Journal of Convex Analysis
Volume 17 (2010), No. 3&4, 765–780

On Some Curvature-Dependent Steplength for the Gradient Method

B. Baji
Université Montpellier II, Place Eugène Bataillon,
34095 Montpellier Cedex 05, France
baji19@free.fr

A. Cabot
Université Montpellier II, Place Eugène Bataillon,
34095 Montpellier Cedex 05, France
acabot@math.univ-montp2.fr

Dedicated to Hedy Attouch on the occasion of his 60th birthday.

Received: October 13, 2008

The aim of this paper is to show the interest of taking into account the notion of curvature in gradient methods. More precisely, given a Hilbert space H and a strictly convex function $\phi : H \to \mathbb{R}$ of class $\mathcal{C}^2$, we consider the following algorithm

$$(\star) \qquad x_{n+1} = x_n - \lambda_n \nabla\phi(x_n), \quad \text{with } \lambda_n = \frac{|\nabla\phi(x_n)|^2}{\langle \nabla^2\phi(x_n).\nabla\phi(x_n), \nabla\phi(x_n)\rangle}.$$

We obtain results of linear convergence for the above algorithm, even without strong convexity. Some variants of $(\star)$ are also considered, with different expressions of the curvature-dependent steplength λ_n. A large part of the paper is devoted to the study of an implicit version of $(\star)$, falling into the field of the proximal point iteration. All these algorithms are clearly related to the Barzilai-Borwein method and numerical illustrations at the end of the paper allow to compare these different schemes.

Keywords: Unconstrained convex optimization, steepest descent, gradient method, proximal point algorithm, Barzilai-Borwein stepsize

1991 Mathematics Subject Classification: 65K10, 90C25, 49M25

1. Introduction

Throughout this paper, we denote by H a Hilbert space endowed with the scalar product $\langle .,. \rangle$ and the corresponding norm $|\,.\,|$. Given a smooth convex objective function $\phi : H \to \mathbb{R}$, let us consider the gradient method with variable stepsize

$$x_{n+1} = x_n - \lambda_n \nabla\phi(x_n), \tag{1}$$

where (λ_n) is a sequence of positive scalars. For every $n \in \mathbb{N}$, define the map $q : \mathbb{R} \to \mathbb{R}$ by $q(\lambda) = \phi(x_n - \lambda\nabla\phi(x_n))$. The optimal step length along the gradient direction is given by the minimum of the map q. If $q''(0) > 0$, it is direct to verify that the quadratic

ISSN 0944-6532 / $ 2.50

approximation of q on a neighbourhood of 0 attains its minimum at $\lambda_n = -q'(0)/q''(0)$, i.e.

$$\lambda_n = \frac{|\nabla\phi(x_n)|^2}{\langle \nabla^2\phi(x_n).\nabla\phi(x_n), \nabla\phi(x_n)\rangle}. \tag{2}$$

When the function ϕ is quadratic, the above expression for the stepsize λ_n corresponds exactly to the classical steepest descent method, also called the Cauchy method [5, 6, 7, 14]. Notice that in dimension one, algorithm (1)–(2) reduces to the classical Newton algorithm $x_{n+1} = x_n - \phi'(x_n)/\phi''(x_n)$, for which we refer to [7, 14] and the references therein.

The above gradient algorithm (1)–(2) can be viewed as the time discretization of the following "Steepest Descent with Curvature" system

$$(SDC) \qquad \dot{x}(t) + \frac{|\nabla\phi(x(t))|^2}{\langle \nabla^2\phi(x(t)).\nabla\phi(x(t)), \nabla\phi(x(t))\rangle}\,\nabla\phi(x(t)) = 0, \quad t \geq 0,$$

which was intensively studied in [1]. The trajectories of (SDC) are simply the steepest descent ones but they are described with different speeds. The most remarkable property of (SDC) lies in the fact that $|\nabla\phi(x(t))| = |\nabla\phi(x(0))|e^{-t}$, for every $t \geq 0$. This is a nice scale invariant property because the rate of convergence is independent of ϕ. The use of second-order information about ϕ gives a normalized exponential decay under no strong convexity condition. The aim of this paper is to show that such a property remains true for the discrete dynamical system (1)–(2). Under suitable conditions, we show that the sequence $(\nabla\phi(x_n))$ converges linearly[1] toward 0, see Theorem 4.2. More generally, we consider the following expression for λ_n

$$\lambda_n = \frac{h\,|\nabla\phi(x_n)|^p}{\langle \nabla^2\phi(x_n).\nabla\phi(x_n), \nabla\phi(x_n)\rangle}, \tag{3}$$

for some coefficient $h > 0$ and some exponent $p > 0$. In this case, the convergence rate of the gradient norm depends on the exponent p, see Theorem 4.3.

In view of practical implementation, we have to consider numerical approximations for the hessian term in (2). This remark leads us to introduce the following alternative expression for λ_n

$$\lambda_n = \frac{h\,\varepsilon_n\,|\nabla\phi(x_n)|^2}{\langle \nabla\phi(x_n + \varepsilon_n\nabla\phi(x_n)) - \nabla\phi(x_n), \nabla\phi(x_n)\rangle}, \tag{4}$$

where (ε_n) is a sequence of positive scalars. It is shown that the above-mentioned result of linear convergence for $(\nabla\phi(x_n))$ still holds for algorithm (1)–(4), see Theorem 4.4.

Barzilai and Borwein proposed in [3] the following choice for the steplength associated with the gradient method

$$\lambda_n = \frac{|x_n - x_{n-1}|^2}{\langle \nabla\phi(x_n) - \nabla\phi(x_{n-1}), x_n - x_{n-1}\rangle}. \tag{5}$$

[1]Let us recall that a sequence (ξ_n) converges linearly to $\bar{\xi}$ if there exists $q \in [0, 1[$ such that $|\xi_{n+1} - \bar{\xi}| \leq q\,|\xi_n - \bar{\xi}|$ for n large enough.

This choice was shown to speed up the convergence of the method and to avoid the poor behavior of the steepest descent algorithm. Raydan [15] proved global convergence for convex quadratic functions, and Dai-Liao [9] established a result of linear convergence. We show throughout the paper that expressions like (2) or (4) for the stepsize λ_n are strongly related to the Barzilai-Borwein choice (5). Numerical experiments at the end of the paper allow to compare these algorithms.

Another way of discretizing a continuous system like (SDC) consists in considering the following implicit scheme

$$x_{n+1} = x_n - \lambda_n \nabla\phi(x_{n+1}), \tag{6}$$

where the expression of λ_n is still given by (2). This algorithm falls into the field of proximal point methods proposed in [12, 13]. Such methods have been intensively studied during the last decades and there is a significant amount of results concerning this type of algorithms, ranging from abstract convergence theorems to applications in nonlinear programming (see for example [2, 8, 10, 11, 16, 17]). In the previous algorithm, the iterate x_{n+1} is uniquely determined by $x_{n+1} = J_{\lambda_n}^{\nabla\phi}(x_n)$, where $J_\lambda^{\nabla\phi} = (I + \lambda\nabla\phi)^{-1}$ is the resolvent of parameter λ of the maximal monotone operator $\nabla\phi$. One advantage of the prox-method (6) versus the gradient one (1) lies in its stability properties. For the sake of presentation, we will discuss first the asymptotic properties of the proximal algorithm (6), which is closer to the corresponding continuous dynamical system. The analysis of the implicit case will then serve as a guideline in the study of the explicit algorithm.

2. Basic assumptions

Throughout the paper, we assume the following set of hypotheses

$$(\mathcal{H}) \quad \begin{cases} \text{The function } \phi : H \to \mathbb{R} \text{ is of class } \mathcal{C}^2. \\ \text{There exist } \bar{x} \in \operatorname{argmin}\phi,\ q \geq 0 \text{ and } m, M > 0 \text{ such that} \end{cases}$$

$$\forall x \in H, \forall y \in H \setminus \{0\}, \quad m\,|x - \bar{x}|^q \leq \frac{\langle \nabla^2\phi(x).y, y\rangle}{|y|^2} \leq M\,|x - \bar{x}|^q\,. \tag{7}$$

Assumption $(\mathcal{H})$ implies that the function ϕ is strictly convex, hence has a unique minimum. When $H = \mathbb{R}^n$, inequality (7) can be rewritten as:

$$\forall x \in \mathbb{R}^n, \quad m\,|x - \bar{x}|^q \leq \mu_1(\nabla^2\phi(x)) \leq \mu_n(\nabla^2\phi(x)) \leq M\,|x - \bar{x}|^q\,,$$

where $\mu_1(\nabla^2\phi(x))$ and $\mu_n(\nabla^2\phi(x))$ denote respectively the smallest and the largest eigenvalue of the matrix $\nabla^2\phi(x)$.

Remark 2.1. Assumption $(\mathcal{H})$ is rather stringent but it provides a basic situation where strong convexity does not hold. This last one corresponds here to the limiting case $q = 0$. More sophisticated models without strong convexity could be considered but they are out of the scope of this paper.

Let us now state two fundamental inequalities under assumption $(\mathcal{H})$.

Lemma 2.2. *Let $\phi : H \to \mathbb{R}$ be a function satisfying assumption $(\mathcal{H})$. Then there exists $c > 0$ such that for every x, $\xi \in H$,*

$$\langle \nabla\phi(x+\xi) - \nabla\phi(x), \xi\rangle \geq m\,c\,|\xi|^2\,|x-\bar{x}|^q \tag{8}$$

$$|\nabla\phi(x+\xi) - \nabla\phi(x)| \leq M\,|\xi|\,(|x-\bar{x}| + |\xi|)^q. \tag{9}$$

Proof. Observing that

$$\nabla\phi(x+\xi) - \nabla\phi(x) = \int_0^1 \nabla^2\phi(x+t\,\xi).\xi\,dt, \tag{10}$$

the left inequality of (7) implies that

$$\langle \nabla\phi(x+\xi) - \nabla\phi(x), \xi\rangle \geq m|\xi|^2 \int_0^1 |x-\bar{x}+t\,\xi|^q\,dt. \tag{11}$$

From Claim 2.3 below, there exists $c > 0$ (independent of x, ξ) such that $\int_0^1 |x-\bar{x}+t\,\xi|^q\,dt \geq c\,|x-\bar{x}|^q$. Inequality (8) then follows immediately.

Let us now prove (9). Starting from equality (10), the right inequality of (7) implies that

$$|\nabla\phi(x+\xi) - \nabla\phi(x)| \leq \int_0^1 \left|\nabla^2\phi(x+t\,\xi).\xi\right|\,dt \leq M\,|\xi| \int_0^1 |x-\bar{x}+t\,\xi|^q\,dt.$$

Since $|x-\bar{x}+t\,\xi| \leq |x-\bar{x}| + |\xi|$ for every $t \in [0,1]$, we deduce inequality (9). $\square$

Let us now establish the result that we used in the above proof.

Claim 2.3. *There exists $c > 0$ such that for all $u, v \in H$,*

$$\int_0^1 |tu + (1-t)v|^q\,dt \geq c\,\max(|u|^q, |v|^q). \tag{12}$$

Proof. If $u = v = 0$, the above inequality is clearly verified for $c = 1$. Without loss of generality, we can then suppose that $|u| \leq |v|$ and $v \neq 0$. For all $t \in [0,1]$, we have $|tu + (1-t)v| \geq |v| - t(|u| + |v|)$. The right member of the previous inequality is nonnegative if and only if $t \leq \frac{|v|}{|u|+|v|}$. Hence we derive that

$$\begin{aligned}\int_0^1 |tu + (1-t)v|^q\,dt &\geq \int_0^{\frac{|v|}{|u|+|v|}} (|v| - t(|u|+|v|))^q\,dt \\ &= \frac{1}{q+1}\frac{|v|^{q+1}}{|u|+|v|} \geq \frac{1}{2\,(q+1)}\,|v|^q.\end{aligned}$$

Thus, the announced inequality is obtained with $c = \frac{1}{2\,(q+1)}$. $\square$

Remark 2.4. If $x = \bar{x}$ inequality (8) provides no information. However in this case, inequality (11) tells us that

$$\langle \nabla\phi(\bar{x}+\xi), \xi\rangle \geq m|\xi|^2 \int_0^1 |t\,\xi|^q\,dt = \frac{m}{q+1}\,|\xi|^{q+2}.$$

Since $\langle \nabla\phi(\bar{x}+\xi), \xi\rangle \leq |\nabla\phi(\bar{x}+\xi)|\,|\xi|$, we infer that for every $\xi \in H$,

$$|\nabla\phi(\bar{x}+\xi)| \geq \frac{m}{q+1}\,|\xi|^{q+1}. \tag{13}$$

3. Proximal point with curvature algorithm

3.1. Algorithm (PPC)

Let $\phi : H \to \mathbb{R}$ be a convex function of class C^1 and let us consider the sequence (x_n) defined by the following implicit algorithm

$$x_{n+1} = x_n - \lambda_n \nabla\phi(x_{n+1}), \tag{14}$$

for some positive sequence (λ_n). Iteration (14) falls into the category of the proximal point algorithms. For classical results on the convergence of proximal algorithms, the reader is referred to [4, 10, 11, 16].

Considering any sequence (x_n) generated by algorithm (14), it is immediate to check that the gradient norm $|\nabla\phi(x_n)|$ is decreasing as $n \to +\infty$. In the next lemma, we evaluate its decay rate when the function ϕ satisfies assumption $(\mathcal{H})$.

Lemma 3.1. *Let $\phi : H \to \mathbb{R}$ be a function satisfying assumption $(\mathcal{H})$. Given a sequence (λ_n) of positive scalars, consider any sequence (x_n) generated by the proximal point algorithm* (14)*. For every $n \in \mathbb{N}$, we have*

$$|\nabla\phi(x_{n+1})|^2 - |\nabla\phi(x_n)|^2 \leq -2mc\lambda_n\,|x_n - \bar{x}|^q\,|\nabla\phi(x_{n+1})|^2, \tag{15}$$

where $c > 0$ is the scalar given by Lemma 2.3.

Proof. Let us observe that

$$\begin{aligned}
&|\nabla\phi(x_{n+1})|^2 - |\nabla\phi(x_n)|^2 \\
&= 2\,\langle \nabla\phi(x_{n+1}) - \nabla\phi(x_n), \nabla\phi(x_{n+1})\rangle - |\nabla\phi(x_{n+1}) - \nabla\phi(x_n)|^2 \\
&\leq\; -\frac{2}{\lambda_n}\,\langle \nabla\phi(x_{n+1}) - \nabla\phi(x_n), x_{n+1} - x_n\rangle\,.
\end{aligned} \tag{16}$$

By applying inequality (8) with $x = x_n$ and $\xi = x_{n+1} - x_n$, we find

$$\langle \nabla\phi(x_{n+1}) - \nabla\phi(x_n), x_{n+1} - x_n\rangle \geq mc\,|x_{n+1} - x_n|^2\,|x_n - \bar{x}|^q\,.$$

Recalling that $|x_{n+1} - x_n| = \lambda_n\,|\nabla\phi(x_{n+1})|$, we deduce inequality (15). $\square$

Given $h > 0$ and a function $\phi : H \to \mathbb{R}$ satisfying assumption $(\mathcal{H})$, we consider algorithm (14) with the following steplength

$$\lambda_n = \frac{h\,|\nabla\phi(x_n)|^2}{\langle \nabla^2\phi(x_n).\nabla\phi(x_n), \nabla\phi(x_n)\rangle},$$

while $x_n \neq \bar{x}$. If $x_{n_0} = \bar{x}$ for some $n_0 \in \mathbb{N}$, then we set by convention $x_n = \bar{x}$ for every $n \geq n_0$. This algorithm will be referred to as the "Proximal Point with Curvature" algorithm, PPC for short.

Let us establish a result of linear convergence for the sequence (x_n) when the function ϕ satisfies assumption $(\mathcal{H})$.

Theorem 3.2. *Let $\phi : H \to \mathbb{R}$ be a function satisfying assumption $(\mathcal{H})$. Given $h > 0$, consider any sequence (x_n) defined by algorithm (PPC). Then there exists $r \in [0, 1[$ such that $|\nabla\phi(x_{n+1})| \leq r\,|\nabla\phi(x_n)|$ for every $n \in \mathbb{N}$.*

Proof. From the definition of λ_n and the right inequality of (7), we have

$$\lambda_n \geq \frac{h}{M}\frac{1}{|x_n - \bar{x}|^q}.$$

By using formula (15) of Lemma 3.1, we deduce that for every $n \in \mathbb{N}$

$$|\nabla\phi(x_{n+1})|^2 - |\nabla\phi(x_n)|^2 \leq -2hc\frac{m}{M}\,|\nabla\phi(x_{n+1})|^2.$$

We conclude that $|\nabla\phi(x_{n+1})| \leq \left[1 + 2hc\frac{m}{M}\right]^{-1/2} |\nabla\phi(x_n)|$ for every $n \in \mathbb{N}$. □

The main interest of Theorem 3.2 is to give a result of linear convergence for algorithm (PPC), under no strong convexity condition.

Remark 3.3. Under assumption $(\mathcal{H})$, the above result of linear convergence for the sequence $(\nabla\phi(x_n))$ immediately implies a corresponding result of linear convergence for the sequence $(|x_n - \bar{x}|)$. Indeed, we infer from inequality (13) that

$$\forall n \in \mathbb{N}, \quad |x_n - \bar{x}|^{q+1} \leq \frac{q+1}{m}\,|\nabla\phi(x_n)| \leq \frac{q+1}{m}\,|\nabla\phi(x_0)|\,r^n.$$

Setting $\rho = r^{\frac{1}{q+1}}$, we conclude that $|x_n - \bar{x}| = O(\rho^n)$ when $n \to +\infty$.

3.2. Some variant for the stepsize of algorithm (PPC)

Given a positive parameter $h > 0$ and an exponent $p \in \mathbb{R}$, let us consider the following expression of the steplength λ_n

$$\lambda_n = \frac{h\,|\nabla\phi(x_n)|^p}{\langle \nabla^2\phi(x_n).\nabla\phi(x_n), \nabla\phi(x_n)\rangle}. \tag{17}$$

Theorem 3.4. *Let $\phi : H \to \mathbb{R}$ be a function satisfying assumption $(\mathcal{H})$. Given $h > 0$ and $p \in \mathbb{R}$, consider any sequence (x_n) defined by algorithm (14)–(17). Then the following properties hold:*

(*a*) *If $p < 2$, the order of convergence of the sequence $(\nabla\phi(x_n))$ toward 0 equals $s = \frac{4-p}{2} > 1$: there exists $r \in [0,1[$ such that $|\nabla\phi(x_n)| = O\left(r^{s^n}\right)$ when $n \to +\infty$.*

(*b*) *If $p > 2$, we have $|\nabla\phi(x_n)| = O\left(n^{-\frac{1}{p-2}}\right)$ when $n \to +\infty$.*

Proof. By adapting the computations of Theorem 3.2, we immediately obtain that for every $n \in \mathbb{N}$

$$|\nabla\phi(x_{n+1})|^2 \leq |\nabla\phi(x_n)|^2 \left[1 + 2hc\frac{m}{M}|\nabla\phi(x_n)|^{p-2}\right]^{-1}. \tag{18}$$

The sequence $(|\nabla\phi(x_n)|)_{n\in\mathbb{N}}$ is clearly nonincreasing and it is immediate to prove that $\lim_{n\to+\infty}|\nabla\phi(x_n)| = 0$.

(*a*) Assume that $p < 2$. Inequality (18) implies $|\nabla\phi(x_{n+1})|^2 \leq \frac{M}{2hcm}|\nabla\phi(x_n)|^{4-p}$ for every $n \in \mathbb{N}$. Setting $A = \left(\frac{M}{2hcm}\right)^{1/2}$ and $s = \frac{4-p}{2} > 1$, we obtain $|\nabla\phi(x_{n+1})| \leq A\,|\nabla\phi(x_n)|^s$ for every $n \in \mathbb{N}$. We deduce that for all n_0, $n \in \mathbb{N}$ such that $n_0 \leq n$,

$$|\nabla\phi(x_n)| \leq A^{1+s+\ldots+s^{n-n_0-1}}|\nabla\phi(x_{n_0})|^{s^{n-n_0}} \leq \left[A^{\frac{1}{s-1}}|\nabla\phi(x_{n_0})|\right]^{s^{n-n_0}}.$$

Let us choose $n_0 \in \mathbb{N}$ so that $A^{\frac{1}{s-1}}|\nabla\phi(x_{n_0})| < 1$. The conclusion then follows by setting $r = \left[A^{\frac{1}{s-1}}|\nabla\phi(x_{n_0})|\right]^{s^{-n_0}}$.

(*b*) Assume that $p > 2$. Since $\lim_{n\to+\infty}|\nabla\phi(x_n)|^{p-2} = 0$, we infer that, for n large enough

$$\begin{aligned}\left[1 + 2hc\frac{m}{M}|\nabla\phi(x_n)|^{p-2}\right]^{-1} &= 1 - 2hc\frac{m}{M}|\nabla\phi(x_n)|^{p-2} + o(|\nabla\phi(x_n)|^{p-2})\\ &\leq 1 - hc\frac{m}{M}|\nabla\phi(x_n)|^{p-2}.\end{aligned}$$

In view of inequality (18) we then have

$$|\nabla\phi(x_{n+1})|^2 \leq |\nabla\phi(x_n)|^2\left(1 - hc\frac{m}{M}|\nabla\phi(x_n)|^{p-2}\right).$$

Lemma 3.5 below applied to the sequence $u_n = |\nabla\phi(x_n)|^2$ enables us to derive that $|\nabla\phi(x_n)| = O(n^{-\frac{1}{p-2}})$ as $n \to +\infty$. $\square$

Lemma 3.5. *Given $a > 0$ and $s > 0$, let (u_n) be a sequence of positive numbers such that*

$$\forall n \in \mathbb{N}, \quad u_{n+1} \leq u_n(1 - au_n^s). \tag{19}$$

Then, we have $u_n = O(n^{-\frac{1}{s}})$ when $n \to +\infty$.

For the proof of this result, the reader is referred to the book of Polyak [14, Lemma 6, Chapter 2].

Remark 3.6. The results of Theorem 3.4 on the sequence $(\nabla\phi(x_n))$ have immediate consequences on the rate of convergence of the sequence (x_n) itself. When $p < 2$, we deduce from Theorem 3.4(a) and inequality (13) that $|x_n - \bar{x}| = O(\rho^{s^n})$ when $n \to +\infty$, with $\rho = r^{\frac{1}{(q+1)}}$. When $p > 2$, we deduce in the same way from Theorem 3.4(b) that $|x_n - \bar{x}| = O\left(n^{-\frac{1}{(p-2)(q+1)}}\right)$ when $n \to +\infty$.

3.3. Numerical approximation of the stepsize of algorithm (PPC)

For numerical purposes, it is fundamental to observe that the Hessian term $\nabla^2\phi(x_n).\nabla\phi(x_n)$ can be approximated as follows

$$\nabla^2\phi(x_n).\nabla\phi(x_n) \approx \frac{1}{\varepsilon}\left[\nabla\phi(x_n + \varepsilon\nabla\phi(x_n)) - \nabla\phi(x_n)\right]$$

for ε sufficiently small. This remark leads us to consider the following expression for λ_n

$$\lambda_n = \frac{h\,\varepsilon_n\,|\nabla\phi(x_n)|^2}{\langle\nabla\phi(x_n + \varepsilon_n\nabla\phi(x_n)) - \nabla\phi(x_n), \nabla\phi(x_n)\rangle}, \tag{20}$$

where (ε_n) is a sequence of positive scalars. By applying inequalities (8)–(9) with $x = x_n$ and $\xi = \varepsilon_n\,\nabla\phi(x_n)$, we find

$$\begin{aligned} m\,c\,\varepsilon_n\,|x_n - \bar{x}|^q &\le \frac{\langle\nabla\phi(x_n + \varepsilon_n\nabla\phi(x_n)) - \nabla\phi(x_n), \nabla\phi(x_n)\rangle}{|\nabla\phi(x_n)|^2} \\ &\le M\,\varepsilon_n\,(|x_n - \bar{x}| + \varepsilon_n\,|\nabla\phi(x_n)|)^q. \end{aligned} \tag{21}$$

The left inequality shows that algorithm (14)–(20) is well-defined while $x_n \ne \bar{x}$. If $x_{n_0} = \bar{x}$ for some $n_0 \in \mathbb{N}$, then we set by convention $x_n = \bar{x}$ for every $n \ge n_0$.

On the other hand, by applying inequality (9) with $x = \bar{x}$ and $\xi = x_n - \bar{x}$, we find $|\nabla\phi(x_n)| \le M\,|x_n - \bar{x}|^{q+1}$. Hence in view of inequalities (21), we infer that

$$\frac{h}{M}\,\frac{1}{|x_n - \bar{x}|^q\,\left[1 + \varepsilon_n\,M\,|x_n - \bar{x}|^q\right]^q} \le \lambda_n \le \frac{h}{mc}\,\frac{1}{|x_n - \bar{x}|^q}. \tag{22}$$

We are now able to prove the following result.

Theorem 3.7. *Let $\phi : H \to \mathbb{R}$ be a function satisfying assumption $(\mathcal{H})$. Given $h > 0$ and a sequence (ε_n) of positive scalars, consider any sequence (x_n) defined by algorithm* (14)–(20)*. Assume moreover that one of the following cases holds:*

(i) *The sequence (ε_n) is bounded.*

(ii) *The sequence (ε_n) is defined by $\varepsilon_n = \lambda_{n-1}$ (Barzilai-Borwein choice[2]).*

Then there exists $r \in [0, 1[$ such that $|\nabla\phi(x_{n+1})| \le r\,|\nabla\phi(x_n)|$ for every $n \in \mathbb{N}$.

Proof. (i) First of all, it is direct to check that the sequence $(|x_n - \bar{x}|)$ is nonincreasing[3]. If the sequence (ε_n) is majorized by some $\bar{\varepsilon} > 0$, then we derive from the left inequality

[2]see Remark 3.8 below.

[3]This is a general feature shared by proximal point methods. Combined with the Opial lemma, this property is a key ingredient to prove the weak convergence of the proximal algorithm.

of (22) that

$$\lambda_n \geq \frac{h}{MC}\,\frac{1}{|x_n - \bar{x}|^q},$$

where $C = [1 + \bar{\varepsilon}\, M\, |x_0 - \bar{x}|^q]^q$. By using formula (15) of Lemma 3.1, we deduce that for every $n \in \mathbb{N}$

$$|\nabla\phi(x_{n+1})|^2 - |\nabla\phi(x_n)|^2 \leq -2h\frac{mc}{MC}\,|\nabla\phi(x_{n+1})|^2,$$

and consequently $|\nabla\phi(x_{n+1})| \leq |\nabla\phi(x_n)|\left[1 + 2h\frac{mc}{MC}\right]^{-1/2}$.

(ii) In view of the right inequality of (22), we have

$$\varepsilon_n = \lambda_{n-1} \leq \frac{h}{mc}\,\frac{1}{|x_{n-1} - \bar{x}|^q} \leq \frac{h}{mc}\,\frac{1}{|x_n - \bar{x}|^q}.$$

Hence the left inequality of (22) gives in turn

$$\lambda_n \geq \frac{h}{MC'}\,\frac{1}{|x_n - \bar{x}|^q},$$

where $C' = \left[1 + \frac{hM}{mc}\right]^q$. Then, it suffices to conclude as in the proof of (i) with C' in place of C. $\square$

Remark 3.8. Let us now comment on the particular choice corresponding to $\varepsilon_n = \lambda_{n-1}$ for every $n \geq 1$. From the definition of algorithm (14), we have $x_n + \lambda_{n-1}\,\nabla\phi(x_n) = x_{n-1}$, so that expression (20) of λ_n becomes

$$\lambda_n = \frac{h\,|x_n - x_{n-1}|^2}{\langle\nabla\phi(x_n) - \nabla\phi(x_{n-1}), x_n - x_{n-1}\rangle}. \tag{23}$$

It is remarkable that such a choice of ε_n leads exactly to the Barzilai-Borwein steplength.

In the sequel, we consider an alternative expression for the steplength λ_n, based on a different approximation of the Hessian term. It is elementary to check that

$$\langle\nabla^2\phi(x_n).\nabla\phi(x_n), \nabla\phi(x_n)\rangle \approx [\phi(x_n + \varepsilon\nabla\phi(x_n)) + \phi(x_n - \varepsilon\nabla\phi(x_n)) - 2\phi(x_n)]/\varepsilon^2$$

for ε sufficiently small. This remark leads us to consider the following expression for λ_n

$$\lambda_n = \frac{h\,\varepsilon_n^2\,|\nabla\phi(x_n)|^2}{\phi(x_n + \varepsilon_n\nabla\phi(x_n)) + \phi(x_n - \varepsilon_n\nabla\phi(x_n)) - 2\phi(x_n)}, \tag{24}$$

where (ε_n) is a sequence of positive scalars. A suitable use of the second-order Taylor formula with integral remainder shows that

$$\begin{aligned}&\phi(x_n + \varepsilon_n\nabla\phi(x_n)) + \phi(x_n - \varepsilon_n\nabla\phi(x_n)) - 2\phi(x_n)\\ &= \varepsilon_n^2\int_{-1}^{1}(1 - |t|)\,\langle\nabla^2\phi(x_n + t\,\varepsilon_n\,\nabla\phi(x_n)).\nabla\phi(x_n), \nabla\phi(x_n)\rangle\,dt.\end{aligned}$$

From inequalities (7), we derive that

$$\begin{aligned}
&m\,\varepsilon_n^2 \int_{-1}^{1} (1-|t|)\,|x_n - \bar{x} + t\,\varepsilon_n\,\nabla\phi(x_n)|^q\,dt \\
&\le \frac{\phi(x_n+\varepsilon_n\nabla\phi(x_n)) + \phi(x_n-\varepsilon_n\nabla\phi(x_n)) - 2\phi(x_n)}{|\nabla\phi(x_n)|^2} \\
&\le M\varepsilon_n^2 \int_{-1}^{1} (1-|t|)\,|x_n - \bar{x} + t\,\varepsilon_n\,\nabla\phi(x_n)|^q\,dt. \qquad (25)
\end{aligned}$$

If the numerator of the middle term equals 0, then the left integral is also equal to 0. Since its integrand is a nonnegative continuous function, it is equal to 0 everywhere on $[-1,1]$. In particular, for $t=0$ we find $x_n = \bar{x}$. This shows that algorithm (14)–(24) is well-defined while $x_n \neq \bar{x}$. If $x_{n_0} = \bar{x}$ for some $n_0 \in \mathbb{N}$, then we set by convention $x_n = \bar{x}$ for every $n \ge n_0$.

We are going to show that under suitable conditions, any sequence defined by algorithm (14)–(24) satisfies the same property of linear convergence as algorithm (PPC).

Theorem 3.9. *Let $\phi : H \to \mathbb{R}$ be a function satisfying assumption $(\mathcal{H})$. Given $h>0$ and a bounded sequence (ε_n) of positive scalars, consider any sequence (x_n) defined by algorithm (14)–(24). Then there exists $r \in [0,1[$ such that $|\nabla\phi(x_{n+1})| \le r\,|\nabla\phi(x_n)|$ for every $n \in \mathbb{N}$.*

Proof. Recalling that $|\nabla\phi(x_n)| \le M\,|x_n-\bar{x}|^{q+1}$, we find for every $t\in[0,1]$,

$$\begin{aligned}
|x_n - \bar{x} + t\,\varepsilon_n\nabla\phi(x_n)| &\le |x_n-\bar{x}| + \varepsilon_n\,|\nabla\phi(x_n)| \\
&= |x_n-\bar{x}|\,\left[1+\varepsilon_n\,M\,|x_n-\bar{x}|^q\right].
\end{aligned}$$

In view of the right inequality of (25) and the expression of λ_n, we deduce that

$$\lambda_n \ge \frac{h}{M}\,\frac{1}{|x_n-\bar{x}|^q\,\left[1+\varepsilon_n\,M\,|x_n-\bar{x}|^q\right]^q}.$$

It suffices then to use the same arguments as in the proof of Theorem 3.7(i). □

4. Gradient with curvature method

4.1. Algorithm (GCM)

Given a differentiable function $\phi : H \to \mathbb{R}$, a standard minimization technique consists in considering the following gradient algorithm with variable stepsize

$$x_{n+1} = x_n - \lambda_n \nabla\phi(x_n). \qquad (26)$$

The purpose of the next lemma is to give a first upper bound for the ratio $|\nabla\phi(x_{n+1})|/|\nabla\phi(x_n)|$ when the function ϕ satisfies assumption $(\mathcal{H})$. It plays the same role as Lemma 3.1 of Section 3.

Lemma 4.1. *Let $\phi : H \to \mathbb{R}$ be a function satisfying assumption $(\mathcal{H})$. Given a sequence (λ_n) of positive scalars, consider any sequence (x_n) generated by the gradient algorithm* (26)*. For every $n \in \mathbb{N}$, we have*

$$\begin{aligned} &|\nabla\phi(x_{n+1})|^2 \\ \leq\ & |\nabla\phi(x_n)|^2 \Big[1 - 2m\,c\,\lambda_n\,|x_n - \bar{x}|^q + M^2\,\lambda_n^2\,|x_n - \bar{x}|^{2q}\,[1 + M\lambda_n\,|x_n - \bar{x}|^q]^{2q}\Big], \end{aligned} \tag{27}$$

where $c > 0$ is the scalar given by Lemma 2.3*.*

Proof. Let us evaluate the difference $|\nabla\phi(x_{n+1})|^2 - |\nabla\phi(x_n)|^2$:

$$\begin{aligned} &|\nabla\phi(x_{n+1})|^2 - |\nabla\phi(x_n)|^2 \\ =\ & 2\,\langle\nabla\phi(x_{n+1}) - \nabla\phi(x_n), \nabla\phi(x_n)\rangle + |\nabla\phi(x_{n+1}) - \nabla\phi(x_n)|^2 \\ =\ & -\frac{2}{\lambda_n}\,\langle\nabla\phi(x_{n+1}) - \nabla\phi(x_n), x_{n+1} - x_n\rangle + |\nabla\phi(x_{n+1}) - \nabla\phi(x_n)|^2 . \end{aligned} \tag{28}$$

By arguing as in the proof of Lemma 3.1, we find

$$\langle\nabla\phi(x_{n+1}) - \nabla\phi(x_n), x_{n+1} - x_n\rangle \geq m\,c\,\lambda_n^2\,|\nabla\phi(x_n)|^2\,|x_n - \bar{x}|^q . \tag{29}$$

By applying inequality (9) with $x = x_n$ and $\xi = x_{n+1} - x_n$, we find

$$|\nabla\phi(x_{n+1}) - \nabla\phi(x_n)| \leq M\,\lambda_n\,|\nabla\phi(x_n)|\,\left(|x_n - \bar{x}| + \lambda_n\,|\nabla\phi(x_n)|\right)^q.$$

Recalling that $|\nabla\phi(x_n)| \leq M\,|x_n - \bar{x}|^{q+1}$, we derive that

$$|\nabla\phi(x_{n+1}) - \nabla\phi(x_n)| \leq M\,\lambda_n\,|\nabla\phi(x_n)|\,|x_n - \bar{x}|^q\,[1 + M\lambda_n\,|x_n - \bar{x}|^q]^q . \tag{30}$$

Coming back to formula (28), we infer from inequalities (29) and (30) that

$$\begin{aligned} &|\nabla\phi(x_{n+1})|^2 - |\nabla\phi(x_n)|^2 \\ \leq\ & -2m\,c\,\lambda_n\,|\nabla\phi(x_n)|^2\,|x_n - \bar{x}|^q + M^2\,\lambda_n^2\,|\nabla\phi(x_n)|^2\,|x_n - \bar{x}|^{2q}\,[1 + M\lambda_n\,|x_n - \bar{x}|^q]^{2q} , \end{aligned}$$

and the conclusion immediately follows. □

Given $h > 0$ and a function $\phi : H \to \mathbb{R}$ satisfying assumption $(\mathcal{H})$, we consider algorithm (26) with the following steplength

$$\lambda_n = \frac{h\,|\nabla\phi(x_n)|^2}{\langle\nabla^2\phi(x_n).\nabla\phi(x_n), \nabla\phi(x_n)\rangle},$$

while $x_n \neq \bar{x}$. If $x_{n_0} = \bar{x}$ for some $n_0 \in \mathbb{N}$, then we set by convention $x_n = \bar{x}$ for every $n \geq n_0$. This algorithm will be referred to as the "Gradient with Curvature Method" algorithm, GCM for short.

Theorem 4.2. *Let $\phi : H \to \mathbb{R}$ be a function satisfying assumption $(\mathcal{H})$. Then for $h > 0$ sufficiently small, there exists $r \in [0, 1[$ such that any sequence (x_n) defined by algorithm (GCM) satisfies $|\nabla\phi(x_{n+1})| \leq r\,|\nabla\phi(x_n)|$ for every $n \in \mathbb{N}$.*

Proof. By definition of λ_n, we immediately have in view of inequalities (7):

$$\forall n \in \mathbb{N}, \quad \frac{h}{M} \frac{1}{|x_n - \bar{x}|^q} \leq \lambda_n \leq \frac{h}{m} \frac{1}{|x_n - \bar{x}|^q}.$$

Taking into account (27), we obtain for every $n \in \mathbb{N}$

$$|\nabla\phi(x_{n+1})|^2 \leq |\nabla\phi(x_n)|^2 \left[1 - 2h\, c\, \frac{m}{M} + \underbrace{h^2 \frac{M^2}{m^2} \left(1 + h\frac{M}{m}\right)^{2q}}_{\theta(h)} \right].$$

It is clear that the above quantity $\theta(h)$ is negligible with respect to h when $h \to 0$. Hence there exists $h_0 > 0$ such that for every $h \in]0, h_0[$ and for every $n \in \mathbb{N}$

$$|\nabla\phi(x_{n+1})|^2 \leq |\nabla\phi(x_n)|^2 \left[1 - h\, c\, \frac{m}{M}\right],$$

and the conclusion is satisfied with $r := (1 - h\, c\, \frac{m}{M})^{1/2}$. □

4.2. Some variant for the stepsize of algorithm (GCM)

Given a positive parameter $h > 0$ and a positive exponent $p > 0$, let us consider the following expression of the steplength λ_n

$$\lambda_n = \frac{h\, |\nabla\phi(x_n)|^p}{\langle \nabla^2\phi(x_n).\nabla\phi(x_n), \nabla\phi(x_n)\rangle}. \tag{31}$$

Before treating the general case, let us consider the particular case corresponding to the function $\phi : \mathbb{R} \to \mathbb{R}$ defined by $\phi(x) = |x|^a$, for some $a \geq 2$. In this case, algorithm (26)–(31) reads as

$$\begin{aligned} x_{n+1} &= x_n - h\, \frac{|\phi'(x_n)|^{p-2}}{\phi''(x_n)}\, \phi'(x_n) \\ &= x_n \left(1 - h\, \frac{a^{p-2}}{a-1}\, |x_n|^{(a-1)(p-2)}\right). \end{aligned}$$

If $x_0 = 0$, algorithm (26)–(31) definitively stops. Let us now assume that $x_0 \neq 0$. Setting $\theta = h\, \frac{a^{p-2}}{a-1}\, |x_0|^{(a-1)(p-2)}$, we let the reader check that if $p \geq 2$ the following properties hold

- If $\theta < 2$ then $|x_n| \searrow 0$ as $n \to +\infty$.
- If $\theta = 2$ then $x_n = (-1)^n\, x_0$.
- If $\theta > 2$ then $|x_n| \nearrow +\infty$ as $n \to +\infty$.

On the other hand, when $p < 2$, algorithm (26)–(31) is always divergent.

Coming back to the general case, the next proposition gives an estimate for the convergence rate of the gradient norm in the case $p > 2$.

Theorem 4.3. *Let $\phi : H \to \mathbb{R}$ be a function satisfying assumption $(\mathcal{H})$. Given $p > 2$, there exists $h_0 > 0$ such that for every $h \in]0, h_0[$, any sequence (x_n) defined by algorithm* (26)–(31) *satisfies $|\nabla\phi(x_n)| = O(n^{-\frac{1}{p-2}})$ when $n \to +\infty$.*

Proof. By definition of λ_n, we immediately have in view of inequalities (7):

$$\forall n \in \mathbb{N}, \quad \frac{h\,|\nabla\phi(x_n)|^{p-2}}{M\,|x_n - \bar{x}|^q} \leq \lambda_n \leq \frac{h\,|\nabla\phi(x_n)|^{p-2}}{m\,|x_n - \bar{x}|^q}. \tag{32}$$

Taking into account (27), we obtain for every $n \in \mathbb{N}$

$$|\nabla\phi(x_{n+1})|^2$$

$$\leq |\nabla\phi(x_n)|^2 \left[1 - 2h\,c\,\frac{m}{M}\,|\nabla\phi(x_n)|^{p-2} + \underbrace{h^2\frac{M^2}{m^2}\,|\nabla\phi(x_n)|^{2(p-2)}\left(1+h\frac{M}{m}\,|\nabla\phi(x_n)|^{p-2}\right)^{2q}}_{\theta(h)}\right].$$

It is clear that the above quantity $\theta(h)$ is negligible with respect to h when $h \to 0$. Hence there exists $h_0 > 0$ such that for every $h \in]0, h_0[$ and for every $n \in \mathbb{N}$

$$|\nabla\phi(x_{n+1})|^2 \leq |\nabla\phi(x_n)|^2 \left[1 - h\,c\,\frac{m}{M}\,|\nabla\phi(x_n)|^{p-2}\right].$$

It suffices now to apply Lemma 3.5 to the sequence $u_n = |\nabla\phi(x_n)|^2$ with $a = h\,c\,\frac{m}{M}$ and $s = \frac{p-2}{2}$. □

4.3. Numerical approximation of the stepsize of algorithm (GCM)

Guided by numerical purposes, we consider as in Section 3 the following alternative expression for λ_n

$$\lambda_n = \frac{h\,\varepsilon_n\,|\nabla\phi(x_n)|^2}{\langle\nabla\phi(x_n + \varepsilon_n\nabla\phi(x_n)) - \nabla\phi(x_n), \nabla\phi(x_n)\rangle}, \tag{33}$$

where (ε_n) is a sequence of positive scalars.

Theorem 4.4. *Let $\phi : H \to \mathbb{R}$ be a function satisfying assumption $(\mathcal{H})$. Let (ε_n) be a bounded sequence of positive scalars. Then for $h > 0$ sufficiently small, there exists $r \in [0, 1[$ such that any bounded sequence (x_n) defined by algorithm* (26)–(33) *satisfies $|\nabla\phi(x_{n+1})| \leq r\,|\nabla\phi(x_n)|$ for every $n \in \mathbb{N}$.*

Proof. Let us first recall that the stepsize λ_n satisfies the double inequality (22). Since the sequences (ε_n) and (x_n) are supposed to be bounded, there exists $C > 0$ such that $[1 + \varepsilon_n\,M\,|x_n - \bar{x}|^q]^q \leq C$ for every $n \in \mathbb{N}$. Hence we infer from (22) that

$$\frac{h}{MC}\,\frac{1}{|x_n - \bar{x}|^q} \leq \lambda_n \leq \frac{h}{mc}\,\frac{1}{|x_n - \bar{x}|^q}.$$

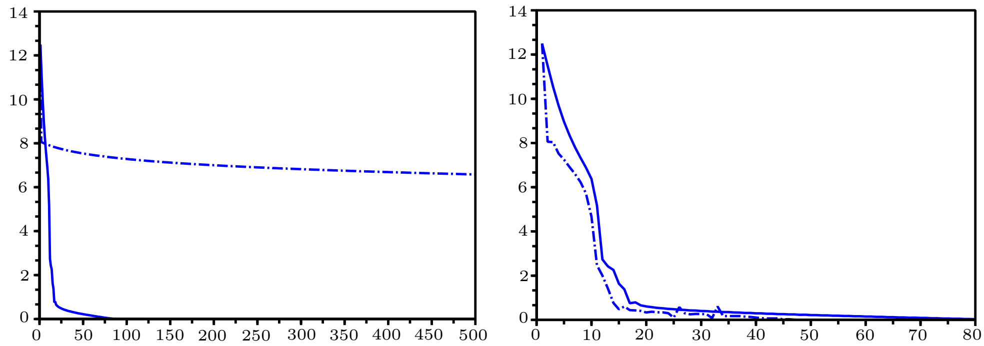

Figure 4.1: Convergence history for ($|x_n - x_{\min}|$) during the minimization of the Beale function with algorithms (A), (B) and (C). Left: comparison of algorithms (A) (solid line) and (C) (dashed line). Right: comparison of algorithms (A) (solid line) and (B) (dashed line). Initial point: $x_0 = (-4, -5)$, parameters: $h = 1$, $\lambda = 10^{-5}$, $\varepsilon = 10^{-6}$.

By using formula (27) of Lemma 4.1, we obtain for every $n \in \mathbb{N}$

$$|\nabla\phi(x_{n+1})|^2 \leq |\nabla\phi(x_n)|^2 \left[1 - 2h\,\frac{mc}{MC} + \underbrace{h^2 \frac{M^2}{m^2c^2}\left(1 + h\frac{M}{mc}\right)^{2q}}_{\theta(h)} \right].$$

It is clear that the above quantity $\theta(h)$ is negligible with respect to h when $h \to 0$. Hence there exists $h_0 > 0$ such that for every $h \in]0, h_0[$ and for every $n \in \mathbb{N}$

$$|\nabla\phi(x_{n+1})|^2 \leq |\nabla\phi(x_n)|^2 \left[1 - h\,\frac{mc}{MC}\right],$$

and the conclusion is satisfied with $r := (1 - h\,\frac{mc}{MC})^{1/2}$. $\square$

4.4. Numerical Experiments

In this paragraph, we are going to test the (GCM) algorithm and to compare it with two other gradient methods, respectively with fixed stepsize and with Barzilai-Borwein stepsize. The approximation of the Hessian term by a finite difference scheme leads us to consider the expression of λ_n given by (33). In the sequel, we implement the following algorithm:

$$(A) \qquad x_{n+1} = x_n - \frac{h\,\varepsilon\,|\nabla\phi(x_n)|^2}{|\langle\nabla\phi(x_n + \varepsilon\nabla\phi(x_n)) - \nabla\phi(x_n), \nabla\phi(x_n)\rangle|}\nabla\phi(x_n),$$

with $h > 0$ and $\varepsilon > 0$. The absolute value at the denominator permits to deal with objective functions that are either locally convex or concave. Algorithm (A) will be compared to the Barzilai-Borwein gradient method[4]

$$(B) \qquad x_{n+1} = x_n - \frac{h\,|x_n - x_{n-1}|^2}{|\langle\nabla\phi(x_n) - \nabla\phi(x_{n-1}), x_n - x_{n-1}\rangle|}\nabla\phi(x_n),$$

[4]In the original Barzilai-Borwein method, the parameter h equals 1 and there is no absolute value.

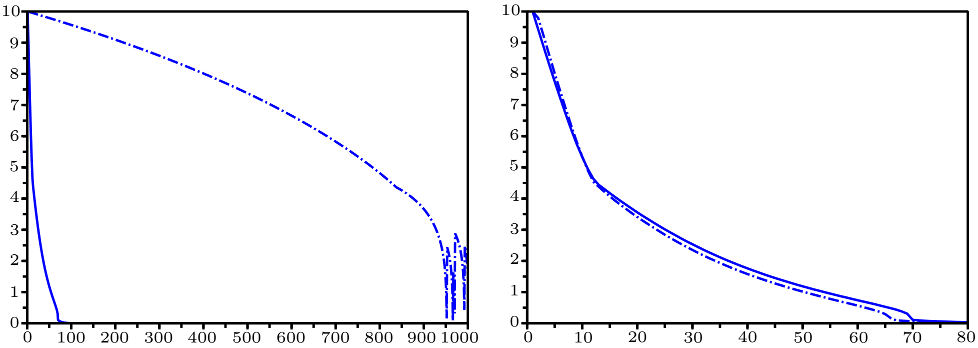

Figure 4.2: Convergence history for $(|x_n - x_{\min}|)$ during the minimization of the Shekel function with algorithms (A), (B) and (C). Left: comparison of algorithms (A) (solid line) and (C) (dashed line). Right: comparison of algorithms (A) (solid line) and (B) (dashed line). Initial point: $x_0 = (5, 0, 5, 0)$, parameters: $h = \lambda = 0.1$, $\varepsilon = 10^{-5}$.

and also to the basic gradient method with fixed stepsize $\lambda > 0$

$$(C) \qquad x_{n+1} = x_n - \lambda \nabla\phi(x_n).$$

We use two objective functions that are well-known in optimization theory: the Beale function and the Shekel function.

Example 4.5. Let us start with the Beale function

$$\phi(x_1, x_2) = (1.5 - x_1 + x_1 x_2)^2 + (2.25 - x_1 + x_1 x_2^2)^2 + (2.625 - x_1 + x_1 x_2^3)^2.$$

This function admits a unique global minimum point $x_{\min} = (3, 0.5)$ on the subset $[-5, 5] \times [-5, 5]$ and its minimum value is $\phi(x_{\min}) = 0$. It is very rough and steep specially at the point $(-5, -5)$, leading algorithm (C) to fail. If one chooses a small stepsize $\lambda << 1$, the discrete trajectories (x_n) generated by (C) are described very slowly and they do not converge toward $x_{\min}$ because the function is very flat in a neighbourhood of this point. On the other hand, the choice of a large stepsize in (C) gives sequences that are very unstable and strongly oscillating. On the contrary, any sequence (x_n) generated by algorithms (A) or (B) converges to $x_{\min}$ very fast (see Figure 4.1).

Example 4.6. We now consider the Shekel function

$$\phi(x) = -\sum_{i=1}^{5} 1/(|x - d_i|^2 + c_i),$$

where $x \in [0, 10]^4$. The coefficients (c_i) and (d_i) are respectively given by:

$$(c_1, c_2, c_3, c_4, c_5) = (0.1, 0.2, 0.2, 0.4, 0.4),$$

$$d_1 = (4, 4, 4, 4),\ \ d_2 = (1, 1, 1, 1),\ \ d_3 = (8, 8, 8, 8),\ \ d_4 = (6, 6, 6, 6),\ \ d_5 = (3, 7, 3, 7).$$

The Shekel function has four local minima and one global minimum $x_{\min} = (4, 4, 4, 4)$. This function is very flat except in neighbourhoods of the points d_1, d_2, d_3, d_4 and d_5

where it is very steep. Figure 4.2 shows the evolution of the sequence $(|x_n - x_{\min}|)$ for each gradient method (A), (B), and (C). As in the previous example, iterates of (C) do not converge while those of (A) and (B) converge very fast. Notice that in the above two examples, the speeds of convergence of algorithms (A) and (B) are very similar.

Acknowledgements. The second author would like to thank M. Teboulle (Tel-Aviv University) for fruitful discussions and pertinent remarks about the paper.

References

[1] F. Alvarez, A. Cabot: Steepest descent with curvature dynamical system, J. Optimization Theory Appl. 120 (2004) 247–273.

[2] A. Auslender: Numerical methods for nondifferentiable convex optimization, Math. Program. Study 30 (1987) 102–126.

[3] J. Barzilai, J. M. Borwein: Two point step size gradient methods, IMA J. Numer. Anal. 8 (1988) 141–148.

[4] H. Brézis, P. L. Lions: Produits infinis de résolvantes, Isr. J. Math. 29 (1978) 329–345.

[5] A. Cauchy: Méthodes générales pour la résolution des systèmes d'équations simultanées, C. R. Acad. Sci., Paris 25 (1847) 536–538.

[6] J. Céa: Optimisation. Théorie et Algorithmes, Dunod, Paris (1971).

[7] P. G. Ciarlet: Introduction à l'Analyse Numérique Matricielle et à l'Optimisation, Masson, Paris (1990).

[8] R. Cominetti: Coupling the proximal point algorithm with approximation methods, J. Optimization Theory Appl. 95 (1997) 581–600.

[9] Y. H. Dai, L. Z. Liao: R-linear convergence of the Barzilai and Borwein gradient method, IMA J. Numer. Anal. 22 (2002) 1–10.

[10] O. Güler: On the convergence of the proximal point algorithm for convex minimization, SIAM J. Control Optimization 29 (1991) 403–419.

[11] B. Lemaire: About the convergence of the proximal method, in: Advances in Optimization (Lambrecht, 1991), W. Oettli, D. Pallaschke (eds.), Lect. Notes Econ. Math. Syst. 382, Springer, Berlin (1992) 39–51.

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

[13] B. Martinet: Détermination approchée d'un point fixe d'une application pseudo-contractante, C. R. Acad. Sci., Paris, Sér. A 274 (1972) 163–165.

[14] B. T. Polyak: Introduction to Optimization, Optimization Software, New York (1987).

[15] M. Raydan: On the Barzilai and Borwein choice of steplength for the gradient method, IMA J. Numer. Anal. 13 (1993) 321–326.

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

[17] M. Teboulle: Convergence of proximal-like algorithms, SIAM J. Optim. 7 (1997) 1069–1083.