МАТЕМАТИЧКИ ВЕСНИК
49 (1997), 245–248

UDK 517.982.2
оригинални научни рад
research paper

ON CONCTRACTIBILITY OF THE OPERATOR $I - t\nabla f$

Vladimir Janković and Milan Jovanović

Abstract. We study the set $K(f)$ of positive numbers t for which the operator $I - t\nabla f$ is contractible, where f is a differentiable function defined on a convex subset of the Hilbert space (I is the identity operator of that Hilbert space). The set $K(f)$ is interesting for a problem of minimization of strongly convex functions when the method of contractible mappings is applied.

1. Introduction

Let C be a convex subset of the Hilbert space H and let $f\colon C \to R$ be a differentiable function on the set C. (That means f can be extended to some open superset D of the set C, so that its extension is differentiable in every point of the set C.) The function f is a strongly convex function if there exists a positive number r such that

$$f((1-\lambda)x+\lambda y) \leq (1-\lambda)f(x)+\lambda f(y)-(1-\lambda)\lambda r\|x-y\|^2, \quad x,y \in C, 0 \leq \lambda \leq 1. \tag{1}$$

The number r satisfying (1) is called a constant of strong convexity of the function f. Maximal constant of strong convexity is denoted by r_0. The gradient ∇f satisfies Lipschitz condition if there exists a nonnegative number L such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|, \quad x,y \in C. \tag{2}$$

The number L satisfying (2) is called a Lipschitz constant of the gradient ∇f. Minimal Lipschitz constant is denoted by L_0. The operator $I - t\nabla f$ is contractible if there exists a number q, $0 < q < 1$, such that

$$\|(x - t\nabla f(x)) - (y - t\nabla f(y))\| \leq q\|x-y\|, \quad x,y \in C. \tag{3}$$

We introduce the following set

$$K(f) = \{\, t > 0 \mid \text{operator } I - t\nabla f \text{ is contractible} \,\}.$$

AMS Subject Classification: 26 B 25

Communicated at the 4th Symposium on Mathematical Analysis and Its Applications, Aranđelovac 1997.

The set $K(f)$ is interesting for the method of contractible mappings and our primary concern is to determine it.

In the sequel the following known theorems will be used (their statements can be found in [1]).

THEOREM A. *Positive number r is a constant of strong convexity of the function f if and only if*

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 2r\|x-y\|^2, \quad x, y \in C. \tag{4}$$

THEOREM B. *If the gradient of the function f satisfies*

$$\|\nabla f(x) - \nabla f(y)\|^2 \leq L\langle \nabla f(x) - \nabla f(y), x - y \rangle, \quad x, y \in C, \tag{5}$$

then f is convex and its gradient satisfies the Lipschitz condition (2).

THEOREM C. *If* $\operatorname{int} C \neq \phi$, *$f$ is convex, and the gradient of f satisfies the Lipschitz condition (2), then the condition (5) is fulfilled.*

Theorem A can be found in the third paragraph of the fourth chapter of [1]. In the second paragraph of the same chapter there is a theorem according to which a function f satisfies the condition (5) if and only if it is convex and its gradient satisfies the Lipschitz condition (2). However, the convexity of f together with the Lipschitz condition (2) do not imply the condition (5) in the general case. That can be shown using the function from the example in the next paragraph of this article. Actually, there is a mistake in the last part of the proof of that theorem (the step from the case $\operatorname{int} C \neq \phi$ to the general case). All statements from [1] that we use here are concerned with the finite dimensional Hilbert space, but they can easily be generalized to arbitrary Hilbert space.

2. Properties of the set $K(f)$

THEOREM 1. *If the set $K(f)$ is non-empty then f is strongly convex and its gradient satisfies the Lipschitz condition.*

Proof. Let $t > 0$ and $0 < q < 1$. Then (3) is equivalent to

$$\begin{aligned}(1-q^2)\|x-y\|^2 + t^2\|\nabla f(x) - \nabla f(y)\|^2 &\leq \\ &\leq 2t\langle \nabla f(x) - \nabla f(y), x-y\rangle, \quad x, y \in C. \end{aligned}\tag{6}$$

If $t \in K(f)$, then there exists q, $0 < q < 1$, such that (3) holds and consequently (6) holds too. It follows that

$$(1-q^2)\|x-y\|^2 \leq 2t\langle \nabla f(x) - \nabla f(y), x-y\rangle, \quad x, y \in C,$$

so that (4) is fulfilled for $r = (1-q^2)/4t$. From (6) it follows that

$$t\|\nabla f(x) - \nabla f(y)\|^2 \leq 2\langle \nabla f(x) - \nabla f(y), x-y\rangle, \quad x, y \in C. \tag{7}$$

This implies that (5), and consequently (2), holds with $L = 2/t$. ∎

THEOREM 2. *If f is a strongly convex function and if its gradient satisfies the Lipschitz condition, then*

$$\left(0, \frac{4r_0}{L_0^2}\right) \subseteq K(f) \subseteq \left(0, \frac{2}{L_0}\right).$$

Proof. Since (2) holds for $L = L_0$, then (6) is a consequence of the inequality

$$\left(\frac{1-q^2}{2t} + \frac{tL_0^2}{2}\right)\|x-y\|^2 \leq \langle \nabla f(x) - \nabla f(y), x-y\rangle, \quad x, y \in C. \tag{8}$$

If $0 < t < 4r_0/L_0^2$, then $tL_0^2/2 < 2r_0$ and there exists q, $0 < q < 1$, such that $(1-q^2)/2t + tL_0^2/2 \leq 2r_0$ holds. For such q (8) is valid, and therefore (6) i.e. (3) hold; therefore $t \in K(f)$.

Suppose $t \in K(f)$. For some q, $0 < q < 1$, the inequality (6) is valid. From this inequality it follows (7), and from the inequality (7) it follows (5) i.e. (2) with $L = 2/t$. Consequently, $L_0 \leq 2/t$, i.e. $t \leq 2/L_0$. ∎

THEOREM 3. *Suppose* $\operatorname{int} C \neq \phi$. *If f is a strongly convex function and if its gradient satisfies the Lipschitz condition, then*

$$K(f) = \left(0, \frac{2}{L_0}\right).$$

Proof. Suppose conditions (4) and (2) hold. Then $f(x) - r\|x\|^2$ is a convex function and the Lipschitz condition is fulfilled for its gradient with the constant $L + 2r$. If we apply Theorem C we get

$$\begin{aligned} 2r(L+4r)\|x-y\|^2 + \|\nabla f(x) - \nabla f(y)\|^2 &\leq \\ \leq (L+6r)\langle \nabla f(x) - \nabla f(y), x-y\rangle&, \quad x, y \in C. \end{aligned} \tag{9}$$

Let $t < 2/L_0$. It can be shown that there exists r, $0 < r \leq r_0$, such that $t \leq 2/(L_0 + 6r)$ $(2/(L_0+6r) \to 2/L_0, r \to 0_+)$. Also there exists a number $L \geq L_0$ such that $t = 2/(L+6r)$. Put $q = (L+2r)/(L+6r)$. Then $0 < q < 1$. For such L, r and q the inequality (9) reduces to (6). It follows that $t \in K(f)$. ∎

NOTE. If $\operatorname{int} C = \phi$ but $\operatorname{ri} C \neq \phi$, then the set $K(f)$ can be determined by applying the above theorem on the function f with the domain C which is considered as the subset of $\operatorname{aff} C$. In that case it is useful to have in mind that the gradient of the restriction of the function f on some subspace is a projection of the gradient of the function f onto this subspace. If the dimension of the space H is finite, the condition $\operatorname{ri} C \neq \phi$ is fulfilled.

EXAMPLE. Suppose $H = R^2$, $C = \{\,(x,x) \mid x \in R\,\}$ and the function f is given by

$$f(x,y) = ax^2 + by^2, \quad a + b > 0.$$

Then $\nabla f(x,y) = (2ax, 2by)$, $\|\nabla f(x,x) - \nabla f(u,u)\|^2 = 4(a^2+b^2)(x-u)^2$, $\langle \nabla f(x,x) - \nabla f(u,u), (x,x)-(u,u)\rangle = 2(a+b)(x-u)^2$, $\|(x,x)-(u,u)\|^2 = 2(x-u)^2$. It follows that $L_0 = \sqrt{2(a^2+b^2)}$, $r_0 = \frac{a+b}{2}$. Since

$$\|[(x,x) - t\nabla f(x,x)] - [(u,u) - t\nabla f(u,u)]\|^2 = [(1-2at)^2 + (1-2bt)^2](x-u)^2,$$

then $t \in K(f)$ if and only if $(1-2at)^2 + (1-2bt)^2 < 2$, which is equivalent to $t < \dfrac{a+b}{a^2+b^2}$. It follows that

$$K(f) = \left(0, \frac{a+b}{a^2+b^2}\right) = \left(0, \frac{4r_0}{L_0^2}\right).$$

3. Method of contractible mappings

THEOREM 4. *Let H be a Hilbert space and let $f\colon H \to R$ be a strongly convex function. If f is differentiable and if its gradient satisfies the Lipschitz condition (2), then the sequence (x_n) of the points from H, satisfying*

$$x_{n+1} = x_n - t\nabla f(x_n), n = 1, 2, 3, \dots,$$

converges to $\hat{x}$, the point in which f reaches the minimum, providing $0 < t < 2/L$.

This theorem is the consequence of the Theorem 3. In [2] the convergence is proved under the condition that $0 < t < 4r/M^2$, where M is the upper bound of the norm of the second derivative of f. In [3] the convergence is proved under slightly more general assumption that $0 < t < 4r/L^2$. According to the Theorem 2 we have that $4r/L^2 \le 2/L$, and hence the condition $0 < t < 2/L$, which provides convergence in the Theorem 4, is weaker than the corresponding conditions in older theorems.

REFERENCES

[**1**] Ф. П. Васильев, *Численные методы решения екстремальных задач*, Наука, Москва, 1980.

[**2**] Ж. Сеа, *Оптимизация. Теория и алгоритмы*, Мир, Москва, 1973. (translated from French)

[**3**] D. Blebea, G. Dinca, *Remarque sur une methode de contraction a minimiser les fonctionnelles convexes sur les espaces de Hilbert*, Bull. Math. Soc. Sci. Math. R. S. Roumanie (N.S.), **23(71)**, 3, 1979, 227–229.

(received 11.09.1997.)

Vladimir Janković, Mathematical Faculty, Studentski trg 16, 11000 Beograd

Milan Jovanović, Faculty for Electrical Engineering, Patre 5, 78000 Banja Luka