Int. J. Math. And Appl., 6(4)(2018), 51–57
ISSN: 2347-1557
Available Online: http://ijmaa.in/

International Journal of Mathematics And its Applications

Diagonal Bundle Methods for Convex Minimization : Theoritical Aspect

M. Amin Bahraoui[1,*]

1 Département de Mathématique, Faculty of Sciences and Techniques of Tangier, B.P.416, Marocco.

Abstract: The bundle method can be seen as an approximate form of the proximal method. We study in this paper the convergence of general bundle method and we give a diagonal version of it.

Keywords: Convex minimization, bundle method, proximal method, Mosco-convergence, diagonalization.

Accepted on: 22.10.2018

1. Introduction

Bundle methods ([1, 5, 9, 10, 13]) are iterative methods for non smooth convex minimization in a Hilbert space. A crucial assumption in their convergence analysis is that the convex function to be minimized be everywhere finite. Constrained problems can be reduced to that situation by penalty techniques, but in general only an approximation of the given problem can be solved. Therefore it is natural to consider a diagonal version of such a method, mixing the basic method with a sequence of approximations everywhere finite, changing the approximation at each iteration. The aim of the present work is to analyses the convergence of such a diagonal process under assumptions general enough to include the cases of external and exponential penalty approximations. The approach is based upon the unified framework of bundle methods given in [1] which actually appears as an approximate form of the proximal method of Martinet-Rockafellar. Section 2 is devoted to the diagonal version of this form. In section 3, we present the general algorithm of [1] for the approximation of a proximal point which the general bundle method is based upon. This method is presented in section 4 as a special instance of the method of section 2.

2. A Diagonal Approximate Version of the Prox Method

In this section, we improve a result of [6] (section 3). Let X be a real Hilbert space equipped with the inner product $\langle\cdot,\cdot\rangle$ and the associated norm $\|\cdot\|$. For $n = 1, 2, \cdots$, let $f^n \in \Gamma_0(X)$ set of proper closed convex functions on X, $\lambda_n > 0$, $\epsilon_n \geq 0$ and $x_0 \in X$. The sequence $\{x_n\}$ in X is defined recursively by

$$x_n \in (I + \lambda_n \partial_{\epsilon_n} f^n)^{-1} x_{n-1}, \quad n = 1, 2, \cdots \tag{1}$$

* *E-mail: mbahraoui@gmail.com*

which is equivalent to

$$\frac{x_{n-1} - x_n}{\lambda_n} \in \partial_{\epsilon_n} f^n(x_n), \quad n = 1, 2, \cdots \tag{2}$$

Remark 2.1. *Such a sequence $\{x_n\}$ is (not uniquely) well defined, for example take $x_n := prox_{\lambda_n f^n} x_{n-1}$, and $f^n(x_n)$ is finite. The main object of the following sections is to show how to build up x_n from x_{n-1} according to (1), (2) in finitely many steps when ϵ_n is positive.*

The two following lemmas will be of importance for the following.

Lemma 2.2. *for all a, b, x in X, we have $2\langle a - b, x - b\rangle = \|x - b\|^2 - \|x - a\|^2 + \|b - a\|^2$.*

Lemma 2.3. *Let $\{\lambda_n\}$ be a sequence of positive reals and $\{\alpha_n\}$ a sequence of reals. Let us set*

$$t_n := \sum_{k=1}^{n} \lambda_k \quad \text{and} \quad b_n := \frac{\sum_{k=1}^{n} \lambda_k \alpha_k}{t_n}.$$

If $\lim_{n\to+\infty} t_n = +\infty$, then

(i). $\liminf_{n\to+\infty} \alpha_n \le \liminf_{n\to+\infty} b_n \le \limsup_{n\to+\infty} b_n \le \limsup_{n\to+\infty} \alpha_n$.

(ii). if $\alpha := \lim_{n\to+\infty} \alpha_n$ exists, then $b_n \to \alpha$ (Silverman-Toeplitz theorem [2]).

Proof. (i). It is enough to prove the first inequality. If $\liminf_{n\to+\infty} \alpha_n = -\infty$ we are done. Otherwise, for all integers $n > p > 1$,

$$\frac{\sum_{k=p}^{n} \lambda_k}{t_n} \inf_{k\ge p} \alpha_k \le \frac{\sum_{k=p}^{n} \lambda_k \alpha_k}{t_n} = b_n - \frac{\sum_{k=1}^{p-1} \lambda_k \alpha_k}{t_n}.$$

Passing to the limit as $n \to +\infty$, we get $\forall\, p \ge 1$, $\inf_{k\ge p} \alpha_k \le \liminf_{n\to+\infty} b_n$. Then let $p \to +\infty$.

(ii). is a direct consequence of (i). □

Theorem 2.4. *Let us assume there exists a proper function g on X such that*

(i). $\forall n$, $f^n \le g$ or $\forall n$, $dom f^n \supset dom\ g$ and $f^n \overset{p.w.}{\to} g$.

(ii). $\inf_X g \le \liminf_{n\to+\infty} (\inf_X f^n)$.

(iii). $t_n := \sum_{k=1}^{n} \lambda_k \longrightarrow +\infty$.

(iv). $\lim_{n\to+\infty} \epsilon_n = 0$.

Then

$$\frac{\liminf_{n\to+\infty} f^n(x_n) = \lim_{n\to+\infty} \sum_{k=1}^{n} \lambda_k f^k(x_k)}{t_n = \inf_X g.}$$

Proof. We follow here the technique of [3]. We have, from (2), $\forall x \in X$, $\forall\, k \in I\!N$,

$$\lambda_k f^k(x) \ge \lambda_k f^k(x_k) + \langle x_{k-1} - x_k, x - x_k\rangle - \lambda_k \epsilon_k.$$

Thanks to Lemma 2.1,

$$2\lambda_k f^k(x) \ge 2\lambda_k f^k(x_k) + \|x - x_k\|^2 - \|x - x_{k-1}\|^2 + \|x_{k-1} - x_k\|^2 - 2\lambda_k \epsilon_k.$$

Summing from $k=1$ to n,

$$\sum_{k=1}^{n}\lambda_k f^k(x) \geq \sum_{k=1}^{n}\lambda_k f^k(x_k) - \frac{1}{2}\|x-x_0\|^2 - \sum_{k=1}^{n}\lambda_k\epsilon_k.$$

Dividing by t_n,

$$\frac{\sum_{k=1}^{n}\lambda_k f^k(x_k)}{t_n} \leq \frac{\sum_{k=1}^{n}\lambda_k f^k(x)}{t_n} + \frac{\|x-x_0\|^2}{2t_n} + \frac{\sum_{k=1}^{n}\lambda_k\epsilon_k}{t_n}.$$

From assumptions and Lemma 2.2, we have, $\forall x \in domg$

$$\begin{aligned}
\inf_X g &\leq \liminf_{n\to+\infty} f^n(x_n)\\
&\leq \liminf_{n\to+\infty} \frac{\sum_{k=1}^{n}\lambda_k f^k(x_k)}{t_n} \leq \limsup_{n\to+\infty} \frac{\sum_{k=1}^{n}\lambda_k f^k(x_k)}{t_n}\\
&\leq \limsup_{n\to+\infty}\Big[\frac{\sum_{k=1}^{n}\lambda_k f^k(x)}{t_n} + \frac{\|x-x_0\|^2}{2t_n} + \frac{\sum_{k=1}^{n}\lambda_k\epsilon_k}{t_n}\Big]\\
&\leq g(x).
\end{aligned}$$

□

Corollary 2.5. *Under assumptions (iii), (iv) of Theorem 2.1, if we assume*

(i). $\{f^n\}$ *is increasing,* $f := \sup_n f^n$ *and* $\lim_{n\to+\infty}(\inf_X f^n) = \inf_X f$ *or*

(ii). $\{f^n\}$ *is decreasing,* $\forall n \ domf^n = domf$ *and* $f = cl(\inf_n f^n)$,

then

$$\liminf_{n\to+\infty} f^n(x_n) = \inf_X f.$$

Remark 2.6. *(ii) of Corollary 2.1 has already been obtained in [6] (Corollary 3.1). If* $\sum_{n=1}^{+\infty}\epsilon_n < +\infty$, *then* $\liminf$ *can be replaced by* $\lim$. *Moreover* $\lim_{n\to+\infty} f(x_n) = \inf_X f$.

Theorem 2.7. *Let us assume*

(i). $Argminf \neq \emptyset$.

(ii). $\forall\ x \in Argminf, \exists\ \theta_n(x) \geq 0, \sum_{n=1}^{+\infty}\lambda_n\theta_n(x) < +\infty,$ s.t. $f^n(x) \leq \inf_X f^n + \theta_n$.

(iii). $\sum_{n=0}^{+\infty}\lambda_n\epsilon_n < +\infty$.

(iv). $\{f^n\}$ *is decreasing,* $\forall n \ domf^n = domf, f = cl(\inf_n f^n), \sum_{n=1}^{+\infty}\lambda_n = +\infty, \sum_{n=1}^{+\infty}\epsilon_n < +\infty$, *or* $0 < \underline{\lambda} \leq \lambda_n$ *and* f^n *Mosco converges to* f.

Then $f^n(x_n) \to \inf_X f$ *and* $\{x_n\}$ *weakly converges to some* $\overline{x} \in Argminf$.

Proof. We have (cf. proof of Theorem 2.1), for all $x \in X$,

$$f^n(x) \geq f^n(x_n) + \frac{1}{2\lambda_n}(\|x-x_n\|^2 - \|x-x_{n-1}\|^2 + \|x_{n-1}-x_n\|^2) - \epsilon_n.$$

Then, $\forall x \in Argminf$,

$$\|x_n - x\|^2 \leq \|x_{n-1}-x\|^2 - \|x_{n-1}-x_n\|^2 + 2\lambda_n(\theta_n + \epsilon_n).$$

As $\sum_{n=1}^{+\infty} \lambda_n(\theta_n + \epsilon_n) < +\infty$, then $\|x_{n-1} - x_n\| \to 0$, $\lim_{n\to+\infty} \|x_n - x\|$ exists and the sequence $\{x_n\}$ is bounded. Let $\{x_{n_k}\}$ be a subsequence which weakly converges to some $\overline{x}$. In the first case of (iv), that $\overline{x}$ belongs to $Argminf$ is a consequence of Remark 2.2 and the weak lower semi-continuity of f. Otherwise, from (2),

$$\forall \;\; x \in X, \;\; f^{n_k}(x) \geq f^{n_k}(x_{n_k}) + \langle \frac{x_{n_k-1} - x_{n_k}}{\lambda_{n_k}}, x - x_{n_k}\rangle - \epsilon_{n_k}.$$

As $0 < \underline{\lambda} \leq \lambda_n$ then $\frac{x_{n-1} - x_n}{\lambda_n} \xrightarrow{s} 0$. Therefore thanks to Mosco convergence we get $\overline{x} \in Argminf$ and $f^n(x_n) \to \inf_X f$. Finally, the uniqueness of weak accumulation point is standard [8, 12]. □

3. Approximation of $\text{prox}_{\lambda f} x$

(From [1]). Let $f \in \Gamma_0(X)$, $\lambda > 0$ and $x \in X$.

Choose φ^0 in $H_0(f) := \{\varphi \in \Gamma_0(X); \varphi \leq f\}$

For $k := 0, 1, 2, \cdots$

compute $y_k := prox_{\lambda\varphi^k} x$,

compute $\gamma_k := \frac{x - y_k}{\lambda} (\in \partial\varphi^k(y_k))$,

choose $g_k \in \partial f(y_k)$,

choose $\varphi^{k+1} \in H_{k+1}(f) := \{\varphi \in \Gamma_0(X) \;\; \text{satisfying } (1), (2), (3)\}$

(1). $\varphi \leq f$.

(2). $l_k(\cdot) := \varphi^k(y_k) + \langle \gamma_k, \cdot - y_k\rangle \leq \varphi(\cdot)$.

(3). $f(y_k) + \langle g_k, \cdot - y_k\rangle \leq \varphi(\cdot)$.

Example 3.1 (from [1]). $\varphi^0 := f(x) + \langle g_x, . - x\rangle$ *where* $g_x \in \partial f(x)$. *Therefore* $y_0 = x - \lambda g_x$.

(a). $\varphi^{k+1} := \max\{\varphi^k, f(y_k) + \langle g_k, \cdot - y_k\rangle\}$.

(b). $\varphi^{k+1} := \max\{l_k, f(y_k) + \langle g_k, \cdot - y_k\rangle\}$.

(c). $\varphi^{k+1} := \max\{l_k, f(y_i) + \langle g_i, \cdot - y_i\rangle, i \in I^k\}$.

where $I^k \subset \{1, \cdots, k\}$ *and containing* k. *All these examples are based upon the cutting plane idea and lead to the very bundle methods ([1]).*

In the following we give a more direct proof of Proposition 4.3 of [1]. Let us note :

$$\tilde{\varphi}^k(\cdot) := \varphi^k(\cdot) + \frac{1}{2\lambda}\| \cdot - x\|^2 \tag{3}$$

Proposition 3.2. *If* f *is Lipschitz-continuous on bounded sets, for instance* $domf = X$ *and* X *is finite dimensional, then*

(i). $0 \leq f(y_k) - \varphi^k(y_k) =: \alpha_k \to 0$.

(ii). $y_k \xrightarrow{s} prox_{\lambda f} x$.

Proof. (i). We have

$$\begin{aligned} f(x) &\geq \varphi^{k+1}(x) = \tilde{\varphi}^{k+1}(x) \\ &\geq \tilde{\varphi}^{k+1}(y_{k+1}) \\ &\geq \tilde{\varphi}^{k}(y_k) + \frac{1}{2\lambda}\|y_{k+1} - y_k\|^2. \end{aligned}$$

$\{\tilde{\varphi}^k(y_k)\}$ is then an increasing sequence and, as x is fixed, $\tilde{\varphi}^k(y_k)$ is bounded above. Therefore $\{\tilde{\varphi}^k(y_k)\}$ converges in $I\!R$ and

$$\|y_{k+1} - y_k\| \to 0. \tag{4}$$

Moreover, for fixed y, we have

$$\tilde{\varphi}^k(y_k) + \frac{1}{2\lambda}\|y - y_k\|^2 - \frac{1}{2\lambda}\|y - x\|^2 = l_k(y) \leq \varphi^{k+1}(y) \leq f(y).$$

$\{y_k\}$ is then bounded and thanks to (1) and (3),

$$f(y_{k+1}) - f(y_k) \geq \varphi^{k+1}(y_{k+1}) - f(y_k) \geq \langle g_k, y_{k+1} - y_k\rangle.$$

Therefore, from (4) and Lipschitz-continuity of f on bounded sets, we get $\varphi^{k+1}(y_{k+1}) - f(y_k) \to 0$. But $0 \leq f(y_k) - \varphi^k(y_k) = f(y_k) - f(y_{k-1}) + f(y_{k-1}) - \varphi^k(y_k)$. Using anew the Lipschitz-continuity of f on bounded sets, we get the result.

(ii). We have $\frac{x - y_k}{\lambda} \in \partial\varphi^k(y_k)$ which, thanks to (1), implies $\frac{x - y_k}{\lambda} \in \partial_{\alpha_k} f(y_k)$. Let $\overline{y} := prox_{\lambda f} x$. We have $\frac{x-\overline{y}}{\lambda} \in \partial f(\overline{y})$. Therefore, [4] (Lemma 5.3.1), $0 \leq \|y_k - \overline{y}\|^2 \leq \lambda\alpha_k$. □

4. General Bundle Method

(From [1]). Let f be a closed proper convex function on X and $\{\lambda_n\}$ be a sequence of positive reals. Let $x_0 \in X$ and $m \in]0,1[$.

The sequence $\{x_n\}$ is recursively defined as follows:

$\underline{x_{n-1} \longrightarrow x_n :}$

for all $k := 0, 1, \cdots,$

choose $\varphi^k \in H_k(f)$

compute $y_k := prox_{\lambda_n \varphi^k} x_{n-1}$.

Let k_n be the least integer $k \geq 0$ such that

$(*)\quad f(x_{n-1}) - f(y_k) \geq m[f(x_{n-1}) - \varphi^k(y_k)]\ \ (\geq 0).$

If $(*)$ is never satisfied, $k_n := +\infty$, **stop**.

Otherwise, $x_n := y_{k_n}$ and $n := n + 1$.

Remark 4.1. *If, for all k, we choose $\varphi^k = f$, then $\forall n \in I\!N,\ k_n = n$ and we recover the prox method applied to f:*

$$x_n := prox_{\lambda_n f} x_{n-1},\ n = 1, \cdots.$$

Actually we will show that this method is a special instance of the approximate form of the prox method introduced in section 2 (with $f^n = f,\ \forall\ n$), and therefore, convergence results will follow directly from Theorems 2.1. et 2.2.

Theorem 4.2. *Let us consider the sequence $\{x_n\}$ generated by the general bundle method.*

(i). If there exists n such that $k_n := +\infty$, then if $domf = X$, x_{n-1} minimizes f.

(ii). Otherwise if $\sum_{n=1}^{+\infty} \lambda_n = +\infty$, then $f(x_n) \to \inf f$. If $0 < \underline{\lambda} \leq \lambda_n \leq \overline{\lambda} < +\infty$ and $Argminf \neq \emptyset$, then $x_n \overset{w}{\to} \overline{x} \in Argminf$.

Proof. (i). Is a consequence of Proposition 3.1 taking $\lambda = \lambda_n$ and $x = x_{n-1}$. Indeed, if we assume that from some $n-1$, the test $(*)$ is never more satisfied, we get

$$f(x_{n-1}) - f(y_k) < m(f(x_{n-1}) - \varphi^k(y_k)), \quad \forall k \geq 0$$

which is equivalent to

$$(1-m)[f(x_{n-1}) - f(y_k)] < m[f(y_k) - \varphi^k(y_k)], \quad \forall k \geq 0.$$

Then, passing to the limit when $k \to +\infty$, we have (cf. Proposition 3.1)

$$(1-m)[f(x_{n-1}) - f(prox_{\lambda_n f} x_{n-1})] \leq 0$$

from which we get $f(x_{n-1}) \leq f(prox_{\lambda_n f} x_{n-1})$. Therefore x_{n-1} is a fixed point of $\text{prox}_{\lambda_n f}$ viz. minimizes f.

(ii). As

$$\gamma_{k_n} := \frac{x_{n-1} - x_n}{\lambda_n} \in \partial \varphi^{k_n}(x_n)$$

we have

$$\frac{x_{n-1} - x_n}{\lambda_n} \in \partial_{\alpha_n} f(x_n)$$

where $0 \leq \alpha_n := f(x_n) - \varphi^{k_n}(x_n)$. In words, $\{x_n\}$ satisfied scheme (2) with $\epsilon_n := \alpha_n$. From $(*)$ we get

$$0 \leq \alpha_n \leq \frac{1-m}{m}[f(x_{n-1}) - f(x_n)]$$

therefore $\{f(x_n)\}$ is decreasing. If $\{f(x_n)\}$ is not bounded from below then $f(x_n) \to -\infty = \inf_X f$. Otherwise apply Theorems 2.1 et 2.2 with $f^n = f = g$, because $\sum_{n=1}^{+\infty} \alpha_n < +\infty$. □

Remark 4.3. *Theorem 4.1 above is nothing but Theorem 4.4 in [1]. Point (ii) is obtained here more simply.*

References

[1] R. Correa and C. Lemaréchal, *Convergence of some algorithms for convex minimization*, Mathematical Programming, 62(1993), 261-276.

[2] N. Dunford and J. Schwartz, *Linear Operators. Part I: General theory*, New-York, (1988).

[3] O. Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIAM J. Control and Optimization, 29(2)(1991), 403-419.

[4] S. Kabbadj, *Méthodes proximales entropiques*, Thèse, Université Montpellier, II(1994).

[5] K.C. Kiwiel, *Proximity control in bundle methods for convex non differentiable minimization*, Mathematical Programming, 46(1)(1990), 105-122.

[6] B. Lemaire, *About the convergence of the proximal method*, Advance in optimization, Proceedings Lambrecht 1991, Lecture Notes in Economics and Mathematical Systems 382, Springer-Verlag, (1992), 39-51.

[7] B. Lemaire, *Coupling optimization methods and variational convergence*, in Trends in Mathematical Optimization, International Series of Numerical Mathematics, 84(c)(1988), 163-179.

[8] B. Lemaire, *The proximal algorithm*, International Series of Numerical Mathematics, 87(1989), 73-86.

[9] C. Lemaréchal, *An extension of Davidon methods to non differentiable problems, in Non differentiable Optimization*, M.L.Balinski, P.Wolfe, eds, Mathematical Programming Study, 3(1975), 95-409.

[10] C. Lemaréchal, J.J. Strodiot and A. Bihain, *On a bundle algorithm for non smooth optimization*, Nonlinear Programming 4, O.L. Mangasarian, R.R.Meyer, S.M. Robinson eds., Academic Press, (1981), 245-282.

[11] B. Martinet, *Perturbation des méthodes d'optimisation*, RAIRO, Analyse numérique, 12(2), 153-171.

[12] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control and Optimization, 14(5)(1975), 877-898.

[13] H. Schramm and J. Zowe, *A version of the bundle idea for minimizing a non smooth function: conceptual idea, convergence analysis, numerical results*, SIAM Journal on Optimization.