Research report manuscript No.
(will be inserted by the editor)

A strongly convergent proximal bundle method for convex minimization in Hilbert spaces

W. van Ackooij · J.Y. Bello Cruz · W. de Oliveira

December 6, 2013

Abstract A key procedure in proximal bundle methods for convex minimization problems is the definition of stability centers, which are points generated by the iterative process that successfully decrease the objective function. In this paper we study a different stability-center classification rule for proximal bundle methods. We show that the proposed bundle variant has three particularly interesting features: (i) the sequence of stability centers generated by the method converges strongly to the solution that lies closest to the initial point; (ii) the entire sequence of stability centers is contained in a ball with diameter equal to the distance between the initial point and the solution set; (iii) if the sequence of stability centers is finite, $\hat{x}$ being its last element, then the sequence of non-stability centers (null steps) converges strongly to $\hat{x}$. Property (i) is useful in some practical applications in which a minimal norm solution is requested. We show the interest of this property on several instances of a full sized unit-commitment problem.

Keywords Convex optimization, Nonsmooth optimization, Proximal bundle method, Strong convergence

Mathematics Subject Classification (2010) MSC 49M37 · MSC 65K05 · MSC 90C25

1 Introduction

In this paper we deal with problems of the form

$$f_* := \min_{x \in X} f(x)\,, \tag{1}$$

where X is a nonempty convex and closed subset of a real Hilbert space $\mathcal{H}$, endowed with the norm $\|\cdot\|$ induced by an inner product $\langle \cdot, \cdot \rangle$. Throughout this paper we assume that $f : \mathcal{H} \to \mathbb{R}$ is a lower

Wim van Ackooij
EDF R&D. OSIRIS
1, avenue du Général de Gaulle, 92141 Clamart Cedex, France
E-mail: wim.van-ackooij@edf.fr

José Yunier Bello Cruz
Instituto de Matemática e Estatística, Universidade Federal de Goiás, Goiânia, Brazil
E-mail: yunier@impa.br

Welington de Oliveira
Instituto Nacional de Matemática Pura e Aplicada - IMPA
Rua Dona Castorina 110, 22460-320, Rio de Janeiro, Brazil
E-mail: wlo@impa.br

semicontinuous and convex function, possibly nondifferentiable, and the solution set $\mathcal{S}_*$ of problem (1) is nonempty.

Bundle methods are designed to solve nonsmooth convex optimization problems like (1) (mainly when $\mathcal{H} = \mathbb{R}^n$) by only making use of first-order information on f. Such methods are well known by their robustness and by having an efficient stopping test. Moreover, differently from cutting-plane methods [17], most bundle methods, e.g., [18,21,5,25,9,10] have limited memory: the bundle of information built along the iterations can be kept bounded. In this way, we can save computer memory without impairing convergence. This feature is particularly interesting for large-scale optimization problems.

At any given iteration k, a general bundle algorithm makes use of a convex model f_k^M (in general a cutting-plane approximation) of f satisfying $f_k^M \leq f$, and a stability center $\hat{x}_\ell \in X$. In particular, the new iterate x_{k+1} of a proximal bundle method depends on $\hat{x}_\ell$ and f_k^M in the following manner:

$$x_{k+1} := \arg\min \left\{ f_k^M(x) + \frac{1}{2\mathtt{t}_k} \|x - \hat{x}_\ell\|^2 : \; x \in X \right\} , \tag{2}$$

where $\mathtt{t}_k > 0$ is the so-called proximal parameter; see for instance [9].

Generally, in proximal bundle methods the stability center $\hat{x}_\ell$ is some previous iterate, usually the "best" point generated by the iterative process so far. A classification rule decides when to update the stability center $\hat{x}_\ell$. Invariably, such a rule depends on the decrease of the function in the following manner, for given $v_k \geq 0$ and $\gamma \in (0,1)$:

$$\textbf{If } f(x_{k+1}) \leq f(\hat{x}_\ell) - \gamma v_k, \textbf{ then } \hat{x}_{\ell+1} \leftarrow x_{k+1} \textbf{ and } \ell \leftarrow \ell + 1\,. \tag{3}$$

In general, the *predicted decrease* v_k is defined as $v_k := f(\hat{x}_\ell) - f_k^M(x_{k+1})$, but other alternatives are also possible; see for instance [10]. As mentioned in [10], the inequality in (3) is a sort of Armijo rule, used in line-searches or trust-region algorithms for smooth optimization. When the inequality in (3) is not satisfied, then x_{k+1} is said to be a *null iterate*: the algorithm performs an unsuccessful step towards improving (significantly) the threshold $f(\hat{x}_\ell)$. However, null steps are useful to update the model f_k^M. A *descent step* is said to be performed when inequality (3) is satisfied: in this case, the iterate x_{k+1} is a point that improves the threshold $f(\hat{x}_\ell)$. As we are interested in minimizing f, it makes sense to take the best known value $f(x_{k+1})$ as the new threshold after a descent step. Therefore, we may take x_{k+1} as the new stability center: $\hat{x}_{\ell+1} \leftarrow x_{k+1}$. By construction, rule (3) ensures that the sequence of stability centers (or descent iterates) $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ is a subsequence of $(x_k)_{k\in\mathbb{N}}$, the sequence of iterates generated by solving subproblem (2).

Instead of employing rule (3), the present paper proposes a more general classification rule aiming to provide strong results on the sequence of stability centers:

$$\textbf{If } f(x_{k+1}) \leq f(\hat{x}_\ell) - \gamma v_k \textbf{ then } \hat{x}_{\ell+1} \leftarrow P_{\mathbb{X}_k}(x_0) \textbf{ and } \; \ell \leftarrow \ell + 1\,. \tag{4}$$

In rule (4), $P_{\mathbb{X}_k}(x_0)$ stands for the projection of a given initial point $x_0 \in X$ onto a well-chosen subset $\mathbb{X}_k \subseteq X$ (see (17) below).

Notice that by taking $\mathbb{X}_k = \{x_{k+1}\}$ rule (4) coincides with (3) for any given x_0 in $\mathcal{H}$. However, other alternatives to $\mathbb{X}_k$ are also possible and may imply that $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ is not a subsequence of $(x_k)_{k\in\mathbb{N}}$. We are particularly interested in projection sets $\mathbb{X}_k$ that provide the following properties:

(i) the sequence of stability centers $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ generated by the method converges strongly to $P_{\mathcal{S}_*}(x_0)$;

(ii) the entire sequence of stability centers is contained in a ball with diameter equal to the distance between the initial point and the solution set;

(iii) if the sequence of stability centers is finite, $\hat{x}$ being its last element, then $\hat{x}$ is the solution of problem (1) closest to x_0, and the sequence of null steps $(x_k)_{k\in\mathbb{N}}$ converges strongly to $\hat{x}$.

The above three properties improve the results given in [9, Proposition 4.3 and Theorem 4.4] for proximal bundle methods employing (3), that we recall here:

(i') if the sequence of stability center is infinite, then the sequence $(f(\hat{x}_\ell))_{\ell\in\mathbb{N}}$ converges to f_*. Moreover, if $\mathcal{H}$ a is finite-dimensional space, then $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ converges to a solution of problem (1);

(iii') if the sequence of stability centers is finite, $\hat{x}$ being its last element, then $\hat{x}$ is the solution of problem (1), and the sequence of null steps $(x_k)_{k\in\mathbb{N}}$ converges strongly to $\hat{x}$.

Notice that item (i') from [9] does not mention anything about the behavior of $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ when $\mathcal{H}$ is an infinite-dimensional space. Moreover, item (ii) above is not present in [9], that employs rule (3) to define the stability centers. Since the proposed rule (4) only has an impact over the sequence of stability centers $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$, it is natural to expect that when $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ has finitely many elements, strong convergence of $(x_k)_{k\in\mathbb{N}}$ to the last stability center $\hat{x}$ follows from [9, Theorem 4.4], see item (iii'). In contrast to (iii), item (iii') from [9] does not state that $\hat{x}$ is the solution of problem (1) closest to x_0.

We emphasize that property (i) establishes a connection between convex nondifferentiable minimization problems and *minimal norm solution problems*, that have been mostly studied in the differentiable setting; see [4,16] and references therein. Notice that property (i) states that the limit of the sequence $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ is a point in the solution set $\mathcal{S}_*$ of problem (1) that also solves the following *second-stage problem*

$$\min \|x - x_0\| \ \text{ s.t. } \ x \in \mathcal{S}_* \,, \tag{5}$$

known as the minimal norm solution problem of (1) (mainly when $x_0 = 0$ belongs to X).

The main contributions of the present paper are the following:

- improvement of the proximal bundle method given in [9], showing strong convergence of the whole sequence of stability centers (compare properties (i) and (i') above);
- we show that as long as problem (1) has a solution, the sequence of stability centers is contained in a certain ball with fixed radius;
- strong converge is shown without knowing the optimal value f_* of problem (1). This differs from the analysis of [6];
- provided that x_0 belongs to X, the proposed bundle algorithm solves the minimal norm solution problem associated to (1) without requiring either differentiability or Lipschitz continuity of f, in contrast to the methods given in [4].

This paper is organized as follows. Section 2 presents the connection of minimal norm solution problems and property (i). In Section 3 we rely on the proximal bundle method given in [9] to provide our algorithm. Moreover, we propose an easy manner to define the projection set $\mathbb{X}_k$ in (4) in order to ensure the stated properties (i)-(iii). Section 4 gives the convergence analysis of the proposed method. The interest of property (i) is highlighted in Section 5 on several unit-commitment problem instances.

Notation

Our notation is standard. For any points $x, y \in \mathcal{H}$, $\langle x, y\rangle$ stands for an inner product in a real Hilbert space $\mathcal{H}$, and $\|\cdot\|$ for its associated norm, i.e., $\|x\| = \sqrt{\langle x, x\rangle}$. The closed ball centered at x with radius R is denoted by $B[x, R]$. For a set $X \subset \mathcal{H}$, we denote by i_X its indicator function, i.e., $i_X(x) = 0$ if $x \in X$ and $i_X(x) = +\infty$ otherwise. The relative interior of X, denoted by $\mathtt{ri}(X)$, is the interior of X relative to its affine hull. The affine hull of X is the intersection of all affine sets that contain X. Given a convex function f, we remind the definition of the subdifferential of f at the point x, i.e., $\partial f(x) = \{g \in \mathcal{H} \mid f(y) \geqslant f(x) + \langle g, y - x\rangle \ \forall\, y \in \mathcal{H}\}$. Throughout this paper we assume that f satisfies the following property:

$$\partial f \text{ is bounded on bounded sets,} \tag{6}$$

which holds automatically when $\mathcal{H}$ is a finite-dimensional space in view of Theorem 4.6.1(ii) in [8]. For some equivalences with condition (6) see for instance [3, Proposition 16.17]. Assumption (6) has been considered in the convergence analysis of many optimization methods in infinite-dimensional spaces; see for instance [1,5,6] and references therein. In [9], the assumption (6) is also used, although this hypothesis is not explicitly stated in [9, Proposition 4.3].

2 Minimal norm solution problem

Let $\mathcal{S}_*$ be the solution set of problem (1), which is convex and closed (remember that X is a convex and closed set, and f is a convex and lower semicontinuous function). Suppose that problem (1) has multiple optimal solutions. In this case, one may be interested in finding a point in $\mathcal{S}_*$ satisfying some specific criterion: for instance, the minimal norm solution with respect to the initial point, x_0, that is the unique solution of the second-stage problem (5).

The recent work [4] proposes a gradient-based method for solving minimal norm solution problems assuming in (1) that $\mathcal{H} = \mathbb{R}^n$, x_0 is an arbitrary point, f is differentiable, and Lipschitz continuous on $\mathbb{R}^n$. As in this paper, [4] finds the optimal solution of problem (5) dealing explicitly only with the *core problem* (1). In what follows we present an example illustrating the usefulness of property (i).

2.1 Unit-commitment problem

Consider a hydro-thermal system composed of m power plants (hydro-valleys, thermal or nuclear-plants), and the task of determining which power plants are to be used in order to generate enough power to satisfy an electrical load demand $d \in \mathbb{R}^T$, with minimum operating cost and reliability. Depending on the time-horizon T of planning, this problem is known as a unit-commitment problem, and it can be written schematically as:

$$\begin{aligned} \min_{p\in\mathbb{R}^n} \quad & \sum_{i=1}^m c^i(p^i) \\ s.t. \quad & \sum_{i=1}^m h^i(p^i) \geq d \\ & p^i \in P^i \subseteq \mathbb{R}^{n_i}, \end{aligned} \tag{7}$$

where a unit is represented by i, with $c^i : \mathbb{R}^{n_i} \to \mathbb{R}$ and $h^i : \mathbb{R}^{n_i} \to \mathbb{R}^T$ its cost and production functions, respectively. The decision variable p^i stands for the production schedule, which lies in the feasible set of technical constraints P^i, and has operating cost $c^i(p^i)$. The constraint $\sum_{i=1}^m h^i(p^i) \geq d$ links the decision variables in order to satisfy the offer-demand equilibrium constraints. These can also cover spinning reserve requirements (e.g., [14, 11]). The main difficulties of the problem reside in its large scale (m, n) and heterogeneity: hydro valleys have fundamentally different constraints (e.g., [13, 28]) than thermal units (e.g., [19, 20]) and potentially stochastic nature; see for instance [27, 7, 23] and references therein.

Due to the presence of multiple coupling constraints, Lagrangian Relaxation appears as a natural approach for dealing with this kind of problems; see [24] and references therein. The dual problem of (7) can be represented as (1) by taking

$$f(x) = \langle x, d\rangle - \sum_{i=1}^m \left(\min_{p^i\in P^i} c^i(p^i) - \langle x, h^i(p^i)\rangle \right), \tag{8}$$

and $X = \mathbb{R}^T_+$. The key advantage of moving to the dual is that the above inner problem is separable and can be solved by special dedicated approaches that can exploit specific structures.

Let $\bar{x}$ be an optimal dual solution. Due to its economic interpretation, $\bar{x}$ represents the marginal costs of power production. Since the price of electricity is decided by taking into account the marginal costs, power companies and/or governmental entities may wish to keep $\bar{x}$ as small as possible (high marginal costs imply a high price for the electricity). Therefore, for the unit-commitment problem the aim is to find a solution to the dual problem like (1) with the minimal norm, i.e., finding x_* solution of the second stage problem (5) with reference point given by $x_0 = 0$ in $\mathbb{R}^n$. In other words, one wishes to find $x_* = P_{\mathcal{S}_*}(x_0)$. Alternatively, x_0 could be a specific smooth signal and we may wish to single out a dual solution closest to such a signal. The suggested method can thus be interpreted as a natural way of obtaining stabilized prices, without penalizing the oscillation behavior of x in the dual objective f, as proposed in [29].

We emphasize that the proposed algorithm that solves problem (1) finds the optimal solution of (5) without explicitly dealing with the latter problem. This is a consequence of property (i).

3 Strongly convergent proximal bundle method

The proposed method for solving problem (1) generates two sequences of feasible iterates: $(x_k)_{k\in\mathbb{N}}$ and $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ both contained in X. For each point x_k (respectively $\hat{x}_\ell$) an oracle is called to compute $f(x_k)$ and a subgradient $g_k \in \partial f(x_k)$ (respectively $f(\hat{x}_\ell)$ and $\hat{g}_\ell \in \partial f(\hat{x}_\ell)$). With such information, the method creates the linearization

$$f_k^L(x) := f(x_k) + \langle g_k, x - x_k\rangle \le f(x)\,, \tag{9}$$

where the inequality follows from the convexity of f. At iteration k a polyhedral *cutting-plane* model of f is available:

$$f_k^M(x) := \max_{j\in\mathcal{B}_k} f_j^L(x) \;\text{ with } \mathcal{B}_k \text{ an index set, e.g., } \;\mathcal{B}_k \subset \{1,\dots,k\}\,. \tag{10}$$

The index set $\mathcal{B}_k$ gathers the bundle of information: $\{x_j,\, f(x_j), g_j\}_{j\in\mathcal{B}_k}$.

As already mentioned, given a stability center $\hat{x}_\ell$ and a prox-parameter $\mathtt{t}_k > 0$, the next iterate x_{k+1} is obtained by solving subproblem (2). Some properties of the minimizer x_{k+1} are given in the following lemma.

Lemma 1 *Suppose that one of the following constraint qualifications holds: X is a polyhedral set or $\boldsymbol{ri}(X) \neq \emptyset$. Denoting by i_X the indicator function of the set X, the optimality conditions for* (2) *imply*

$$x_{k+1} = \hat{x}_\ell - \mathtt{t}_k p_k \;\text{ with }\; p_k = p_f^k + p_X^k\,, \tag{11}$$

where $p_f^k \in \partial f_k^M(x_{k+1})$ and $p_X^k \in \partial i_X(x_{k+1})$. Moreover, the affine function

$$f_{k^a}^L(x) := f_k^M(x_{k+1}) + \langle p_k, x - x_{k+1}\rangle \tag{12}$$

is an underestimate of the model: $f_{k^a}^L(x) \le f_k^M(x)$ for all $x \in X$.

Proof Under the current constraint qualification, relation (11) comes from the optimality conditions of the problem

$$\min_x f_k^M(x) + \frac{1}{2\mathtt{t}_k}\|x - \hat{x}_k\|^2 + i_X(x)\,,$$

which is equivalent to problem (2) in terms of optimal value and solution x_{k+1}.

Since $p_f^k \in \partial f_k^M(x_{k+1})$ and $p_X^k \in \partial i_X(x_{k+1})$, the subgradient inequality gives

$$\begin{aligned} f^M(x_{k+1}) + \langle p_f^k, x - x_{k+1}\rangle &\le f_k^M(x)\\ \langle p_X^k, x - x_{k+1}\rangle &\le i_X(x)\end{aligned}$$

for all $x \in X$. By summing the above inequalities and using (11) we get (12). □

We next provide useful connections between the predicted decrease v_k and the aggregate linearization error defined as:

$$\hat{e}_k := f(\hat{x}_\ell) - f_{k^a}^L(\hat{x}_\ell)\,. \tag{13}$$

We also establish a key relation that will be the basis for the subsequent convergence analysis.

Proposition 1 *Let $\ell \geq 0$, $v_k = f(\hat{x}_\ell) - f_k^M(x_{k+1})$ in (4), and K^ℓ be an index set gathering the iterates issued by the ℓ-th stability center $\hat{x}_\ell$, i.e.,*

$$K^\ell := \{k \in \mathbb{N} \mid x_{k+1} = \hat{x}_\ell - \mathtt{t}_k p_k\}.$$

It holds that

$$\hat{e}_k \geq 0, \quad \hat{e}_k + \mathtt{t}_k \|p_k\|^2 = v_k \geq 0 \ \textit{ for all } \ k \in K^\ell. \tag{14}$$

Moreover, the following inequality holds:

$$f(\hat{x}_\ell) \leq f(x) + \hat{e}_k + \|p_k\| \|\hat{x}_\ell - x\| \quad \textit{for all } x \in X \textit{ and } k \in K^\ell. \tag{15}$$

Proof Let $k \in K^\ell$. The fact that $\hat{e}_k \geq 0$ follows directly from (9) and (12). To show (14), note that

$$\begin{aligned}
\hat{e}_k &= f(\hat{x}_\ell) - f_{k^a}^L(\hat{x}_\ell) \\
&= f(\hat{x}_\ell) - (f_k^M(x_{k+1}) + \langle p_k, \hat{x}_\ell - x_{k+1}\rangle) \\
&= v_k - \langle p_k, \hat{x}_\ell - x_{k+1}\rangle \\
&= v_k - \mathtt{t}_k \|p_k\|^2,
\end{aligned}$$

where the last equality follows from (11). This completes the proof of (14).

We next verify (15). Using again (9), for all $x \in X$, it holds that

$$\begin{aligned}
f(x) &\geq f_{k^a}^L(x) \\
&= f_k^M(x_{k+1}) + \langle p_k, x - x_{k+1}\rangle \\
&= f(\hat{x}_\ell) - (f(\hat{x}_\ell) - f_k^M(x_{k+1})) + \langle p_k, x - \hat{x}_\ell\rangle + \langle p_k, \hat{x}_\ell - x_{k+1}\rangle \\
&= f(\hat{x}_\ell) - v_k + \langle p_k, \hat{x}_\ell - x_{k+1}\rangle + \langle p_k, x - \hat{x}_\ell\rangle \\
&= f(\hat{x}_\ell) - \hat{e}_k + \langle p_k, x - \hat{x}_\ell\rangle
\end{aligned}$$

and (15) follows by taking into account (14) and the Cauchy-Schwartz inequality. □

Provided that the stability sequence $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$ is bounded (this feature comes from item (ii) to be shown in Proposition 3 below), it follows from inequality (15) that convergence amounts to obtaining the following property:

$$\text{a subsequence } ((\hat{e}_k, p_k))_{k \in \mathcal{A}} \text{ such that } \lim_{k \in K} \hat{e}_k = 0 \text{ and } \lim_{k \in K} \|p_k\| = 0. \tag{16}$$

Hence, a proximal bundle algorithm can stop with a satisfactory solution $\hat{x}_\ell$ when both $\hat{e}_k$ and $\|p_k\|$ are small.

3.1 Definition of the projection set

In order to obtain the properties (i)-(iii) stated in the introduction, a crucial requirement for the projection set $\mathbb{X}_k$ is the inclusion $\mathcal{S}_* \subseteq \mathbb{X}_k \subseteq X$ for all k, where $\mathcal{S}_*$ is the solution set of problem (1). With the aim of providing property (ii), the projection set $\mathbb{X}_k$ must contain the points $x \in X$ that satisfy the *angular constraint* $\langle x_0 - \hat{x}_\ell, x - \hat{x}_\ell\rangle \leq 0$, for a given initial point $x_0 \in X$ (see Proposition 3 below). We will provide a certain *minimal requirement* for the projection set, which is crucial to obtain (i)-(iii). We will refer to this set as $\mathbb{X}_k^{\mathtt{min}}$. Then convergence will be shown for all sets $\mathbb{X}_k$ satisfying $\mathcal{S}_* \subseteq \mathbb{X}_k \subseteq \mathbb{X}_k^{\mathtt{min}}$. The minimal projection set is defined as:

$$\mathbb{X}_k^{\mathtt{min}} := \left\{ x \in X \,\middle|\, \begin{array}{ll} \langle \hat{g}_\ell, x - \hat{x}_\ell\rangle & \leq f_k^{\text{best}} - f(\hat{x}_\ell) \\ \langle x_0 - \hat{x}_\ell, x - \hat{x}_\ell\rangle & \leq 0 \end{array} \right\} \quad \text{with } \ f_k^{\text{best}} = \min_{1 \leq j \leq k+1} f(x_j). \tag{17}$$

A tighter projection set can be obtained by inserting the cutting-plane model in definition (17), that is

$$\mathbb{X}_k := \left\{ x \in X \,\middle|\, \begin{array}{ll} f_k^M(x) & \leq f_k^{\text{best}} \\ \langle \hat{g}_\ell, x - \hat{x}_\ell\rangle & \leq f_k^{\text{best}} - f(\hat{x}_\ell) \\ \langle x_0 - \hat{x}_\ell, x - \hat{x}_\ell\rangle & \leq 0 \end{array} \right\} \subseteq \mathbb{X}_k^{\mathtt{min}}, \tag{18}$$

with f_k^{best} defined in (17). An even tighter set is available when the optimal value f_* of (1) is known: just replace f_k^{best} in (18) by f_*. Naturally, the tighter is $\mathbb{X}_k \supseteq \mathcal{S}_*$ the better is the stability center $\hat{x}_{\ell+1} = P_{\mathbb{X}_k}(x_0)$. Ideally, one wishes to have $\mathbb{X}_k = \mathcal{S}_*$ so that $\hat{x}_{\ell+1}$ would be the solution of problem (1) that is nearest to the initial point, solving in this way a sort of minimal normal solution problem. However, having $\mathbb{X}_k = \mathcal{S}_*$ is, of course, impossible in practice.

We recall that we have assumed that $\mathcal{S}_* \neq \emptyset$. We now proceed to show that the sequence of points generated by rule (4) is well defined. In order to do so, we follow the lead of [5] to define the following set

$$V := \{x \in \mathcal{H} \,|\, \langle x_0 - x, z - x\rangle \leq 0, \ \text{ for all } \ z \in \mathcal{S}_*\}\,,$$

which is nonempty because x_0 belongs to V.

Proposition 2 *Let k be an iteration index such that the point x_{k+1} satisfies the inequality in (4). Moreover, let $\mathbb{X}_k$ be defined in (17), and $\hat{x}_{\ell+1} = P_{\mathbb{X}_k}(x_0)$. If $\hat{x}_\ell \in V$, then*

(a) $\mathcal{S}_* \subseteq \mathbb{X}_k \subseteq X$ and $\hat{x}_{\ell+1} \in V$;
(b) as a result of (a), the whole sequence $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ is contained in V, and the inclusion $\mathcal{S}_* \subseteq \mathbb{X}_k$ holds.

Proof Let $z \in \mathcal{S}_*$ be a solution of problem (1). It follows from the convexity of f that

$$\langle g_{\hat{x}_\ell}, z - \hat{x}_\ell\rangle \leq f(z) - f(\hat{x}_\ell) \leq f_k^{\text{best}} - f(\hat{x}_\ell)\,,$$

where the last inequality follows from definition of f_k^{best}. This shows that z satisfies the first inequality in (17). Since $\hat{x}_\ell$ belongs to V by assumption, we have $\langle x_0 - \hat{x}_\ell, z - \hat{x}_\ell\rangle \leq 0$ and z satisfies the second inequality in (17) as well. Hence, we have shown that $\mathcal{S}_* \subseteq \mathbb{X}_k$. Since $\mathcal{S}_* \neq \emptyset$, then $\mathbb{X}_k$ is nonempty and thus $\hat{x}_{\ell+1} = P_{\mathbb{X}_k}(x_0)$ is well defined. The inclusion $\mathbb{X}_k \subseteq X$ is trivial. Since $\hat{x}_{\ell+1}$ is the orthogonal projection of x_0 onto the projection set $\mathbb{X}_k$, the inequality

$$\langle x - \hat{x}_{\ell+1}, x_0 - \hat{x}_{\ell+1}\rangle \leq 0 \ \text{ holds true for all } \ x \in \mathbb{X}_k\,.$$

In particular, $\langle z - \hat{x}_{\ell+1}, x_0 - \hat{x}_{\ell+1}\rangle \leq 0$ for all $z \in \mathcal{S}_*$, i.e., $\hat{x}_{\ell+1}$ belongs to V, showing item (a). In order to show item (b), notice that $x_0 \in V$. To prove that the whole sequence $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ is contained in V apply item (a) inductively. Since x_{k+1} satisfying (4) was arbitrary, we have $\mathcal{S}_* \subseteq \mathbb{X}_k \subseteq X$ as it was to be shown. □

Corollary 1 *The statements of Proposition 2 also hold when $\mathbb{X}_k$ is defined as in (18).*

Proof In order to show that $\mathcal{S}_* \subseteq \mathbb{X}_k$, we need only establish $f_k^M(z) \leq f_k^{\text{best}}$ for all $z \in \mathcal{S}_*$. The latter inequality readily follows from $f_k^M(z) \leq f(z) \leq f_k^{\text{best}}$. The remainder of the proof is a verbatim copy of that of Proposition 2. □

Remark 1 Obviously convexity of $\mathbb{X}_k$, orthogonal projection, and $\mathcal{S}_* \subseteq \mathbb{X}_k$ are the important properties for deriving a result similar to Proposition 2. The choice $\mathbb{X}_k = \{x_{k+1}\}$, which results in rule (3), can not be shown to have properties (a) and (b) above.

3.2 Strong proximal bundle algorithm

The strong proximal-bundle algorithm can be written as follows:

Algorithm 1 Strongly convergent proximal-bundle algorithm

Step 0. (Initialization) Select $\gamma \in (0,1)$ and $\mathtt{t}_1, \mathtt{t}_{\mathtt{min}}$ such that $\mathtt{t}_1 \geq \mathtt{t}_{\mathtt{min}} > 0$. Choose $x_0 \in X$ and stopping tolerances $\text{Tol}_{\hat{e}}, \text{Tol}_p > 0$. Call an oracle to compute $(f(x_0), g_0)$.
Set $\hat{x}_0 \leftarrow x_0$, $k \leftarrow 0$, $\ell \leftarrow 0$ and $\mathcal{B}_0 \leftarrow \{0\}$,

Step 1. (Next iterate) Obtain x_{k+1} by solving (2)
Set $p_k \leftarrow (\hat{x}_\ell - x_{k+1})/\mathtt{t}_k$, $v_k \leftarrow f(\hat{x}_\ell) - f_k^M(x_{k+1})$, and $\hat{e}_k \leftarrow v_k - \mathtt{t}_k\|p_k\|^2$
Step 2. (Stopping test) If $\hat{e}_k \leq \mathrm{Tol}_{\hat{e}}$ and $\|p_k\| \leq \mathrm{Tol}_p$, stop
Step 3. (Oracle call) Call the oracle to compute $(f(x_{k+1}), g_{k+1})$
Step 3.1.(Descent test) **If** $f(x_{k+1}) \leq f(\hat{x}_\ell) - \gamma v_k$, **then**
Set $\mathbb{X}_k$ as in (17) (or as in (18)) and obtain $\hat{x}_{\ell+1} \leftarrow P_{\mathbb{X}_k}(x_0)$
Call the oracle again to compute $(f(\hat{x}_{\ell+1}), \hat{g}_{\ell+1})$
Set $\ell \leftarrow \ell + 1$ and choose $\mathtt{t}_{k+1} \geq \mathtt{t}_k$
Step 3.2.(Null step) **Else**, choose $\mathtt{t}_{k+1} \in [\mathtt{t}_{\mathtt{min}}, \mathtt{t}_k]$
Step 4 (Bundle management) Choose $\mathcal{B}_{k+1} \supset \{k+1, k^a\}$
Set $k = k+1$ and go back to Step 1.

The aim of the algorithm is to estimate a solution to problem (1) as accurately as possible. The stability centers $\hat{x}_\ell$ provide such an estimation.

We emphasize that the aggregate linearization $f_{k^a}^L$ must enter into the bundle (i.e., $k^a \in \mathcal{B}_{k+1}$) only when some active linearization f_j^L is excluded (we say that f_j^L is an active linearization at iteration k when $f_j^L(x_{k+1}) = f_k^M(x_{k+1})$). This is the so called *bundle compression mechanism*, that allows for a limit-memory algorithm.

Essentially Algorithm 1 is a classical proximal bundle method, except for Step 3.1. Notice that if we choose $\mathbb{X}_k = \{x_{k+1}\}$ for all k, then the next stability $\hat{x}_{\ell+1}$ coincides with x_{k+1}, and we have the classical rule (3) to define the stability center. Convergence can then be established following classic lines of proof (see for instance, [9] and [10]). However, when $\mathbb{X}_k$ is defined as in (17) (or (18)), then $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ converges strongly to $P_{\mathcal{S}_*}(x_0)$, as we will show in the next section.

4 Convergence analysis

We start this section by showing that property (ii) stated in Section 1 holds.

Proposition 3 *Let $x_* \in \mathcal{S}_*$ be the projection of the initial iterate x_0 onto $\mathcal{S}_*$, i.e., $x_* = P_{\mathcal{S}_*}(x_0)$. Moreover let $\mathbb{X}_k$ be such that $\mathcal{S}_* \subseteq \mathbb{X}_k$. Then, the sequence $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ is contained in the closed ball centered in $(x_0 + x_*)/2$ with radius $\|x_* - x_0\|/2$, i.e.,*

$$(\hat{x}_\ell)_{\ell\in\mathbb{N}} \subset B\left[\frac{x_0 + x_*}{2}, \frac{\|x_* - x_0\|}{2}\right]. \tag{19}$$

Proof Define the points $y_\ell = \hat{x}_\ell - \frac{1}{2}(x_0 + x_*)$ and $y_* = x_* - \frac{1}{2}(x_0 + x_*)$. With this definition we can write x_0 as $x_0 = -y_* + \frac{1}{2}(x_0 + x_*)$. Since $x_* \in \mathcal{S}_* \subseteq \mathbb{X}_k$, it follows from the angular inequality defining $\mathbb{X}_k$ that

$$\begin{aligned}
0 &\geq \langle x_* - \hat{x}_\ell, x_0 - \hat{x}_\ell\rangle \\
&= \big\langle [y_* + \frac{1}{2}(x_0 + x_*)] - [y_\ell + \frac{1}{2}(x_0 + x_*)], [-y_* + \frac{1}{2}(x_0 + x_*)] - [y_\ell + \frac{1}{2}(x_0 + x_*)]\big\rangle \\
&= \langle y_* - y_\ell, -y_* - y_\ell\rangle = \|y_\ell\|^2 - \|y_*\|^2.
\end{aligned}$$

Hence, we have shown that

$$\left\|\hat{x}_\ell - \frac{x_0 + x_*}{2}\right\| \leq \left\|x_* - \frac{x_0 + x_*}{2}\right\| = \frac{\|x_* - x_0\|}{2},$$

establishing (19). □

In what follows we assume that $\mathrm{Tol}_{\hat{e}} = \mathrm{Tol}_p = 0$ and that Algorithm 1 does not stop. If the algorithm stops, then by (15) the last stability center $\hat{x}_\ell$ is a solution to problem (1), and by Proposition 3 all the stated properties (i)-(iii) are automatically satisfied (since the sequences $(\hat{x}_\ell)_{\ell\in\mathbb{N}}$ and $(x_k)_{k\in\mathbb{N}}$ are finite). In order to establish the stated properties (i) and (iii), assuming that the algorithm does not stop, we split our analysis into the two exclusive cases below:

– either the stability center $\hat{x}_\ell = \hat{x}$ remains unchanged after a finite number of iterations;
– or there are infinitely many stability centers.

The first case is addressed in [9, Proposition 4.3], since Algorithm 1 coincides with the one given in [9] when finitely many descent steps are performed. However, since assumption (6) is required in [9, Proposition 4.3] but not explicitly stated, we provide the complete proof here.

4.1 Finitely many stability centers

Notice that for iterates in which the stability center $\hat{x}_\ell$ is fixed, i.e., *null steps*, Algorithm 1 works like the classical proximal bundle method. We assume in this subsection that there exists a last stability center $\hat{x}_\ell = \hat{x}$, i.e., the descent step counter ℓ is no longer updated for all iteration indices k greater than a certain $\bar{k}$. Let $k \geq \bar{k}$; we start by defining the following two useful functions:

$$f_k^S(x) := f_k^M(x) + \frac{1}{2\mathtt{t}_k}\|x-\hat{x}\|^2 \tag{20}$$

$$\begin{aligned} f_{k^a}^S(x) &:= f_{k^a}^L(x) + \frac{1}{2\mathtt{t}_k}\|x-\hat{x}\|^2 \\ &= f_k^M(x_{k+1}) + \langle p_k, x - x_{k+1}\rangle + \frac{1}{2\mathtt{t}_k}\|x-\hat{x}\|^2 . \end{aligned} \tag{21}$$

Notice that $f_{k^a}^S$ is twice differentiable:

$$\nabla f_{k^a}^S(x) = p_k + \frac{x-\hat{x}}{\mathtt{t}_k} \quad \text{and} \quad \nabla^2 f_{k^a}^S(x) = \frac{1}{\mathtt{t}_k} I ,$$

where I is the identity operator in $\mathcal{H}$, i.e., $I(x) = x$ for all $x \in \mathcal{H}$. It follows from (11) that $\nabla f_{k^a}^S(x_{k+1}) = 0$, i.e., the point x_{k+1} is the unique solution of the problem $\min_{x\in\mathcal{H}} f_{k^a}^S(x)$. Moreover, since $f_{k^a}^S$ is quadratic, its second-order Taylor expansion is exact:

$$\begin{aligned} f_{k^a}^S(x) &= f_{k^a}^S(x_{k+1}) + \langle \nabla f_{k^a}^S(x_{k+1}), x - x_{k+1}\rangle + \frac{1}{2}\langle \nabla^2 f_{k^a}^S(x_{k+1})(x - x_{k+1}), x - x_{k+1}\rangle \\ &= f_{k^a}^S(x_{k+1}) + \langle 0, x - x_{k+1}\rangle + \frac{1}{2\mathtt{t}_k}\langle I(x - x_{k+1}), x - x_{k+1}\rangle \\ &= f_{k^a}^S(x_{k+1}) + \frac{1}{2\mathtt{t}_k}\|x - x_{k+1}\|^2 \\ &= f_k^S(x_{k+1}) + \frac{1}{2\mathtt{t}_k}\|x - x_{k+1}\|^2 , \end{aligned} \tag{22}$$

where the last equality follows from (20) and (21). The above development is crucial to show the following Lemma, which is essentially a reformulation of [10, Lemma 6.3] to our setting.

Lemma 2 *Let f_k^S be the model given in (20) and x_{k+1} be a null iterate, obtained from the stability center $\hat{x}$. If $\mathtt{t}_{k+1} \leq \mathtt{t}_k \leq \mathtt{t}_{k-1}$, then*

(a) *the sequence $(f_k^S(x_{k+1}))_{k\in\mathbb{N}}$ is nondecreasing and satisfies*

$$f_k^S(x_{k+1}) + \frac{1}{2\,\mathtt{t}_k}\|x_{k+2} - x_{k+1}\|^2 \leq f_{k+1}^S(x_{k+2}) ;$$

(b) *the sequence* $(f^S_k(x_{k+1}))_{k\in\mathbb{N}}$ *is bounded from above:*

$$f^S_k(x_{k+1}) + \frac{1}{2\,\mathtt{t}_k}\|\hat{x} - x_{k+1}\|^2 \le f(\hat{x})\,;$$

(c) *the inequality*

$$f^M_{k+1}(x_{k+1}) - f^M_{k-1}(x_k) \le f^S_k(x_{k+1}) - f^S_{k-1}(x_k) + \frac{1}{\mathtt{t}_{min}}\|x_{k+1} - x_k\|\|\hat{x} - x_k\|$$

holds true.

Proof As the aggregate index k^a enters the bundle in Step 4 of Algorithm 1, it follows that $f^L_{k^a}(x) \le f^M_{k+1}(x)$ for all $x \in \mathcal{H}$. Thus, using (21), we have

$$\begin{aligned} f^S_{k^a}(x) &= f^L_{k^a}(x) + \frac{1}{2\mathtt{t}_k}\|x - \hat{x}\|^2 \\ &\le f^M_{k+1}(x) + \frac{1}{2\mathtt{t}_k}\|x - \hat{x}\|^2 \\ &\le f^M_{k+1}(x) + \frac{1}{2\mathtt{t}_{k+1}}\|x - \hat{x}\|^2 = f^S_{k+1}(x)\,, \end{aligned}$$

where the last inequality follows from the assumption $\mathtt{t}_{k+1} \le \mathtt{t}_k$. Set $x = x_{k+2}$ in (22) to obtain (a), and $x = \hat{x}$ to obtain (b) (it follows from definition (20), with k replaced by $k+1$, that $f^S_{k+1}(\hat{x}) = f^M_{k+1}(\hat{x}) \le f(\hat{x})$). In order to show (c), note that

$$\begin{aligned} f^S_k(x_{k+1}) - f^M_k(x_{k+1}) &= \frac{1}{2\mathtt{t}_k}\|x_{k+1} - x_k + x_k - \hat{x}\|^2 \\ &\ge \frac{1}{2\mathtt{t}_k}\|x_k - \hat{x}\|^2 + \frac{1}{\mathtt{t}_k}\langle x_{k+1} - x_k, x_k - \hat{x}\rangle \\ &\ge \frac{1}{2\mathtt{t}_{k-1}}\|x_k - \hat{x}\|^2 + \frac{1}{\mathtt{t}_k}\langle x_{k+1} - x_k, x_k - \hat{x}\rangle \\ &\ge f^S_{k-1}(x_k) - f^M_{k-1}(x_k) + \frac{1}{\mathtt{t}_k}\langle x_{k+1} - x_k, x_k - \hat{x}\rangle\,, \end{aligned}$$

where the second inequality is due to $\mathtt{t}_k \le \mathtt{t}_{k-1}$. The result follows by applying the Cauchy-Schwartz inequality and assumption $\mathtt{t}_k \ge \mathtt{t}_{\mathtt{min}} > 0$. □

Items (a)-(c) in Lemma 2 are useful in the next lemma, which shows that the cutting-plane model evaluated at the iterate x_k asymptotically approximates the value of the function.

Lemma 3 *In addition to the setting of Lemma 2, assume* (6)*. Then the following holds*

$$\lim_{k\to\infty}[f(x_k) - f^M_{k-1}(x_k)] = 0\,.$$

Proof By Lemma 2(a) the sequence $(f^S_k(x_{k+1}))_{k\in\mathbb{N}}$ is nondecreasing. Thus, there exists a constant $C > 0$ such that $f^S_k(x_{k+1}) \ge -C$ for all k. Using Lemma 2(b), we conclude that

$$\frac{1}{2\mathtt{t}_k}\|\hat{x} - x_{k+1}\|^2 \le f(\hat{x}) - f^S_k(x_{k+1}) \le f(\hat{x}) + C\,,$$

showing that the sequences $(\|\hat{x} - x_k\|)_{k\in\mathbb{N}}$ and $(x_k)_{k\in\mathbb{N}}$ are bounded, since $(\mathtt{t}_k)_{k\in\mathbb{N}}$ is nonincreasing. By assumption (6), the subdifferential of f is bounded on bounded sets. It thus results that $(g_k)_{k\in\mathbb{N}}$ is a bounded sequence.

By Lemma 2(b) the sequence $(f^S_k(x_{k+1}))_{k\in\mathbb{N}}$ is bounded from above by $f(\hat{x})$. Lemma 2(a) shows that $(f^S_k(x_{k+1}))_{k\in\mathbb{N}}$ is nondecreasing and hence

$$\lim_{k\to\infty}[f^S_{k+1}(x_{k+2}) - f^S_k(x_{k+1})] = 0 \quad\text{and}\quad \lim_{k\to\infty}\|x_{k+2} - x_{k+1}\|^2 = 0\,, \tag{23}$$

where the second limit follows from Lemma 2(a) and the assumption that $\mathtt{t}_k$ is nonincreasing. Since $k \in \mathcal{B}_k$, we have:

$$f(x_k) + \langle g_k, x - x_k\rangle = f_k^L(x) \le f_k^M(x) \ \text{ for all } x \in \mathcal{H}.$$

Setting $x = x_{k+1}$ in the above inequality, we get $f(x_k) = f_k^L(x_{k+1}) + \langle g_k, x_k - x_{k+1}\rangle$. Therefore,

$$\begin{aligned} f(x_k) - f_{k-1}^M(x_k) &= f_k^L(x_{k+1}) + \langle g_k, x_k - x_{k+1}\rangle - f_{k-1}^M(x_k) \\ &\le f_k^M(x_{k+1}) + \langle g_k, x_k - x_{k+1}\rangle - f_{k-1}^M(x_k) \\ &\le f_k^S(x_{k+1}) - f_{k-1}^S(x_k) + \tfrac{1}{\mathtt{t}_{\mathtt{min}}}\|x_{k+1} - x_k\|\|\hat{x} - x_k\| + \langle g_k, x_k - x_{k+1}\rangle \,, \end{aligned}$$

where the last inequality is due to Lemma 2(c). Applying the limit with $k \to \infty$ in the above inequalities and taking into account (23), remembering that $(x_k)_{k\in\mathbb{N}}$ and $(g_k)_{k\in\mathbb{N}}$ are bounded sequences, we conclude that $\limsup_{k\to\infty}[f(x_k) - f_{k-1}^M(x_k)] \le 0$. Since f is convex, we have $f(x_k) \ge f_{k-1}^M(x_k)$ and the result follows. □

We can now show that property (iii) holds:

Proposition 4 (Finitely many stability center) *Assume that* (6) *holds and suppose that there exists a last stability center $\hat{x}$ such that the descent test* (4) *does not hold for k large enough. Then* (16) *holds, $\hat{x} = P_{\mathcal{S}_*}(x_0)$, and the sequence $(x_k)_{k\in\mathbb{N}}$ converges strongly to $\hat{x}$.*

Proof Lemma 3 ensures that $\lim_{k\to\infty}[f(x_k) - f_{k-1}^M(x_k)] = 0$. Since the descent test (4) does not hold for k large enough, it follows from (4) that

$$0 = \lim_{k\to\infty}[f(x_k) - f_{k-1}^M(x_k)] \ge \lim_{k\to\infty}(f(\hat{x}) - \gamma v_{k-1} - f_{k-1}^M(x_k)) = \lim_{k\to\infty}(1-\gamma)v_{k-1}\,,$$

i.e., $v_k \to 0$ because $v_k \ge 0$ (c.f. Proposition 1). Since $\mathtt{t}_k \ge \mathtt{t}_{\mathtt{min}} > 0$, it follows from (14) that

$$0 = \lim_{k\to\infty} v_k = \hat{e}_k + \mathtt{t}_k\|p_k\| \ge \hat{e}_k + \mathtt{t}_{\min}\|p_k\| \ge 0\,, \tag{24}$$

i.e., property (16) holds because $\hat{e}_k \ge 0$. Applying the limit with $k \to \infty$ in (15) for a fixed $\hat{x}_\ell = \hat{x}$, we conclude that

$$f(\hat{x}) \le f(x) + \lim_{k\to\infty}[\hat{e}_k + \|p_k\|\|x - \hat{x}\|] = f(x)\,.$$

Since $x \in X$ is arbitrary, we conclude that $\hat{x}$ is a solution to problem (1).

Moreover, it follows from Proposition 3 that $\hat{x}$ belongs to the ball B defined in (19), whose diameter is $\|P_{\mathcal{S}_*}(x_0) - x_0\|$. It is readily seen that x_0 belongs to B as well. Hence, by defining $x_* = P_{\mathcal{S}_*}(x_0)$ we have the inequality

$$\|\hat{x} - x_0\| \le \|x_* - x_0\|\,.$$

Since the solution set $\mathcal{S}_*$ is convex and the projection is orthogonal, then x_* satisfies the strict inequality:

$$\|x_* - x_0\| < \|z - x_0\| \ \text{ for all } \ z \in \mathcal{S}_*\backslash\{x_*\}\,.$$

It thus follows from the two inequalities above that $\hat{x} = x_* = P_{\mathcal{S}_*}(x_0)$, because $\hat{x} \in \mathcal{S}_*$.

In order to show the last assertion, note that by (24), $\lim_{k\to\infty}\mathtt{t}_k\|p_k\| = 0$. By using (11) we conclude that $\lim_{k\to\infty}\|x_{k+1} - \hat{x}\| = 0$, i.e., $(x_k)_{k\in\mathbb{N}}$ converges strongly to $\hat{x}$. □

In what follows we deal with the case of infinitely many stability centers. The choice of $\mathbb{X}_k$ then becomes crucial.

4.2 Infinitely many stability centers

We start by providing the following useful result.

Lemma 4 *Let $\mathbb{X}_k$ be such that $\mathbb{X}_k \subseteq \mathbb{X}_k^{\mathtt{min}}$, and let $k(\ell) = k$ be the iteration in which the $(\ell+1)$-th stability center was determined by Algorithm 1, i.e., $\hat{x}_{\ell+1} = P_{\mathbb{X}_{k(\ell)}}(x_0)$. Then, it holds that*

$$\|x_0 - \hat{x}_{\ell+1}\|^2 \geq \|x_0 - \hat{x}_\ell\|^2 + \frac{\gamma^2}{\|\hat{g}_\ell\|^2} v_{k(\ell)}^2, \tag{25}$$

for all $\ell \geq 0$ and with $\gamma \in (0,1)$.

Proof Since $\mathbb{X}_k \subseteq \mathbb{X}_k^{\mathtt{min}}$, any $x \in \mathbb{X}_k$ satisfies the two inequalities (17) defining $\mathbb{X}_k^{\mathtt{min}}$. Hence, since $\hat{x}_{\ell+1} = P_{\mathbb{X}_{k(\ell)}}(x_0)$, it follows by the first inequality in (17) that

$$\langle \hat{g}_\ell, \hat{x}_{\ell+1} - \hat{x}_\ell \rangle \leq f_{k(\ell)}^{\text{best}} - f(\hat{x}_\ell)\,.$$

Since by definition $f_{k(\ell)}^{\text{best}} \leq f(x_{k+1}) = f(x_{k(\ell)+1})$, we conclude that

$$\langle \hat{g}_\ell, \hat{x}_{\ell+1} - \hat{x}_\ell \rangle \leq f(x_{k(\ell)+1}) - f(\hat{x}_\ell) \leq -\gamma v_{k(\ell)}\,,$$

where the last inequality is due to (4). By changing the sign and applying the Cauchy-Schwartz inequality, we get

$$\|\hat{g}_\ell\| \|\hat{x}_\ell - \hat{x}_{\ell+1}\| \geq \gamma v_{k(\ell)}\,. \tag{26}$$

It also follows from definition (17) of $\mathbb{X}_{k(\ell)}$ that $\langle x_0 - \hat{x}_\ell, \hat{x}_{\ell+1} - \hat{x}_\ell \rangle \leq 0$. Therefore, by developing the squares in $\|x_0 - \hat{x}_\ell + \hat{x}_\ell - \hat{x}_{\ell+1}\|$, we have

$$\|x_0 - \hat{x}_{\ell+1}\|^2 \geq \|x_0 - \hat{x}_\ell\|^2 + \|\hat{x}_\ell - \hat{x}_{\ell+1}\|^2\,.$$

Inequality (25) then follows from (26) and the above inequality. □

Lemma 5 *Let $\mathbb{X}_k$ be such that $\mathcal{S}_* \subseteq \mathbb{X}_k \subseteq \mathbb{X}_k^{\mathtt{min}}$. Assume moreover that ∂f is bounded on bounded sets, i.e.,* (6) *holds. Then, condition* (16) *holds and each weak cluster point of the sequence $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$ belongs to the solution set $\mathcal{S}_*$.*

Proof We first recall that by Proposition 3 the sequence $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$ is bounded. Let $k(\ell)$ be the iteration in which the $(\ell+1)^{th}$ stability center was determined, i.e., $\hat{x}_{\ell+1} = P_{\mathbb{X}_{k(\ell)}}(x_0)$. Recalling that the expected decrease v_k is nonnegative by (14), it follows from Lemma 4 that the sequence $(\|x_0 - \hat{x}_\ell\|)_{\ell \in \mathbb{N}}$ is nondecreasing. It follows from Proposition 3 that the sequence $(\|x_0 - \hat{x}_\ell\|)_{\ell \in \mathbb{N}}$ is bounded (from above), and hence convergent by combining these two features. Therefore, by using (25)

$$0 \leq v_{k(\ell)}^2 \frac{\gamma^2}{\|\hat{g}_\ell\|^2} \leq \|\hat{x}_{\ell+1} - x_0\|^2 - \|\hat{x}_\ell - x_0\|^2,$$

we conclude that $\lim_{\ell \to \infty} \frac{\gamma^2}{\|\hat{g}_\ell\|^2} v_{k(\ell)}^2 = 0$.

Since $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$ is bounded, assumption (6) ensures that $(\hat{g}_\ell)_{\ell \in \mathbb{N}}$ is also bounded. Thus,

$$0 = \lim_{\ell \to \infty} v_{k(\ell)} = \lim_{\ell \to \infty} (\hat{e}_{k(\ell)} + \mathtt{t}_{k(\ell)} \|p_{k(\ell)}\|) \geq \lim_{\ell \to \infty} (\hat{e}_{k(\ell)} + \mathtt{t}_{\min} \|p_{k(\ell)}\|) \geq 0\,,$$

and (16) holds for $\mathcal{A} := \{k(1), k(2), \ldots\}$, because $\hat{e}_k \geq 0$ for all k. Let $(\hat{x}_{\ell'})_{\ell' \in \mathbb{N}}$ be any weakly convergent subsequence of $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$, with weak cluster point $\bar{x}$ (since $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$ is a bounded sequence in a Hilbert space $\mathcal{H}$, the existence of such a subsequence is ensured). Let $x \in X$ be an arbitrary point and $K^{\ell'}$ be an index set gathering the iterates issued by the ℓ'-th stability center, as defined in Proposition 1. It follows from the definition that $k(\ell')$ belongs to $K^{\ell'}$. Writing (15) with ℓ replaced by ℓ' gives:

$$f(\hat{x}_{\ell'}) \leq f(x) + \hat{e}_{k(\ell')} + \|p_{k(\ell')}\| \|\hat{x}_{k(\ell')} - x\|\,.$$

We recall that f is lower semicontinuous by assumption, and hence f is also weakly lower semicontinuous. Therefore,

$$f(\bar{x}) \leq \liminf_{\ell' \to \infty} f(\hat{x}_{\ell'}) \leq f(x) + \liminf_{\ell' \to \infty} [\hat{e}_{k(\ell')} + \|p_{k(\ell')}\| \|\hat{x}_{k(\ell')} - x\|] = f(x)\,,$$

because $(\hat{x}_{\ell'})_{\ell' \in \mathbb{N}}$ is a bounded subsequence of $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$. Therefore, we have shown that any weak cluster point $\bar{x}$ of $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$ belongs to the solution set $\mathcal{S}_*$ of problem (1). □

The following result shows that the sequence of stability centers converges strongly to the solution that lies closest to the initial point.

Proposition 5 (Infinitely many stability centers) *Let $\mathbb{X}_k$ be such that $\mathcal{S}_* \subseteq \mathbb{X}_k \subseteq \mathbb{X}_k^{\mathtt{min}}$. Assume moreover that ∂f is bounded on bounded sets, i.e.,* (6) *holds. Then $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$ converges strongly to x_*, where x_* is defined as $x_* = P_{\mathcal{S}_*}(x_0)$.*

Proof The point x_* is well defined, since $\mathcal{S}_* \neq \emptyset$ is a closed and convex set. Definition of $\hat{x}_\ell$ gives the following inequality:

$$\|\hat{x}_\ell - x_0\| \leq \|x - x_0\| \quad \text{for all} \quad x \in \mathbb{X}_{k(\ell-1)}\,, \tag{27}$$

with $k(\ell-1)$ the iteration in which the ℓ-th stability center was determined, i.e., $\hat{x}_\ell = P_{\mathbb{X}_{k(\ell-1)}}(x_0)$. In particular, it follows from Proposition 3 that $\mathcal{S}_* \subseteq \mathbb{X}_{k(\ell-1)}$ and thus,

$$\|\hat{x}_\ell - x_0\| \leq \|x_* - x_0\|, \quad \text{for all } \ell \geq 0. \tag{28}$$

It follows from Lemma 5 that every weak cluster point of the sequence $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$ belongs to $\mathcal{S}_*$. Let $(\hat{x}_{\ell'})_{\ell' \in \mathbb{N}}$ be a subsequence of $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$ that converges weakly to $\bar{x} \in \mathcal{S}_*$. By adding and subtracting x_0 in the term $\|\hat{x}_{\ell'} - x_*\|^2$ we get $\|\hat{x}_{\ell'} - x_*\|^2 = \|\hat{x}_{\ell'} - x_0\|^2 + \|x_* - x_0\|^2 - 2\langle \hat{x}_{\ell'} - x_0, x_* - x_0 \rangle$. Therefore, inequality (28) provides

$$\|\hat{x}_{\ell'} - x_*\|^2 \leq 2\|x_* - x_0\|^2 - 2\langle \hat{x}_{\ell'} - x_0, x_* - x_0 \rangle\,.$$

By applying the limit with $\ell' \to \infty$ and recalling $(\hat{x}_{\ell'})_{\ell' \in \mathbb{N}}$ converges weakly to $\bar{x} \in \mathcal{S}_*$, we obtain

$$\limsup_{\ell' \to \infty} \|\hat{x}_{\ell'} - x_*\|^2 \leq 2(\|x_* - x_0\|^2 - \langle \bar{x} - x_0, x_* - x_0 \rangle). \tag{29}$$

We now proceed to show that the right-hand side of the above inequality is nonpositive. Since $x_* = P_{\mathcal{S}_*}(x_0)$, then $\langle x_0 - x_*, \bar{x} - x_* \rangle \leq 0$. This inequality gives

$$\begin{aligned} 0 &\geq \langle x_0 - x_*, \bar{x} - x_* \rangle = \langle x_0 - x_*, x_0 - x_* \rangle + \langle x_0 - x_*, \bar{x} - x_0 \rangle \\ &= \|x_* - x_0\|^2 - \langle x_* - x_0, \bar{x} - x_0 \rangle\,. \end{aligned}$$

Thus, it follows from (29) that $\limsup_{\ell' \to \infty} \|\hat{x}_{\ell'} - x_*\|^2 \leq 0$, showing that $(\hat{x}_{\ell'})_{\ell' \in \mathbb{N}}$ converges strongly to x_*. As $(\hat{x}_{\ell'})_{\ell' \in \mathbb{N}}$ is an arbitrary weakly convergent subsequence of $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$, we have shown that every weakly convergent subsequence of $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$ converges strongly to x_*. Hence, the whole sequence $(\hat{x}_\ell)_{\ell \in \mathbb{N}}$ converges strongly to $x_* = P_{\mathcal{S}_*}(x_0)$. □

4.3 Convergence

We summarize the results of this paper in the following theorem.

Theorem 1 *Consider problem* (1) *with f a convex and lower semicontinuous function having ∂f bounded on bounded sets. Suppose that X is a convex and closed subset of a real Hilbert space $\mathcal{H}$. Let* $\mathrm{Tol}_{\hat{e}} = \mathrm{Tol}_p = 0$ *in Algorithm 1, and let $\mathbb{X}_k$ be defined such that $\mathcal{S}_* \subseteq \mathbb{X}_k \subseteq \mathbb{X}_k^{\mathtt{min}}$, where $\mathbb{X}_k^{\mathtt{min}}$ is defined in* (17)*. In addition, suppose that the solution set $\mathcal{S}_*$ of problem* (1) *is nonempty. Then* (16) *is satisfied, and properties (i)-(iii) stated in the Introduction hold true.*

Proof Proposition 5 gives item (i). Item (ii) is ensured by Proposition 3. Finally, Proposition 4 gives item (iii) (see also Proposition 4.3 in [9], assuming (6)). □

The following Corollary follows directly from Corollary 1:

Corollary 2 *Theorem 1 holds for* $\mathbb{X}_k$ *defined as* (18).

5 Some numerical experiments in unit-commitment

In this section we will present some numerical results obtained by solving specific unit-commitment problems (7). The instances are those coming from EDF (Électricité de France). Modeling details can be found in [11]. The main difficulty in solving the problem lies in the fact that P^i is a very complex (essentially discrete) set for several power units. The inner problem in (8) requires specific methods for its solution. Dynamic programming methods are used for solving the subproblems related to thermal units and special interior-point methods are used for solving the hydro-valley subproblems [15].

Beyond the derived production schedule $\bar{p}$, which is obviously of great importance, the optimal dual signal $\bar{x}$ is important on its own. We begin by recalling that in order to derive $\bar{p}$, additional Lagrangian heuristics (e.g., [26,12,11]) or Augmented Lagrangian based heuristics [2] are required. We will not investigate $\bar{p}$ further, but focus on $\bar{x}$ instead. This optimal dual solution $\bar{x}$ can be interpreted as a price signal explaining the optimal solution. Its information can be exploited to make minor changes to $\bar{p}$ in an economic way. For this reason properties of $\bar{x}$, such as "visual" smoothness, are desirable features. Price stabilization approaches such as those investigated in [29] are one way to obtain such features.

In the numerical tests we have used a standard proximal bundle Algorithm with elementary updating of the proximal parameter $\mathtt{t}_k$. This is basically Algorithm 1 except that the projection set is $\mathbb{X}_k = \{x_{k+1}\}$. Consequently, the need for a second oracle in Step 3.1 is eliminated. We refer to this algorithm as *Reference* algorithm. The second version of Algorithm 1, hereafter named *Strong Bundle*, uses definition (18) for $\mathbb{X}_k$, which we believe is more promising. The minimal requirement set $\mathbb{X}_k^{\mathtt{min}}$ was experimented on several academic cases and it was found to lead to significantly slower convergence. This choice was therefore not retained.

The interest of the specific projection step in Algorithm 1 is well demonstrated for two different EDF unit-commitment problems in Figures 1 and 2, where it is readily shown that the obtained price signal is indeed closer to the initial price signal (as it should be): $x_{\mathtt{ref}}$ and $x_{\mathtt{strong}}$ stand for the solutions obtained by the Reference and Strong Bundle algorithms, respectively. In particular, in Figure 1, the price spike showing at time step 86 is reduced significantly (by 10%). This is of interest for the operator, which can smoothen the initial price signal in order to have control over the smoothness of the optimal dual signal provided by the algorithm. In figure 2 the prices are significantly reduced over time steps 65 to 82.

In table 1 we present some numerical results obtained with both the Reference and Strong Bundle algorithms for the first instance. This table shows that the Strong Bundle algorithm requires significantly more iterations and oracle calls than the Reference algorithm in order to satisfy the stopping tests given at Step 2 of Algorithm 1. In Table 1 and 2, m is the Armijo-like constant employed in the descent test (4).

Table 1 A comparison of the Strong Bundle algorithm with the Reference algorithm (nne^x stands for $nn10^x$)

Method	Nb. Iter	Nb. Oracle Calls	f^*	Parameters
Reference	57	57	$-3.4842e^7$	$m = 0.1$
Strong Bundle	161	267	$-3.4837e^7$	$m = 0.1$
Strong Bundle	168	265	$-3.4837e^7$	$m = 0.2$

Similarly, in Table 2 we show the result for the second instance of the problem. As in Figure 2, the initial point was set as zero: $x_0 = 0$.

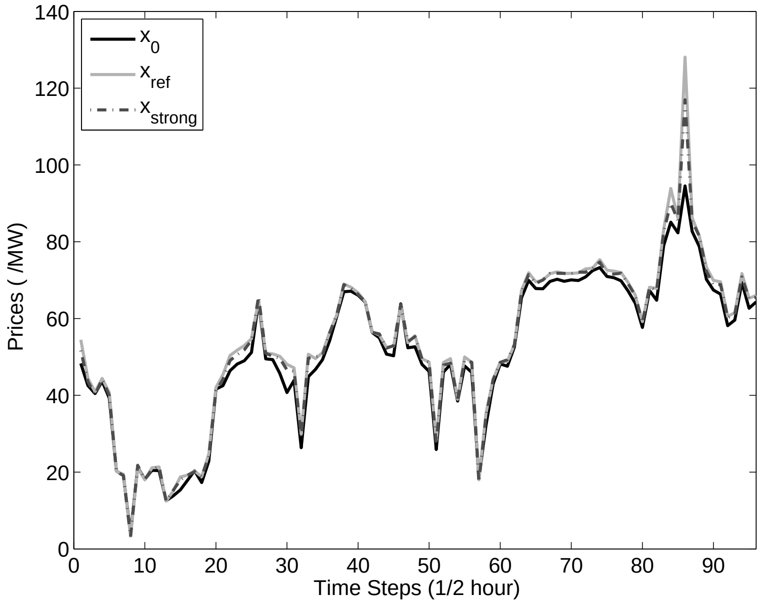

Fig. 1 A comparison of prices. Initial point x_0 carefully chosen for the first problem.

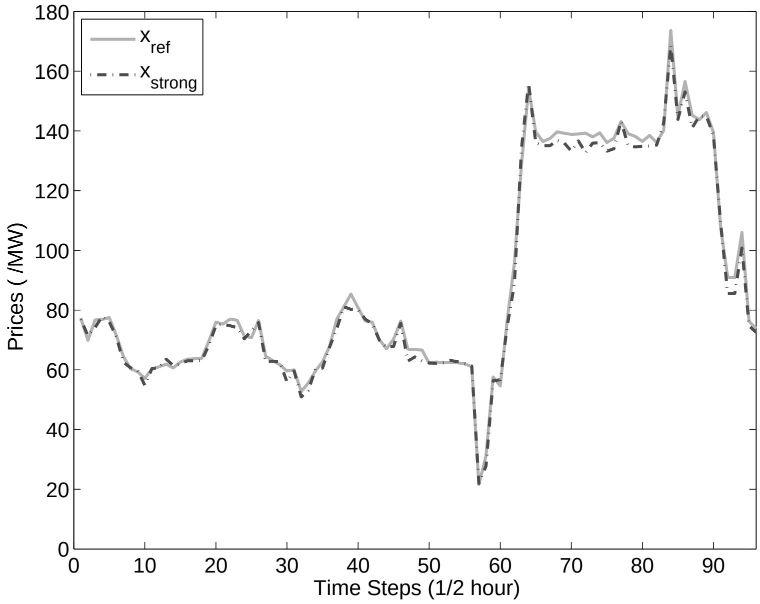

Fig. 2 A comparison of prices. Initial point $x_0 = 0$ for the second problem.

Table 2 A comparison of the Strong Bundle algorithm with the Reference algorithm (nne^x stands for $nn10^x$)

Method	Nb. Iter	Nb. Oracle Calls	f^*	Parameters
Reference	40	40	$-5.68001e^7$	$m = 0.1$
Strong Bundle	128	243	$-5.68024e^7$	$m = 0.1$
Strong Bundle	111	207	$-5.67896e^7$	$m = 0.2$

We have verified in our numerical experiments that the Strong Bundle algorithm provides a similar function value even when stopped after a maximum number of oracle calls. For this reason we have looked at 28 reputedly very hard instances and set the maximum number of oracle calls to 500. This means that the Reference algorithm performs 500 iterations, whereas the Strong Bundle algorithm can roughly perform only 250 iterations. Results are provided in Figure 3. We can observe here that the Strong

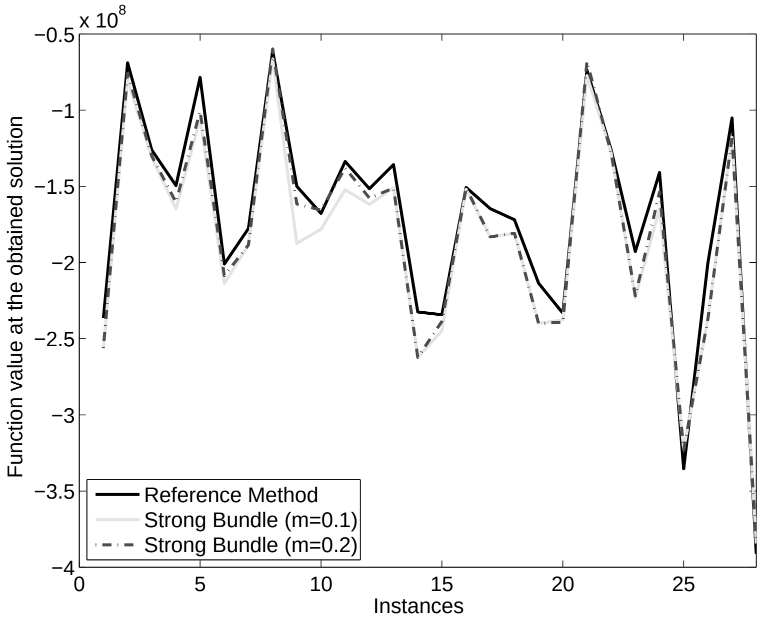

Fig. 3 A comparison of approximate optimal values.

Bundle algorithm provides a similar dual value (even slightly better). Altogether we have sketched the interest of using the Strong Bundle algorithm despite the increased number of iterations, that can be effectively reduced by considering a hybrid approach, as described below.

5.1 Hybrid algorithm

A hybrid approach combining both Reference and Strong Bundle algorithms could be of interest to reduce the number of iterations and oracle calls, bringing these quantities down to a more reasonable level. Such a hybrid approach could simply consist of the following two phases:

First phase: run a classical proximal bundle method for several iterations or until satisfying a loose stopping test;

Second phase: run and warm start the strong bundle method. For warm starting purposes, we can use the previously obtained cutting-plane model and the (better) estimate f^{best} for f_*, constructed in the first phase.

In both phases, it is probably desirable to employ a more efficient rule to update the prox-parameter $\mathtt{t}_k$ (e.g., poor-man Newton formulæ as in [22]).

In Table 3, we provide more detailed statistics on the number of iterations needed to obtain convergence. We have also included the above described hybrid approach for comparaison. These statistics were obtained on 50 easier instances. Consequently we have set the maximum number of oracle calls to 1000. The results of Table 3 clearly show that by combining both the Reference and Strong Bundle algorithms, in addition to an efficient rule for updating the prox-parameter, one should not be concerned to much about the increase in the number of oracle calls of the basic Strong Bundle algorithm versus the Reference one.

Table 3 Statistics on the number of oracle calls on 50 easy instances

Method	Average	Standard deviation	Minimum	Maximum	Parameters
Reference	157.5	70.8	57	392	$m = 0.1$
Strong Bundle	388.9	138.0	189	762	$m = 0.1$
Strong Bundle	390.0	134.2	189	702	$m = 0.2$
Strong Bundle	398.5	143.2	189	784	$m = 0.3$
Strong Bundle	406.4	149.6	189	861	$m = 0.4$
Hybrid	65.9	66.9	19	327	$m = 0.1$

6 Concluding Remarks

We have proposed a strongly convergent proximal bundle method in a real Hilbert space. Other strongly convergent proximal variants may be obtained by changing the definitions of the predicted decrease v_k and the projection set $\mathbb{X}_k$ in (4). In order to obtain the stated properties (i)-(iii) is crucial to choose the projection so that the results of Proposition 2 remain valid. Essentially, the projection set must be selected at each iteration in such a way to contain the solution set of the optimization problem, and to be contained in the *minimal requirement* set defined in (17).

Among the three properties of the proposed method, we emphasize that property (i) is of great interest in some practical applications, such as unit-commitment in which a minimal norm solution is requested. In particular, this feature was highlighted on unit-commitment problems coming from EDF.

Finally, we mention that a hybrid approach combining a classical proximal bundle method with the proposed strong bundle method was shown to be competitive with the former method, yet providing the interesting properties of the latter one.

Acknowledgment

J.Y. Bello-Cruz was partially supported by CNPq, by projects UNIVERSAL "Otimização Contínua: Teoria e Algoritmos", PROCAD-nf - UFG/UnB/IMPA and CAPES-MES-CUBA 226/2012 "Modelos de Otimização e Aplicações".

References

1. Ya. I. Alber, A. N. Iusem, and M. V. Solodov, *On the projected subgradient method for nonsmooth convex optimization in a hilbert space*, Mathematical Programming, 81 (1998), pp. 23–35.
2. J. Batut and A. Renaud, *Daily scheduling with transmission constraints: A new class of algorithms*, IEEE Transactions on Power Systems, 7 (1992), pp. 982–989.
3. H.H. Bauschke and P.L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books in mathematics, Springer Berlin, 2011.
4. A. Beck and S. Sabach, *A first order method for finding minimal norm-like solutions of convex optimization problems*, Mathematical Programming, 7 (1992), pp. 1–22.
5. J. Y. Bello-Cruz and W. de Oliveria, *Level bundle-like algorithms for convex optimization*, To appear in Journal of Global Optimization, (2013), pp. 1–23.
6. J. Y. Bello-Cruz and A. N. Iusem, *A strongly convergent method for nonsmooth convex minimization in hilbert spaces*, Numerical Functional Analysis and Optimization, 32 (2011), pp. 1009–1018.
7. D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, *Adaptive robust optimization for the security constrained unit commitment problem*, IEEE Transactions on Power Systems, 28 (2013), pp. 52–63.

8. R. S. Burachik and A. N. Iusem, *Set-Valued Mappings and Enlargements of Monotone Operators*, vol. 8 of Springer Optimization and Its Applications, Springer Berlin, 2008.
9. R. Correa and C. Lemaréchal, *Convergence of some algorithms for convex minimization*, Mathematical Programming, 62 (1993), pp. 261–275.
10. W. de Oliveira, C. Sagastizábal, and C. Lemaréchal, *Proximal bundle methods in depth: a unified analysis for inexact oracles*, Optimization Oline Report 3792, (2013).
11. L. Dubost, R. Gonzalez, and C. Lemaréchal, *A primal-proximal heuristic applied to french unitcommitment problem*, Mathematical programming, 104 (2005), pp. 129–151.
12. S. Feltenmark and K.C. Kiwiel, *Dual applications of proximal bundle methods, including lagrangian relaxation of nonconvex problems*, SIAM Journal on Optimization, 10 (2000), pp. 697–721.
13. E.C. Finardi and E.L. Da Silva, *Solving the hydro unit commitment problem via dual decomposition and sequential quadratic programming*, Power Systems, IEEE Transactions on, 21 (2006), pp. 835–844.
14. E.C. Finardi, E.L. Da Silva, and C.A. Sagastizábal, *Solving the unit commitment problem of hydropower plants via Lagrangian relaxation and sequential quadratic programming*, Computational & Applied Mathematics, 24 (2005), pp. 317–341.
15. R. Gonzalez, M.P. Bongrain, and A. Renaud, *Unit commitment handling transmission constraints with an interior point method*, in 13th PSCC conference, Trondheim, vol. 2, 1999, pp. 715–723.
16. C. Kanzow, H. Qi, and L. Qi, *On the minimum norm solution of linear programs*, Journal of Optimization Theory and Applications, 116 (2003), pp. 333–345.
17. J.E. Kelley, *The cutting-plane method for solving convex programs*, Journal of the Society for Industrial and Applied Mathematics, 8 (1960), pp. 703–712.
18. K.C. Kiwiel, *Proximity control in bundle methods for convex nondiferentiable minimization*, Math. Programming, 46 (1990), pp. 105–122.
19. N. Langrene, W. van Ackooij, and F. Bréant, *Dynamic constraints for aggregated units: Formulation and application*, IEEE transactions on Power Systems, 26 (2011), pp. 1349–1356.
20. K.D. Le, R.R. Jackups, J. Feinstein, and J.S. Griffith, *Operational aspects of generation cycling*, IEEE Transactions on Power Systems, 5 (1990), pp. 1194–1203.
21. C. Lemaréchal, *Constructing bundle methods for convex optimization*, in Fermat Days 85: Mathematics for Optimization, vol. 129 of North-Holland Mathematics Studies, North-Holland, 1986, pp. 201–240.
22. C. Lemaréchal and C. Sagastizábal, *An approach to variable metric bundle methods*, Lecture Notes in Control and Information Science, 197 (1994), pp. 144–162.
23. R. Nürnberg and W. Römisch, *A two-stage planning model for power scheduling in a hydro-thermal system under uncertainty*, Optimization and Engineering, 3 (2003), pp. 355–378.
24. C. Sagastizábal, *Divide to conquer: Decomposition methods for energy optimization*, Mathematical Programming, 134 (2012), pp. 187–222.
25. M. V. Solodov, *On approximations with finite precision in bundle methods for nonsmooth optimization*, Journal of Optimization Theory and Applications, 119 (2003), pp. 151–165.
26. S. Takriti and J.R. Birge, *Using integer programming to refine lagrangian-based unit commitment solutions*, IEEE Transactions on Power Systems, 15 (2000), pp. 151–156.
27. W. van Ackooij, *Decomposition approaches for block-structured chance-constrained programs with application to hydro-thermal unit-commitment*, Submitted ; Preprint CR-2012-08, (2012).
28. W. van Ackooij, R. Henrion, A. Möller, and R. Zorgati, *Joint chance constrained programming for hydro reservoir management*, to Appear in Optimization and Engineering, (2011).
29. S. Zaourar and J. Malick, *Prices stabilization for inexact unit-commitment problems*, Mathematical Methods of Operations Research, to Appear, (2013), pp. 1–19.