44/2006

Raport Badawczy
Research Report

RB/58/2006

A method of centers with approximate subgradient linearizations for nonsmooth convex optimization

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

A method of centers with approximate subgradient linearizations for nonsmooth convex optimization

Krzysztof C. Kiwiel*

August 29, 2006

Abstract

We give a proximal bundle method for constrained convex optimization. It only requires evaluating the problem functions and their subgradients with an unknown accuracy ϵ. Employing a combination of the classic method of centers' improvement function with an exact penalty function, it does not need a feasible starting point. It asymptotically finds points with at least ϵ-optimal objective values that are ϵ-feasible. When applied to the solution of LP programs via column generation, it allows for ϵ-accurate solutions of column generation subproblems.

Key words. Nondifferentiable optimization, convex programming, proximal bundle methods, approximate subgradients, column generation.

1 Introduction

We are concerned with the solution of the following convex programming problem

$$f_* := \inf\{ f(u) : h(u) \leq 0,\ u \in C \}, \tag{1.1}$$

where C is a closed convex set in the Euclidean space $\mathbb{R}^m$ with inner product $\langle \cdot, \cdot \rangle$ and norm $|\cdot|$, f and h are convex real-valued functions, and there exists a *Slater point*

$$\mathring{u} \in C \quad \text{such that} \quad h(\mathring{u}) < 0. \tag{1.2}$$

Further, we assume that for fixed (and possibly unknown) *accuracy tolerances* $\epsilon_f, \epsilon_h \geq 0$, for each $u \in C$ we can find *approximate values* f_u, h_u and *approximate subgradients* g_f^u, g_h^u that produce the *approximate linearizations* of f and h:

$$\bar{f}_u(\cdot) := f_u + \langle g_f^u, \cdot - u \rangle \leq f(\cdot) \quad \text{with} \quad \bar{f}_u(u) = f_u \geq f(u) - \epsilon_f, \tag{1.3a}$$

$$\bar{h}_u(\cdot) := h_u + \langle g_h^u, \cdot - u \rangle \leq h(\cdot) \quad \text{with} \quad \bar{h}_u(u) = h_u \geq h(u) - \epsilon_h. \tag{1.3b}$$

*Systems Research Institute, Newelska 6, 01–447 Warsaw, Poland (kiwiel@ibspan.waw.pl)

Thus $f_u \in [f(u) - \epsilon_f, f(u)]$ estimates $f(u)$, while $g_f^u \in \partial_{\epsilon_f} f(u)$, i.e., g_f^u is a member of

$$\partial_{\epsilon_f} f(u) := \{g : f(\cdot) \geq f(u) - \epsilon_f + \langle g, \cdot - u\rangle\},$$

the ϵ_f-subdifferential of f at u; similar relations hold for f replaced by h.

This paper modifies the phase 1 - phase 2 method of centers of [Kiw85, §5.7] and extends it to approximate linearizations. We first discuss the exact case of $\epsilon_f = \epsilon_h = 0$. For an infeasible starting point, in phase 1 this method reduces the constraint violation while keeping the objective increase as small as possible; this is reasonable especially if the starting point is close to a solution. Once a feasible point is found, in phase 2 the method reduces the objective while maintaining feasibility. Both phases employ the same improvement function, and each iterate solves a subproblem with f and h approximated via accumulated linearizations, stabilized by a quadratic term centered at the best point found so far. For phase 1, the analysis of [Kiw85, §5.7] established optimality of all cluster points of the iterates, without discussing their existence. A nontrivial sufficient condition for their existence was recently given in [SaS05, Prop. 4.3(ii)] for a modified variant. We show that this condition may be expected to hold only if problem (1.1) has a Lagrange multiplier $\bar{\mu} \leq 1$ (cf. Rem. 3.11(ii)), and we extend this condition to $\bar{\mu} > 1$ by combining the standard improvement function with an exact penalty function for penalty parameters $\hat{c} \geq \bar{\mu} - 1$. In effect, our results (cf. Thms. 3.6, 3.7 and 3.10) extend the main convergence results of [Kiw85, Thm. 5.7.4] and [SaS05, Thms. 4.4–4.5]. It is crucial for large-scale implementations that our results hold for various aggregation schemes that control the size of each quadratic programming (QP) subproblem, including the schemes of [Kiw85, §5.7] and [SaS05] (see Rem. 4.1).

Our combination of improvement and penalty functions with suitable penalty parameter updates seems to be necessary for our extension to inexact evaluations (otherwise, the method could jam at phase 1 when the standard improvement function can't be reduced by more than $\max\{\epsilon_f, \epsilon_h\}$ for the tolerances ϵ_f, ϵ_h of (1.3)). Our method generates iterates in the set C, having f-values of at most $f_* + \epsilon_f$ and h-values of at most ϵ_h asymptotically (cf. Thms. 3.6–3.8), without *any* additional boundedness assumptions (such as boundedness of the feasible set, or the sufficient conditions discussed above). In a sense, this is the strongest convergence result one could hope for. Our algorithmic constructions and analysis combine the inexact linearization framework of [Kiw06c] (in a simplified version that highlights its crucial ingredients; cf. [Kiw06d]) with fairly intricate properties of improvement and penalty functions which have not been used so far in bundle methods.

As for other bundle methods, we note that the exact penalty function methods of [Kiw87, Kiw91] require additionally that the set C be bounded, and may converge slowly when their penalty parameter estimates are too high. The level methods of [LNN95] (also see [Kiw95, Fáb00, BTN05]) need boundedness of the set C as well. Similar boundedness assumptions are employed in the filter methods of [FlL99, KRSS05]. Except for [Fáb00], all these methods work with exact linearizations. We show elsewhere how to handle inexact linearizations in an exact penalty method [Kiw06b] and a filter method [Kiw06a], the latter being based on the present paper.

Our work was partly motivated by possible applications in column generation approaches to integer programming problems [LüD04], which lead to linear programming

(LP) problems with huge numbers of columns. When the dual LP problems can be formulated as (1.1) (cf. [BLM+05, LüD04, Sav97]), our approach allows for ϵ_h-accurate solutions of column generation subproblems, as well as for recovering approximate solutions to the primal problems. (See [Kiw05] for related developments and numerical results.)

The paper is organized as follows. In §2, after reviewing basic properties of penalty and improvement functions, we present our bundle method. Its convergence is analyzed in §3. Several modifications are given in §4. Applications to column generation for LP programs are studied in §5.

2 The proximal bundle method of centers

2.1 Lagrange multipliers and exact penalties

We first recall some basic duality results for problem (1.1) (cf. [Ber99, §§5.1 and 5.3]).

Consider the *Lagrangian* $L(\cdot;\mu) := f(\cdot) + \mu h(\cdot)$ with $\mu \in \mathbb{R}$, the *dual function* $q(\mu) := \inf_C L(\cdot;\mu)$ and the *dual problem* $q_* := \sup_{\mathbb{R}_+} q$ of (1.1). Under our assumptions, $f_* = q_*$. If $f_* > -\infty$, the *dual optimal set* $M := \operatorname{Arg\,max}_{\mathbb{R}_+} q$ is nonempty and compact, and consists of *Lagrange multipliers* $\mu \geq 0$ such that $q(\mu) = f_*$; if $f_* = -\infty$, $M := \emptyset$. Thus, the quantity $\bar{\mu} := \inf_{\mu\in M} \mu$ is the *minimal* Lagrange multiplier if $f_* > -\infty$, $\bar{\mu} = \infty$ otherwise.

For a *penalty parameter* $c \geq 0$, the *exact penalty function*

$$\pi(\cdot;c) := f(\cdot) + ch(\cdot)_+ \quad \text{with} \quad h(\cdot)_+ := \max\{h(\cdot), 0\} \tag{2.1}$$

satisfies $\inf_C \pi(\cdot;c) = f_* > -\infty$ iff $c \geq \bar{\mu}$ (cf. [Ber99, §5.4.5]).

2.2 Improvement functions

We associate with problem (1.1) the *improvement functions* defined for $\tau \in \mathbb{R}$ by

$$e(\cdot;\tau) := \max\{f(\cdot) - \tau, h(\cdot)\}, \quad e_C(\cdot;\tau) := e(\cdot;\tau) + i_C(\cdot), \quad E(\tau) := \inf e_C(\cdot;\tau), \tag{2.2}$$

where i_C is the *indicator* function of C ($i_C(u) = 0$ if $u \in C$, ∞ if $u \notin C$). In our context, τ will be an asymptotic estimate of f_* generated by our method, and to prove that $\tau \leq f_*$, we shall need the main property of the function E given in part (vi) of the lemma below.

Lemma 2.1. (i) *The function E defined by* (2.2) *is nonincreasing and convex.*

(ii) *If E is improper, then $E(\cdot) = f_* = -\infty$ for f_* given by* (1.1).

(iii) *If E is proper, then E is Lipschitzian with modulus* 1.

(iv) *If E is proper and $f_* = -\infty$, then $E(\cdot) = \inf_C h \in (-\infty, 0)$.*

(v) *If $f_* > -\infty$, then $E(\tau) > 0$ for $\tau < f_*$, $E(f_*) = 0$, and $E(\tau) < 0$ for $f_* < \tau$.*

(vi) *If $E(\tau) \geq 0$ for some $\tau \in \mathbb{R}$, then $\tau \leq f_*$.*

Proof. (i) Monotonicity is obvious, and convexity follows from [Roc70, Thm. 5.7].

(ii) Since $\operatorname{dom} E = \mathbb{R}$, $E(\cdot) = -\infty$ by [Roc70, Thm. 7.2] and then $f_* = -\infty$ by (1.1).

(iii) E is finite on $\operatorname{dom} E = \mathbb{R}$, and $e(\cdot;\tau') \leq e(\cdot;\tau) + |\tau - \tau'|$ for any τ and τ'.

(iv) Since $f_* = -\infty$ implies $E(\cdot) \le 0$, $E(\cdot)$ is constant and finite by [Roc70, Cor. 8.6.2], i.e., $E(\cdot) = \alpha \in \mathbb{R}$. Then, on the one hand, $\alpha \ge \inf_C h$ by (2.2). On the other hand, for $u \in C$ and $\tau \ge f(u) - h(u)$, the fact that $e(u;\tau) \le h(u)$ yields $\alpha \le \inf_C h < 0$ by (1.2).

(v) We have $E(f_*) \le 0$ by (1.1), and $E(f_*) \ge 0$ (otherwise $f(u) < f_*$ and $h(u) < 0$ for some $u \in C$ would contradict (1.1)); thus $E(f_*) = 0$. By (1.2), for $\mathring{\tau} := f(\mathring{u}) - h(\mathring{u}) > f(\mathring{u}) \ge f_*$, $e(\mathring{u};\mathring{\tau}) = h(\mathring{u}) < 0$ implies $E(\mathring{\tau}) < 0$, so by convexity, we have $E(\tau) > 0$ for $\tau < f_*$, $E(\tau) < 0$ for $\tau \in (f_*, \mathring{\tau}]$, as well as $E(\tau) < 0$ for $\tau > \mathring{\tau}$ by monotonicity.

(vi) E is proper by (ii), $f_* > -\infty$ by (iv), and (v) yields the conclusion. □

Let $U := \{u \in C : h(u) \le 0\}$ and $U_* := \operatorname{Arg\,min}_U f$ denote the *feasible* and *optimal* sets of problem (1.1). We shall need the following extension of [Kiw85, Lem. 1.2.16].

Lemma 2.2. *Let $\bar{u} \in C$, $\bar{c} \ge 0$, $\bar{\tau} := \pi(\bar{u};\bar{c})$ (cf. (2.1)). Then the following are equivalent:*
(a) *$\bar{u} \in U_*$ (i.e., $\bar{u}$ solves problem (1.1));*
(b) *$E(\bar{\tau}) = e_C(\bar{u};\bar{\tau})$ (i.e., $\bar{u}$ minimizes $e(\cdot;\bar{\tau})$ over C);*
(c) *$0 \in \partial e_C(\bar{u};\bar{\tau})$ (i.e., $0 \in \partial\psi(\bar{u})$, where $\psi(\cdot) := e_C(\cdot;\bar{\tau})$).*

Proof. First, (a) implies $\bar{\tau} = f(\bar{u}) = f_*$, $e(\bar{u};\bar{\tau}) = 0$, $E(\bar{\tau}) = 0$ by Lemma 2.1(v), and hence (b). Since (b) means $\bar{u} \in \operatorname{Arg\,min} e_C(\cdot;\bar{\tau})$, (b) and (c) are equivalent. Next, note that

$$\partial e_C(\bar{u};\bar{\tau}) = \partial i_C(\bar{u}) + \begin{cases} \partial f(\bar{u}) & \text{if } f(\bar{u}) - \bar{\tau} > h(\bar{u}), \\ \operatorname{co}\{\partial f(\bar{u}) \cup \partial h(\bar{u})\} & \text{if } f(\bar{u}) - \bar{\tau} = h(\bar{u}), \\ \partial h(\bar{u}) & \text{if } f(\bar{u}) - \bar{\tau} < h(\bar{u}). \end{cases} \tag{2.3}$$

Finally, (c) implies $h(\bar{u}) \le 0$ (otherwise $h(\bar{u}) > 0 \ge f(\bar{u}) - \bar{\tau}$ and $0 \in \partial e_C(\bar{u};\bar{\tau}) = \partial h(\bar{u}) + \partial i_C(\bar{u})$ would give $\min_C h = h(\bar{u}) > 0$, contradicting (1.2)), so the facts that $\bar{\tau} = f(\bar{u})$ and $E(\bar{\tau}) = e(\bar{u};\bar{\tau}) = 0$ yield $\bar{\tau} = f_*$ by Lemma 2.1(v), and hence (a). □

Lemma 2.2 suggests the following algorithmic scheme: Given the current iterate $\hat{u} \in C$ and the target $\hat{\tau} := \pi(\hat{u};\hat{c})$ for a penalty parameter $\hat{c} \ge 0$, find an approximate minimizer u of $e_C(\cdot;\hat{\tau})$, replace $\hat{u}$ by u, and repeat. Note that if $e_C(u;\hat{\tau}) < e_C(\hat{u};\hat{\tau})$, then u is better than $\hat{u}$: either $f(u) < f(\hat{u})$ and $u \in U$ if $\hat{u} \in U$, or $h(u) < h(\hat{u})$ if $\hat{u} \notin U$. To progress towards the optimal set U_*, it helps if $e_C(\bar{u};\hat{\tau}) \le e_C(\hat{u};\hat{\tau})$ for any optimal $\bar{u} \in U_*$; the sufficient condition given below employs the minimal multiplier $\bar{\mu}$ of §2.1.

Lemma 2.3. *Let $\bar{u} \in U_*$, $\hat{u} \in C$, $\hat{c} \ge 0$, $\hat{\tau} := \pi(\hat{u};\hat{c})$. Then $e(\hat{u};\hat{\tau}) = h(\hat{u})_+$, and $e(\bar{u};\hat{\tau}) \le e(\hat{u};\hat{\tau})$ iff $f(\bar{u}) \le \pi(\hat{u};\hat{c}+1)$. In particular, $f(\bar{u}) \le \pi(\hat{u};\hat{c}+1)$ if $\hat{c} \ge \bar{\mu} - 1$.*

Proof. First, $\hat{\tau} = f(\hat{u})$ and $e(\hat{u};\hat{\tau}) = 0$ if $h(\hat{u}) \le 0$, $e(\hat{u};\hat{\tau}) = h(\hat{u})$ if $h(\hat{u}) > 0$. Next,

$$e(\bar{u};\hat{\tau}) - e(\hat{u};\hat{\tau}) = \max\{f(\bar{u}) - \pi(\hat{u};\hat{c}+1), h(\bar{u}) - h(\hat{u})_+\}$$

is nonpositive iff $f_* = f(\bar{u}) \le \pi(\hat{u};\hat{c}+1)$; the latter holds if $\hat{c}+1 \ge \bar{\mu}$ (see §2.1). □

2.3 An overview of the method

Our method generates a sequence of *trial points* $\{u^k\}_{k=1}^\infty \subset C$ for evaluating the approximate values $f_u^k := f_{u^k}$, $h_u^k := h_{u^k}$, subgradients $g_f^k := g_f^{u^k}$, $g_h^k := g_h^{u^k}$ and linearizations $f_k := \bar{f}_{u^k}$, $h_k := \bar{h}_{u^k}$ of f and h at u^k, respectively, such that

$$f_k(\cdot) = f_u^k + \langle g_f^k, \cdot - u^k \rangle \leq f(\cdot) \quad \text{with} \quad f_k(u^k) = f_u^k \geq f(u^k) - \epsilon_f, \tag{2.4a}$$

$$h_k(\cdot) = h_u^k + \langle g_h^k, \cdot - u^k \rangle \leq h(\cdot) \quad \text{with} \quad h_k(u^k) = h_u^k \geq h(u^k) - \epsilon_h, \tag{2.4b}$$

as stipulated in (1.3). At iteration k, the polyhedral *cutting-plane models* of f and h

$$\check{f}_k(\cdot) := \max_{j \in J_f^k} f_j(\cdot) \leq f(\cdot) \quad \text{with} \quad k \in J_f^k \subset \{1, \ldots, k\}, \tag{2.5a}$$

$$\check{h}_k(\cdot) := \max_{j \in J_h^k} h_j(\cdot) \leq h(\cdot) \quad \text{with} \quad k \in J_h^k \subset \{1, \ldots, k\}, \tag{2.5b}$$

which stem from the accumulated linearizations, yield the relaxed version of problem (1.1)

$$\check{f}_*^k := \inf\{\check{f}_k(u) : u \in \check{H}_k \cap C\} \quad \text{with} \quad \check{H}_k := \{u : \check{h}_k(u) \leq 0\}, \tag{2.6}$$

in which $\check{H}_k$ is an outer approximation of $H := \{u : h(u) \leq 0\}$. The current *prox* (or *stability*) *center* $\hat{u}^k := u^{k(l)} \in C$ for some $k(l) \leq k$ has the values $f_{\hat{u}}^k = f_u^{k(l)}$ and $h_{\hat{u}}^k = h_u^{k(l)}$,

$$f_{\hat{u}}^k \in [f(\hat{u}^k) - \epsilon_f, f(\hat{u}^k)] \quad \text{and} \quad h_{\hat{u}}^k \in [h(\hat{u}^k) - \epsilon_h, h(\hat{u}^k)]. \tag{2.7}$$

As in (2.2) and Lemma 2.2, our improvement function for subproblem (2.6) is given by

$$\check{e}_k(\cdot) := \max\{\check{f}_k(\cdot) - \tau_k, \check{h}_k(\cdot)\} \quad \text{with} \quad \tau_k := f_{\hat{u}}^k + c_k[h_{\hat{u}}^k]_+ \tag{2.8}$$

for some penalty coefficient $c_k \geq 0$ and $[\cdot]_+ := \max\{\cdot, 0\}$. We solve a proximal version of the relaxed improvement problem $\check{E}_k := \inf \check{e}_C^k$ with $\check{e}_C^k := \check{e}_k + i_C$ by finding the trial point

$$u^{k+1} := \arg\min \left\{ \phi_k(\cdot) := \check{e}_k(\cdot) + i_C(\cdot) + \tfrac{1}{2t_k} |\cdot - \hat{u}^k|^2 \right\}, \tag{2.9}$$

where $t_k > 0$ is a *stepsize* that controls the size of $|u^{k+1} - \hat{u}^k|$. For deciding whether u^{k+1} is better than $\hat{u}^k$, we use approximate values of the improvement function $e(\cdot; \tau_k)$. Thus, $e(\hat{u}^k; \tau_k)$ is approximated by $[h_{\hat{u}}^k]_+$, and $e(\hat{u}^k; \tau_k) - \check{e}_k(u^{k+1})$ by the *predicted decrease*

$$v_k := [h_{\hat{u}}^k]_+ - \check{e}_k(u^{k+1}). \tag{2.10}$$

When $f_{\hat{u}}^k < \check{f}_k(\hat{u}^k)$ or $h_{\hat{u}}^k < \check{h}_k(\hat{u}^k)$ due to inexact evaluations, v_k may be nonpositive; if necessary, we increase t_k, as well as c_k in (2.8) if $h_{\hat{u}}^k > 0$, and recompute u^{k+1} to decrease $\check{e}_k(u^{k+1})$ until $v_k \geq |u^{k+1} - \hat{u}^k|^2/2t_k$ (as motivated below). Of course, $e(u^{k+1}; \tau_k)$ is approximated by $\max\{f_u^{k+1} - \tau_k, h_u^{k+1}\}$. A *descent* step to $\hat{u}^{k+1} := u^{k+1}$ occurs if $\max\{f_u^{k+1} - \tau_k, h_u^{k+1}\} \leq [h_{\hat{u}}^k]_+ - \kappa v_k$ for a fixed $\kappa \in (0, 1)$. Otherwise, a *null* step $\hat{u}^{k+1} := \hat{u}^k$ improves the next models $\check{f}_{k+1}$, $\check{h}_{k+1}$ with the new linearizations f_{k+1} and h_{k+1} (cf. (2.5)).

2.4 Aggregate linearizations and an optimality estimate

Extending the approach of [Kiw06c], we now use optimality conditions for subproblem (2.9) to derive aggregate linearizations (i.e., affine minorants) of the problem functions at u^{k+1}, as well as an optimality estimate (see (2.22) below) related to Lemma 2.1(vi).

Lemma 2.4. (i) *There exist subgradients* p_f^k, p_h^k, p_C^k *and a multiplier* ν_k *such that*

$$p_f^k \in \partial \check{f}_k(u^{k+1}),\ p_h^k \in \partial \check{h}_k(u^{k+1}),\ p_C^k \in \partial i_C(u^{k+1}), \tag{2.11}$$

$$\nu_k p_f^k + (1-\nu_k)p_h^k + p_C^k = -(u^{k+1} - \hat{u}^k)/t_k, \tag{2.12}$$

$$\nu_k \in [0,1],\ \nu_k[\check{e}_k(u^{k+1}) - \check{f}_k(u^{k+1}) + \tau_k] = 0,\ (1-\nu_k)[\check{e}_k(u^{k+1}) - \check{h}_k(u^{k+1})] = 0. \tag{2.13}$$

(ii) *These subgradients determine the following* aggregate linearizations

$$\bar{f}_k(\cdot) := \check{f}_k(u^{k+1}) + \langle p_f^k, \cdot - u^{k+1}\rangle \le \check{f}_k(\cdot) \le f(\cdot), \tag{2.14}$$

$$\bar{h}_k(\cdot) := \check{h}_k(u^{k+1}) + \langle p_h^k, \cdot - u^{k+1}\rangle \le \check{h}_k(\cdot) \le h(\cdot), \tag{2.15}$$

$$\bar{\imath}_C^k(\cdot) := i_C(u^{k+1}) + \langle p_C^k, \cdot - u^{k+1}\rangle \le i_C(\cdot), \tag{2.16}$$

$$\bar{e}_C^k(\cdot) := \nu_k[\bar{f}_k(\cdot) - \tau_k] + (1-\nu_k)\bar{h}_k(\cdot) + \bar{\imath}_C^k(\cdot) \le \check{e}_C^k(\cdot) \le e_C(\cdot;\tau_k). \tag{2.17}$$

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

$$p^k := \nu_k p_f^k + (1-\nu_k)p_h^k + p_C^k = (\hat{u}^k - u^{k+1})/t_k \quad \textit{and} \quad \epsilon_k := [h_{\hat{u}}^k]_+ - \bar{e}_C^k(\hat{u}^k), \tag{2.18}$$

and the optimality measure

$$V_k := \max\{|p^k|, \epsilon_k + \langle p^k, \hat{u}^k\rangle\}, \tag{2.19}$$

we have

$$\bar{e}_C^k(\cdot) = \check{e}_k(u^{k+1}) + \langle p^k, \cdot - u^{k+1}\rangle, \tag{2.20}$$

$$[h_{\hat{u}}^k]_+ - \epsilon_k + \langle p^k, \cdot - \hat{u}^k\rangle = \bar{e}_C^k(\cdot) \le \check{e}_C^k(\cdot) \le e_C(\cdot;\tau_k), \tag{2.21}$$

$$e_C(u;\tau_k) \ge \check{e}_C^k(u) \ge [h_{\hat{u}}^k]_+ - V_k(1+|u|) \quad \textit{for all } u. \tag{2.22}$$

Proof. (i) Use the optimality condition $0 \in \partial\phi_k(u^{k+1})$ for (2.9) and the form (2.8) of $\check{e}_k$.

(ii) The first inequalities in (2.14)–(2.15) stem from (2.11), and the final ones from (2.5). Similarly, (2.11) gives (2.16) with $i_C(u^{k+1}) = 0$. Then (2.17) follows from the facts that $\nu \in [0,1]$ (cf. (2.13)) yields $\nu_k(\bar{f}_k - \tau_k) + (1-\nu_k)\bar{h}_k \le \check{e}_k$ by using $\bar{f}_k \le \check{f}_k$ and $\bar{h}_k \le \check{h}_k$ in (2.8), and that $\check{e}_C^k := \check{e}_k + i_C \le e_C(\cdot;\tau_k)$ by using $\check{f}_k \le f$ and $\check{h}_k \le h$ in (2.2).

(iii) For (2.20), use (2.12)–(2.13) and the definitions in (2.14)–(2.18); since $\bar{e}_C^k$ is affine, its expression in (2.21) follows from (2.18). Finally, since by the Cauchy-Schwarz inequality,

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

in (2.21), we obtain (2.22) from the definition of V_k in (2.19). □

Observe that V_k is indeed an optimality measure: if $V_k = 0$ in (2.22), then $E(\tau_k) \ge 0$ gives $f_{\hat{u}}^k \le \tau_k \le f_*$ by Lemma 2.1(vi); similar relations hold asymptotically.

2.5 Ensuring sufficient predicted decrease

In view of the optimality estimate (2.22), we would like V_k to vanish asymptotically. Hence it is crucial to bound V_k via the predicted decrease v_k, since normally bundling and descent steps drive v_k to 0. The necessary bounds are given below.

Lemma 2.5. (i) *In the notation of* (2.18), *the predicted decrease* v_k *of* (2.10) *satisfies*

$$v_k = t_k|p^k|^2 + \epsilon_k. \tag{2.23}$$

(ii) *We have* $v_k \geq -\epsilon_k \Leftrightarrow t_k|p^k|^2/2 \geq -\epsilon_k \Leftrightarrow v_k \geq t_k|p^k|^2/2 = |u^{k+1} - \hat{u}^k|/2t_k$.

(iii) *For the* maximal evaluation error $\epsilon_{\max} := \max\{\epsilon_f, \epsilon_h\}$, *we have*

$$-\epsilon_k \leq \epsilon_{\max}. \tag{2.24}$$

(iv) *The optimality measure of* (2.19) *satisfies* $V_k \leq \max\{|p^k|, \epsilon_k\}(1 + |\hat{u}^k|)$. *Moreover,*

$$v_k \geq \max\{t_k|p^k|^2/2, |\epsilon_k|\} \qquad \text{if} \quad v_k \geq -\epsilon_k, \tag{2.25}$$

$$V_k \leq \max\{(2v_k/t_k)^{1/2}, v_k\}(1 + |\hat{u}^k|) \quad \text{if} \quad v_k \geq -\epsilon_k, \tag{2.26}$$

$$V_k < (2\epsilon_{\max}/t_k)^{1/2}(1 + |\hat{u}^k|) \qquad \text{if} \quad v_k < -\epsilon_k. \tag{2.27}$$

Proof. (i) We have $\langle p^k, u^{k+1} - \hat{u}^k\rangle = -t_k|p^k|^2$ by (2.18), whereas by (2.20),

$$\check{e}_k(u^{k+1}) = \bar{e}_C^k(u^{k+1}) = \bar{e}_C^k(\hat{u}^k) + \langle p^k, u^{k+1} - \hat{u}^k\rangle,$$

so $v_k := [h_{\hat{u}}^k]_+ - \check{e}_k(u^{k+1}) = \epsilon_k + t_k|p^k|^2$ by (2.18). Note that $v_k \geq \epsilon_k$.

(ii) This follows from (2.23) and the first part of (2.18).

(iii) By the definitions of $\bar{e}_C^k$ and ϵ_k in (2.17)–(2.18), we may express $-\epsilon_k$ as follows

$$-\epsilon_k = \nu_k[\bar{f}_k(\hat{u}^k) - \tau_k] + (1 - \nu_k)\bar{h}_k(\hat{u}^k) + \bar{\imath}_C^k(\hat{u}^k) - [h_{\hat{u}}^k]_+,$$

where $\nu_k \in [0, 1]$ by (2.13), $\bar{f}_k(\hat{u}^k) \leq f(\hat{u}^k) \leq f_{\hat{u}}^k + \epsilon_f$, $\bar{h}_k(\hat{u}^k) \leq h(\hat{u}^k) \leq h_{\hat{u}}^k + \epsilon_h$ and $\bar{\imath}_C^k(\hat{u}^k) \leq i_C(\hat{u}^k) = 0$ by (2.14)-(2.16) and (2.7), and $\tau_k \geq f_{\hat{u}}^k$ by (2.8). Therefore, we have

$$-\epsilon_k \leq \nu_k\epsilon_f + (1 - \nu_k)h(\hat{u}^k) - (1 - \nu_k)[h_{\hat{u}}^k]_+ \leq \nu_k\epsilon_f + (1 - \nu_k)\epsilon_h \leq \epsilon_{\max}.$$

(iv) Since $V_k \leq \max\{|p^k|, \epsilon_k\}(1 + |\hat{u}^k|)$ by (2.19) and the Cauchy-Schwarz inequality, the bounds follow from the equivalences in statement (ii), using $v_k \geq \epsilon_k$ and (2.24). □

The bound (2.27) will imply that if $\tau_k > f_*$ (so that $E(\tau_k) < 0$ and V_k can't vanish in (2.22) as t_k increases), then $v_k \geq -\epsilon_k$ and the bound (2.26) hold for t_k large enough.

2.6 Linearization selection

For choosing the sets J_f^{k+1} and J_h^{k+1}, note that (2.4)–(2.5) and (2.11) yield the existence of multipliers α_j^k for the pieces f_j, $j \in J_f^k$, and β_j^k for the pieces h_j, $j \in J_h^k$, such that

$$(p_f^k, 1) = \sum_{j \in J_f^k} \alpha_j^k (\nabla f_j, 1) \ \alpha_j^k \geq 0, \ \alpha_j^k[\check{f}_k(u^{k+1}) - f_j(u^{k+1})] = 0, \ j \in J_f^k, \tag{2.28a}$$

$$(p_h^k, 1) = \sum_{j \in J_h^k} \beta_j^k (\nabla h_j, 1) \ \beta_j^k \geq 0, \ \beta_j^k[\check{h}_k(u^{k+1}) - h_j(u^{k+1})] = 0, \ j \in J_h^k. \tag{2.28b}$$

Denote the indices of linearizations f_j and h_j that are "strongly" active at u^{k+1} by

$$\hat{J}_f^k := \{j \in J_f^k : \alpha_j^k \neq 0\} \quad \text{and} \quad \hat{J}_h^k := \{j \in J_h^k : \beta_j^k \neq 0\}. \tag{2.29}$$

These linearizations embody all the information contained in the aggregates $\bar{f}_k$ and $\bar{h}_k$ (which are actually their convex combinations; cf. (2.14)–(2.15) and (2.28)). To save storage and work per iteration, we may drop the remaining linearizations.

2.7 The method

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

Algorithm 2.6.
Step 0 (*Initialization*). Select $u^1 \in C$, a *descent parameter* $\kappa \in (0,1)$, an *infeasibility contraction bound* $\kappa_h \in (0,1]$, a *stepsize bound* $t_{\min} > 0$, a stepsize $t_1 \geq t_{\min}$ and a penalty coefficient $c_1 \geq 0$. Set $\hat{u}^1 := u^1$, $f_{\hat{u}}^1 := f_u^1 := f_{u^1}$, $g_f^1 := g_f^{u^1}$, $h_{\hat{u}}^1 := h_u^1 := h_{u^1}$, $g_h^1 := g_h^{u^1}$ (cf. (2.4)), $J_f^1 := J_h^1 := \{1\}$, $i_t^1 := 0$, $k := k(0) := 1$, $l := 0$ ($k(l) - 1$ will denote the iteration of the lth descent step).

Step 1 (*Trial point finding*). For $\check{e}_k$ given by (2.8), find u^{k+1} (cf. (2.9)) and multipliers α_j^k, β_j^k such that (2.28) holds. Set v_k by (2.10), $p^k := (\hat{u}^k - u^{k+1})/t_k$ and $\epsilon_k := v_k - t_k|p^k|^2$.

Step 2 (*Stopping criterion*). If $V_k = 0$ (cf. (2.19)) and $h_{\hat{u}}^k \leq 0$, stop ($f_{\hat{u}}^k \leq f_*$).

Step 3 (*Phase 1 stepsize correction*). If $h_{\hat{u}}^k \leq 0$ or $\epsilon_{\max} = 0$ or $v_k \geq \kappa_h h_{\hat{u}}^k$, go to Step 4. Set $t_k := 10t_k$, $i_t^k := k$. If $c_k > 0$, set $c_k := 2c_k$; otherwise, pick $c_k > 0$. Go back to Step 1.

Step 4 (*Stepsize correction*). If $v_k \geq -\epsilon_k$, go to Step 5. Set $t_k := 10t_k$, $i_t^k := k$. If $h_{\hat{u}}^k > 0$, set $c_k := 2c_k$ if $c_k > 0$, pick $c_k > 0$ otherwise. Go back to Step 1.

Step 5 (*Descent test*). Evaluate f_{k+1} and h_{k+1} (cf. (2.4)). If the *descent test* holds:

$$\max\{f_u^{k+1} - \tau_k, h_u^{k+1}\} \leq [h_{\hat{u}}^k]_+ - \kappa v_k, \tag{2.30}$$

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

Step 6 (*Bundle selection*). For the active sets $\hat{J}_f^k$ and $\hat{J}_h^k$ given by (2.29), choose

$$J_f^{k+1} \supset \hat{J}_f^k \cup \{k+1\} \quad \text{and} \quad J_h^{k+1} \supset \hat{J}_h^k \cup \{k+1\}. \tag{2.31}$$

Step 7 (*Stepsize updating*). If $k(l) = k+1$ (i.e., after a descent step), select $t_{k+1} \geq t_k$ and $c_{k+1} \geq 0$; otherwise, set $c_{k+1} := c_k$ and 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.

Remarks 2.7. (i) When the set C is polyhedral, Step 1 may use the QP method of [Kiw94], which can solve efficiently sequences of related subproblems (2.9).

(ii) Step 2 may also use the test $\inf \check{e}_C^k \geq 0$ and $h_{\hat{u}}^k \leq 0$ (see Lemma 2.8(i) below).

(iii) Step 3 is needed in phase 1 (for $h_{\hat{u}}^k > 0$) when inaccuracies occur ($\epsilon_{\max} > 0$); it increases t_k and τ_k (via c_k) to obtain $v_k \geq \kappa_h h_{\hat{u}}^k$, so that eventually a descent step (cf. (2.30)) will reduce the constraint violation significantly: $h_{\hat{u}}^{k+1} \leq (1-\kappa\kappa_h) h_{\hat{u}}^k$.

(iv) In the case of exact evaluations ($\epsilon_{\max} = 0$), Step 4 is redundant, since $v_k \geq \epsilon_k \geq 0$ (cf. (2.23)–(2.24)). When inexactness is discovered via $v_k < -\epsilon_k$, t_k is increased to produce descent or confirm that $\hat{u}^k$ is almost optimal. Namely, when $\hat{u}^k$ is bounded in (2.27), increasing t_k drives V_k to 0, so that $f_{\hat{u}}^k \leq \tau_k \leq f_*$ asymptotically. Whenever t_k is increased at Steps 3 or 4, 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 5). Otherwise, decreasing t_k at Step 7 aims at collecting more local information about f and h at null steps.

(v) When $\epsilon_{\max} := \max\{\epsilon_f, \epsilon_h\} = 0$, our method employs the exact function values

$$f_{\hat{u}}^k = f(\hat{u}^k), \quad h_{\hat{u}}^k = h(\hat{u}^k), \quad \tau_k = \pi(\hat{u}^k; c_k) \geq f(\hat{u}^k) \quad \text{and} \quad [h_{\hat{u}}^k]_+ = e(\hat{u}^k; \tau_k) \tag{2.32}$$

(cf. (2.7), (2.1), (2.8) and Lem. 2.3), and the aggregate inequality (2.21) means that

$$p^k \in \partial_{\epsilon_k} e_C(\hat{u}^k; \tau_k) \quad \text{with} \quad \epsilon_k \geq 0. \tag{2.33}$$

Thus, if $V_k = 0$ in (2.19), then $|p^k| = \epsilon_k = 0$ imply that $0 \in \partial e_C(\hat{u}^k; \tau_k)$ and hence that $\hat{u}^k \in U_*$ by Lemma 2.2; in particular, in this case we have $h_{\hat{u}}^k = h(\hat{u}^k) \leq 0$.

(vi) At Step 5, we have $v_k > 0$ (using (2.26) and $V_k > 0$ at Step 2 if $h_{\hat{u}}^k \leq 0$; otherwise $v_k \geq \kappa_h h_{\hat{u}}^k > 0$ by Step 3 if $\epsilon_{\max} > 0$, $V_k > 0$ by item (v) if $\epsilon_{\max} = 0$). When a descent step occurs, the descent test (2.30) with the target τ_k given by (2.8) implies that

$$h_{\hat{u}}^{k+1} \leq h_{\hat{u}}^k - \kappa v_k \qquad \text{if } h_{\hat{u}}^k > 0, \tag{2.34a}$$

$$f_{\hat{u}}^{k+1} \leq f_{\hat{u}}^k - \kappa v_k \quad \text{and} \quad h_{\hat{u}}^{k+1} \leq 0 \quad \text{if } h_{\hat{u}}^k \leq 0. \tag{2.34b}$$

Thus at phase 1 (i.e., when $h_{\hat{u}}^k > 0$), we have reduction in the constraint violation, whereas at phase 2 the objective value is decreased while preserving (approximate) feasibility.

(vii) An active-set method for solving (2.9) (cf. [Kiw94]) will produce $|\hat{J}_f^k| + |\hat{J}_h^k| \leq m+1$ (cf. (2.29)). Hence Step 6 can keep $|J_f^{k+1}| + |J_h^{k+1}| \leq \bar{m}$ for any given bound $\bar{m} \geq m+3$.

(viii) Step 7 may use the procedure of [Kiw90, §2] for updating the proximity weight $1/t_k$, with obvious modifications.

We now show that, in phase 2, the loop between Steps 1 and 4 is infinite iff $0 \le \inf \check{e}_C^k < \check{e}_k(\hat{u}^k)$, in which case $\hat{u}^k$ is *approximately optimal:* $f(\hat{u}^k) \le f_* + \epsilon_f$ and $h(\hat{u}^k) \le \epsilon_h$.

Lemma 2.8. *Assuming $h_{\hat{u}}^k \le 0$, recall that $\check{E}_k := \inf \check{e}_C^k$ with $\check{e}_C^k := \check{e}_k + i_C$. Then:*

(i) *If $\check{E}_k \ge 0$, then $f(\hat{u}^k) - \epsilon_f \le f_{\hat{u}}^k \le f_*$ and $h(\hat{u}^k) \le \epsilon_h$.*

(ii) *Step 2 terminates, i.e., $V_k := \max\{|p^k|, \epsilon_k + \langle p^k, \hat{u}^k\rangle\} = 0$, iff $0 \le \check{E}_k = \check{e}_k(\hat{u}^k)$.*

(iii) *If the loop between Steps 1 and 4 is infinite, then $\check{E}_k \ge 0$ and $V_k \to 0$.*

(iv) *If $\check{E}_k \ge 0$ at Step 1 and Step 2 does not terminate (i.e., $\check{E}_k < \check{e}_k(\hat{u}^k)$; cf. (ii)), then an infinite loop between Steps 4 and 1 occurs.*

Proof. (i) We have $E(\tau_k) \ge \check{E}_k$ and $\tau_k = f_{\hat{u}}^k$ (cf. (2.2), (2.8), (2.14)–(2.15)), so $f_{\hat{u}}^k \le f_*$ by Lemma 2.1(vi), whereas $f(\hat{u}^k) \le f_{\hat{u}}^k + \epsilon_f$ and $h(\hat{u}^k) \le h_{\hat{u}}^k + \epsilon_h$ by (2.7).

(ii) "$\Rightarrow$": Since $|p^k| = 0 \ge \epsilon_k$, (2.18) and (2.21) yield $u^{k+1} = \hat{u}^k$, $\bar{e}_C^k(\hat{u}^k) \le \check{e}_C^k(\cdot)$ and $0 \le \bar{e}_C^k(\hat{u}^k)$, whereas by (2.20), $\bar{e}_C^k(\hat{u}^k) = \check{e}_k(u^{k+1}) = \check{e}_k(\hat{u}^k)$. "$\Leftarrow$": Since $\check{e}_C^k(\hat{u}^k) = \min \check{e}_C^k$, using $\phi_k(\hat{u}^k) = \min \check{e}_C^k \le \phi_k(u^{k+1}) \le \phi_k(\hat{u}^k)$ in (2.9) gives $u^{k+1} = \hat{u}^k$, so again $\bar{e}_C^k(\hat{u}^k) = \check{e}_C^k(\hat{u}^k)$ by (2.20), and (2.18) yields $p^k = 0$ and $\epsilon_k = -\check{e}_C^k(\hat{u}^k) \le 0$.

(iii) At Step 4 during the loop the facts that $V_k < (2\epsilon_{\max}/t_k)^{1/2}(1 + |\hat{u}^k|)$ (cf. (2.27)) and $t_k \uparrow \infty$ as the loop continues give $V_k \to 0$, so $\check{e}_C^k(\cdot) \ge 0$ by (2.22).

(iv) We have $\check{e}_k(u^{k+1}) \ge \inf \check{e}_C^k \ge 0$. Thus $v_k = -\check{e}_k(u^{k+1}) \le 0$ (cf. (2.10)) and (cf. (2.23)) $v_k = t_k|p^k|^2 + \epsilon_k$ yield $\epsilon_k \le -t_k|p^k|^2$ at Step 4 with $p^k \ne 0$ (since $\max\{|p^k|, \epsilon_k + \langle p^k, \hat{u}^k\rangle\} =: V_k > 0$ at Step 2). Hence $\epsilon_k < -\frac{t_k}{2}|p^k|^2$, so $v_k < -\epsilon_k$ (cf. (2.23)) and Step 4 loops back to Step 1, after which Step 2 can't terminate due to (ii). □

3 Convergence

In view of Lemma 2.8, we may suppose that the algorithm neither terminates nor loops infinitely between Steps 1 and 4 at phase 2 (otherwise $\hat{u}^k$ is approximately optimal). For phase 1, our analysis will imply that any loop between Steps 1 and 3 or 4 is finite. We shall show that the algorithm generates points that are approximately optimal asymptotically by establishing upper bounds on the values $f_{\hat{u}}^k$ and $h_{\hat{u}}^k$. We first bound $f_{\hat{u}}^k$ via V_k.

Lemma 3.1. *Let $K \subset \mathbb{N}$ be such that $V_k \xrightarrow{K} 0$. Then $\overline{\lim}_{k\in K} f_{\hat{u}}^k \le \overline{\lim}_{k\in K} \tau_k \le f_*$.*

Proof. Pick $K' \subset K$ such that $\tau_k \xrightarrow{K'} \bar{\tau} := \overline{\lim}_{k\in K} \tau_k$. Since $f_{\hat{u}}^k \le \tau_k$ by (2.8), we need only show that $\bar{\tau} \le f_*$ when $\bar{\tau} > -\infty$. Note that $\bar{\tau} < \infty$, since otherwise for $\tau_k \ge f(\mathring{u}) - h(\mathring{u})$, the fact that $e(\mathring{u}; \tau_k) = h(\mathring{u}) < 0$ (cf. (2.2), (1.2)) and the bound (2.22) would yield

$$0 > h(\mathring{u}) = e_C(\mathring{u}; \tau_k) \ge -V_k(1 + |\mathring{u}|) \xrightarrow{K'} 0,$$

a contradiction. Thus $\bar{\tau}$ is finite. Since $e_C(u; \cdot)$ is continuous, letting $k \xrightarrow{K'} \infty$ in (2.22) gives $e_C(\cdot; \bar{\tau}) \ge 0$. Therefore, we have $E(\bar{\tau}) \ge 0$, and hence $\bar{\tau} \le f_*$ by Lemma 2.1(vi). □

The upper bound of Lemma 3.1 is complemented below with a lower bound (which is highly useful for the "dual" applications in §4.3 and §5).

Lemma 3.2. *If* $\overline{\lim}_k h_{\hat{u}}^k \le 0$, *then for the minimal multiplier* $\bar{\mu} := \inf_{\mu \in M} \mu$ *(cf. §2.1),*

$$\underline{\lim}_k f_{\hat{u}}^k + \epsilon_f \ge \underline{\lim}_k f(\hat{u}^k) \ge f_* - \bar{\mu}\epsilon_h \quad \textit{and} \quad \overline{\lim}_k h(\hat{u}^k) \le \epsilon_h. \tag{3.1}$$

Proof. For all k, $f(\hat{u}^k) \le f_{\hat{u}}^k + \epsilon_f$, $h(\hat{u}^k) \le h_{\hat{u}}^k + \epsilon_h$ by (2.7), $L(\hat{u}^k; \bar{\mu}) = f(\hat{u}^k) + \bar{\mu} h(\hat{u}^k) \ge f_*$ with $\hat{u}^k \in C$ and $0 \le \bar{\mu} < \infty$ if $f_* > -\infty$, $\bar{\mu} = \infty$ if $f_* = -\infty$; the conclusion follows. □

We first consider the case where only finitely many descent steps occur. After the last descent step, only null steps occur and $\{t_k\}$ becomes eventually monotone, since once Steps 3 or 4 increase t_k, Step 7 can't decrease t_k; thus the limit $t_\infty := \lim_k t_k$ exists. After showing that $t_\infty = \infty$ may occur only at phase 2 in Lemma 3.3 below, we deal with the cases of $t_\infty = \infty$ in Lemma 3.4 and $t_\infty < \infty$ in Lemma 3.5.

Lemma 3.3. *Suppose there exists* $\bar{k}$ *such that* $h_{\hat{u}}^{\bar{k}} > 0$ *and only null steps occur for all* $k \ge \bar{k}$. *Then Steps* 3 *and* 4 *can increase* t_k *only a finite number of times.*

Proof. For contradiction, suppose $t_k \to \infty$. Since $\tau_k \to \infty$ (cf. Steps 3, 4 and (2.8)), we may assume $\tau_k \ge \mathring{\tau} := f(\mathring{u}) - h(\mathring{u})$ for the Slater point $\mathring{u}$ of (1.2) and $k \ge \bar{k}$; then using the minorants $\check{f}_k \le f$ and $\check{h}_k \le h$ (cf. (2.4)) in the definitions (2.8) and (2.2) yields

$$\check{e}_k(\mathring{u}) \le \max\{\check{f}_k(\mathring{u}) - \mathring{\tau}, \check{h}_k(\mathring{u})\} \le e(\mathring{u}; \mathring{\tau}) = h(\mathring{u}) < 0 \quad \text{with} \quad \mathring{u} \in C. \tag{3.2}$$

At Step 1, (2.9) gives the proximal projection property for the level set of $\check{e}_C^k := \check{e}_k + i_C$

$$u^{k+1} = \arg\min\{\tfrac{1}{2}|u - \hat{u}^k|^2 : \check{e}_C^k(u) \le \check{e}_C^k(u^{k+1})\}, \tag{3.3}$$

whereas before Step 3 increases t_k, $v_k < \kappa_h h_{\hat{u}}^k$ yields $\check{e}_k(u^{k+1}) > (1 - \kappa_h) h_{\hat{u}}^k \ge 0$ by (2.10), so for $k \ge \bar{k}$, (3.2) and (3.3) give $|u^{k+1} - \hat{u}^k| \le r := |\mathring{u} - \hat{u}^k|$ and hence $|p^k| \le r/t_k$ by (2.18). Therefore, if Step 3 increases t_k at infinitely many iterations, indexed by K say, then $t_k \to \infty$ yields $p^k \xrightarrow{K} 0$, and by (2.21), (2.20) and Cauchy-Schwarz, we get

$$0 > h(\mathring{u}) \ge \check{e}_C^k(\mathring{u}) \ge \bar{e}_C^k(\mathring{u}) = \check{e}_k(u^{k+1}) + \langle p^k, \mathring{u} - u^{k+1} \rangle \ge \langle p^k, \mathring{u} - u^{k+1} \rangle \xrightarrow{K} 0,$$

a contradiction. Similarly, if Step 4 is entered with $v_k < -\epsilon_k$ for infinitely many iterations indexed by K (say), then $t_k \to \infty$ and (2.27) give $V_k \xrightarrow{K} 0$, and we get from (2.22)

$$0 > h(\mathring{u}) \ge \check{e}_C^k(\mathring{u}) \ge -V_k(1 + |\mathring{u}|) \xrightarrow{K} 0,$$

another contradiction. The conclusion follows. □

The case where the stepsize t_k keeps growing at a fixed prox center is quite simple.

Lemma 3.4. *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* $V_k \xrightarrow{K} 0$ *and* $h_{\hat{u}}^{\bar{k}} \le 0$.

Proof. We have $h_{\hat{u}}^{\bar{k}} \le 0$ (otherwise Lemma 3.3 would imply $t_\infty < \infty$, a contradiction). For $k \in K$, before t_k is increased at Step 4 on the last loop to Step 1, we have $V_k < (2\epsilon_{\max}/t_k)^{1/2}(1 + |\hat{u}^{\bar{k}}|)$ by (2.27), so $t_k \to \infty$ gives $V_k \xrightarrow{K} 0$. □

The case where the stepsize t_k doesn't grow at a fixed prox center is analyzed as in [Kiw06c]. After showing that the optimal value $\phi_k(u^{k+1})$ of subproblem (2.9) is nondecreasing and bounded above, u^{k+1} is bounded and $u^{k+2} - u^{k+1} \to 0$, we invoke the descent test (2.30) to get $v_k \to 0$; the rest follows from the bounds (2.25)–(2.26).

Lemma 3.5. *Suppose there exists $\bar{k}$ such that for all $k \geq \bar{k}$, only null steps occur and Steps 3 and 4 don't increase t_k. Then $V_k \to 0$ and $h_{\hat{u}}^{\bar{k}} \leq 0$.*

Proof. Fix $k \geq \bar{k}$. We first show that the aggregate $\bar{e}_C^k$ minorizes the next model $\check{e}_C^{k+1}$:

$$\bar{e}_C^k(\cdot) \leq \check{e}_C^{k+1}(\cdot) := \check{e}_{k+1}(\cdot) + i_C(\cdot). \tag{3.4}$$

Consider the selected model $\hat{f}_k := \max_{j \in \hat{J}_f^k} f_j$ of $\check{f}_k := \max_{j \in J_f^k} f_j$; then $\hat{f}_k \leq \check{f}_k$. Using (2.29) in the expression (2.28a) of p_f^k gives $\hat{f}_k(u^{k+1}) = \check{f}_k(u^{k+1})$ and $p_f^k \in \partial \hat{f}_k(u^{k+1})$ (cf. [HUL93, Ex. VI.3.4]). Thus $\bar{f}_k \leq \hat{f}_k$ by (2.14), so the choice of $\hat{J}_f^k \subset J_f^{k+1}$ implies that $\bar{f}_k \leq \hat{f}_k \leq \check{f}_{k+1}$. Similarly, for $\hat{h}_k := \max_{j \in \hat{J}_h^k} h_j$, (2.28b) yields $\bar{h}_k \leq \hat{h}_k \leq \check{h}_{k+1}$. Then, using the definition (2.17) of $\bar{e}_C^k$ with $\nu_k \in [0,1]$ (cf. (2.13)), the minorization $\bar{\imath}_C^k \leq i_C$ of (2.16) and the fact that $\tau_{k+1} = \tau_k$ (by (2.8) and Steps 3 and 4) gives the required bound

$$\bar{e}_C^k \leq \nu_k[\check{f}_{k+1} - \tau_k] + (1-\nu_k)\check{h}_{k+1} + i_C \leq \max\{\check{f}_{k+1} - \tau_{k+1}, \check{h}_{k+1}\} + i_C = \check{e}_C^{k+1}.$$

(Note that this bound only needs the minorizations $\bar{f}_k \leq \check{f}_{k+1} + i_C$ and $\bar{h}_k \leq \check{h}_{k+1} + i_C$; this will be important when selection is replaced by aggregation in §4.2.)

Next, consider the following partial linearization of the objective ϕ_k of (2.9):

$$\bar{\phi}_k(\cdot) := \bar{e}_C^k(\cdot) + \tfrac{1}{2t_k}|\cdot - \hat{u}^k|^2. \tag{3.5}$$

We have $\bar{e}_C^k(u^{k+1}) = \check{e}_k(u^{k+1})$ by (2.20) and $\nabla\bar{\phi}_k(u^{k+1}) = 0$ from $\nabla \bar{e}_C^k = p^k = (\hat{u}^k - u^{k+1})/t_k$ (cf. (2.20), (2.18)); hence $\bar{\phi}_k(u^{k+1}) = \phi_k(u^{k+1})$ by (2.9), and by Taylor's expansion

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

To bound $\bar{\phi}_k(\hat{u}^k)$ from above, notice that (3.5), (2.18) and (2.24) imply that

$$\bar{\phi}_k(\hat{u}^k) = \bar{e}_C^k(\hat{u}^k) = [h_{\hat{u}}^k]_+ - \epsilon_k \leq [h_{\hat{u}}^k]_+ + \epsilon_{\max}.$$

Then by (3.6),

$$\phi_k(u^{k+1}) + \tfrac{1}{2t_k}|u^{k+1} - \hat{u}^k|^2 = \bar{\phi}_k(\hat{u}^k) \leq [h_{\hat{u}}^{\bar{k}}]_+ + \epsilon_{\max}. \tag{3.7}$$

Now, using the facts that $\hat{u}^{k+1} = \hat{u}^k$ and $t_{k+1} \leq t_k$ and the model minorization property (3.4) in the definitions (3.5) of $\bar{\phi}_k$ and (2.9) of ϕ_{k+1} gives $\bar{\phi}_k \leq \phi_{k+1}$. Hence by (3.6),

$$\phi_k(u^{k+1}) + \tfrac{1}{2t_k}|u^{k+2} - u^{k+1}|^2 = \bar{\phi}_k(u^{k+2}) \leq \phi_{k+1}(u^{k+2}). \tag{3.8}$$

Thus the nondecreasing sequence $\{\phi_k(u^{k+1})\}_{k \geq \bar{k}}$, being bounded above by (3.7) with $\hat{u}^k = \hat{u}^{\bar{k}}$ for $k \geq \bar{k}$, must have a limit, say $\phi_\infty \leq [h_{\hat{u}}^{\bar{k}}]_+ + \epsilon_{\max}$. Moreover, since the stepsizes satisfy $t_k \leq t_{\bar{k}}$ for $k \geq \bar{k}$, we deduce from the bounds (3.7)–(3.8) that

$$\phi_k(u^{k+1}) \uparrow \phi_\infty, \quad u^{k+2} - u^{k+1} \to 0, \tag{3.9}$$

and the sequence $\{u^{k+1}\}$ is bounded. Then the sequence $\{g_f^{k+1}\}$ is bounded as well, since $g_f^k \in \partial_{\epsilon_f} f(u^k)$ by (2.4), whereas the mapping $\partial_{\epsilon_f} f$ is locally bounded [HUL93, §XI.4.1]; similarly, the sequence $\{g_h^{k+1}\}$ is bounded, since $g_h^k \in \partial_{\epsilon_h} h(u^k)$ by (2.4).

For $v_k := [h_{\hat{u}}^k]_+ - \check{e}_k(u^{k+1})$ (cf. (2.10)) and the following linearization of $e(\cdot;\tau_k)$ at u^{k+1}

$$e_{k+1}(\cdot) := \begin{cases} f_{k+1}(\cdot) - \tau_k & \text{if } f_u^{k+1} - \tau_k \geq h_u^{k+1}, \\ h_{k+1}(\cdot) & \text{otherwise,} \end{cases} \tag{3.10}$$

the descent test (2.30) reads $e_{k+1}(u^{k+1}) \leq [h_{\hat{u}}^k]_+ - \kappa v_k$ or equivalently

$$\tilde{\epsilon}_k := e_{k+1}(u^{k+1}) - \check{e}_k(u^{k+1}) \leq (1-\kappa) v_k. \tag{3.11}$$

We now show that this approximation error $\tilde{\epsilon}_k \to 0$. First, note that the linearization gradients $g_e^{k+1} := \nabla e_{k+1}$ are bounded, since $|g_e^{k+1}| \leq \max\{|g_f^{k+1}|, |g_h^{k+1}|\}$ by (2.4). Further, the minorizations $f_{k+1} \leq \check{f}_{k+1}$ and $h_{k+1} \leq \check{h}_{k+1}$ due to $k+1 \in J_f^{k+1} \cap J_h^{k+1}$ (cf. (2.5)) yield $e_{k+1} \leq \check{e}_{k+1}$ by (2.8), since $\tau_{k+1} = \tau_k$. Using the linearity of e_{k+1}, the bound $e_{k+1} \leq \check{e}_{k+1}$, the Cauchy-Schwarz inequality and (2.9) with $\hat{u}^k = \hat{u}^{\bar{k}}$ for $k \geq \bar{k}$, we estimate

$$\begin{aligned} \tilde{\epsilon}_k &:= e_{k+1}(u^{k+1}) - \check{e}_k(u^{k+1}) = e_{k+1}(u^{k+2}) - \check{e}_k(u^{k+1}) + \langle g_e^{k+1}, u^{k+1} - u^{k+2} \rangle \\ &\leq \check{e}_{k+1}(u^{k+2}) - \check{e}_k(u^{k+1}) + |g_e^{k+1}||u^{k+1} - u^{k+2}| \\ &= \phi_{k+1}(u^{k+2}) - \phi_k(u^{k+1}) + \Delta_k + |g_e^{k+1}||u^{k+1} - u^{k+2}|, \end{aligned} \tag{3.12}$$

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 $t_{\min} \leq t_{k+1} \leq t_k$ for $k \geq \bar{k}$ (cf. Step 7), $|u^{k+1} - \hat{u}^{\bar{k}}|^2$ is bounded, $u^{k+2} - u^{k+1} \to 0$ by (3.9), and

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

Hence, using (3.9) and the boundedness of $\{g_e^{k+1}\}$ in (3.12) yields $\overline{\lim}_k \tilde{\epsilon}_k \leq 0$. On the other hand, for $k \geq \bar{k}$, the descent test written as (3.11) fails: $(1-\kappa)v_k < \tilde{\epsilon}_k$, where $\kappa < 1$ and $v_k > 0$; it follows that $\tilde{\epsilon}_k \to 0$ and $v_k \to 0$.

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 (2.26), $\epsilon_k \to 0$ and $|p^k| \to 0$ by (2.25). It remains to prove that $h_{\hat{u}}^{\bar{k}} \leq 0$. If $\epsilon_{\max} > 0$, but $h_{\hat{u}}^{\bar{k}} > 0$, then the facts that $v_k \to 0$ with $v_k \geq \kappa_h h_{\hat{u}}^k$ (cf. Step 3), $\kappa_h > 0$ and $h_{\hat{u}}^k = h_{\hat{u}}^{\bar{k}}$ for $k \geq \bar{k}$ give in the limit $h_{\hat{u}}^{\bar{k}} \leq 0$, a contradiction. Finally, for $\epsilon_{\max} = 0$, recalling Remark 2.7(v) and using $\epsilon_k, |p^k| \to 0$ in (2.21) yields $e_C(\hat{u}^{\bar{k}}; \tau_{\bar{k}}) \leq e_C(\cdot; \tau_{\bar{k}})$. In other words, we have $0 \in \partial e_C(\hat{u}^{\bar{k}}; \tau_{\bar{k}})$, so $\hat{u}^{\bar{k}} \in U_*$ by Lemma 2.2 and thus $h_{\hat{u}}^{\bar{k}} = h(\hat{u}^{\bar{k}}) \leq 0$. □

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

Theorem 3.6. *Suppose 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 $V_k \xrightarrow{K} 0$, $f_{\hat{u}}^{\bar{k}} \leq f_*$ and $h_{\hat{u}}^{\bar{k}} \leq 0$. Moreover, the bounds of (3.1) hold.*

Proof. Steps 3, 4, 5 and 7 ensure that $\{t_k\}$ is monotone for large k. We have $V_k \xrightarrow{K} 0$ and $h_{\hat{u}}^{\bar{k}} \leq 0$ from either Lemma 3.4 if $t_\infty = \infty$, or Lemma 3.5 if $t_\infty < \infty$. Then $f_{\hat{u}}^{\bar{k}} \leq f_*$ by Lemma 3.1 (since $\tau_k = f_{\hat{u}}^k = f_{\hat{u}}^{\bar{k}}$ for $k \geq \bar{k}$). The final assertion stems from Lemma 3.2. □

Next, we analyze the case of infinitely many descent steps in phase 2.

Theorem 3.7. *Suppose infinitely many descent steps occur, and $h_{\hat{u}}^{\bar{k}} \le 0$ for some $\bar{k}$. Let $f_{\hat{u}}^{\infty} := \lim_k f_{\hat{u}}^k$ and $K := \{k \ge \bar{k} : f_{\hat{u}}^{k+1} < f_{\hat{u}}^k\}$. Then either $f_{\hat{u}}^{\infty} = f_* = -\infty$, or $-\infty < f_{\hat{u}}^{\infty} \le f_*$ and $\underline{\lim}_{k\in K} V_k = 0$. Moreover, the bounds of (3.1) hold. In particular, if $\{\hat{u}^k\}$ is bounded, then $f_{\hat{u}}^{\infty} > -\infty$ and $V_k \xrightarrow{K} 0$.*

Proof. For $k \ge \bar{k}$, we have $h_{\hat{u}}^k \le 0$, $\tau_k = f_{\hat{u}}^k$ (cf. (2.8)) and $f_{\hat{u}}^{k+1} \le f_{\hat{u}}^k$, since the descent test (2.30) becomes $\max\{f_u^{k+1} - f_{\hat{u}}^k, h_u^{k+1}\} \le -\kappa v_k$. First, suppose that $f_{\hat{u}}^{\infty} > -\infty$.

We have $0 < \kappa v_k \le f_{\hat{u}}^k - f_{\hat{u}}^{k+1}$ if $k \in K$, $f_{\hat{u}}^{k+1} = f_{\hat{u}}^k$ otherwise, so $\sum_{k\in K} \kappa v_k \le f_{\hat{u}}^{\bar{k}} - f_{\hat{u}}^{\infty} < \infty$ gives $v_k \xrightarrow{K} 0$ and hence $\epsilon_k, t_k|p^k|^2 \xrightarrow{K} 0$ by (2.25), as well as $|p^k| \xrightarrow{K} 0$, using $t_k \ge t_{\min}$. Now, for the descent iterations $k \in K$, we have $\hat{u}^{k+1} - \hat{u}^k = -t_k p^k$ by (2.18) and therefore

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

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

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

(since otherwise $|\hat{u}^k|^2 \to -\infty$, which is impossible). Combining this with $t_k|p^k|^2 \xrightarrow{K} 0$ gives $\underline{\lim}_{k\in K}\langle p^k, \hat{u}^k\rangle \le 0$. Since also $\epsilon_k, |p^k| \xrightarrow{K} 0$, we have $\underline{\lim}_{k\in K} V_k = 0$ by (2.19).

Then using $\underline{\lim}_{k\in K} V_k = 0$ and $\tau_k \to f_{\hat{u}}^{\infty}$ in Lemma 3.1 shows that $f_{\hat{u}}^{\infty} \le f_*$.

For the case of $f_{\hat{u}}^{\infty} = -\infty$ and the assertion on (3.1), invoke Lemma 3.2.

For the final assertion, if $\{\hat{u}^k\} \subset C$ is bounded, then $\inf_k f(\hat{u}^k) > -\infty$ (f is closed on C) implies that $f_{\hat{u}}^{\infty} > -\infty$ by (3.1), so we have $\epsilon_k, |p^k| \xrightarrow{K} 0$ as above. Combining this with the fact that $V_k \le \max\{|p^k|, \epsilon_k\}(1 + |\hat{u}^k|)$ by Lemma 2.5(iv) gives $V_k \xrightarrow{K} 0$. □

We now deal with the case of infinitely many descent steps at phase 1 for $\epsilon_{\max} > 0$.

Theorem 3.8. *Suppose infinitely many descent steps occur, $h_{\hat{u}}^k > 0$ for all k, and $\epsilon_{\max} > 0$. Let $K := \{k : h_{\hat{u}}^{k+1} < h_{\hat{u}}^k\}$. Then we have the following statements.*

(i) $h_{\hat{u}}^k \downarrow 0$.

(ii) $\underline{\lim}_{k\in K} V_k = 0$.

(iii) *Let $K' \subset \mathbb{N}$ be such that $V_k \xrightarrow{K'} 0$. Then $\overline{\lim}_{k\in K'} f_{\hat{u}}^k \le \overline{\lim}_{k\in K'} \tau_k \le f_*$.*

(iv) *If $\{\hat{u}^k\}$ is bounded, then $V_k \xrightarrow{K} 0$, and we may take $K' = K$ in* (iii) *above.*

(v) *The bounds of* (3.1) *hold, and $\underline{\lim}_k \tau_k \ge f_* - \epsilon_f - \bar{\mu}\epsilon_h$.*

Proof. (i) We have $0 < \kappa v_k \le h_{\hat{u}}^k - h_{\hat{u}}^{k+1}$ by (2.30) if $k \in K$, $h_{\hat{u}}^{k+1} = h_{\hat{u}}^k$ otherwise, so $\sum_{k\in K} \kappa v_k \le h_{\hat{u}}^1$ gives $v_k \xrightarrow{K} 0$; hence the fact that $v_k \ge \kappa_h h_{\hat{u}}^k$ (cf. Step 3) yields $h_{\hat{u}}^k \downarrow 0$.

(ii) Use $v_k \xrightarrow{K} 0$ as in the proof of Theorem 3.7 to get $\underline{\lim}_{k\in K} V_k = 0$ and $\epsilon_k, |p^k| \xrightarrow{K} 0$.

(iii) This follows from Lemma 3.1.

(iv) Invoke Lemma 2.5(iv) and the relations $\epsilon_k, |p^k| \xrightarrow{K} 0$ from the proof of item (ii).

(v) This follows from item (i), Lemma 3.2 and the fact that $\tau_k \ge f_{\hat{u}}^k$ for all k. □

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

Remarks 3.9. (i) Inspection of the preceding proofs reveals that Theorems 3.6–3.8 require only convexity and finiteness of f and h on C, and *local boundedness* of the approximate subgradient mappings g_f of f and g_h of h on C. In particular, it suffices to assume that f and h are finite convex on a neighborhood of C.

(ii) Using the *evaluation errors* $\epsilon_f^k := f(u^k) - f_u^k$ and $\epsilon_h^k := h(u^k) - h_u^k$, our results are sharpened as follows; cf. [Kiw06d, §4.2]. In general, $f(\hat{u}^k) = f_{\hat{u}}^k + \epsilon_f^{k(l)}$ and $h(\hat{u}^k) = h_{\hat{u}}^k + \epsilon_h^{k(l)}$, where $k(l) - 1$ denotes the iteration number of the lth descent step. Hence ϵ_f and ϵ_h in the bounds of (3.1) for Theorems 3.6–3.8 may be replaced by the *asymptotic errors* ϵ_f^∞ and ϵ_h^∞, where ϵ_f^∞ equals the final $\epsilon_f^{k(l)}$ if only finitely many descent steps occur, $\overline{\lim}_l \epsilon_f^{k(l)}$ otherwise, and ϵ_h^∞ is defined analogously.

(iii) Concerning Theorem 3.8(iv), note that the sequence $\{\hat{u}^k\}$ is bounded if the feasible set U is bounded. Indeed, $h(\hat{u}^k) \leq h_{\hat{u}}^k + \epsilon_h$ (cf. (2.7)) with $h_{\hat{u}}^k \leq h_{\hat{u}}^1$ imply that $\{\hat{u}^k\}$ lies in the set $\{u \in C : h(u) \leq h_{\hat{u}}^1 + \epsilon_h\}$, which is bounded, since such is U.

Finally, we analyze infinitely many descent steps in the exact case of $\epsilon_{\max} = 0$.

Theorem 3.10. *Suppose infinitely many descent steps occur and* $\epsilon_{\max} = 0$. *Let* $K := \{k(l) - 1\}_{l=1}^\infty$ *index the descent iterations (cf. Step* 5*), and let* $\bar{k} := \inf\{k : h(\hat{u}^k) \leq 0\}$ *(so that phase* 2 *starts at iteration* $k = \bar{k}$ *iff* $\bar{k} < \infty$*). Then we have the following statements.*

(i) *If* $\bar{k} < \infty$, *then* $f(\hat{u}^k) \to f_*$, $\tau_k \to f_*$, $h(\hat{u}^k)_+ \to 0$ *and each cluster point of the sequence* $\{\hat{u}^k\}$ *(if any) lies in the optimal set* U_*; *moreover,* $\underline{\lim}_{k\in K} V_k = 0$ *if* $f_* > -\infty$.

(ii) *If* $\inf_k f(\hat{u}^k) > -\infty$ *or* $\bar{k} = \infty$, *then* $\sum_{k\in K} v_k < \infty$, $\epsilon_k \xrightarrow{K} 0$ *and* $p^k \xrightarrow{K} 0$.

(iii) *If the sequence* $\{\hat{u}^k\}$ *is bounded, then all its cluster points lie in the optimal set* U_*, *and we have* $f(\hat{u}^k) \to f_* > -\infty$, $\tau_k \to f_*$, $h(\hat{u}^k)_+ \to 0$ *and* $V_k \xrightarrow{K} 0$.

(iv) *If the sequence* $\{\hat{u}^k\}$ *has a cluster point* $\bar{u}$, *then* $\bar{u} \in U_*$, $h(\hat{u}^k)_+ \to 0$ *and* $\underline{\lim}_k \tau_k \geq \underline{\lim}_k f(\hat{u}^k) \geq f_* > -\infty$; *moreover, if* $K' \subset K$ *is such that* $\hat{u}^k \xrightarrow{K'} \bar{u}$, *then* $V_k \xrightarrow{K'} 0$.

(v) *The sequence* $\{\hat{u}^k\}$ *has a cluster point if the set* U_* *is nonempty and bounded.*

(vi) *The sequence* $\{\hat{u}^k\}$ *is bounded if such is the feasible set* $U := \{u \in C : h(u) \leq 0\}$.

(vii) *Suppose that* $\bar{u} \in U_*$ *and there exists an iteration index* k' *such that*

$$f(\bar{u}) \leq \pi(\hat{u}^k; c_k + 1) \quad \text{for all } k \geq k', k \in K. \tag{3.13}$$

In particular, (3.13) *holds if* $\hat{u}^{k'} \in U$ *for some* k', *or* $c_k \geq \bar{\mu} - 1$ *for all* $k \geq k', k \in K$. *Further, suppose* $\overline{\lim}_{k\in K} t_k < \infty$. *Then the sequence* $\{\hat{u}^k\}$ *converges to a point in* U_*.

(viii) *Suppose* $\{\hat{u}^k\}$ *is bounded, but we only have* $\sum_{k\in K} t_k = \infty$ *instead of* $\inf_{k\in K} t_k \geq t_{\min}$. *Then* $\{\hat{u}^k\}$ *has a cluster point in* U_*. *Moreover, assertion* (vii) *still holds.*

Proof. First, recalling the basic "exact" relations (2.32)–(2.33), note that $\epsilon_k \geq 0$ and

$$e_C(\cdot; \tau_k) \geq e_C(\hat{u}^k; \tau_k) + \langle p^k, \cdot - \hat{u}^k \rangle - \epsilon_k \quad \text{with} \quad e_C(\hat{u}^k; \tau_k) = h(\hat{u}^k)_+. \tag{3.14}$$

By Remark 2.7(vi), the descent test (2.30) ensures that $0 < h(\hat{u}^{k+1}) \leq h(\hat{u}^k)$ for all k if $\bar{k} = \infty$, $f_* \leq f(\hat{u}^{k+1}) \leq f(\hat{u}^k)$ and $h(\hat{u}^k) \leq 0$ for all $k \geq \bar{k}$ otherwise.

(i) Use $f_{\hat{u}}^\infty = \lim_k f(\hat{u}^k) = \lim_k \tau_k$ in Theorem 3.7 and the closedness of C, f and h.

(ii) Use the proof of Theorem 3.7 if $\bar{k} < \infty$, or of Theorem 3.8(i,ii) otherwise.

(iii) First, suppose that $\bar{k} = \infty$, i.e., consider phase 1 with $h(\hat{u}^k) > 0$ for all k.

Let $\bar{u}$ be a cluster point of $\{\hat{u}^k\}$. Then $\bar{u} \in C$, since $\{\hat{u}^k\} \subset C$ and C is closed. Pick $K' \subset K$ such that $\hat{u}^k \xrightarrow{K'} \bar{u}$. Then $f(\hat{u}^k) \xrightarrow{K'} f(\bar{u})$, $h(\hat{u}^k) \xrightarrow{K'} h(\bar{u}) \geq 0$ (f, h are continuous on C). Since $\epsilon_k, |p^k| \xrightarrow{K} 0$ by (ii), Lemma 2.5(iv) yields $V_k \xrightarrow{K'} 0$. Let $\bar{\tau}$ be any cluster point of $\{\tau_k\}_{k\in K'}$. Pick $K'' \subset K'$ such that $\tau_k \xrightarrow{K''} \bar{\tau}$. We have $\bar{\tau} \geq f(\bar{u})$ ($\tau_k \geq f(\hat{u}^k)$) and $\bar{\tau} < \infty$, since otherwise for large $k \in K''$, $\tau_k \geq f(\mathring{u}) - h(\mathring{u})$ would give $e(\mathring{u}; \tau_k) = h(\mathring{u}) < 0$ by (2.2) and (1.2), and (3.14) with $\epsilon_k, |p^k| \xrightarrow{K} 0$ would yield

$$0 > h(\mathring{u}) = e_C(\mathring{u}; \tau_k) \geq h(\hat{u}^k)_+ + \langle p^k, \mathring{u} - \hat{u}^k \rangle - \epsilon_k \xrightarrow{K''} h(\bar{u})_+ \geq 0,$$

a contradiction. Since e_C is continuous on $C \times \mathbb{R}$, letting $k \xrightarrow{K''} \infty$ in (3.14) gives $e_C(\cdot; \bar{\tau}) \geq e_C(\bar{u}; \bar{\tau})$, i.e., $0 \in \partial e_C(\bar{u}; \bar{\tau})$. Since $h(\bar{u}) \geq 0$ and $\bar{\tau} \geq f(\bar{u})$, $0 \in \partial e_C(\bar{u}; \bar{\tau})$ in (2.3) implies $\bar{\tau} = f(\bar{u})$ and $h(\bar{u}) = 0$ (otherwise for $h_C := h + i_C$, $0 \in \partial h_C(\bar{u})$ would give $\min_C h \geq 0$, contradicting (1.2)). Hence, $\bar{u} \in U_*$ by Lemma 2.2 (using $\bar{\tau} = \pi(\bar{u}; \bar{c})$ for any $\bar{c} \geq 0$) and $f(\bar{u}) = f_*$. Since $h(\bar{u}) = 0$ and $\{h(\hat{u}^k)\}$ is nonincreasing, we obtain that $h(\hat{u}^k) \to 0$.

By considering any convergent subsequences, we deduce that $V_k \xrightarrow{K} 0$, and that f_* is the unique cluster point of $\{\tau_k\}_{k\in K}$ and $\{f(\hat{u}^k)\}_{k\in K}$. Hence, $\lim_l \tau_{k(l)-1} = \lim_l f(\hat{u}^{k(l)-1}) = f_*$. Since $f(\hat{u}^{k(l)}) \leq \tau_k \leq \tau_{k(l+1)-1}$ for $k(l) \leq k < k(l+1)$ by Steps 3, 4 and 7, we obtain $\lim_k f(\hat{u}^k) = \lim_k \tau_k = f_*$. Finally, for the remaining case of $\bar{k} < \infty$, use the monotonicity of $\{\tau_k = f(\hat{u}^k)\}_{k \geq \bar{k}}$ and the relations $\bar{\tau} = f(\bar{u})$, $h(\bar{u}) \leq 0$ in the preceding arguments.

(iv) Use the proof of (iii), getting $\underline{\lim}_k f(\hat{u}^k) \geq f_*$ from $h(\hat{u}^k)_+ \to 0$ as in Lemma 3.2.

(v) If $\bar{k} < \infty$, the set $\{u \in C : f(u) \leq f(\hat{u}^{\bar{k}}), h(u) \leq 0\}$ is bounded (such is U_*) and contains $\{\hat{u}^k\}_{k \geq \bar{k}}$. Suppose $\bar{k} = \infty$. By the proof of Thm. 3.8(ii,iii), there is $K' \subset K$ such that $\overline{\lim}_{k\in K'} f(\hat{u}^k) \leq f_*$. Hence, for infinitely many k, $\hat{u}^k$ lies in the set $\{u \in C : f(u) \leq f_* + 1, h(u) \leq h(u^1)_+\}$, which is bounded (such is U_*). Therefore, $\{\hat{u}^k\}$ has a cluster point.

(vi) The set $\{u \in C : h(u) \leq h(u^1)_+\}$ is bounded (such is U) and contains $\{\hat{u}^k\}$.

(vii) If $\bar{k} < \infty$, then for $k \geq \bar{k}$, $\hat{u}^k \in U$ implies $f(\bar{u}) = f_* \leq f(\hat{u}^k) = \pi(\hat{u}^k; c_k + 1)$; together with Lemma 2.3, this validates our claim below (3.13). Let $k \in K$, $k \geq k'$. Since (3.13) implies $e_C(\bar{u}; \tau_k) \leq e_C(\hat{u}^k; \tau_k)$ by Lemma 2.3, (3.14) yields $\langle p^k, \bar{u} - \hat{u}^k \rangle \leq \epsilon_k$. Then, using the facts that $\hat{u}^{k+1} - \hat{u}^k = -t_k p^k$ by (2.18) and $v_k = t_k |p^k|^2 + \epsilon_k$ by (2.23), we get

$$\begin{aligned} |\hat{u}^{k+1} - \bar{u}|^2 &= |\hat{u}^k - \bar{u}|^2 + 2\langle \hat{u}^{k+1} - \hat{u}^k, \hat{u}^k - \bar{u} \rangle + |\hat{u}^{k+1} - \hat{u}^k|^2 \\ &\leq |\hat{u}^k - \bar{u}|^2 + 2t_k\epsilon_k + 2t_k^2 |p^k|^2 = |\hat{u}^k - \bar{u}|^2 + 2t_k v_k. \end{aligned}$$

Therefore, since $\overline{\lim}_{k\in K} t_k < \infty$, $\sum_{k\in K} v_k < \infty$ by (ii), and $|\hat{u}^{k+1} - \bar{u}|^2 = |\hat{u}^k - \bar{u}|^2$ if $k \notin K$, we deduce from [Pol83, Lem. 2.2.2] that the sequence $\{|\hat{u}^k - \bar{u}|\}$ converges. Thus the sequence $\{\hat{u}^k\}$ is bounded, and using (iii) we may choose $\bar{u} \in U_*$ as a cluster point of $\{\hat{u}^k\}$, in which case the sequence $\{|\hat{u}^k - \bar{u}|\}$ must converge to zero, i.e., $\hat{u}^k \to \bar{u}$.

(viii) Argue as for (ii) to get $\sum_{k\in K} v_k < \infty$. Since $v_k = t_k|p^k|^2 + \epsilon_k$ (cf. (2.23)) and $\epsilon_k \geq 0$, we have $\underline{\lim}_{k\in K} |p^k|^2 = 0$ (using $\sum_{k\in K} t_k = \infty$) and $\lim_{k\in K} \epsilon_k = 0$. Thus, there is $\bar{K} \subset K$ such that $\epsilon_k, |p^k| \xrightarrow{\bar{K}} 0$. Let $\bar{u}$ be a cluster point of $\{\hat{u}^k\}_{k\in\bar{K}}$. To see that $\bar{u} \in U_*$, replace K by $\bar{K}$ in the proof of (iii). Hence, this point $\bar{u}$ may be used in the final part of the proof of (vii). $\square$

Remarks 3.11. (i) The condition $\epsilon_{\max} = 0$ in Theorem 3.10 means that the linearizations are exact *and* Step 3 is inactive. If we drop this condition in Step 3, so that Step 3 ensures $v_k \geq \kappa_h h^k_{\hat{u}}$ when $h^k_{\hat{u}} > 0$ in the exact case as well, then for $\epsilon_{\max} = 0$, both Theorem 3.10 and Theorem 3.8 hold with $\epsilon_f = \epsilon_h = 0$ in the bounds of (3.1).

(ii) Condition (3.13) was used in [SaS05, Prop. 4.3(ii)] with $c_k \equiv 0$. Since in this case, $f_* = \inf_C \pi(\cdot, c_k + 1)$ iff $\bar{\mu} \leq 1$ (cf. §2.1), we conclude that at phase 1 ($\bar{k} = \infty$) condition (3.13) with $c_k \equiv 0$ may be expected to hold only if $\bar{\mu} \leq 1$. (Also see §4.4.)

4 Modifications

4.1 Alternative descent tests

As in [Kiw06c, §4.3], at Steps 4 and 5 we may replace the predicted decrease $v_k = t_k|p^k|^2 + \epsilon_k$ (cf. (2.23)) by the smaller quantity $w_k := t_k|p^k|^2/2 + \epsilon_k$. Then Lemma 2.5(ii) is replaced by the fact that

$$w_k \geq -\epsilon_k \quad \Longleftrightarrow \quad t_k|p^k|^2/4 \geq -\epsilon_k \quad \Longleftrightarrow \quad w_k \geq t_k|p^k|^2/4.$$

Hence, $w_k \geq -\epsilon_k$ at Step 5 implies $w_k \leq v_k \leq 3w_k$ and $v_k \geq -\epsilon_k$ for the bounds (2.25)–(2.26), whereas for Step 4, the bound (2.27) is replaced by the fact that

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

The preceding results extend easily (in the proof of Lemma 3.5, $e_{k+1}(u^{k+1}) > [h^k_{\hat{u}}]_+ - \kappa w_k$ implies $e_{k+1}(u^{k+1}) > [h^k_{\hat{u}}]_+ - \kappa v_k$, whereas in the proofs of Theorems 3.7 and 3.8(i), we have $\sum_{k \in K} v_k \leq 3 \sum_{k \in K} w_k < \infty$). We add that [SaS05, Alg. 3.1] uses w_k instead of v_k.

As in [Kiw85, p. 227], we may replace the descent test (2.30) by the two-part test

$$h^{k+1}_u \leq h^k_{\hat{u}} - \kappa v_k \qquad\qquad \text{if } h^k_{\hat{u}} > 0, \tag{4.1a}$$

$$f^{k+1}_u \leq f^k_{\hat{u}} - \kappa v_k \quad \text{and} \quad h^{k+1}_u \leq 0 \quad \text{if } h^k_{\hat{u}} \leq 0. \tag{4.1b}$$

Since (2.30) implies (4.1), the latter test may produce faster convergence. In particular, at phase 2 ($h^k_{\hat{u}} \leq 0$) the additional requirement $h^{k+1}_u \leq -\kappa v_k$ of (2.30) may hinder progress of $\{\hat{u}^k\}$ towards the boundary of the feasible set. The preceding convergence results are not affected (since if (4.1) fails at a null step, then so does (2.30), whereas the requirements of (4.1) suffice for descent steps).

In connection with (4.1b), we add that if $h^1_{\hat{u}} \leq 0$, i.e., the starting point is approximately feasible, then the objective linearizations needn't be defined at infeasible points. Specifically, if $h^{k+1}_u > 0$ in (4.1b), then a null step must occur, so we may skip evaluating f^{k+1}_u and choose $J^{k+1}_f \supset \hat{J}^k_f$ at Step 6 (without requiring $J^{k+1}_f \ni k+1$). In the proof of Lemma 3.5, using $v_k = -\check{e}_k(u^{k+1})$ (cf. (2.10)) and replacing (3.10) by

$$e_{k+1}(\cdot) := \begin{cases} f_{k+1}(\cdot) - f^k_{\hat{u}} & \text{if } h^{k+1}_u \leq 0, \\ h_{k+1}(\cdot) & \text{otherwise,} \end{cases} \tag{4.2}$$

we see that (4.1b) can be expressed as $e_{k+1}(u^{k+1}) \leq -\kappa v_k$ or equivalently by (3.11); this suffices for the proof. Similarly, if $h^{k+1}_u \leq 0$, then we may skip finding the subgradient g^{k+1}_h, and choose $J^{k+1}_h \supset \hat{J}^k_h$ at Step 6 (omitting $\check{h}_k(\cdot) := -\infty$ in (2.8) if $J^k_h = \emptyset$).

4.2 Linearization aggregation

To trade off storage and work per iteration for speed of convergence, one may replace selection with aggregation, so that only $\bar{m} \geq 4$ subgradients are stored. To this end, we note that the preceding results remain valid if, for each k, $\check{f}_{k+1}$ and $\check{h}_{k+1}$ are closed convex functions such that $0 \in \partial\phi_k(u^{k+1})$ implies (2.11)–(2.13) for k increased by 1, and

$$\max\{\bar{f}_k(u), f_{k+1}(u)\} \leq \check{f}_{k+1}(u) \leq f(u) \quad \forall u \in C, \tag{4.3a}$$

$$\max\{\bar{h}_k(u), h_{k+1}(u)\} \leq \check{h}_{k+1}(u) \leq h(u) \quad \forall u \in C. \tag{4.3b}$$

The max terms above are needed only after null steps in the proof of Lemma 3.5, $\bar{f}_k$ is not needed if $\nu_k = 0$, and $\bar{h}_k$ is not needed if $\nu_k = 1$. The aggregate linearizations may be treated like the oracle linearizations. Indeed, letting $f_{-j} := \bar{f}_j$, $h_{-j} := \bar{h}_j$ for $j = 1\colon k$, to ensure that $\bar{f}_k \leq \check{f}_{k+1}$ and $\bar{h}_k \leq \check{h}_{k+1}$, we may work with $J_f^{k+1}, J_h^{k+1} \subset \{-k\colon k+1\}$ in (2.31), replacing the set $\hat{J}_f^k$ or $\hat{J}_h^k$ by $\{-k\}$ when $\hat{J}_f^k$ or $\hat{J}_h^k$ is "too large".

To illustrate, consider the following scheme with *minimal aggregation*. First, suppose $|J_f^k| + |J_h^k| = \bar{m}$. If $|\hat{J}_f^k| + |\hat{J}_h^k| \leq \bar{m} - 2$, remove from J_f^k or J_h^k two indices in $J_f^k \setminus \hat{J}_f^k$ or $J_h^k \setminus \hat{J}_h^k$. If $|\hat{J}_f^k| + |\hat{J}_h^k| = \bar{m} - 1$, set $J_f^k := \hat{J}_f^k$, $J_h^k := \hat{J}_h^k$; if $|\hat{J}_h^k| \geq 2$, remove two indices from $\hat{J}_h^k$ and set $J_h^k := \hat{J}_h^k \cup \{-k\}$, otherwise remove two indices from $\hat{J}_f^k$ and set $J_f^k := \hat{J}_f^k \cup \{-k\}$. If $|\hat{J}_f^k| + |\hat{J}_h^k| = \bar{m}$, remove four indices from $\hat{J}_f^k$ or $\hat{J}_h^k$, and set $J_f^k := \hat{J}_f^k \cup \{-k\}$, $J_h^k := \hat{J}_h^k \cup \{-k\}$. Next, suppose $|J_f^k| + |J_h^k| = \bar{m} - 1$. If $|\hat{J}_f^k| + |\hat{J}_h^k| = \bar{m} - 1$, proceed as in the second case above. If $|\hat{J}_f^k| + |\hat{J}_h^k| \leq \bar{m} - 2$, remove from J_f^k or J_h^k one index in $J_f^k \setminus \hat{J}_f^k$ or $J_h^k \setminus \hat{J}_h^k$. At this stage, $|J_f^k| + |J_h^k| \leq \bar{m} - 2$, so set $J_f^{k+1} := J_f^k \cup \{k+1\}$, $J_h^{k+1} := J_h^k \cup \{k+1\}$. This scheme employs aggregation only where needed; for $\bar{m} \geq m+3$, it reduces to selection (cf. Rem. 2.7(vii)).

In practice, without storing the points u^j for $j \geq 1$, we may use the representations

$$f_j(\cdot) = f_j(\hat{u}^k) + \langle \nabla f_j, \cdot - \hat{u}^k \rangle \quad \text{and} \quad h_j(\cdot) = h_j(\hat{u}^k) + \langle \nabla h_j, \cdot - \hat{u}^k \rangle,$$

since after a descent step, we can update the linearization values

$$f_j(\hat{u}^{k+1}) = f_j(\hat{u}^k) + \langle \nabla f_j, \hat{u}^{k+1} - \hat{u}^k \rangle \quad \text{for } j \in J_f^{k+1}, \tag{4.4a}$$

$$h_j(\hat{u}^{k+1}) = h_j(\hat{u}^k) + \langle \nabla h_j, \hat{u}^{k+1} - \hat{u}^k \rangle \quad \text{for } j \in J_h^{k+1}. \tag{4.4b}$$

Let us now consider a variant with *total aggregation*, in which only two linearizations need be stored. Let $J_e^1 := \{1\}$, define e_1 by (3.10) with $k = 0$ and $\tau_0 := \tau_1$, and replace $\check{e}_k$ in (2.8) by the "overall" model

$$\check{e}_k(\cdot) := \max_{j \in J_e^k} e_j(\cdot) \tag{4.5}$$

of $e(\cdot; \tau_k)$; thus we no longer maintain separate models of f and h. Then the optimality condition $0 \in \partial\phi_k(u^{k+1})$ yields the existence of a subgradient $p_e^k \in \partial\check{e}_k(u^{k+1})$ such that p_e^k replaces $\nu_k p_f^k + (1 - \nu^k) p_h^k$ in (2.12) and (2.18), and using the aggregate linearization

$$\bar{e}_k(\cdot) := \check{e}_k(u^{k+1}) + \langle p_e^k, \cdot - u^{k+1} \rangle \leq \check{e}_k(\cdot) \leq e(\cdot; \tau_k), \tag{4.6}$$

we may replace the definition (2.17) of the linearization $\bar e_C^k$ and its expression (2.20) by

$$\bar e_C^k(\cdot) := \bar e_k(\cdot) + \bar\imath_C^k(\cdot) = \check e_k(u^{k+1}) + \langle p^k, \cdot - u^{k+1}\rangle. \tag{4.7}$$

For linearization selection, we may use multipliers γ_j^k of the pieces e_j, $j \in J_e^k$, such that

$$(p_e^k, 1) = \sum_{j\in J_e^k} \gamma_j^k(\nabla e_j, 1),\ \gamma_j^k \ge 0,\ \gamma_j^k[\check e_k(u^{k+1}) - e_j(u^{k+1})] = 0,\ j \in J_e^k, \tag{4.8}$$

to choose the set $J_e^{k+1} \supset \hat J_e^k \cup \{k+1\}$ with $\hat J_e^k := \{j \in J_e^k : \gamma_j^k \ne 0\}$ and e_{k+1} given by (3.10). For aggregation (cf. (4.3)), after a null step the next model $\check e_{k+1}$ should satisfy

$$\max\{\bar e_k(u), e_{k+1}(u)\} \le \check e_{k+1}(u) \le e(u;\tau_k) \quad \forall u \in C, \tag{4.9}$$

and it suffices to choose $J_e^{k+1} \supset \{-k, k+1\}$ with $e_{-k} := \bar e_k$. Note that (4.6) and the minorization $e_{k+1}(\cdot) \le e(\cdot;\tau_k)$ (cf. (3.10)) yield $\check e_{k+1}(\cdot) \le e(\cdot;\tau_k)$. To ensure that $e(\cdot;\tau_k)$ is still minorized by each $e_j(\cdot) = e_j(\hat u^k) + \langle \nabla e_j, \cdot - \hat u^k\rangle$ after a descent step, we may update

$$e_j(\hat u^{k+1}) := e_j(\hat u^k) + \langle \nabla e_j, \hat u^{k+1} - \hat u^k\rangle - (\tau_{k+1} - \tau_k)_+, \tag{4.10}$$

since $e(\cdot;\tau_{k+1}) \ge e(\cdot;\tau_k) - (\tau_{k+1} - \tau_k)_+$ (cf. (2.2)). Similarly, when τ_k increases to τ_k' say, at Steps 3 or 4, the update $e_j(\hat u^k) := e_j(\hat u^k) - \tau_k' + \tau_k$ provides the minorization $e_j(\cdot) \le e(\cdot;\tau_k')$.

Although total aggregation needs only $\bar m \ge 2$ linearizations, whereas separate aggregation described below (4.3) needs $\bar m \ge 4$, in practice this difference is immaterial, since larger values of $\bar m$ are required for faster convergence anyway. On the other hand, total aggregation has a serious drawback: its update (4.10), being based on a crude pessimistic estimate, tends to make the linearizations e_j lower than necessary when $\tau_{k+1} \ne \tau_k$. In contrast, separate aggregation is not sensitive to changes of τ_k, since it employs the natural updates of (4.4) and accounts for the current τ_k explicitly in its model $\check e_k$ of (2.8). In other words, it pays to maintain separate models of f and h instead of ignoring the structure of $e(\cdot, \tau_k)$ in the overall model (4.5); thus, total aggregation is of theoretical interest only.

Similar techniques can be applied to the *composite model*

$$\check e_k(\cdot) := \max\{\max_{j\in J_f^k} f_j(\cdot) - \tau_k, \max_{j\in J_h^k} h_j(\cdot), \max_{j\in J_e^k} e_j(\cdot)\}. \tag{4.11}$$

For instance, (4.9) holds if $J_f^{k+1} \ni k+1$, $J_h^{k+1} \ni k+1$, $J_e^{k+1} \ni -k$, but many other choices are possible. We skip the details, because in practice separate selection or aggregation of the linearizations of f and h is more efficient, due to avoiding the update of (4.10).

Remark 4.1. We add that [SaS05, Alg. 3.1] employs the composite model (4.11) with

$$J_f^k := \{\, j \in J^k : f_u^j - \tau_k \ge h_u^j \,\} \quad \text{and} \quad J_h^k := \{\, j \in J^k : f_u^j - \tau_k < h_u^j \,\} \tag{4.12}$$

for an additional "oracle" set $J^k \subset \{1\colon k\}$; then J^k and J_e^k are reduced if necessary so that $2|J^k| + |J_e^k| \le \bar m - 3$ for a given $\bar m \ge 3$, and $J^{k+1} := J^k \cup \{k+1\}$, $J_e^{k+1} := J_e^k \cup \{-k\}$. First, this scheme is quite unusual: although $|J^k|$ "original" linearizations of f and h are maintained ($2|J^k|$ in total), only half of them are selected via (4.12) for the model (4.11) (this selection is unnecessary in the sense that even for $J_f^k = J_h^k = J^k$, the model (4.11) still satisfies $\check e_k(\cdot) \le e(\cdot, \tau_k)$). Second, its storage requirement of $\bar m \ge 3$ places it between total aggregation and separate aggregation. Third, and most importantly, this scheme employs the crude update of (4.10), and hence is less efficient than separate aggregation.

4.3 Estimating Lagrange multipliers

Suppose $f_* > -\infty$, so that the dual optimal set $M := \operatorname{Arg\,max}_{\mathbb{R}_+} q$ is nonempty (cf. §2.1). For $\bar\epsilon \ge 0$, the set of *$\bar\epsilon$-optimal dual solutions* is defined by

$$M_{\bar\epsilon} := \{\, \mu \in \mathbb{R}_+ : q(\mu) \ge f_* - \bar\epsilon \,\}. \tag{4.13}$$

We now develop conditions under which the *Lagrange multiplier estimates*

$$\mu_k := (1 - \nu_k)/\nu_k \tag{4.14}$$

converge to the set $M_{\bar\epsilon}$ for a suitable $\bar\epsilon \ge 0$, where ν_k is the multiplier of (2.12)–(2.13).

Since $\nu_k \in [0,1]$ by (2.13), (2.14)–(2.19) yield the sharper version of (2.22)

$$\nu_k[f(u) - \tau_k] + (1 - \nu_k)h(u) \ge [h_{\hat u}^k]_+ - V_k(1 + |u|) \quad \text{for all } u \in C. \tag{4.15}$$

If $\nu_k > 0$ (e.g., $V_k < -h(\mathring{u})/(1 + |\mathring{u}|)$), then (4.14) with $\mu_k \in \mathbb{R}_+$ and (4.15) give

$$f(u) + \mu_k h(u) \ge \tau_k - V_k(1 + |u|)/\nu_k \quad \text{for all } u \in C. \tag{4.16}$$

Lemma 4.2. (i) *Suppose $f_* > -\infty$. Let $K' \subset \mathbb{N}$ be such that $V_k \xrightarrow{K'} 0$ and*

$$\varliminf_{k\in K'} \tau_k \ge f_* - \epsilon_f - \bar\mu\epsilon_h, \tag{4.17}$$

where $\bar\mu := \inf_{\mu\in M} \mu$ (cf. §2.1). Then $\varlimsup_{k\in K'} \mu_k < \infty$ and $V_k/\nu_k \xrightarrow{K'} 0$. Moreover, the sequence $\{\mu_k\}_{k\in K'}$ converges to the set $M_{\bar\epsilon}$ given by (4.13) for $\bar\epsilon := \epsilon_f + \bar\mu\epsilon_h$.

(ii) *If $f_* > -\infty$, then a set K' satisfying the requirements of* (i) *exists under the assumptions of Theorems* 3.6, 3.7 *or* 3.8, *or those of Theorem* 3.10 *if additionally either* $\inf\{k : h(\hat u^k) \le 0\} < \infty$ *or* $|\hat u^k| \not\to \infty$ *(e.g., the optimal set U_* is nonempty and bounded).*

Proof. (i) By (4.17), $\tau_\infty := \varliminf_{k\in K'} \tau_k \ge f_* - \bar\epsilon$. If we had $\varliminf_{k\in K'} \nu_k = 0$, for $u = \mathring{u}$, (4.15) would yield in the limit $0 > h(\mathring{u}) \ge 0$, a contradiction. Hence, $\varliminf_{k\in K'} \nu_k > 0$, so that $V_k/\nu_k \xrightarrow{K'} 0$ and $\varlimsup_{k\in K'} \mu_k < \infty$ by (4.14). Let μ_∞ be any cluster point of $\{\mu_k\}_{k\in K'}$; then $\mu_\infty \in \mathbb{R}_+$. Passing to the limit in (4.16) bounds the Lagrangian values as follows

$$L(u; \mu_\infty) := f(u) + \mu_\infty h(u) \ge \tau_\infty \quad \text{for all } u \in C.$$

Hence, $q(\mu_\infty) \ge \tau_\infty \ge f_* - \bar\epsilon$ implies $\mu_\infty \in M_{\bar\epsilon}$ by (4.13). Since μ_∞ was an arbitrary cluster point of $\{\mu_k\}_{k\in K'} \subset \mathbb{R}_+ \cup \{\infty\}$ and $\varlimsup_{k\in K'} \mu_k < \infty$, the conclusion follows.

(ii) In Theorem 3.6, $\tau_k = f_{\hat u}^k$ for all $k \ge \bar k$ (and we may take $K' = K$). In Theorem 3.7, $\tau_k \to f_{\hat u}^\infty \in [f_* - \epsilon_f - \bar\mu\epsilon_h, f_*]$ and $\varliminf_{k\in K} V_k = 0$. For the rest, see Theorem 3.8(ii,v) and Theorem 3.10(i,iv,v), noting that $|\hat u^k| \not\to \infty$ iff $\{\hat u^k\}$ has a cluster point. □

4.4 Updating the penalty coefficient in the exact case

We first show how to choose the penalty coefficient c_k by using the Lagrange multiplier estimate μ_k of (4.14) to ensure the "convergence" condition (3.13) of Theorem 3.10(vii).

Lemma 4.3. *Under the assumptions of Theorem* 3.10, *suppose* $|\hat{u}^k| \not\to \infty$. *Suppose for all large* k, *after a descent step, Step 7 chooses* $c_{k+1} \geq \max\{\mu_k, c_k\}$ *if* $\mu_k < \infty$, $c_{k+1} \geq c_k$ *otherwise. Then there exists* k' *such that condition* (3.13) *holds for any* $\bar{u} \in U_*$.

Proof. By Theorem 3.10(iv), the assumptions of Lemma 4.2(i) hold for some $K' \subset K$, $\epsilon_f = \epsilon_h = \bar{\epsilon} = 0$; thus, $\{\mu_k\}_{k\in K'}$ converges to $M_0 = M$, and $\underline{\lim}_{k\in K'} \mu_k \geq \bar{\mu} := \inf_{\mu\in M} \mu$ implies $\mu_k \geq \bar{\mu} - 1$ for all large $k \in K'$. Hence, since $\{c_k\}$ is nondecreasing for large k, we have $c_k \geq \bar{\mu} - 1$ for all large k, and the conclusion follows from Theorem 3.10(vii). □

Remark 4.4. Variations on the strategy of Lemma 4.3 are possible. For instance, if $\{\hat{u}^k\}$ is bounded (e.g., U is bounded), Step 7 may choose $c_{k+1} \geq \mu_k$ after each descent step when $\mu_k < \infty$; this suffices for the proof of Lemma 4.3 with $K' = K$ by Theorem 3.10(iii).

We shall exploit the following elementary property of the exact penalty function (2.1).

Lemma 4.5. *If* $c \geq \bar{\mu}$, *then* $\pi(u; c) \geq f_* + (c - \bar{\mu})h(u)_+$ *for all* $u \in C$.

Proof. By (2.1), $\pi(u;c) = L(u;\bar{\mu}) + (c-\bar{\mu})h(u)_+ + \bar{\mu}[h(u)_+ - h(u)]$ for each $u \in C$, where $L(u;\bar{\mu}) \geq q(\bar{\mu}) = f_*$ (cf. §2.1), $\bar{\mu} \geq 0$ and $h(u)_+ \geq h(u)$. □

For phase 1 in the exact case (when Step 3 is inactive), the main difficulty lies in ensuring $h(\hat{u}^k) \downarrow 0$. Complementing Theorem 3.10, we now show that it suffices if the penalty parameter c_k majorizes strictly the minimal Lagrange multiplier $\bar{\mu}$ asymptotically, and we give a specific update of c_k, based on a simple idea: increase the penalty coefficient if the constraint violation is large relative to the optimality measure (cf. [Kiw91]).

Lemma 4.6. *Under the assumptions of Thm.* 3.10, *suppose* $h(\hat{u}^k) > 0$ *for all* k. *Then*:

(i) *There exists* $K' \subset K$ *such that* $V_k \xrightarrow{K'} 0$ *and* $\overline{\lim}_{k\in K'} f(\hat{u}^k) \leq \overline{\lim}_{k\in K'} \tau_k \leq f_*$.

(ii) *If* $c_\infty := \underline{\lim}_k c_k > \bar{\mu}$, *then* $h(\hat{u}^k) \downarrow 0$.

(iii) *Suppose for all large* k, *after a descent step, Step 7 chooses* $c_{k+1} \geq 2c_k$ *if* $h(\hat{u}^{k+1}) > V_k$, $c_{k+1} \geq c_k$ *otherwise,* $c_{k+1} > 0$ *when* $h(\hat{u}^{k+1}) > 0$. *If* $f_* > -\infty$, *then* $h(\hat{u}^k) \downarrow 0$.

(iv) *If* $h(\hat{u}^k) \downarrow 0$, *then* $\underline{\lim}_k \tau_k \geq \underline{\lim}_k f(\hat{u}^k) \geq f_*$, *and* $\lim_{k\in K'} f(\hat{u}^k) = f_*$ *in* (i) *above.*

Proof. (i) This follows from the proof of Theorem 3.8(ii,iii), using $\tau_k = \pi(\hat{u}^k; c_k)$.

(ii) By (i) and Lemma 4.5, $f_* \geq \underline{\lim}_k \tau_k \geq f_* + (c_\infty - \bar{\mu})\, \underline{\lim}_k h(\hat{u}^k)_+$ with $c_\infty > \bar{\mu}$ yield $\underline{\lim}_k h(\hat{u}^k)_+ = 0$. Hence, $h(\hat{u}^k) \downarrow 0$, using $0 < h(\hat{u}^{k+1}) \leq h(\hat{u}^k)$ for all k.

(iii) If $c_\infty := \lim_k c_k < \infty$, then $h(\hat{u}^{k+1}) \leq V_k$ for all large $k \in K$, so by (i), $V_k \xrightarrow{K'} 0$ yields $h(\hat{u}^k) \downarrow 0$. Otherwise, $c_\infty = \infty > \bar{\mu}$ (from $f_* > -\infty$) and assertion (ii) applies.

(iv) Invoke Lemma 3.2 with $\epsilon_f = \epsilon_h = 0$, and use the fact that $\tau_k \geq f(\hat{u}^k)$. □

5 Column generation for LP programs

In this section we consider the following primal-dual pair of LP problems

$$\min\ c\lambda \quad \text{s.t.} \quad A\lambda \geq b,\ \lambda \geq 0, \tag{5.1}$$

$$\max\ ub \quad \text{s.t.} \quad uA \leq c,\ u \geq 0, \tag{5.2}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. We assume that $c > 0$. Let A_i denote column i of A for $i \in I := \{1{:}n\}$. When the number of columns is huge, problems (5.1)–(5.2) may be solved by column generation, provided that for each $u \geq 0$, one can solve the *column generation subproblem* of finding $i_u \in \operatorname{Arg\,max}_{i \in I}(uA_i - c_i)$. We show that this subproblem may be solved inexactly when our method is applied to the dual problem (5.2) formulated as (1.1), and that approximate solutions to (5.1) can be recovered at no extra cost.

To ease subsequent notation, let us rewrite the LP programs (5.1)–(5.2) as follows

$$\max\ \psi_0(\lambda) := -c\lambda \quad \text{s.t.} \quad \psi(\lambda) := A\lambda - b \geq 0,\ \lambda \in \mathbb{R}^n_+, \tag{5.3}$$

$$\min\ f(u) := -ub \quad \text{s.t.} \quad uA \leq c,\ u \in \mathbb{R}^m_+. \tag{5.4}$$

The dual problem (5.4) is formulated as (1.1) with $C := \mathbb{R}^m_+$ and the constraint function

$$h(\cdot) := \max_{i \in I}(\langle A_i, \cdot\rangle - c_i). \tag{5.5}$$

Since $c > 0$, $\mathring{u} := 0$ may serve as the Slater point. For our method applied to (1.1), we assume that f is evaluated exactly (i.e., $\epsilon_f = 0$ and $f_k = f$), whereas the approximate linearization condition (2.4b) boils down to finding an index $i_k \in I$ such that

$$h_k(\cdot) = \langle A_{i_k}, \cdot\rangle - c_{i_k} \quad \text{with} \quad h_k(u^k) \geq h(u^k) - \epsilon_h. \tag{5.6}$$

By duality, f_* is the common optimal value of (5.3) and (5.4). In view of Lemma 4.2, we assume that $f_* > -\infty$ and let $K' \subset \mathbb{N}$ be the set such that $V_k \xrightarrow{K'} 0$ and (4.17) holds; then $\nu_k > 0$ and $\mu_k := (1 - \nu_k)/\nu_k < \infty$ for large $k \in K'$. We shall show that the corresponding subsequence of the multipliers $\{\mu_k \beta_j^k\}_{j \in J_h^k}$ of (2.28b) solves the primal problem (5.3) approximately; thus, below we consider only $k \in K'$ such that $\nu_k > 0$.

The multipliers $\{\mu_k \beta_j^k\}_{j \in J_h^k}$ define an *approximate primal solution* $\hat{\lambda}^k \in \mathbb{R}^n_+$ via

$$\hat{\lambda}_i^k := \mu_k \sum_{j \in J_h^k : i_j = i} \beta_j^k \quad \text{for each } i \in I.$$

Let $\underline{1} := (1, \dots, 1) \in \mathbb{R}^n$. In this notation, using the form (5.6) of the linearizations h_j in (2.28b) and the fact that $\mu_k \check{h}_k(u^{k+1}) = \mu_k \check{e}_k(u^{k+1})$ (cf. (2.13)) yields the relations

$$\mu_k p_h^k = A\hat{\lambda}^k,\ \mu_k = \underline{1}\hat{\lambda}^k,\ \hat{\lambda}^k \geq 0,\ (u^{k+1}A - c)\hat{\lambda}^k = \mu_k \check{e}_k(u^{k+1}). \tag{5.7}$$

We first derive useful expressions for the primal function values $\psi_0(\hat{\lambda}^k)$ and $\psi(\hat{\lambda}^k)$.

Lemma 5.1. $\psi_0(\hat{\lambda}^k) = \tau_k + ([h_{\hat{u}}^k]_+ - \epsilon_k - \langle p^k, \hat{u}^k\rangle)/\nu_k$, $\psi(\hat{\lambda}^k) = (p^k - p_C^k)/\nu_k \geq p^k/\nu_k$.

Proof. Since $p_f^k = \nabla f = -b$ (cf. (2.11), (5.4)), $\mu_k p_h^k = A\hat{\lambda}^k$ by (5.7), and $\nu_k\mu_k = 1 - \nu_k$ by (4.14), the definitions of $\psi(\lambda)$ in (5.3) and of p^k in (2.18) give

$$\nu_k\psi(\hat{\lambda}^k) = \nu_k(A\hat{\lambda}^k - b) = \nu_k p_f^k + (1-\nu_k)p_h^k = p^k - p_C^k,$$

where $p_C^k \in \partial i_{\mathbb{R}^m_+}(u^{k+1})$ implies $p_C^k \le 0$ and $\langle p_C^k, u^{k+1}\rangle = 0$. Next, by (5.7) and (2.18),

$$\nu_k c\hat{\lambda}^k + (1-\nu_k)\check{e}_k(u^{k+1}) = \langle \nu_k\mu_k p_h^k, u^{k+1}\rangle = \langle (1-\nu_k)p_h^k + p_C^k, u^{k+1}\rangle = \langle p^k - \nu_k p_f^k, u^{k+1}\rangle,$$

where $\nu_k\langle p_f^k, u^{k+1}\rangle = \nu_k\check{f}_k(u^{k+1}) = \nu_k\check{e}_k(u^{k+1}) + \nu_k\tau_k$ by (2.13); hence, by (2.20)–(2.21),

$$-\nu_k c\hat{\lambda}^k - \nu_k\tau_k = \check{e}_k(u^{k+1}) - \langle p^k, u^{k+1}\rangle = \bar{e}_C^k(0) = [h_{\hat{u}}^k]_+ - \langle p^k, \hat{u}^k\rangle - \epsilon_k.$$

Dividing by ν_k gives the required expression of $\psi_0(\hat{\lambda}^k) := -c\hat{\lambda}^k$; for $\psi(\hat{\lambda}^k)$, see above. □

In terms of the optimality measure V_k of (2.19), the bounds of Lemma 5.1 imply

$$\hat{\lambda}^k \ge 0 \quad \text{with} \quad \psi_0(\hat{\lambda}^k) \ge \tau_k - V_k/\nu_k, \quad \psi_i(\hat{\lambda}^k) \ge -V_k/\nu_k, \quad i = 1{:}\, m. \tag{5.8}$$

We now show that $\{\hat{\lambda}^k\}_{k\in K'}$ converges to the set of *$\bar{\epsilon}$-optimal primal solutions* of (5.3)

$$\Lambda_{\bar{\epsilon}} := \{\, \lambda \in \mathbb{R}^n_+ : \psi_0(\lambda) \ge f_* - \bar{\epsilon}, \psi(\lambda) \ge 0 \,\}, \tag{5.9}$$

where $\bar{\epsilon} := \bar{\mu}\epsilon_h$, with $\bar{\mu}$ being the minimal Lagrange multiplier of (1.1); in our context, we may as well take (a possibly larger) $\bar{\mu} := \underline{1}\bar{\lambda}$ for any primal solution $\bar{\lambda}$ of (5.3).

Theorem 5.2. *Suppose $f_* > -\infty$. Let $K' \subset \mathbb{N}$ be such that $V_k \xrightarrow{K'} 0$ and* (4.17) *holds* (*see Lem.* 4.2(ii) *for sufficient conditions*). *Then the following statements hold.*

(i) *The sequence $\{\hat{\lambda}^k\}_{k\in K'}$ is bounded and all its cluster points lie in $\mathbb{R}^n_+$.*

(ii) *Let $\hat{\lambda}^\infty$ be a cluster point of $\{\hat{\lambda}^k\}_{k\in K'}$. Then $\hat{\lambda}^\infty \in \Lambda_{\bar{\epsilon}}$.*

(iii) $d_{\Lambda_{\bar{\epsilon}}}(\hat{\lambda}^k) := \inf_{\lambda\in\Lambda_{\bar{\epsilon}}} |\hat{\lambda}^k - \lambda| \xrightarrow{K'} 0$.

Proof. By Lemma 4.2, $\overline{\lim}_{k\in K'}\mu_k < \infty$ and $V_k/\nu_k \xrightarrow{K'} 0$. Since $\underline{\lim}_{k\in K'}\tau_k \ge f_* - \bar{\epsilon}$ by (4.17), the bounds of (5.8) yield $\underline{\lim}_{k\in K'}\psi_0(\hat{\lambda}^k) \ge f_* - \bar{\epsilon}$ and $\underline{\lim}_{k\in K'}\min_{i=1}^n \psi_i(\hat{\lambda}^k) \ge 0$.

(i) This follows from $\overline{\lim}_{k\in K'}\underline{1}\hat{\lambda}^k = \overline{\lim}_{k\in K'}\mu_k < \infty$ (cf. (5.7)) and $\{\hat{\lambda}^k\}_{k\in K'} \subset \mathbb{R}^n_+$.

(ii) We have $\hat{\lambda}^\infty \ge 0$, $\psi_0(\hat{\lambda}^\infty) \ge f_* - \bar{\epsilon}$ and $\psi(\hat{\lambda}^\infty) \ge 0$ by continuity of ψ_0 and ψ.

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

Remarks 5.3. (i) By Remark 3.9(ii), we may use $\bar{\epsilon} := \bar{\mu}\epsilon_h^\infty$ in (5.9) for Theorem 5.2.

(ii) By Lemma 2.8(iii) and the proof of Theorem 5.2, if an infinite loop between Steps 1 and 4 occurs, then $V_k \to 0$ yields $d_{\Lambda_{\bar{\epsilon}}}(\hat{\lambda}^k) \to 0$. Similarly, if Step 2 terminates with $V_k = 0$, then $\hat{\lambda}^k \in \Lambda_{\bar{\epsilon}}$. In both cases, we may take $\bar{\epsilon} := \bar{\mu}\epsilon_h^{k(l)}$ by Remark 3.9(ii).

(iii) Given two tolerances $\epsilon_F, \epsilon_{tol} > 0$, the method may stop if $h_{\hat{u}}^k \le \epsilon_F$,

$$\psi_0(\hat{\lambda}^k) \ge f(\hat{u}^k) - \epsilon_{tol} \quad \text{and} \quad \psi_i(\hat{\lambda}^k) \ge -\epsilon_{tol}, \quad i = 1{:}\, m.$$

Then $\psi_0(\hat{\lambda}^k) \ge f_* - \bar{\mu}(\epsilon_h + \epsilon_F) - \epsilon_{tol}$ from $f(\hat{u}^k) \ge f_* - \bar{\mu}(\epsilon_h + \epsilon_F)$, so $\hat{\lambda}^k$ is an approximate solution of (5.3). This stopping criterion will be met when $V_k/\nu_k \le \epsilon_{tol}$ in (5.8).

References

[Ber99] D. P. Bertsekas, *Nonlinear Programming*, second ed., Athena Scientific, Belmont, MA, 1999.

[BLM+05] O. Briant, C. Lemaréchal, Ph. Meurdesoif, S. Michel, N. Perrot and F. Vanderbeck, *Comparison of bundle and classical column generation*, Research Report RR-5453, INRIA, Montbonnot, France, 2005.

[BTN05] A. Ben-Tal and A. Nemirovski, *Non-euclidean restricted memory level method for large-scale convex optimization*, Math. Programming **102** (2005) 407–456.

[Fáb00] C. I. Fábián, *Bundle-type methods for inexact data*, Central European J. Oper. Res. **8** (2000) 35–55.

[FlL99] R. Fletcher and S. Leyffer, *A bundle filter method for nonsmooth nonlinear optimization*, Numerical Analysis Report NA/195, Dept. of Math., Univ. of Dundee, Dundee, 1999.

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

[Kiw85] K. C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics 1133, Springer-Verlag, Berlin, 1985.

[Kiw87] ———, *A constraint linearization method for nondifferentiable convex minimization*, Numer. Math. **51** (1987) 395–414.

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

[Kiw91] ———, *Exact penalty functions in proximal bundle methods for constrained convex nondifferentiable minimization*, Math. Programming **52** (1991) 285–302.

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

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

[Kiw05] ———, *An inexact bundle approach to cutting-stock problems*, Tech. report, Systems Research Institute, Warsaw, 2005.

[Kiw06a] ———, *An inexact filter bundle method for nonsmooth convex optimization*, Tech. report, Systems Research Institute, Warsaw, 2006. In preparation.

[Kiw06b] ———, *A penalty bundle method with inexact subgradient linearizations*, Tech. report, Systems Research Institute, Warsaw, 2006. In preparation.

[Kiw06c] ———, *A proximal bundle method with approximate subgradient linearizations*, SIAM J. Optim. **16** (2006) 1007–1023.

[Kiw06d] ———, *A proximal-projection bundle method for Lagrangian relaxation, including semidefinite programming*, SIAM J. Optim. ? (2006). To appear.

[KRSS05] E. Karas, A. Ribeiro, C. Sagastizábal and M. Solodov, *A bundle-filter method for nonsmooth convex constrained optimization*, Tech. report, IMPA, Rio de Janeiro, Brasil, 2005.

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

[LüD04] M. E. Lübbecke and J. Desrosiers, *Selected topics in column generation*, Oper. Res. ? (2004). To appear.

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

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

[SaS05] C. Sagastizábal and M. V. Solodov, *An infeasible bundle method for nonsmooth convex constrained optimization without a penalty function or a filter*, SIAM J. Optim. **16** (2005) 146–169.

[Sav97] M. W. P. Savelsbergh, *A branch-and-price algorithm for the generalized assignment problem*, Oper. Res. **45** (1997) 831–841.