SIAM J. OPTIM.
Vol. 16, No. 4, pp. 1007–1023

A PROXIMAL BUNDLE METHOD WITH APPROXIMATE SUBGRADIENT LINEARIZATIONS*

KRZYSZTOF C. KIWIEL†

Abstract. We give a proximal bundle method for minimizing a convex function f over a closed convex set. It only requires evaluating f and its subgradients with an accuracy $\epsilon > 0$, which is fixed but possibly unknown. It asymptotically finds points that are ϵ-optimal. When applied to Lagrangian relaxation, it allows for ϵ-accurate solutions of Lagrangian subproblems and finds ϵ-optimal solutions of convex programs.

Key words. nondifferentiable optimization, convex programming, proximal bundle methods, approximate subgradients, Lagrangian relaxation

AMS subject classifications. 65K05, 90C25

DOI. 10.1137/040603929

1. Introduction.

We consider the convex constrained minimization problem

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

where S is a nonempty closed convex set in the Euclidean space $\mathbb{R}^n$ with inner product $\langle \cdot, \cdot \rangle$ and norm $|\cdot|$, and $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function. We assume that for fixed *accuracy tolerances* $\epsilon_f \geq 0$ and $\epsilon_g \geq 0$, for each $y \in S$ we can find an *approximate value* f_y and an *approximate subgradient* g_y of f that produce the *approximate linearization* of f:

$$\bar{f}_y(\cdot) := f_y + \langle g_y, \cdot - y \rangle \leq f(\cdot) + \epsilon_g \quad \text{with} \quad \bar{f}_y(y) = f_y \geq f(y) - \epsilon_f. \tag{1.2}$$

Thus $f_y \in [f(y) - \epsilon_f, f(y) + \epsilon_g]$ estimates $f(y)$, while $g_y \in \partial_\epsilon f(y)$ for the *total accuracy tolerance* $\epsilon := \epsilon_f + \epsilon_g$; i.e., g_y is a member of the ϵ-subdifferential of f at y,

$$\partial_\epsilon f(y) := \{\, g : f(\cdot) \geq f(y) - \epsilon + \langle g, \cdot - y \rangle \,\}\,.$$

The above assumption is realistic in many applications. For instance, if f is a max-type function of the form

$$f(y) := \sup\{\, F_z(y) : z \in Z \,\}\,, \tag{1.3}$$

where each $F_z : \mathbb{R}^n \to \mathbb{R}$ is convex and Z is an infinite set, then it may be impossible to calculate $f(y)$. However, we may still consider the following two cases. In the first case of *controllable accuracy*, for each positive $\tilde{\epsilon}$ one can find an $\tilde{\epsilon}$-maximizer of (1.3), i.e., an element $z_y \in Z$ satisfying $F_{z_y}(y) \geq f(y) - \tilde{\epsilon}$; in the second case, this may be possible only for some fixed (and possibly unknown) $\tilde{\epsilon} < \infty$. In both cases we may set $f_y := F_{z_y}(y)$ and take g_y as any subgradient of F_{z_y} at y to satisfy (1.2) with $\epsilon_f := \tilde{\epsilon}$, $\epsilon_g := 0$; then $\epsilon = \tilde{\epsilon}$.

*Received by the editors February 6, 2004; accepted for publication (in revised form) July 1, 2004; published electronically January 26, 2006. This research was supported by the State Committee for Scientific Research under grant 4T11A00622.
http://www.siam.org/journals/siopt/16-4/60392.html
†Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01–447 Warsaw, Poland (kiwiel@ibspan.waw.pl).

A special case of (1.3) arises in *Lagrangian relaxation* [Ber99, section 5.5.3], [HUL93, Chap. XII], where problem (1.1) with $S := \mathbb{R}^n_+$ is the Lagrangian dual of the primal problem

$$\sup \ \psi_0(z) \quad \text{s.t.} \quad \psi_j(z) \geq 0,\ j = 1\colon n,\ z \in Z, \tag{1.4}$$

with $F_z(y) := \psi_0(z) + \langle y, \psi(z)\rangle$ for $\psi := (\psi_1, \dots, \psi_n)$. Then, for each multiplier $y \geq 0$, we need only find $z_y \in Z$ such that $f_y := F_{z_y}(y) \geq f(y) - \epsilon$ in (1.3) to use $g_y := \psi(z_y)$. For instance, if (1.4) is a semidefinite program with each ψ_j affine and Z the set of symmetric positive semidefinite matrices of order m with unit trace, then $f(y)$ is the maximum eigenvalue of a symmetric matrix $M(y)$ depending affinely on y [Tod01, section 6.3], and z_y can be found by computing an approximate eigenvector corresponding to the maximum eigenvalue of $M(y)$ via the Lanczos method [HeK02], [HeR00].

This paper extends the proximal bundle method of [Kiw90] and its variants [Hin01], [ScZ92], [HUL93, section XV.3] to the inexact setting of (1.2) with *unknown* ϵ_f and ϵ_g. Our extension is natural and simple: the original method is run as if the linearizations were exact until a *predicted descent test* discovers their inaccuracy; then the method is restarted with a decreased proximity weight. Since our descent test (or a similar one) is employed as a stopping criterion by the existing implementations of proximal bundle methods, our analysis also sheds light on the implementations' behavior in the inexact case (cf. section 4.5).

We show that our method asymptotically estimates the optimal value f_* of (1.1) with accuracy ϵ and finds ϵ-optimal points. In Lagrangian relaxation, under standard convexity and compactness assumptions on problem (1.4) (see section 5), it finds ϵ-optimal primal solutions by combining partial Lagrangian solutions, even when Lagrange multipliers don't exist. This seems to be the first such result on primal recovery in Lagrangian relaxation.

We now comment briefly on other relations with the literature.

The setting of (1.2) subsumes those in [Hin01], [Kiw85], [Kiw95a]. Indeed, suppose that for some nonnegative tolerances $\tilde{\epsilon}_f^-$, $\tilde{\epsilon}_f^+$, and $\tilde{\epsilon}_g$, for each $y \in S$ we can find some

$$f_y \in \left[\, f(y) - \tilde{\epsilon}_f^-, f(y) + \tilde{\epsilon}_f^+ \,\right] \quad \text{and} \quad g_y \in \partial_{\tilde{\epsilon}_g} f(y). \tag{1.5}$$

Then (1.2) holds with $\epsilon_f := \tilde{\epsilon}_f^-$ and $\epsilon_g := \tilde{\epsilon}_f^+ + \tilde{\epsilon}_g$. We add that $\tilde{\epsilon}_f^- = \tilde{\epsilon}_f^+ = \tilde{\epsilon}_g$ in [Kiw85], [Hin01] uses $\tilde{\epsilon}_f^- = \tilde{\epsilon}_f^+ = 0$, i.e., exact values $f_y = f(y)$, whereas [Kiw95a] employs (1.2) with $\epsilon_g = 0$ (corresponding to $\tilde{\epsilon}_f^- := \tilde{\epsilon}_g := \epsilon_f = \epsilon$ and $\tilde{\epsilon}_f^+ := 0$ in (1.5)).

First, our method is more widely applicable than those in [Hin01], [Kiw85], [Kiw95a], since [Kiw85], [Kiw95a] assume that the $\tilde{\epsilon}$-tolerances in (1.5) are controllable and can be driven to 0, whereas [Hin01] needs exact f-values. Thus only our method can handle Lagrangian relaxation with subproblem solutions of unknown accuracy. Second, our convergence results are stronger than those in [Hin01], since they handle constraints and practicable stopping criteria (cf. section 4.2). Third, our method is much simpler than that of [Hin01].

Finally, the method of [Sol03] works in the setting of (1.2) with $\epsilon_g = 0$ and known (possibly varying) tolerances ϵ_f employed in its stopping criterion and the descent test. If the tolerances are below a fraction of a stopping threshold $\Delta > 0$, the method terminates, ensuring that the traditional stopping criterion of bundle methods is met for this Δ. In turn, the framework of [Mil01, section 4.5] is related to those in [Kiw85], [Kiw95a].

The paper is organized as follows. In section 2 we present our proximal bundle method. Its convergence is analyzed in section 3. Several modifications are given in section 4. Applications to Lagrangian relaxation of convex and nonconvex programs are studied in section 5.

2. The inexact proximal bundle method. We may regard (1.1) as an unconstrained problem $f_* = \min f_S$ with the *essential objective*

$$f_S := f + i_S, \tag{2.1}$$

where i_S is the *indicator* function of S ($i_S(x) = 0$ if $x \in S$, ∞ if $x \notin S$).

Our method generates a sequence of *trial points* $\{y^k\}_{k=1}^{\infty} \subset S$ for evaluating the approximate values $f_y^k := f_{y^k}$, subgradients $g^k := g_{y^k}$, and linearizations $f_k := \bar{f}_{y^k}$ such that

$$f_k(\cdot) = f_y^k + \langle g^k, \cdot - y^k\rangle \le f(\cdot) + \epsilon_g \quad \text{with} \quad f_k(y^k) = f_y^k \ge f(y^k) - \epsilon_f, \tag{2.2}$$

as stipulated in (1.2). Iteration k uses the polyhedral *cutting-plane model* of f

$$\check{f}_k(\cdot) := \max_{j \in J^k} f_j(\cdot) \quad \text{with} \quad k \in J^k \subset \{1, \dots, k\} \tag{2.3}$$

for finding

$$y^{k+1} := \arg\min \left\{ \phi_k(\cdot) := \check{f}_k(\cdot) + i_S(\cdot) + \tfrac{1}{2t_k}|\cdot - x^k|^2 \right\}, \tag{2.4}$$

where $t_k > 0$ is a *stepsize* that controls the size of $|y^{k+1} - x^k|$ and the *prox center* $x^k := y^{k(l)}$ has the value $f_x^k := f_y^{k(l)}$ for some $k(l) \le k$ (usually $f_x^k = \min_{j=1}^k f_y^j$). Note that, by (2.2),

$$f(x^k) - \epsilon_f \le f_x^k \le f(x^k) + \epsilon_g. \tag{2.5}$$

However, we may have $f_x^k < \check{f}_k(x^k) = \phi_k(x^k)$ in (2.4), in which case the *predicted descent*

$$v_k := f_x^k - \check{f}_k(y^{k+1}) \tag{2.6}$$

may be nonpositive; then t_k is increased and y^{k+1} is recomputed to decrease $\check{f}_k(y^{k+1})$ until $v_k > 0$ (specific tests on v_k for increasing t_k are discussed below and in section 4.3). A *descent* step to $x^{k+1} := y^{k+1}$ with $f_x^{k+1} := f_y^{k+1}$ occurs if $f_y^{k+1} \le f_x^k - \kappa v_k$ for a fixed $\kappa \in (0,1)$. Otherwise, a *null* step $x^{k+1} := x^k$ improves the next model $\check{f}_{k+1}$ with f_{k+1} (cf. (2.3)).

For choosing J^{k+1}, note that by the optimality condition $0 \in \partial\phi_k(y^{k+1})$ for (2.4),

$$\exists p_f^k \in \partial\check{f}_k(y^{k+1}) \text{ such that } p_S^k := -(y^{k+1} - x^k)/t_k - p_f^k \in \partial i_S(y^{k+1}) \tag{2.7}$$

and there are multipliers ν_j^k, $j \in J^k$, also known as *convex weights,* such that

$$p_f^k = \sum_{j \in J^k} \nu_j^k g^j, \ \sum_{j \in J^k} \nu_j^k = 1, \ \nu_j^k \ge 0, \ \nu_j^k \left[\check{f}_k(y^{k+1}) - f_j(y^{k+1})\right] = 0, \ j \in J^k. \tag{2.8}$$

Let $\hat{J}^k := \{j \in J^k : \nu_j^k \ne 0\}$. To save storage without impairing convergence, it suffices to choose $J^{k+1} \supset \hat{J}^k \cup \{k+1\}$ (i.e., we may drop inactive linearizations f_j with $\nu_j^k = 0$ that do not contribute to the trial point y^{k+1}).

The subgradient relations in (2.7) enable us to derive an optimality estimate from the following *aggregate linearizations* of $\check{f}_k$ and f, i_S, $\check{f}_S^k := \check{f}_k + i_S$ and f_S, respectively:

$$\bar{f}_k(\cdot) := \check{f}_k(y^{k+1}) + \langle p_f^k, \cdot - y^{k+1} \rangle \leq \check{f}_k(\cdot) \leq f(\cdot) + \epsilon_g, \tag{2.9}$$

$$\bar{\imath}_S^k(\cdot) := \langle p_S^k, \cdot - y^{k+1} \rangle \leq i_S(\cdot), \tag{2.10}$$

$$\bar{f}_S^k(\cdot) := \bar{f}_k(\cdot) + \bar{\imath}_S^k(\cdot) \leq \check{f}_S^k(\cdot) := \check{f}_k(\cdot) + i_S(\cdot) \leq f_S(\cdot) + \epsilon_g, \tag{2.11}$$

where the final inequalities follow from (2.1)–(2.3). Adding (2.9)–(2.10) and using (2.11) and the linearity of

$$\bar{f}_S^k(\cdot) = \check{f}_k(y^{k+1}) + \langle p_f^k + p_S^k, \cdot - y^{k+1} \rangle, \tag{2.12}$$

we get

$$f_x^k + \langle p^k, \cdot - x^k \rangle - \alpha_k = \bar{f}_S^k(\cdot) \leq \check{f}_S^k(\cdot) \leq f_S(\cdot) + \epsilon_g, \tag{2.13}$$

where

$$p^k := p_f^k + p_S^k = (x^k - y^{k+1})/t_k \quad \text{and} \quad \alpha_k := f_x^k - \bar{f}_S^k(x^k) \tag{2.14}$$

are the *aggregate subgradient* (cf. (2.7)) and the *aggregate linearization error*, respectively. The aggregate subgradient inequality (2.13) yields the *optimality estimate*

$$f_x^k \leq f(x) + \epsilon_g + |p^k||x - x^k| + \alpha_k \quad \text{for all } x \in S. \tag{2.15}$$

Combined with $f(x^k) - \epsilon_f \leq f_x^k$ (cf. (2.5)), the optimality estimate (2.15) says that the point x^k is ϵ*-optimal* (i.e., $f(x^k) - f_* \leq \epsilon := \epsilon_f + \epsilon_g$) if the *optimality measure*

$$V_k := \max \left\{ |p^k|, \alpha_k \right\} \tag{2.16}$$

is zero; x^k is approximately ϵ-optimal if V_k is small.

Thus we would like V_k to vanish asymptotically. Hence it is crucial to bound V_k via the predicted descent v_k, since normally bundling and descent steps drive v_k to 0. To this end, we first highlight some elementary properties of α_k and v_k; see Figure 2.1.

In other words, (2.13) and (2.5) mean that the model $\check{f}_S^k$ and its linearization $\bar{f}_S^k$ may overshoot the objective f_S by at most ϵ_g, whereas f_x^k may underestimate $f(x^k)$ by at most ϵ_f. Hence the linearization error α_k of (2.14) may drop below 0 by no more than $\epsilon := \epsilon_f + \epsilon_g$:

$$\alpha_k \geq f_x^k - \check{f}_S^k(x^k) \geq f_x^k - f(x^k) - \epsilon_g \geq -\epsilon_f - \epsilon_g = -\epsilon. \tag{2.17}$$

The predicted descent v_k (cf. (2.6)) may be expressed in terms of p^k and α_k as

$$v_k = t_k|p^k|^2 + \alpha_k = |d^k|^2/t_k + \alpha_k \quad \text{with} \quad d^k := y^{k+1} - x^k = -t_k p^k \tag{2.18}$$

being the *search direction*. Indeed, $|y^{k+1} - x^k|^2/t_k = t_k|p^k|^2$ by (2.14), whereas by (2.12)

$$\check{f}_k(y^{k+1}) = \bar{f}_S^k(y^{k+1}) = \bar{f}_S^k(x^k) + \langle p^k, y^{k+1} - x^k \rangle = \bar{f}_S^k(x^k) - |y^{k+1} - x^k|^2/t_k,$$

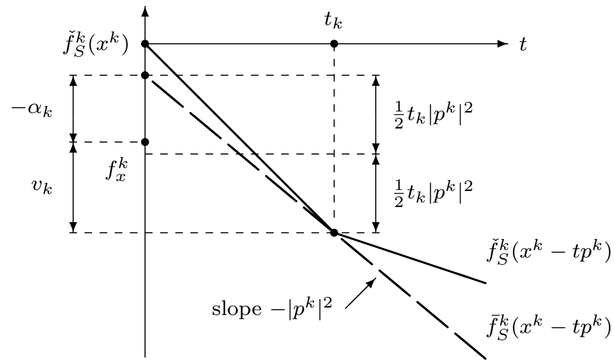

FIG. 2.1. *Predicted descent domination:* $v_k \geq -\alpha_k \Leftrightarrow \frac{1}{2}t_k|p^k|^2 \geq -\alpha_k \Leftrightarrow v_k \geq \frac{1}{2}t_k|p^k|^2$.

so $v_k := f_x^k - \check{f}_k(y^{k+1}) = \alpha_k + t_k|p^k|^2$ by (2.14). Note that $v_k \geq \alpha_k$.

Since $V_k := \max\{|p^k|, \alpha_k\}$, $v_k = t_k|p^k|^2 + \alpha_k$, and $-\alpha_k \leq \epsilon$ (cf. (2.16)–(2.18)), we have

$$V_k = \max\left\{ [(v_k - \alpha_k)/t_k]^{1/2}, \alpha_k \right\}, \tag{2.19}$$

$$V_k \leq \max\left\{ (2v_k/t_k)^{1/2}, v_k \right\} \qquad \text{if} \quad v_k \geq -\alpha_k, \tag{2.20}$$

$$V_k < (-2\alpha_k/t_k)^{1/2} \leq (2\epsilon/t_k)^{1/2} \quad \text{if} \quad v_k < -\alpha_k. \tag{2.21}$$

The bound (2.21) will imply that if x^k isn't ϵ-optimal (so that V_k can't vanish as t_k increases), then $v_k \geq -\alpha_k$ and the bound (2.20) hold for t_k large enough; on the other hand, the bound (2.20) suggests that t_k shouldn't decrease unless V_k is small enough.

We now have the necessary ingredients to state our method in detail.

ALGORITHM 2.1.

Step 0 (initialization). Select $x^1 \in S$, a *descent parameter* $\kappa \in (0,1)$, a *stepsize bound* $T_1 > 0$, and a stepsize $t_1 \in (0, T_1]$. Set $y^1 := x^1$, $f_x^1 := f_y^1$ (cf. (2.2)), $g^1 := g_{y^1}$, $J^1 := \{1\}$, $i_t^1 := 0$, $k := k(0) := 1$, $l := 0$ ($k(l) - 1$ will denote the iteration of the lth descent step).

Step 1 (trial point finding). Find y^{k+1} and multipliers ν_j^k such that (2.7)–(2.8) hold.

Step 2 (stopping criterion). If $V_k = 0$ (cf. (2.15)–(2.16)), stop ($f_x^k \leq f_* + \epsilon_g$).

Step 3 (stepsize correction). If $v_k < -\alpha_k$, set $t_k := 10t_k$, $T_k := \max\{T_k, t_k\}$, $i_t^k := k$ and loop back to Step 1; else set $T_{k+1} := T_k$.

Step 4 (descent test). Evaluate f_y^{k+1} and g^{k+1} (cf. (2.2)). If the *descent test* holds:

$$f_y^{k+1} \leq f_x^k - \kappa v_k, \tag{2.22}$$

set $x^{k+1} := y^{k+1}$, $f_x^{k+1} := f_y^{k+1}$, $i_t^{k+1} := 0$, $k(l+1) := k+1$ and increase l by 1 (*descent step*); else set $x^{k+1} := x^k$, $f_x^{k+1} := f_x^k$, and $i_t^{k+1} := i_t^k$ (*null step*).

Step 5 (bundle selection). Choose $J^{k+1} \supset \hat{J}^k \cup \{k+1\}$, where $\hat{J}^k := \{j \in J^k : \nu_j^k \neq 0\}$.

Step 6 (stepsize updating). If $k(l) = k+1$ (i.e., after a descent step), select $t_{k+1} \in [t_k, T_{k+1}]$; otherwise, either set $t_{k+1} := t_k$, or choose $t_{k+1} \in [0.1t_k, t_k]$ if $i_t^{k+1} = 0$ and

$$f_x^k - f_{k+1}(x^k) \geq V_k := \max\left\{ |p^k|, \alpha_k \right\}. \tag{2.23}$$

Step 7 (loop). Increase k by 1 and go to Step 1.

A few comments on the method are in order.

Remark 2.2.

(i) When the feasible set S is polyhedral, Step 1 may use the quadratic programming (QP) method of [Kiw94], which can efficiently solve sequences of related subproblems (2.4).

(ii) Step 2 may also use the test $f_x^k \leq \inf \check{f}_S^k$ (cf. Lemma 2.3(i)); more practicable stopping criteria are discussed in section 4.2.

(iii) In the case of exact evaluations ($\epsilon = 0$), we have $v_k \geq \alpha_k \geq 0$ (cf. (2.17)–(2.18)), Step 3 is redundant, and Algorithm 2.1 becomes essentially that of [Kiw90].

(iv) To see the need for increasing t_k at Step 3, suppose $n = 1$, $f(x) = -x$, $S = \mathbb{R}$, $x^1 = 0$, $t_1 = \epsilon = 1$, $f_x^1 = g^1 = -1$, $f_2(x) = -x$. If Step 3 were omitted and null steps were taken when $v_k \leq 0$, the method would jam with $y^{k+1} = 1$ for $k \geq 1$. Also note that decreasing t_k would not help. In fact decreasing t_k at Step 6 aims at collecting more local information about f at null steps, whereas in such cases t_k must be increased to produce descent or confirm that x^k is ϵ-optimal (let $f(x) = \max\{-x, x-2\}$ above). Hence whenever t_k is increased at Step 3, the *stepsize indicator* $i_t^k \neq 0$ prevents Step 6 from decreasing t_k after null steps until the next descent step occurs (cf. Step 4).

(v) At Step 5, one may let $J^{k+1} := J^k \cup \{k+1\}$ and then, if necessary, drop from J^{k+1} an index $j \in J^k \setminus \hat{J}^k$ with the smallest $f_j(x^k)$ to keep $|J^{k+1}| \leq M$ for some $M \geq n+2$.

(vi) Step 6 may use the procedure of [Kiw90, section 2] for updating the proximity weight $u_k := 1/t_k$, with obvious modifications.

We now show that the loop between Steps 1 and 3 is infinite iff $f_x^k \leq \inf \check{f}_S^k < \check{f}_k(x^k)$, in which case the current iterate x^k is already ϵ-optimal.

LEMMA 2.3.

(i) *If* $f_x^k \leq \inf \check{f}_S^k$, *then* $f(x^k) - \epsilon_f \leq f_x^k \leq f_* + \epsilon_g$ *and* $f(x^k) \leq f_* + \epsilon$.

(ii) *Step 2 terminates, i.e.,* $V_k := \max\{|p^k|, \alpha_k\} = 0$, *iff* $f_x^k \leq \min \check{f}_S^k = \check{f}_S^k(x^k)$.

(iii) *If the loop between Steps* 1 *and* 3 *is infinite, then* $f_x^k \leq \inf \check{f}_S^k$ $(< \check{f}_S^k(x^k)$; *cf.* (ii)). *Moreover, in this case we have* $\check{f}_S^k(y^{k+1}) \downarrow \inf \check{f}_S^k$ *as* $t_k \uparrow \infty$.

(iv) *If* $f_x^k \leq \inf \check{f}_S^k$ *at Step* 1 *and Step* 2 *does not terminate* (*i.e.,* $\inf \check{f}_S^k < \check{f}_S^k(x^k)$; *cf.* (ii)), *then an infinite loop between Steps* 3 *and* 1 *occurs.*

Proof. (i) Combine $f_* = \inf f_S$ (cf. (1.1), (2.1)) with $\inf \check{f}_S^k \leq \inf f_S + \epsilon_g$ (cf. (2.13)) and $f(x^k) - \epsilon_f \leq f_x^k$ (cf. (2.5)), and use $\epsilon := \epsilon_f + \epsilon_g$ for the second inequality.

(ii) "$\Rightarrow$": Since $|p^k| = 0 \geq \alpha_k$, (2.13)–(2.14) yield $\bar{f}_S^k(x^k) \leq \check{f}_S^k(\cdot)$, $y^{k+1} = x^k$, and $f_x^k \leq \bar{f}_S^k(x^k)$, whereas by (2.12), $\bar{f}_S^k(x^k) = \check{f}_k(y^{k+1}) = \check{f}_S^k(x^k)$. "$\Leftarrow$": Since $\check{f}_S^k(x^k) = \min \check{f}_S^k$, using $\phi_k(x^k) = \min \check{f}_S^k \leq \phi_k(y^{k+1}) \leq \phi_k(x^k)$ in (2.4) gives $y^{k+1} = x^k$, so again $\bar{f}_S^k(x^k) = \check{f}_S^k(x^k)$ by (2.12), and (2.14) yields $p^k = 0$ and $\alpha_k = f_x^k - \check{f}_S^k(x^k) \leq 0$.

(iii) At Step 3 during the loop the facts $V_k < (2\epsilon/t_k)^{1/2}$ (cf. (2.21)) and $t_k \uparrow \infty$ give $\max\{|p^k|, \alpha_k\} =: V_k \to 0$, so (2.13) yields $f_x^k \leq \inf \check{f}_S^k$. The fact that $\check{f}_S^k(y^{k+1}) \downarrow \inf \check{f}_S^k$ as $t_k \uparrow \infty$ in (2.4) is well known; see, e.g., [Kiw95b, Lem. 2.1].

(iv) By (2.11), $\check{f}_k(y^{k+1}) = \check{f}_S^k(y^{k+1}) \geq \inf \check{f}_S^k$. Thus (cf. (2.6)) $v_k \leq f_x^k - \inf \check{f}_S^k \leq 0$ and (cf. (2.18)) $v_k = t_k|p^k|^2 + \alpha_k$ yield $\alpha_k \leq -t_k|p^k|^2$ at Step 3 with $p^k \neq 0$ (since

$\max\{|p^k|, \alpha_k\} =: V_k > 0$ at Step 2). Hence $\alpha_k < -\frac{t_k}{2}|p^k|^2$, so (cf. (2.18)) $v_k < -\alpha_k$ and Step 3 loops back to Step 1, after which Step 2 can't terminate due to (ii). ∎

Remark 2.4. By Lemma 2.3, the algorithm may terminate if $f_x^k \le \inf \check{f}_S^k$. When S is polyhedral, then either $\inf \check{f}_S^k = -\infty$, or there is $\check{t}_k$ such that $\check{f}_S^k(y^{k+1}) = \min \check{f}_S^k$ whenever $t_k \ge \check{t}_k$; either case may be discovered by a parametric QP method [Kiw95b], and the algorithm may stop if $f_x^k \le \min \check{f}_S^k$, thus forestalling an infinite loop in Steps 1 through 3.

3. Convergence. In view of Lemma 2.3, we may suppose that the algorithm neither terminates nor loops infinitely between Steps 1 and 3 (otherwise x^k is ϵ-optimal). At Step 4, $y^{k+1} \in S$ and $v_k > 0$ (by (2.20), since $V_k > 0$ at Step 2), so $x^{k+1} \in S$ and $f_x^{k+1} \le f_x^k$ for all k.

Let $f_x^\infty := \lim_k f_x^k$. We shall show that $f_x^\infty \le f_* + \epsilon_g$. Because the proof is quite complex, it is broken into a series of lemmas, starting with the following two simple results. To handle loops, let V_k' denote the minimum value of V_k at each iteration k.

LEMMA 3.1. *If* $\underline{\lim}_k V_k' = 0$ *(e.g.,* $\underline{\lim}_k V_k = 0$*) and* $\{x^k\}$ *is bounded, then* $f_x^\infty \le f_* + \epsilon_g$.

Proof. Pick $K \subset \{1, 2, \dots\}$ such that $V_k' \xrightarrow{K} 0$. Fix $x \in S$. Letting $k \in K$ tend to infinity in (2.15)–(2.16) with $V_k = V_k'$ yields $f_x^\infty \le f(x) + \epsilon_g$, so $f_x^\infty \le \inf_S f + \epsilon_g = f_* + \epsilon_g$. ∎

LEMMA 3.2. *Let* $T_\infty := \lim_k T_k$ *at Step* 4. *If* $T_\infty = \infty$*, then* $\underline{\lim}_k V_k' = 0$.

Proof. Let $K \subset \{1, 2, \dots\}$ index iterations k that increase T_k at Step 3. For $k \in K$, at Step 3 on the last loop to Step 1 we have $V_k < (2\epsilon/t_k)^{1/2}$ (cf. (2.21)) with t_k such that $10t_k$ becomes the final T_k, so the facts $0 \le V_k' \le V_k$ and $T_k \xrightarrow{K} \infty$ give $V_k' \xrightarrow{K} 0$. ∎

In view of Lemmas 3.1–3.2, we may assume that $T_\infty < \infty$ when $\{x^k\}$ is bounded, e.g., only finitely many descent steps occur. This case is analyzed below.

LEMMA 3.3. *Suppose there exists* $\bar{k}$ *such that for all* $k \ge \bar{k}$*, Step* 3 *doesn't increase* t_k *and only null steps occur with* $t_{k+1} \le t_k$ *determined by Step* 6. *Then* $v_k \to 0$.

Proof. Fix $k \ge \bar{k}$. We first show that $\check{f}_S^{k+1} \ge \bar{f}_S^k$. Let $\hat{f}_k := \max_{j \in \hat{J}^k} f_j$. Since $\hat{J}^k := \{j \in J^k : \nu_j^k \ne 0\}$ and $g^j = \nabla f_j$, $\hat{f}_k \le \max_{j \in J^k} f_j =: \check{f}_k$ and (2.8) yield $\hat{f}_k(y^{k+1}) = \check{f}_k(y^{k+1})$ and $p_f^k \in \partial \hat{f}_k(y^{k+1})$. Thus $\bar{f}_k \le \hat{f}_k$ by (2.9), so $\hat{f}_k \le \check{f}_{k+1}$ ($\hat{J}^k \subset J^{k+1}$) gives $\bar{f}_k \le \check{f}_{k+1}$. Hence (2.10)–(2.11) yield $\bar{f}_S^k := \bar{f}_k + \bar{\imath}_S^k \le \check{f}_{k+1} + i_S =: \check{f}_S^{k+1}$.

Next, consider the following partial linearization of the objective ϕ_k of (2.4):

$$\bar{\phi}_k(\cdot) := \bar{f}_S^k(\cdot) + \tfrac{1}{2t_k}|\cdot - x^k|^2. \tag{3.1}$$

We have $\nabla \bar{\phi}_k(y^{k+1}) = 0$ from $\nabla \bar{f}_S^k = p^k = (x^k - y^{k+1})/t_k$ (cf. (2.13)–(2.14)), and $\bar{f}_S^k(y^{k+1}) = \check{f}_k(y^{k+1})$ by (2.12), so $\bar{\phi}_k(y^{k+1}) = \phi_k(y^{k+1})$ (cf. (2.4)) and by Taylor's expansion,

$$\bar{\phi}_k(\cdot) = \phi_k(y^{k+1}) + \tfrac{1}{2t_k}|\cdot - y^{k+1}|^2. \tag{3.2}$$

By (3.1) and (2.11), we have $\bar{\phi}_k(x^k) = \bar{f}_S^k(x^k) \le f(x^k) + \epsilon_g$ (using $x^k \in S$); hence by (3.2),

$$\phi_k(y^{k+1}) + \tfrac{1}{2t_k}|y^{k+1} - x^k|^2 = \bar{\phi}_k(x^k) \le f(x^k) + \epsilon_g. \tag{3.3}$$

Now, using $x^{k+1} = x^k$, $t_{k+1} \le t_k$, and $\check{f}_S^{k+1} \ge \bar{f}_S^k$ in (2.4) and (3.1) gives $\phi_{k+1} \ge \bar{\phi}_k$, so

$$\phi_k(y^{k+1}) + \tfrac{1}{2t_k}|y^{k+2} - y^{k+1}|^2 \le \phi_{k+1}(y^{k+2}) \tag{3.4}$$

by (3.2). Since $x^k = x^{\bar{k}}$ and $t_k \le t_{\bar{k}}$ for $k \ge \bar{k}$, by (3.3)–(3.4) there exists $\phi_\infty \le f(x^{\bar{k}}) + \epsilon_g$ such that

$$\phi_k(y^{k+1}) \uparrow \phi_\infty, \quad y^{k+2} - y^{k+1} \to 0, \tag{3.5}$$

and $\{y^{k+1}\}$ is bounded. Then $\{g^k\}$ is bounded as well, since $g^k \in \partial_\epsilon f(y^k)$ with $\epsilon := \epsilon_f + \epsilon_g$ by (2.2), whereas $\partial_\epsilon f$ is locally bounded [HUL93, section XI.4.1].

We now show that the *approximation error* $\check{\epsilon}_k := f_y^{k+1} - \check{f}_k(y^{k+1})$ vanishes. Using the form (2.2) of f_{k+1}, the bound $f_{k+1} \le \check{f}_{k+1}$ (cf. (2.3)), the Cauchy–Schwarz inequality, and (2.4) with $x^k = x^{\bar{k}}$ and $t_{k+1} \le t_k$ for $k \ge \bar{k}$, we estimate

$$\begin{aligned}
\check{\epsilon}_k &:= f_y^{k+1} - \check{f}_k(y^{k+1}) = f_{k+1}(y^{k+2}) - \check{f}_k(y^{k+1}) + \langle g^{k+1}, y^{k+1} - y^{k+2} \rangle \\
&\le \check{f}_{k+1}(y^{k+2}) - \check{f}_k(y^{k+1}) + |g^{k+1}||y^{k+1} - y^{k+2}| \\
&= \phi_{k+1}(y^{k+2}) - \phi_k(y^{k+1}) + |g^{k+1}||y^{k+1} - y^{k+2}| \\
&\qquad - \tfrac{1}{2t_{k+1}}|y^{k+2} - x^{\bar{k}}|^2 + \tfrac{1}{2t_k}|y^{k+1} - x^{\bar{k}}|^2 \\
&\le \phi_{k+1}(y^{k+2}) - \phi_k(y^{k+1}) + |g^{k+1}||y^{k+1} - y^{k+2}| + \Delta_k, \qquad (3.6)
\end{aligned}$$

where

$$\begin{aligned}
\Delta_k &:= \tfrac{1}{2t_k}\left(|y^{k+1} - x^{\bar{k}}|^2 - |y^{k+2} - x^{\bar{k}}|^2 \right) \\
&\le \tfrac{1}{2t_k}\left(|y^{k+1} - y^{k+2}|^2 + 2|y^{k+2} - y^{k+1}||y^{k+2} - x^{\bar{k}}| \right) \\
&\le \tfrac{1}{2t_k}|y^{k+1} - y^{k+2}|^2 + \left(\tfrac{1}{t_k}|y^{k+1} - y^{k+2}|^2 \tfrac{1}{t_{k+1}}|y^{k+2} - x^{\bar{k}}|^2 \right)^{1/2}.
\end{aligned}$$

We have $\overline{\lim}_k \Delta_k \le 0$, since $\frac{1}{2t_k}|y^{k+1} - y^{k+2}|^2 \to 0$ by (3.4)–(3.5), whereas $\frac{1}{t_{k+1}}|y^{k+2} - x^{\bar{k}}|^2$ is bounded by (3.3). Hence using (3.5) and the boundedness of $\{g^{k+1}\}$ in (3.6) yields $\overline{\lim}_k \check{\epsilon}_k \le 0$. On the other hand, the null step condition $f_y^{k+1} > f_x^k - \kappa v_k$ for $k \ge \bar{k}$ gives

$$\check{\epsilon}_k = \left[f_y^{k+1} - f_x^k \right] + \left[f_x^k - \check{f}_k(y^{k+1}) \right] > -\kappa v_k + v_k = (1-\kappa) v_k \ge 0,$$

where $\kappa < 1$ by Step 0; thus $\check{\epsilon}_k \to 0$ and $v_k \to 0$. ∎

Using (2.18) we may relate the descent $v_k := f_x^k - \check{f}_k(y^{k+1})$ predicted by $\check{f}_k$ with the descent predicted by the augmented model ϕ_k in subproblem (2.4):

$$w_k := f_x^k - \phi_k(y^{k+1}) = v_k - \tfrac{1}{2} t_k |p^k|^2 \tag{3.7a}$$

$$= \tfrac{1}{2} t_k |p^k|^2 + \alpha_k = \tfrac{1}{2}|d^k|^2 / t_k + \alpha_k. \tag{3.7b}$$

The above relations are convenient in showing that $|d^k| = O(t_k^{1/2})$ during a series of null steps that decrease t_k; this will be useful when $\underline{\lim}_k t_k = 0$.

LEMMA 3.4. *If Step 4 is entered with $i_t^k = 0$, then $|d^k|^2 \le \left(t_{k(l)} |g^{k(l)}|^2 + 2\epsilon \right) t_k$.*

Proof. First, suppose $k = k(l)$. Then (cf. Steps 0 and 4) $x^k = y^k$ and $f_x^k = f_y^k$, so using the bound $\check{f}_k \ge f_k$ (cf. (2.3)) in subproblem (2.4) and the form (2.2) of f_k gives

$$\phi_k(y^{k+1}) \ge \min \left\{ f_k(\cdot) + \tfrac{1}{2t_k} |\cdot - x^k|^2 \right\} = f_x^k - \tfrac{t_k}{2}|g^k|^2.$$

Thus $w_{k(l)} \le \frac{t_{k(l)}}{2}|g^{k(l)}|^2$ by (3.7a). Next, suppose $k > k(l)$. Then (cf. Steps 3, 4, 6) $x^{j+1} = x^{k(l)}$ and $t_{j+1} \le t_j$ for $j = k(l)\colon k-1$ due to $i_t^k = 0$, and hence $w_{j+1} \le w_j$ by

(3.4) and (3.7a). Thus $w_k \le w_{k(l)}$, and by (3.7b) and (2.17), $\frac{1}{2t_k}|d^k|^2 = w_k - \alpha_k \le w_{k(l)} + \epsilon$. ∎

We now use the safeguard (2.23) for analyzing the case of diminishing stepsizes.

LEMMA 3.5. *Suppose* $\underline{\lim}_k t_k = 0$ *at Step* 6 *and either only finitely many descent steps occur, or* $\sup_l t_{k(l)} < \infty$ *and* $\{x^k\}$ *is bounded. Then* $\underline{\lim}_k V_k = 0$ *at Step* 6.

Proof. Let C be the supremum of $t_{k(l)}|g^{k(l)}|^2 + 2\epsilon$ over the generated values of l. Note that $C < \infty$ since, if l is unbounded, then $\{g^{k(l)}\}$ is bounded because for $k = k(l)$ we have $x^k = y^k$ and $g^k \in \partial_\epsilon f(y^k)$ with $\epsilon := \epsilon_f + \epsilon_g$ by (2.2), whereas $\partial_\epsilon f$ is locally bounded.

Since $\underline{\lim}_k t_k = 0$, there is $K \subset \{1, 2, \dots\}$ such that $t_{k+1} \xrightarrow{K} 0$ at Step 6 with $t_{k+1} < t_k$ for all $k \in K$; thus $t_k \xrightarrow{K} 0$, since $t_k \le 10t_{k+1}$ at Step 6. For $k \in K$, at Step 6 we have (2.23), and at Step 4 we have $f_y^{k+1} > f_x^k - \kappa v_k$ and $i_t^k = 0$. Using $i_t^k = 0$, the definition of C and $t_k \xrightarrow{K} 0$ in Lemma 3.4 yields $|d^k|^2 \le C t_k \xrightarrow{K} 0$, i.e., $d^k \xrightarrow{K} 0$. Thus, since $\{x^k\}$ is bounded, so are $\{y^{k+1} = x^k + d^k\}_{k\in K}$ and $\{g^{k+1} \in \partial_\epsilon f(y^{k+1})\}_{k\in K}$ because $\partial_\epsilon f$ is locally bounded.

Let $k \in K$ at Step 6. Since $f_y^{k+1} > f_x^k - \kappa v_k$ and $y^{k+1} = x^k + d^k$, using (2.2) gives

$$f_x^k - f_{k+1}(x^k) = f_x^k - f_y^{k+1} - \langle g^{k+1}, x^k - y^{k+1}\rangle \le \kappa v_k + |g^{k+1}||d^k|. \tag{3.8}$$

Now, (2.23), (3.8), and the fact $v_k = |d^k||p^k| + \alpha_k$ (cf. (2.18)) imply

$$\begin{aligned} V_k := \max\left\{ |p^k|, \alpha_k \right\} &\le f_x^k - f_{k+1}(x^k) \le \kappa\left(|d^k||p^k| + \alpha_k\right) + |g^{k+1}||d^k| \\ &\le \kappa(1 + |d^k|)\max\left\{ |p^k|, \alpha_k \right\} + |g^{k+1}||d^k| = \kappa(1+|d^k|)V_k + |g^{k+1}||d^k|. \end{aligned} \tag{3.9}$$

Therefore, since $\kappa < 1$, $d^k \xrightarrow{K} 0$, and $\{g^{k+1}\}_{k\in K}$ is bounded, for large $k \in K$,

$$0 \le V_k \le |g^{k+1}||d^k| / \left[1 - \kappa(1 + |d^k|)\right] \xrightarrow{K} 0.$$

Thus $\lim_{k\in K} V_k = 0$. ∎

We may now finish the case of infinitely many consecutive null steps.

LEMMA 3.6. *Suppose there exists* $\bar{k}$ *such that only null steps occur for all* $k \ge \bar{k}$. *Then either* $T_\infty = \infty$ *and* $\underline{\lim}_k V_k' = 0$, *or* $T_\infty < \infty$ *and* $\underline{\lim}_k V_k = 0$ *at Step* 4.

Proof. If $\underline{\lim}_k t_k = 0$ at Step 6, then $\underline{\lim}_k V_k = 0$ by Lemma 3.5, so assume $\underline{\lim}_k t_k > 0$. Next, if $T_\infty = \infty$, then $\underline{\lim}_k V_k' = 0$ by Lemma 3.2, so assume $T_\infty < \infty$.

If Step 3 increases t_k for some $k = k' \ge \bar{k}$, then $t_k \ge 10t_{k-1}$ and $i_t^k \ne 0$, whereas for $k \ge k'$ Step 4 keeps $i_t^{k+1} = i_t^k \ne 0$ and Step 6 sets $t_{k+1} = t_k$, so the number of such increases must be finite (otherwise $t_k \to \infty$ and $T_\infty = \infty$, a contradiction). Hence we may assume that Step 3 doesn't increase t_k for $k \ge \bar{k}$. Then Lemma 3.3 gives $v_k \to 0$. Since (cf. (2.20)) $V_k \le \max\{(2v_k/t_k)^{1/2}, v_k\}$ and $\underline{\lim}_k t_k > 0$, we get $V_k \to 0$. ∎

For analyzing the remaining case of infinitely many descent steps, we shall use the *descent indicator* i_k defined by $i_k := 1$ if (2.22) holds and use $i_k := 0$ otherwise.

LEMMA 3.7.

(i) *If* $f_x^\infty > -\infty$, *then* $i_k v_k \to 0$ *at Step* 4.

(ii) *If* $f_x^\infty > f_* + \epsilon_g$, *then* $\{x^k\}$ *is bounded.*

Proof. (i) At Step 4, $0 \le \kappa i_k v_k \le f_x^k - f_x^{k+1}$, so $\sum_k i_k v_k \le (f_x^1 - f_x^\infty)/\kappa < \infty$.

(ii) Pick $x \in S$ and $\gamma > 0$ such that $f_x^k > f(x) + \epsilon_g + \gamma$ for all k. Since $\langle p^k, x - x^k\rangle \le \alpha_k - \gamma$ by (2.13), $x^{k+1} - x^k = -i_k t_k p^k$ and $v_k = t_k|p^k|^2 + \alpha_k$ by (2.18), we deduce that

$$\begin{aligned}|x^{k+1}-x|^2 &= |x^k-x|^2+2\langle x^{k+1}-x^k,x^k-x\rangle+|x^{k+1}-x^k|^2\\ &\le |x^k-x|^2+2i_kt_k(\alpha_k-\gamma)+2i_kt_k^2|p^k|^2\\ &= |x^k-x|^2+2i_kt_k(v_k-\gamma).\end{aligned}$$

Since $i_kv_k\to 0$ by (i), there is k_γ such that for all $k\ge k_\gamma$, $i_k(v_k-\gamma)\le 0$ above, and hence $|x^{k+1}-x|\le|x^k-x|$. Thus $\{x^k\}$ is bounded. ∎

LEMMA 3.8. *If infinitely many descent steps occur, then $f_x^\infty\le f_*+\epsilon_g$.*

Proof. Suppose for contradiction $f_x^\infty>f_*+\epsilon_g$. By Lemma 3.7(ii), $\{x^k\}$ is bounded. Further, $T_\infty<\infty$, since otherwise Lemmas 3.1 and 3.2 would yield $f_x^\infty\le f_*+\epsilon_g$, a contradiction. Similarly, $\underline{\lim}_k t_k>0$, since otherwise Lemmas 3.1 and 3.5 would yield a contradiction. Let $K:=\{k:i_k=1\}$. Using $\underline{\lim}_k t_k>0$ and $v_k\xrightarrow{K}0$ (cf. Lemma 3.7(i)) in the bound $V_k\le\max\{(2v_k/t_k)^{1/2},v_k\}$ (cf. (2.20)) yields $V_k\xrightarrow{K}0$. Hence $\underline{\lim}_k V_k=0$, and Lemma 3.1 again gives a contradiction. ∎

We may now prove our principal result. Note that $f_x^k\downarrow f_x^\infty\ge f_*-\epsilon_f$ by (2.5).

THEOREM 3.9. *We have $f_x^k\downarrow f_x^\infty\le f_*+\epsilon_g$. Moreover, $\overline{\lim}_k f(x^k)\le f_*+\epsilon$ for $\epsilon:=\epsilon_f+\epsilon_g$ so that each cluster point x^* of $\{x^k\}$ (if any) satisfies $x^*\in S$ and $f(x^*)\le f_*+\epsilon$.*

Proof. To get $f_x^\infty\le f_*+\epsilon_g$, invoke Lemmas 3.1 and 3.6 in the case of finitely many descent steps, and invoke Lemma 3.8 otherwise. By (2.5), $\overline{\lim}_k f(x^k)\le\lim_k f_x^k+\epsilon_f\le f_*+\epsilon_f+\epsilon_g$. The final assertion follows from the fact $\{x^k\}\subset S$ and the closedness of S and f. ∎

It is instructive to examine the assumptions of the preceding results.

Remark 3.10.

(i) Inspection of the proofs of Lemmas 3.3 and 3.5 reveals that Lemmas 3.3–3.8 and Theorem 3.9 require only convexity, finiteness, and closedness of f on S and *local boundedness* of the approximate subgradient mapping $g.$ on S. In particular, it suffices to assume that f is finite convex on a neighborhood of S, since $g.\in\partial_\epsilon f(\cdot)$.

(ii) For Lemma 3.5, it suffices to assume boundedness of $\{g^k\}$ instead of local boundedness of $g.$ and boundedness of $\{x^k\}$. Note that $\{x^k\}$ is bounded if f_S is coercive, since then the level set $\{x\in S:f(x)\le f_x^1+\epsilon_f\}$ is bounded and contains $\{x^k\}$ by (2.5).

The next result will justify the stopping criteria of section 4.2.

LEMMA 3.11. *Suppose $f_*>-\infty$, and either $\{g^k\}$ is bounded, or $g.$ is locally bounded and $\{x^k\}$ is bounded (e.g., f_S is coercive). Then $\underline{\lim}_k V_k'=0$.*

Proof. If only finitely many descent steps occur, then the proofs of Lemma 3.6 and Remark 3.10 yield $\underline{\lim}_k V_k'=0$. Hence suppose for contradiction that $\underline{\lim}_k V_k'>0$ for infinitely many descent steps.

We have $T_\infty<\infty$, since otherwise Lemma 3.2 would yield $\underline{\lim}_k V_k'=0$. Similarly, $\underline{\lim}_k t_k>0$, since otherwise Lemma 3.5 and Remark 3.10(ii) would imply $\underline{\lim}_k V_k=0$. Next, $f_x^k\ge f(x^k)-\epsilon_f\ge f_*-\epsilon_f>-\infty$ (cf. (2.5)) gives $f_x^\infty>-\infty$. Let $K:=\{k:i_k=1\}$. Using $\underline{\lim}_k t_k>0$ and $v_k\xrightarrow{K}0$ (cf. Lemma 3.7(i)) in the bound $V_k\le\max\{(2v_k/t_k)^{1/2},v_k\}$ (cf. (2.20)) yields $V_k\xrightarrow{K}0$ and hence $\underline{\lim}_k V_k'=0$, a contradiction. ∎

4. Modifications.

4.1. Subgradient aggregation.

4.1. Subgradient aggregation. To trade off storage and work per iteration for speed of convergence, one may replace subgradient selection with aggregation as in

[Kiw90] so that only $M \geq 2$ subgradients are stored. To this end, we note that the preceding results remain valid if, for each k, $\check{f}_{k+1}$ is a closed convex function such that $\partial(\check{f}_{k+1} + i_S) = \partial \check{f}_{k+1} + \partial i_S$ (cf. (2.7)) and

$$\max\left\{ \bar{f}_k(x), f_{k+1}(x) \right\} \leq \check{f}_{k+1}(x) \leq f(x) + \epsilon_g \quad \text{for all } x \in S. \tag{4.1}$$

Examples include $\check{f}_{k+1} = \max\{\bar{f}_k, f_{k+1}\}$, or $\check{f}_{k+1} = \max\{\bar{f}_k, f_j : j \in J^{k+1}\}$ with $k+1 \in J^{k+1} \subset \{1\colon k+1\}$, and possibly some f_j replaced by $\bar{f}_j$ for $j \leq k$. In fact the aggregate linearization $\bar{f}_k$ may be omitted in (4.1) after a descent step.

4.2. Optimality measures and stopping criteria. In practice Step 2 may use the stopping criterion $V_k \leq \epsilon_{\mathrm{opt}}$, where $\epsilon_{\mathrm{opt}} > 0$ is an *optimality tolerance*. Then any loop between Steps 1 and 3 is finite (cf. the proof of Lemma 2.3(iii)), whereas Lemma 3.11 gives conditions that ensure finite termination.

It may be more appropriate to replace V_k by the *modified optimality measure*

$$\hat{V}_k := R|p^k| + \alpha_k^+ \quad \text{with} \quad \alpha_k^+ := \max\{\alpha_k, 0\}, \tag{4.2}$$

where $R > 0$ is the "radius of the picture" [HUL93, Note XIV.3.4.3[6]], because the optimality estimate (2.15) combined with $f(x^k) \leq f_x^k + \epsilon_f$ (cf. (2.5)) gives the bounds

$$f(x^k) - \min_{|x-x^k|\leq R} f_S(x) - \epsilon \leq f_x^k - \min_{|x-x^k|\leq R} f_S(x) - \epsilon_g \leq R|p^k| + \alpha_k. \tag{4.3}$$

Since $\min\{R, 1\}V_k \leq \hat{V}_k \leq (R+1)V_k$ by (2.16) and (4.2), the preceding results hold with V_k replaced by $\hat{V}_k$, also in the safeguard (2.23) of Step 6, since (3.9) may be replaced by

$$\begin{aligned} \hat{V}_k &:= R|p^k| + \alpha_k^+ \leq f_x^k - f_{k+1}(x^k) \leq \kappa\left(|d^k||p^k| + \alpha_k\right) + |g^{k+1}||d^k| \\ &\leq \kappa(1 + |d^k|/R)(R|p^k| + \alpha_k^+) + |g^{k+1}||d^k| = \kappa(1 + |d^k|/R)\hat{V}_k + |g^{k+1}||d^k|. \end{aligned} \tag{4.4}$$

In view of (4.3), another optimality measure $\bar{V}_k := R|p^k| + \alpha_k$ may replace V_k both in the stopping criterion (since $\bar{V}_k \leq \hat{V}_k \leq (R+1)V_k$) and in the safeguard (2.23), which becomes

$$f_x^k - f_{k+1}(x^k) \geq \bar{V}_k := R|p^k| + \alpha_k. \tag{4.5}$$

LEMMA 4.1. *Suppose Step* 6 *employs the safeguard* (4.5) *instead of* (2.23)*. Then Lemma* 3.5*, Remark* 3.10*, and Lemma* 3.11 *remain true.*

Proof. We give only two replacements for (3.9). First, for $k \in K_+ := \{k \in K : \alpha_k \geq 0\}$, we have $\bar{V}_k = \hat{V}_k$ in (4.5), so (4.4) holds. Hence if K_+ is infinite, then $\hat{V}_k \xrightarrow{K_+} 0$ by the previous argument, and thus $V_k \xrightarrow{K_+} 0$ because $V_k \leq \hat{V}_k / \min\{R, 1\}$. Otherwise $K_- := \{k \in K : \alpha_k < 0\}$ is infinite. Let $k \in K_-$. Then $V_k := \max\{|p^k|, \alpha_k\} = |p^k|$, whereas $v_k \geq -\alpha_k$ and (2.18) yield $\alpha_k \geq -\frac{1}{2}t_k|p^k|^2 = -\frac{1}{2}|d^k||p^k|$, so $\bar{V}_k := R|p^k| + \alpha_k \geq (R - \frac{1}{2}|d^k|)V_k$. Hence using (4.5) we may replace (3.9) by

$$(R - \tfrac{1}{2}|d^k|)V_k \leq f_x^k - f_{k+1}(x^k) \leq \kappa|d^k||p^k| + |g^{k+1}||d^k| = \kappa|d^k|V_k + |g^{k+1}||d^k|$$

to get $V_k \xrightarrow{K_-} 0$ as before. ∎

4.3. Tests for stepsize expansion and descent. Consider replacing the test $v_k \geq -\alpha_k$ of Step 3 by the stronger test $\kappa_v v_k \geq -\alpha_k$ with a fixed coefficient $\kappa_v \in (0,1)$. The preceding results are not impaired, since (2.20)–(2.21) are replaced by

$$V_k \leq \max\left\{ [(1+\kappa_v)v_k/t_k]^{1/2}, v_k \right\} \qquad \text{if} \quad \kappa_v v_k \geq -\alpha_k,$$

$$V_k < [-(1+\kappa_v^{-1})\alpha_k/t_k]^{1/2} \leq [(1+\kappa_v^{-1})\epsilon/t_k]^{1/2} \quad \text{if} \quad \kappa_v v_k < -\alpha_k.$$

Further, the facts that $v_k = t_k|p^k|^2 + \alpha_k$ (cf. (2.18)), $w_k = \frac{1}{2}t_k|p^k|^2 + \alpha_k$ (cf. (3.7b)), and $\kappa_v v_k \geq -\alpha_k$ at Step 4 yield the bounds

$$w_k \leq v_k \leq \tfrac{2}{1-\kappa_v} w_k. \tag{4.6}$$

These bounds allow us to replace v_k by w_k in the descent test (2.22), thus bringing it closer to those of [HUL93, Alg. XV.3.1.4] and [Kiw90, section 5]. Again the preceding results extend easily (in the proof of Lemma 3.3, $f_y^{k+1} > f_x^k - \kappa w_k$ implies $f_y^{k+1} > f_x^k - \kappa v_k$, whereas in the proof of Lemma 3.7(i), $\sum_k i_k v_k \leq \frac{2}{1-\kappa_v}\sum_k i_k w_k < \infty$).

For $\kappa_v = \frac{1}{3}$, we have $w_k \leq v_k \leq 3w_k$ by (4.6), whereas the test $\kappa_v v_k \geq -\alpha_k$ is equivalent to $w_k \geq -\alpha_k$. Note that $w_k \geq 0$ is equivalent to the original test $v_k \geq -\alpha_k$.

4.4. Zigzag searches. Our analysis may accommodate zigzag searches (cf. [HUL93, section XV.3.3], [Hin01], [Kiw96], [ScZ92]), which amount to trying possibly more than one value of t_k at each iteration.

We first consider stepsize expansion at descent steps. Suppose that the descent test (2.22) holds, but $t_k < T_k$ and some other tests, e.g., $f_y^{k+1} \leq f_x^k - \bar{\kappa} v_k$ or $\langle g^{k+1}, d^k \rangle < -\bar{\kappa} v_k$ with $\bar{\kappa} \in (\kappa, 1)$, indicate that larger descent might occur if t_k were increased. Letting $\underline{t}_k := t_k$, we may choose a larger $t_k \in (\underline{t}_k, T_k]$ and go back to Step 1. If (2.22) fails when Step 4 is reentered, then a descent step must be made with t_k reset to $\underline{t}_k$. Otherwise, either a descent step with the current t_k is accepted, or a larger stepsize may be tested as above.

One may use simple safeguards, such as $1.1\underline{t}_k \leq T_k$ and $t_k \geq 1.1\underline{t}_k$, to ensure finiteness of the loop between Steps 1 and 4. (If Step 3 drove t_k and T_k to ∞, the conclusions of Lemma 2.3(iii) would hold by its proof, so a cycle between Steps 1 and 3 would occur by Lemma 2.3(iv).) In effect, the preceding results are not affected by such modifications.

To enable zigzag searches at null steps, it suffices to redefine $\check{f}_{k+1}$ after Step 6 as

$$\check{f}_{k+1} := \check{f}_k \quad \text{if} \quad t_{k+1} \leq 0.9t_k. \tag{4.7}$$

Then "$t_{k+1} \leq t_k$" in Lemma 3.3 must be replaced by "$0.9t_k < t_{k+1} \leq t_k$," but this is enough for the proof of Lemma 3.6, since if $\underline{\lim}_k t_k > 0$ and $t_{k+1} \leq t_k$ for $k \geq \bar{k}$, then $t_{k+1} > 0.9t_k$ for all large k. The remaining results are not affected.

4.5. Ad hoc modification. Our analysis also sheds light on the behavior of the original proximal bundle method [Kiw90], [HUL93, section XV.3] in the inexact case.

Consider the following crippled version of Algorithm 2.1 with the safeguard (2.23) replaced by (4.5). Suppose Step 2 employs any of the stopping criteria of section 4.2 with a positive optimality tolerance ϵ_{opt}, whereas Step 3 is replaced by

Step 3′ (inaccuracy detection). If $w_k < 0$, then stop; else set $T_{k+1} := T_k$.

This version is an ad hoc modification of the method of [Kiw90] that employs only the additional stopping criterion $w_k < 0$; in fact most existing implementations use this criterion anyway (to detect QP inaccuracy or erroneous subgradients).

As for convergence of this modification, there are three cases. First, if no termination occurs, then the results of section 3 apply (with $T_\infty = T_1$); in view of Lemma 3.11, this case is quite unlikely. Second, termination at Step 2 means a satisfactory solution has been found. Third, termination at Step 3′ implies $V_k < (2\epsilon/t_k)^{1/2}$ (cf. (2.21)); thus x^k is a satisfactory solution if t_k is "large enough"; otherwise a failure occurs.

The above analysis suggests that the existing bundle codes may behave reasonably well in the inexact case, provided large enough stepsizes are used (most codes allow the user to choose the initial stepsize and its updating strategies). Of course, in case of failure, the user may choose a larger stepsize, disallow stepsize decreases, and restart the algorithm at Step 1; such a "natural" strategy reinvents Algorithm 2.1! Finally, note that the existing codes won't face any trouble until the predicted descent v_k falls below the oracle's error ϵ (since $w_k < 0$ implies $v_k < -\alpha_k \le \epsilon$ by (3.7b), (2.18), and (2.17)).

5. Lagrangian relaxation. In this section we consider the special case where problem (1.1) with $S := \mathbb{R}^n_+$ is the Lagrangian dual problem of the *primal* convex optimization problem

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

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

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

can be evaluated with *accuracy* $\epsilon \ge 0$ by finding a *partial Lagrangian ϵ-solution*

$$z(y) \in Z \quad \text{such that} \quad f_y := \psi_0(z(y)) + \langle y, \psi(z(y))\rangle \ge f(y) - \epsilon. \tag{5.3}$$

Thus f is finite convex and has an ϵ-subgradient mapping $g_\cdot := \psi(z(\cdot))$ on S. In view of Remark 3.10(i), we suppose that $\psi(z(\cdot))$ is locally bounded on S (e.g., $\psi(z(S))$ is bounded if $\inf_Z \min_{j=1}^n \psi_j > -\infty$, or ψ is continuous on Z). Finally, we assume that f_S is coercive, i.e., $\operatorname{Arg\,min}_S f$ is nonempty and bounded (e.g., Slater's condition holds: $\psi(\check z) > 0$ for some $\check z \in Z$).

In effect, assuming $k \to \infty$, the results of section 3 hold with $\epsilon_f := \epsilon$ and $\epsilon_g := 0$, $f_* > -\infty$, $\{x^k\}$ is bounded (cf. Remark 3.10(ii)), and Lemma 3.11 yields $\underline{\lim}_k V_k' = 0$. In particular, the partial Lagrangian solutions $z^k := z(y^k)$ (cf. (5.3)) and their constraint values $g^k := \psi(z^k)$ determine the linearizations (2.2) as Lagrangian pieces of f in (5.2):

$$f_k(\cdot) = \psi_0(z^k) + \langle \cdot, \psi(z^k)\rangle. \tag{5.4}$$

Using their weights $\{\nu_j^k\}_{j \in J^k}$ (cf. (2.8)), we may estimate solutions to (5.1) via *aggregate primal solutions*

$$\tilde z^k := \sum_{j \in J^k} \nu_j^k z^j. \tag{5.5}$$

We now derive useful bounds on $\psi_0(\tilde z^k)$ and $\psi(\tilde z^k)$ as in [Kiw95a, Lem. 4.1].

LEMMA 5.1. *$\tilde z^k \in Z$, $\psi_0(\tilde z^k) \ge f_x^k - \alpha_k - \langle p^k, x^k\rangle$, $\psi(\tilde z^k) \ge p_f^k \ge p^k$.*

Proof. We have (cf. (2.8)) $\sum_{j\in J^k} \nu_j^k = 1$ with $\nu_j^k \ge 0$. Hence $\tilde z^k \in \mathrm{co}\{z^j\}_{j\in J^k} \subset Z$, $\psi_0(\tilde z^k) \ge \sum_j \nu_j^k \psi_0(z^j)$, $\psi(\tilde z^k) \ge \sum_j \nu_j^k \psi(z^j)$ by convexity of Z and concavity of ψ_0, ψ. Since (cf. (2.7)) $p_S^k \in \partial i_S(y^{k+1})$ with $S := \mathbb{R}^n_+$, we have $p_S^k \le 0$ and $\langle p_S^k, y^{k+1}\rangle = 0$, so (cf. (2.14)) $p_f^k = p^k - p_S^k \ge p^k$. Next, using (2.8) and (5.4) with $\psi(z^j) =: g^j$, we get $\sum_j \nu_j^k \psi(z^j) = \sum_j \nu_j^k g^j = p_f^k$ and

$$\begin{aligned}\check f_k(y^{k+1}) &= \sum\nolimits_j \nu_j^k f_j(y^{k+1}) = \sum\nolimits_j \nu_j^k \left[\psi_0(z^j) + \langle y^{k+1}, \psi(z^j)\rangle\right]\\ &= \sum\nolimits_j \nu_j^k \psi_0(z^j) + \langle y^{k+1}, p_f^k\rangle.\end{aligned}$$

Rearranging and using $\langle p_S^k, y^{k+1}\rangle = 0$, $p^k := p_f^k + p_S^k$ (cf. (2.14)), (2.12), and (2.13) gives

$$\sum\nolimits_j \nu_j^k \psi_0(z^j) = \check f_k(y^{k+1}) - \langle p_f^k + p_S^k, y^{k+1}\rangle = \bar f_S^k(0) = f_x^k - \alpha_k - \langle p^k, x^k\rangle.$$

Combining the preceding relations yields the conclusion. $\square$

The bounds of Lemma 5.1 are expressed in terms of the *primal-dual* optimality measure

$$\check V_k := \max\left\{ \max_{j=1:\,n} [-p_f^k]_j, \alpha_k + \langle p^k, x^k\rangle \right\} \tag{5.6}$$

as $\psi_0(\tilde z^k) \ge f_x^k - \check V_k$, $\min_{j=1}^n \psi_j(\tilde z^k) \ge -\check V_k$. Hence we may generate *record* measures $\check V_k^*$ and primal solutions $\tilde z_*^k$ as follows. At Step 0, set $\check V_1^* := \infty$. At Step 1, if $\check V_k < \check V_k^*$, set $\check V_k^* := \check V_k$, $\tilde z_*^k := \tilde z^k$. At Step 4 set $\check V_{k+1}^* := \check V_k^*$, $\tilde z_*^{k+1} := \tilde z_*^k$. In effect, $\check V_k^*$ (the current minimum of $\check V_j$ for $j \le k$) measures the quality of the primal iterate

$$\tilde z_*^k \in Z \quad \text{with} \quad \psi_0(\tilde z_*^k) \ge f_x^k - \check V_k^*, \quad \psi_j(\tilde z_*^k) \ge -\check V_k^*, \quad j = 1:n. \tag{5.7}$$

We now show that $\{\tilde z_*^k\}$ converges to the set of *ϵ-optimal primal solutions* of (5.1)

$$Z_\epsilon := \{\, z \in Z : \psi_0(z) \ge \psi_0^{\max} - \epsilon, \psi(z) \ge 0 \,\}. \tag{5.8}$$

THEOREM 5.2.

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

(ii) *$\lim_k f_x^k =: f_x^\infty \ge f_* - \epsilon$ and $\lim_k \check V_k^* \le 0$.*

(iii) *Let $\tilde z_*^\infty$ be a cluster point of $\{\tilde z_*^k\}$. Then $\tilde z_*^\infty \in Z_\epsilon$.*

(iv) *$d_{Z_\epsilon}(\tilde z_*^k) := \inf_{z\in Z_\epsilon} |\tilde z_*^k - z| \to 0$ as $k \to \infty$.*

Proof. (i) By (5.7), $\{\tilde z_*^k\}$ lies in the set Z, which is compact by our assumption.

(ii) By (2.5), $f_x^k \ge f(x^k) - \epsilon_f$ with $\epsilon_f := \epsilon$ gives $f_x^\infty \ge f_* - \epsilon$. Next, since $p_f^k \ge p^k$ (cf. Lemma 5.1) implies $\max_j [-p_f^k]_j \le |p^k|$, using (5.6) and (2.16) yields

$$\check V_k \le \max\{\, |p^k|, \alpha_k + \langle p^k, x^k\rangle \,\} \le \max\{\, |p^k|, \alpha_k \,\} + |p^k||x^k| \le V_k\left(1 + |x^k|\right); \tag{5.9}$$

hence by construction $\check V_k^* \le \min_{j=1}^k V_j'(1 + |x^j|)$. Recall that under our assumptions on (5.1), $\underline{\lim}_k V_k' = 0$ and $\{x^k\}$ is bounded. Therefore, $\lim_k \check V_k^* \le 0$ by monotonicity.

(iii) By (i), $\tilde z_*^\infty \in Z$. Using (ii) in (5.7) gives $\psi_0(\tilde z_*^\infty) \ge f_x^\infty$, $\psi(\tilde z_*^\infty) \ge 0$ by closedness of ψ_0, ψ. Since $f_x^\infty \ge f_* - \epsilon$ by (ii), where $f_* \ge \psi_0^{\max}$ by weak duality (cf. (1.1), (5.1), (5.2)), we have $\psi_0(\tilde z_*^\infty) \ge \psi_0^{\max} - \epsilon$. Thus $\tilde z_*^\infty \in Z_\epsilon$ by the definition (5.8).

(iv) This follows from (i), (iii), and the continuity of the distance function d_{Z_ϵ}. ∎

Remark 5.3.

(i) By the proofs of Lemma 2.3(iii) and Theorem 5.2, if an infinite loop between Steps 1 and 3 occurs, then $V_k \to 0$ yields $\max\{\breve{V}_k, 0\} \to 0$ and $d_{Z_\epsilon}(\tilde{z}^k) \to 0$. Similarly, if Step 2 terminates with $V_k = 0$, then $\breve{V}_k \leq 0$ and $\tilde{z}^k \in Z_\epsilon$.

(ii) Theorem 5.2 holds for $\{\tilde{z}^k_*\}$ replaced by $\{\tilde{z}^k\}_{k\in K}$ for any $K \subset \{1, 2, \dots\}$ such that $\lim_{k\in K} \max\{\breve{V}_k, 0\} = 0$; i.e., other selections could be considered.

(iii) Given a tolerance $\epsilon_{\mathrm{tol}} > 0$, the method may stop if

$$\psi_0(\tilde{z}^k) \geq f^k_x - \epsilon_{\mathrm{tol}} \quad \text{and} \quad \psi_j(\tilde{z}^k) \geq -\epsilon_{\mathrm{tol}}, \quad j = 1\colon n.$$

Then $\psi_0(\tilde{z}^k) \geq \psi_0^{\max} - \epsilon - \epsilon_{\mathrm{tol}}$ from $f^k_x \geq f_* - \epsilon$ (cf. (2.5)) and $f_* \geq \psi_0^{\max}$ (weak duality), so $\tilde{z}^k \in Z$ is an approximate solution of (5.1). This stopping criterion will be satisfied for some k (cf. (5.7) and Theorem 5.2(ii)).

No longer assuming coercivity of f_S, we still have the following.

THEOREM 5.4. *Theorem* 5.2 *holds if* $f_* > -\infty$ *and* $t_k \geq t_{\min} > 0$ *for all* k.

Proof. In view of the proof of Theorem 5.2, we need only to show that $\lim_k \breve{V}^*_k \leq 0$ when infinitely many descent steps occur (since otherwise $\{x^k\}$ is bounded, whereas $\underline{\lim}_k V'_k = 0$ by Lemma 3.11).

Let $K := \{k : i_k = 1\}$. Since $v_k \xrightarrow{K} 0$ (cf. Lemma 3.7(i)) with $v_k = t_k|p^k|^2 + \alpha_k$ (cf. (2.18)) and $v_k \geq |\alpha_k|$ at Step 4, we have $\alpha_k \xrightarrow{K} 0$ and $t_k|p^k|^2 \xrightarrow{K} 0$. By (2.18), $x^{k+1} - x^k = -i_k t_k p^k$, so

$$|x^{k+1}|^2 - |x^k|^2 = i_k t_k \left\{ t_k|p^k|^2 - 2\langle p^k, x^k\rangle \right\}.$$

Sum up and use the fact $\sum_k i_k t_k \geq \sum_{k\in K} t_{\min} = \infty$ to get

$$\overline{\lim_{k\in K}} \left\{ t_k|p^k|^2 - 2\langle p^k, x^k\rangle \right\} \geq 0$$

(since otherwise $|x^{k+1}|^2 \to -\infty$, which is impossible). Combining this with $t_k|p^k|^2 \xrightarrow{K} 0$ yields $\underline{\lim}_{k\in K}\langle p^k, x^k\rangle \leq 0$, as well as $|p^k|^2 \xrightarrow{K} 0$ by using the fact $t_k \geq t_{\min}$. Since also $\alpha_k \xrightarrow{K} 0$, we have $\underline{\lim}_{k\in K} \breve{V}_k \leq 0$ by (5.9). Then the fact $\breve{V}^*_k \leq \breve{V}_k$ implies $\lim_k \breve{V}^*_k \leq 0$. ∎

Remark 5.5.

(i) For Theorem 5.4, we may impose a lower bound $t_{\min} > 0$ on t_{k+1} at Step 6, whereas $f_* > -\infty$ if problem (5.1) is feasible (by weak duality). Thus, in contrast with [FeK00], [Kiw95a], our primal recovery works even if (5.1) has no Lagrange multipliers.

(ii) Remark 5.3 remains valid under the assumptions of Theorem 5.4.

In the remainder of this section we allow the primal problem (5.1) to be nonconvex. As before, our standing assumptions are that $\{\psi_j\}_{j=0}^n$ are finite and upper semicontinuous on the compact set Z, $\psi(z(\cdot))$ is locally bounded on S, and either f_S is coercive or $f_* > -\infty$ and $t_k \geq t_{\min} > 0$ as in Theorem 5.4 (cf. Remark 5.5(i)).

Since problem (5.1) may be nonconvex, consider its *relaxed convexified version*

(5.10)

$$\psi_0^{\mathrm{rel}} := \max_{(\nu_j, z^j)_{j=1}^M} \sum_{j=1}^M \nu_j \psi_0(z^j) \quad \text{s.t.} \quad \sum_{j=1}^M \nu_j \psi(z^j) \geq 0,\ \sum_{j=1}^M \nu_j = 1,\ z^j \in Z,\ \nu_j \geq 0,$$

where $M := n+1$. Both (5.1) and (5.10) have the same dual (1.1) with $f_* = \psi_0^{\rm rel} \geq \psi_0^{\max}$; see [FeK00], [LeR01], [MSW76]. Similarly to (5.8), let $\tilde{Z}_\epsilon$ denote the set of ϵ-optimal solutions of (5.10). Such solutions may be estimated by $(\nu_j^k, z^j)_{j\in\hat{J}^k}$ with $\hat{J}^k := \{j \in J^k : \nu_j^k \neq 0\}$ as follows. Since the QP routine of [Kiw94] delivers $|\hat{J}^k| \leq M$, whereas any (ν_j^k, z^j) can be split into two elements $(\nu_j^k/2, z^j)$, we may assume $|\hat{J}^k| = M$. Denoting $(\nu_j^k, z^j)_{j\in\hat{J}^k}$ as $(\hat{\nu}_j^k, \hat{z}^{jk})_{j=1}^M$, the proof of Lemma 5.1 yields

$$\sum_{j=1}^M \hat{\nu}_j^k \psi_0(\hat{z}^{jk}) = f_x^k - \alpha_k - \langle p^k, x^k\rangle \quad \text{and} \quad \sum_{j=1}^M \hat{\nu}_j^k \psi(\hat{z}^{jk}) = p_f^k \geq p^k. \tag{5.11}$$

The record solutions $(\tilde{\nu}_j^k, \tilde{z}^{jk})_{j=1}^M$ are generated just like $\tilde{z}_*^k$ by setting $(\tilde{\nu}_j^k, \tilde{z}^{jk})_{j=1}^M := (\hat{\nu}_j^k, \hat{z}^{jk})_{j=1}^M$ at Step 1 if $\check{V}_k < \check{V}_k^*$, and $(\tilde{\nu}_j^{k+1}, \tilde{z}^{j,k+1})_{j=1}^M := (\tilde{\nu}_j^k, \tilde{z}^{jk})_{j=1}^M$ at Step 4. We now show that $(\tilde{\nu}_j^k, \tilde{z}^{jk})_{j=1}^M$ converges to $\tilde{Z}_\epsilon$, thus extending [FeK00, Thm. 6.2].

THEOREM 5.6.

(i) $\{(\tilde{\nu}_j^k, \tilde{z}^{jk})_{j=1}^M\}$ *lies in a compact set.*

(ii) $\lim_k f_x^k =: f_x^\infty \geq f_* - \epsilon$ *and* $\lim_k \check{V}_k^* \leq 0$.

(iii) *Let* $(\tilde{\nu}_j, \tilde{z}^j)_{j=1}^M$ *be a cluster point of* $\{(\tilde{\nu}_j^k, \tilde{z}^{jk})_{j=1}^M\}$. *Then* $(\tilde{\nu}_j, \tilde{z}^j)_{j=1}^M \in \tilde{Z}_\epsilon$.

(iv) $d_{\tilde{Z}_\epsilon}((\tilde{\nu}_j^k, \tilde{z}^{jk})_{j=1}^M) \to 0$ *as* $k \to \infty$.

Proof. (i) By construction (cf. (2.8)), $\sum_j \tilde{\nu}_j^k = 1$, $\tilde{\nu}_j^k > 0$, $\tilde{z}^{jk} \in Z$, a compact set.

(ii) The proofs of Theorems 5.2(ii) and 5.4 remain valid.

(iii) By (i), $\sum_j \tilde{\nu}_j = 1$, $\tilde{\nu}_j \geq 0$, $\tilde{z}^j \in Z$, $j = 1\colon M$. Next, using (ii) with $\check{V}_k^* = \check{V}_k$ (cf. (5.6)) for k such that $(\hat{\nu}_j^k, \hat{z}^{jk}) = (\tilde{\nu}_j^k, \tilde{z}^{jk})$ in (5.11) and the upper semicontinuity of ψ_0, ψ gives

$$\sum_{j=1}^M \tilde{\nu}_j \psi_0(\tilde{z}^j) \geq f_x^\infty \geq f_* - \epsilon \quad \text{and} \quad \sum_{j=1}^M \tilde{\nu}_j \psi(\tilde{z}^j) \geq 0.$$

Since $(\tilde{\nu}_j, \tilde{z}^j)_{j=1}^M$ is feasible in (5.10) and $f_* \geq \psi_0^{\rm rel}$ by weak duality (cf. (1.1), (5.2), (5.10)), we have $\sum_{j=1}^M \tilde{\nu}_j \psi_0(\tilde{z}^j) \geq \psi_0^{\rm rel} - \epsilon$; i.e., $(\tilde{\nu}_j, \tilde{z}^j)_{j=1}^M$ is an ϵ-optimal solution of (5.10).

(iv) This follows from (i), (iii), and the continuity of $d_{\tilde{Z}_\epsilon}$. ∎

Extensions to separable problems are easily developed as in [FeK00, section 6].

Acknowledgments. I would like to thank the associate editor and the two anonymous referees for helpful comments.

REFERENCES

[Ber99] D. P. BERTSEKAS, *Nonlinear Programming*, 2nd ed., Athena Scientific, Belmont, MA, 1999.

[FeK00] S. FELTENMARK AND K. C. KIWIEL, *Dual applications of proximal bundle methods, including Lagrangian relaxation of nonconvex problems*, SIAM J. Optim., 10 (2000), pp. 697–721.

[HeK02] C. HELMBERG AND K. C. KIWIEL, *A spectral bundle method with bounds*, Math. Programming, 93 (2002), pp. 173–194.

[HeR00] C. HELMBERG AND F. RENDL, *A spectral bundle method for semidefinite programming*, SIAM J. Optim., 10 (2000), pp. 673–696.

[Hin01] M. HINTERMÜLLER, *A proximal bundle method based on approximate subgradients*, Comput. Optim. Appl., 20 (2001), pp. 245–266.

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

[Kiw85] K. C. Kiwiel, *An algorithm for nonsmooth convex minimization with errors*, Math. Comp., 45 (1985), pp. 173–180.

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

[Kiw94] K. C. Kiwiel, *A Cholesky dual method for proximal piecewise linear programming*, Numer. Math., 68 (1994), pp. 325–340.

[Kiw95a] K. C. Kiwiel, *Approximations in proximal bundle methods and decomposition of convex programs*, J. Optim. Theory Appl., 84 (1995), pp. 529–548.

[Kiw95b] K. C. Kiwiel, *Finding normal solutions in piecewise linear programming*, Appl. Math. Optim., 32 (1995), pp. 235–254.

[Kiw96] K. C. Kiwiel, *Restricted step and Levenberg–Marquardt techniques in proximal bundle methods for nonconvex nondifferentiable optimization*, SIAM J. Optim., 6 (1996), pp. 227–249.

[LeR01] C. Lemaréchal and A. Renaud, *A geometric study of duality gaps, with applications*, Math. Programming, 90 (2001), pp. 399–427.

[Mil01] S. A. Miller, *An Inexact Bundle Method for Solving Large Structured Linear Matrix Inequalities*, Ph.D. thesis, Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, 2001.

[MSW76] T. L. Magnanti, J. F. Shapiro, and M. H. Wagner, *Generalized linear programming solves the dual*, Management Sci., 22 (1976), pp. 1195–1203.

[ScZ92] H. Schramm and J. Zowe, *A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results*, SIAM J. Optim., 2 (1992), pp. 121–152.

[Sol03] M. V. Solodov, *On approximations with finite precision in bundle methods for nonsmooth optimization*, J. Optim. Theory Appl., 119 (2003), pp. 151–165.

[Tod01] M. J. Todd, *Semidefinite optimization*, Acta Numer., 10 (2001), pp. 515–560.