z \text{ with } \psi(\check{z}) = 0. \tag{7.2}$$

7.2 Dual approach

In the framework of Rem. 5.6(i), letting $\psi_0(z) := -\breve{\psi}_0(z_0)$ and $\breve{S} := \mathbb{R}^n$, we have $f(x) = \sum_{i=0}^m f^i(x)$ and $z(x) = (z_0(x), \ldots, z_m(x))$ in (5.2)–(5.3) with $f^0(x) = \sum_{j=1}^n f_j^0(x_j)$ and

$$f_j^0(x_j) := \max_t \{\, x_j t - \breve{\psi}_{0j}(t) \,\} =: \breve{\psi}_{0j}^*(x_j), \quad j = 1\colon n, \tag{7.3a}$$

$$z_{0j}(x) := \arg\min_t \{\, \breve{\psi}_{0j}(t) - x_j t \,\} = \nabla \breve{\psi}_{0j}^*(x_j) = \nabla f_j^0(x_j), \quad j = 1\colon n, \tag{7.3b}$$

$$f^i(x) := \max\{\, -\langle x, z_i \rangle : E z_i = s_i, 0 \le z_i \le \bar{z}_i \,\}, \quad i = 1\colon m, \tag{7.4a}$$

$$z_i(x) \in \operatorname{Arg\,min}\{\, \langle x, z_i \rangle : E z_i = s_i, 0 \le z_i \le \bar{z}_i \,\} = -\partial f^i(x), \quad i = 1\colon m, \tag{7.4b}$$

where $\breve{\psi}_{0j}^*$, the conjugate of $\breve{\psi}_{0j}$, is continuously differentiable due to the strict convexity of $\breve{\psi}_{0j}$. Thus $z(\cdot)$ and $g_f(\cdot) := \psi(z(\cdot))$ are locally bounded.

The set $\breve{S}_* := \operatorname{Arg\,min} f$ of Lagrange multipliers of problem (7.1) is nonempty. Indeed, since each $\breve{\psi}_{0j}$ is increasing on its domain, problem (7.1) is equivalent to the following:

$$\breve{\psi}_0^{\min} := \min \ \breve{\psi}_0(z_0) \quad \text{s.t.} \quad \psi(z) \ge 0, \ z \in Z, \tag{7.5}$$

whereas (7.2) is equivalent to Slater's condition for (7.5) ($\psi(z) > 0$ for some $z \in Z$ with $z_0 \in \operatorname{dom} \breve{\psi}_0$), so (cf. [Roc70, Cors. 28.4.1 and 29.1.5]) the set $\tilde{S}_* := \operatorname{Arg\,min}_{\mathbb{R}^n_+} f$ of Lagrange multipliers of problem (7.5) is nonempty and bounded; clearly, $\tilde{S}_* \subset \breve{S}_*$.

Following [Fuk84a], we now consider using the *restricted dual feasible* set

$$S := \{\, x : x \ge x^{\text{low}} \,\} \quad \text{with} \quad x_j^{\text{low}} := \breve{\psi}'_{0j}(0) \text{ for } j = 1\colon n. \tag{7.6}$$

With $S_* := \operatorname{Arg\,min}_S f$, we have $\emptyset \ne S_* \subset \tilde{S}_*$. Indeed, $x_j^{\text{low}} \ge 0$, since each $\breve{\psi}_{0j}$ is nondecreasing on its domain, whereas for $x_j \le x_j^{\text{low}}$, $f^0(x)$ is constant and each $f^i(x)$, $i = 1\colon m$, is nonincreasing (cf. (7.3b), (7.4a)). Thus S_* is bounded (so is $\tilde{S}_*$).

Further, the conclusions of Fact 4.3 hold. Specifically, $\partial f(x) \cap -\mathcal{N}_{\breve{S}}(x) = \{0\}$ for each $x \in \breve{S}_*$, and although $\breve{S}_* \cap \operatorname{int} S = \emptyset$ is possible, we still have $\mathcal{G} := \partial f(x) \cap -\mathcal{N}_S(x) = \{0\}$

for every $x \in S_*$. Indeed, let $x \in S_*$, $g \in \mathcal{G}$. If $x_j > x_j^{\rm low}$ then $g_j = 0$ from $-g \in \mathcal{N}_S(x)$, otherwise $x_j = x_j^{\rm low}$ gives $0 = z_{0j}(x) = [\nabla f^0(x)]_j$ by (7.3b), so $\partial f(x) = \sum_{i=0}^m \partial f^i(x)$ with $\partial f^i(x) \subset -\mathbb{R}^n_+$ by (7.4b) yield $g_j \le 0$, and hence $g_j = 0$ from $-g \in \mathcal{N}_S(x)$.

Suppose $\bar z_{ij} \ge r_i$ for all i and j. Then for $x \in S$, ignoring $\bar z_i$ in (7.4b), we may find $z_i(x)$ by solving a *shortest path* problem with nonnegative arc lengths, since $z_{ij}(x) \le r_i$ for all j. Thus, if the bounds $\bar z_i$ are omitted in (7.1d) and (7.4b), as happens in many applications, then $z_i(\cdot)$ are still bounded on S (although ∂f^i may be unbounded).

Since $\breve\psi_0$ is strictly convex, the *primal solution* set of (7.1) has the form

$$Z_* = \{z_0^*\} \times Z_*^f \quad \text{with} \quad Z_*^f := \left\{ (z_1, \ldots, z_m) \in Z_1 \times \cdots \times Z_m : z_0^* = \sum_{i=1}^m z_i \right\}, \tag{7.7}$$

where z_0^* is the unique optimal total flow. On the other hand, if each $\breve\psi_{0j}$ is differentiable on $(0, \infty)$, then (cf. (7.3b)) $\nabla f_j^0(x_j) = \nabla \breve\psi_{0j}^*(x_j)$ is increasing for $x_j > \breve\psi'_{0j}(0)$ (since $\nabla\breve\psi_{0j}^* = (\nabla\breve\psi_{0j})^{-1}$); in effect, f^0 and f are strictly convex on S, and S_* is singleton.

7.3 Algorithmic constructions and convergence

The convergence results of §5 may be specialized as follows. Although the kth aggregate $\tilde z^k \in Z$ (cf. Lem. 5.1(ii)) need not be feasible in the primal problem (7.1), we may use the kth *aggregate total flow*

$$\check z_0^k := \sum_{i=1}^m \tilde z_i^k \tag{7.8}$$

to produce the *primal feasible aggregate* $\check z^k := (\check z_0^k, \tilde z_1^k, \ldots, \tilde z_m^k) \in Z$ with $\psi(\check z^k) = 0$. Then (cf. (5.1), (7.5)) we have $-\breve\psi_0^{\rm min} = \psi_0^{\rm max} \le f(x_{\rm rec}^k)$ (weak duality), and hence

$$\breve\psi_0(\check z_0^k) - \breve\psi_0^{\rm min} \le \breve\psi_0(\check z_0^k) + f(x_{\rm rec}^k). \tag{7.9}$$

Proposition 7.1. (i) *$\{x^{k(l)}\}$, $\{g_f^k\}$, $\{z^k\}$, $\{\tilde z^k\}$ and $\{\check z^k\}$ are bounded, and all the cluster points of $\{\tilde z^k\}$ and $\{\check z^k\}$ lie in Z.*

(ii) *$f(x^{k(l)}) \downarrow f_*$, $\tilde\epsilon_k + \langle \nabla \tilde f_S^k, x^{k(l)} \rangle \xrightarrow{K} 0$, and $\tilde z_0^k - \check z_0^k = \psi(\tilde z^k) = \nabla \tilde f_k \xrightarrow{K} 0$.*

(iii) *All the cluster points of $\{\tilde z^k\}_{k\in K}$ and $\{\check z^k\}_{k \in K}$ lie in Z_*. Further, $\psi_0^{\rm max} = f_*$.*

(iv) *$\tilde z_0^k \xrightarrow{K} z_0^*$, $\check z_0^k \xrightarrow{K} z_0^*$, $d_{Z_*^f}((\tilde z_1^k, \ldots, \tilde z_m^k)) \xrightarrow{K} 0$, and $f(x^{k(l)}) \downarrow -\breve\psi_0^{\rm min} = -\breve\psi_0(z_0^*)$.*

(v) *If $z_0^* \in \prod_{j=1}^n [0, \kappa_j)$, then $\breve\psi_0(\check z_0^k) \xrightarrow{K} \breve\psi_0^{\rm min}$ and $\breve\psi_0(\check z_0^k) + f(x_{\rm rec}^k) \xrightarrow{K} 0$.*

Proof. (i) Since S_* is bounded, so is $\{x^{k(l)}\}$ (cf. Rem. 2.2(vi)). But $|x^k - x^{k(l)}| \le 2R$ at Step 1 (Rem. 2.2(v)), whereas $z(\cdot)$ and $g_f(\cdot)$ are locally bounded, so the conclusion follows from the definitions (5.5)–(5.6) and (7.8), since $\tilde z^k \in \mathrm{co}\{z^j\}_{j \le k} \subset Z$ and Z is closed.

(ii) By Lem. 3.5, $f(x^{k(l)}) \downarrow f_*$, $\tilde\epsilon_k + \langle \nabla \tilde f_S^k, x^{k(l)} \rangle \xrightarrow{K} 0$. As discussed in §7.2, the conclusions of Fact 4.3 hold with $\mathcal{G} = \{0\}$, so $\nabla \tilde f_k \xrightarrow{K} 0$ by Cor. 4.2, where $\nabla \tilde f_k = \psi(\tilde z^k) = \tilde z_0^k - \check z_0^k$ by Rem. 5.6(i), (7.1b) and (7.8).

(iii) Argue as for Thm 5.2(iii), with $\psi(\tilde z^\infty) = 0$ and $|\tilde z^k - \check z^k| = |\tilde z_0^k - \check z_0^k| \xrightarrow{K} 0$ by (ii).

(iv) By (i–iii) and the continuity of d_{Z_*}, $f(x^{k(l)}) \downarrow \psi_0^{\rm max} = -\breve\psi_0^{\rm min}$ and $d_{Z_*}(\check z^k) \xrightarrow{K} 0$. Hence the conclusion follows from the form (7.7) of Z_*, since $\tilde z_0^k - \check z_0^k \xrightarrow{K} 0$ by (ii).

(v) This follows from (iv), since $\breve\psi_0$ is continuous on $\prod_{j=1}^n [0, \kappa_j)$. □

7.4 Extension to linear costs

Retaining the remaining assumptions, suppose for a fixed $\breve{n} < n$ and all $j > \breve{n}$, each cost $\breve{\psi}_{0j}$ is *linear*: $\breve{\psi}_{0j}(t) := \breve{\psi}'_{0j}(0)t$ if $t \geq 0$, ∞ otherwise, with $\breve{\psi}'_{0j}(0) > 0$. Then (cf. (7.3), (7.6)) $f_j^0(x_j) = 0$ and $z_{0j}(x) = 0$ if $x_j < x_j^{\text{low}}$, $f_j^0(x_j) = \infty$ and $z_{0j}(x)$ is undefined if $x_j > x_j^{\text{low}}$, but for $x_j = x_j^{\text{low}}$, $f_j^0(x_j) = 0$ and $z_{0j}(x)$ could be arbitrary in $\mathbb{R}_+$. Exploiting this freedom, we may restrict attention to the following subset of S (cf. (7.6)):

$$\hat{S} := \{\, x : x_j \geq x_j^{\text{low}} \text{ for } j \leq \breve{n},\ x_j = x_j^{\text{low}} \text{ for } j > \breve{n} \,\}, \tag{7.10}$$

letting

$$z_{0j}(x) := \sum_{i=1}^{m} z_{ij}(x) \quad \text{if} \quad x \in \hat{S},\ j > \breve{n}. \tag{7.11}$$

Thus $g_{fj}(x) := \psi_j(x) = 0$ if $x \in \hat{S}$, $j > \breve{n}$. Hence, for an initial point $x^1 \in \hat{S}$, by induction we always have $x^k \in \hat{S}$, $g_{fj}^k := \psi_j(x^k) = 0$ and hence (cf. (5.6), (7.8)) $\psi_j(\tilde{z}^k) = 0$ and $\check{z}_{0j}^k = \tilde{z}_{0j}^k$ for $j > \breve{n}$. In other words, for arcs with linear costs, the multipliers are fixed at their optimal values, and the aggregate flows are primal feasible. Clearly, $z(\cdot)$ and $g_f(\cdot) := \psi(z(\cdot))$ are locally bounded on $\hat{S}$ (so are $z_i(\cdot)$ and $z_{0j}(\cdot)$ for $j \leq \breve{n}$ as before, and then by (7.11) also $z_{0j}(\cdot)$ for $j > \breve{n}$).

The above observations suffice for proving parts (i,ii,iii) of Prop. 7.1. In part (iv), since now $Z_* = \{(z_{01}^*, \ldots, z_{0\breve{n}}^*)\} \times \hat{Z}_*$ with $\hat{Z}_*$ polyhedral, we have $\tilde{z}_{0j}^k \xrightarrow{K} z_{0j}^*$, $\check{z}_{0j}^k \xrightarrow{K} z_{0j}^*$, $j \leq \breve{n}$, $d_{\hat{Z}_*}((\check{z}_{0,\breve{n}+1}^k, \ldots, \check{z}_{0n}^k, \check{z}_1^k, \ldots, \check{z}_m^k)) \xrightarrow{K} 0$. In part (v), it can be shown that $\psi_0(\tilde{z}^k) \xrightarrow{K} \psi_0^{\max}$ (hint: if $\tilde{z}^k \xrightarrow{K'} \tilde{z}^\infty$ then $\psi_0(\tilde{z}^\infty) \geq \overline{\lim}_{k\in K'} \psi_0(\tilde{z}^k) \geq \underline{\lim}_{k\in K'} \psi_0(\tilde{z}^k) \geq f(x^\infty)$ in the proof of Thm 5.2(iii)); thus $\breve{\psi}_0^{\min} \xleftarrow{K} \breve{\psi}_0(\check{z}_0^k) = \psi_0(\tilde{z}_0^k) + \sum_{j\leq\breve{n}}[\breve{\psi}_{0j}(\check{z}_{0j}^k) - \breve{\psi}_{0j}(\tilde{z}_{0j}^k)]$ with $\breve{\psi}_{0j}(\tilde{z}_{0j}^k), \breve{\psi}_{0j}(\check{z}_{0j}^k) \xrightarrow{K'} \breve{\psi}_{0j}(z_{0j}^*)$, since $\tilde{z}_{0j}^k, \check{z}_{0j}^k \xrightarrow{K} z_{0j}^*$ and $\breve{\psi}_{0j}$ are continuous on $[0, \kappa_j)$ for $j \leq \breve{n}$, so $\breve{\psi}_0(\check{z}_0^k) \xrightarrow{K} \breve{\psi}_0^{\min}$, as desired.

7.5 Numerical results

Our method was programmed in Fortran 77 and run on a notebook PC (Pentium II 400 MHz, 256 MB RAM). We used $\beta = \frac{1}{2}$, $\delta_1 = \frac{1}{2}\delta_0$ and $R_l := R(\delta_l/\delta_0)^\beta$ with $\delta_0 = R|\breve{g}^1|$ for consistency with [KLL99b, §8], $t_k \equiv 1$, the third projection of §3.5 and the aggregate subgradient strategy of Ex. 6.5, updating the total flows (cf. (6.2), (6.3), (7.8))

$$\check{z}_0^k = (\bar{\nu}_k/\tilde{\nu}_f^k)\check{z}_{\phi 0}^k + (1 - \bar{\nu}_k/\tilde{\nu}_f^k)\check{z}_0^{k-1} \quad \text{with} \quad \check{z}_{\phi 0}^k := \sum_{i=1}^{m} z_{\phi i}^k = (1-\alpha_k)\sum_{i=1}^{m} z_i^k + \alpha_k \check{z}_{\phi 0}^{k-1},$$

where $\check{z}_{\phi 0}^0 := \check{z}_0^0 := \sum_{i=1}^m z_i^1$. Letting $\check{z}_{\text{rec}}^1 := \check{z}^1$, every tenth iteration or when l increased at Steps 3 or 6, we set $\check{z}_{\text{rec}}^k := \check{z}^k$ if $\breve{\psi}_0(\check{z}_0^k) < \breve{\psi}_0(\check{z}_{\text{rec}0}^k)$, $\check{z}_{\text{rec}}^k := \check{z}_{\text{rec}}^{k-1}$ otherwise. Our stopping criterion $\breve{\psi}_0(\check{z}_{\text{rec}0}^k) + f(x_{\text{rec}}^k) \leq \epsilon_{\text{opt}}[1 + |\breve{\psi}_0(\check{z}_{\text{rec}0}^k)|]$ (cf. (7.9)) ensured a relative objective accuracy of $100\epsilon_{\text{opt}}$ in %. We used $\epsilon_{\text{opt}} = 10^{-i/2}$ for $i = 4, 5, 6$.

We first give results for the CNET collection of [OMV00], 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

Table 7.1: Results for the CNET instances, with $R = 10$.

m	Load	$\epsilon_{\text{opt}} = 10^{-2}$ Delay	k	Time	$\epsilon_{\text{opt}} = 10^{-2.5}$ Delay	k	Time	$\epsilon_{\text{opt}} = 10^{-3}$ Delay	k	Time	Optimal Delay
4452	1.0	12.6131	110	1	12.5881	180	2	12.5856	600	6	12.5847
	1.5	19.1949	150	2	19.1832	310	3	19.1814	620	6	19.1799
	2.0	25.9926	190	2	25.9833	300	3	25.9788	460	4	25.9755
	2.5	33.0326	200	2	33.0012	336	3	32.9835	1300	11	32.9809
	3.0	40.2486	220	2	40.2177	490	4	40.2125	1450	12	40.2072
6678	1.0	19.6691	170	2	19.6513	350	4	19.6494	490	5	19.6481
	1.5	30.2017	240	3	30.1828	460	5	30.1795	983	9	30.1776
	2.0	41.2699	160	2	41.2265	354	3	41.2095	697	6	41.2066
	2.5	52.8099	270	3	52.7880	726	7	52.7830	1390	12	52.7790
	3.0	64.9984	386	4	64.9754	550	5	64.9520	1539	14	64.9460
8904	1.0	26.4872	230	3	26.4872	238	3	26.4745	1080	11	26.4730
	1.5	41.0065	230	3	40.9884	390	5	40.9785	550	6	40.9742
	2.0	56.4728	460	4	56.4316	658	6	56.4295	779	7	56.4233
	2.5	73.0581	380	4	72.9577	560	6	72.9456	864	8	72.9392
	3.0	90.8231	416	4	90.7027	630	6	90.6700	1260	12	90.6620
11130	1.0	33.5348	190	2	33.4984	410	5	33.4952	680	7	33.4931
	1.5	52.4137	200	2	52.2741	640	7	52.2721	766	8	52.2677
	2.0	72.6954	470	5	72.6571	710	7	72.6474	1090	11	72.6434
	2.5	95.0557	325	4	94.9119	690	8	94.8916	1470	15	94.8838
	3.0	119.406	1240	13	119.353	1340	14	119.313	2280	23	119.306

that scale up the standard requirements r_i. The costs are Kleinrock's *average delays* $\breve{\psi}_{0j}(z_{0j}) := z_{0j}/(\kappa_j - z_{0j})$ on $[0, \kappa_j)$. We used $x_j^1 := \kappa_j^{-1}(1-\rho_*)^{-2}$, with $\rho_* = \frac{1}{4}$ estimating the maximum traffic intensity $\max_j z_{0j}^*/\kappa_j$ [Gof87] (which sometimes exceeded $\frac{1}{2}$). In Table 7.1, Delay $:= \breve{\psi}_0(\breve{z}_{\text{rec}0}^k)$, 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 the stopping criterion; e.g., for $\epsilon_{\text{opt}} = 10^{-3}$, $[\breve{\psi}_0(\breve{z}_{\text{rec}0}^k) - \breve{\psi}_0^{\min}]/\breve{\psi}_0^{\min} < 10^{-4}$ for the unit load instances. Since each instance had 106 common sources, most work per iteration went into solving 106 shortest path subproblems via subroutine L2QUE of [GaP88]. Our machine is about three times faster than the one employed in [OMV00]. Hence Table 7.1 suggests that our method is highly competitive with all the methods tested in [OMV00, 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 7.2. These problems have nonlinear *BTR delays* $\breve{\psi}_{0j}(z_{0j}) := \alpha_j z_{0j} + \beta_j z_{0j}^{\gamma_j}$ on $\mathbb{R}_+$ with $\alpha_j \geq 0$, $\beta_j > 0$, $\gamma_j > 1$, as well as linear costs $\breve{\psi}_{0j}(z_{0j}) := \alpha_j z_{0j}$ with $\alpha_j > 0$. The first three medium-sized problems were used in [LaP92, LPR97] ([HLV87] solved a slightly different version of Winnipeg). The Chicago problem [TEB98] is much bigger than the largest (random) problems considered in [GSV97, OMV00]. The Skåne problem (not reported so far) is really huge. We used $x^1 = x^{\text{low}}$ and $R = 100$, except $R = 10^4$ for Linköping. Concerning Table 7.3, we add that again for $\epsilon_{\text{opt}} = 10^{-3}$ the final accuracy was quite high: 1.5e-4 for Barcelona, 2.8e-4 for Linköping, 4.5e-4 for Winnipeg, 1.8e-4 for Chicago, 5.1e-4

Table 7.2: 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

Table 7.3: Results for the traffic assignment problems

	$\epsilon_{\text{opt}} = 10^{-2}$			$\epsilon_{\text{opt}} = 10^{-2.5}$			$\epsilon_{\text{opt}} = 10^{-3}$		
Problem	Delay	k	Time	Delay	k	Time	Delay	k	Time
Barcelona	1.27279e+6	140	11	1.26938e+6	320	25	1.26865e+6	910	72
Linköping	4.06050e+8	120	4	4.05774e+8	150	4	4.05716e+8	720	21
Winnipeg	8.89731e+5	56	5	8.86426e+5	116	10	8.85725e+5	220	20
Chicago	4.06489e+6	80	74	4.04218e+6	130	120	4.04004e+6	269	249
Skåne	7.64631e+7	20	172	7.63957e+7	44	377	7.63712e+7	80	685

for Skåne.

Acknowledgment. We 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

[AnW93] 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.

[BaA98] F. Barahona and R. Anbil, *On some difficult linear programs coming from set partitioning*, Research Report RC 21410, IBM Research Division, T.J. Watson Research Center, Yorktown Heights, New York, 1998.

[BaA00] ———, *The volume algorithm: Producing primal solutions with a subgradient method*, Math. Programming **87** (2000) 385–399.

[BaC96] E. Balas and M. C. Carrera, *A dynamic subgradient-based branch-and-bound procedure for set covering*, Oper. Res. **44** (1996) 875–890.

[BaC00a] F. Barahona and D. Chudak, *Solving large scale uncapacitated facility location problems*, in Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems, P. M. Pardalos, ed., Kluwer, Dordrecht, 2000, pp. 48–62.

[BaC00b] F. Barahona and F. Chudak, *Near-optimal solutions to large scale facility location problems*, Research Report RC 21606, IBM Research Division, T.J. Watson Research Center, Yorktown Heights, New York, 2000.

[BaL01] F. Barahona and L. Ladányi, *Branch and cut based on the volume algorithm: Steiner trees in graphs and max-cut*, Research Report RC ?, IBM Research Division, T.J. Watson Research Center, Yorktown Heights, New York, 2001.

[BBP00] L. Bahiense, F. Barahona and O. Porto, *Solving Steiner tree problems in graphs with Lagrangian relaxation*, Research Report RC 21846, IBM Research Division, T.J. Watson Research Center, Yorktown Heights, New York, 2000.

[Bea93] J. E. Beasley, *Lagrangean relaxation*, in Modern Heuristic Techniques for Combinatorial Problems, C. R. Reeves, ed., Blackwell Scientific Publications, Oxford, 1993, pp. 243–303.

[Ber98] D. P. Bertsekas, *Network Optimization: Continuous and Discrete Models*, Athena Scientific, Belmont, MA, 1998.

[Ber99] ———, *Nonlinear Programming*, second ed., Athena Scientific, Belmont, MA, 1999.

[BMS01] L. Bahiense, N. Maculan and C. A. Sagastizábal, *The volume algorithm revisited: Relation with bundle methods*, Tech. report, COPPE, UFRJ, Rio de Janeiro, Brasil, 2001.

[BSS93] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, second ed., Wiley, New York, 1993.

[BuF91] J. V. Burke and M. C. Ferris, *Characterization of solutions sets of convex programs*, Oper. Res. Lett. **10** (1991) 57–60.

[CFT96] A. Caprara, M. Fischetti and P. Toth, *A heuristic algorithm for the set covering problem*, in Integer Programming and Combinatorial Optimization, W. H. Cunningham, S. T. McCormick and M. Queyranne, eds., Lecture Notes in Computer Science 1084, Springer-Verlag, Berlin, 1996, pp. 72–84.

[CNS98] S. Ceria, P. Nobili and A. Sassano, *A Lagrangian-based heuristic for large-scale set covering problems*, Math. Programming **81** (1998) 215–228.

[FeK00] 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.

[Fuk84a] M. Fukushima, *A nonsmooth optimization approach to nonlinear multicommodity network flow problems*, J. Oper. Res. Soc. Japan **27** (1984) 151–176.

[Fuk84b] ———, *On a dual approach to the traffic assignment problem*, Transportation Res. **18** (1984) 235–245.

[GaP88] G. Gallo and S. Pallotino, *Shortest path algorithms*, Ann. Oper. Res. **13** (1988) 3–79.

[GGSV96] 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.

[Gof87] J.-L. Goffin, *The ellipsoid method and its predecessors*, in 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.

[GSV97] 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.

[HLV87] D. W. Hearn, S. Lawphongpanich and J. A. Ventura, *Restricted simplicial decomposition: Computation and extensions*, Math. Programming Stud. **31** (1987) 99–118.

[HUL93] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer, Berlin, 1993.

[Kiw90] K. C. Kiwiel, *Proximity control in bundle methods for convex nondifferentiable minimization*, Math. Programming **46** (1990) 105–122.

[Kiw95] ———, *Proximal level bundle methods for convex nondifferentiable optimization, saddle-point problems and variational inequalities*, Math. Programming **69** (1995) 89–109.

[Kiw96] ———, *The efficiency of subgradient projection methods for convex optimization, part II: Implementations and extensions*, SIAM J. Control Optim. **34** (1996) 677–697.

[Kiw00] ———, *Efficiency of proximal bundle methods*, J. Optim. Theory Appl. **104** (2000) 589–603.

[Kiw03] ———, *Convergence of approximate and incremental subgradient methods for convex optimization*, SIAM J. Optim. ? (2003). To appear.

[KLL99a] 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.

[KLL99b] ———, *The efficiency of ballstep subgradient level methods for convex optimization*, Math. Oper. Res. **24** (1999) 237–254.

[LaL97] T. Larsson and Z. Liu, *A Lagrangean relaxation scheme for structured linear programs with application to multicommodity network flows*, Optimization **40** (1997) 247–284.

[LaP92] T. Larsson and M. Patriksson, *Simplicial decomposition with disaggregated representation for the traffic assignment problem*, Transportation Sci. **26** (1992) 4–17.

[LLP97] T. Larsson, Z. Liu and M. Patriksson, *A dual scheme for traffic assignment problems*, Optimization **42** (1997) 323–358.

[LNN95] C. Lemaréchal, A. S. Nemirovskii and Yu. E. Nesterov, *New variants of bundle methods*, Math. Programming **69** (1995) 111–147.

[LPR97] T. Larsson, M. Patriksson and C. Rydergren, *Applications of simplicial decomposition with nonlinear column generation to nonlinear network flows*, in 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.

[LPS98] T. Larsson, M. Patriksson and A.-B. Strömberg, *Ergodic convergence in subgradient optimization*, Optim. Methods Softw. **9** (1998) 93–120.

[LPS99] ———, *Ergodic, primal convergence in dual subgradient schemes for convex programming*, Math. Programming **86** (1999) 283–312.

[Min86] M. Minoux, *Mathematical Programming, Theory and Algorithms*, John Wiley and Sons, Chichester, 1986.

[OMV00] A. Ouorou, P. Mahey and J.-Ph. Vial, *A survey of algorithms for convex multicommodity flow problems*, Management Sci. **46** (2000) 126–147.

[Pol69] 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.

[Pol83] ———, *Introduction to Optimization*, Nauka, Moscow, 1983. English transl., Optimization Software Inc., New York, 1987.

[Roc70] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[SeS86] S. Sen and H. D. Sherali, *A class of convergent primal-dual subgradient algorithms for decomposable convex programs*, Math. Programming **35** (1986) 279–297.

[ShC96] 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.

[Sho79] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, Naukova Dumka, Kiev, 1979 (Russian). English transl., Springer-Verlag, Berlin, 1985.

[TEB98] 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.

[Zhu77] 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 (Russian).

A Appendix

Proof of Lem. 3.7. The inequalities $f_* \geq \min_S \tilde{f}_k \geq \tilde{f}^k_{\min}$ follow from (cf. (3.19))

$$f(x) \geq \tilde{f}_k(x) = \tilde{f}^k_S(x) - \tilde{i}^k_S(x) \geq \tilde{f}^k_S(x) \quad \text{for each } x \text{ in } S$$

(since $\tilde{\imath}_S^k(x) \le \imath_S(x) = 0\ \forall x \in S$). Let $\tilde{x}^k \in \operatorname{Arg\,min}_S \tilde{f}_S^k$, so that $\tilde{f}_S^k(\tilde{x}^k) = \tilde{f}_{\min}^k \le f_* \le f(x^{k(l)})$ for $k = k(l+1)$. Set $x = \tilde{x}^k$ in (3.16) and use the Cauchy-Schwarz inequality together with $\tilde{\epsilon}_k, |\nabla \tilde{f}_S^k| \xrightarrow{K} 0$ (cf. Lem. 3.5(i)) and boundedness of $\{x^k\}, \{\tilde{x}^k\} \subset S$ to get

$$0 \le f(x^{k(l)}) - \tilde{f}_{\min}^k = f(x^{k(l)}) - \tilde{f}_S^k(\tilde{x}^k) = \tilde{\epsilon}_k - \left\langle \nabla \tilde{f}_S^k, \tilde{x}^k - x^{k(l)} \right\rangle \le \tilde{\epsilon}_k + |\nabla \tilde{f}_S^k||\tilde{x}^k - x^{k(l)}| \xrightarrow{K} 0.$$

But $f(x^{k(l)}) \downarrow f_*$ (Rem. 2.2(vi)), so the preceding relation gives $\tilde{f}_{\min}^k \xrightarrow{K} f_*$. □

Proof of Thm 4.1. (i) By (3.4)–(3.5), $\nabla \tilde{f}_k \in \operatorname{co}\{g_f^j\}_{j=k(l)}^k$, so $\{\nabla \tilde{f}_k\}_{k \in K'}$ is bounded by Lem. 3.5(ii). Next, $\nabla \tilde{f}_k \in \partial_{\tilde{\epsilon}_f^k} f(x^{k(l)})$ (Lem. 3.3) with $x^{k(l)} \xrightarrow{L'} x^\infty$ and $\tilde{\epsilon}_f^k \xrightarrow{K'} 0$ (Lem. 3.5(i)) imply that each cluster point of $\{\nabla \tilde{f}_k\}_{k \in K'}$ lies in $\partial f(x^\infty)$, since the approximate subdifferential mapping $(x, \epsilon) \mapsto \partial_\epsilon f(x)$ is closed [HUL93, §XI.4.1].

(ii) Using (3.5), the facts that $\tilde{\epsilon}_f^k := f(x^{k(l)}) - \tilde{f}_k(x^{k(l)}) \to 0$ and $f(x^{k(l)}) \downarrow f(x^\infty)$ (cf. (3.15) and Lem. 3.5), and our assumption $\nabla \tilde{f}_k \xrightarrow{K''} \nabla \tilde{f}_\infty$, we obtain

$$\tilde{f}_k(\cdot) = f(x^{k(l)}) - \tilde{\epsilon}_f^k + \left\langle \nabla \tilde{f}_k, \cdot - x^{k(l)} \right\rangle \xrightarrow{K''} f(x^\infty) + \left\langle \nabla \tilde{f}_\infty, \cdot - x^\infty \right\rangle =: \tilde{f}_\infty(\cdot).$$

By (i), $\nabla \tilde{f}_\infty \in \partial f(x^\infty)$. Next, $\nabla \tilde{f}_S^k - \nabla \tilde{f}_k = \nabla \tilde{\imath}_S^k \in \partial_{\tilde{\epsilon}_S^k} \imath_S(x^{k(l)})$ (cf. (3.5) and Lem. 3.3) with $\nabla \tilde{f}_S^k \xrightarrow{K} 0$, $\tilde{\epsilon}_S^k \to 0$ (cf. Lem. 3.5(i)) yield $\nabla \tilde{\imath}_S^k \xrightarrow{K''} -\nabla \tilde{f}_\infty \in \partial \imath_S(x^\infty)$ by the closedness of $\partial_\epsilon \imath_S(x)$. Since $\tilde{\epsilon}_S^k := -\tilde{\imath}_S^k(x^{k(l)})$ (cf. (3.15)), $\tilde{\imath}_S^k(\cdot) = \tilde{\imath}_S^k(x^{k(l)}) + \left\langle \nabla \tilde{\imath}_S^k, \cdot - x^{k(l)} \right\rangle \xrightarrow{K''} \tilde{\imath}_S^\infty(\cdot)$.

(iii) $\{\nabla \tilde{\imath}_S^k = \nabla \tilde{f}_S^k - \nabla \tilde{f}_k\}_{k \in K'}$ is bounded, since $\nabla \tilde{f}_S^k \xrightarrow{K} 0$ by Lem. 3.5(i) and $\{\nabla \tilde{f}_k\}_{k \in K'}$ is bounded by (i). If $\{\nabla \tilde{\imath}_S^k\}_{k \in K'}$ has a cluster point $\nabla \tilde{\imath}_S^\infty$, then by (i,ii), $\{\nabla \tilde{f}_k\}_{k \in K'}$ has a cluster point $\nabla \tilde{f}_\infty$ such that $\nabla \tilde{\imath}_S^\infty = -\nabla \tilde{f}_\infty \in \mathcal{N}_S(x^\infty)$.

(iv) This follows from (i)–(iii) and the continuity of $d_{\mathcal{G}}$ (e.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 $d_{\mathcal{G}}(\nabla \tilde{f}_k) \xrightarrow{K''} 0$). □