Journal of Convex Analysis
Volume 12 (2005), No. 1, 95–111

A Bundle Interior Proximal Method for Solving Convex Minimization Problems

Nguyen Thi Thu Van
Department of Mathematics and Computer Sciences, Faculty of Natural Sciences, National University of Vietnam, HoChiMinh City, Vietnam

Van Hien Nguyen
Unité d'Optimisation, Département de Mathématique, Facultés Universitaires Notre Dame de la Paix, 5000 Namur, Belgium

Jean-Jacques Strodiot
Unité d'Optimisation, Département de Mathématique, Facultés Universitaires Notre Dame de la Paix, 5000 Namur, Belgium

Dedicated to Dr. Claude Lemaréchal for his 60th anniversary

Received October 20, 2003

In this paper we extend the standard bundle proximal method for finding the minimum of a convex not necessarily differentiable function on the nonnegative orthant. The strategy consists in approximating the objective function by a piecewise linear convex function and using distance–like functions based on second order homogeneous kernels. First we prove the convergence of this new bundle interior proximal method under the same assumptions as for the standard bundle method and then we report some preliminary numerical experiences for a particular distance function.

1. Introduction

In this paper we consider the convex minimization problem

$$(\mathrm{P}) \quad \begin{cases} \text{minimize} \quad f(x) \\ \text{subject to } x \in X, \end{cases} \tag{1}$$

where $f : I\!\!R^p \to I\!\!R \cup \{+\infty\}$ is a closed proper convex function and X is a closed convex subset of $I\!\!R^p$ (see, e.g., [7] and [14] for all the definitions concerning Convex Analysis). When $X = I\!\!R^p$, a very well-known method, called the proximal method, has been introduced by Martinet [12] and Rockafellar [15]. This method generates a sequence $\{x^k\}$ via the scheme

$$x^{k+1} = \arg\min\{f(x) + \frac{1}{2\lambda_k}\|x - x^k\|^2 : x \in I\!\!R^p\}, \tag{2}$$

where $\{\lambda_k\}$ is a sequence of positive numbers. Applied to the dual of a convex mathematical programming problem, this method gives rise to the classical multiplier method. When $X \neq I\!\!R^p$, several authors have proposed to replace the quadratic term $\|x - x^k\|^2$ in (2) by another distance between x and x^k. The aim is that the new unconstrained minimization problem (2) yields automatically a feasible solution. This avoids to solving constrained subproblems. Among the distances recently proposed in the literature, let us mention two classes of distances. The first class is based on Bregman functions (see, e.g., Burachik and Iusem [5] and Kiwiel [9]) and the second class on logarithmic–entropy

ISSN 0944-6532 / $ 2.50

functions (see, e.g., Auslender and Haddou [1], Ben–Tal and Zibulesky [4] and Tseng and Bertsekas [16] and the references therein). When $X = I\!R^p_+ = \{x \in I\!R^p_+ : x \geq 0\}$, a distance-like function corresponding to the second class has the form

$$d_\varphi(x, y) = \sum_{j=1}^{p} y_j \, \varphi\left(\frac{x_j}{y_j}\right), \tag{3}$$

where $x = (x_1, ..., x_p) \in I\!R^p_+$, $y = (y_1, ..., y_p) \in I\!R^p_{++}$ and φ is a strictly convex function. With this distance, the scheme (2) becomes

$$x^{k+1} = \arg\,\min\{f(x) + \frac{1}{\lambda_k} d_\varphi(x, x^k) : x \in I\!R^p_+\}. \tag{4}$$

Moreover, by choosing φ in an appropriate class of functions, all the iterates x^k belong to $I\!R^p_{++} = \{x \in I\!R^p : x > 0\}$. Hence the name "interior proximal method". However, when applied within the dual framework, the corresponding primal sequence can be shown to converge only in an ergodic sense to an optimal solution of the primal problem. The same difficulties are encountered with methods based on Bregman functions (see Kiwiel [9] for more details). To avoid these drawbacks, Auslender, Teboulle and Ben-Tiba [2], proposed to modify the distance–function (3) as follows

$$d_\varphi(x, y) = \sum_{j=1}^{p} {y_j}^2 \varphi\left(\frac{x_j}{y_j}\right). \tag{5}$$

They proved the global convergence of the corresponding scheme (4) - (5) and although they also consider inexact versions of this scheme, their algorithms remain conceptual. In a remark ([2], Remark 2.1), they suggest to extend a bundle type algorithm as developed for the classical proximal algorithm (see, e.g., Correa and Lemaréchal [6]) to the class of interior proximal methods. Motivated by this remark, we propose in this paper a new approximate interior proximal method to make implementable the scheme (4) - (5). We prove the convergence of the new algorithm and we examine its behavior on some test–functions. Let us mention that such an extension has already been done in [10] but for the distance–like functions based on Bregman functions.

The paper is organized as follows: in Section 2, we define the class of functions φ we will use in our approximate interior proximal method and we display its main properties. Then we briefly recall the convergence results obtained in [2] for the interior proximal method. In Section 3 we present our bundle proximal method for solving a constrained convex minimization problem. The nonsmooth convex function is approximated by a piecewise linear convex function and conditions on this approximation are given to ensure the convergence of the method. Finally in the last section, we examine how to implement the method and we report some preliminary numerical experiences using the Matlab environment. For a very comprehensive survey on the standard proximal bundle methods, we refer the reader to the book by Hiriart–Urruty and Lemaréchal ([7]).

2. Interior Proximal Method

Let $f : I\!R^p \to I\!R \cup \{+\infty\}$ be a closed proper convex function such that $I\!R^p_+ \subseteq \text{int dom} f$. The problem is to find the minimum of f on $I\!R^p_+$. For solving this problem, we use the proximal–like scheme defined by (4) where $\{\lambda_k\}$ is a sequence of positive real numbers and $d_\varphi(x,y)$ is the distance–function (5) based on a function φ. In order to define it properly, we first consider a class of functions denoted by Φ. This class Φ contains all the closed, proper and convex functions $\varphi : I\!R \to I\!R \cup \{+\infty\}$ that satisfy the following properties:

(i) $\operatorname{dom} \varphi \subseteq [0, +\infty)$;
(ii) φ is twice continuously differentiable on $\operatorname{int}(\operatorname{dom}\varphi) = (0, +\infty)$;
(iii) φ is strictly convex on its domain;
(iv) $\lim_{t \to 0^+} \varphi'(t) = -\infty$;
(v) $\varphi(1) = \varphi'(1) = 0$ and $\varphi''(1) > 0$.

Let $\varphi \in \Phi$. Then the distance d_φ is defined by

$$d_\varphi(x,y) = \sum_{j=1}^{p} {y_j}^2 \varphi\left(\frac{x_j}{y_j}\right) \quad \forall x, y \in I\!R^p_{++}.$$

It is easy to see that the function d_φ has the following basic properties:

- d_φ is an homogeneous function of order 2, i.e.,
  $d_\varphi(\alpha x, \alpha y) = \alpha^2\, d_\varphi(x,y)$, $\forall \alpha > 0, \quad \forall x, y \in I\!R^p_{++}$
- $d_\varphi(x,y) \geq 0$, $\forall x, y \in I\!R^p_{++}$
- $d_\varphi(x,y) = 0$ if and only if $x = y$.

The function φ being differentiable and convex on $(0, +\infty)$, the function $d_\varphi(\cdot, y)$ is differentiable and convex on $I\!R^p_{++}$ for any $y \in I\!R^p_{++}$. Hence x^k is a minimum of the function $f + d_\varphi(\cdot, x^{k-1})$ if and only if

$$0 \in \partial f(x^k) + \frac{1}{\lambda_k}\Psi(x^k, x^{k-1}),$$

where ∂f denotes the subdifferential of f and

$$\Psi(a,b) = \left(b_1\ \varphi'\left(\frac{a_1}{b_1}\right),\ b_2\ \varphi'\left(\frac{a_2}{b_2}\right), ..., b_p\ \varphi'\left(\frac{a_p}{b_p}\right)\right) \quad \forall a, b \in I\!R^p_{++}. \tag{6}$$

With these definitions, the basic iteration scheme (BIS) introduced by Auslender and al. for finding the minimum of f on $I\!R^p_+$ can be expressed as

Given $\varphi \in \Phi$, $x^0 \in I\!R^p_{++}$, $\varepsilon_k \geq 0, \lambda_k > 0$, generate the sequences $\{x^k\} \subseteq I\!R^p_{++}$ and $\{g^k\}$ satisfying

$$g^k \in \partial_{\varepsilon_k} f(x^k) \quad \textit{and} \quad \lambda_k g^k + \Psi(x^k, x^{k-1}) = 0, \tag{7}$$

where $\partial_{\varepsilon_k} f(x^k)$ denotes the ε_k–subdifferential of f at x^k.

From (7), we have that

$$0 \in \partial_{\epsilon_k} f(x^k) + \frac{1}{\lambda_k}\Psi(x^k, x^{k-1}).$$

This means that x^k is an ϵ_k–minimum of the function $f + d_\varphi(\cdot, x^{k-1})$.

Our aim is to present a bundle version of this algorithm and to study its convergence. In that purpose, we need to introduce a subclass of Φ defined by

$$\Phi_0 = \{h \in \Phi : h''(1)(1 - \frac{1}{t}) \leq h'(t) \leq h''(1)(t-1), \forall t > 0\}, \tag{8}$$

and to consider a specific choice for the functions φ we will use, namely,

$$\varphi(t) := \mu h(t) + \frac{\nu}{2}(t-1)^2, \tag{9}$$

where $\mu > 0$, $\nu > 0$ and $h \in \Phi_0$.
The kernel h is used to enforce the iterates to stay in the interior of the nonnegative orthant while the quadratic term $(t-1)^2$ gives rise to the usual term used in "regularization" (see, e.g., [13]). It is easy to see that the following functions belong to Φ_0:

$$\begin{aligned} h_1(t) &= t \log t - t + 1, \quad && \operatorname{dom} h_1 = [\,0, +\infty); \\ h_2(t) &= -\log t + t - 1, \quad && \operatorname{dom} h_2 = (0, +\infty); \\ h_3(t) &= 2(\sqrt{t} - 1)^2, \quad && \operatorname{dom} h_3 = [0, +\infty). \end{aligned}$$

When the functions φ are defined by (9) with $h \in \Phi_0$ and $\nu \geq \mu h''(1) > 0$, Auslender and al. ([2], Theorem 3.2) proved that the sequence $\{x^k\}$ generated by the BIS algorithm converges to a minimum of f on $\mathbb{R}^p_+$ provided that $\sum \lambda_k = +\infty$ and $\sum \lambda_k \varepsilon_k < +\infty$.

3. Bundle Interior Proximal Method

Let $f : \mathbb{R}^p \to \mathbb{R}$ be a convex function. Since f can be nondifferentiable, we observe that finding x^{k+1} by using (4) is often as difficult as finding the minimum of f over $\mathbb{R}^p_+$. So the strategy to get an implementable algorithm, is to approximate f at iteration k by a simpler convex function in such a way that the resulting problem is easy to solve. The same strategy has been used for the classical proximal method. First we recall the corresponding algorithm for the sake of completeness.

Approximate Proximal Algorithm

Let $\sigma \in (0, 1)$ be a tolerance and $\{\lambda_k\}_{k \in \mathbb{N}}$ be a sequence of positive numbers. Choose a starting point x^0 and set $y^0 = x^0, k = 0$ and $i = 1$.

Step 1. Choose a convex function $\widehat{f^i} : \mathbb{R}^p \to \mathbb{R}$ and solve the problem

$$(\mathrm{SP}_i) \quad \min\{\widehat{f^i}(y) + \frac{1}{2\lambda_k}\|y - x^k\|^2\}$$

to get the unique optimal solution y^i as well as the aggregate subgradient

$$\gamma^i = \frac{1}{\lambda_k}(x^k - y^i) \in \partial \widehat{f^i}(y^i).$$

Step 2. Compute $f(y^i)$. If $f(x^k) - f(y^i) \geq \sigma[f(x^k) - \widehat{f^i}(y^i)]$, then set $x^{k+1} = y^i$ and increase k by 1.

Step 3. Increase i by 1 and go to Step 1.

If we consider (4) instead of (2), this algorithm becomes

Approximate Interior Proximal Method

Let $\sigma \in (0,1)$ be a tolerance and $\{\lambda_k\}_{k\in\mathbb{N}}$ be a sequence of positive numbers. Choose a starting point $x^0 \in \mathbb{R}^p_{++}$ and set $y^0 = x^0, k = 0$ and $i = 1$.

Step 1. Choose a convex function $\widehat{f^i} : \mathbb{R}^p \to \mathbb{R}, \widehat{f^i} \le f$. Solve

$$\begin{cases} \text{minimize } \widehat{f^i}(y) + \frac{1}{\lambda_k} d_\varphi(y, x^k) \\ \text{subject to } y \in \mathbb{R}^p_{++} \end{cases}$$

to get $y^i \in \mathbb{R}^p_{++}$ as well as the aggregate subgradient

$$\gamma^i = -\frac{1}{\lambda_k}\Psi(y^i, x^k) \in \partial\widehat{f^i}(y^i).$$

Step 2. Compute $f(y^i)$. If $f(x^k) - f(y^i) \ge \sigma[f(x^k) - \widehat{f^i}(y^i)]$, then set $x^{k+1} = y^i$ and increase k by 1.

Step 3. Increase i by 1 and go to Step 1.

When the reduction on f is sufficient, i.e., when $f(x^k) - f(y^i) \ge \sigma[f(x^k) - \widehat{f^i}(y^i)]$, we say that a serious step is done; otherwise, the step is called a null step and a new $\widehat{f^{i+1}}$ is chosen to improve the approximation $\widehat{f^i}$ of f. The way to choose the approximate functions $\widehat{f^i}$ is crucial to get an implementable algorithm. So, for instance, if $\widehat{f^i}$ has the following form

$$\widehat{f^i}(y) = \max_{1\le j\le m}\{a_j^T y + b_j\}, \quad \forall y \in \mathbb{R}^p,$$

then the subproblem in Step 1 is equivalent to the linearly constrained problem

$$\begin{cases} \text{minimize } \; v + \frac{1}{\lambda_k} d_\varphi(y, x^k) \\ \text{subject to } \; a_j^T y + b_j \le v, \quad j = 1, \dots, m. \end{cases}$$

We will discuss the way for solving this problem in the next section.

In order to give rules for choosing $\widehat{f^i}$, we need to define the function

$$l^i : \mathbb{R}^p \to \mathbb{R} \quad l^i(y) = \widehat{f^i}(y^i) + \langle \gamma^i, y - y^i \rangle \quad \forall y \in \mathbb{R}^p.$$

This function satisfies $l^i \le \widehat{f^i}$. Indeed, since $\gamma^i \in \partial\widehat{f^i}(y^i)$, we have, for all y, that

$$\widehat{f^i}(y) \ge \widehat{f^i}(y^i) + \langle \gamma^i, y - y^i \rangle = l^i(y).$$

As usual in the bundle methods, we impose the following assumption on the function f:
(A) At each $x \in \mathbb{R}^p$, one subgradient of f at x can be computed (this subgradient is denoted by $s(x)$ in the sequel).
This assumption is realistic because computing the whole subdifferential is often very

expensive or impossible while obtaining one subgradient is often easy. This situation occurs, for instance, if the function f is the dual function associated with a mathematical programming problem.

Now to obtain the convergence of the bundle method, we impose the following conditions on the sequence $\{\widehat{f}^i\}$,

(A1) $\widehat{f}^i \le f$ for all $i = 1, 2, \dots$

(A2) If the i^{th} iteration gives rise to a null–step, then

$$\begin{array}{ll} \text{(i)} & l^i \le \widehat{f}^{i+1}, \\ \text{(ii)} & f(y^i) + \langle s(y^i), \cdot - y^i \rangle \le \widehat{f}^{i+1}, \end{array}$$

where $s(y^i)$ denotes the subgradient of f available at y^i. These conditions have already been used in [6] for the standard proximal method.

Here are some examples of functions $\widehat{f}^i$ satisfying conditions (A1) and (A2). In the first example, we suppose that all the subgradients collected at the previous iterates are kept in mind. So, for $i = 0, 1, \dots,$ we define

$$\widehat{f}^{i+1}(y) = \max\{\, f(y^j) + \langle s(y^j), y - y^j \rangle \mid j = 0, \dots, i \,\}.$$

Conditions (A1) and (A2)(ii) are obviously satisfied by $\widehat{f}^{i+1}$. Moreover, since $l^i \le \widehat{f}^i \le \widehat{f}^{i+1}$, condition (A2)(i) is also satisfied. In the second example, we only keep the function l^i and the latest subgradient $s(y^i)$ to have

$$\widehat{f}^{i+1}(y) = \max\, \{l^i(y), f(y^i) + \langle s(y^i), y - y^i \rangle\}. \tag{10}$$

Conditions (A1) and (A2) are obviously satisfied.

Remark 3.1. Observe that the reduction predicted by the model $\widehat{f}^i$, namely $f(x^k) - \widehat{f}^i(y^i)$ is nonnegative. Indeed, since $f \ge \widehat{f}^i$ and $\gamma^i \in \partial \widehat{f}^i(y^i)$, we have

$$\begin{array}{rcl} f(x^k) - \widehat{f}^i(y^i) & \ge & \widehat{f}^i(x^k) - \widehat{f}^i(y^i) \ge \langle \gamma^i, x^k - y^i \rangle \\ & = & -\frac{1}{\lambda_k} \langle \Psi(y^i, x^k), x^k - y^i \rangle \ge 0. \end{array}$$

To prove the convergence of the bundle interior proximal method, we also need to introduce the following notations

$$\tilde{l}^i(y) = l^i(y) + \lambda_k^{-1} d_\varphi(y, x^k),$$

$$\tilde{f}^i(y) = \widehat{f}^i(y) + \lambda_k^{-1} d_\varphi(y, x^k).$$

Then we have

$$\tilde{f}^i(x^k) = \widehat{f}^i(x^k) \;\text{ and }\; \tilde{l}^i(y^i) = \tilde{f}^i(y^i). \tag{11}$$

Indeed, $d_\varphi(x^k, x^k) = 0$ and

$$\tilde{l}^i(y^i) = l^i(y^i) + \lambda_k^{-1} d_\varphi(y^i, x^k) = \widehat{f}^i(y^i) + \lambda_k^{-1} d_\varphi(y^i, x^k) = \tilde{f}^i(y^i).$$

Lemma 3.2. *There exists $\beta > 0$ such that, for all i,*

$$\tilde{l}^i(y) \geq \tilde{l}^i(y^i) + \frac{\beta}{2\lambda_k}\|y - y^i\|^2.$$

Proof. By definition of $\tilde{l}^i$, we have

$$\begin{aligned} \tilde{l}^i(y) - \tilde{l}^i(y^i) &= l^i(y) + \lambda_k^{-1} d_\varphi(y, x^k) - l^i(y^i) - \lambda_k^{-1} d_\varphi(y^i, x^k) \\ &= \langle \gamma^i, y - y^i \rangle + \lambda_k^{-1} [\, d_\varphi(y, x^k) - d_\varphi(y^i, x^k)\,]. \end{aligned} \tag{12}$$

Since $d_\varphi(x, x^k) = \sum_{i=1}^{p} (x_i^k)^2 \varphi(\frac{x_i}{x_i^k})$ and φ is strongly convex on $\{t \in I\!R \,|\, t > 0\}$, the function d_φ is itself strongly convex on $I\!R^p_{++}$, i.e., there exists $\beta > 0$ such that, for all $y \in I\!R^p_{++}$,

$$d_\varphi(y, x^k) - d_\varphi(y^i, x^k) \geq \langle \Psi(y^i, x^k), y - y^i \rangle + \frac{\beta}{2}\|y - y^i\|^2.$$

Using this inequality in (12) and noting that $\Psi(y^i, x^k) = -\lambda_k \gamma^i$, we obtain

$$\tilde{l}^i(y) - \tilde{l}^i(y^i) \geq \frac{\beta}{2\lambda_k}\|y - y^i\|^2.$$

□

Proposition 3.3. *Suppose that after x^k has been obtained in the bundle interior proximal algorithm, the test of sufficient reduction is suppressed : only null–steps are made. If the sequence $\{\widehat{f}^i\}$ satisfies conditions $(A1)$ and $(A2)$, then*

$$\begin{aligned} &(1) \quad f(y^i) - \widehat{f}^i(y^i) \to 0, \\ &(2) \quad y^i \to y^* = \arg\min_{x>0} \{f(x) + \lambda_k^{-1} d_\varphi(x, x^k)\}. \end{aligned}$$

Proof

(1) We use three steps to prove this part.

(i) $\tilde{l}^i(y^i)$ is convergent and $y^{i+1} - y^i \to 0$.
For $i = 1, \ldots,$ we have

$$\begin{array}{rcll} f(x^k) & \geq & \widehat{f}^{i+1}(x^k) & \text{by (A1)} \\ & = & \tilde{f}^{i+1}(x^k) & \text{by (11)} \\ & \geq & \tilde{f}^{i+1}(y^{i+1}) & \text{by definition of } y^{i+1} \\ & = & \tilde{l}^{i+1}(y^{i+1}) & \text{by (11)} \\ & \geq & \tilde{l}^{i}(y^{i+1}) & \text{by (A2)(i)} \\ & \geq & \tilde{l}^i(y^i) + \frac{\beta}{2\lambda_k}\|y^{i+1} - y^i\|^2 & \text{by Lemma 3.2 with } y = y^{i+1}. \end{array}$$

From these relations, we have for all i,

$$\tilde{l}^i(y^i) \leq \tilde{l}^{i+1}(y^{i+1}) \text{ and } \tilde{l}^i(y^i) \leq f(x^k).$$

Hence the sequence $\{\tilde{l}^i(y^i)\}$ is convergent in $I\!R$. Moreover, by Lemma 3.2, we have

$$\tilde{l}^{i+1}(y^{i+1}) - \tilde{l}^i(y^i) \geq \frac{\beta}{2\lambda_k}\|y^{i+1} - y^i\|^2 \geq 0.$$

Hence $y^{i+1} - y^i \to 0$.

(ii) The sequence $\{y^i\}$ is bounded.
We have (for y fixed)

$$\begin{array}{rcll} f(y) + \frac{1}{\lambda_k} d_\varphi(y, x^k) & \geq & \widehat{f}^{i+1}(y) + \frac{1}{\lambda_k} d_\varphi(y, x^k) & \text{by (A1)} \\ & = & \tilde{f}^{i+1}(y) & \text{by definition of } \tilde{f}^{i+1} \\ & \geq & \tilde{l}^i(y) & \text{by (A2)} \\ & \geq & \tilde{l}^i(y^i) + \frac{\beta}{2\lambda_k}\|y - y^i\|^2 & \text{by Lemma 3.2.} \end{array}$$

Since the sequence $\{\tilde{l}^i(y^i)\}$ is convergent, it is bounded and thus also the sequences $\{\|y - y^i\|^2\}$ and $\{y^i\}$.

(iii) $f(y^{i+1}) - \widehat{f}^{i+1}(y^{i+1}) \to 0$.
By definition of $s(y^{i+1})$, we have

$$\langle s(y^i), y^{i+1} - y^i\rangle \leq \widehat{f}^{i+1}(y^{i+1}) - f(y^i) \leq f(y^{i+1}) - f(y^i) \leq \langle s(y^{i+1}), y^{i+1} - y^i\rangle.$$

Since the subdifferential ∂f is bounded on bounded subsets of $I\!R^p_{++}$ and the sequence $\{y^i\}$ is bounded, then the sequence $\{s(y^i)\}$ is also bounded. Taking the limit of the opposite sides of the previous inequalities, we obtain

$$\langle s(y^i), y^{i+1} - y^i\rangle \to 0 \quad \text{and} \quad \langle s(y^{i+1}), y^{i+1} - y^i\rangle \to 0,$$

and hence

$$\widehat{f}^{i+1}(y^{i+1}) - f(y^i) \to 0 \quad \text{and} \quad f(y^{i+1}) - f(y^i) \to 0.$$

So

$$f(y^{i+1}) - \widehat{f}^{i+1}(y^{i+1}) = f(y^{i+1}) - f(y^i) + f(y^i) - \widehat{f}^{i+1}(y^{i+1}) \ \to \ 0.$$

(2) We also use three steps to prove this part.

(i) Any limit point $\bar{y}$ of $\{y^i\}$ is such that $\bar{y}_j > 0, \quad \forall j = 1, \ldots, p$.
Let $\{y^i\}_{i\in K}$ be a subsequence of $\{y^i\}$ converging to $\bar{y}$ and suppose, to get a contradiction that $J := \{\, j \mid \bar{y}_j = 0 \,\}$ is nonempty. By definition of γ^i, we can write

$$\gamma^i = -\lambda_k^{-1}\Psi(y^i, x^k) \in \partial\widehat{f}^i(y^i).$$

Then, since $f \geq \widehat{f}^i$, we have

$$\forall y \in I\!R^p \quad f(y) \geq \widehat{f}^i(y) \geq \widehat{f}^i(y^i) + \langle\gamma^i, y - y^i\rangle,$$

i.e.,

$$\forall y \in I\!R^p \quad f(y) \geq \widehat{f}^i(y^i) - \lambda_k^{-1}\langle\Psi(y^i, x^k), y - y^i\rangle. \tag{13}$$

On the other hand

$$\Psi(y^i, x^k) = \left(x_1^k \varphi'\left(\frac{y_1^i}{x_1^k}\right), ..., x_j^k \varphi'\left(\frac{y_j^i}{x_j^i}\right), \dots, x_p^k \varphi'\left(\frac{y_p^i}{x_p^k}\right)\right).$$

Then it is easy to see that for all $j \in J$, we have

$$\frac{y_j^i}{x_j^k} \ \rightarrow \ 0^+ \quad \text{and} \quad \varphi'\left(\frac{y_j^i}{x_j^k}\right) \ \rightarrow \ -\infty \text{ (by property (iv) of } \varphi) \tag{14}$$

while, for $j \notin J$, we obtain

$$\varphi'\left(\frac{y_j^i}{x_j^k}\right) \quad \rightarrow \quad \varphi'\left(\frac{\bar{y}_j^i}{x_j^k}\right) \in I\!R. \tag{15}$$

Choose $y = (1, \dots, 1)^T$. Then, for all i, we deduce from (13) that

$$f(y) \geq \widehat{f}^i(y^i) - f(y^i) + f(y^i) - \lambda_k^{-1} \sum_{j \in J} x_j^k \varphi'\left(\frac{y_j^i}{x_j^k}\right)(1 - y_j^i) - \lambda_k^{-1} \sum_{j \notin J} x_j^k \varphi'\left(\frac{y_j^i}{x_j^k}\right)(1 - y_j^i).$$

Taking the limit as $i \rightarrow +\infty$, using part (1), the continuity of f, (14) and (15), we obtain

$$f(y) \geq +\infty.$$

This is impossible, so J is empty.

(ii) Any limit point $\bar{y}$ of $\{y^i\}$ is a solution of

$$\begin{cases} \text{minimize} \;\; f(x) + \lambda_k^{-1} d_\varphi(x, x^k), \\ \text{subject to } x \;\; > 0. \end{cases}$$

Let $y^i \rightarrow \bar{y}$, $i \in K \subseteq I\!N$. By part 2(i), $\bar{y}_j > 0$ for all $j = 1, \dots, p$. To obtain that $\bar{y}$ is a minimum of $f + \lambda_k^{-1} d_\varphi(\cdot, x^k)$, we have to prove that $0 \in \partial f(\bar{y}) + \lambda_k^{-1}\Psi(\bar{y}, x^k)$, i.e.,

$$\forall y \in I\!R^p \quad f(y) \geq f(\bar{y}) - \lambda_k^{-1}\langle \Psi(\bar{y}, x^k), y - \bar{y}\rangle. \tag{16}$$

Let then $y \in I\!R^p$. By definition of $\gamma^i \in \partial \widehat{f}^i(y^i)$ and since $f \geq \widehat{f}^i$, we have

$$f(y) \geq \widehat{f}^i(y) \geq \widehat{f}^i(y^i) - \lambda_k^{-1}\langle \Psi(y^i, x^k), y - y^i\rangle,$$

i.e.,

$$f(y) \geq \widehat{f}^i(y^i) - f(y^i) + f(y^i) - \lambda_k^{-1}\langle \Psi(y^i, x^k), y - y^i\rangle.$$

Taking the limit as $i \rightarrow +\infty$, using part (1)(iii), the continuity of f and $\Psi(\cdot, x^k)$, we obtain the required inequality (16).

(iii) $y^i \rightarrow y^* = \arg\min\{f(x) + \lambda_k^{-1} d_\varphi(x, x^k)\}$ when $i \ \rightarrow \ +\infty$.

By part 2(ii), any limit point of $\{y^i\}$ is a solution of the problem

$$\begin{cases} \text{minimize} \quad f(x) + \lambda_k^{-1} d_\varphi(x, x^k) \\ \text{subject to} \quad x > 0. \end{cases}$$

However, this problem has exactly one solution because $d_\varphi(\cdot, x^k)$ is strongly convex. So, all the limit points of $\{y^i\}$ coincide and thus the whole sequence $\{y^i\}$ converges to y^*. □

Now we can apply these results to prove the convergence of the bundle interior proximal method. But first we need a lemma.

Lemma 3.4. *If $x^k = \arg\min\{f(x) + \lambda_k^{-1} d_\varphi(x, x^k)/x > 0\}$ then x^k is a minimum of f on $\mathbb{R}_+^p$.*

Proof. By optimality of x^k, we have

$$0 \in \partial f(x^k) + \lambda_k^{-1} \Psi(x^k, x^k).$$

Since $\Psi(x^k, x^k) = 0$ by definition of φ, we obtain $0 \in \partial f(x^k)$ and, since $x^k > 0$, x^k is a minimum of f over $\mathbb{R}_+^p$. □

Theorem 3.5. *Let $\varphi(t) = \mu h(t) + (\nu/2)(t-1)^2$, with $h \in \Phi_0$, $\mu > 0$ and $\nu \geq \mu h''(1) > 0$. Then in the bundle interior proximal algorithm, there are two possibilities*
(1) The index k remains fixed, i.e., only null steps are made from x^k. In this case, x^k is a minimum of f on $\mathbb{R}_+^p$.
(2) The index $k \to +\infty$. Then

- $\displaystyle\sum_{k=1}^{+\infty} \lambda_k = +\infty \Longrightarrow f(x^k) \to \bar{f} := \inf_{x \in \mathbb{R}_+^p} f(x).$
- *If, in addition, $\{\lambda_k\}$ is bounded, then $x^k \to x^*$, minimum of f (if there exists some minimum).*

Proof.
(1) Let i_k be the iteration index that has produced x^k. Since only null–steps are made from x^k, we have

$$\forall i > i_k \quad f(x^k) - f(y^i) < \sigma[f(x^k) - \widehat{f}^i(y^i)]. \tag{17}$$

By Proposition 3.3, we have $y^i \to y^* \equiv \arg\min\{f(x) + \lambda_k^{-1} d_\varphi(x, x^k)\}$ and $f(y^i) - \widehat{f}^i(y^i) \to 0$. Taking the limit in (17), we obtain

$$f(x^k) - f(y^*) \leq \sigma[f(x^k) - f(y^*)],$$

because $\widehat{f}^i(y^i) = \widehat{f}^i(y^i) - f(y^i) + f(y^i) \to f(y^*)$ and f is continuous. Hence

$$(1 - \sigma)[f(x^k) - f(y^*)] \leq 0.$$

Since $1 - \sigma > 0$, we have $f(x^k) \leq f(y^*)$, or again by definition of $d_\varphi(., x^k)$,

$$f(x^k) + \lambda_k^{-1} d_\varphi(x^k, x^k) = f(x^k) \leq f(y^*) \leq f(y^*) + \lambda_k^{-1} d_\varphi(y^*, x^k).$$

Since the solution y^* is unique, we deduce that $x^k = y^*$ and by Lemma 3.4, x^k is a minimum of f on $\mathbb{R}_+^p$.

(2) Denote by $i(k)$ the iteration index where x^k is updated. Then we have $y^{i(k)} = x^{k+1}$. Let us also define $\gamma^k \equiv \gamma^{i(k)} \in \partial \widehat{f}^{i(k)}(x^{k+1})$. We know that

$$\gamma^k = -\lambda_k^{-1} \Psi(x^{k+1}, x^k).$$

With these notations, we prove the following assertions.

a. $\{f(x^k)\}$ is nonincreasing.
Since

$$f(x^k) - f(x^{k+1}) \geq \sigma[f(x^k) - \widehat{f}^{\imath(k)}(x^{k+1})] \tag{18}$$

and since, by Remark 3.1, the reduction $f(x^k) - \widehat{f}^{\imath(k)}(x^{k+1})$ predicted by the model is nonnegative, it follows that $\{f(x^k)\}$ is nonincreasing. In the sequel, we suppose that $\{f(x^k)\}$ is bounded from below (otherwise $f(x^k) \to -\infty$ and the proof is finished).

b. $\gamma^k \in \partial_{\varepsilon_k} f(x^k)$ with

$$\varepsilon_k = f(x^k) - \widehat{f}^{\imath(k)}(x^{k+1}) + \lambda_k^{-1}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle.$$

By definition of γ^k, we observe immediately that

$$\varepsilon_k = f(x^k) - \widehat{f}^{\imath(k)}(x^{k+1}) - \langle \gamma^k, x^k - x^{k+1}\rangle.$$

Moreover, since $f \geq \widehat{f}^{\imath(k)}$ and $\gamma^k \in \partial \widehat{f}^{\imath(k)}(x^{k+1})$, we have for all y, that

$$f(y) \geq \widehat{f}^{\imath(k)}(y) \geq \widehat{f}^{\imath(k)}(x^{k+1}) + \langle \gamma^k, y - x^{k+1}\rangle. \tag{19}$$

In particular, for $y = x^k$, we obtain that $\varepsilon_k \geq 0$. Now, from (19), we also have for all y, that

$$f(y) \geq f(x^k) + \widehat{f}^{\imath(k)}(x^{k+1}) - f(x^k) + \langle \gamma^k, y - x^k\rangle + \langle \gamma^k, x^k - x^{k+1}\rangle,$$

i.e., $\gamma^k \in \partial_{\varepsilon_k} f(x^k)$.

c. $\sum_{k=1}^{+\infty}\{\varepsilon_k - \lambda_k^{-1}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle\} < +\infty$.
By (18) we have

$$\begin{aligned} \varepsilon_k &= f(x^k) - \widehat{f}^{\imath(k)}(x^{k+1}) + \lambda_k^{-1}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle \\ &\leq \sigma^{-1}[f(x^k) - f(x^{k+1})] + \lambda_k^{-1}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle. \end{aligned}$$

Thus

$$\begin{aligned} \sum_{k=1}^{n}\{\varepsilon_k - \lambda_k^{-1} < \Psi(x^{k+1}, x^k), x^k - x^{k+1} >\} &\leq \sigma^{-1}\sum_{k=1}^{n}[f(x^k) - f(x^{k+1})] \\ &= \sigma^{-1}[f(x^1) - f(x^{n+1})]. \end{aligned}$$

Since f is bounded from below, then

$$\sum_{k=1}^{+\infty}\{\varepsilon_k - \lambda_k^{-1}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle\} < +\infty.$$

d. $f(x^k) \;\to\; \bar{f} = \inf\{f(x) \,|\, x \geq 0\}$.
Since the sequence $\{f(x^k)\}$ is nonincreasing, it converges to some $\bar{f}$. Suppose now, to get

a contradiction, that $\bar{f} > f^* := \inf_{x\geq 0} f(x)$. Then there exist $y \in \mathbb{R}^p_+$ and $\delta > 0$ such that, for all k, $f(y) + \delta < f(x^k)$. Since

$$\varepsilon_k - \lambda_k^{-1}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle \to \ 0,$$

there exists k_0 such that, for $k \geq k_0$,

$$\varepsilon_k - \lambda_k^{-1}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle < \frac{\delta}{2}. \tag{20}$$

Using Lemma 3.4 of [2] with $a = x^k, b = x^{k+1}$ and $c = y$, we obtain

$$\begin{aligned} \|y - x^{k+1}\|^2 - \|y - x^k\|^2 &\leq -\theta\langle y - x^{k+1}, \Psi(x^{k+1}, x^k)\rangle \\ = -\,\theta\langle y - x^k, \Psi(x^{k+1}, x^k)\rangle &- \theta\langle x^k - x^{k+1}, \Psi(x^{k+1}, x^k)\rangle, \end{aligned} \tag{21}$$

where $\theta = [(\nu + \mu h''(1))/2]^{-1}$.
By (20), we have immediately

$$-\theta\langle x^k - x^{k+1}, \Psi(x^{k+1}, x^k)\rangle < \lambda_k\theta\left(\frac{\delta}{2} - \varepsilon_k\right). \tag{22}$$

On the other hand, by definition of $\gamma^k = -\lambda_k^{-1}\Psi(x^{k+1}, x^k) \in \partial_{\varepsilon_k} f(x^k)$, and by part b., we have successively

$$-\theta\langle y - x^k, \Psi(x^{k+1}, x^k)\rangle = \lambda_k\theta\langle \gamma^k, y - x^k\rangle \tag{23}$$

and

$$f(x^k) - \delta > f(y) \geq f(x^k) + \langle \gamma^k, y - x^k\rangle - \varepsilon_k. \tag{24}$$

Combining (23) and (24) yields

$$-\theta\langle y - x^k, \Psi(x^{k+1}, x^k)\rangle < \lambda_k\theta[-\delta + \epsilon_k]. \tag{25}$$

Finally we obtain, from (21), (22) and (25), that

$$\|y - x^{k+1}\|^2 \leq \|y - x^k\|^2 + \lambda_k\theta[\frac{\delta}{2} - \varepsilon_k - \delta + \varepsilon_k] = \|y - x^k\|^2 - \lambda_k\theta\frac{\delta}{2}.$$

Summing up, we obtain, for all $k > k_0$,

$$0 \leq \|x^k - y\|^2 \leq \|x^{k_0} - y\|^2 - \frac{\delta\theta}{2}\sum_{k=k_0}^{k-1}\lambda_k.$$

Taking the limit as $k \to +\infty$, we have $\sum_{k=k_0}^{+\infty}\lambda_k \leq 2\theta^{-1}\delta^{-1}\|x^{k_0} - y\|^2 < +\infty$ which contradicts the assumption $\sum_{k=1}^{+\infty}\lambda_k = +\infty$.
Now suppose that f has a minimum $\bar{x}$ on $\mathbb{R}^p_+$ and that the sequence $\{\lambda_k\}$ is bounded.

e. $\{x^k\}$ is bounded.
Using inequality (21) with $y = \bar{x}$, we obtain

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \theta\langle \bar{x} - x^k, \Psi(x^{k+1}, x^k)\rangle - \theta\langle x^k - x^{k+1}, \Psi(x^{k+1}, x^k)\rangle.$$

By definition of $\gamma^k = -\lambda_k^{-1}\Psi(x^{k+1}, x^k) \in \partial_{\varepsilon_k} f(x^k)$, we have

$$\begin{aligned} -\theta\langle \bar{x} - x^k, \Psi(x^{k+1}, x^k)\rangle &= \lambda_k\theta\langle \gamma^k, \bar{x} - x^k\rangle \\ &\leq \lambda_k\theta[f(\bar{x}) - f(x^k) + \varepsilon_k] \leq \lambda_k\theta\varepsilon_k. \end{aligned}$$

So

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 + \lambda_k\theta[\epsilon_k - \lambda_k^{-1}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle]. \tag{26}$$

Since $\{\lambda_k\}$ is bounded (by assumption) and since, by part c., we have

$$\sum_{k=1}^{+\infty} \lambda_k\theta[\varepsilon_k - \lambda_k^{-1}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle] < +\infty.$$

Using Lemma 3.1 of [2], we deduce that the sequence $\{\|x^k - x\|\}_k$ is convergent. Thus $\{x^k\}$ is bounded.

f. Any limit point x^* of $\{x^k\}$ is a minimum of f on $\mathbb{R}^p_+$ and $x^k \to x^*$.
Let $x^{n_k} \to x^*$. By continuity of f, we have $f(x^{n_k}) \to f(x^*)$. By part d., $f(x^k) \to \bar{f} = \inf_{x\geq 0} f(x)$. So $f(x^{n_k}) \to \bar{f}$ and, by the uniqueness of the limit, $f(x^*) = \bar{f}$. Since $x^* \in \mathbb{R}^p_+$, then x^* is a minimum of f on $\mathbb{R}^p_+$. Using (26) with x^* instead of $\bar{x}$, we obtain

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 + \delta_k,$$

where

$$\delta_k = \lambda_k\theta[\varepsilon_k - \lambda_k^{-1}\langle \Psi(x^{k+1}, x^k), x^k - x^{k+1}\rangle].$$

Since $\sum_{k=1}^{+\infty} \delta_k < +\infty$, it follows from Proposition 1.3 of [6] that the whole sequence $\{x^k\}$ converges to x^*. □

4. Numerical Results

To obtain that the approximate interior proximal algorithm is implementable, it remains to explain how to solve the subproblem

$$\begin{cases} \text{minimize} \;\; \widehat{f}^i(y) + \frac{1}{\lambda_k} d_\varphi(y, x^k) \\ \text{subject to } y \;\in\; \mathbb{R}^p_{++}. \end{cases}$$

Since $\widehat{f}^i(y) = \max\{f(y^j) + \langle s(y^j), y - y^j\rangle \mid j = 0, \ldots, i-1\}$, this problem is equivalent to

$$(SP)_{k,i} \begin{cases} \text{minimize} & v + \frac{1}{\lambda_k} d_\varphi(y, x^k) \\ \text{subject to} & v \geq f(y^j) + \langle s(y^j), y - y^j\rangle \quad j = 0, \ldots, i-1 \\ & y \in \mathbb{R}^p_{++}. \end{cases}$$

Observe that if (y^i, v^i) is a solution of this problem, then

$$v^i = \max_{0 \le j \le i-1} \{f(y^j) + \langle s(y^j), y - y^j \rangle\}$$

so that the stopping criterion for the inner iterations is

$$f(x^k) - f(y^i) \ge \sigma[f(x^k) - v^i].$$

Since $\varphi(t) = \mu h(t) + \frac{\nu}{2}(t-1)^2$, the objective function of $(SP)_{k,i}$ is highly nonlinear and finding the solution of $(SP)_{k,i}$ can be very hard. However, if we observe that

$$d_\varphi(y, x^k) = \sum_{m=1}^{p} (x_m^k)^2 \, \varphi\left(\frac{y_m}{x_m^k}\right),$$

then the objective function is separable and one way of solving such a problem is to solve its dual. Setting $z_m = \frac{y_m}{x_m^k}$ for all $m = 1, \ldots, p$ and $z = (z_m)$, problem $(SP)_{k,i}$ can be expressed as

$$(MSP)_{k,i} \quad \begin{cases} \text{minimize} & v + \displaystyle\sum_{m=1}^{p} \alpha_m \varphi(z_m) \\ \text{subject to} & \langle s^j, z \rangle - v \le b_j \quad j = 0, \ldots, i-1, \end{cases}$$

where $\alpha_m = \lambda_k^{-1}(x_m^k)^2, s_m^j = s(y^j)_m x_m^k$, $m = 1, \ldots, p$ and $b_j = \langle s(y^j), y^j \rangle - f(y^j), j = 0, \ldots, i-1$. Then the Lagrangian function associated with $(MSP)_{k,i}$ is

$$L(v, z, \lambda) = v + \sum_{m=1}^{p} \alpha_m \varphi(z_m) + \sum_{j=0}^{i-1} \lambda_j \left[\langle s^j, z \rangle - v - b_j \right]$$

and the dual function

$$\begin{aligned} d(\lambda) &= \inf L(v, z, \lambda) \\ &= \begin{cases} \inf v + \displaystyle\sum_{m=1}^{p} \alpha_m \varphi(z_m) + \sum_{j=0}^{i-1} \lambda_j [\langle s^j, z \rangle - v - b_j] & \text{if } \sum \lambda_j = 1, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

So the dual problem is

$$(D) \begin{cases} \text{maximize} & d(\lambda) \\ \text{subject to} & \sum \lambda_j = 1 \quad \lambda_j \ge 0, \; j = 0, \ldots, i-1, \end{cases}$$

where

$$d(\lambda) = \sum_{m=1}^{p} d_m(\lambda) - \sum_{j=0}^{i-1} \lambda_j b_j \quad \text{with} \quad d_m(\lambda) = \inf_{z_m} \{\alpha_m \varphi(z_m) + \sum_{j=0}^{i-1} \lambda_j s_m^j z_m\}.$$

Moreover, each function d_m is differentiable and

$$\nabla d(\lambda) = \left(\sum_{m=1}^{p} s_m^j \tilde{z}_m \right)_{0 \le j \le i-1} - b,$$

where, for each m, $\tilde{z}_m = \arg\min_{z_m}\{\alpha_m\varphi(z_m) + \sum_{j=0}^{i-1}\lambda_j s^j_m z_m\}$.

Since (D) is a smooth problem whose objective function is easily evaluated, we can use any classical method for solving it. Let λ^* be the solution of (D). Then the vector $z^* = (z^*_m)$ where, for each m,

$$z^*_m = \arg\min\ \{\alpha^*_m\varphi(z_m) + \sum_{j=0}^{i-1}\lambda^*_j s^j_m z_m\}$$

and the scalar

$$v^* = \langle\sum_{j=0}^{i-1}\lambda^*_j s^j, z^*\rangle - \sum_{j=0}^{i-1}\lambda^*_j b_j$$

are solutions of problem $(SP)_{k,i}$. Indeed by the complementarity conditions

$$\sum_{j=0}^{i-1}\lambda^*_j[\langle s^j, z^*\rangle - v^* - b_j] = 0,$$

i.e., since $\sum_{j=0}^{i-1}\lambda^*_j = 1$,

$$v^* = \sum_{j=0}^{i-1}\lambda^*_j v^* = \langle\sum_{j=0}^{i-1}\lambda^*_j s^j, z^*\rangle - \sum_{j=0}^{i-1}\lambda^*_j b_j.$$

The computational results presented here are obtained by using the MATLAB environment. The function f used in the tests, is defined on $I\!R^{10}$ and is the maximum of five quadratic functions:

$$q_j(x) = x^T C^j x - d^{jT}x,\ j = 1, \ldots, 5,$$

where C^j is a 10×10 symmetric matrix defined by

$$C^j_{ik} = \exp(\frac{i}{k})\cos(ik)\sin j,\ i < k \quad \text{and} \quad C^j_{ii} = \frac{i}{n} \mid \sin j \mid + \sum_{i\neq k} \mid C^j_{ik} \mid,$$

and d^j is a vector in $I\!R^{10}$ whose components are $d^j_i = \exp(i/k)\sin(ij)$. This function f is well–known in nonsmooth optimization ([11], Test problem 1: Maxquad, p.151).

The parameters of the method and the function h are chosen as follows: $\nu = 2$, $\mu = 1$, $\lambda_k = 0.1$ for all k and $h(t) = -\log t + t - 1$ for all $t > 0$ so that the function φ becomes

$$\varphi(t) = h(t) + (t-1)^2 = t^2 - t - \log t \quad \forall t > 0.$$

The stopping criterion for the outer iterations is $\|x^{k+1} - x^k\| \leq \varepsilon$ where $\varepsilon = 10^{-3}$. Two values for the parameter σ are used in the numerical experiences, $\sigma = 0.1$ and $\sigma = 0.05$. The results are reported in Table 1 where for each outer iteration (denoted by k), the number of subproblems to be solved is mentioned. As in nonsmooth convex optimization, let us mention that it is possible to limit the size of the bundle, i.e., the number of

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$\sigma = 1$	4	1	3	1	2	3	4	5	5	6	6	10	8	12	9	10	15	12	49
$\sigma = 0.5$	3	1	2	1	1	2	2	3	2	4	5	6	7	8	13	8	10	16	26

Table 4.1: The bundle interior proximal method. Number of inner iterations (for $\sigma = 1$ and $\sigma = 0.5$) for each outer iteration denoted by k.

constraints in the subproblems, by using aggregation ([8]). Although our convergence theorems allow us to use this technique (see (10)), we have not applied it to illustrate the behavior of our method given the small size of the test problems.

From this table, we can observe that the number of subproblems per outer iteration is relatively small. Furthermore, for fixed k, each subproblem is the previous one with an additional linear inequality constraint. So, these problems can be solved very efficiently if the solution of a subproblem is the starting point of the next one. We also observe that the number of subproblems become smaller when the value of the parameter σ is reduced. The smaller is the value of σ, the faster is the stopping criterion satisfied for inner iterations. Contrary to the standard proximal methods, the subproblems are no more quadratic and the way to solve them is crucial for the rate of convergence of the algorithm. The preliminary results are encouraging but more efforts should be devoted to design appropriate numerical methods for solving them. This could be the subject of a future research.

Acknowledgements. Nguyen Thi Thu Van gratefully acknowledges the support of the Optimization and Applied Mathematics Program of the Belgian CIUF-CUD (CUI-UNV02 and CUI-MT03) and of the University of Namur (FUNDP CA 03-1062).

References

[1] A. Auslender, M. Haddou: An interior proximal method for convex linearly constrained problems and its extension to variational inequalities, Math. Programming 71 (1995) 77–100.

[2] A. Auslender, M. Teboulle, S. Ben-Tiba: Interior proximal and multiplier methods based on second order homogeneous kernels, Math. Oper. Res 24(3) (1999) 654–668.

[3] A. Auslender, R. Cominetti, M. Haddou: Asymptotic analysis for penalty and barrier methods in convex and linear programming, Math. Oper. Res 22 (1997) 43–62.

[4] A. Ben-Tal, M. Zibulevsky: Penalty / barrier multiplier methods for convex programming problems, SIAM J. Optim. 7 (1997) 347–366.

[5] R. S. Burachik, A. N. Iusem: A generalized proximal point algorithm for the variational inequality problem in a Hilbert space, SIAM J. Optim. 8 (1998) 197–216.

[6] R. Correa, C. Lemaréchal: Convergence of some algorithms for convex minimization, Math. Programming 62 (1993) 261–275.

[7] J. B. Hiriart-Urruty, C. Lemaréchal: Convex Analysis and Minimization Algorithms, Springer, Berlin (1993).

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

[9] K. C. Kiwiel: Proximal minimization methods with generalized Bregman functions, SIAM J. Control Optim. 35 (1997) 1142–1168.

[10] K. C. Kiwiel: A bundle Bregman proximal method for convex nondifferentiable minimization, Math. Programming 85 (1999) 241–258.

[11] C. Lemaréchal, R. Mifflin: Nonsmooth Optimization, IIASA Proceedings Series, Volume 3, Pergamon Press, Oxford (1978).

[12] B. Martinet: Régularisation d'inéquations variationnelles par approximations successives, Rev. Française Automat. Inform. Rech. Opér. 4 (1970) 154–159.

[13] J. J. Moreau: Proximité et dualité dans un espace Hilbertien, Bull. Soc. Math. France 93 (1965) 273–299.

[14] R. T. Rockafellar: Convex Analysis, Princeton University Press, Princeton (1976).

[15] R. T. Rockafellar: Monotone operators and the proximal point algorithm, SIAM J. Control Optim. 14(5) (1976) 877–898.

[16] P. Tseng, D. P. Bertsekas: On the convergence of the exponential multiplier method for convex programming, Math. Programming 60 (1993) 1–19.