46/2006

Raport Badawczy

Research Report

RB/60/2006

A proximal-projection bundle method for Lagrangian relaxation, including semidefinite programming

K. C. Kiwiel

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 8373578

fax: (+48) (22) 8372772

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

Warszawa 2006

SIAM J. OPTIM.
Vol. 0, No. 0, pp. 000–000

A PROXIMAL-PROJECTION BUNDLE METHOD FOR LAGRANGIAN RELAXATION, INCLUDING SEMIDEFINITE PROGRAMMING*

KRZYSZTOF C. KIWIEL†

Abstract. We give a proximal bundle method for minimizing a convex function f over a convex set C. It requires evaluating f and its subgradients with a fixed but possibly unknown accuracy $\epsilon > 0$. Each iteration involves solving an unconstrained proximal subproblem and projecting a certain point onto C. The method asymptotically finds points that are ϵ-optimal. In Lagrangian relaxation of convex programs, it allows for ϵ-accurate solutions of Lagrangian subproblems and finds ϵ-optimal primal solutions. For semidefinite programming problems, it extends the highly successful spectral bundle method to the case of inexact eigenvalue computations.

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

AMS subject classifications. 65K05, 90C25

DOI. 10.1137/050639284

1. Introduction.

We consider the convex constrained minimization problem

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

where C 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 a fixed *accuracy tolerance* $\epsilon_f \geq 0$, for each $u \in C$ we can find an *approximate value* f_u and an *approximate subgradient* g_u of f that produce the *approximate linearization* of f:

$$\bar{f}_u(\cdot) := f_u + \langle g_u, \cdot - u \rangle \leq f(\cdot) \quad \text{with} \quad \bar{f}_u(u) = f_u \geq f(u) - \epsilon_f. \tag{1.2}$$

Thus $f_u \in [f(u) - \epsilon_f, f(u)]$ estimates $f(u)$, while $g_u \in \partial_{\epsilon_f} f(u)$; i.e., g_u is a member of the ϵ_f-subdifferential $\partial_{\epsilon_f} f(u) := \{g : f(\cdot) \geq f(u) - \epsilon_f + \langle g, \cdot - u \rangle\}$ of f at u.

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

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

where each $F_z : \mathbb{R}^n \to \mathbb{R}$ is convex and Z is an infinite set, then it may be impossible to compute $f(u)$. However, if for some fixed (and possibly *unknown*) tolerance ϵ_f we can find an ϵ_f-maximizer of (1.3), i.e., an element $z_u \in Z$ satisfying $F_{z_u}(u) \geq f(u) - \epsilon_f$, then we may set $f_u := F_{z_u}(u)$ and take g_u as any subgradient of F_{z_u} at u to satisfy (1.2). An important special case arises in *Lagrangian relaxation* [HUL93, Chap. XII], [Lem01], where problem (1.1) with $C := \mathbb{R}^n_+$ is the Lagrangian dual of the primal problem

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

*Received by the editors August 30, 2005; accepted for publication (in revised form) July 20, 2006; published electronically DATE.
http://www.siam.org/journals/siopt/x-x/63928.html

†Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01–447 Warsaw, Poland (kiwiel@ibspan.waw.pl).

with $F_z(u) := \psi_0(z) + \langle u, \psi(z)\rangle$ for $\psi := (\psi_1, \dots, \psi_n)$. Then, for each multiplier $u \geq 0$, we need only find $z_u \in Z$ such that $f_u := F_{z_u}(y) \geq f(u) - \epsilon_f$ in (1.3) to use $g_u := \psi(z_u)$. For instance, if (1.4) is a *semidefinite program* (SDP) with each ψ_i affine and Z the set of symmetric positive semidefinite matrices of order m with a bounded trace, then $f(u)$ is the maximum eigenvalue of a symmetric matrix $M(u)$ depending affinely on u [Tod01, sect. 6.3], and z_u can be found by computing an approximate eigenvector corresponding to the maximum eigenvalue of $M(u)$ via the Lanczos method [HeK02, HeR00, Nay05].

The recent paper [Kiw06b] extended the proximal bundle methods of [Kiw90] and [HUL93, sect. XV.3] to the inexact setting of (1.2) (see [Hin01, Kiw85, Kiw95, Mil01, Sol03] for earlier related developments, and [Kiw05] for numerical tests). Such methods at each iteration find a trial point that minimizes over C a polyhedral model of f built from accumulated linearizations, stabilized by a quadratic *prox* term centered at a point which is usually the best iterate found so far. Solving this subproblem can require much work for large n even when the set C is polyhedral, including the simplest case of $C = \mathbb{R}^n_+$ used in Lagrangian relaxation.

This paper extends the projection-proximal method of [Kiw99] to the case of inexact linearizations. For this method, we may regard (1.1) as an unconstrained problem $f_* = \inf f_C$ with the *essential objective*

$$f_C := f + i_C, \tag{1.5}$$

where i_C is the *indicator* function of C ($i_C(u) = 0$ if $u \in C$, ∞ otherwise). In its simplest form, the method generates the trial point in two steps. The first *proximal* step minimizes a polyhedral model $\check{f}$ of f, augmented with a quadratic proximal term and a linearization of i_C obtained at the previous iteration, to produce a linearization of $\check{f}$. The second *projection* step minimizes over C this linearization augmented with the proximal term; this amounts to projecting a certain point onto C to produce the trial point and the next linearization of i_C. Thus the standard bundle subproblem is replaced by two subproblems, where the first "unconstrained" subproblem is much easier to solve, and the projection is straightforward if the set C is "simple." Our development is related to the *alternating linearization* approach of [KRR99], in which the prox subproblem for the sum of two functions, such as (1.5), is approximated by two subproblems in which the functions are alternately represented by linear models.

Our extension of [Kiw99] is natural and simple: the original method is run as if the objective linearizations were exact until a test on predicted descent discovers their inaccuracy; then the proximity weight is decreased to produce descent or confirm that the current prox center is ϵ_f-optimal. We show that our method asymptotically estimates the optimal value f_* of (1.1) with accuracy ϵ_f and finds ϵ_f-optimal points. In Lagrangian relaxation, under standard convexity and compactness assumptions on problem (1.4) (see section 5), it finds ϵ_f-optimal primal solutions by combining partial Lagrangian solutions, even when Lagrange multipliers don't exist. These features are essentially "inherited" from the inexact framework of [Kiw06b] (although some technical developments are nontrivial). On the other hand, this paper reorganizes and simplifies the convergence framework of [Kiw06b] and sheds light on several important issues not discussed in there (such as the "true" impact of inexact evaluations, the possible use of "more inexact" null steps, primal recovery for Lagrangian relaxation with subgradient aggregation, and Lagrangian relaxation of equality constraints).

For the important special case where the functions ψ_i of the primal problem (1.4) are affine, we show how to employ *nonpolyhedral* models of f. Each model has the form $\check{f}(\cdot) := \sup_{z\in\check{Z}} F_z(\cdot)$ stemming from (1.3), where $\check{Z}$ is a closed convex subset of Z. Then the proximal step can be implemented by solving a dual subproblem of minimizing a convex quadratic function over $\check{Z}$ (e.g., via interior-point methods when $\check{Z}$ is simple enough), and the projection on $C := \mathbb{R}^n_+$ is trivial. Further, the dual subproblem solutions estimate ϵ_f-optimal primal solutions asymptotically as above. In particular, our framework extends the highly successful methods of [FGRS06, sect. 3.2] and [ReS06, sect. 3] (see Remark 5.6).

Finally, for SDP (see below (1.4)) our general framework yields extensions of several variants of the spectral bundle method [Hel03, Hel04, HeK02, HeR00, Nay99]. This method employs the nonpolyhedral models discussed above, with $\check{Z}$ constructed from accumulated eigenvectors of the dual objective matrix $M(u)$. The original version of [HeR00] could handle only equality-constrained SDPs. Its extension [HeK02] to inequality-constrained SDPs can be seen as a specialization of the method of [Kiw99]; this helps in distinguishing its "driving force" from "implementation details" (although the latter are, of course, crucial for its performance in practice). Hence the primal recovery result of [Hel04, Thm. 3.6] also follows from our more general results (see Theorems 3.7 and 5.2); in fact, we don't need the assumption of [Hel04, Thm. 3.6] that the dual problem has a solution (see Remark 5.7(i)). Our extension to the case of approximate eigenvectors (see below (1.4)) is relevant for both theory and practice. Namely, while the existing version [HeK02] already employs approximate eigenvectors at so-called null steps (and this saves much work in practice [Hel03, HeK02, Nay99, Nay05]), it requires exact eigenvalues at the remaining descent steps. Our theoretical results show what to expect if approximate eigenvectors are used at descent steps as well, thus opening room for more efficient implementations.

The paper is organized as follows. In section 2 we present our method for general objective models. Its convergence is analyzed in section 3. Various modifications and model choices are given in section 4. Applications to Lagrangian relaxation are studied in section 5.

Our notation is fairly standard. $P_C(u) := \arg\min_C |\cdot - u|$ is the *projector* onto C.

2. The proximal-projection bundle method. Our method generates a sequence of *trial points* $\{u^k\}_{k=1}^\infty \subset C$ for evaluating the approximate values $f_u^k := f_{u^k}$, subgradients $g^k := g_{u^k}$, and linearizations $f_k := \bar{f}_{u^k}$ such that

$$f_k(\cdot) = f_u^k + \langle g^k, \cdot - u^k \rangle \le f(\cdot) \quad \text{with} \quad f_k(u^k) = f_u^k \ge f(u^k) - \epsilon_f, \tag{2.1}$$

as stipulated in (1.2). At iteration k, the current *prox* (or *stability*) *center* $\hat{u}^k := u^{k(l)} \in C$ for some $k(l) \le k$ has the value $f_{\hat{u}}^k := f_u^{k(l)}$ (usually $f_{\hat{u}}^k = \min_{j=1}^k f_u^j$); note that, by (2.1),

$$f_{\hat{u}}^k \in [f(\hat{u}^k) - \epsilon_f, f(\hat{u}^k)]. \tag{2.2}$$

For a model $\check{f}_k \le f$, the next point u^{k+1} approximately solves the prox subproblem

$$\min\ \check{f}_k(\cdot) + i_C(\cdot) + \frac{1}{2t_k}|\cdot - \hat{u}^k|^2, \tag{2.3}$$

where $t_k > 0$ is a *stepsize* that controls the size of $|u^{k+1} - \hat{u}^k|$. To this end, two partial linearizations of (2.3) are employed. First, replacing i_C by its past linearization $\bar{\imath}_C^{k-1} \le i_C$ in (2.3), we find its solution $\check{u}^{k+1}$ and a linearization $\bar{f}_k \le \check{f}_k$ such that $\check{u}^{k+1}$ solves (2.3) with $\check{f}_k$, i_C replaced by $\bar{f}_k$, $\bar{\imath}_C^{k-1}$. Next, replacing $\check{f}_k$ by $\bar{f}_k$ in (2.3), we find its solution u^{k+1} and a linearization $\bar{\imath}_C^k \le i_C$ such that u^{k+1} solves (2.3) with $\check{f}_k$, i_C replaced by $\bar{f}_k$, $\bar{\imath}_C^k$. Due to evaluation errors, we may have $f_{\hat{u}}^k < \check{f}_k(\hat{u}^k)$, in which case the *predicted descent* $v_k := f_{\hat{u}}^k - \bar{f}_k(u^{k+1})$ may be nonpositive; then t_k is increased and u^{k+1} is recomputed to decrease $\bar{f}_k(u^{k+1})$ until $v_k > 0$. A *descent* step to $\hat{u}^{k+1} := u^{k+1}$ is taken if $f_u^{k+1} \le f_{\hat{u}}^k - \kappa v_k$ for a fixed $\kappa \in (0,1)$. Otherwise, a *null* step $\hat{u}^{k+1} := \hat{u}^k$ occurs; then $\bar{f}_k$ and the new linearization f_{k+1} are used to produce a better model $\check{f}_{k+1} \ge \max\{\bar{f}_k, f_{k+1}\}$ (e.g., $\check{f}_{k+1} = \max\{\bar{f}_k, f_{k+1}\}$).

Specific rules of our method will be discussed after its formal statement below.

ALGORITHM 2.1.

Step 0 (initialization). Select $u^1 \in C$, a *descent parameter* $\kappa \in (0,1)$, a *stepsize bound* $t_{\min} > 0$, and a *stepsize* $t_1 \ge t_{\min}$. Set $\bar{f}_0 := f_1$ (cf. (2.1)), $\bar{\imath}_C^0 := \langle p_C^0, \cdot - u^1 \rangle$ with $p_C^0 := 0$, $\hat{u}^1 := u^1$, $f_{\hat{u}}^1 := f_u^1 := f_{u^1}$, $g^1 := g_{u^1}$ (cf. (2.1)), $i_t^1 := 0$, $k := k(0) := 1$, $l := 0$ ($k(l) - 1$ will denote the iteration of the lth descent step).

Step 1 (model selection). Choose $\check{f}_k : \mathbb{R}^n \to \mathbb{R}$ closed convex and such that

$$\max\{\bar{f}_{k-1}, f_k\} \le \check{f}_k \le f_C. \tag{2.4}$$

Step 2 (proximal point finding). Set

$$\check{u}^{k+1} := \arg\min \left\{ \phi_f^k(\cdot) := \check{f}_k(\cdot) + \bar{\imath}_C^{k-1}(\cdot) + \frac{1}{2t_k}|\cdot - \hat{u}^k|^2 \right\}, \tag{2.5}$$

$$\bar{f}_k(\cdot) := \check{f}_k(\check{u}^{k+1}) + \langle p_f^k, \cdot - \check{u}^{k+1} \rangle \quad \text{with} \quad p_f^k := \frac{\hat{u}^k - \check{u}^{k+1}}{t_k} - p_C^{k-1}. \tag{2.6}$$

Step 3 (projection). Set

$$u^{k+1} := \arg\min \left\{ \phi_C^k(\cdot) := \bar{f}_k(\cdot) + i_C(\cdot) + \frac{1}{2t_k}|\cdot - \hat{u}^k|^2 \right\} = P_C(\hat{u}^k - t_k p_f^k), \tag{2.7}$$

$$\bar{\imath}_C^k(\cdot) := \langle p_C^k, \cdot - u^{k+1} \rangle \quad \text{with} \quad p_C^k := \frac{\hat{u}^k - u^{k+1}}{t_k} - p_f^k, \tag{2.8}$$

$$v_k := f_{\hat{u}}^k - \bar{f}_k(u^{k+1}), \quad p^k := \frac{\hat{u}^k - u^{k+1}}{t_k}, \quad \text{and} \quad \epsilon_k := v_k - t_k|p^k|^2. \tag{2.9}$$

Step 4 (stopping criterion). If $\max\{|p^k|, \epsilon_k\} = 0$, stop ($f_{\hat{u}}^k \le f_*$).

Step 5 (stepsize correction). If $v_k < -\epsilon_k$, set $t_k := 10t_k$, $i_t^k := k$, and go back to Step 2.

Step 6 (descent test). Evaluate f_u^{k+1} and g^{k+1} (cf. (2.1)). If the *descent test* holds,

$$f_u^{k+1} \le f_{\hat{u}}^k - \kappa v_k, \tag{2.10}$$

set $\hat{u}^{k+1} := u^{k+1}$, $f_{\hat{u}}^{k+1} := f_u^{k+1}$, $i_t^{k+1} := 0$, $k(l+1) := k+1$, and increase l by 1 (*descent step*); otherwise, set $\hat{u}^{k+1} := \hat{u}^k$, $f_{\hat{u}}^{k+1} := f_{\hat{u}}^k$, and $i_t^{k+1} := i_t^k$ (*null step*).

Step 7 (stepsize updating). If $k(l) = k+1$ (i.e., after a descent step), select $t_{k+1} \ge t_k$; otherwise, either set $t_{k+1} := t_k$ or choose $t_{k+1} \in [t_{\min}, t_k]$ if $i_t^{k+1} = 0$.

Step 8 (loop). Increase k by 1 and go to Step 1.

Several comments on the method are in order. Step 1 may choose the simplest model $\check{f}_k = \max\{\bar{f}_{k-1}, f_k\}$; more efficient choices are given in section 4.4. For a polyhedral model $\check{f}_k$, subproblem (2.5) can be handled via simple QP solvers [Kiw86]; in contrast, the more difficult subproblem (2.3) employed in [Kiw06b] requires more sophisticated solvers even for a polyhedral set C [Kiw94]. The projection of (2.7) is easily found if the set C is "simple" (e.g., the Cartesian product of boxes, simplices, and ellipsoids).

We now use the relations of Steps 2 and 3 to derive an optimality estimate, which involves the *aggregate linearization* $\bar{f}_C^k := \bar{f}_k + \bar{\imath}_C^k$ and the *optimality measure*

$$V_k := \max\{|p^k|, \epsilon_k + \langle p^k, \hat{u}^k\rangle\}. \tag{2.11}$$

LEMMA 2.2. (i) *The vectors p_f^k and p_C^k defined in* (2.6) *and* (2.8) *are in fact subgradients,*

$$p_f^k \in \partial \check{f}_k(\check{u}^{k+1}) \quad and \quad p_C^k \in \partial i_C(u^{k+1}), \tag{2.12}$$

and the linearizations $\bar{f}_k$ and $\bar{\imath}_C^k$ defined in (2.6) *and* (2.8) *provide the minorizations*

$$\bar{f}_k \le \check{f}_k, \quad \bar{\imath}_C^k \le i_C, \quad and \quad \bar{f}_C^k := \bar{f}_k + \bar{\imath}_C^k \le f_C. \tag{2.13}$$

(ii) *The* aggregate subgradient p^k *defined in* (2.9) *and the linearization $\bar{f}_C^k$ above satisfy*

$$p^k = p_f^k + p_C^k = \frac{\hat{u}^k - u^{k+1}}{t_k}, \tag{2.14}$$

$$\bar{f}_C^k(\cdot) = \bar{f}_k(u^{k+1}) + \langle p^k, \cdot - u^{k+1}\rangle. \tag{2.15}$$

(iii) *The predicted descent v_k and the* aggregate linearization error ϵ_k *of* (2.9) *satisfy*

$$v_k = t_k|p^k|^2 + \epsilon_k \quad and \quad \epsilon_k = f_{\hat{u}}^k - \bar{f}_C^k(\hat{u}^k). \tag{2.16}$$

(iv) *The aggregate linearization $\bar{f}_C^k$ is expressed in terms of p^k and ϵ_k as follows:*

$$f_{\hat{u}}^k - \epsilon_k + \langle p^k, \cdot - \hat{u}^k\rangle = \bar{f}_C^k(\cdot) \le f_C(\cdot). \tag{2.17}$$

(v) *The optimality measure V_k of* (2.11) *satisfies* $V_k \le \max\{|p^k|, \epsilon_k\}(1 + |\hat{u}^k|)$ *and*

$$f_{\hat{u}}^k \le f_C(u) + V_k\big(1 + |u|\big) \quad for\ all\ u. \tag{2.18}$$

(vi) *We have $v_k \ge -\epsilon_k \Leftrightarrow t_k|p^k|^2/2 \ge -\epsilon_k \Leftrightarrow v_k \ge t_k|p^k|^2/2$. Moreover, $v_k \ge \epsilon_k$, $-\epsilon_k \le \epsilon_f$, and*

$$v_k \ge \max\left\{\frac{t_k|p^k|^2}{2}, |\epsilon_k|\right\} \qquad if \quad v_k \ge -\epsilon_k, \tag{2.19}$$

$$V_k \le \max\left\{\left(\frac{2v_k}{t_k}\right)^{1/2}, v_k\right\}(1 + |\hat{u}^k|) \quad if \quad v_k \ge -\epsilon_k, \tag{2.20}$$

$$V_k < \left(\frac{2\epsilon_f}{t_k}\right)^{1/2}(1 + |\hat{u}^k|) \qquad if \quad v_k < -\epsilon_k. \tag{2.21}$$

Proof. (i) By (2.5)–(2.6), the optimality condition (using $\nabla \bar{\imath}_C^{k-1} = p_C^{k-1}$; cf. (2.8))

$$0 \in \partial\phi_f^k(\check{u}^{k+1}) = \partial\check{f}_k(\check{u}^{k+1}) + p_C^{k-1} + \frac{\check{u}^{k+1} - \hat{u}^k}{t_k} = \partial\check{f}_k(\check{u}^{k+1}) - p_f^k$$

and the equality $\bar{f}_k(\check{u}^{k+1}) = \check{f}_k(\check{u}^{k+1})$ yield $p_f^k \in \partial\check{f}_k(\check{u}^{k+1})$ and $\bar{f}_k \le \check{f}_k$. By (2.7)–(2.8),

$$0 \in \partial\phi_C^k(u^{k+1}) = p_f^k + \partial i_C(u^{k+1}) + \frac{u^{k+1} - \hat{u}^k}{t_k} = \partial i_C(u^{k+1}) - p_C^k$$

(using $\nabla \bar{f}_k = p_f^k$) and $\bar{\imath}_C^k(u^{k+1}) = i_C(u^{k+1}) = 0$ give $p_C^k \in \partial i_C(u^{k+1})$ and $\bar{\imath}_C^k \le i_C$. Combining both minorizations, we obtain that $\bar{f}_k + \bar{\imath}_C^k \le \check{f}_k + i_C \le f_C$ by (2.4) and (1.5).

(ii) Use the linearity of $\bar{f}_C^k := \bar{f}_k + \bar{\imath}_C^k$, (2.6), (2.8) with $\bar{\imath}_C^k(u^{k+1}) = 0$, and (2.9).

(iii) Rewrite (2.9), using the fact that $\bar{f}_C^k(\hat{u}^k) = \bar{f}_k(u^{k+1}) + t_k|p^k|^2$, by (ii).

(iv) We have $f_{\hat{u}}^k - \epsilon_k = \bar{f}_C^k(\hat{u}^k)$ by (iii), and $\bar{f}_C^k$ is affine by (ii) and minorizes f_C by (i).

(v) Use the Cauchy–Schwarz inequality in the definition (2.11) and in (iv).

(vi) The equivalences follow from the expression of $v_k = t_k|p^k|^2 + \epsilon_k$ in (iii); in particular, $v_k \ge \epsilon_k$. Next, by (2.16), (2.13), and (2.2) with $f_C(\hat{u}^k) = f(\hat{u}^k)$ ($\hat{u}^k \in C$), we have

$$-\epsilon_k = \bar{f}_C^k(\hat{u}^k) - f_{\hat{u}}^k \le f_C(\hat{u}^k) - f_{\hat{u}}^k = f(\hat{u}^k) - f_{\hat{u}}^k \le \epsilon_f.$$

Finally, to obtain the bounds (2.19)–(2.21), use the equivalences together with the facts that $v_k \ge \epsilon_k$, $-\epsilon_k \le \epsilon_f$ and the bound on V_k from assertion (v). □

The optimality estimate (2.18) justifies the stopping criterion of Step 4: $V_k = 0$ yields $f_{\hat{u}}^k \le \inf f_C = f_*$; thus, the point $\hat{u}^k$ is ϵ_f*-optimal*; i.e., $f(\hat{u}^k) \le f_* + \epsilon_f$ by (2.2). In the case of exact evaluations ($\epsilon_f = 0$), we have $v_k \ge \epsilon_k \ge 0$ by Lemma 2.2(vi), Step 5 is redundant, and Algorithm 2.1 becomes essentially that of [Kiw99, Alg. 3.1]. When inexactness is discovered via $v_k < -\epsilon_k$, the stepsize t_k is increased to produce descent or confirm that $\hat{u}^k$ is ϵ_f-optimal. Namely, when $\hat{u}^k$ is bounded in (2.21), increasing t_k drives V_k to 0, so that $f_{\hat{u}}^k \le f_*$ asymptotically. Whenever t_k is increased at Step 5, the *stepsize indicator* $i_t^k \ne 0$ prevents Step 7 from decreasing t_k after null steps until the next descent step occnrs (cf. Step 6). Otherwise, decreasing t_k at Step 7 aims at collecting more local information about f at null steps.

We now show that an infinite cycle between Steps 2 and 5 means that $\hat{u}^k$ is ϵ_f-optimal.

LEMMA 2.3. *If an infinite cycle between Steps* 2 *and* 5 *occurs, then* $f_{\hat{u}}^k \le f_*$ *and* $V_k \to 0$.

Proof. At Step 5 during the cycle the facts that $V_k < (2\epsilon_f/t_k)^{1/2}(1 + |\hat{u}^k|)$ by (2.21) and $t_k \uparrow \infty$ as the cycle continues give $V_k \to 0$, so that $f_{\hat{u}}^k \le \inf f_C = f_*$ by (2.18). □

3. Convergence. In view of Lemma 2.3, we may suppose that the algorithm neither terminates nor cycles infinitely between Steps 2 and 5 (otherwise $\hat{u}^k$ is ϵ_f-optimal). At Step 6, we have $u^{k+1} \in C$ and $v_k > 0$ (by (2.19), since $\max\{|p^k|, \epsilon_k\} > 0$ at Step 4), so that $\hat{u}^{k+1} \in C$ and $f_{\hat{u}}^{k+1} \le f_{\hat{u}}^k$ for all k. We shall show that the asymptotic value $f_{\hat{u}}^\infty := \lim_k f_{\hat{u}}^k$ satisfies $f_{\hat{u}}^\infty \le f_*$. As in [Kiw99, sect. 4], we assume that the model subgradients $p_f^k \in \partial\check{f}_k(\check{u}^{k+1})$ in (2.12) satisfy

$$\{p_f^k\} \text{ is bounded if } \{u^k\} \text{ is bounded.} \tag{3.1}$$

It will be seen in Remark 4.4 that typical models $\check f_k$ satisfy this condition automatically.

We first consider the case where only finitely many descent steps occur. After the last descent step, only null steps occur, and the sequence $\{t_k\}$ eventually becomes monotone, since once Step 5 increases t_k, Step 7 can't decrease t_k; thus the limit $t_\infty := \lim_k t_k$ exists. We deal with the cases of $t_\infty = \infty$ in Lemma 3.1 and $t_\infty < \infty$ in Lemma 3.2 below.

LEMMA 3.1. *Suppose there exists $\bar k$ such that only null steps occur for all $k \ge \bar k$, and $t_\infty := \lim_k t_k = \infty$. Let $K := \{k \ge \bar k : t_{k+1} > t_k\}$. Then $V_k \xrightarrow{K} 0$ at Step 5.*

Proof. At iteration $k \in K$, before Step 5 increases t_k for the last time, we have $V_k < (2\epsilon_f/t_k)^{1/2}(1+|\hat u^{\bar k}|)$ by (2.21); consequently, $t_k \to \infty$ gives $V_k \xrightarrow{K} 0$. □

LEMMA 3.2. *Suppose there exists $\bar k$ such that, for all $k \ge \bar k$, only null steps occur and Step 5 doesn't increase t_k. Then $V_k \to 0$.*

Proof. First, using partial linearizations of subproblems (2.5) and (2.7), we show that their optimal values $\phi_f^k(\check u^{k+1}) \le \phi_C^k(u^{k+1})$ are nondecreasing and bounded above.

Fix $k \ge \bar k$. By the definitions in (2.5)–(2.6), we have $\bar f_k(\check u^{k+1}) = \check f_k(\check u^{k+1})$ and

$$\check u^{k+1} = \arg\min\left\{ \bar\phi_f^k(\cdot) := \bar f_k(\cdot) + \bar\imath_C^{k-1}(\cdot) + \tfrac{1}{2t_k}|\cdot - \hat u^k|^2 \right\} \tag{3.2}$$

from $\nabla\bar\phi_f^k(\check u^{k+1}) = 0$. Since $\bar\phi_f^k$ is quadratic and $\bar\phi_f^k(\check u^{k+1}) = \phi_f^k(\check u^{k+1})$, by Taylor's expansion

$$\bar\phi_f^k(\cdot) = \phi_f^k(\check u^{k+1}) + \tfrac{1}{2t_k}|\cdot - \check u^{k+1}|^2. \tag{3.3}$$

Similarly, by the definitions in (2.7)–(2.8), we have $\bar\imath_C^k(u^{k+1}) = i_C(u^{k+1}) = 0$,

$$u^{k+1} = \arg\min\left\{ \bar\phi_C^k(\cdot) := \bar f_k(\cdot) + \bar\imath_C^k(\cdot) + \tfrac{1}{2t_k}|\cdot - \hat u^k|^2 \right\}, \tag{3.4}$$

$$\bar\phi_C^k(\cdot) = \phi_C^k(u^{k+1}) + \tfrac{1}{2t_k}|\cdot - u^{k+1}|^2. \tag{3.5}$$

Next, to bound the objective values of the linearized subproblems (3.2) and (3.4) from above, we use the minorizations $\bar f_k \le f_C$ and $\bar\imath_C^{k-1}, \bar\imath_C^k \le i_C$ of (2.13) with $\hat u^k \in C$:

$$\phi_f^k(\check u^{k+1}) + \tfrac{1}{2t_k}|\check u^{k+1} - \hat u^k|^2 = \bar\phi_f^k(\hat u^k) \le f(\hat u^k), \tag{3.6a}$$

$$\phi_C^k(u^{k+1}) + \tfrac{1}{2t_k}|u^{k+1} - \hat u^k|^2 = \bar\phi_C^k(\hat u^k) \le f(\hat u^k), \tag{3.6b}$$

where the equalities stem from (3.3) and (3.5). Due to the minorization $\bar\imath_C^{k-1} \le i_C$, the objectives of subproblems (3.2) and (2.7) satisfy $\bar\phi_f^k \le \phi_C^k$. On the other hand, since $\hat u^{k+1} = \hat u^k$, $t_{k+1} \le t_k$ (cf. Step 7), and $\bar f_k \le \check f_{k+1}$ by (2.4), the objectives of (3.4) and the next subproblem (2.5) satisfy $\bar\phi_C^k \le \phi_f^{k+1}$. Altogether, by (3.3) and (3.5), we see that

$$\phi_f^k(\check u^{k+1}) + \tfrac{1}{2t_k}|u^{k+1} - \check u^{k+1}|^2 = \bar\phi_f^k(u^{k+1}) \le \phi_C^k(u^{k+1}), \tag{3.7a}$$

$$\phi_C^k(u^{k+1}) + \tfrac{1}{2t_k}|\check u^{k+2} - u^{k+1}|^2 = \bar\phi_C^k(\check u^{k+2}) \le \phi_f^{k+1}(\check u^{k+2}). \tag{3.7b}$$

In particular, the inequalities $\phi_f^k(\check u^{k+1}) \le \phi_C^k(u^{k+1}) \le \phi_f^{k+1}(\check u^{k+2})$ imply that the nondecreasing sequences $\{\phi_f^k(\check u^{k+1})\}_{k\ge\bar k}$ and $\{\phi_C^k(u^{k+1})\}_{k\ge\bar k}$, which are bounded above by (3.6) with $\hat{u}^k = \hat{u}^{\bar{k}}$ for all $k \geq \bar{k}$, must have a common limit, say $\phi_\infty \leq f(\hat{u}^{\bar{k}})$. Moreover, since the stepsizes satisfy $t_k \leq t_{\bar{k}}$ for all $k \geq \bar{k}$, we deduce from the bounds (3.6)–(3.7) that

$$\phi_f^k(\check{u}^{k+1}),\ \phi_C^k(u^{k+1}) \uparrow \phi_\infty, \qquad \check{u}^{k+2} - u^{k+1} \to 0, \tag{3.8}$$

and the sequences $\{\check{u}^{k+1}\}$ and $\{u^{k+1}\}$ are bounded. Then the sequence $\{p_f^k\}$ is bounded by (3.1), and the sequence $\{g^k\}$ is bounded as well, since $g^k \in \partial_{\epsilon_f} f(u^k)$ by (2.1), whereas the mapping $\partial_{\epsilon_f} f$ is locally bounded [HUL93, sect. XI.4.1].

We now show that the *approximation error* $\check{\epsilon}_k := f_u^{k+1} - \bar{f}_k(u^{k+1})$ vanishes. Using the form (2.1) of f_{k+1}, the minorization $f_{k+1} \leq \check{f}_{k+1}$ of (2.4), the Cauchy–Schwarz inequality, and the optimal values of subproblems (2.5) and (2.7) with $\hat{u}^k = \hat{u}^{\bar{k}}$ for $k \geq \bar{k}$, we estimate

$$\begin{aligned}
\check{\epsilon}_k &:= f_u^{k+1} - \bar{f}_k(u^{k+1}) = f_{k+1}(\check{u}^{k+2}) - \bar{f}_k(u^{k+1}) + \langle g^{k+1}, u^{k+1} - \check{u}^{k+2}\rangle \\
&\leq \check{f}_{k+1}(\check{u}^{k+2}) - \bar{f}_k(u^{k+1}) + |g^{k+1}||u^{k+1} - \check{u}^{k+2}| \\
&= \phi_f^{k+1}(\check{u}^{k+2}) - \phi_C^k(u^{k+1}) + \Delta_k - \bar{\imath}_C^k(\check{u}^{k+2}) + |g^{k+1}||u^{k+1} - \check{u}^{k+2}|,
\end{aligned} \tag{3.9}$$

where $\Delta_k := |u^{k+1} - \hat{u}^{\bar{k}}|^2/2t_k - |\check{u}^{k+2} - \hat{u}^{\bar{k}}|^2/2t_{k+1}$. To see that $\Delta_k \to 0$, note that

$$|\check{u}^{k+2} - \hat{u}^{\bar{k}}|^2 = |u^{k+1} - \hat{u}^{\bar{k}}|^2 + 2\langle \check{u}^{k+2} - u^{k+1}, u^{k+1} - \hat{u}^{\bar{k}}\rangle + |\check{u}^{k+2} - u^{k+1}|^2,$$

$|u^{k+1} - \hat{u}^{\bar{k}}|^2$ is bounded, $\check{u}^{k+2} - u^{k+1} \to 0$ by (3.8), and $t_{\min} \leq t_{k+1} \leq t_k$ for $k \geq \bar{k}$ by Step 7. These properties also give $\bar{\imath}_C^k(\check{u}^{k+2}) \to 0$, since by (2.8) and the Cauchy–Schwarz inequality, we have

$$|\bar{\imath}_C^k(\check{u}^{k+2})| \leq |p_C^k||\check{u}^{k+2} - u^{k+1}| \quad \text{with} \quad |p_C^k| \leq |u^{k+1} - \hat{u}^{\bar{k}}|/t_k + |p_f^k|,$$

where $\{p_f^k\}$ is bounded. Hence, using (3.8) and the boundedness of $\{g^{k+1}\}$ in (3.9) yields $\overline{\lim}_k \check{\epsilon}_k \leq 0$. On the other hand, for $k \geq \bar{k}$ the null step condition $f_u^{k+1} > f_{\hat{u}}^k - \kappa v_k$ gives

$$\check{\epsilon}_k = \left[f_u^{k+1} - f_{\hat{u}}^k\right] + \left[f_{\hat{u}}^k - \bar{f}_k(u^{k+1})\right] > -\kappa v_k + v_k = (1-\kappa)v_k \geq 0,$$

where $\kappa < 1$ by Step 0; we conclude that $\check{\epsilon}_k \to 0$ and $v_k \to 0$. Finally, since $v_k \to 0$, $t_k \geq t_{\min}$ (cf. Step 7), and $\hat{u}^k = \hat{u}^{\bar{k}}$ for $k \geq \bar{k}$, we have $V_k \to 0$ by (2.20). □

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

LEMMA 3.3. *Suppose that there exists $\bar{k}$ such that only null steps occur for all $k \geq \bar{k}$. Let $K := \{k \geq \bar{k} : t_{k+1} > t_k\}$ if $t_k \to \infty$, $K := \{k : k \geq \bar{k}\}$ otherwise. Then $V_k \xrightarrow{K} 0$.*

Proof. Steps 5–7 ensure that the sequence $\{t_k\}$ is monotone for large k. We have $V_k \xrightarrow{K} 0$ from either Lemma 3.1 if $t_\infty = \infty$, or Lemma 3.2 if $t_\infty < \infty$. □

It remains to analyze the case of infinitely many descent steps.

LEMMA 3.4. *Suppose that infinitely many descent steps occur and $f_{\hat{u}}^\infty := \lim_k f_{\hat{u}}^k > -\infty$. Let $K := \{k : f_{\hat{u}}^{k+1} < f_{\hat{u}}^k\}$. Then $\underline{\lim}_{k\in K} V_k = 0$. Moreover, if $\{\hat{u}^k\}$ is bounded, then $V_k \xrightarrow{K} 0$.*

Proof. We have $0 < \kappa v_k \leq f_{\hat{u}}^k - f_{\hat{u}}^{k+1}$ if $k \in K$, $f_{\hat{u}}^{k+1} = f_{\hat{u}}^k$ otherwise (see Step 6). Thus $\sum_{k\in K} \kappa v_k \leq f_{\hat{u}}^1 - f_{\hat{u}}^\infty < \infty$ gives $v_k \xrightarrow{K} 0$ and hence $\epsilon_k, t_k|p^k|^2 \xrightarrow{K} 0$ by (2.19) and $|p^k| \xrightarrow{K} 0$, using $t_k \geq t_{\min}$ (cf. Step 7). For $k \in K$, $\hat{u}^{k+1} - \hat{u}^k = -t_k p^k$ by (2.9), so

$$|\hat{u}^{k+1}|^2 - |\hat{u}^k|^2 = t_k\{t_k|p^k|^2 - 2\langle p^k, \hat{u}^k\rangle\}.$$

Sum up and use the facts that $\hat{u}^{k+1} = \hat{u}^k$ if $k \notin K$, $\sum_{k\in K} t_k \geq \sum_{k\in K} t_{\min} = \infty$ to get

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

(since otherwise $|\hat{u}^k|^2 \to -\infty$, which is impossible). Combining this with $t_k|p^k|^2 \xrightarrow{K} 0$ gives $\underline{\lim}_{k\in K}\langle p^k, \hat{u}^k\rangle \leq 0$. Since also $\epsilon_k, |p^k| \xrightarrow{K} 0$, we have $\underline{\lim}_{k\in K} V_k = 0$ by (2.11).

If $\{\hat{u}^k\}$ is bounded, using $\epsilon_k, |p^k| \xrightarrow{K} 0$ in Lemma 2.2(v) gives $V_k \xrightarrow{K} 0$. □

We may now state and prove our principal result.

THEOREM 3.5. (i) *We have* $f_{\hat{u}}^k \downarrow f_{\hat{u}}^\infty \leq f_*$, *and additionally* $\underline{\lim}_k V_k = 0$ *if* $f_* > -\infty$.

(ii) $f_* \leq \underline{\lim}_k f(\hat{u}^k) \leq \overline{\lim}_k f(\hat{u}^k) \leq f_{\hat{u}}^\infty + \epsilon_f$.

Proof. The inequalities in (ii) stem from the facts that $f_* = \inf_C f$, $\{\hat{u}^k\} \subset C$, and $f(\hat{u}^k) \leq f_{\hat{u}}^k + \epsilon_f$ for all k by (2.2). By (ii), if $f_{\hat{u}}^\infty = -\infty$, then $f_* = -\infty$ in (i). Hence, suppose $f_* > -\infty$. Then $f_{\hat{u}}^\infty \geq f_* - \epsilon_f > -\infty$ by (ii). We have $\underline{\lim}_k V_k = 0$ by Lemma 3.3 in the case of finitely many descent steps, or by Lemma 3.4 otherwise. Finally, using $\underline{\lim}_k V_k = 0$ in the estimate (2.18) gives $f_{\hat{u}}^\infty \leq \inf f_C = f_*$. □

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

Remark 3.6. (i) Inspection of the preceding proofs reveals that Theorem 3.5 requires only convexity and finiteness of f on C, and *local boundedness* of the approximate subgradient mapping $u \mapsto g_u$ of f on C (see below (3.8)). In particular, it suffices to assume that f is finite convex on a neighborhood of C.

(ii) The requirement $\max\{\bar{f}_{k-1}, f_k\} \leq \check{f}_k$ of (2.4) is needed only after null steps in the proof of Lemma 3.2. After a descent step (when $k = k(l)$), Step 1 may take any $\check{f}_k \leq f_C$.

We now show that for exact evaluations ($\epsilon_f = 0$), our algorithm has the usual strong convergence properties of typical bundle methods. Instead of requiring that $\inf_k t_k \geq t_{\min} > 0$, as before, we give more general stepsize conditions in the theorem below.

THEOREM 3.7. *Suppose that* $\epsilon_f = 0$. *Let* $U_* := \operatorname{Arg\,min}_C f$ *denote the (possibly empty) solution set of problem* (1.1). *Then we have the following statements:*

(i) *If only* $l < \infty$ *descent steps occur and* $t_k \downarrow t_\infty > 0$, *then* $\hat{u}^{k(l)} \in U_*$ *and* $V_k \to 0$.

(ii) *Assuming that infinitely many descent steps occur, suppose that* $\sum_{k\in K} t_k = \infty$ *for* $K := \{k : f(\hat{u}^{k+1}) < f(\hat{u}^k)\}$. *Then* $f(\hat{u}^k) \downarrow f_*$. *Moreover, we have the following.*

(a) Let $\check{\epsilon}_k := f(\hat{u}^{k+1}) - \bar{f}_k(\hat{u}^{k+1})$ for $k \in K$. If $U_* \neq \emptyset$ and $\sum_{k\in K} t_k\check{\epsilon}_k < \infty$ (e.g., $\sup_{k\in K} t_k < \infty$), then $\hat{u}^k \to \hat{u}^\infty \in U_*$, and $V_k \xrightarrow{K} 0$ if $\inf_{k\in K} t_k > 0$.

(b) If $U_* = \emptyset$, then $|\hat{u}^k| \to \infty$.

Proof. Since $\epsilon_f = 0$, Step 5 is inactive, and Algorithm 2.1 fits the framework of [Kiw99, Alg. 3.1]. For $l \not\to \infty$, the conclusion follows from Lemma 3.2 and Theorem 3.5. For $l \to \infty$, combine [Kiw99, Thm. 4.4] and the proof of Lemma 3.4. □

4. Modifications.

4.1. Looping between subproblems. To obtain a more accurate solution to the prox subproblem (2.3), we may cycle between subproblems (2.5) and (2.7), updating their data as if null steps occur without changing the model $\check{f}_k$. Specifically, for a given *subproblem accuracy threshold* $\check{\kappa} \in (0,1)$, suppose that the following step is inserted after Step 5.

Step 5′ (subproblem accuracy test). If

$$\check{f}_k(u^{k+1}) > f_{\hat{u}}^k - \check{\kappa} v_k, \tag{4.1}$$

set $\bar{\imath}_C^{k-1}(\cdot) := \bar{\imath}_C^k(\cdot)$, $p_C^{k-1} := p_C^k$ and go back to Step 2.

We now give two motivations for the test (4.1) written as (cf. (2.9))

$$\bar{\epsilon}_k := \check{f}_k(u^{k+1}) - \bar{f}_k(u^{k+1}) > (1 - \check{\kappa}) v_k.$$

First, when $\bar{\epsilon}_k$ is small relative to v_k, $\check{f}_k$ is correctly approximated by $\bar{f}_k$, so the loop can be broken. Second, since $\bar{f}_k \le \check{f}_k$ (Lemma 2.2(i)) in (2.7), by standard arguments [Kiw99, p. 145], the distance from u^{k+1} to the prox solution of (2.3) is at most $\sqrt{2t_k\bar{\epsilon}_k}$.

The analysis of this modification is given in the following remarks.

Remark 4.1. (i) For any k, each execution of Steps 2 through 5′ is called a loop. First, suppose that finitely many loops occur for each k. By its proof, Lemma 2.2 holds at Step 4 for the current quantities. This suffices for the proofs of Lemmas 2.3, 3.1, and 3.4, whereas the proofs of Lemma 3.3 and Theorem 3.5 will go through once Lemma 3.2 is established. The proof of Lemma 3.2 is modified as follows. For each $k \ge \bar{k}$, (3.6) and (3.7a) hold at each loop, and (3.7b) holds for the final loop. For any preceding loop, letting $\check{u}_{\text{next}}^{k+1}$ and $\phi_{f,\text{next}}^k$ stand for $\check{u}^{k+1}$ and ϕ_f^k produced by Step 2 on the next loop, use the minorization $\bar{f}_k \le \check{f}_k$ of (2.13) in subproblems (3.4) and (2.7) to get $\bar{\phi}_C^k \le \phi_{f,\text{next}}^k$ and, by (3.5),

$$\phi_C^k(u^{k+1}) + \tfrac{1}{2t_k}|\check{u}_{\text{next}}^{k+1} - u^{k+1}|^2 = \bar{\phi}_C^k(\check{u}_{\text{next}}^{k+1}) \le \phi_{f,\text{next}}^k(\check{u}_{\text{next}}^{k+1}). \tag{4.2}$$

Then, replacing (3.7b) by (4.2) for all nonfinal loops, we deduce that the optimal values $\phi_f^k(\check{u}^{k+1}) \le \phi_C^k(u^{k+1})$ can't decrease during the loops or when k grows; hence (3.8) and the boundedness of $\{\check{u}^{k+1}\}$ and $\{u^{k+1}\}$ follow as before. For the rest of the proof, let $\check{u}^{k+2}$ in (3.9) stand for the point produced by Step 2 on the first loop at iteration $k+1$, and argue as before.

(ii) Next, suppose that infinitely many loops occur at iteration $k = \check{k}$, for some $\check{k}$. If Step 5 drives t_k to ∞, $f_{\hat{u}}^k \le f_*$ and $V_k \to 0$ by the proof of Lemma 2.3. Hence we may assume that Step 5 doesn't increase t_k at all. To show that $V_k \to 0$ (in which case $f_{\hat{u}}^{\check{k}} \le f_*$ by (2.18)), we suppose that the subdifferential $\partial\check{f}_{\check{k}}$ is locally bounded, and we use a subgradient mapping $C \ni u \mapsto \check{g}_u \in \partial\check{f}_{\check{k}}(u)$. Consider the following modification of Algorithm 2.1. Starting from the first loop at iteration $k = \check{k}$, omit Step 5′; at Step 6 set $f_u^{k+1} := \check{f}_{\check{k}}(u^{k+1})$, $g^{k+1} := \check{g}_{u^{k+1}}$, and $\kappa := \check{\kappa}$; at Step 7, set $t_{k+1} := t_k$; finally, when Step 1 is reached, set $\check{f}_k := \check{f}_{k-1}$. This modification only translates loops into additional iterations with a constant model $\check{f}_k = \check{f}_{\check{k}}$; in particular, only null steps occur, because the descent test (2.10) can't hold with $f_u^{k+1} := \check{f}_k(u^{k+1})$ and $\kappa := \check{\kappa}$ due to the model test (4.1). Further, the "new" linearization $f_{k+1}(\cdot) := f_u^{k+1} + \langle g^{k+1}, \cdot - u^{k+1}\rangle$ satisfies $f_{k+1} \le \check{f}_{k+1}$. Hence, to get $V_k \to 0$, we may use the proof of Lemma 3.2, obtaining boundedness of $\{p_f^k\}$, $\{g^{k+1}\}$ from the boundedness of $\{\check{u}^{k+1}\}$, $\{u^{k+1}\}$ and the local boundedness of $\partial\check{f}_{\check{k}}$.

Note that having $\bar{\imath}_C^{k-1}$ as a model of i_C in subproblem (2.5) is essential only after null steps or loops due to Step 5′. Otherwise, a better model may be constructed as follows. After Step 5 increases t_k, we can set $\bar{\imath}_C^{k-1}(\cdot) := \bar{\imath}_C^k(\cdot)$, $p_C^{k-1} := p_C^k$, or use the more efficient update $u^k := P_C(\hat{u}^k - t_k p_f^k)$, $p_C^{k-1} := (\hat{u}^k - u^k)/t_k - p_f^k$, and $\bar{\imath}_C^{k-1}(\cdot) := \langle p_C^{k-1}, \cdot - u^k\rangle$, which corresponds to resolving subproblem (2.7) before going back to Step 2. Similarly, if $\hat{u}^{k+1} \neq \hat{u}^k$ after Step 7, we may use $\tilde{u} := P_C(\hat{u}^{k+1} - t_{k+1}p_f^k)$, $p_C^k := (\hat{u}^{k+1} - \tilde{u})/t_{k+1} - p_f^k$, and $\bar{\imath}_C^k(\cdot) := \langle p_C^k, \cdot - \tilde{u}\rangle$, where $\tilde{u}$ plays the rôle of u^{k+1}.

4.2. Evaluation errors and relaxed null-step requirements. We now inspect the impact of inexact evaluations on our preceding results, in order to obtain weaker convergence conditions and to provide some practical recommendations.

Our assumption (1.2) on the error tolerance ϵ_f means $\epsilon_f := \sup_{u \in C}[f(u) - f_u] < \infty$. In fact, we need only the weaker condition that $\epsilon_f := \sup_k \epsilon_f^k < \infty$ for the *evaluation errors* $\epsilon_f^k := f(u^k) - f_u^k$ (cf. (2.1)). Thus, for $\epsilon_f := \sup_k \epsilon_f^k$, Theorem 3.5 says that our method produces solutions that are as good as the supplied linearizations.

In fact, the asymptotic accuracy depends *only* on the errors that occur at descent steps. Indeed, at Step 1 we have $\hat{u}^k = u^{k(l)}$ and $f(\hat{u}^k) = f_{\hat{u}}^k + \epsilon_f^{k(l)}$, where $k(l) - 1$ is the iteration number of the lth (i.e., latest) descent step (see Steps 0 and 6). Hence the tolerance ϵ_f in Theorem 3.5(ii) may be replaced by the *asymptotic error*

$$\epsilon_f^\infty := \begin{cases} \epsilon_f^{k(l)} & \text{if only } l < \infty \text{ descent steps occur,} \\ \overline{\lim}_l\, \epsilon_f^{k(l)} & \text{otherwise.} \end{cases} \tag{4.3}$$

In particular, $\epsilon_f^\infty = 0$ if all descent steps happen to be exact. On the other hand, whenever an inexact descent step occurs, then $\epsilon_f^{k+1} := f(u^{k+1}) - f_u^{k+1}$ may potentially determine ϵ_f^∞ (only if $f_u^{k+1} \le f_*$, since $f_{\hat{u}}^\infty \le f_*$ by Theorem 3.5).

Since the asymptotic error is not influenced by the errors occurring at null steps, let us now discuss the case where infinitely many successive null steps occur. Then, by the proof of Lemma 3.2, instead of the requirement $\sup_k \epsilon_f^k < \infty$ (which may be difficult to check for some oracles), it suffices if the following *relaxed null-step requirements* are met:

(a) the sequence $\{g^k\}$ is bounded whenever the sequence $\{u^k\}$ is bounded;

(b) a null step implies that $f_u^{k+1} > f_{\hat{u}}^k - \bar{\kappa} v_k$ for some fixed parameter $\bar{\kappa} \in [\kappa, 1)$.

Condition (a) holds if the mapping $u \mapsto g_u$ is locally bounded on C (cf. Remark 3.6(i)). Condition (b) means that the new linearization f_{k+1} may have any accuracy, as long as it improves the next model sufficiently at u^{k+1}. For $\bar{\kappa} > \kappa$, the oracle may set an indicator $i_R := 1$ when $\bar{\kappa}$ should replace κ in the descent test (2.10) to accept a shallower null step; $i_R := 0$ otherwise (i.e., when (2.10) is not modified). Of course, shallow cuts may slow down convergence, but this may be offset by saving the oracle's work per call. To illustrate these requirements, consider the following generalization of the setting of [HeK02].

Example 4.2. Suppose that the objective f has the form $f(\cdot) := \sup_{z \in Z} F_z(\cdot)$ of (1.3) with $F_z(\cdot)$ convex and $\partial F_z(\cdot)$ locally bounded on C, uniformly w.r.t. $z \in Z$. Suppose for each k that the oracle used for approximate evaluation of $f(u^{k+1})$ generates points $z^{(i)} \in Z$, $i = 1, 2, \ldots$, stopping for some i to deliver $f_u^{k+1} := F_{z^{(i)}}(u^{k+1})$ and some $g^{k+1} \in \partial F_{z^{(i)}}(u^{k+1})$. To meet the relaxed null-step requirements, the oracle may stop when $F_{z^{(i)}}(u^{k+1}) > f_{\hat{u}}^k - \bar{\kappa} v_k$ holds, possibly together with other conditions, setting $i_R := 1$ to force a null step.

Remark 4.3. For an SDP (cf. section 5.6), Example 4.2 accommodates the "inexact null steps" of [HeK02], which can save much work in eigenvalue computations [Hel03, Nay99, Nay05]. In general, when the relaxed null-step requirements are met and the descent steps are exact, then $\epsilon_f^\infty = 0$ in (4.3) and Theorem 3.7 holds (by its proof). In particular, Theorem 3.7 holds for the method of [HeK02].

Insisting that all descent steps be exact may be unrealistic (e.g., as in [Hel03, HeK02, Nay05], where this issue is ignored) or too expensive (cf. [Kiw05]).

For the oracle of Example 4.2, additional stopping criteria may be employed to make a "too inexact" descent step less likely. The general idea is to make the oracle work harder before a descent step is accepted. We distinguish the following two cases.

Case 1. Suppose that the oracle's underestimates $F_{z^{(i)}}(u^{k+1})$ of $f(u^{k+1})$ improve when i grows. Then for a given iteration limit $i_{\max}$ the oracle may stop when either $F_{z^{(i)}}(u^{k+1}) > f_{\hat{u}}^k - \bar{\kappa} v_k$ and $i \le i_{\max}$ (setting $i_R := 1$ to force a null step), or $F_{z^{(i)}}(u^{k+1}) \le f_{\hat{u}}^k - \kappa v_k$ and $i = i_{\max}$ (setting $i_R := 0$ for a descent step).

Case 2. In addition to the assumptions of Case 1, suppose that the oracle generates upper bounds $f_{\rm up}^{(i)} \ge f(u^{k+1})$ such that $\bar{f}_{\rm up}^{(i)} - F_{z^{(i)}}(u^{k+1}) \to 0$ if $i \to \infty$. Then the oracle may also stop as soon as for some $i \le i_{\max}$, $\bar{f}_{\rm up}^{(i)} < f_{\hat{u}}^k$, or $f_{\rm up}^{(i)} - F_{z^{(i)}}(u^{k+1}) \le \epsilon_r |F_{z^{(i)}}(u^{k+1})|$ for a given *relative accuracy tolerance* $\epsilon_r > 0$, setting $i_R := 0$ to promote a descent step.

We add that Case 2 covers oracles employing branch and bound in Lagrangian relaxation of integer programming problems. Then, for difficult Lagrangian subproblems, it pays to use rather loose accuracy requirements, because tighter criteria (e.g., small ϵ_r) may force the oracle to work too long on some calls (see, e.g., [Kiw05]). Fortunately, a typical branch-and-bound oracle generates a good lower bound $F_{z^{(i)}}(u^{k+1})$ quickly (although improving the upper bound $f_{\rm up}^{(i)}$ may need much time). Then the stopping criterion of Case 2 with a moderate tolerance ϵ_r (or another heuristic criterion) may still ensure that the actual error $\epsilon_f^{k+1} := f(u^{k+1}) - f_u^{k+1}$ is small enough. Thus our framework is especially suitable for applications with oracles that deliver reasonably accurate linearizations most of the time, although explicit control of their accuracy might be too costly. (We add that the preceding remarks apply also to the method of [Kiw06b], and they partly explain the good numerical results of [Kiw05].)

4.3. A weaker descent test. As in [Kiw06b, sect. 4.3], at Steps 5 and 6 we may replace the predicted decrease $v_k = t_k|p^k|^2 + \epsilon_k$ (cf. (2.16)) by the smaller quantity $w_k := t_k|p^k|^2/2 + \epsilon_k$. Then the equivalences in Lemma 2.2(vi) are replaced by the fact that

$$w_k \ge -\epsilon_k \quad \Longleftrightarrow \quad \frac{t_k|p^k|^2}{4} \ge -\epsilon_k \quad \Longleftrightarrow \quad w_k \ge \frac{t_k|p^k|^2}{4}.$$

Hence, $w_k \ge -\epsilon_k$ at Step 6 implies $w_k \le v_k \le 3w_k$ and $v_k \ge -\epsilon_k$ for the bounds (2.19)–(2.20), whereas for Step 5, the bound (2.21) is replaced by the fact that

$$V_k < \left(\frac{4\epsilon_{\max}}{t_k}\right)^{1/2} (1 + |\hat{u}^k|) \quad \text{if} \quad w_k < -\epsilon_k.$$

The preceding results extend easily. (In the proof of Lemma 3.2, $f_u^{k+1} > f_{\hat{u}}^k - \kappa w_k$ implies $f_u^{k+1} > f_{\hat{u}}^k - \kappa v_k$, whereas in the proof of Lemma 3.4, $\sum_{k\in K} v_k \le 3\sum_{k\in K} w_k < \infty$.)

4.4. Linearization accumulation, selection, and aggregation. There are three basic choices of polyhedral models satisfying relation (2.4) rewritten as

$$\max\{\bar{f}_k, f_{k+1}\} \le \check{f}_{k+1} \le f_C. \tag{4.4}$$

First, *accumulation* takes $\check{f}_{k+1} := \max\{\check{f}_k, f_{k+1}\}$, $\check{f}_1 := f_1$; then we may replace f_C by f in (4.4), using the minorizations $\bar{f}_k \le \check{f}_k$ of (2.13) and $f_{k+1} \le f$ of (2.1). In other words, here $\check{f}_k = \max_{j=1}^k f_j$ is the richest model stemming from all the past linearizations, but its storage requirements and QP work per iteration grow with k, so the other choices discussed below are more attractive in practice.

Second, *selection* retains only selected linearizations for its kth model,

$$\check{f}_k(\cdot) := \max_{j \in J_k} f_j(\cdot) \quad \text{with} \quad k \in J_k \subset \{1, \ldots, k\}. \tag{4.5}$$

Then $\check{f}_k \le f$ by (2.1), so, in view of (4.4), we need only show how to choose the set J_{k+1} so that $\bar{f}_k \le \check{f}_{k+1}$. Since $p_f^k \in \partial \check{f}_k(\check{u}^{k+1})$ by (2.12) and each f_j is affine in (4.5), there exist multipliers ν_j^k, $j \in J_k$, also known as *convex weights*, such that (cf. [HUL93, Ex. VI.3.4])

$$(p_f^k, 1) = \sum_{j \in J_k} \nu_j^k (\nabla f_j, 1), \quad \nu_j^k \ge 0, \quad \nu_j^k \big[\check{f}_k(\check{u}^{k+1}) - f_j(\check{u}^{k+1})\big] = 0, \quad j \in J_k. \tag{4.6}$$

Then, using relations (2.6) and (4.6), it is easy to obtain the following expansion:

$$(\bar{f}_k, 1) = \sum_{j \in \hat{J}_k} \nu_j^k (f_j, 1) \quad \text{with} \quad \hat{J}_k := \{j \in J_k : \nu_j^k > 0\}. \tag{4.7}$$

In other words, the aggregate linearization $\bar{f}_k$ is a convex combination of the "ordinary" linearizations f_j selected by the *active* set $\hat{J}_k$. Since $\bar{f}_k \le \max_{j \in \hat{J}_k} f_j$, it suffices to choose

$$J_{k+1} \supset \hat{J}_k \cup \{k+1\}. \tag{4.8}$$

Active-set methods for solving subproblem (2.5) [Kiw86, Kiw94] find multipliers ν_j^k such that $|\hat{J}_k| \le n+1$. Hence we can keep $|J_{k+1}| \le \bar{n}$ for any given upper bound $\bar{n} \ge n+2$.

Third, *aggregation* treats the past aggregate linearizations $\bar{f}_j$ like the "ordinary" linearizations f_j, defining $f_{-j} := \bar{f}_j$ for $j = 0\colon k-1$ to replace (4.5) by the aggregate model

$$\check{f}_k(\cdot) := \max_{j \in J_k} f_j(\cdot) \quad \text{with} \quad k \in J_k \subset \{1-k\colon k\}, \quad f_j := \bar{f}_{-j} \text{ for } j \le 0. \tag{4.9}$$

The weights ν_j^k of (4.6) produce $f_{-k} := \bar{f}_k$ via (4.7), and relation (4.8) is replaced by

$$J_{k+1} \supset \{-k, k+1\}, \tag{4.10}$$

so that only $\bar{n} \ge 2$ linearizations may be kept. Formally, if $f_j \le f$ for all $j \in J_k$, then $f_{-k} := \bar{f}_k \le f$ by (4.7); hence, by induction, (4.9)–(4.10) yield (4.4) for all k. Of course, the selection requirement (4.8) may replace (4.10) whenever $|\hat{J}_k| \le \bar{n} - 1$. After a descent step, we can replace (4.8) and (4.10) by $J_{k+1} \ni k+1$ (cf. Remark 3.6(ii)).

Remark 4.4. In the proof of Lemma 3.2, condition (3.1) holds *automatically* for the models discussed above. Indeed, by (4.6) (and induction for aggregation), we have $p_f^k \in \mathrm{co}\{g^j\}_{j=1}^k$ and hence $|p_f^k| \le \max_{j=1}^k |g^j|$, whereas the sequence $\{g^k\}$ is bounded. Similarly, each model $\check{f}_k$ has a bounded subdifferential, as required in Remark 4.1(ii).

5. Lagrangian relaxation.

5.1. The primal problem. Let $\mathcal{Z}$ be a real inner-product space with a finite dimension $\bar{m}$. (We could, of course, always identify $\mathcal{Z}$ with $\mathbb{R}^{\bar{m}}$, but a less concrete approach helps our future development.) In this section we consider the special case where problem (1.1) with $C := \mathbb{R}^n_+$ is the Lagrangian dual problem of the following *primal* convex optimization problem in $\mathcal{Z}$:

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

where $\emptyset \neq Z \subset \mathcal{Z}$ is compact and convex, and each ψ_i is concave and closed (upper semicontinuous) with $\operatorname{dom}\psi_i \supset Z$. The Lagrangian of (5.1) has the form $\psi_0(z) + \langle u, \psi(z)\rangle$, where $\psi := (\psi_1, \ldots, \psi_n)$ and u is a multiplier. Suppose that, at each $u \in C$, the *dual function*

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

can be evaluated with *accuracy* $\epsilon_f \geq 0$ by finding a *partial Lagrangian* ϵ_f*-solution*

$$z(u) \in Z \quad \text{such that} \quad f_u := \psi_0(z(u)) + \langle u, \psi(z(u))\rangle \geq f(u) - \epsilon_f. \tag{5.3}$$

Thus f is finite convex and has an ϵ_f-subgradient mapping $g_u := \psi(z(u))$ for $u \in C$. In view of Remark 3.6(i), we suppose that $\psi(z(\cdot))$ is locally bounded on C. (Note that the whole set $\psi(z(C))$ is bounded if $\inf_Z \min_{i=1}^n \psi_i > -\infty$, or the function ψ is continuous on Z.)

5.2. Primal recovery with selection. We first consider our method with linearization selection (cf. section 4.4).

The partial Lagrangian solutions $z^k := z(u^k)$ (cf. (5.3)) and their constraint values $g^k := \psi(z^k)$ determine the linearizations (2.1) 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. (4.6)), we may estimate a solution to (5.1) via the *aggregate primal solution*

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

By (4.7), this convex combination is associated with the aggregate linearization $\bar{f}_k$ via

$$(\bar{f}_k, \hat{z}^k, 1) = \sum_{j\in \hat{J}_k} \nu_j^k (f_j, z^j, 1) \quad \text{with} \quad \hat{J}_k := \{j \in J_k : \nu_j^k > 0\}. \tag{5.6}$$

We now derive useful bounds on $\psi_0(\hat{z}^k)$ and $\psi(\hat{z}^k)$, generalizing [Kiw06b, Lem. 5.1].

LEMMA 5.1. *$\hat{z}^k \in Z$, $\psi_0(\hat{z}^k) \geq f_{\hat{u}}^k - \epsilon_k - \langle p^k, \hat{u}^k\rangle$, and $\psi(\hat{z}^k) \geq p_f^k \geq p^k$.*

Proof. By (5.6), $\hat{z}^k \in \operatorname{co}\{z^j\}_{j\in\hat{J}_k} \subset Z$, $\psi_0(\hat{z}^k) \geq \sum_j \nu_j^k \psi_0(z^j)$, and $\psi(\hat{z}^k) \geq \sum_j \nu_j^k \psi(z^j)$ by convexity of Z and concavity of ψ_0, ψ. Since $p_C^k \in \partial i_{\mathbb{R}^n_+}(u^{k+1})$ by (2.12), we have $p_C^k \leq 0$ and $\langle p_C^k, u^{k+1}\rangle = 0$ [HUL93, Ex. III.5.2.6(b)], so $p_f^k = p^k - p_C^k \geq p^k$ by (2.14). Next, using (5.6) with $p_f^k = \nabla \bar{f}_k$ by (2.6) and $\nabla f_j = \psi(z^j)$ by (5.4), we get $\bar{f}_k(0) = \sum_j \nu_j^k \psi_0(z^j)$ and $p_f^k = \sum_j \nu_j^k \psi(z^j)$. Since $\bar{f}_k(0) = \bar{f}_C^k(0) - \bar{\imath}_C^k(0)$ with

$\bar{\imath}_C^k(0) = -\langle p_C^k, u^{k+1}\rangle = 0$ from (2.8), we have $\bar{f}_k(0) = \bar{f}_C^k(0) = f_{\hat{u}}^k - \epsilon_k - \langle p^k, \hat{u}^k\rangle$ by (2.17). Combining the preceding relations yields the conclusion. □

In terms of the optimality measure V_k of (2.11), the bounds of Lemma 5.1 imply

$$\hat{z}^k \in Z \quad \text{with} \quad \psi_0(\hat{z}^k) \geq f_{\hat{u}}^k - V_k, \quad \psi_i(\hat{z}^k) \geq -V_k, \quad i = 1\colon n. \tag{5.7}$$

We now show that $\{\hat{z}^k\}$ has cluster points in the set of *ϵ_f-optimal primal solutions* of (5.1),

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

unless this set is empty, i.e., the primal problem is infeasible.

THEOREM 5.2. *Either $f_* = -\infty$ and $f_{\hat{u}}^k \downarrow -\infty$, in which case the primal problem (5.1) is infeasible, or $f_* > -\infty$, $f_{\hat{u}}^k \downarrow f_{\hat{u}}^\infty \in [f_* - \epsilon_f, f_*]$, $\overline{\lim}_k f(\hat{u}^k) \leq f_{\hat{u}}^\infty + \epsilon_f$, and $\underline{\lim}_k V_k = 0$. In the latter case, let $K' \subset \mathbb{N}$ be a subsequence such that $V_k \xrightarrow{K'} 0$. Then we have the following:*

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

(ii) *Let $\hat{z}^\infty$ be a cluster point of the sequence $\{\hat{z}^k\}_{k\in K'}$. Then $\hat{z}^\infty \in Z_{\epsilon_f}$.*

(iii) *$d_{Z_{\epsilon_f}}(\hat{z}^k) := \inf_{z\in Z_{\epsilon_f}} |\hat{z}^k - z| \xrightarrow{K'} 0$.*

Proof. The first assertion follows from Theorem 3.5 (since $f_* = -\infty$ implies primal infeasibility by weak duality). In the second case, using $f_{\hat{u}}^k \downarrow f_{\hat{u}}^\infty \geq f_* - \epsilon_f$ and $V_k \xrightarrow{K'} 0$ in the bounds of (5.7) yields $\underline{\lim}_{k\in K'} \psi_0(\hat{z}^k) \geq f_* - \epsilon_f$ and $\underline{\lim}_{k\in K'} \min_{i=1}^n \psi_i(\hat{z}^k) \geq 0$.

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

(ii) We have $\hat{z}^\infty \in Z$, $\psi_0(\hat{z}^\infty) \geq f_* - \epsilon_f$, and $\psi(\hat{z}^\infty) \geq 0$ by the closedness of ψ_0 and ψ. Since $f_* \geq \psi_0^{\max}$ by weak duality (cf. (1.1), (5.1), (5.2)), we get $\psi_0(\hat{z}^\infty) \geq \psi_0^{\max} - \epsilon_f$. Thus $\hat{z}^\infty \in Z_{\epsilon_f}$ by the definition (5.8).

(iii) This follows from (i), (ii), and the continuity of the distance function $d_{Z_{\epsilon_f}}$. □

Remark 5.3. (i) For Theorem 5.2, we can replace ϵ_f in (5.8) by ϵ_f^∞ (cf. (4.3)).

(ii) By the proofs of Lemma 2.3 and Theorem 5.2, if an infinite cycle between Steps 2 and 5 occurs, then $V_k \to 0$ yields $d_{Z_{\epsilon_f}}(\hat{z}^k) \to 0$. Similarly, if Step 4 terminates with $V_k = 0$, then $\hat{z}^k \in Z_{\epsilon_f}$. In both cases, we can replace ϵ_f with ϵ_f^∞ (cf. (4.3)).

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

$$\psi_0(\hat{z}^k) \geq f_{\hat{u}}^k - \epsilon_{\text{tol}} \quad \text{and} \quad \psi_i(\hat{z}^k) \geq -\epsilon_{\text{tol}}, \quad i = 1\colon n.$$

Then $\psi_0(\hat{z}^k) \geq \psi_0^{\max} - \epsilon_f - \epsilon_{\text{tol}}$ from $f_{\hat{u}}^k \geq f_* - \epsilon_f$ (cf. (2.2)) and $f_* \geq \psi_0^{\max}$ (weak duality), so that the point $\hat{z}^k \in Z$ is an approximate primal solution of (5.1). This stopping criterion will be satisfied for some k if $f_* > -\infty$ (cf. (5.7) and Theorem 5.2).

5.3. Primal recovery with aggregation. Let us now consider the variant with aggregation based on (4.9), where each linearization f_j has an associated primal point z^j, with $f_j := \bar{f}_{-j}$ and $z^j := \hat{z}^{-j}$ for $j < 0$. Letting $z^0 := z^1$, suppose for induction that $(f_j, z^j) \in \operatorname{co}\{(f_i, z^i)\}_{i=0}^{|j|}$ for $j \in J_k$. For the convex weights ν_j^k satisfying (4.7), let $z^{-k} := \hat{z}^k$ for the aggregate primal solution $\hat{z}^k$ given by (5.6). Since a convex combination of convex combinations of given points is a convex combination of those points, we deduce the existence of convex weights $\bar{\nu}_j^k$ such that

$$(f_{-k}, z^{-k}, 1) := (\bar{f}_k, \hat{z}^k, 1) = \sum_{0\leq j\leq k} \bar{\nu}_j^k (f_j, z^j, 1) \quad \text{with} \quad \bar{\nu}_j^k \geq 0,\ j = 0\colon k. \tag{5.9}$$

In other words, $(f_{-k}, z^{-k}) \in \operatorname{co}\{(f_i, z^i)\}_{i=0}^k$, as required for induction. Replacing (5.6) by (5.9) for Lemma 5.1, we conclude that the preceding convergence results remain valid.

5.4. Handling primal equality constraints. Consider the primal problem (5.1) with additional equality constraints of the form

$$\psi_0^{\max} := \max\ \psi_0(z) \quad \text{s.t.} \quad \psi_{\mathcal{I}}(z) \ge 0,\ \psi_{\mathcal{E}}(z) = 0,\ z \in Z, \tag{5.10}$$

where $\mathcal{I} \cup \mathcal{E} = \{1\colon n\}$, $\mathcal{I} \cap \mathcal{E} = \emptyset$, and $\psi_{\mathcal{E}}$ is affine. For $C := \mathbb{R}_+^{|\mathcal{I}|} \times \mathbb{R}^{|\mathcal{E}|}$, the final bound in Lemma 5.1 becomes $\psi_{\mathcal{I}}(\hat{z}^k) \ge p_{f,\mathcal{I}}^k \ge p_{\mathcal{I}}^k$, $\psi_{\mathcal{E}}(\hat{z}^k) = p_{f,\mathcal{E}}^k = p_{\mathcal{E}}^k$ (using $p_{C,\mathcal{I}}^k \le 0$, $p_{C,\mathcal{E}}^k = 0$, $\langle p_C^k, u^{k+1}\rangle = 0$ as before); the final inequalities in (5.7) are replaced by $\min_{i\in\mathcal{I}} \psi_i(\hat{z}^k) \ge -V_k$, $\max_{i\in\mathcal{E}} |\psi_i(\hat{z}^k)| \le V_k$, and $\psi(z) \ge 0$ in (5.8) by $\psi_{\mathcal{I}}(z) \ge 0$, $\psi_{\mathcal{E}}(z) = 0$. With these replacements, the proof of Theorem 5.2 extends easily (since $\underline{\lim}_{k\in K'} \max_{i\in\mathcal{E}} |\psi_i(\hat{z}^k)| = 0$ yields $\psi_{\mathcal{E}}(\hat{z}^\infty) = 0$ in (ii)).

Remark 5.4. We add that the ideas of sections 4.2, 5.3, and 5.4 can be translated into additional properties of the method of [Kiw06b]. Further, a simplified variant of the latter method is obtained by modifying relations (2.5)–(2.8) as follows. Letting u^{k+1} solve the prox subproblem (2.3), for the subgradients $p_f^k \in \partial \check{f}_k(u^{k+1})$ and $p_C^k \in \partial i_C(u^{k+1})$ such that $p_f^k + p_C^k = (\hat{u}^k - u^{k+1})/t_k$, define $\bar{f}_k$ by (2.6) with $\check{u}^{k+1} := u^{k+1}$ and $\bar{\imath}_C^k$ by (2.8). Then Lemma 2.2 holds by construction, and the proof of Lemma 3.2 simplifies to that of [Kiw06b, Lem. 3.3]. In effect, except for section 4.1, all the preceding results hold for this variant as well.

5.5. Nonpolyhedral objective models. In addition to the assumptions of section 5.1, suppose ψ is affine: $\psi(z) := b - Az$ for some given $b \in \mathbb{R}^n$ and a linear mapping $A : Z \to \mathbb{R}^n$. Then the Lagrangian of (5.1) has the form

$$L(z,u) := \psi_0(z) + \langle u, \psi(z)\rangle = \psi_0(z) + \langle u, b - Az\rangle \tag{5.11}$$

and $f(\cdot) := \max_{z\in Z} L(z,\cdot)$. Suppose Step 1 selects the (possibly) *nonpolyhedral* model

$$\check{f}_k(\cdot) := \max_{z\in Z_k} L(z,\cdot) \quad \text{with} \quad z^k \in Z_k \subset Z, \tag{5.12}$$

where the set Z_k is closed convex. Since $f_k(\cdot) = L(z^k,\cdot)$ by (5.4), we have $f_k \le \check{f}_k \le f$. Thus, to meet the requirement of (4.4), we need only show how to choose a set $Z_{k+1} \ni z^{k+1}$ so that $\bar{f}_k \le \check{f}_{k+1}$. First, for solving subproblem (2.5) with the model $\check{f}_k$ given by (5.12), we employ the *Lagrangian* $\bar{L} : \mathbb{R}^n \times Z_k \to \mathbb{R}$ of subproblem (2.5) defined by

$$\bar{L}(u,z) := L(z;u) + \langle p_C^{k-1}, u - u^k\rangle + \frac{1}{2t_k}|u - \hat{u}^k|^2, \tag{5.13}$$

so that

$$\phi_f^k(\cdot) = \max\{\bar{L}(\cdot,z) : z \in Z_k\}. \tag{5.14}$$

For each primal point $z \in Z_k$, the (unique) *Lagrangian solution*

$$u_z := \arg\min \bar{L}(\cdot,z) = \hat{u}^k - t_k[\psi(z) + p_C^{k-1}] \tag{5.15}$$

substituted for u in (5.13) gives the value of the *dual function* $q : Z_k \to \mathbb{R}$ defined by

$$q(z) := \min \bar{L}(\cdot,z) = \psi_0(z) + \langle \psi(z), \hat{u}^k\rangle + \langle p_C^{k-1}, \hat{u}^k - u^k\rangle - \frac{t_k}{2}|\psi(z) + p_C^{k-1}|^2. \tag{5.16}$$

Since q is closed and Z_k is compact, the *dual problem* $\max_{Z_k} q$ has at least one solution:

$$\hat{z}^k \in \operatorname{Arg\,max}\{q(z) : z \in Z_k\}. \tag{5.17}$$

LEMMA 5.5. *Given a dual solution $\hat{z} := \hat{z}^k$ of (5.17), define the Lagrangian solution $\check{u} := u_{\hat{z}}$ by (5.15). Then we have the following statements:*

(i) *The pair $(\check{u}, \hat{z})$ is a saddle-point of the Lagrangian $\bar{L}$ defined by (5.13):*

$$\bar{L}(\check{u}, z) \le \bar{L}(\check{u}, \hat{z}) \le \bar{L}(u, \hat{z}) \quad \forall u \in \mathbb{R}^n, z \in Z_k. \tag{5.18}$$

(ii) *For $\check{u}^{k+1}$, $\check{f}_k$, and p_f^k defined by (2.5)–(2.6), we have $\check{u}^{k+1} = \check{u}$, $p_f^k = \psi(\hat{z}^k)$,*

$$\check{u}^{k+1} = \hat{u}^k - t_k\left[\psi(\hat{z}^k) + p_C^{k-1}\right], \tag{5.19}$$

$$\check{f}_k(\cdot) = \psi_0(\hat{z}^k) + \langle \cdot, \psi(\hat{z}^k)\rangle. \tag{5.20}$$

Proof. (i) $\bar{L}$ is convex-concave on $\mathbb{R}^n \times Z_k$, Z_k is compact, and for each $z \in Z_k$, $\bar{L}(u, z) \to \infty$ when $|u| \to \infty$. Hence $\bar{L}$ has a saddle-point $(\bar{u}, \bar{z})$ [HUL93, Thm. VII.4.3.1]. Since $\hat{z} \in \operatorname{Arg\,max}_{Z_k} \min_u \bar{L}(u, \cdot)$ by (5.16)–(5.17), $(\bar{u}, \hat{z})$ is a saddle-point as well [HUL93, Thm. VII.4.2.5]. Then $\bar{L}(\bar{u}, \hat{z}) \le \bar{L}(u, \hat{z})$ $\forall u$ yields $\bar{u} = u_{\hat{z}} = \check{u}$ by (5.15), so that (5.18) holds.

(ii) By (2.5) and (5.14), (5.18) implies $\check{u}^{k+1} = \check{u}$ [HUL93, Thm. VII.4.2.5]. Then (2.6) and (5.15) with $z = \hat{z}$ yield $p_f^k = \psi(\hat{z}^k)$. The left inequality in (5.18) combined with (5.11)–(5.13) gives $\check{f}_k(\check{u}^{k+1}) = \psi_0(\hat{z}^k) + \langle \check{u}^{k+1}, \psi(\hat{z}^k)\rangle$, and then (2.6) yields (5.20). □

In view of (5.12) and (5.20), the requirement of (4.4) is met if the set Z_{k+1} satisfies

$$Z_{k+1} \supset \{\hat{z}^k, z^{k+1}\}, \tag{5.21}$$

in addition to being a closed convex subset of Z. Further, condition (3.1) holds (with $p_f^k = \psi(\hat{z}^k)$, $\hat{z}^k \in Z_k$, Z_k compact, ψ continuous), and the aggregate representation (5.20) can be seen as a special case of (5.6) (with $\hat{J}_k := \{k\}$ and z^k replaced by $\hat{z}^k$ in (5.4)). In effect, the results of section 5.2 hold for this variant as well.

Remark 5.6. (i) We add that for $p_f^k = \psi(\hat{z}^k)$ (and $C := \mathbb{R}^n_+$), (2.7)–(2.8) simplify to

$$u^{k+1} = \max\{\hat{u}^k - t_k(b - A\hat{z}^k), 0\} \quad \text{and} \quad p_C^k = \min\left\{\frac{\hat{u}^k}{t_k} - b + A\hat{z}^k, 0\right\}. \tag{5.22}$$

In general, $\langle p_C^{k-1}, u^k\rangle = 0$ from $p_C^{k-1} \in \partial i_C(u^k)$, so we can omit u^k in (5.13) and (5.16). A dual interpretation of (5.22) follows. Since $i_C(\cdot) = \sup\{-\langle \eta, \cdot\rangle : \eta \in \mathbb{R}^n_+\}$, using a dual variable $\eta \in \mathbb{R}^n_+$ for subproblem (2.3), its Lagrangian $\bar{L}(u, z, \eta)$, relaxed solution $u_{z,\eta}$, and dual function $q(z, \eta)$ are given by (5.13), (5.15), and (5.16) with p_C^{k-1} replaced by $-\eta$. Let $\eta^k := -p_C^{k-1}$. The dual problem $\max_{Z_k \times \mathbb{R}^n_+} q$ is treated in a Gauss–Seidel fashion by finding $\hat{z}^k \in \operatorname{Arg\,max}_{Z_k} q(\cdot, \eta^k)$ (cf. (5.17)) and then $\eta^{k+1} := \arg\max_{\mathbb{R}^n_+} q(\hat{z}^k, \cdot)$, for which $u^{k+1} = u_{\hat{z}^k, \eta^{k+1}}$ and $\eta^{k+1} = -p_C^k$ by (5.22). Thus alternating linearizations of subproblem (2.3) correspond to coordinatewise maximizations of its dual function.

(ii) Suppose that ψ_0 is linear and $Z_k := \operatorname{co}\{z^j\}_{j=1}^k$. Then $z \in Z_k$ iff $z = \sum_j \nu_j z^j$ for a weight vector ν in $N := \{\nu \in \mathbb{R}^k_+ : \sum_j \nu_j = 1\}$. For $F := [\psi_0(z^1), \ldots, \psi_0(z^k)]$ and $G := [g^1, \ldots, g^k]$, we have $\psi_0(z) = F\nu$ and $\psi(z) = G\nu$. Using these representations in (5.16)–(5.17), we may take $\hat{z}^k = \sum_j \nu_j^k z^j$ for any solution ν^k to the dual QP subproblem

$$\nu^k \in \operatorname{Arg\,max}\left\{F\nu + \nu^T G^T \hat{u}^k - \frac{t_k}{2}|G\nu - p_C^{k-1}|^2 : \nu \in N\right\}. \tag{5.23}$$

In effect, our framework comprises the method of [FGRS06, sect. 3.2], which requires exact evaluations. Note that the similarity of $\hat{z}^k$ above to (5.5) is not accidental: the model (5.12) with $Z_k := \operatorname{co}\{z^j\}_{j=1}^k$ is *equivalent* to the polyhedral model (4.5) with $J_k := \{1\colon k\}$ (cf. (5.11) and (5.4)). Other choices of J_k from section 4.4 correspond to $Z_k := \operatorname{co}\{z^j\}_{j\in J_k}$.

(iii) For problem (5.10) with mixed constraints, formula (5.22) is valid for components indexed by $\mathcal{I}$, whereas $u_{\mathcal{E}}^{k+1} = \hat{u}_{\mathcal{E}}^k - t_k(b - A\hat{z}^k)_{\mathcal{E}}$ and $p_{C,\mathcal{E}}^k = 0$. Then the setting of (ii) above comprises the method of [ReS06, sect. 3] (for exact evaluations).

(iv) By Remark 4.1, the results of section 5.2 hold when Step 5′ is used as well, since each $\check{f}_k$ has bounded subgradients (by (5.11)–(5.12) and the compactness of $Z_k \subset Z$).

5.6. SDP via eigenvalue optimization. To discuss applications in SDP, we need the following notation.

We consider the Euclidean space S^m of $m \times m$ real symmetric matrices with the Frobenius inner product $\langle x, y\rangle = \operatorname{tr} xy$ (we use lowercase notation for the elements of S^m for consistency with the rest of the text). S_+^m is the cone of positive semidefinite matrices. The maximum eigenvalue $\lambda_{\max}(y)$ of a matrix $y \in S^m$ and its positive part $\lambda_{\max}^+(y) := \max\{\lambda_{\max}(y), 0\}$ satisfy (see, e.g., [LeO96, Tod01])

$$\lambda_{\max}(y) = \max\bigl\{\langle y, x\rangle : x \in \Sigma^m\bigr\} \quad \text{with} \quad \Sigma^m := \bigl\{x \in S_+^m : \operatorname{tr} x = 1\bigr\}, \tag{5.24a}$$

$$\lambda_{\max}^+(y) = \max\bigl\{\langle y, x\rangle : x \in \Sigma_{\le}^m\bigr\} \quad \text{with} \quad \Sigma_{\le}^m := \bigl\{x \in S_+^m : \operatorname{tr} x \le 1\bigr\}. \tag{5.24b}$$

Let $a > 0$, $b \in \mathbb{R}^n$, $c \in S^m$, and $A : S^m \to \mathbb{R}^n$ be linear. Consider the *SDPs*

$$(P_=) : \quad \max\ \langle c, x\rangle \quad \text{s.t.} \quad Ax \le b, \quad x \in S_+^m, \quad \operatorname{tr} x = a, \tag{5.25}$$

$$(P_\le) : \quad \max\ \langle c, x\rangle \quad \text{s.t.} \quad Ax \le b, \quad x \in S_+^m, \quad \operatorname{tr} x \le a. \tag{5.26}$$

Any SDP can be formulated as $(P_\le)$ without the final trace condition. If we know or simply guess an upper bound a on the trace of some optimal solution, we may use $(P_\le)$. (For a wrong guess, our method will produce dual values going to $-\infty$, thus indicating primal infeasibility.) Of course, $(P_\le)$ can be formulated as $(P_=)$ by adding a slack variable, but this is not really necessary, since our method can handle both. $(P_=)$ is natural in many combinatorial applications, where the trace of all feasible solutions is known [HeR00]; $(P_\le)$ is employed in [Nay05] for equality-constrained SDPs.

We can regard $(P_=)$ as an instance of (5.1) with $\mathcal{Z} := S^m$, $\psi_0(z) := \langle c, z\rangle$, $\psi(z) := b - Az$, and $Z := a\Sigma^m$. Then, by (5.2) and (5.24a), the dual function f satisfies

$$f(u) = a\lambda_{\max}(c - A^*u) + \langle b, u\rangle \quad \forall\, u, \tag{5.27}$$

where A^* is the adjoint of A (defined by $\langle z, A^*u\rangle = \langle Az, u\rangle\ \forall\, z \in S^m,\ u \in \mathbb{R}^n$). For each u, the approximate evaluation condition (5.3) is met by $z(u) := ar(u)r(u)^T$, where $r(u) \in \mathbb{R}^m$ is an (ϵ_f/a)-*eigenvector* of the matrix $s(u) := c - A^*u \in S^m$ satisfying

$$r(u)^T s(u) r(u) \geq \lambda_{\max}(s(u)) - \frac{\epsilon_f}{a}, \qquad r(u)^T r(u) = 1. \tag{5.28}$$

Then the ϵ_f-subgradient mapping $u \to g_u := \psi(z(u)) = b - Az(u)$ is bounded on $\mathbb{R}^n$.

Thus we can use the setting of section 5.5 with models $\check{f}_k$ given by (5.12) for sets Z_k satisfying (5.21). In effect, the results of section 5.2 and Remark 5.6 hold for this variant as well.

Remark 5.7. (i) Our dual problem $f_* := \inf_C f$ is equivalent to the standard dual of $(P_=)$, which is strictly feasible. Hence (cf. [Tod01, Thm. 4.1]) if $(P_=)$ is feasible, then its optimal value is finite and equals f_*, although the dual problem need not have solutions. Thus, even for exact evaluations, Theorem 5.2 improves upon [Hel04, Thm. 3.6], which assumes that $\operatorname{Arg\,min}_C f \neq \emptyset$. We show elsewhere [Kiw06a] how to extend a related result of [Hel04, Thm. 4.8], without assuming that $\operatorname{Arg\,min}_C f$ is nonempty and bounded.

(ii) Condition (5.28) is particularly useful when approximate eigenvectors are found by iterative methods (such as the Lanczos method [Hel03, Nay05]) that employ only matrix-vector multiplications to exploit the structure of the matrix $s(u) := c - A^*u$. This condition has the following meaning in the setting of Example 4.2 with $u = u^{k+1}$, $s^{k+1} := s(u^{k+1})$. Suppose that an iterative method generates approximate eigenvectors $r^{(i)} \in \mathbb{R}^m$, $|r^{(i)}| = 1$, $i = 1, 2, \ldots$, stopping for some i to deliver $z^{k+1} := ar^{(i)}r^{(i)T}$. To meet the relaxed null-step requirements, the method may stop when $ar^{(i)T}s^{k+1}r^{(i)} + \langle b, u^{k+1}\rangle > f_{\hat{u}}^k - \bar{\kappa}v_k$. If a descent step occurs, then $\epsilon_f^{k+1} = a\lambda_{\max}(s^{k+1}) - ar^{(i)T}s^{k+1}r^{(i)}$ may potentially determine the asymptotic error ϵ_f^∞ of (4.3). To ensure that ϵ_f^{k+1} is not "too large," we can employ additional stopping criteria based on upper estimates of $\lambda_{\max}(s^{k+1})$ generated as in [Nay05].

(iii) We may employ the following choice of the set Z_k due to [Nay99, Nay05]:

$$Z_k := \left\{ \sum_{j=1}^{\jmath} \nu_j \check{z}^j + pvp^T : \nu \in \mathbb{R}_+^{\jmath}, v \in S_+^r, \sum_{j=1}^{\jmath} \nu_j + \operatorname{tr} v = a \right\}, \tag{5.29}$$

where each $\check{z}^j \in \Sigma^m$ and p is an $m \times r$ orthonormal matrix. The resulting model

$$\check{f}_k(u) = a \max\Big\{ \max_{j=1:\jmath} \langle c - A^*u, \check{z}^j \rangle, \lambda_{\max}(p^T(c - A^*u)p) \Big\} + \langle b, u \rangle \tag{5.30}$$

attempts to strike a balance between being easy to handle (the polyhedral part) and accurate enough for fast convergence (the semidefinite part). Then the dual subproblem (5.17) can be cast as a conic optimization problem and handled by specialized solvers. Two efficient updates of Z_k satisfying (5.21) are given in [Nay99, sect. 4.4.2] (although they update AZ_k, they can update Z_k as well). For $\jmath = 1$, (5.29) reduces to the original choice of [HeR00]; again, (5.17) can be solved efficiently as a quadratic SDP [HeK02], and efficient updates of Z_k are given in [Hel03, HeK02].

(iv) For problem $(P_\leq)$ of (5.26), we can take $Z := a\Sigma_\leq^m$. Then (cf. (5.24)), $\lambda_{\max}^+$ replaces $\lambda_{\max}$ in (5.27), and we can take $r(u) := 0$ if $\lambda_{\max}(s(u)) < 0$, using (5.28) otherwise. We can thus stop an iterative eigenvalue computation whenever an upper bound indicates that $\lambda_{\max}(s(u)) < 0$. Of course, the final "=" in (5.29) is replaced by "$\leq$".

Acknowledgments. I would like to thank the Associate Editor, the two anonymous referees, and Claude Lemaréchal for helpful comments.

REFERENCES

[FGRS06] I. Fischer, G. Gruber, F. Rendl, and R. Sotirov, *Computational experience with a bundle approach for semidefinite cutting plane relaxations of Max-Cut and equipartition*, Math. Program., 105 (2006), pp. 451–469.

[HeK02] C. Helmberg and K. C. Kiwiel, *A spectral bundle method with bounds*, Math. Program., 93 (2002), pp. 173–194.

[Hel03] C. Helmberg, *Numerical evaluation of SBmethod*, Math. Program., 95 (2003), pp. 381–406.

[Hel04] C. Helmberg, *A cutting plane algorithm for large scale semidefinite relaxations*, The Sharpest Cut, The Impact of Manfred Padberg and His Work, M. Grötschel, ed., MPS-SIAM Ser. Optim. 4, SIAM, Philadelphia, 2004, pp. 233–256.

[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.

[Kiw86] K. C. Kiwiel, *A method for solving certain quadratic programming problems arising in nonsmooth optimization*, IMA J. Numer. Anal., 6 (1986), pp. 137–152.

[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.

[Kiw95] K. C. Kiwiel, *Approximations in proximal bundle methods and decomposition of convex programs*, J. Optim. Theory Appl., 84 (1995), pp. 529–548.

[Kiw99] K. C. Kiwiel, *A projection-proximal bundle method for convex nondifferentiable minimization*, in Ill-posed Variational Problems and Regularization Techniques, M. Théra and R. Tichatschke, eds., Lecture Notes in Econom. Math. Systems 477, Springer-Verlag, Berlin, 1999, pp. 137–150.

[Kiw05] K. C. Kiwiel, *An Inexact Bundle Approach to Cutting-Stock Problems*, Technical report, Systems Research Institute, Warsaw, 2005.

[Kiw06a] K. C. Kiwiel, *Inexact Dynamic Bundle Methods*, Technical report, Systems Research Institute, Warsaw, 2006.

[Kiw06b] K. C. Kiwiel, *A proximal bundle method with approximate subgradient linearizations*, SIAM J. Optim., 16 (2006), pp. 1007–1023.

[KRR99] K. C. Kiwiel, C. H. Rosa, and A. Ruszczyński, *Proximal decomposition via alternating linearization*, SIAM J. Optim., 9 (1999), pp. 668–689.

[Lem01] C. Lemaréchal, *Lagrangian relaxation*, in Computational Combinatorial Optimization, M. Jünger and D. Naddef, eds., Lecture Notes in Comput. Sci. 2241, Springer-Verlag, Berlin, 2001, pp. 112–156.

[LeO96] A. S. Lewis and M. L. Overton, *Eigenvalue optimization*, Acta Numer., 5 (1996), pp. 149–190.

[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.

[Nay99] M. V. Nayakkankuppam, *Optimization Over Symmetric Cones*, Ph.D. thesis, Department of Computer Science, New York University, New York, NY, 1999.

[Nay05] M. V. Nayakkankuppam, *Solving Large-scale Semidefinite Programs in Parallel*, Technical report, Department of Mathematics & Statistics, University of Maryland, Baltimore County, MD, 2005.

[ReS06] F. Rendl and R. Sotirov, *Bounds for the quadratic assignment problem using the bundle method*, Math. Program., (2006), to appear.

[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.