Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 2, No. 2, pp. 319–326, May 2006

PARTIAL LINEARIZATION METHOD FOR CONVEX OPTIMIZATION PROBLEMS IN BANACH SPACES

M.S.S. Ali

Abstract: In this paper, we consider an optimization problem in Banach space, whose cost function can be represented as the sum of a convex and a uniformly convex functions. We propose to solve this problem by a partial linearization method and prove its strong convergence to a solution. We show that the method has certain advantages over the usual gradient methods.

Key words: *convex optimization, Banach spaces, partial linearization method, strong convergence*

Mathematics Subject Classification: *46E15, 49M37, 65K10, 90C30*

---

1 Introduction

A great number of problems arising in mathematical physics, economics, engineering and other fields can be formulated as the optimization problem: Find a point x^* in a convex and closed subset D of a Banach space E such that

$$\psi(x^*) \le \psi(x) \quad \forall x \in D,$$

or briefly,

$$\min_{x \in D} \to \psi(x), \tag{1}$$

where $\psi : E \to R$ is a convex function; see e.g. [1], [2] and [8] and the references therein. It is well known that taking into account the essential features of the problem under consideration may enhance properties of the solution method essentially. To obtain strong convergence of iterative methods to a solution of problem (1) one needs certain strengthened convexity properties, otherwise, a regularization type approach is necessary; see e.g. [9] and [10]. In this paper, we intend to construct an iterative method for problem (1) where ψ can be represented as the sum of a convex and a uniformly convex functions without using any auxiliary functions as in projection type methods; see [1] and [8]. The method belongs to a class of partial linearization ones. Such methods are known to be efficient if the problem has a special structure, however, only weak convergence results were established for them; see [4], [6] and [11]. We propose to combine the partial linearization technique with an Armijo type linesearch and prove the strong convergence result for a Banach space setting. An example of applications of the method is also given.

2 Properties of Uniformly Convex Optimization Problems

In this section, we give preliminary properties of rather a general class of convex functions. First we recall several definitions. A function $h : E \to R$ is said to be uniformly convex (see [5]) if there exists a continuously increasing function $\theta : R \to R$ such that $\theta(0) = 0$ and that for all $x, y \in E$ and for each $\lambda \in [0, 1]$, we have

$$h(\lambda x + (1-\lambda)y) \leq \lambda h(x) + (1-\lambda)h(y) - \lambda(1-\lambda)\theta(\|x-y\|)\|x-y\|.$$

If $\theta(\tau) = \kappa\tau$ for $\kappa > 0$, then h is called a strongly convex function. One can see that the class of uniformly convex functions is rather broad. These functions possess several very useful properties (see [5] and also [10]), which are listed in the following proposition.

Proposition 2.1. *Suppose that $h : E \to R$ is a uniformly convex and lower semicontinuous function. Then:*

(i) h is bounded from below on E;

(ii) for each μ, the level set $X_\mu = \{x \in E \mid h(x) \leq \mu\}$ is bounded;

(iii) If h is differentiable, then, for each pair of points $x, y \in E$, we have

$$h(y) - h(x) \geq \langle \nabla h(x), y - x\rangle + \theta(\|x-y\|)\|x-y\| \tag{2}$$

and

$$\langle \nabla h(y) - \nabla h(x), y - x\rangle \geq 2\theta(\|x-y\|)\|x-y\|. \tag{3}$$

These properties allow us to easily deduce the existence and uniqueness results of solutions for uniformly convex optimization problems.

Proposition 2.2. *[10, Chapter 1, Section 3, Theorem 9] Suppose that D is a nonempty, convex and closed subset of a reflexive Banach space $E, h : E \to R$ is a uniformly convex and lower semicontinuous function. Then the optimization problem*

$$\min_{x \in D} \to h(x) \tag{4}$$

has the unique solution x^*, moreover,

$$h(x) - h(x^*) \geq \theta(\|x - x^*\|)\|x - x^*\| \quad \forall x \in D. \tag{5}$$

In what follows, we shall use the following basic assumptions on problem (1).

(A1) *D is a nonempty, convex, and closed subset of a reflexive Banach space $E; \psi : E \to R$ is of the form*

$$\psi(x) = f(x) + \varphi(x).$$

(A2) *$f : E \to R$ is convex and has the Lipschitz continuous gradient map $\nabla f : E \to E^*$, where E^* is the conjugate space; $\varphi : E \to R$ is uniformly convex and has the Lipschitz continuous gradient map $\nabla \varphi : E \to E^*$.*

We have introduced the differentiability assumption for substantiation of the method. Note that the differentiability implies the lower semicontinuity of f and φ, so that ψ is a uniformly convex and lower semicontinuous function. That is, we consider the optimization problem

$$\min_{x \in D} \to f(x) + \varphi(x). \tag{6}$$

3 Properties of Auxiliary Problems

In this section, we establish several results which will be used for construction and substantiation of an iterative solution method for problem (6). Recall that, on account of Proposition 2.2, this problem has a unique solution, if assumptions **(A1)** and **(A2)** are fulfilled.

We start our considerations from the standard equivalence results for problem (6), which are modifications of those in [2, Chapter 2, Propositions 2.1 and 2.2].

Lemma 3.1. *Suppose* **(A1)** *and* **(A2)** *are fulfilled. Then problem (6) or (1) is equivalent to each of the following variational inequalities:*

(i) Find $x^* \in D$ *such that*

$$\langle \nabla f(x^*), x - x^* \rangle + \varphi(x) - \varphi(x^*) \geq 0 \quad \forall x \in D, \tag{7}$$

and

(ii) Find $x^* \in D$ *such that*

$$\langle \nabla f(x^*) + \nabla \varphi(x^*), x - x^* \rangle \geq 0 \quad \forall x \in D. \tag{8}$$

Proof. Obviously, (8) is the classical necessary and sufficient condition of optimality for (6) (see e.g. [1], Chapter I, Theorem 0.4). Utilizing the standard inequality for the differentiable convex function φ:

$$\varphi(x) - \varphi(x^*) \geq \langle \nabla \varphi(x^*), x - x^* \rangle$$

in (8) (see also (2)), we obtain the implication (8)$\Rightarrow$ (7). Utilizing the same property with respect to f in (7), we obtain (7) $\Rightarrow$ (6), but (6) $\Leftrightarrow$ (8). Therefore, the assertion is true.

Let us introduce the auxiliary function

$$\Phi(x, y) = \langle \nabla f(x), y - x \rangle + \varphi(y) - \varphi(x).$$

Since $\Phi(x, \cdot)$ is clearly uniformly convex and lower semicontinuous under **(A2)**, the optimization problem

$$\min_{y \in D} \to \Phi(x, y) \tag{9}$$

has a unique solution, which will be denoted by $y(x)$. Observe that the function Φ in (9) can be in principle replaced with the simplified expression

$$\tilde{\Phi}(x, y) = \langle \nabla f(x), y \rangle + \varphi(y).$$

Anyway, taking into account Lemma 3.1, we obtain immediately the fixed point characterization of the solution of (6).

Proposition 3.1. *If* **(A1)** *and* **(A2)** *are fulfilled, then* x^* *is a solution of (6) if and only if* $x^* = y(x^*)$.

Observe that

$$\Psi(x) = -\min_{y \in D} \Phi(x, y) = -\Phi(x, y(x))$$

can be regarded as the primal gap function for the problem (7). The approach based on the auxiliary problem (9) enables us to avoid including additional parameters and functions in descent methods and to propose very flexible computational schemes.

We can specialize and strengthen the previous result by establishing an error bound for the auxiliary problem (6).

Proposition 3.2. *Suppose* **(A1)** *and* **(A2)** *are fulfilled. Then, we have*

$$2\theta(\|x-x^*\|) \le (L_f + 2L_\varphi)\|x-y(x)\| \quad \forall x \in D, \tag{10}$$

where L_f and L_φ are the Lipschitz constants for ∇f and $\nabla\varphi$, respectively.

Proof. Fix $x \in D$ and for brevity set $y = y(x)$. Then, due to Lemma 3.1, we have

$$\langle \nabla\psi(x^*), y - x^*\rangle \ge 0$$

and

$$\langle \nabla f(x) + \nabla\varphi(y), x^* - y\rangle \ge 0, \tag{11}$$

where the second inequality follows from the same optimality criterion applied to problem (9). Adding these inequalities gives

$$\langle \nabla\psi(x^*) - \nabla\psi(x), y - x^*\rangle + \langle \nabla\varphi(y) - \nabla\varphi(x), x^* - y\rangle \ge 0.$$

Using **(A2)** and Proposition 2.1 (iii), we have

$$\begin{aligned} & 2\theta(\|x-x^*\|)\|x-x^*\| \le \langle \nabla\psi(x) - \nabla\psi(x^*), x - x^*\rangle \\ & = \langle \nabla\psi(x) - \nabla\psi(x^*), x - y\rangle + \langle \nabla\psi(x) - \nabla\psi(x^*), y - x^*\rangle. \end{aligned}$$

Applying now the above inequality and the monotonicity of $\nabla\varphi$ yields

$$\begin{aligned} 2\theta(\|x-x^*\|)\|x-x^*\| &\le \langle \nabla\psi(x) - \nabla\psi(x^*), x - y\rangle + \langle \nabla\varphi(y) - \nabla\varphi(x), x^* - x\rangle \\ &\quad + \langle \nabla\varphi(y) - \nabla\varphi(x), x - y\rangle \\ &\le \langle \nabla f(x) - \nabla f(x^*), x - y\rangle + \langle \nabla\varphi(x) - \nabla\varphi(x^*), x - y\rangle \\ &\quad + \langle \nabla\varphi(y) - \nabla\varphi(x), x^* - x\rangle \\ &\le L_f\|x-x^*\|\|x-y\| + L_\varphi\|x-x^*\|\|x-y\| \\ &\quad + L_\varphi\|y-x\|\|x^*-x\|, \end{aligned}$$

hence

$$2\theta(\|x-x^*\|) \le (L_f + 2L_\varphi)\|x-y\|,$$

and the result follows.

So, if x is not optimal, then $x \ne y(x)$ and we can may try to obtain the basic descent property which utilizes the direction $y(x) - x$.

Proposition 3.3. *Suppose* **(A1)** *and* **(A2)** *are fulfilled. Then, we have*

$$\begin{aligned} \langle \nabla\psi(x), y(x) - x\rangle &\le -\langle \nabla\varphi(y(x)) - \nabla\varphi(x), y(x) - x\rangle \\ &\le -2\theta(\|y(x)-x\|)\|y(x)-x\| \quad \forall x \in D. \end{aligned} \tag{12}$$

Proof. Fix $x \in D$ and again set, for brevity, $y = y(x)$. Then, writing the optimality condition for the problem (9), we have

$$\langle \nabla f(x) + \nabla\varphi(y), z - y\rangle \ge 0, \quad \forall z \in D, \tag{13}$$

(cf. (11)). Setting $z = x$ in this inequality gives

$$\langle \nabla\psi(x), x - y\rangle + \langle \nabla\varphi(y) - \nabla\varphi(x), x - y\rangle \ge 0.$$

Thus the first inequality in (12) holds. The second inequality now follows from Proposition 2.1 (iii).

Additionally, we give the continuity property for the mapping $x \mapsto y(x)$.

Proposition 3.4. *If* **(A1)** *and* **(A2)** *are fulfilled, then $x \mapsto y(x)$ is continuous.*

Proof. Take arbitrary points $x', x'' \in D$ and set $y' = y(x')$, $y'' = y(x'')$. Using the optimality condition (13) gives

$$\langle \nabla f(x') + \nabla \varphi(y'), y'' - y' \rangle \geq 0$$

and

$$\langle \nabla f(x'') + \nabla \varphi(y''), y' - y'' \rangle \geq 0.$$

Adding these inequalities and taking into account (3), we obtain

$$\begin{aligned} \langle \nabla f(x') - \nabla f(x''), y'' - y' \rangle &\geq \langle \nabla \varphi(y'') - \nabla \varphi(y'), y'' - y' \rangle \\ &\geq 2\theta(\|y'' - y'\|)\|y'' - y'\|, \end{aligned}$$

hence

$$L_f \|x' - x''\| \geq 2\theta(\|y' - y''\|),$$

which implies the continuity of $x \mapsto y(x)$.

4 Descent Method for Convex Optimization

The properties of the mapping $x \mapsto y(x)$ established in the previous section enable us to develop a descent partial linearization algorithm for solving the problem (1) (or, equivalently, (6)). Unlike the usual algorithmic schemes (see e.g. [3, 7]) it does not involve any auxiliary functions.

Algorithm (PL).
Step 0. Choose a point $x^0 \in D$ and numbers $\alpha \in (0,1)$ and $\beta \in (0,1)$. Set $i = 0$.
Step 1. Compute $y^i = y(x^i)$ and set $d^i = y^i - x^i$.
Step 2. Find m as the smallest nonnegative integer such that

$$\psi(x^i + \beta^m d^i) \leq \psi(x^i) - \alpha\beta^m \langle \nabla \varphi(y^i) - \nabla \varphi(x^i), d^i \rangle, \tag{14}$$

set $\lambda_i = \beta^m, x^{i+1} = x^i + \lambda_i d^i, i = i + 1$ and go to Step 1.

The next theorem presents a convergence result for Algorithm (PL).

Theorem 4.1. *Suppose that assumptions* **(A1)** *and* **(A2)** *are satisfied. Then any sequence $\{x^i\}$, generated by Algorithm (PL), converges strongly to the unique solution x^* of problem (1)(or(6)).*

Proof. It has been mentioned that (1)(or(6)) has a unique solution due to Proposition 2.2. By **(A2)**, $\nabla \psi$ is Lipschitz continuous with constant $L = L_f + L_\varphi$, where L_f and L_φ are the corresponding Lipschitz constants for ∇f and $\nabla \varphi$, respectively. It follows that the well-known inequality (see e.g. [10, Chapter 2, Section 3]) holds:

$$\psi(x^i + \lambda d^i) - \psi(x^i) \leq \lambda \langle \nabla \psi(x^i), d^i \rangle + 0.5\lambda^2 L \|d^i\|^2.$$

In view of (12), we have

$$\psi(x^i + \lambda d^i) - \psi(x^i) \leq -\lambda \langle \nabla \varphi(y^i) - \nabla \varphi(x^i), d^i \rangle + 0.5\lambda^2 L \|d^i\|^2.$$

If

$$\langle \nabla \varphi(y^i) - \nabla \varphi(x^i), d^i \rangle - 0.5\lambda L \|d^i\|^2 \geq \alpha \langle \nabla \varphi(y^i) - \nabla \varphi(x^i), d^i \rangle \tag{15}$$

is satisfied for a positive λ, then the linesearch procedure in Algorithm (PL) becomes implementable. However, (15) is equivalent to

$$\lambda \leq 2(1-\alpha)\langle \nabla\varphi(y^i) - \nabla\varphi(x^i), d^i\rangle/(L\|d^i\|^2)$$

and (see (12)) we have

$$\langle \nabla\varphi(y^i) - \nabla\varphi(x^i), d^i\rangle \geq 2\theta(\|d^i\|)\|d^i\|.$$

So that (15) is satisfied if

$$\lambda \leq 4(1-\alpha)\theta(\|d^i\|)/(L\|d^i\|).$$

By **(A1)** and **(A2)**, ψ is uniformly convex, hence it is bounded from below on D and the level set $D_0 = \{x \in D \mid \psi(x) \leq \psi(x^0)\}$ is also bounded because of Proposition 2.1. But the sequence $\{\psi(x^i)\}$ is nonincreasing due to (14), hence, by Proposition 3.4,

$$\|d^i\| \leq C < +\infty \quad \text{for} \quad i = 0, 1, \ldots$$

Suppose that

$$\|d^i\| \geq l > 0 \quad \text{for} \quad i = 0, 1, \ldots$$

Choosing $\lambda \leq \lambda' = 4(1-\alpha)\theta(l)/(LC)$, we see that

$$\psi(x^i + \lambda d^i) - \psi(x^i) \leq -2\alpha\lambda\theta(\|d^i\|)\|d^i\|,$$

i.e., $\lambda_i \geq \lambda'' = \min\{\beta, \beta\lambda'\}$. It follows that

$$\psi(x^i) - \psi(x^{i+1}) \geq 2\alpha\lambda''\theta(\|d^i\|)\|d^i\| \to 0$$

as $i \to +\infty$, which is a contradiction. Therefore, there exists a subsequence $\{i_s\}$ such that $\|d^{i_s}\| \to 0$ as $i_s \to +\infty$, i.e. $\|y(x^{i_s}) - x^{i_s}\| \to 0$, and, by (10), $\|x^{i_s} - x^*\| \to 0$ as $i_s \to +\infty$. It means that x^* is a strong limit point of $\{x^i\}$. Since $\{\psi(x^i)\}$ is nonincreasing, it means that

$$\lim_{i\to\infty} \psi(x^i) = \psi(x^*).$$

Applying now (5) , we conclude that the whole sequence $\{x^i\}$ converges strongly to x^* and the result follows.

Being based on the above result, we can approximate the solution x^* of problem (1)(or(6)) with any prescribed accuracy in a finite number of iterations.

5 Example of Applications

Let us consider problem (1) satisfying assumptions **(A1)** and **(A2)**, which admits an additional partition of the space. More precisely, suppose that the initial space is a Cartesian product of spaces, i.e.,

$$E = E_1 \times \cdots \times E_m,$$

correspondingly,

$$D = D_1 \times \cdots \times D_m,$$

where D_i is a convex closed subset of E_i, and that

$$\varphi(x) = \sum_{i=1}^{m} \varphi(x_i), \tag{16}$$

where $x_i \in E_i, E_i$ is a real reflexive Banach space. Then the basic problem (9) of finding the point $y(x)$ can be replaced by a sequence of m independent subproblems:

$$\min_{y_i \in D_i} \to \langle (\nabla f(x))_i, y_i \rangle + \varphi_i(y_i) \tag{17}$$

with the solutions $y_i(x)$, where

$$\nabla f(x) = (\nabla f(x))_1 \times \cdots \times (\nabla f(x))_m$$

is the corresponding partition of the gradient of f, i.e., $(\nabla f(x))_i$ is in the conjugate space E_i^*. Clearly, each function φ_i in (16) is uniformly convex, hence, each problem (17) has a unique solution. The simplest form of subproblem (17) is obtained in case $E = R^m$, then $E_i = R$ for $i = 1, .., m$ and (17) is an one-dimensional problem, whose solution may be found easily.

This approach can be extended for more general classes of problems. For instance, let us consider the problem of minimization a convex function f subject to the following constraints:

$$x \in D = D_1 \times \cdots \times D_, m$$
$$h_k(x) \leq 0 \ k \in K_1, \ h_k(x) = 0 \ k \in K_2;$$

where h_k are separable functions corresponding to the structure of the set D. By utilizing the Lagrangian multipliers methods we can replace this problem by a sequence of problems of form (6) and make use of a partial linearization method described to find their solution in primal variables with addition regularization terms if necessary.

6 Conclusions

Thus, we have presented a general method with Armijo linesearch for convex optimization problems in Banach spaces and proved its strong convergence. Although this method is used for differentiable problems, the above approach may be viewed as a basis for non-differentiable problems.

Acknowledgement

The author is grateful to Prof. I.V. Konnov for his suggestion of the subject of this paper and for useful discussions of the results.

References

[1] J. Cea, *Optimization: Theorie et Algorithms*, Dunod, Paris, 1971.

[2] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, 1976.

[3] V.G. Karmanov, *Mathematical Programming*, Nauka, Moscow, 1986 (in Russian).

[4] T. Larsson and A. Migdalas, An algorithm for nonlinear programs over Cartesian product sets, *Optimization* 21 (1990) 535–542.

[5] E.S. Levitin and B.T. Polyak, Constrained minimization methods, *USSR Comput. Mathem. and Physics* 6 (1966) 1–50.

[6] H. Mine and M. Fukushima, A minimization method for the sum of a convex function and a continuously differentiable function, *J. Optim. Theory Appl.* 33 (1981) 9–23.

[7] M. Patriksson, Cost approximation: a unified framework of descent algorithms for nonlinear programs, *SIAM J. Optim.* 8 (1998) 561–582.

[8] B.T. Polyak, *Introduction to Optimization*, Optimization Software, New York, 1987.

[9] A.N. Tikhonov and V.Ya. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley and Sons, New York, 1977.

[10] F.P. Vasil'yev, *Methods of Solution of Extremal Problems*, Nauka, Moscow, 1981 (in Russian).

[11] D.L. Zhu and P. Marcotte, Coupling the auxiliary problem principle with descent methods of pseudoconvex programming, *Eur. J. Oper. Res.* 83 (1995) 670–685.

Manuscript received 13 September 2005
revised 10 February 2006
accepted for publication 13 February 2006

M.S.S. Ali
Department of Mathematics, Faculty of Education, Ain Shams University, Cairo, Egypt
E-mail address: @mssali5@hotmail.com