Mathematical Programming manuscript No.
(will be inserted by the editor)

Inexact Dynamic Bundle Methods

K.C. Kiwiel

Received: December 16, 2010 / Accepted: date

Abstract We give a proximal bundle method for minimizing a convex function f over $\mathbb{R}^n_+$. It requires evaluating f and its subgradients with a possibly unknown accuracy $\varepsilon \geq 0$, and maintains a set of free variables I to simplify its prox subproblems. 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 programs with exponentially many constraints, it adopts a relax-and-cut approach where the set I is extended only if a separation oracle finds a sufficiently violated constraint. In a simplified version, each iteration involves solving an unconstrained prox subproblem. For semidefinite programming problems, we extend the spectral bundle method to the case of weaker conditions on its optimization and separation oracles, and on the original primal problem.

Keywords Nondifferentiable optimization · Convex programming · Proximal bundle methods · Lagrangian relaxation · Semidefinite programming

Mathematics Subject Classification (2000) 65K05 · 90C25 · 90C27

1 Introduction

We consider the convex constrained minimization problem

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

where $C := \mathbb{R}^n_+$ is the nonnegative orthant in $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function.

We are interested in bundle methods, which at each trial point in C call an *oracle* to produce a *linearization* of f, given by a tuple in $\mathbb{R} \times \mathbb{R}^n$. At the current iteration k of such a method, the oracle has been called at trial points $u^1, \ldots, u^k$ in C, and has returned the corresponding tuples $\{(f_u^j, g^j)\}_{j=1}^k$ in $\mathbb{R} \times \mathbb{R}^n$. For an *exact* oracle,

K.C. Kiwiel
Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01–447 Warsaw, Poland, E-mail: kiwiel@ibspan.waw.pl

$f_u^j = f(u^j)$ and $g^j \in \partial f(u^j)$ denote the exact objective value and a subgradient at u^j. An *inexact* oracle may return $f_u^j = f(u^j) - \varepsilon_f^j$ and $g^j \in \partial_{\varepsilon_f^j} f(u^j)$ with *error* $\varepsilon_f^j \geq 0$, where $\partial_\varepsilon f(u) := \{g : f(\cdot) \geq f(u) - \varepsilon + \langle g, \cdot - u\rangle\}$ is the ε-subdifferential of f at u; in other words, it delivers the linearization

$$f_j(\cdot) := f_u^j + \langle g^j, \cdot - u^j\rangle \leq f(\cdot) \quad \text{with} \quad f_j(u^j) = f_u^j = f(u^j) - \varepsilon_f^j. \tag{1.2}$$

The errors are unknown, but bounded by some (unknown) constant $\varepsilon_f^{\max}$:

$$\varepsilon_f^j \leq \varepsilon_f^{\max} \quad \text{for all } j. \tag{1.3}$$

For instance, in many applications f is a max-type function of the form

$$f(u) := \sup\{F_z(u) : z \in Z\}, \tag{1.4}$$

where each $F_z : \mathbb{R}^n \to \mathbb{R}$ is convex and Z is an arbitrary set. If, for $u = u^j$, the oracle finds a possibly inexact maximizer $z^j \in Z$ of (1.4), sets $f_u^j := F_{z^j}(u^j)$ and takes g^j as any subgradient of F_{z^j} at u^j, then (1.2) holds with $\varepsilon_f^j = f(u^j) - F_{z^j}(u^j)$. An important special case arises in *Lagrangian relaxation* [8, Chap. XII], [21], where problem (1.1) with $C := \mathbb{R}_+^n$ is the Lagrangian dual of the primal problem

$$\sup \ \psi_0(z) \quad \text{s.t.} \quad \psi_i(z) \geq 0, \ i = 1, \ldots, n, \ z \in Z, \tag{1.5}$$

with $F_z(u) := \psi_0(z) + \langle u, \psi(z)\rangle$ for $\psi := (\psi_1, \ldots, \psi_n)$. Then, for each multiplier $u^j \geq 0$, picking $z^j \in Z$, we may take $f_u^j := F_{z^j}(u^j)$ and $g^j := \psi(z^j)$.

In [14] we extended the proximal bundle methods of [10] and [8, Sect. XV.3] to the inexact setting of (1.2); see [15,17,18] for further developments, [16,20] for constrained extensions, and [18,19] for numerical tests. Such methods maintain:

- a closed convex *model* $\check{f}_k \leq f$ (typically $\check{f}_k = \max_{j \in J_k} f_j$ with $J_k \subset \{1, \ldots, k\}$),
- a *stability center* $\hat{u}^k = u^{k'}$ for some $k' \leq k$ that has the value $f_{\hat{u}}^k = f_u^{k'}$, and
- a *proximity stepsize* $t_k > 0$ that controls the distance from $\hat{u}^k$ to the next trial point

$$u^{k+1} := \arg\min \left\{ \check{f}_k(u) + \tfrac{1}{2t_k}|u - \hat{u}^k|^2 : u \in C \right\}. \tag{1.6}$$

After the oracle called at u^{k+1} produces f_u^{k+1}, a *descent* step to $\hat{u}^{k+1} := u^{k+1}$ is taken if the objective reduction is at least a given fraction $\kappa \in (0,1)$ of the *predicted decrease*

$$v_k := f_{\hat{u}}^k - \check{f}_k(u^{k+1}), \tag{1.7}$$

i.e., if

$$f_u^{k+1} \leq f_{\hat{u}}^k - \kappa v_k. \tag{1.8}$$

Otherwise, a *null* step $\hat{u}^{k+1} := \hat{u}^k$ occurs; then the new linearization f_{k+1} is used to produce a better model $\check{f}_{k+1} \geq f_{k+1}$ for the next iteration. This summarizes exact bundle. The inexact extension of [14] is based on the observation that having the majorization

$$v_k \geq |u^{k+1} - \hat{u}^k|^2 / 2t_k \tag{1.9}$$

suffices for convergence. Hence, if necessary, t_k is increased and u^{k+1} is recomputed to decrease $\check{f}_k(u^{k+1})$ until (1.9) holds. As for convergence, [14] showed that the *asymptotic objective value* $f_{\hat{u}}^{\infty} := \lim_k f_{\hat{u}}^k$ estimates the optimal value f_* of (1.1): $f_{\hat{u}}^{\infty} \in [f_* - \varepsilon_f^{\max}, f_*]$. Next, [15, Sect. 4.2] observed that in fact the asymptotic accuracy depends only on the errors that occur at descent steps. Specifically, let $\ell(k) - 1$ index the last descent iteration prior to k, and denote by

$$\varepsilon_f^{\infty} := \overline{\lim_{k\to\infty}}\, \varepsilon_f^{\ell(k)} \tag{1.10}$$

the *asymptotic oracle error* at descent steps; then $f_{\hat{u}}^{\infty} \in [f_* - \varepsilon_f^{\infty}, f_*]$ (see [15]).

Since solving subproblem (1.6) can require much work for large n due to the complicating constraint $u \geq 0$, in this paper we augment the bundle method with an *active-set* strategy. At each iteration, we choose a (hopefully small) subset of *free* variables I_k, keeping the remaining variables fixed to simplify the bundle prox subproblem. Specifically, the *working set* $I_k \subset N := \{1\colon n\}$ and the complementary *fixed set* $I_k' := N \setminus I_k$ are such that the prox center $\hat{u}^k$ lies in the *restricted feasible set*

$$C_k := \left\{ u \in C : u_{I_k'} = 0 \right\} = \left\{ u \in \mathbb{R}_+^n : u_{I_k'} = 0 \right\}. \tag{1.11}$$

Replacing C by C_k in (1.6) yields the simplified bundle prox subproblem

$$u^{k+1} := \arg\min \left\{ \check{f}_k(u) + \tfrac{1}{2t_k}|u - \hat{u}^k|^2 : u \in C_k \right\}. \tag{1.12}$$

The set I_k is extended if necessary to ensure that our method asymptotically estimates the optimal value f_* with accuracy ε_f^{∞}. In Lagrangian relaxation, under standard convexity and compactness assumptions on problem (1.5) (see Sect. 5), it finds ε_f^{∞}-optimal primal solutions by combining partial Lagrangian solutions. These features are “inherited” from [14]. Yet, we give new results on *finite* convergence in the case where f is polyhedral and only finitely many different linearization are available. This covers the important case of LP relaxations of integer programming problems of the form (1.5) with ψ_0, ψ affine and Z finitely generated (see Rem. 6.3(7)).

We aim to keep the working set I_k as small as possible to ease the solution of subproblem (1.12). Further, extending the setting of [15] (which generalized the proximal-projection framework of [7, 13] to the inexact case), we give a simplified variant with *unconstrained* prox subproblems (see Sect. 8.1).

Our active-set strategy translates into a *relax-and-cut* approach for solving a convex problem of the form (1.5) with a huge number of constraints. The working set I_k gives the current set of dualized constraints, and our bundle subproblem needs only the corresponding components of the past subgradients (cf. (1.2)). The working set is extended if, for the current primal proposal $\hat{z}^k \in Z$, a *separation oracle* finds a *sufficiently violated* constraint $\psi_{\hat{\imath}}(\hat{z}^k) < -V_k$ with $\hat{\imath} \notin I_k$, where the threshold $V_k > 0$ is initially large, and diminishes as optimality is approached. This helps in keeping the set I_k as small as possible. In particular, even for exact evaluations ($\varepsilon_f^j \equiv 0$), our results improve on those in [1], where stronger requirements are imposed on the oracle and the primal problem (see Rem. 7.1). For semidefinite programming problems (SDP for short), we extend the cutting-plane spectral bundle method [6] to the case of weaker conditions on its optimization and separation oracles, and the original primal problem (see Rem. 8.4). Moreover, our results may serve as guidlines for possible improvement of the heuristic approaches of [4,5] and [2,22,23] (see Rem. 8.3), which have been highly successful in practice.

The paper is organized as follows. In Sect. 2 we present our method for general objective models. Its convergence is analyzed in Sect. 3. Various model choices are given in Sect. 4. Applications to Lagrangian relaxation are studied in Sect. 5. Finite termination in the polyhedral case is discussed in Sect. 6. Our relax-and-cut framework is exposed in Sect. 7. The proximal-projection variant and its SDP applications are discussed in Sect. 8.

Our notation is fairly standard. $|\cdot|$ is the Euclidean norm associated with the standard inner product $\langle\cdot,\cdot\rangle$ in $\mathbb{R}^n$, $|\cdot|_1$ is the l_1-norm, and $|\cdot|_\infty$ is the l_∞-norm. $P_C(u) := \arg\min_C |\cdot - u|$ is the *projector* onto C, and e_i is the ith unit vector in $\mathbb{R}^n$.

2 The inexact dynamic bundle method

Before stating our method, we summarize basic properties of subproblem (1.12).

2.1 Aggregate linearizations and an optimality estimate

We regard (1.1) as an unconstrained problem $f_* = \inf f_C$ with the *essential objective*

$$f_C := f + i_C, \tag{2.1}$$

where i_C is the *indicator* function of the set C ($i_C(u) = 0$ if $u \in C$, ∞ otherwise). We may then rewrite the trial point finding subproblem (1.12) as

$$u^{k+1} := \arg\min\left\{ \phi_k(\cdot) := \check{f}_k(\cdot) + i_{C_k}(\cdot) + \tfrac{1}{2t_k}|\cdot - \hat{u}^k|^2 \right\}. \tag{2.2}$$

Recall from Sect. 1 that $\check{f}_k \le f$ is closed convex, $t_k > 0$ and $\hat{u}^k \in C_k$ above; further, the stability center $\hat{u}^k = u^{\ell(k)}$ obtained at the last descent iteration $\ell(k) - 1$ prior to k has the value $f_{\hat{u}}^k = f_u^{\ell(k)}$, so that the oracle property (1.2) yields

$$f_{\hat{u}}^k = f(\hat{u}^k) - \varepsilon_f^{\ell(k)}. \tag{2.3}$$

We now exhibit aggregate linearizations (i.e., affine minorants) at u^{k+1} of the functions in subproblem (2.2), as well as an optimality estimate (see (2.12) below) over the current restricted set C_k.

Lemma 2.1 (1) *There exist subgradients $\hat{g}^k$ and ν^k such that*

$$\hat{g}^k \in \partial \check{f}_k(u^{k+1}), \quad \nu^k \in \partial i_{C_k}(u^{k+1}) \quad \text{and} \quad \hat{g}^k + \nu^k = (\hat{u}^k - u^{k+1})/t_k. \tag{2.4}$$

(2) *These subgradients determine the following three* aggregate linearizations *of the functions $\check{f}_k$ and f, i_{C_k}, $\check{f}_C^k := \check{f}_k + i_{C_k}$ and $f_{C_k} := f + i_{C_k}$, respectively*:

$$\bar{f}_k(\cdot) := \check{f}_k(u^{k+1}) + \langle \hat{g}^k, \cdot - u^{k+1}\rangle \le \check{f}_k(\cdot) \le f(\cdot), \tag{2.5}$$

$$\bar{\imath}_C^k(\cdot) := i_{C_k}(u^{k+1}) + \langle \nu^k, \cdot - u^{k+1}\rangle \le i_{C_k}(\cdot), \tag{2.6}$$

$$\bar{f}_C^k(\cdot) := \bar{f}_k(\cdot) + \bar{\imath}_C^k(\cdot) \le \check{f}_C^k(\cdot) \le f_{C_k}(\cdot). \tag{2.7}$$

(3) *For the* aggregate subgradient *and the* aggregate linearization error *given by*

$$p^k := \hat{g}^k + \nu^k = (\hat{u}^k - u^{k+1})/t_k \quad \textit{and} \quad \varepsilon_k := f_{\hat{u}}^k - \bar{f}_C^k(\hat{u}^k), \tag{2.8}$$

and the optimality measure

$$V_k := \max\big\{|p^k|, \varepsilon_k + \langle p^k, \hat{u}^k\rangle\big\}, \tag{2.9}$$

we have

$$\bar{f}_C^k(\cdot) = \check{f}_k(u^{k+1}) + \langle p^k, \cdot - u^{k+1}\rangle, \tag{2.10}$$

$$f_{\hat{u}}^k - \varepsilon_k + \langle p^k, \cdot - \hat{u}^k\rangle = \bar{f}_C^k(\cdot) \le \check{f}_C^k(\cdot) \le f_{C_k}(\cdot), \tag{2.11}$$

$$f_{\hat{u}}^k \le f_{C_k}(u) + V_k\big(1 + |u|\big) \quad \textit{for all } u. \tag{2.12}$$

Proof (1) Use the optimality condition $0 \in \partial\phi_k(u^{k+1})$ for subproblem (2.2).

(2) The inequalities in (2.5)–(2.6) stem from (2.4) as subgradient inequalities, and from our assumption that $\check{f}_k \le f$; note that $i_{C_k}(u^{k+1}) = 0$ in (2.6). Adding (2.5) and (2.6) gives (2.7).

(3) The first equalities in (2.5)–(2.8) yield (2.10); since $\bar{f}_C^k$ is affine, its expression in (2.11) follows from (2.8). Finally, since $|a||b| + c \le \max\{|a|, c\}(1 + |b|)$ for any scalars a, b, c, so in (2.11) for $a = |p^k|$ and $b = |u|$, by the Cauchy-Schwarz inequality,

$$-\langle p^k, u\rangle + \varepsilon_k + \langle p^k, \hat{u}^k\rangle \le |p^k||u| + \varepsilon_k + \langle p^k, \hat{u}^k\rangle \le \max\{|p^k|, \varepsilon_k + \langle p^k, \hat{u}^k\rangle\}(1 + |u|),$$

and we obtain (2.12) from the definition of V_k in (2.9). □

Note that V_k is indeed an optimality measure for the current working set: if $V_k = 0$ in (2.12), then $f_{\hat{u}}^k \le \inf f_{C_k}$; extending the estimate (2.12) to f_C is discussed below.

2.2 Ensuring sufficient predicted decrease

To ensure that the optimality measure V_k vanishes asymptotically, it is crucial to bound V_k via the predicted decrease v_k, since bundling and descent steps drive v_k to 0. This is done in parts (1) through (3) of Lemma 2.2 below; the remaining parts, which extend (2.12) to an optimality estimate (see (2.18) below) over the set C (instead of merely C_k), will suggest a suitable update of the working set I_k.

Lemma 2.2 (1) *In the notation of* (2.8)*, the predicted decrease* v_k *of* (1.7) *satisfies*

$$v_k = t_k|p^k|^2 + \varepsilon_k. \tag{2.13}$$

(2) *The optimality measure* V_k *of* (2.9) *satisfies* $V_k \le \max\{|p^k|, \varepsilon_k\}(1 + |\hat{u}^k|)$.

(3) *We have the equivalences*

$$v_k \ge -\varepsilon_k \;\Leftrightarrow\; t_k|p^k|^2/2 \ge -\varepsilon_k \;\Leftrightarrow\; v_k \ge t_k|p^k|^2/2 \;\Leftrightarrow\; v_k \ge |u^{k+1} - \hat{u}^k|^2/2t_k.$$

Moreover, $v_k \ge \varepsilon_k$*. Finally, for* $\varepsilon_f^{\ell(k)}$ *in* (2.3)*, we have* $-\varepsilon_k \le \varepsilon_f^{\ell(k)}$ *and*

$$v_k \ge \max\big\{t_k|p^k|^2/2, |\varepsilon_k|\big\} \qquad \textit{if} \quad v_k \ge -\varepsilon_k, \tag{2.14}$$

$$V_k \le \max\big\{(2v_k/t_k)^{1/2}, v_k\big\}\big(1 + |\hat{u}^k|\big) \quad \textit{if} \quad v_k \ge -\varepsilon_k, \tag{2.15}$$

$$V_k < \Big(2\varepsilon_f^{\ell(k)}/t_k\Big)^{1/2}\big(1 + |\hat{u}^k|\big) \qquad \textit{if} \quad v_k < -\varepsilon_k. \tag{2.16}$$

(4) *For the following* measure of nearness *of the subgradient* ν^k *of* (2.4) *to* $\partial i_C(u^{k+1})$

$$\zeta_k := \max\{-\hat{g}_i^k : i \in I_k^-\} \quad \text{with} \quad I_k^- := \{i \in I_k' : \hat{g}_i^k < 0\}, \tag{2.17}$$

we have

$$f_{\hat{u}}^k \le f(u) + \max\{V_k, \zeta_k\}\big(1 + |u|_1\big) \quad \text{for all } u \in C. \tag{2.18}$$

(5) *We have* $\nu^k \in \partial i_C(u^{k+1}) \Leftrightarrow \hat{g}_{I_k'}^k \ge 0 \Leftrightarrow I_k^- = \emptyset$*. Further, if* $I_k^- = \emptyset$*, then we can replace* C_k *by* C *in relations* (2.2)*,* (2.4)*,* (2.6)*,* (2.7)*,* (2.11) *and* (2.12)*.*

Proof (1) By (2.10) and (2.8), $\bar{f}_C^k(\hat{u}^k) = \check{f}_k(u^{k+1}) + t_k|p^k|^2$. Rewrite (1.7), using (2.8).

(2) Using the Cauchy-Schwarz inequality in (2.9), we have

$$V_k \le \max\{|p^k|, \varepsilon_k + |p^k||\hat{u}^k|\} \le \max\{|p^k|, \varepsilon_k\} + |p^k||\hat{u}^k| \le \max\{|p^k|, \varepsilon_k\}\big(1 + |\hat{u}^k|\big).$$

(3) The equivalences follow from (2.13); in particular, $v_k \ge \varepsilon_k$. Next, by (2.11) and (2.3) with $f_{C_k}(\hat{u}^k) = f(\hat{u}^k)$ ($\hat{u}^k \in C_k$), we have $-\varepsilon_k \le f_{C_k}(\hat{u}^k) - f_{\hat{u}}^k = \varepsilon_f^{\ell(k)}$. Finally, to obtain the bounds (2.14)–(2.16), use the equivalences together with the facts that $v_k \ge \varepsilon_k$, $-\varepsilon_k \le \varepsilon_f^{\ell(k)}$ and the bound on V_k from assertion (2). For instance, $v_k < -\varepsilon_k$ yields $0 \le t_k|p^k|^2/2 < -\varepsilon_k$ and $|p^k| < (-2\varepsilon_k/t_k)^{1/2}$ for (2.16).

(4) The indicator functions of the sets $C_k := \{u \in \mathbb{R}_+^n : u_{I_k'} = 0\}$ and $C := \mathbb{R}_+^n$ have the following subdifferentials at the point $u^{k+1} \in C_k \subset C$ (cf. [8, Ex. III.5.2.6(b)])

$$\partial i_{C_k}(u^{k+1}) = \{\nu \in \mathbb{R}^n : \nu_{I_k} \le 0, \langle \nu_{I_k}, u_{I_k}^{k+1}\rangle = 0\}, \tag{2.19}$$

$$\partial i_C(u^{k+1}) = \{\nu \in \mathbb{R}^n : \nu \le 0, \langle \nu, u^{k+1}\rangle = 0\}. \tag{2.20}$$

Since the points $\hat{u}^k, u^{k+1}$ in C_k satisfy $\hat{u}_{I_k'}^k = u_{I_k'}^{k+1} = 0$, the first part of (2.8) yields

$$p_{I_k'}^k = 0 \quad \text{and} \quad \nu_{I_k'}^k = -\hat{g}_{I_k'}^k. \tag{2.21}$$

Further, with $\nu^k \in \partial i_{C_k}(u^{k+1})$ by (2.4), (2.19) gives $\nu_{I_k}^k \le 0$ and $\langle \nu_{I_k}^k, u_{I_k}^{k+1}\rangle = 0$. Using the equality $u_{I_k'}^{k+1} = 0$, it then follows from (2.20) and (2.21) that the vector

$$\hat{\nu}^k := \min\{\nu^k, 0\} \quad \text{with} \quad \hat{\nu}_{I_k}^k = \nu_{I_k}^k \quad \text{and} \quad \hat{\nu}_{I_k'}^k = \min\{-\hat{g}_{I_k'}^k, 0\} \tag{2.22}$$

lies in $\partial i_C(u^{k+1})$; in fact, it is the projection of ν^k onto the set $\partial i_C(u^{k+1})$ described as

$$\partial i_C(u^{k+1}) = \{\nu \in \mathbb{R}^n : \nu \le 0, \langle \nu_{I_k}, u_{I_k}^{k+1}\rangle = 0\},$$

and we have $|\nu^k - \hat{\nu}^k|_\infty = \zeta_k$ if $\zeta_k > -\infty$ in (2.17), $\nu^k = \hat{\nu}^k$ otherwise. Now, since $\hat{\nu}^k \in \partial i_C(u^{k+1})$, the linearization $\hat{\imath}_C^k(\cdot) := \langle \hat{\nu}^k, \cdot - u^{k+1}\rangle \le i_C(\cdot)$ satisfies (cf. (2.6))

$$\hat{\imath}_C^k(\hat{u}^k) = \langle \hat{\nu}_{I_k}^k, \hat{u}_{I_k}^k - u_{I_k}^{k+1}\rangle + \langle \hat{\nu}_{I_k'}^k, \hat{u}_{I_k'}^k - u_{I_k'}^{k+1}\rangle = \langle \nu_{I_k}^k, \hat{u}_{I_k}^k - u_{I_k}^{k+1}\rangle = \bar{\imath}_C^k(\hat{u}^k)$$

with $\hat{u}_{I_k'}^k = u_{I_k'}^{k+1} = 0$, and letting $\hat{f}_C^k := \bar{f}_k + \hat{\imath}_C^k$ and $\hat{p}^k := \hat{g}^k + \hat{\nu}^k$, we obtain the estimate

$$f_{\hat{u}}^k - \varepsilon_k + \langle \hat{p}^k, \cdot - \hat{u}^k\rangle = \hat{f}_C^k(\cdot) \le \check{f}_k(\cdot) + i_C(\cdot) \le f_C(\cdot) \tag{2.23}$$

as in (2.7)–(2.11), using the equality $\hat{f}^k_C(\hat{u}^k) = \bar{f}^k_C(\hat{u}^k)$, a consequence of $\hat{\imath}^k_C(\hat{u}^k) = \bar{\imath}^k_C(\hat{u}^k)$. Next, since $\hat{p}^k := \hat{g}^k + \hat{\nu}^k$ and $p^k := \hat{g}^k + \nu^k$ by (2.8), (2.21) and (2.22) yield

$$\hat{p}^k_{I_k} = p^k_{I_k} \quad \text{and} \quad \hat{p}^k_{I'_k} = p^k_{I'_k} + \hat{\nu}^k_{I'_k} - \nu^k_{I'_k} = \min\{0, \hat{g}^k_{I'_k}\}.$$

Hence, we have $\langle \hat{p}^k, \hat{u}^k\rangle = \langle p^k, \hat{u}^k\rangle$ ($\hat{u}^k_{I'_k} = 0$) and $\langle \hat{p}^k, u\rangle = \langle p^k_{I_k}, u_{I_k}\rangle + \sum_{i\in I^-_k} \hat{g}^k_i u_i$ for any u by the definition of I^-_k in (2.17). In effect, it follows from (2.23) that

$$f^k_{\hat{u}} \le f(u) + \varepsilon_k + \langle p^k, \hat{u}^k\rangle - \langle p^k_{I_k}, u_{I_k}\rangle - \sum_{i\in I^-_k} \hat{g}^k_i u_i \quad \text{for all } u \in C. \tag{2.24}$$

Let $\eta_k := \max\{-p^k_i : i \in I_k\}$; then $\eta_k \le |p^k|$. For any $u \ge 0$, by our choice of η_k, we have $-\langle p^k_{I_k}, u_{I_k}\rangle \le \max\{\eta_k, 0\} \sum_{i\in I_k} u_i$; on the other hand, by (2.17), we also have $-\sum_{i\in I^-_k} \hat{g}^k_i u_i \le \max\{\zeta_k, 0\} \sum_{i\in I'_k} u_i$. Plugging both estimates into (2.24) gives

$$f^k_{\hat{u}} \le f(u) + \varepsilon_k + \langle p^k, \hat{u}^k\rangle + \max\{\eta_k, \zeta_k, 0\}|u|_1 \quad \text{for all } u \in C. \tag{2.25}$$

Finally, (2.25) and (2.9) with $V_k \ge |p^k| \ge \eta_k$ yield the desired conclusion.

(5) The equivalences follow from the proof of (4), where $\nu^k = \hat{\nu}^k$ iff $\hat{g}^k_{I'_k} \ge 0$. If $\nu^k \in \partial i_C(u^{k+1})$, then (2.4) implies that u^{k+1} solves (2.2) with C_k replaced by C; the rest follows by construction. □

The optimality estimate (2.18) involves both the optimality measure V_k and the nearness measure ζ_k. However, once we ensure that $\zeta_k \le V_k$ by extending the working set I_k if necessary, having $V_k \to 0$ will suffice for asymptotic optimality.

2.3 The method

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

Algorithm 2.3 (inexact dynamic bundle method)

Step 0 (*Initiation*). Select $u^1 \in C$, a *descent parameter* $\kappa \in (0,1)$, a *stepsize bound* $t_{\min} > 0$ and a *stepsize* $t_1 \ge t_{\min}$. Choose $I_1 \supset \{i \in N : u^1_i > 0\}$. Call the oracle at u^1 to obtain f^1_u and g^1 of (1.2), and set $\bar{f}_0 := f_1$. Set $\hat{u}^1 := u^1$, $f^1_{\hat{u}} := f^1_u$, $i^1_t := 0$, $\ell(1) = 1$ and $k := 1$.

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

$$\max\{\bar{f}_{k-1}, f_k\} \le \check{f}_k \le f. \tag{2.26}$$

Step 2 (*Trial point finding*). Find the solution u^{k+1} of subproblem (2.2). Set v_k by (1.7), $p^k := (\hat{u}^k - u^{k+1})/t_k$, $\varepsilon_k := v_k - t_k|p^k|^2$ and V_k by (2.9).

Step 3 (*Working set extension*). If $\zeta_k \le V_k$ for ζ_k given by (2.17), go to Step 4. Otherwise, pick a subset $\emptyset \ne \check{I}'_k \subset I'_k$, set $I_k := I_k \cup \check{I}'_k$ and go back to Step 2.

Step 4 (*Stopping criterion*). If $V_k = 0$, stop ($f^k_{\hat{u}} \le f_*$).

Step 5 (*Stepsize correction*). If $v_k < -\varepsilon_k$, set $t_k := 10t_k$, $i^k_t := k$ and return to Step 2.

Step 6 (*Oracle call and descent test*). Call the oracle at u^{k+1} to obtain f_u^{k+1} and g^{k+1} of (1.2). If the descent test (1.8) holds, set $\hat{u}^{k+1} := u^{k+1}$, $f_{\hat{u}}^{k+1} := f_u^{k+1}$, $i_t^{k+1} := 0$ and $\ell(k+1) := k+1$ (*descent step*); otherwise, set $\hat{u}^{k+1} := \hat{u}^k$, $f_{\hat{u}}^{k+1} := f_{\hat{u}}^k$, $i_t^{k+1} := i_t^k$ and $\ell(k+1) := \ell(k)$ (*null step*).

Step 7 (*Stepsize updating*). If $\ell(k+1) = k+1$ (i.e., after a descent step), select $t_{k+1} \geq t_{\min}$; 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 (*Working set updating*). If $\ell(k+1) = k+1$ (i.e., after a descent step), select I_{k+1} containing $\hat{I}_{k+1} := \{i \in N : \hat{u}_i^{k+1} > 0\}$; otherwise, set $I_{k+1} := I_k$.

Step 9 (*Loop*). Increase k by 1 and go to Step 1.

A few comments on the method are in order.

Remark 2.4 (1) Step 1 may choose the simplest model $\check{f}_k := \max\{\bar{f}_{k-1}, f_k\}$; more efficient choices are given in Sect. 4. If $\check{f}_{k+1}$ is chosen after Step 8, (2.26) becomes

$$\max\{\bar{f}_k, f_{k+1}\} \leq \check{f}_{k+1} \leq f. \tag{2.27}$$

(2) For a polyhedral model $\check{f}_k$, Step 2 may use the QP method of [11], which can efficiently solve sequences of related subproblems (2.2).

(3) Step 3 extends the working set I_k if necessary to maintain the crucial relation $\zeta_k \leq V_k$ for the optimality estimate (2.18). In fact, this relation is needed only when V_k diminishes, so Step 3 may be skipped if $V_k > \tau_V$ for a fixed $\tau_V > 0$. Thus, to keep the working set as small as possible, a classical *conservative* strategy is to wait until approximate optimality over the current set is detected ($V_k \leq \tau_V$), and then release the constraint $u_i = 0$ with the largest "wrong" multiplier $\nu_i^k > 0$ (setting $\check{I}'_k := \{i\}$).

(4) The stopping criterion of Step 4 is justified by the optimality estimate (2.18): $V_k = 0$ yields $f_{\hat{u}}^k \leq \inf f_C =: f_*$; thus, by (2.3), the point $\hat{u}^k$ is ε-*optimal* for $\varepsilon = \varepsilon_f^{\ell(k)}$, i.e., $f(\hat{u}^k) \leq f_* + \varepsilon$. Step 4 may stop if $V_k \leq \tau_{\text{opt}}$ for a tolerance $\tau_{\text{opt}} > 0$. Sect. 3 below will show that this stopping criterion will be met, unless $f_{\hat{u}}^k \to -\infty$.

(5) In the case of exact evaluations ($\varepsilon_f^j \equiv 0$), Step 5 is redundant, since $v_k \geq \varepsilon_k \geq 0$ by Lemma 2.2(3). When inexactness is discovered via the test $v_k < -\varepsilon_k$, the stepsize t_k is increased to produce a "safe" v_k for the descent test or confirm that the stability center $\hat{u}^k$ is already ε-optimal.

(6) At Step 6, we have $u^{k+1} \in C$ and $v_k > 0$ (by (2.15), since $V_k > 0$ after Step 4); hence Step 6 produces $\hat{u}^{k+1} \in C$ and $f_{\hat{u}}^{k+1} \leq f_{\hat{u}}^k$.

(7) Whenever t_k is increased at Step 5, the *stepsize indicator* $i_t^k \neq 0$ prevents Step 7 from decreasing t_k after null steps until the next descent step occurs (cf. Step 6). Otherwise, decreasing t_k at Step 7 aims at collecting more local information about the objective f at null steps.

(8) After a descent step, Step 8 allows us to reduce the working set to the minimal set $\hat{I}_{k+1}$, ensuring $\hat{u}^{k+1} \in C_{k+1}$ (cf. (1.11)). However, to prevent oscillations, I_{k+1} may also keep $i \in I_k$ if g_i^{k+1} is "significantly" negative relative to $\hat{g}_i^k$ or V_k (otherwise the constraint $u_i = 0$ may need to be released again), or if i has entered I_k recently.

3 Convergence

Our analysis splits into several cases.

3.1 The case of an infinite cycle due to oracle errors

Note that any cycle between Steps 2 and 3 must be finite, since the working set I_k can grow at Step 3 at most n times. For our purposes, it suffices to give a very simple result on cycles between Steps 2 and 5 (see [14, Lem. 2.3] for more).

Lemma 3.1 *If an infinite cycle between Steps* 2 *and* 5 *occurs, then* $f_{\hat{u}}^k \le f_*$, $f(\hat{u}^k) \le f_* + \varepsilon_f^{\ell(k)}$, *and* $V_k \to 0$.

Proof At Step 5 during the cycle (2.16) with $t_k \uparrow \infty$ gives $V_k \to 0$. Hence (2.18) with $\zeta_k \le V_k$ by Step 3 yields $f_{\hat{u}}^k \le \inf f_C =: f_*$, and the conclusion follows from (2.3). □

In view of Lemma 3.1, from now on we assume (unless stated otherwise) that the algorithm neither terminates nor cycles infinitely between Steps 2 and 5.

3.2 The case of finitely many descent steps

We now consider the case where only finitely many descent steps occur. After the last descent step, only null steps occur and the sequence $\{t_k\}$ becomes eventually monotone, since once Step 5 increases t_k, Step 7 cannot decrease t_k; thus the limit $t_\infty := \lim_k t_k$ exists. We first deal with the case of $t_\infty = \infty$.

Lemma 3.2 *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 the bound (2.16) with constant $\varepsilon_f^{\ell(k)} = \varepsilon_f^{\ell(\bar{k})}$; hence, $t_k \to \infty$ gives $V_k \xrightarrow[K]{} 0$. □

For the remaining case of $t_\infty < \infty$, we now recall from [18, Lem. 3.3] a fairly abstract result which shows that the *approximation errors*

$$\check{\gamma}_k := f_u^{k+1} - \check{f}_k(u^{k+1}) \tag{3.1}$$

vanish asymptotically when, instead of assuming that $\varepsilon_f^j \le \varepsilon_f^{\max}$ in (1.3) as in [14], we suppose that the oracle is *locally bounded* on C in the sense that

$$\text{the sequence } \{g^k\} \text{ is bounded whenever the sequence } \{u^k\} \subset C \text{ is bounded.} \tag{3.2}$$

Note that the former condition implies the latter, since for $\varepsilon = \varepsilon_f^{\max}$, the mapping $\partial_\varepsilon f$ is locally bounded (see, e.g., [8, Sect. XI.4.1]).

Lemma 3.3 *Suppose there exists* $\bar{k}$ *such that for all* $k \ge \bar{k}$, *we have* $\hat{u}^k = \hat{u}^{\bar{k}}$ *and* $t_{\min} \le t_{k+1} \le t_k$. *Further, assume that the oracle is locally bounded in the sense of* (3.2). *Then the approximation errors of* (3.1) *satisfy* $\overline{\lim}_k \check{\gamma}_k \le 0$. *Moreover, if the descent criterion* (1.8) *fails for all* $k \ge \bar{k}$, *then* $V_k \to 0$.

Proof First, using partial linearizations $\bar\phi_k$ of the objectives ϕ_k of consecutive subproblems (2.2), we show that their optimal values $\phi_k(u^{k+1})$ are nondecreasing and bounded above. Since the set I_k can grow at most n times at Step 3, we can assume that the sets I_k and C_k are constant for all $k \ge \bar k$.

Fix $k \ge \bar k$. By (2.10), (2.2) and (2.8), we have $\bar f^k_C(u^{k+1}) = \check f_k(u^{k+1})$ and

$$u^{k+1} = \arg\min\left\{ \bar\phi_k(\cdot) := \bar f^k_C(\cdot) + \tfrac{1}{2t_k}|\cdot - \hat u^k|^2 \right\} \tag{3.3}$$

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

$$\bar\phi_k(\cdot) = \phi_k(u^{k+1}) + \tfrac{1}{2t_k}|\cdot - u^{k+1}|^2. \tag{3.4}$$

Next, since $\bar f^k_C(\hat u^k) \le f(\hat u^k)$ by (2.7) with $\hat u^k \in C_k$, relations (3.4) and (3.3) above yield

$$\phi_k(u^{k+1}) + \tfrac{1}{2t_k}|u^{k+1} - \hat u^k|^2 = \bar\phi_k(\hat u^k) \le f(\hat u^k). \tag{3.5}$$

Now, the minorizations $\bar f_k \le \check f_{k+1}$ of (2.27), $\bar\imath^k_C \le i_{C_k}$ of (2.6) and the equality $C_{k+1} = C_k$ give $\bar f^k_C := \bar f_k + \bar\imath^k_C \le \check f_{k+1} + i_{C_{k+1}}$; since we also have $\hat u^{k+1} = \hat u^k$ and $t_{k+1} \le t_k$ (cf. Step 7), the objectives of (3.3) and the next subproblem (2.2) satisfy $\bar\phi_k \le \phi_{k+1}$. Hence by (3.4),

$$\phi_k(u^{k+1}) + \tfrac{1}{2t_k}|u^{k+2} - u^{k+1}|^2 = \bar\phi_k(u^{k+2}) \le \phi_{k+1}(u^{k+2}). \tag{3.6}$$

Thus the nondecreasing sequence $\{\phi_k(u^{k+1})\}_{k\ge\bar k}$, being bounded above by (3.5) with $\hat u^k = \hat u^{\bar k}$ for $k \ge \bar k$, must have a limit, say $\phi_\infty \le f(\hat u^{\bar k})$. Moreover, since the stepsizes satisfy $t_k \le t_{\bar k}$ for $k \ge \bar k$, we deduce from the bounds (3.5)–(3.6) that

$$\phi_k(u^{k+1}) \uparrow \phi_\infty, \quad u^{k+2} - u^{k+1} \to 0, \tag{3.7}$$

and the sequence $\{u^{k+1}\}$ is bounded. Then the sequence $\{g^k\}$ is bounded as well, since by our assumption the oracle is locally bounded in the sense of (3.2).

We now show that the approximation error $\check\gamma_k$ of (3.1) vanishes. Using the form (1.2) of f_{k+1}, the minorization $f_{k+1} \le \check f_{k+1}$ of (2.27), the Cauchy-Schwarz inequality, and the optimal values of subproblems (2.2) with $\hat u^k = \hat u^{\bar k}$ for $k \ge \bar k$, we estimate

$$\begin{aligned}\check\gamma_k &:= f_u^{k+1} - \check f_k(u^{k+1}) = f_{k+1}(u^{k+2}) - \check f_k(u^{k+1}) + \langle g^{k+1}, u^{k+1} - u^{k+2}\rangle \\ &\le \check f_{k+1}(u^{k+2}) - \check f_k(u^{k+1}) + |g^{k+1}||u^{k+1} - u^{k+2}| \\ &= \phi_{k+1}(u^{k+2}) - \phi_k(u^{k+1}) + \Delta_k + |g^{k+1}||u^{k+1} - u^{k+2}|,\end{aligned} \tag{3.8}$$

where $\Delta_k := |u^{k+1} - \hat u^{\bar k}|^2/2t_k - |u^{k+2} - \hat u^{\bar k}|^2/2t_{k+1}$. We have $\Delta_k \to 0$, since $t_{\min} \le t_{k+1} \le t_k$ for $k \ge \bar k$ by our assumption, $|u^{k+1} - \hat u^{\bar k}|^2$ is bounded, $u^{k+2} - u^{k+1} \to 0$ by (3.7), and thus

$$|u^{k+2} - \hat u^{\bar k}|^2 - |u^{k+1} - \hat u^{\bar k}|^2 = 2\langle u^{k+2} - u^{k+1}, u^{k+1} - \hat u^{\bar k}\rangle + |u^{k+2} - u^{k+1}|^2 \to 0.$$

Hence, using (3.7) and the boundedness of $\{g^{k+1}\}$ in (3.8) yields $\overline{\lim}_k \check\gamma_k \le 0$.

Next, if the descent test (1.8) fails for $k \ge \bar k$, then $f_u^{k+1} > f_{\hat u}^k - \kappa v_k$ gives

$$\check\gamma_k = \left[f_u^{k+1} - f_{\hat u}^k\right] + \left[f_{\hat u}^k - \check f_k(u^{k+1})\right] > -\kappa v_k + v_k = (1-\kappa)v_k \ge 0, \tag{3.9}$$

where $\kappa < 1$ by Step 0; we conclude that $\check\gamma_k \to 0$ and $v_k \to 0$. Finally, since $v_k \to 0$, $t_k \ge t_{\min}$ and $\hat u^k = \hat u^{\bar k}$ for $k \ge \bar k$ by our assumption, we have $V_k \to 0$ by (2.15). □

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

Lemma 3.4 *Suppose there exists $\bar{k}$ such that only null steps occur for all $k \geq \bar{k}$, and the oracle is locally bounded in the sense of* (3.2)*. Let $K := \{k : t_{k+1} > t_k\}$ if $t_k \to \infty$, $K := \mathbb{N}$ otherwise. Then $V_k \xrightarrow[K]{} 0$.*

Proof Steps 5, 6 and 7 ensure that the sequence $\{t_k\}$ is monotone for large k. We have $V_k \xrightarrow[K]{} 0$ from either Lemma 3.2 if $t_\infty = \infty$, or Lemma 3.3 if $t_\infty < \infty$. □

3.3 The case of infinitely many descent steps

Lemma 3.5 *Suppose the set $\mathscr{D}$ of descent iterations is infinite and $f_{\hat{u}}^\infty := \lim_k f_{\hat{u}}^k > -\infty$. Then $\underline{\lim}_{k\in\mathscr{D}} V_k = 0$ and $|p^k| \xrightarrow[\mathscr{D}]{} 0$. Moreover, if $\{\hat{u}^k\}$ is bounded, then $V_k \xrightarrow[\mathscr{D}]{} 0$.*

Proof We have $0 \leq \kappa v_k \leq f_{\hat{u}}^k - f_{\hat{u}}^{k+1}$ if $k \in \mathscr{D}$, $f_{\hat{u}}^{k+1} = f_{\hat{u}}^k$ otherwise (see Step 6). Thus $\sum_{k\in\mathscr{D}} \kappa v_k \leq f_{\hat{u}}^1 - f_{\hat{u}}^\infty < \infty$ gives $v_k \xrightarrow[\mathscr{D}]{} 0$ and hence $\varepsilon_k, t_k|p^k|^2 \xrightarrow[\mathscr{D}]{} 0$ by (2.14) and $|p^k| \xrightarrow[\mathscr{D}]{} 0$, using $t_k \geq t_{\min}$ (cf. Step 7). For $k \in \mathscr{D}$, $\hat{u}^{k+1} - \hat{u}^k = -t_k p^k$ by (2.8), so that

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

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

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

(since otherwise $|\hat{u}^k|^2 \to -\infty$, which is impossible). Combining this with $t_k|p^k|^2 \xrightarrow[\mathscr{D}]{} 0$ gives $\underline{\lim}_{k\in\mathscr{D}}\langle p^k, \hat{u}^k\rangle \leq 0$. Since also $\varepsilon_k, |p^k| \xrightarrow[\mathscr{D}]{} 0$, we have $\underline{\lim}_{k\in\mathscr{D}} V_k = 0$ by (2.9).

If $\{\hat{u}^k\}$ is bounded, using $\varepsilon_k, |p^k| \xrightarrow[\mathscr{D}]{} 0$ in Lemma 2.2(2) gives $V_k \xrightarrow[\mathscr{D}]{} 0$. □

3.4 Synthesis

Our principal result on the asymptotic objective value $f_{\hat{u}}^\infty := \lim_k f_{\hat{u}}^k$ follows.

Theorem 3.6 *Suppose Algorithm* 2.3 *neither terminates nor loops infinitely between Steps* 2 *and* 5 (*so that $k \to \infty$*)*, the oracle is locally bounded in the sense of* (3.2)*, and its asymptotic error ε_f^∞ of* (1.10) *is finite. Let $K := \mathbb{N}$ if $f_{\hat{u}}^\infty := \lim_k f_{\hat{u}}^k = -\infty$; otherwise, let K be such that $V_k \xrightarrow[K]{} 0$* (*such K exists by Lemmas* 3.4 *and* 3.5)*. Then*:

(1) *$f_* \leq \underline{\lim}_k f(\hat{u}^k) \leq \overline{\lim}_k f(\hat{u}^k) = f_{\hat{u}}^\infty + \varepsilon_f^\infty$, where f_* is the optimal value of* (1.1).

(2) *We have $f_{\hat{u}}^k \downarrow f_{\hat{u}}^\infty \leq f_*$, and additionally $V_k \xrightarrow[K]{} 0$ if $f_* > -\infty$.*

Proof (1) For all k, we have $\hat{u}^k \in C$ and $f_* := \inf_C f \leq f(\hat{u}^k) = f_{\hat{u}}^k + \varepsilon_f^{\ell(k)}$ by (2.3). Pass to the limit, with $f_{\hat{u}}^k$ converging to $f_{\hat{u}}^\infty$, and $\varepsilon_f^\infty < \infty$ in (1.10) by our assumption.

(2) By (1), if $f_{\hat{u}}^\infty = -\infty$, then $f_* = -\infty$. Hence, suppose that $f_* > -\infty$. Then $f_{\hat{u}}^\infty \geq f_* - \varepsilon_f^\infty > -\infty$ by (1), so Lemmas 3.4 and 3.5 guarantee the existence of K such that $V_k \xrightarrow[K]{} 0$. Pass to the limit in (2.18) with $\zeta_k \leq V_k$ to obtain $f_{\hat{u}}^\infty \leq \inf f_C = f_*$. □

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

Remark 3.7 (1) Inspection of the preceding proofs reveals that Theorem 3.6 requires only convexity and finiteness of f on C, and the local boundedness condition (3.2). If f is finite convex on a neighborhood of C, then again (1.3) implies (3.2). On the other hand, (3.2) holds in the min-max setting of (1.4) if $\partial F_z(\cdot)$ is locally bounded on C, uniformly w.r.t. $z \in Z$; e.g., when Z is finite.

(2) Concerning Lemma 3.5, note that the sequence $\{\hat{u}^k\}$ is bounded if f_C is coercive and $\varepsilon_f^\infty < \infty$ in (1.10), since then the level set $\{u \in C : f(u) \le f_{\hat{u}}^1 + \varepsilon_f^{\sup}\}$ is bounded for $\varepsilon_f^{\sup} := \sup_k \varepsilon_f^{\ell(k)} < \infty$ and contains $\{\hat{u}^k\}$ by (2.3).

(3) What would happen in Theorem 3.6 if Step 3 failed to detect $\zeta_k > V_k$ at some iterations k in the set K such that $V_k \xrightarrow[K]{} 0$? Then assertion (1) still holds (by its proof), but in assertion (2), instead of $f_{\hat{u}}^\infty \le f_*$, (2.18) gives the optimality estimate

$$f_{\hat{u}}^\infty \le f(u) + \max\{\zeta_\infty, 0\}\big(1 + |u|_1\big) \quad \text{for all } u \in C, \tag{3.10}$$

where $\zeta_\infty := \underline{\lim}_{k \in K} \zeta_k$ is the asymptotic "detection" error of Step 3.

4 Linearization accumulation, selection and aggregation

For future use, we recall from [15, Sect. 4.4] that there are three basic choices of polyhedral models satisfying (2.26) or equivalently (2.27).

First, *accumulation* takes $\check{f}_k := \max_{j=1}^k f_j$ as the richest model.

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

$$\check{f}_k(\cdot) := \max_{j \in J_k} f_j(\cdot) \quad \text{with} \quad k \in J_k \subset \{1, \dots, k\}. \tag{4.1}$$

Since $\hat{g}^k \in \partial \check{f}_k(\check{u}^{k+1})$ by (2.4) and each f_j is affine in (4.1), there exist *convex weights* α_j^k, $j \in J_k$, such that (cf. [8, Ex. VI.3.4])

$$(\hat{g}^k, 1) = \sum_{j \in J_k} \alpha_j^k (\nabla f_j, 1),\ \alpha_j^k \ge 0,\ \alpha_j^k \big[\check{f}_k(u^{k+1}) - f_j(u^{k+1})\big] = 0,\ j \in J_k, \tag{4.2}$$

and using relation (2.5), it is easy to obtain the following expansion

$$(\bar{f}_k, 1) = \sum_{j \in \hat{J}_k} \alpha_j^k (f_j, 1) \quad \text{with} \quad \hat{J}_k := \{j \in J_k : \alpha_j^k > 0\}. \tag{4.3}$$

Since $\bar{f}_k \le \max_{j \in \hat{J}_k} f_j$, to meet the requirement (2.27) it suffices to choose

$$J_{k+1} \supset \hat{J}_k \cup \{k+1\}. \tag{4.4}$$

Active-set methods for solving subproblem (2.2) [9,11] find multipliers α_j^k such that $|\hat{J}_k| \le |I_k| + 1$. 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 (4.1) by

$$\check{f}_k(\cdot) := \max_{j \in J_k} f_j(\cdot) \quad \text{with} \quad k \in J_k \subset \{1-k\colon k\}, \quad f_j := \bar{f}_{-j} \text{ for } j \le 0. \tag{4.5}$$

The weights α_j^k of (4.2) produce $f_{-k} := \bar{f}_k$ via (4.3), and relation (4.4) is replaced by $J_{k+1} \supset \{-k, k+1\}$, so that only $\bar{n} \ge 2$ linearizations may be kept.

5 Lagrangian relaxation

5.1 The primal problem

Let $\mathscr{Z}$ be a real inner-product space with a finite dimension $\bar{m}$. 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 $\mathscr{Z}$:

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

where $\emptyset \neq Z \subset \mathscr{Z}$ 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. For the *dual function*

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

at each $u = u^j$, the oracle picks an approximate maximizer $z^j \in Z$ of (5.2) and delivers

$$f_u^j := \psi_0(z^j) + \langle u^j, \psi(z^j) \rangle \quad \text{and} \quad g^j := \psi(z^j). \tag{5.3}$$

Note that $\psi(Z)$ is bounded if $\inf_Z \min_{i=1}^n \psi_i > -\infty$, or the constraint function ψ is continuous on Z; then the ε_f^j-subgradients g^j are bounded.

5.2 Primal recovery with selection

We first consider our method with linearization selection (cf. Sect. 4).

The partial Lagrangian solutions z^j in (5.3) and their constraint values $g^j := \psi(z^j)$ determine the linearizations (1.2) as Lagrangian pieces of f in (5.2):

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

Using their weights $\{\alpha_j^k\}_{j \in J_k}$ (cf. (4.3)), we may estimate a solution to (5.1) via the *aggregate primal solution*

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

this convex combination is associated with the aggregate linearization $\bar{f}_k$ via

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

We now derive useful bounds on $\psi_0(\hat{z}^k)$ and $\psi(\hat{z}^k)$, generalizing [14, Lem. 5.1].

Lemma 5.1 *$\hat{z}^k \in Z$, $\psi_0(\hat{z}^k) \geq f_{\hat{u}}^k - \varepsilon_k - \langle p^k, \hat{u}^k \rangle$ and $\psi_{I_k}(\hat{z}^k) \geq \hat{g}_{I_k}^k \geq p_{I_k}^k$. Moreover,* $\min\{\psi_i(\hat{z}^k) : i \in I_k'\} \geq -\max\{\zeta_k, 0\}$ *for ζ_k given by* (2.17).

Proof By (5.6), $\hat{z}^k \in \operatorname{co}\{z^j\}_{j\in\hat{J}_k} \subset Z$, $\psi_0(\hat{z}^k) \geq \sum_j \alpha_j^k \psi_0(z^j)$, $\psi(\hat{z}^k) \geq \sum_j \alpha_j^k \psi(z^j)$ by convexity of Z and concavity of ψ_0, ψ. Using (5.6) with $\hat{g}^k = \nabla \bar{f}_k$ by (2.5) and (5.4) with $\nabla f_j = \psi(z^j)$, we get $\bar{f}_k(0) = \sum_j \alpha_j^k \psi_0(z^j)$ and $\hat{g}^k = \sum_j \alpha_j^k \psi(z^j)$. Since (2.4) and (2.19) imply $v_{I_k}^k \leq 0$ and $\langle v_{I_k}^k, u_{I_k}^{k+1}\rangle = 0$, we have $\hat{g}_{I_k}^k = p_{I_k}^k - v_{I_k}^k \geq p_{I_k}^k$ by (2.8), whereas $\min_{i\in I_k'} \hat{g}_i^k \geq -\max\{\zeta_k, 0\}$ by (2.17). Next, the equality $\bar{\imath}_C^k(0) = -\langle v^k, u^{k+1}\rangle$ due to (2.6) together with $\langle v_{I_k}^k, u_{I_k}^{k+1}\rangle = 0$ and $u_{I_k'}^{k+1} = 0$ give $\bar{\imath}_C^k(0) = 0$. Since $\bar{f}_k(0) = \bar{f}_C^k(0) - \bar{\imath}_C^k(0)$ by (2.7), we obtain that $\bar{f}_k(0) = \bar{f}_C^k(0) = f_{\hat{u}}^k - \varepsilon_k - \langle p^k, \hat{u}^k\rangle$ by (2.11). Combining the preceding relations yields the conclusion. □

In terms of the optimality measure V_k of (2.9), the bounds of Lemma 5.1 imply

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

Theorem 5.2 *Under the primal assumptions of Sect.* 5.1*, suppose Algorithm* 2.3 *neither terminates nor loops infinitely between Steps* 2 *and* 5 (*so that* $k \to \infty$)*, the oracle is locally bounded in the sense of* (3.2) (*e.g.,* $\psi(Z)$ *is bounded*)*, and its asymptotic error* ε_f^∞ *of* (1.10) *is finite. Then either* $f_{\hat{u}}^k \downarrow f_* = -\infty$*, in which case the primal problem* (5.1) *is infeasible, or* $f_* > -\infty$ *and* $f_{\hat{u}}^k \downarrow f_{\hat{u}}^\infty \in [f_* - \varepsilon_f^\infty, f_*]$*. In the latter case, let* $K \subset \mathbb{N}$ *be such that* $V_k \xrightarrow[K]{} 0$ *as described in Theorem* 3.6*. Then*:

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

(2) *Each cluster point* $\hat{z}^\infty$ *of* $\{\hat{z}^k\}_{k\in K}$ *lies in the set of* ε_f^∞-optimal primal solutions

$$Z_\varepsilon := \{\, z \in Z : \psi_0(z) \geq \psi_0^{\max} - \varepsilon, \psi(z) \geq 0 \,\} \quad \textit{for} \quad \varepsilon = \varepsilon_f^\infty. \tag{5.8}$$

(3) $d_{Z_{\varepsilon_f^\infty}}(\hat{z}^k) := \inf_{z\in Z_{\varepsilon_f^\infty}} |\hat{z}^k - z| \xrightarrow[K]{} 0$.

Proof The first assertion follows from Theorem 3.6, since $f_* = -\infty$ implies primal infeasibility by weak duality. In the second case, using $f_{\hat{u}}^k \downarrow f_{\hat{u}}^\infty \geq f_* - \varepsilon_f^\infty$ and $V_k \xrightarrow[K]{} 0$ with $\zeta_k \leq V_k$ in (5.7) yields $\underline{\lim}_{k\in K} \psi_0(\hat{z}^k) \geq f_* - \varepsilon_f^\infty$ and $\underline{\lim}_{k\in K} \min_{i=1}^n \psi_i(\hat{z}^k) \geq 0$.

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

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

(3) This follows from (1), (2) and the continuity of the distance function $d_{Z_{\varepsilon_f^\infty}}$. □

Remark 5.3 (1) By the proofs of Lemma 3.1 and Theorem 5.2, if an infinite cycle between Steps 2 and 5 occurs, then $V_k \to 0$ yields $d_{Z_\varepsilon}(\hat{z}^k) \to 0$ for $\varepsilon = \varepsilon_f^{\ell(k)}$. Similarly, if Step 4 terminates with $V_k = 0 \geq \zeta_k$, then $\hat{z}^k \in Z_\varepsilon$.

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

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

Then $\psi_0(\hat{z}^k) \geq \psi_0^{\max} - \varepsilon_f^{\ell(k)} - \varepsilon_{\text{tol}}$ from $f_{\hat{u}}^k \geq f_* - \varepsilon_f^{\ell(k)}$ (cf. (2.3)) and $f_* \geq \psi_0^{\max}$ (weak duality), so that the point $\hat{z}^k \in Z$ is an approximate primal solution. This stopping criterion will be satisfied for some k if $f_* > -\infty$ (cf. (5.7) and Theorem 5.2).

(3) What would happen in Theorem 5.2 if Step 3 failed to detect $\zeta_k > V_k$ at some iterations $k \in K$ with $V_k \xrightarrow[K]{} 0$? Then, by Rem. 3.7(3), instead of $f_{\hat{u}}^{\infty} \le f_*$, we only have the estimate (3.10), but assertions (2) and (3) still hold with Z_ε replaced by

$$Z_{\varepsilon_f^\infty, \bar{\zeta}_\infty} := \left\{ z \in Z : \psi_0(z) \ge \psi_0^{\max} - \varepsilon_f^\infty, \psi(z) \ge -\max[\bar{\zeta}_\infty, 0] \right\}, \tag{5.9}$$

where $\bar{\zeta}_\infty := \overline{\lim}_{k\in K} \zeta_k$ (since $\underline{\lim}_{k\in K} \min_{i=1}^n \psi_i(\hat{z}^k) \ge -\max\{0, \bar{\zeta}_\infty\}$ by (5.7)). Thus, only asymptotic primal feasibility is affected; this is the best we can hope for.

5.3 Primal recovery with aggregation

Let us now consider the variant with aggregation based on (4.5), 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 $j < 0$, $f_0 := f_1$ and $z^0 := z^1$. For the convex weights α_j^k satisfying (4.3) and $\hat{z}^k$ given by (5.6), there are convex weights $\bar{\alpha}_j^k$ such that (cf. [15, Sect. 5.3])

$$(f_{-k}, z^{-k}, 1) := (\bar{f}_k, \hat{z}^k, 1) = \sum_{0\le j\le k} \bar{\alpha}_j^k (f_j, z^j, 1) \quad \text{with} \quad \bar{\alpha}_j^k \ge 0,\ j = 0, \ldots, k. \tag{5.10}$$

Replacing (5.6) by (5.10) for Lemma 5.1, we conclude that the preceding convergence results remain valid.

6 Finite termination in the polyhedral case

We now describe a specialized version of Algorithm 2.3 that terminates with $f_{\hat{u}}^k \le f_*$ when the objective f is polyhedral, $f_* > -\infty$ and the oracle can deliver only *finitely many different* linearizations. In addition, we assume that:

(a) the algorithm employs linearization selection (cf. (4.1)–(4.4));
(b) Step 6 uses $\kappa = 1$ in the descent test (1.8);
(c) the oracle's errors $\varepsilon_f^{\ell(k)}$ at descent steps are bounded if $\ell(k) \to \infty$;
(d) Step 7 sets $t_{k+1} = t_k$ after a null step;
(e) Step 5 is replaced by the following step.

Step 5_1 (*Stepsize correction*). If $v_k \ge -\varepsilon_k$, go to Step 6. Set $t_k := 10t_k$ and $i_t^k := k$. If $f_{\hat{u}}^k > \check{f}_k(u^{k+1})$, go back to Step 2. If $u^{k+1} \in \operatorname{Arg\,min}_C \check{f}_k$, stop. Otherwise, pick a subset $\check{I}_k' \subset I_k'$, $\check{I}_k' \ne \emptyset$ unless $I_k = N$, set $I_k := I_k \cup \check{I}_k'$ and go back to Step 2.

We shall need the following elementary result.

Lemma 6.1 *If the function $a : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is convex and polyhedral, then there exists $\varepsilon_a > 0$ such that $p \in \partial a(u)$ and $|p| \le \varepsilon_a$ imply $u \in \operatorname{Arg\,min} a$.*

Theorem 6.2 *Under the above assumptions, the modified algorithm terminates.*

Proof If the algorithm terminates at Step 5_1, then $f_{\hat{u}}^k \le \check{f}_k(u^{k+1}) = \min_C \check{f}_k \le f_*$ from $\check{f}_k \le f$ (cf. (2.26)). For contradiction, suppose the algorithm doesn't terminate.

First, suppose there is $\bar{k}$ such that only null steps occur for all $k \ge \bar{k}$, and $t_\infty = \infty$. As in the proof of Lemma 3.2, at Step 5_1 we have $V_k \xrightarrow[K]{} 0$ for $K := \{k \ge \bar{k} : t_{k+1} > t_k\}$. Hence $p^k \xrightarrow[K]{} 0$ by (2.9), where $p^k \in \partial \check{f}_C^k(u^{k+1})$ by (2.10)–(2.11). Since the number of different $\check{f}_C^k$-s is finite (each $\check{f}_k$ stems from the finitely many oracle's linearizations, and the C_k-s stem from finitely many different I_k-s), we deduce from Lemma 6.1 the existence of $\hat{k}$ such that for each $k \ge \hat{k}$, $k \in K$, the point u^{k+1} minimizes $\check{f}_C^k$:

$$u^{k+1} \in \check{U}_k := \operatorname{Arg}\min_{C_k} \check{f}_k \quad \text{and} \quad \check{f}_k(u^{k+1}) = \check{f}_*^k := \min_{C_k} \check{f}_k. \tag{6.1}$$

For such k, we have $u^{k+1} = P_{\check{U}_k}(\hat{u}^{\bar{k}})$ by (2.2) with $\hat{u}^k = \hat{u}^{\bar{k}}$. Thus, since there are finitely many different $\check{U}_k$-s, there is $r < \infty$ such that $|u^{k+1} - \hat{u}^{\bar{k}}| \le r$ for all such k. Then (2.11) with $\varepsilon_k < 0$ (from $v_k < -\varepsilon_k$ and Lem. 2.2(3)) yield $\langle p^k, u^{k+1} - \hat{u}^k \rangle \le \check{f}_*^k - f_{\hat{u}}^k$ and hence $|p^k| r \ge f_{\hat{u}}^{\bar{k}} - \check{f}_*^k$ by the Cauchy-Schwarz inequality. Now, if Step 5_1 goes back to Step 2 due to $f_{\hat{u}}^k > \check{f}_k(u^{k+1})$ with $f_{\hat{u}}^k = f_{\hat{u}}^{\bar{k}}$ and $\check{f}_k(u^{k+1}) = \check{f}_*^k$ for infinitely many $k \in K$, then, since there are finitely many different $\check{f}_*^k$-s, there is $\varepsilon > 0$ such that $|p^k| r \ge f_{\hat{u}}^{\bar{k}} - \check{f}_*^k \ge \varepsilon$ for such k, which contradicts $p^k \xrightarrow[K]{} 0$. Hence Step 5_1 must eventually produce $I_k = N$, and then for $C_k = C$ above, $u^{k+1} \in \operatorname{Arg}\min_C \check{f}_k$ implies termination, a contradiction. (The same argument shows that any cycle between Steps 2 and 5_1 must be finite.)

Second, suppose there exists $\bar{k}$ such that for all $k \ge \bar{k}$, only null steps occur and Step 5_1 doesn't increase t_k. Then in the proof of Lemma 3.3, we have $\phi_k(u^{k+1}) < \phi_{k+1}(u^{k+2})$ for $k \ge \bar{k}$ (otherwise (3.6) yields $\phi_k(u^{k+1}) = \phi_{k+1}(u^{k+2})$ and $u^{k+2} = u^{k+1}$, so (3.8) with $\Delta_k = 0$ due to $t_{k+1} = t_k$ gives $\check{\gamma}_k \le 0$, which contradicts (3.9)). However, since there are finitely many different $\check{f}_C^k$-s and $t_k = t_{\bar{k}}$ for $k \ge \bar{k}$, there are finitely many optimal values $\phi_k(u^{k+1})$ of (2.2), so they can't keep increasing. This contradiction, combined with the first case above and the fact that $\{t_k\}$ becomes eventually monotone during a series of null steps, show that the number of null steps must be finite.

Finally, suppose infinitely many descent steps occur. Since our assumptions $f_* > -\infty$ and $\sup_k \varepsilon_f^{\ell(k)} < \infty$ yield $f_{\hat{u}}^\infty > -\infty$ by (2.3), Lemma 3.5 gives $p^k \xrightarrow[\mathscr{D}]{} 0$, so as in the first case above we have $\check{f}_k(u^{k+1}) = \check{f}_*^k$ for all large $k \in \mathscr{D}$, with finitely many different $\check{f}_*^k$-s. However, for such k, (1.8) with $\kappa = 1$ yields $f_{\hat{u}}^{k+1} \le \check{f}_k(u^{k+1}) = \check{f}_*^k < f_{\hat{u}}^k$ by (1.7), which implies that there are infinitely many different $\check{f}_*^k$-s, a contradiction. □

We now extend Theorem 6.2 to several algorithmic variations.

Remark 6.3 (1) First, Step 1 can employ finitely many aggregations. Second, at Step 6, it suffices if only finitely many descent steps with $\kappa < 1$ in (1.8) occur between two descent steps with $\kappa = 1$. Third, Step 7 can decrease t_k finitely many times whenever null steps occur. Fourth, Step 5_1 may replace the test "$f_{\hat{u}}^k > \check{f}_k(u^{k+1})$" by "$f_{\hat{u}}^k > \check{f}_k(u^{k+1})$ or $|p^k| > \tau_p$ or $u^{k+1} \notin \operatorname{Arg}\min_{C_k} \check{f}_k$" for a fixed tolerance $\tau_p > 0$. Note that, having solved (2.2), we can check if $u^{k+1} \in \operatorname{Arg}\min_{C_k} \check{f}_k$ via parametric QP [12]. If this test is employed, then, instead of checking if $u^{k+1} \in \operatorname{Arg}\min_C \check{f}_k$ at each execution of Step 5_1, we may wait until $C_k = C$. On the other hand, if the test

$u^{k+1} \in \operatorname{Arg\,min}_C \check{f}_k$ is omitted, then by the proof of Theorem 6.2, either Step 4 terminates, or eventually Step 5_1 yields $t_k \to \infty$ and $f_{\hat{u}}^k \le f_*$ from $V_k \to 0$ in Lemma 3.1 or $V_k \xrightarrow[K]{} 0$ in Lemma 3.2; hence in practice, Step 5_1 may stop once V_k is small enough.

(2) Theorem 6.2 still holds if Step 6 uses the test (1.8) with $\kappa < 1$ for most descent steps, but from time to time, the next descent step occurs only if (1.8) holds with either κ replaced by 1 or $\check{U}_k = \{u^{k+1}\}$ in (6.1) (again, this can be checked via parametric QP [12]). Indeed, there are finitely many different $\check{U}_k$-s, each $\hat{u}^{k+1}$ can have finitely many different linearization values $f_{\hat{u}}^{k+1}$, and $f_{\hat{u}}^{k+1} < f_{\hat{u}}^k$ at descent steps, so the number of descent steps with $\check{U}_k = \{u^{k+1}\}$ must be finite.

(3) In (2) above, condition $\check{U}_k = \{u^{k+1}\}$ can be replaced by $\mathscr{E}_k = \{(u^{k+1}, \check{v}_k)\}$, where for $\check{v}_k := \check{f}_k(u^{k+1})$, $\bar{J}_k := \{j \in J_k : f_j(u^{k+1}) = \check{v}_k\}$ and $\bar{I}_k := \{i \in I_k : u_i^{k+1} = 0\}$,

$$\mathscr{E}_k := \left\{(u, \check{v}) \in \mathbb{R}^{n+1} : f_j(u) - \check{v} = 0, j \in \bar{J}_k,\ u_i = 0, i \in \bar{I}_k \cup I'_k\right\}. \tag{6.2}$$

Clearly, $\mathscr{E}_k = \{(u^{k+1}, \check{v}_k)\}$ iff there are $n+1$ constraints in (6.2) with linearly independent gradients. In particular, $\mathscr{E}_k = \{(u^{k+1}, \check{v}_k)\}$ if for $\hat{J}_k$ given by (4.3) and $\hat{I}_k := \{i \in I_k : \nu_j^k \neq 0\}$, the vectors $\{(\nabla f_j, -1)\}_{j \in \hat{J}_k}$, $\{(e_i, 0)\}_{i \in \hat{I}_k \cup I'_k}$ are linearly independent (as in active-set methods for solving (2.2) [9,11,25]) and $|\hat{J}_k| + |\hat{I}_k| + |I'_k| = n+1$. This modification generalizes the test of [25, Alg. 1] even for exact oracles.

(4) Theorem 6.2 still holds if Step 6 uses $\kappa < 1$ in the descent test (1.8), but the algorithm employs linearization accumulation: $\check{f}_{k+1} := \max\{\check{f}_k, f_{k+1}\}$. Indeed, then there is k' such that $\check{f}_k = \check{f}_{k'}$ for all $k \ge k'$, and descent steps yield $f_{\hat{u}}^{k+1} \le \check{f}_k(u^{k+1})$ as before (otherwise $\check{f}_{k+1}(u^{k+1}) \ge f_{k+1}(u^{k+1}) = f_{\hat{u}}^{k+1} > \check{f}_k(u^{k+1})$, a contradiction).

(5) Theorem 6.2 still holds if additionally Step 3 is replaced by the following

Step 3_1 (*Working set extension*). If $V_k > 0$, go to Step 5_1. If $\zeta_k \le 0$, stop. Otherwise, pick a subset $\emptyset \neq \check{I}'_k \subset I'_k$, set $I_k := I_k \cup \check{I}'_k$ and go back to Step 2.

Here the working set is extended only if $\hat{u}^k \in \operatorname{Arg\,min}_{C_k} \check{f}_k$ (since $V_k = 0$ implies that $0 \in \partial \check{f}_C^k(\hat{u}^k)$, using (2.9), (2.8) and the fact that $p^k \in \partial \check{f}_C^k(u^{k+1})$ by (2.10)–(2.11)).

(6) At Step 7, we can replace $t_{k+1} \ge t_{\min}$ by the milder condition that $\sum_{k \in \mathscr{D}} t_k = \infty$ if $|\mathscr{D}| = \infty$. Then in the proof of Lemma 3.5, $\sum_{k \in \mathscr{D}} v_k < \infty$ gives $\sum_{k \in \mathscr{D}} t_k |p^k|^2 < \infty$ by (2.14), so $|p^k| \xrightarrow[\mathscr{D}']{} 0$ for some $\mathscr{D}' \subset \mathscr{D}$, which replaces $\mathscr{D}$ in the proof of Theorem 6.2. (Incidentally, we needn't have $|p^k| \xrightarrow[\mathscr{D}]{} 0$ as claimed above [3, Eq. (6.2)].)

(7) Our setting is relevant for LP relaxations of integer programming problems of the form (5.1) with ψ_0, ψ affine and Z finitely generated, i.e., $Z := \operatorname{co} \check{Z}$ for some finite set $\check{Z}$. Then typical oracles find linearizations of the form (5.4) with $z^j \in \check{Z}$, i.e., at most $|\check{Z}|$ different linearizations. If Step 4 or Step 3_1 above terminate, then $\hat{z}^k \in Z_\varepsilon$ for $\varepsilon = \varepsilon_f^{\ell(k)}$ by Remark 5.3(1). If Step 5_1 terminates, an ε-optimal solution can be recovered in a similar way. Specifically, $0 \in \partial \check{f}_k(u^{k+1}) + \partial i_C(u^{k+1})$ implies the existence of (possibly different) subgradients $\hat{g}^k$ and ν^k such that (2.4)–(2.12) hold with C_k replaced by C and $\hat{u}^k$ by u^{k+1}; in particular $p^k = 0$, $\varepsilon_k = f_{\hat{u}}^k - \check{f}_k(u^{k+1}) \le 0$ (see Step 5_1) and $V_k = 0$. Hence for convex weights α_j^k satisfying (4.2)–(4.3) and $\hat{z}^k$ defined by (5.5), we get (5.7) as before with ζ_k replaced by $-\infty$. Then $\hat{z}^k \in Z_\varepsilon$ by the proof of Theorem 5.2.

7 A relax-and-cut framework

We now give a specialized variant of Algorithm 2.3 for the primal problem (5.1).

First, note that most of the relations from Sect. 2 can be reduced to the space $\mathbb{R}^{|I_k|}$ of free variables via the extension mapping $E_k : \mathbb{R}^{|I_k|} \to \mathbb{R}^{|I_k|} \times \mathbb{R}^{|I'_k|}$ defined by $E_k(u_{I_k}) := (u_{I_k}, 0_{I'_k})$. E.g., we have $u^{k+1} = E_k(u^{k+1}_{I_k})$ in subproblem (2.2), where

$$u^{k+1}_{I_k} := \arg\min \left\{ \check{f}_k(E_k(u_{I_k})) + \tfrac{1}{2t_k}|u_{I_k} - \hat{u}^k_{I_k}|^2 : u_{I_k} \in \mathbb{R}^{|I_k|}_+ \right\}. \tag{7.1}$$

For selection, $\check{f}_k \circ E_k = \max_{j\in J_k} f_j \circ E_k$ with $f_j \circ E_k(\cdot) = \psi_0(z^j) + \langle \cdot, \psi_{I_k}(z^j)\rangle$ (cf. (5.4)) and the solution of (7.1) doesn't need the components of the subgradients $g^j := \psi(z^j)$ indexed by the set I'_k. However, these components are required to recover the vector $\hat{g}^k_{I'_k}$ via (4.2):

$$\hat{g}^k_{I'_k} = \sum_{j\in\hat{J}_k} \alpha^k_j \psi_{I'_k}(z^j). \tag{7.2}$$

Note that $\hat{g}^k_{I'_k}$ is only needed at Step 3 to find the measure ζ_k of (2.17), and a nonempty subset $\check{I}'_k$ of I'_k if $\zeta_k > V_k$. This motivates the following *relax-and-cut* approach. When the number n of primal constraints is huge, we can work with the *reduced subgradients* $g^j_{I_k} := \psi_{I_k}(z^j)$, evaluating only $g^{k+1}_{I_k} := \psi_{I_k}(z^{k+1})$ at Step 6 (of course, when the set I_k grows, the corresponding subgradient components must be computed). Instead of using the vector $\hat{g}^k_{I'_k}$ directly at Step 3 as above, we can employ a separation oracle which either confirms that $\zeta_k \le V_k$, or delivers a suitable set $\check{I}'_k \ne \emptyset$ if $\zeta_k > V_k$.

These concepts are described formally below, starting with linearization selection.

We suppose that Step 6 employs an *optimization oracle* which, given the working set I_k and the trial point $u^{k+1}_{I_k}$, returns a point $z^{k+1} \in Z$, $\psi_0(z^{k+1})$, $\psi_{I_k}(z^{k+1})$,

$$f^{k+1}_u := \psi_0(z^{k+1}) + \langle u^{k+1}_{I_k}, \psi_{I_k}(z^{k+1})\rangle \quad \text{and} \quad g^{k+1}_{I_k} := \psi_{I_k}(z^{k+1}). \tag{7.3}$$

Since $u^{k+1}_{I'_k} = 0$, this is the same as (5.3), except that $\psi_{I'_k}(z^{k+1})$ needn't be computed.

For Step 3, we assume that we have a *separation oracle* which, given a primal point $z \in Z$, the working set $I_k \subset N$ and the optimality measure $V_k \ge 0$, returns a set

$$\mathscr{O}(z) \subset I'_k := N \setminus I_k, \quad \mathscr{O}(z) \ne \emptyset \quad \text{if } \psi_i(z) < -V_k \text{ for some } i \in I'_k. \tag{7.4}$$

Then the set

$$\mathscr{O}_k := \left\{ \mathscr{O}(z^j) : j \in \hat{J}_k \right\} \tag{7.5}$$

may serve as $\check{I}'_k$ at Step 3, since by (7.2) and (7.4), $\mathscr{O}_k = \emptyset$ implies $\hat{g}^k_i \ge -V_k \ \forall i \in I'_k$ and thus $\zeta_k \le V_k$ (cf. (2.17)) as before. However, before the set I_k grows, the separation oracle (or yet another *augmentation oracle*) should also recover the "missing" subgradient components $\psi_i(z^j)$, $i \in \mathscr{O}_k$, $j \in J_k$. Under these conditions, Step 3 can be replaced by

Step 3_2 (*Working set extension*). If $\mathscr{O}_k \ne \emptyset$, set $I_k := I_k \cup \mathscr{O}_k$ and go back to Step 2.

Our previous convergence results remain valid, since $\zeta_k \le V_k$ at Step 4 as before. In the polyhedral setting of Sect. 6, Step 3_1 of Rem. 6.3(3) can be replaced by

Step 3_3 (*Working set extension*). If $V_k > 0$, go to Step 5_1. If $\mathscr{O}_k = \emptyset$, stop. Otherwise, set $I_k := I_k \cup \mathscr{O}_k$ and go back to Step 2.

Since finding the set $\mathscr{O}_k$ in (7.5) involves $|\hat{J}_k|$ calls of the separation oracle, where $|\hat{J}_k|$ can reach $n+1$, note that just *one* call suffices if the constraint function ψ is affine; then (5.5) and (7.2) imply that $\hat{g}^k_{I'_k} = \psi_{I'_k}(\hat{z}^k)$. Hence, we can replace (7.5) by

$$\mathscr{O}_k := \mathscr{O}(\hat{z}^k) \tag{7.6}$$

at Step 3_2 above, since again $\mathscr{O}_k = \emptyset$ implies that $\hat{g}^k_i \geq -V_k \; \forall i \in I'_k$ and $\zeta_k \leq V_k$.

Similarly for aggregation in Sect. 5.3: when ψ is affine, (5.10) with $\hat{g}^k = \nabla \bar{f}_k$ yield $\hat{g}^k_{I'_k} = \psi_{I'_k}(\hat{z}^k)$, so we can use the single oracle answer (7.6) at Step 3_2 as above.

We now discuss some modifications related to the framework of [1], where the oracle is exact ($\varepsilon^j_f \equiv 0$) and f_C is coercive, so that $\{\hat{u}^k\}$ is bounded as in Rem. 3.7(2).

Remark 7.1 (1) Our requirement (7.4) on the separation oracle seems to be the weakest one needed for convergence results. We may replace V_k in (7.4) by βV_k for a fixed $\beta > 0$; then $\mathscr{O}_k = \emptyset$ yields $\zeta_k \leq \beta V_k$, enough for the optimality estimate (2.18).

(2) Let $\hat{\eta}_k := \max\{-\hat{g}^k_i : i \in I_k\}$, $\eta_k := \max\{-p^k_i : i \in I_k\}$. Since $p^k_{I_k} = \hat{g}^k_{I_k} + v^k_{I_k} \leq \hat{g}^k_{I_k}$ (cf. the proof of Lem. 5.1), we have $\hat{\eta}_k \leq \eta_k \leq |p^k| \leq V_k$. Hence assuming ψ is affine and replacing V_k by $\beta \max\{\hat{\eta}_k, 0\}$ in (7.4) gives the oracle condition of [1], which is stronger in general than replacing V_k by βV_k: $\mathscr{O}_k = \emptyset$ in (7.6) yields $\zeta_k \leq \beta \max\{\hat{\eta}_k, 0\}$ for the former, and thus $\zeta_k \leq \beta V_k$ for the latter. In particular, the maximum violation oracle of [6, Def. 4.1] satisfies the stronger condition with $\beta = 1$.

(3) In general terms, assuming $\varepsilon^j_f \equiv 0$ and ψ is affine, the method of [1, Alg. 1] is obtained from Algorithm 2.3 by removing Steps 3 and 5, stopping in Step 4 if $V_k = 0$ and additionally $\mathscr{O}_k = \emptyset$ in (7.6) (so that $\zeta_k \leq V_k$), and replacing Step 8 by

Step 8_1 (*Working set updating*). After a descent step with $\mathscr{O}_k = \emptyset$, set I_{k+1} to the minimal set $\hat{I}_{k+1} := \{i \in N : \hat{u}^{k+1}_i > 0\}$; otherwise, set $I_{k+1} := I_k \cup \mathscr{O}_k$.

The main difference is that when $\mathscr{O}_k \neq \emptyset$, Algorithm 2.3 with Step 3_2 extends the working set immediately and finds the trial point again. In contrast, this variant delays extension till the next iteration, and its additional f-evaluation is essentially redundant when a null step occurs, and troublesome for convergence otherwise (see below). In particular, to reduce the working set after a descent step, this variant needs the condition $\mathscr{O}_k = \emptyset$, which holds automatically at Step 8 of Algorithm 2.3.

Convergence of this variant may be analyzed as follows.

First, suppose there is $\bar{k}$ such that only null steps occur for $k \geq \bar{k}$. Then $V_k \to 0$ by the proof of Lemma 3.3, with $\mathscr{O}_k = \emptyset$ giving $\zeta_k \leq V_k$ for large k as before; hence we may take $K = \mathbb{N}$ in Theorem 5.2. In contrast, [1, Lem. 7] yields only $\underline{\lim}_k V_k = 0$ under the weaker stepsize conditions of [8, Thm. XV.3.2.4].

In the case of infinitely many descent steps, the set $K := \{k \in \mathscr{D} : \mathscr{O}_k = \emptyset\}$ is of interest (since $\mathscr{O}_k = \emptyset$ implies $\zeta_k \leq V_k$); note that $|K| = \infty$ (otherwise, for k large enough, each descent step would increase I_k, contradicting $|I_k| \leq n$). Suppose $\{\hat{u}^k\}$ is bounded (e.g., f_C is coercive). Since then Lemma 3.5 yields $V_k \xrightarrow[\mathscr{D}]{} 0$, we may take this K in Theorem 5.2; this corresponds to [1, Lem. 8]. Moreover, since $|p^k| \xrightarrow[\mathscr{D}]{} 0$ in Lemma 3.5, Step 8_1 may replace the condition "$\mathscr{O}_k = \emptyset$" with "$\mathscr{O}_k = \emptyset$ or $|p^k| > \tau_p$" for a fixed tolerance $\tau_p > 0$; this corresponds to [1, Rem. 10]. Anyway, since descent steps with $\mathscr{O}_k \neq \emptyset$ may occur, we can't claim that $V_k \xrightarrow[K]{} 0$ when $\{\hat{u}^k\}$ is not bounded (e.g., $|\hat{u}^k| \to \infty$ when $f_{\hat{u}}^k \downarrow f_*$ but the dual problem (1.1) has no solution).

As for finite termination in the polyhedral case, Theorem 6.2 and Remark 6.3 remain valid under the additional assumptions of Sect. 6 (with Steps 3_1 and 5_1 omitted). This subsumes [1, Thm. 11], where f_C is additionally coercive. In fact the result of [1, Thm. 11] is flawed: although it allows selection at null steps, its proof assumes accumulation (as in Remark 6.3(4)).

(4) In (3) above, two additional flaws of [1] should be noted. First, [1, Eq. conv(1)] should additionally assume that the primal objective is affine when its ground set ($\mathscr{Q}$ such that $\operatorname{co}\mathscr{Q} = Z$ in (5.1)) is not convex. Second, [1, Rem. 10] allows reducing I_k at some null steps; this kills the argument that $|K| = \infty$ when $|\mathscr{D}| = \infty$ in (3) above.

We end this section by considering inexact separation oracles.

Remark 7.2 Assuming ψ is affine for simplicity, suppose Step 3_2 employs an *inexact* separation oracle which at some iterations delivers $\mathscr{O}_k = \emptyset$ when $\min_{i \in I'_k} \psi(\hat{z}^k) < -V_k$, failing to detect "significant" primal constraint violation. Such separation blunders correspond to Step 3 failing to detect that $\zeta_k > V_k$, and Remark 5.3(3) exhibits their influence on asymptotic primal (in)feasibility. Apparently, inexact separation oracles have not been analyzed in the literature so far, although they may appear in difficult combinatorial applications, where finding a violated constraint may cost too much, and only heuristic separation oracles are viable.

8 Extensions to proximal-projection bundle methods

8.1 General proximal-projection results

Following [15], we now show how to replace the constrained subproblem (2.2) with an essentially unconstrained subproblem, followed by a projection on the set C_k.

First, note that the set $C_k = C \cap S_k$ (cf. (1.11)), where $S_k := \{u \in \mathbb{R}^n : u_{I'_k} = 0\}$ is the subspace of free variables. Thus $i_{C_k} = i_C + i_{S_k}$ and $\partial i_{C_k} = \partial i_C + \partial i_{S_k}$, where

$$\partial i_{S_k}(\cdot) = S_k^\perp := \left\{ s \in \mathbb{R}^n : s_{I_k} = 0 \right\} \quad \text{on } S_k, \tag{8.1}$$

with $S_k^\perp$ being the subspace orthogonal to S_k. Hence if $S_k = S_{k-1}$, then the linearization $\bar{\imath}_C^{k-1}(\cdot) = \langle \nu^{k-1}, \cdot - u^k \rangle \le i_{C_{k-1}}(\cdot)$ satisfies $\bar{\imath}_C^{k-1} + i_{S_k} \le i_{C_k}$. Then, replacing i_{C_k} by its partial linearization $\bar{\imath}_C^{k-1} + i_{S_k}$ in subproblem (2.2) yields the prox subproblem

$$\check{u}^{k+1} := \arg\min \left\{ \phi_f^k(\cdot) := \check{f}_k(\cdot) + \bar{\imath}_C^{k-1}(\cdot) + i_{S_k}(\cdot) + \tfrac{1}{2t_k} |\cdot - \hat{u}^k|^2 \right\}. \tag{8.2}$$

Since $0 \in \partial \phi_f^k(\check{u}^{k+1})$ above by optimality, there exist subgradients $\hat{g}^k$ and s^k such that

$$\hat{g}^k \in \partial \check{f}_k(\check{u}^{k+1}), \quad s^k \in \partial i_{S_k}(\check{u}^{k+1}) \quad \text{and} \quad \hat{g}^k + \nu^{k-1} + s^k = (\hat{u}^k - \check{u}^{k+1})/t_k. \tag{8.3}$$

Then, as in (2.5), the corresponding aggregate linearization of $\check{f}_k$ and f satisfies

$$\bar{f}_k(\cdot) := \check{f}_k(\check{u}^{k+1}) + \langle \hat{g}^k, \cdot - \check{u}^{k+1} \rangle \le \check{f}_k(\cdot) \le f(\cdot). \tag{8.4}$$

Replacing the model $\check{f}_k$ by its linearization $\bar{f}_k$ in (2.2) gives the second subproblem

$$u^{k+1} := \arg\min \left\{ \phi_C^k(\cdot) := \bar{f}_k(\cdot) + i_{C_k}(\cdot) + \tfrac{1}{2t_k}|\cdot - \hat{u}^k|^2 \right\} = P_{C_k}(\hat{u}^k - t_k\hat{g}^k), \quad (8.5)$$

where the final equality stems from the form (8.4) of $\bar{f}_k$. Since $0 \in \partial\phi_C^k(u^{k+1})$ above, there exists a unique subgradient v^k such that

$$v^k \in \partial i_{C_k}(u^{k+1}) \quad \text{and} \quad \hat{g}^k + v^k = (\hat{u}^k - u^{k+1})/t_k, \quad (8.6)$$

for which the corresponding linearization $\bar{\imath}_C^k$ satisfies (2.6). Finally, comparing (2.2) and (8.5), we see that the predicted decrease of (1.7) must be replaced by

$$v_k := f_{\hat{u}}^k - \bar{f}_k(u^{k+1}). \quad (8.7)$$

To describe this *proximal-projection* variant formally, setting $\bar{\imath}_C^0(\cdot) := \langle v^0, \cdot\rangle$ with $v^0 := 0$ at Step 0, suppose Step 2 is replaced by the following step.

Step 2_1 (*Trial point finding*). Find $\check{u}^{k+1}$ via (8.2) and u^{k+1} via (8.5). Set v_k by (8.7), $p^k := (\hat{u}^k - u^{k+1})/t_k$, $\varepsilon_k := v_k - t_k|p^k|^2$ and V_k by (2.9).

Using the extension mapping E_k as for (7.1), relations (8.2) and (8.5) imply that

$$\check{u}^{k+1} = E_k(\check{u}_{I_k}^{k+1}) \quad \text{with} \quad \check{u}_{I_k}^{k+1} = \arg\min \left\{ (\check{f}_k + \bar{\imath}_C^{k-1}) \circ E_k(\cdot) + \tfrac{1}{2t_k}|\cdot - \hat{u}_{I_k}^k|^2 \right\}, \quad (8.8)$$

$$u^{k+1} = E_k(u_{I_k}^{k+1}) \quad \text{with} \quad u_{I_k}^{k+1} = \max\{\hat{u}_{I_k}^k - t_k\hat{g}_{I_k}^k, 0\}. \quad (8.9)$$

Remark 8.1 (1) For a polyhedral model $\check{f}_k$, subproblem (8.8) is much easier to handle via simple QP solvers [9,11] than the bound-constrained subproblem (7.1).

(2) Note the following connections with the relations of Sect. 2: (1.7) is replaced by (8.7), (2.4) by (8.3) and (8.6), (2.5) by (8.4), and $\check{f}_k$ by $\bar{f}_k$ in (2.10). With these replacements, our basic relations (2.6)–(2.12) remain valid. Then it is easy to check that assertions (1) through (4) of Lemma 2.2 follow as before, whereas a reformulation of assertion (5) means that for $I_k^- = \emptyset$ we are in the framework of [15], with C_k replaced by C in (8.5)–(8.6).

(3) Note that, by (8.6) and (2.19), the linearization $\bar{\imath}_C^k$ of (2.6) has the form

$$\bar{\imath}_C^k(\cdot) = \langle v^k, \cdot\rangle \quad \text{with} \quad v_{I_k}^k = \min\{\hat{u}_{I_k}^k/t_k - \hat{g}_{I_k}^k, 0\},\ v_{I_k'}^k = -\hat{g}_{I_k'}^k. \quad (8.10)$$

Having $\bar{\imath}_C^{k-1}$ as a model of i_{C_k} in subproblem (8.2) is essential only after null steps; after I_k or t_k increases, or a descent step occurs, a better model may be constructed as follows. When Step 3 increases the set I_k, we can reset $v_{I_k}^{k-1} := \min\{v_{I_k}^{k-1}, 0\}$ to keep $\bar{\imath}_C^{k-1}(\cdot) := \langle v^{k-1}, \cdot\rangle \le i_{C_k}(\cdot)$; alternatively, we can set $v_{I_k}^{k-1} := \min\{v_{I_k}^k, 0\}$. Similarly, after Step 5 increases t_k, we can use $v_{I_k}^{k-1} := \min\{\hat{u}_{I_k}^k/t_k - \hat{g}_{I_k}^k, 0\}$ as if subproblem (8.5) were resolved, whereas if $\hat{u}^{k+1} \ne \hat{u}^k$ after Step 7, we can update $v_{I_k}^k := \min\{\hat{u}_{I_k}^{k+1}/t_{k+1} - \hat{g}_{I_k}^k, 0\}$.

(4) As for convergence, in view of the similarities noted in (2) above, it suffices to prove Lemma 3.3, since the remaining proofs of Sect. 3 go through. To this end,

one can adapt the proof of [15, Lem. 3.2], using the following facts: in view of (8.3), we can *a posteriori* replace the function i_{S_k} in subproblem (8.2) by its linearization

$$\bar{\imath}_S^k(\cdot) := \langle s^k, \cdot - \breve{u}^{k+1}\rangle \le i_{S_k}(\cdot), \tag{8.11}$$

for which (see below (8.1)) $\bar{\imath}_C^{k-1} + \bar{\imath}_S^k \le i_{C_k}$ if $C_k = C_{k-1}$. In fact the proof of [15, Lem. 3.2] needs the additional condition that $\{\hat{g}^k\}$ is bounded if such is $\{u^k\}$, but this condition holds automatically for the models $\check{f}_k$ chosen as in Sect. 4 or 8.2 below; cf. [15, Rem. 4.4 and Sect. 5.5]. Similarly, the results of Sect. 4–5 still hold (except that $\breve{u}^{k+1}$ replaces u^{k+1} in (4.2) as in (2.5)). In contrast, the use of the aggregate $\bar{\imath}_C^{k-1}$ in (8.2) spoils the finite-termination framework of Sect. 6. Hence Step 3_3 of Sect. 7 can't be used, but the remaining results of Sect. 7 aren't affected.

(5) To bring u^{k+1} in (8.5) closer to the solution of (2.2), Step 2_1 can be executed more than once at each iteration. Specifically, for a given *subproblem accuracy threshold* $\check{\kappa} \in (0,1)$, if $\check{f}_k(u^{k+1}) > f_{\hat{u}}^k - \check{\kappa} v_k$ after Step 5, then we can go back to Step 2_1 with $\bar{\imath}_C^{k-1}(\cdot) := \bar{\imath}_C^k(\cdot)$. This modification can be analyzed as in [15, Rem. 4.1].

8.2 Lagrangian relaxation with nonpolyhedral objective models

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

$$L(z,u) := \psi_0(z) + \langle u, \psi(z)\rangle = \psi_0(z) + \langle u, b - Az\rangle \tag{8.12}$$

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

$$\check{f}_k(\cdot) := \max_{z\in Z_k} L(z,\cdot) \quad \text{with} \quad z^k \in Z_k \subset Z, \tag{8.13}$$

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 (2.27), 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 (8.2) with the model $\check{f}_k$ given by (8.13), we employ the *Lagrangian* $\bar{L} : \mathbb{R}^n \times Z_k \times S_k^\perp \to \mathbb{R}$ of (8.2) defined by

$$\bar{L}(u;z,s) := L(z;u) + \langle v^{k-1} + s, u\rangle + \tfrac{1}{2t_k}|u - \hat{u}^k|^2, \tag{8.14}$$

so that

$$\phi_f^k(\cdot) = \sup\{\bar{L}(\cdot;z,s) : (z,s) \in Z_k \times S_k^\perp\}. \tag{8.15}$$

For each pair $(z,s) \in Z_k \times S_k^\perp$, the (unique) *Lagrangian solution*

$$u(z,s) := \arg\min \bar{L}(\cdot;z,s) = \hat{u}^k - t_k\left[\psi(z) + v^{k-1} + s\right] \tag{8.16}$$

substituted for u in (8.14) gives the value of the *dual function* $q : Z_k \times S_k^\perp \to \mathbb{R}$,

$$q(z,s) := \min \bar{L}(\cdot;z,s) = \psi_0(z) + \langle \psi(z) + v^{k-1} + s, \hat{u}^k\rangle - \tfrac{t_k}{2}|\psi(z) + v^{k-1} + s|^2. \tag{8.17}$$

Since $s \in S_k^\perp$ implies $s_{I_k} = 0$ by (8.1), whereas $\hat{u}_{I_k'}^k = 0$ by the choice of I_k, we have

$$q(z,s) = \psi_0(z) + \langle \psi_{I_k}(z) + v_{I_k}^{k-1}, \hat{u}_{I_k}^k\rangle - \tfrac{t_k}{2}|\psi_{I_k}(z) + v_{I_k}^{k-1}|^2 - \tfrac{t_k}{2}|\psi_{I_k'}(z) + v_{I_k'}^{k-1} + s_{I_k'}|^2. \tag{8.18}$$

Maximizing $q(z,s)$ above over s in $S_k^\perp$ yields the optimal $s(z)$ and its value $q(z)$:

$$s(z) := \arg\max_{S_k^\perp} q(z,\cdot) \quad \text{with} \quad s_{I_k}(z) = 0 \quad \text{and} \quad s_{I_k'}(z) = -\psi_{I_k'}(z) - v_{I_k'}^{k-1}, \tag{8.19}$$

$$q(z) := q(z,s(z)) = \psi_0(z) + \langle \psi_{I_k}(z) + v_{I_k}^{k-1}, \hat{u}_{I_k}^k \rangle - \tfrac{t_k}{2} |\psi_{I_k}(z) + v_{I_k}^{k-1}|^2. \tag{8.20}$$

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

$$\hat{z}^k \in \operatorname{Arg\,max}\{q(z) : z \in Z_k\}. \tag{8.21}$$

Lemma 8.2 *Given a dual solution $\hat{z} := \hat{z}^k$ of* (8.21)*, define $\hat{s} := s(\hat{z})$ by* (8.19) *and the Lagrangian solution $\check{u} := u(\hat{z},\hat{s})$ by* (8.16)*. Then we have the following statements.*

(1) *The triplet $(\check{u};\hat{z},\hat{s})$ is a saddle-point of the Lagrangian $\bar{L}$ defined by* (8.14):

$$\bar{L}(\check{u};z,s) \le \bar{L}(\check{u};\hat{z},\hat{s}) \le \bar{L}(u;\hat{z},\hat{s}) \quad \forall u \in \mathbb{R}^n, (z,s) \in Z_k \times S_k^\perp. \tag{8.22}$$

(2) *Relations* (8.2)–(8.4) *hold for $\check{u}^{k+1} = \check{u}$, $\hat{g}^k = \psi(\hat{z}^k)$, $s^k = \hat{s}$, and*

$$\check{u}_{I_k}^{k+1} = \hat{u}_{I_k}^k - t_k\big[\psi_{I_k}(\hat{z}^k) + v_{I_k}^{k-1}\big], \quad \check{u}_{I_k'}^{k+1} = 0, \tag{8.23}$$

$$\bar{f}_k(\cdot) = \psi_0(\hat{z}^k) + \langle \cdot, \psi(\hat{z}^k) \rangle. \tag{8.24}$$

Proof (1) By [24, Thm. 37.4], the Lagrangian $\bar{L}$ has a saddle-point $(\bar{u};\bar{z},\bar{s})$. Since the pair $(\hat{z},\hat{s})$ maximizes the function $\min_u \bar{L}(u;\cdot,\cdot)$ over $Z_k \times S_k^\perp$ (cf. (8.16)–(8.21)), $(\bar{u};\hat{z},\hat{s})$ is a saddle-point as well [8, Thm. VII.4.2.5]. Then the fact that $\bar{L}(\bar{u};\hat{z},\hat{s}) \le \bar{L}(u;\hat{z},\hat{s})$ $\forall u$ yields $\bar{u} = u(\hat{z},\hat{s}) = \check{u}$ by (8.16), so that (8.22) holds.

(2) By (8.2) and (8.15), (8.22) yields that $\check{u}^{k+1} = \check{u}$ [8, Thm. VII.4.2.5], $\check{f}_k(\check{u}^{k+1}) = \psi_0(\hat{z}^k) + \langle \check{u}^{k+1}, \psi(\hat{z}^k) \rangle$, and (since $\check{u}$ minimizes $\max_{z \in Z_k} \bar{L}(\cdot;z,\hat{s})$) that (8.2) holds with i_{S_k} replaced by $\langle \hat{s}, \cdot \rangle$; hence the corresponding optimality condition gives the inclusion $(\hat{u}^k - \check{u}^{k+1})/t_k - v^{k-1} - \hat{s} \in \partial \check{f}_k(\check{u}^{k+1})$. Since $\check{u}^{k+1}$ is given by (8.16) with $(z,s) = (\hat{z},\hat{s})$, we see that $\psi(\hat{z}^k) \in \partial \check{f}_k(\check{u}^{k+1})$, (8.3) holds for $\hat{g}^k = \psi(\hat{z}^k)$ and $s^k = \hat{s}$ (cf. (8.1)), and (8.23) follows. Finally, (8.4) with $\check{f}_k(\check{u}^{k+1}) = \psi_0(\hat{z}^k) + \langle \check{u}^{k+1}, \psi(\hat{z}^k) \rangle$ implies (8.24). □

In view of (8.13) and (8.24), the requirement of (2.27) is met if Z_{k+1} satisfies

$$Z_{k+1} \supset \{\hat{z}^k, z^{k+1}\}, \tag{8.25}$$

in addition to being a closed convex subset of Z. Further, the aggregate representation (8.24) 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, recalling Rem. 8.1(4), we conclude that the results of Sect. 5.2 hold for this variant as well, whereas in Sect. 7, since $\hat{g}_{I_k'}^k = \psi_{I_k'}(\hat{z}^k)$ by (8.24), Step 3_2 can call the separation oracle just once to get the set $\mathcal{O}_k := \mathcal{O}(\hat{z}^k)$ of (7.6).

Remark 8.3 (1) In view of (8.5) and (8.10) with $\hat{g}^k = \psi(\hat{z}^k)$, in practice we may use

$$u_{I_k}^{k+1} = \max\{\hat{u}_{I_k}^k - t_k \psi_{I_k}(\hat{z}^k), 0\} \quad \text{and} \quad v_{I_k}^k = \min\{\hat{u}_{I_k}^k / t_k - \psi_{I_k}(\hat{z}^k), 0\}.$$

(2) If ψ_0 is linear and $Z_k := \operatorname{co}\{z^j\}_{j=1}^k$, it can be seen from [15, Rem. 5.6(ii)] that our framework comprises the method of [2, Sect. 3.3], which requires exact evaluations and uses heuristic updates of the set I_k. (In this case the model (8.13) is equivalent to the polyhedral model (4.1) with $J_k := \{1\colon k\}$; cf. (8.12) and (5.4).)

(3) It follows from [15, Rem. 5.6(iv)] that the results of Sect. 5.2 hold for the "looping" modification of Remark 8.1(5); its heuristic variants were used in [2,23].

8.3 SDP via eigenvalue optimization

As in [15], we consider the Euclidean space S^m of $m \times m$ real symmetric matrices with the Frobenius inner product $\langle x,y\rangle = \mathrm{tr}\, xy$ (we use lowercase notation 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 $y \in S^m$ satisfies (see, e.g., [26])

$$\lambda_{\max}(y) = \max\left\{\langle y,x\rangle : x \in \Sigma^m\right\} \quad \text{with} \quad \Sigma^m := \left\{x \in S^m_+ : \mathrm{tr}\, x = 1\right\}. \tag{8.26}$$

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

$$(P): \quad \max\ \langle c,x\rangle \quad \text{s.t.} \quad Ax \le b, \quad x \in S^m_+. \quad \mathrm{tr}\, x = a. \tag{8.27}$$

We can regard (P) as an instance of (5.1) with $\mathscr{Z} := S^m$, $\psi_0(z) := \langle c,z\rangle$, $\psi(z) := b - Az$ and $Z := a\Sigma^m$. Then, by (5.2) and (8.26), the dual function f satisfies

$$f(u) = a\lambda_{\max}(c - A^*u) + \langle b,u\rangle \quad \text{for all } u, \tag{8.28}$$

where A^* is the adjoint of A (defined by $\langle z,A^*u\rangle = \langle Az,u\rangle$ for all $z \in S^m$, $u \in \mathbb{R}^n$). For (5.3), we may use $z^j := ar^j {r^j}^T$, where $r^j \in \mathbb{R}^m$ is an approximate normalized eigenvector of the matrix $s^j := c - A^*u^j \in S^m$. Then the approximate subgradients $g^j := \psi(z^j) = b - Az^j$ are bounded.

Thus we can use the setting of Sect. 8.2 with models $\check{f}_k$ given by (8.13) for sets Z_k satisfying (8.25). In effect, the results of Sect. 5.2 and 8.2 hold for this variant.

Remark 8.4 Our dual problem $f_* := \inf_C f$ is equivalent to the standard dual of (P) which is strictly feasible. Hence (cf. [26, 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 [6, Thm. 4.8], which assumes that $\mathrm{Arg\,min}_C f$ is nonempty and bounded, i.e., f_C is coercive (cf. Rem. 3.7(2)). Further, our separation condition (7.4) can produce a smaller set $\mathscr{O}_k$ in (7.6) than the maximum violation oracle of [6, Def. 4.1].

References

1. Belloni, A., Sagastizábal, C.: Dynamic bundle methods. Math. Program. **120**, 289–311 (2009)
2. Fischer, I., Gruber, G., Rendl, F., Sotirov, R.: Computational experience with a bundle approach for semidefinite cutting plane relaxations of Max-Cut and Equipartition. Math. Program. **105**, 451–469 (2006)
3. Frangioni, A.: Generalized bundle methods. SIAM J. Optim. **13**, 117–156 (2002)
4. Frangioni, A., Gallo, G.: A bundle type dual-ascent approach to linear multi-commodity min cost flow problems. INFORMS J. Comput. **11**, 370–393 (1999)
5. Frangioni, A., Lodi, A., Rinaldi, G.: New approaches for optimizing over the semimetric polytope. Math. Program. **104**, 375–388 (2006)
6. Helmberg, C.: A cutting plane algorithm for large scale semidefinite relaxations. In: M. Grötschel (ed.) The Sharpest Cut: The Impact of Manfred Padberg and His Work, pp. 233–256. MPS-SIAM, Philadelphia (2004)
7. Helmberg, C., Kiwiel, K.C.: A spectral bundle method with bounds. Math. Program. **93**, 173–194 (2002)
8. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms. Springer, Berlin (1993)

9. Kiwiel, K.C.: A method for solving certain quadratic programming problems arising in nonsmooth optimization. IMA J. Numer. Anal. **6**, 137–152 (1986)
10. Kiwiel, K.C.: Proximity control in bundle methods for convex nondifferentiable minimization. Math. Program. **46**, 105–122 (1990)
11. Kiwiel, K.C.: A Cholesky dual method for proximal piecewise linear programming. Numer. Math. **68**, 325–340 (1994)
12. Kiwiel, K.C.: Finding normal solutions in piecewise linear programming. Appl. Math. Optim. **32**, 235–254 (1995)
13. Kiwiel, K.C.: A projection-proximal bundle method for convex nondifferentiable minimization. In: M. Théra, R. Tichatschke (eds.) Ill-posed Variational Problems and Regularization Techniques, Lecture Notes in Economics and Mathematical Systems 477, pp. 137–150. Springer-Verlag, Berlin (1999)
14. Kiwiel, K.C.: A proximal bundle method with approximate subgradient linearizations. SIAM J. Optim. **16**, 1007–1023 (2006)
15. Kiwiel, K.C.: A proximal-projection bundle method for Lagrangian relaxation, including semidefinite programming. SIAM J. Optim. **17**, 1015–1034 (2006)
16. Kiwiel, K.C.: A method of centers with approximate subgradient linearizations for nonsmooth convex optimization. SIAM J. Optim. **18**, 1467–1489 (2007)
17. Kiwiel, K.C.: An alternating linearization bundle method for convex optimization and nonlinear multicommodity flow problems. Math. Program. **?** (2009). DOI 10.1007/s10107-009-0327-0
18. Kiwiel, K.C.: Bundle methods for convex minimization with partially inexact oracles. Tech. rep., Systems Research Institute, Warsaw (2009). Revised April 2010. Available at the URL http://www.optimization-online.org/DB_FILE/2009/03/2257.pdf
19. Kiwiel, K.C.: An inexact bundle approach to cutting-stock problems. INFORMS J. Comput. **22**, 131–143 (2010)
20. Kiwiel, K.C., Lemaréchal, C.: An inexact bundle variant suited to column generation. Math. Program. **118**, 177–206 (2009)
21. Lemaréchal, C.: Lagrangian relaxation. In: M. Jünger, D. Naddef (eds.) Computational Combinatorial Optimization, Lecture Notes in Computer Science 2241, pp. 112–156. Springer-Verlag, Berlin (2001)
22. Rendl, F., Rinaldi, G., Wiegele, A.: Solving Max-Cut to optimality by intersecting semidefinite and polyhedral relaxations. Math. Program. **121**, 307–335 (2010)
23. Rendl, F., Sotirov, R.: Bounds for the quadratic assignment problem using the bundle method. Math. Program. **109**, 505–524 (2007)
24. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton, NJ (1970)
25. Ruszczyński, A.: A regularized decomposition method for minimizing a sum of polyhedral functions. Math. Program. **35**, 309–333 (1986)
26. Todd, M.J.: Semidefinite optimization. Acta Numer. **10**, 515–560 (2001)