Journal of Convex Analysis
Volume 24 (2017), No. 2, 493–500

About the Gradient Projection Algorithm for a Strongly Convex Function and a Proximally Smooth Set*

Maxim V. Balashov
Department of Higher Mathematics, Moscow Institute of Physics and Technology, Institutskii pereulok 9, Dolgoprudny, Moscow region, 141700 Russia
balashov73@mail.ru

Received: March 20, 2015
Accepted: November 11, 2015

In the present work we consider the gradient projection algorithm for a strongly convex function with the Lipschitz continuous gradient and a proximally smooth (nonconvex in general) set in a real Hilbert space. We prove that the problem of minimization of such function on a proximally smooth set has unique solution if the constant of proximal smoothness of the set is sufficiently large. The considered algorithm converges with the rate of geometric progression.

Keywords: Hilbert space, strongly convex set of radius r, proximally smooth set with constant R, Lipschitz continuous gradient, gradient projection algorithm, continuous optimization

2000 Mathematics Subject Classification: Primary 46C05, 52A07, 49J52; Secondary 46N10, 90C26, 26B25

1. Introduction and main notations

We shall consider the next problem in a real Hilbert space $\mathcal{H}$:

$$\min_{x\in A} f(x), \tag{1}$$

where $f : \mathcal{H} \to \mathbb{R}$ is a strongly convex function with Lipschitz continuous gradient, and $A \subset \mathcal{H}$ is a proximally smooth (or prox-regular) set.

Let us recall some definitions and notations.

Denote by (p, x) the *inner product* for vectors $p, x \in \mathcal{H}$, $\|x\| = \sqrt{(x, x)}$. Let $B_r(a) = \{x \in \mathcal{H} \mid \|x - a\| \le r\}$. For a set $A \subset \mathcal{H}$ denote by ∂A and $\operatorname{int} A$ the *boundary and the interior* of the set A, respectively.

*The work was supported by grant RFBR 16-01-00295-a.

ISSN 0944-6532 / $ 2.50

The distance function from the point $x \in \mathcal{H}$ to the subset $A \subset \mathcal{H}$ is defined as follows

$$d_A(x) = \inf_{a \in A} \|x - a\|.$$

The set

$$P_A x = \{a \in A \mid \|x - a\| = d_A(x)\}.$$

is called the *metric projection* of the point $x \in E$ on the subset $A \subset \mathcal{H}$.

The *normal cone* to the convex closed subset $A \subset \mathcal{H}$ at the point $a \in A$ is the set

$$N(A, a) = \{p \in \mathcal{H} \mid (p, a) \geq (p, x), \ \forall x \in A\}.$$

A subset $A \subset \mathcal{H}$ is called *strongly convex of radius* R [8], if there exists a set $X \subset \mathcal{H}$ such that $A = \bigcap_{x \in X} B_R(x)$. It is obvious that if the set A is strongly convex of radius R, then it is strongly convex of radius r for any $r > R$.

Proposition 1.1 ([8], [4, Theorems 1.1, 2.6]). *Let $A \subset \mathcal{H}$ be a strongly convex set of radius $r > 0$. This is equivalent to the next property: for any $a \in \partial A$, $p \in N(A, a)$, $\|p\| = 1$, the next inclusion holds*

$$A \subset B_r\left(a - rp\right).$$

Proposition was proved in [8] in the case $\mathcal{H} = \mathbb{R}^n$.

A continuous function $f : \mathcal{H} \to \mathbb{R}$ is called *strongly convex with constant* $\varkappa > 0$ [11] if for all $x_0, x_1 \in \mathcal{H}$, $\lambda \in [0, 1]$ we have

$$f((1-\lambda)x_0 + \lambda x_1) \leq (1-\lambda)f(x_0) + \lambda f(x_1) - \frac{\lambda(1-\lambda)}{2}\varkappa\|x_0 - x_1\|^2.$$

The equivalent property for the above definition is that the function $f(x) - \frac{\varkappa}{2}\|x\|^2$ is continuous and convex.

We shall say that a function $f : \mathcal{H} \supset U \to \mathbb{R}$ has the *Lipschitz continuous gradient with constant $L > 0$ on the set U,* if for any two points $x_0, x_1 \in U$ we have

$$\|f'(x_0) - f'(x_1)\| \leq L \cdot \|x_0 - x_1\|.$$

Recall that if the function f has the Lipschitz continuous gradient with constant L on the convex subset $U \subset \mathcal{H}$, then the next inequality holds (see [9, Lemma 1.2.3])

$$|f(x_1) - f(x_0) - (f'(x_0), x_1 - x_0)| \leq \frac{L}{2}\|x_1 - x_0\|^2, \quad \forall x_0, x_1 \in U. \tag{2}$$

Also strong convexity of the function f with constant $\varkappa > 0$ gives the inequality (see [9, Definition 2.1.2 and Theorem 2.1.9])

$$f(x_1) \geq f(x_0) + (f'(x_0), x_1 - x_0) + \frac{\varkappa}{2}\|x_1 - x_0\|^2, \quad \forall x_0, x_1 \in U. \tag{3}$$

Thus, for strongly convex functions with Lipschitz continuous gradient we have, in particular, that $\varkappa \le L$.

For a function $f : \mathcal{H} \to \mathbb{R}$ and a number α define the *lower level (Lebesgue's) set*

$$\mathcal{L}_f(\alpha) = \{x \in \mathcal{H} \mid f(x) \le \alpha\}.$$

For a subset $A \subset \mathcal{H}$ and a number $R > 0$ we define the *open R-neighborhood* of the set A as follows

$$U_A(R) = \{x \in \mathcal{H} \mid d_A(x) < R\}.$$

We shall say that a subset $A \subset \mathcal{H}$ satisfies the *supporting condition with constant* $R > 0$, if for any point $x \in U_A(R)\backslash A$ with $P_A x \ne \emptyset$ and $a \in P_A x$, we have the next equality

$$A \bigcap \operatorname{int} B_R\left(a + R\frac{x-a}{\|x-a\|}\right) = \emptyset. \tag{4}$$

A closed subset $A \subset \mathcal{H}$ is called *proximally smooth with constant* R [7, 10], if the distance function $d_A(x)$ is continuously differentiable on the set $U_A(R)\backslash A$.

Proposition 1.2 ([2, Theorem 2.4], [6, Proposition 2.6], [7, Theorem 4.1], [5, Theorem 6.2]). *Let $A \subset \mathcal{H}$ be a closed subset and $R > 0$. The next conditions are equivalent:*

(*1*) *The set A satisfies the supporting condition with constant R.*
(*2*) *The set A is proximally smooth with constant R.*
(*3*) *The metric projection $P_A x$ is singleton and continuous at any point $x \in U_A(R)\backslash A$.*

We shall say that a vector $p \ne 0$ is a *proximal normal* to the proximally smooth set A (with constant of proximal smoothness R) at the point $a \in A$, if there exists $\delta \in (0, R]$ with

$$A \bigcap \operatorname{int} B_\delta\left(a + \delta\frac{p}{\|p\|}\right) = \emptyset.$$

From Proposition 1.2 it is easy to see that

$$A \bigcap \operatorname{int} B_R\left(a + R\frac{p}{\|p\|}\right) = \emptyset.$$

Proposition 1.3. *Suppose that a closed subset $A \subset \mathcal{H}$ is proximally smooth with constant $R > 0$. Then for any point $x \in U_A(R)\backslash A$*

$$P_A x = x - d_A(x) d_A'(x).$$

Proposition 1.3 follows from Proposition 1.2 and [7, Proposition 3.6(2)].

Proposition 1.4 ([7, Theorem 4.8(2)], see also [3, Theorem 2]). *Suppose that a closed subset $A \subset \mathcal{H}$ is proximally smooth with constant $R > 0$, $r \in (0, R)$. Then for any pair of points $x_0, x_1 \in U_A(r)$ and $a_i = P_A x_i$, $i = 0, 1$, the next inequality holds*

$$\|a_0 - a_1\| \le \frac{R}{R-r}\|x_0 - x_1\|.$$

Let us give few remarks concerning Proposition 1.4. Proposition 1.4 was proved in [7, Theorem 4.8(2)] for the case $x_0, x_1 \in U_A(r)\backslash A$. In the case $x_0 \notin A$, $x_1 \in A$ (and $a_0 = P_A x_0$, $a_1 = P_A x_1 = x_1$) consider z, the nearest point to x_0 of the intersection $[x_0, x_1] \cap A$. Let $\{z_k\}_{k=1}^{\infty} \subset U_r(A)\backslash A$ and $z_k \to z$; $a_k = P_A z_k$. Then $a_k \to z$. Passing to the limit $k \to \infty$ in the inequality

$$\|a_k - a_0\| \le \frac{R}{R-r}\|z_k - z\|$$

we get $\|a_0 - z\| \le \frac{R}{R-r}\|x_0 - z\|$. Hence

$$\begin{aligned}\|a_0 - a_1\| \le \|a_0 - z\| + \|z - a_1\| &\le \frac{R}{R-r}\|x_0 - z\| + \|z - a_1\| \\ &\le \frac{R}{R-r}\left(\|x_0 - z\| + \|z - a_1\|\right) = \frac{R}{R-r}\|x_0 - x_1\|.\end{aligned}$$

Proposition 1.5 ([12, Proposition 4.14 (i)]). *Suppose that a function $f : \mathcal{H} \to \mathbb{R}$ is strongly convex with constant $\varkappa > 0$. Let $\mathcal{L}_f(\alpha) \neq \emptyset$ for some $\alpha \in \mathbb{R}$ and*

$$m = \sup_{x \in \mathcal{L}_f(\alpha),\quad f'(x) \in \partial f(x)} \|f'(x)\|$$

(here by $f'(x)$ we denote a subgradient from the subdifferential $\partial f(x)$). Then the set $\mathcal{L}_f(\alpha)$ is strongly convex of radius $r = \frac{m}{\varkappa}$.

2. Gradient projection algorithm

Lemma 2.1. *Let $A \subset \mathcal{H}$ be a closed proximally smooth subset with constant $R > 0$, $f : \mathcal{H} \to \mathbb{R}$ be a strongly convex (with constant $\varkappa > 0$) differentiable function. Let $\alpha \in \mathbb{R}$, $m = \sup_{x \in \mathcal{L}_f(\alpha)} \|f'(x)\|$, $r = \frac{m}{\varkappa} < R$. Suppose that for some point $x_0 \in \mathcal{L}_f(\alpha) \cap A$ the vector $-f'(x_0) \neq 0$ is a proximal normal to the set A at the point x_0. Then the point x_0 is a (unique) solution of the problem (1).*

Proof. Put $\alpha_0 = f(x_0)$, $\alpha_0 \le \alpha$. By Proposition 1.5 the set $\mathcal{L}_f(\alpha_0)$ is strongly convex of radius $r = \frac{m}{\varkappa} < R$.

The gradient $f'(x_0)$ is an element of the normal cone to the set $\mathcal{L}_f(\alpha_0)$ at the point x_0 and from Proposition 1.1 we obtain that

$$\mathcal{L}_f(\alpha_0) \subset B_r\left(x_0 - r\frac{f'(x_0)}{\|f'(x_0)\|}\right). \tag{5}$$

The vector $-f'(x_0)$ is a proximal normal to the set A at the point x_0, hence (taking in mind Proposition 1.2)

$$A \bigcap \operatorname{int} B_R\left(x_0 - R\frac{f'(x_0)}{\|f'(x_0)\|}\right) = \emptyset. \tag{6}$$

Together formulae (5) and (6) mean that

$$\mathcal{L}_f(\alpha_0) \subset B_r\left(x_0 - r\frac{f'(x_0)}{\|f'(x_0)\|}\right) \subset \operatorname{int} B_R\left(x_0 - R\frac{f'(x_0)}{\|f'(x_0)\|}\right) \cup \{x_0\},$$

$$\mathcal{L}_f(\alpha_0) \cap A = \{x_0\}.$$

So, the point x_0 is a unique solution of the problem (1). $\square$

Lemma 2.2. *Let $\varkappa$, m, L, R be positive real numbers and $\frac{m}{\varkappa} < R$, $\varkappa \le L$. Then*

$$\frac{\varkappa - \frac{m}{R}}{L^2 - \frac{\varkappa m}{R}} < \frac{R}{m}, \qquad \frac{\varkappa - \frac{m}{R}}{L^2 - \frac{\varkappa m}{R}} \le \frac{1}{L},$$

and besides numerator and denominator of the above fraction are positive.

Proof. Numerator $\varkappa - \frac{m}{R} = \frac{\varkappa}{R}\left(R - \frac{m}{\varkappa}\right) > 0$, denominator $L^2 - \frac{\varkappa m}{R} > L^2 - \varkappa^2 \ge 0$.

Put $a = \frac{m}{R} < \varkappa \le L$. The first inequality follows from the estimates $\varkappa - a < \frac{1}{a}L^2 - \varkappa$ or $2\varkappa < \frac{1}{a}L^2 + a$ and (note that $\frac{1}{a}\,L^2 \ne a$)

$$\frac{1}{a}\,L^2 + a > 2\sqrt{\frac{L^2}{a}a} = 2L \ge 2\varkappa.$$

The second inequality is equivalent to the inequality $L\varkappa - La \le L^2 - a\varkappa$ or $L\varkappa + a\varkappa \le L^2 + La$ or $\varkappa(L+a) \le L(L+a)$. The last inequality is true. $\square$

Theorem 2.3. *Let $A \subset \mathcal{H}$ be a closed proximally smooth subset with constant of proximal smoothness $R > 0$. Let $f : \mathcal{H} \to \mathbb{R}$ be a strongly convex function with constant of strong convexity $\varkappa > 0$. Let $\alpha \in \mathbb{R}$, $\mathcal{L}_f(\alpha) \cap A \ne \emptyset$ and the function f has the Lipschitz continuous gradient with constant $L > 0$ on the set $\mathcal{L}_f(\alpha)$. Put $m = \sup_{x\in\mathcal{L}_f(\alpha)} \|f'(x)\|$. Suppose that $\frac{m}{\varkappa} < R$.*

Then for any initial point $x_1 \in A \cap \mathcal{L}_f(\alpha)$ *the iteration process*

$$x_{k+1} = P_A\left(x_k - tf'(x_k)\right), \quad t = \frac{\varkappa - \frac{m}{R}}{L^2 - \frac{\varkappa m}{R}}, \tag{7}$$

converges to the unique solution $x_0 \in A$ *of the problem* (1) *with the rate of geometric progression with common ratio*

$$q(t) = \frac{R}{R - tm}\sqrt{1 - 2t\varkappa + t^2L^2} \in (0,1), \quad t = \frac{\varkappa - \frac{m}{R}}{L^2 - \frac{\varkappa m}{R}}, \tag{8}$$

namely, for all $k \geq 1$

$$\|x_{k+1} - x_0\| \leq q(t)\|x_k - x_0\|.$$

Note that t in formula (7) gives the strict global minimum to the (nonconstant) function $q(\cdot)$ on the interval $[0, \frac{R}{m})$ and thus $q(t) < q(0) = 1$.

Proof. Recall (see (2), (3)) that $\varkappa \leq L$. Thus $\frac{m}{R} < \varkappa \leq L$ and, by Lemma 2.2, $t \in [0, \frac{R}{m})$.

Suppose that $x_k \in A \cap \mathcal{L}_f(\alpha)$. Then, using Lemma 2.2,

$$d_A(x_k - tf'(x_k)) \leq t\|f'(x_k)\| \leq tm < R,$$

and x_{k+1} is correctly defined by Proposition 1.2, moreover $\|x_{k+1} - x_k + tf'(x_k)\| = d_A(x_k - tf'(x_k))$. Also $t \leq \frac{1}{L}$ and hence

$$\begin{aligned}&\left\|x_{k+1} - \left(x_k - \frac{1}{L}f'(x_k)\right)\right\| \\ &\leq \|x_{k+1} - x_k + tf'(x_k)\| + \left(\frac{1}{L} - t\right)\|f'(x_k)\| \leq \frac{1}{L}\|f'(x_k)\|.\end{aligned}$$

From formula (2) we obtain that

$$\begin{aligned}f(x_{k+1}) &\leq f(x_k) + (f'(x_k), x_{k+1} - x_k) + \frac{L}{2}\|x_{k+1} - x_k\|^2 \\ &= f(x_k) + \frac{L}{2}\left(\left\|x_{k+1} - \left(x_k - \frac{1}{L}f'(x_k)\right)\right\|^2 - \frac{1}{L^2}\|f'(x_k)\|^2\right) \\ &\leq f(x_k) \leq \alpha\end{aligned}$$

and $x_{k+1} \in A \cap \mathcal{L}_f(\alpha)$. Thus, by induction, $x_k \in A \cap \mathcal{L}_f(\alpha)$ for all k.

From Proposition 1.4

$$\|x_{k+1} - x_k\| \leq \frac{R}{R - tm}\|x_k - tf'(x_k) - x_{k-1} + tf'(x_{k-1})\|.$$

Next, consider

$$\begin{aligned}&\|x_k - tf'(x_k) - x_{k-1} + tf'(x_{k-1})\|^2\\ &= \|x_k - x_{k-1}\|^2 - 2t(x_k - x_{k-1}, f'(x_k) - f'(x_{k-1})) + t^2\|f'(x_k) - f'(x_{k-1})\|^2.\end{aligned}$$

Using the Lipschitz continuity of the gradient f', we have

$$\|f'(x_k) - f'(x_{k-1})\| \leq L\|x_k - x_{k-1}\|.$$

From formula (3) (we should take in formula (3) $x_0 = x_k$, $x_1 = x_{k+1}$ and vice versa and then add the two obtained formulas)

$$-(x_k - x_{k-1}, f'(x_k) - f'(x_{k-1})) \leq -\varkappa\|x_k - x_{k-1}\|^2.$$

Taken together, these results give the following inequality

$$\begin{aligned}\|x_{k+1} - x_k\|^2 &\leq \frac{R^2}{(R-tm)^2}\left(1 - 2t\varkappa + t^2L^2\right)\|x_k - x_{k-1}\|^2\\ &= q^2(t)\|x_k - x_{k-1}\|^2.\end{aligned}$$

Thus, the sequence $\{x_k\}$ converges to some point $x_0 \in A \cap \mathcal{L}_f(\alpha)$.

Passing to the limit $k \to \infty$ in the formula $x_{k+1} = P_A(x_k - tf'(x_k))$, we obtain that $x_0 = P_A(x_0 - tf'(x_0))$.

If $f'(x_0) = 0$, then $x_0 \in A$ is a global minimum (unique!) of the strongly convex function f on $\mathcal{H}$.

If $f'(x_0) \neq 0$, then $-f'(x_0)$ is a proximal normal to the set A at the point $x_0 \in A$ and hence, by Lemma 2.1, x_0 is a unique solution of the problem (1).

If we shall repeat the previous speculations for estimation of the term $\|x_{k+1} - x_0\|$, we shall obtain that

$$\|x_{k+1} - x_0\| \leq q(t)\|x_k - x_0\|, \quad \forall k. \qquad \square$$

Remark 2.4. If we know the distance function $d_A(x)$ then, by Proposition 1.3, we can effectively find elements of the sequence $\{x_k\}$.

Remark 2.5. Dual (to a certain degree) theorem with strongly convex subset $A \subset \mathcal{H}$ and (nonconvex) function with the Lipschitz continuous gradient is considered in the paper [1].

References

[1] M. V. Balashov: Maximization of a function with Lipschitz continuous gradient, J. Math. Sci., New York 209(1) (2015) 12–18.

[2] M. V. Balashov, G. E. Ivanov: Weakly convex and proximally smooth sets in Banach spaces, Izv. Math. 73(3) (2009) 455–499.

[3] M. V. Balashov, G. E. Ivanov: Properties of the metric projection on weakly Vial-convex sets and parametrization of set-valued mappings with weakly convex images, Math. Notes 80(4) (2006) 461–467.

[4] M. V. Balashov, E. S. Polovinkin: M-strongly convex subsets and their generating sets, Sb. Math. 191(1) (2000) 25–60.

[5] F. Bernard, L. Thibault, N. Zlateva: Characterizations of prox-regular sets in uniformly convex Banach spaces, J. Convex Analysis 13(3–4) (2006) 525–559.

[6] A. Canino: On p-convex sets and geodesics, J. Differ. Equations 75 (1988) 118–157.

[7] F. H. Clarke, R. J. Stern, P. R. Wolenski: Proximal smoothness and lower–C^2 property, J. Convex Analysis 2(1–2) (1995) 117–144.

[8] H. Frankowska, Ch. Olech: R-convexity of the integral of the set-valued functions, in: Contributions to Analysis and Geometry (Baltimore, 1980), D. N. Clark et al. (ed.), John Hopkins Univ. Press, Baltimore(1981) 117–129.

[9] Yu. Nesterov: Inroductory Lectures on Convex Optimization: A Basic Course, Kluwer, Boston (2004).

[10] R. A. Poliquin, R. T. Rockafellar, L. Thibault: Local differentiability of distance functions, Trans. Amer. Math. Soc. 352 (2000) 5231–5249.

[11] B. T. Polyak: Existence theorems and convergence of minimizing sequences in extremum problems with restrictions, Sov. Math., Dokl. 7 (1966) 72–75.

[12] J.-Ph. Vial: Strong and weak convexity of sets and functions, Math. Oper. Res. 8(2) (1983) 231–259.