

DESCENT METHODS FOR CONVEX OPTIMIZATION PROBLEMS IN BANACH SPACES

M. S. S. ALI

Received 3 November 2004 and in revised form 22 July 2005

We consider optimization problems in Banach spaces, whose cost functions are convex and smooth, but do not possess strengthened convexity properties. We propose a general class of iterative methods, which are based on combining descent and regularization approaches and provide strong convergence of iteration sequences to a solution of the initial problem.

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

$$f(x^*) \leq f(x) \quad \forall x \in D, \quad (1.1)$$

or briefly,

$$\min_{x \in D} \longrightarrow f(x), \quad (1.2)$$

where $f : E \rightarrow R$ is a convex function; see, for example, [1, 2, 8] and the references therein. It is well known that standard iterative methods for solving (1.2), which are designed for finite-dimensional optimization problems, cannot guarantee strong convergence of their iteration sequences to a solution of the initial problem if the cost function does not possess strengthened convexity properties such as strong convexity. Usually, these methods provide only weak convergence to a solution. However, such a convergence is not satisfactory for many real problems, which are ill-posed in general, since even small perturbations of the initial data may cause great changes in solutions. These questions are crucial for developing stable solution methods. Strong convergence ensuring stability and continuous dependence of the initial data can be obtained via the regularization approach

(see [9]), when the initial problem is replaced with a sequence of perturbed well-posed problems. There exist a great number of applied problems in physics, mechanics, optimal control, economics, engineering, and other fields whose stable solutions require the regularization approach; see [9, 10, 11] and the references therein.

However, each perturbed problem has nonlinear optimization form and its solution can be found only approximately. In order to avoid the difficulty of solving perturbed problems within a prescribed accuracy, some other regularization techniques, such as averaging and iterative regularization, were proposed for obtaining implementable algorithms with strong convergence of their iteration sequences (see, e.g., [11]). However, due to their divergent series step-size rules, convergence may be very slow.

Very recently, another approach to constructing implementable and strongly convergent algorithms for variational inequalities was proposed in [4, 5]. This approach is based on applying a descent algorithm for perturbed problems which are obtained by adding the identity map, and introducing artificial gap functions, which allow for estimating error bounds.

In this paper, motivated by the above approach, we intend to present a rather general class of implementable algorithms for solving the convex optimization problem (1.2) in a reflexive Banach space. These algorithms have a flexible structure in the sense that they can be easily adjusted for essential features of each particular problem under consideration. At the same time, they provide the strong convergence of the iteration sequence if the initial problem is solvable. Namely, we introduce a general class of perturbation functions and utilize its properties for designing gap functions and descent methods.

2. Regularization of optimization problems

In this section, we give a strong convergence result of approximate solutions of perturbed problems. First we recall several definitions and auxiliary properties. A function $\psi : E \rightarrow R$ is said to be uniformly convex (see [6]) if there exists a continuously increasing function $\theta : R \rightarrow R$ such that $\theta(0) = 0$ and that for all $x, y \in E$ and for each $\lambda \in [0, 1]$, we have

$$\psi(\lambda x + (1 - \lambda)y) \leq \lambda\psi(x) + (1 - \lambda)\psi(y) - \lambda(1 - \lambda)\theta(\|x - y\|)\|x - y\|. \quad (2.1)$$

If $\theta(\tau) = \kappa\tau$ for $\kappa > 0$, then ψ 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 [6] and also [10]), which are listed in the following proposition.

PROPOSITION 2.1. *Suppose that $\psi : E \rightarrow R$ is a uniformly convex and lower semicontinuous function. Then,*

- (i) ψ is bounded from below on E ;
- (ii) for each μ , the level set $X_\mu = \{x \in E \mid \psi(x) \leq \mu\}$ is bounded;
- (iii) if ψ is differentiable, then, for each pair of points $x, y \in E$,

$$\psi(y) - \psi(x) \geq \langle \nabla \psi(x), y - x \rangle + \theta(\|x - y\|)\|x - y\|, \quad (2.2)$$

$$\langle \nabla \psi(y) - \nabla \psi(x), y - x \rangle \geq 2\theta(\|x - y\|)\|x - y\|. \quad (2.3)$$

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 , $\psi : E \rightarrow R$ is a uniformly convex and lower semicontinuous function. Then the optimization problem*

$$\min_{x \in D} \psi(x) \tag{2.4}$$

has the unique solution x^* , moreover,

$$\psi(x) - \psi(x^*) \geq \theta(\|x - x^*\|)\|x - x^*\| \quad \forall x \in D. \tag{2.5}$$

In what follows, we will use the following basic assumptions on problem (1.2).

- (A1) D is a nonempty, convex, and closed subset of a reflexive Banach space E .
- (A2) $f : E \rightarrow R$ is a convex and lower semicontinuous function.
- (A3) The solution set D^* of problem (1.2) is nonempty.

In order to apply the Tikhonov regularization approach to problem (1.2), we introduce an auxiliary function $\varphi : E \rightarrow R$, which satisfies the following property.

- (B1) $\varphi : E \rightarrow R$ is a uniformly convex and lower semicontinuous function.

Now we define the perturbed optimization problem

$$\min_{x \in D} f(x) + \varepsilon\varphi(x), \quad \text{for } \varepsilon > 0. \tag{2.6}$$

Note that the choice $\varphi(x) = 0.5\alpha\|x\|^2$, where α is the regularization parameter, is the classical Tikhonov regularization. For brevity, set

$$f_\varepsilon(x) = f(x) + \varepsilon\varphi(x), \tag{2.7}$$

then the function f_ε is clearly uniformly convex and lower semicontinuous. From Proposition 2.2, it follows that problem (2.6) has a unique solution, which will be denoted by x^ε .

We now establish the convergence result for approximate solutions of problem (2.6) to a solution of problem (1.2), which can be viewed as a modification of the known results from [3, 9, 10, 11].

PROPOSITION 2.3. *Suppose that assumptions (A1)–(A3) and (B1) are fulfilled. Then any sequence $\{z^k\}$, which is generated in conformity with the rules*

$$z^k \in D, \quad \|z^k - x^{\varepsilon_k}\| \leq \mu_k, \quad \{\varepsilon_k\} \searrow 0, \quad \{\mu_k\} \searrow 0, \tag{2.8}$$

converges strongly to the unique solution x_n^* of the problem

$$\min_{x \in D^*} \varphi(x). \tag{2.9}$$

Proof. First we note that the set D^* is nonempty, convex, and closed. Hence by Proposition 2.2, problem (2.9) has a unique solution. Next, we proceed to show that

$$\lim_{k \rightarrow \infty} x^{\varepsilon_k} = x_n^*. \tag{2.10}$$

Since f_ε is uniformly convex with function $\varepsilon\theta$, we have from (2.5) that

$$f(x^*) + \varepsilon\varphi(x^\varepsilon) \leq f(x^\varepsilon) + \varepsilon\varphi(x^\varepsilon) \leq f(x^*) + \varepsilon\varphi(x^*) - \varepsilon\tilde{\theta}(\|x^* - x^\varepsilon\|) \tag{2.11}$$

for each $\varepsilon > 0$ and for an arbitrary point $x^* \in D^*$, where $\tilde{\theta}(\|x^* - x^\varepsilon\|) = \theta(\|x^* - x^\varepsilon\|)\|x^* - x^\varepsilon\|$. It follows that

$$\tilde{\theta}(\|x^\varepsilon - x^*\|) \leq \varphi(x^*) - \varphi(x^\varepsilon), \tag{2.12}$$

but φ is bounded from below on E , hence the sequence $\{x^{\varepsilon_k}\}$ is bounded, that is, it has weak limit points. If x' is an arbitrary weak limit point of $\{x^{\varepsilon_k}\}$, then (2.11) yields

$$f(x^*) \leq f(x^\varepsilon) \leq f(x^*) + \varepsilon(\varphi(x^*) - \varphi(x^\varepsilon)), \tag{2.13}$$

that is, $x' \in D^*$. Setting $x^* = x_n^*$ and taking the corresponding limit in (2.12), we now obtain

$$\tilde{\theta}(\|x' - x_n^*\|) \leq \varphi(x_n^*) - \varphi(x') \leq 0, \tag{2.14}$$

hence $x_n^* = x'$ and (2.10) is true. Next, by definition,

$$\|z^k - x_n^*\| \leq \|z^k - x^{\varepsilon_k}\| + \|x^{\varepsilon_k} - x_n^*\| \leq \|x^{\varepsilon_k} - x_n^*\| + \mu_k. \tag{2.15}$$

It follows now from (2.8) and (2.10) that $\{z^k\}$ converges strongly to x_n^* . □

Thus, in order to present an implementable algorithm for solving problem (1.2), we have to find an approximate solution z^k of each perturbed problem (2.6) with the prescribed accuracy μ_k in a finite number of iterations.

3. Properties of the perturbed auxiliary problem

In this section, we establish several results which will be used for construction and substantiation of an iterative solution method for problem (2.6). Recall that, on account of Proposition 2.2, this problem has a unique solution if assumptions (A1), (A2), and (B1) are fulfilled. Nevertheless, in order to construct a convergent solution method, we will make use of the additional differentiability condition. More precisely, we replace (A2) and (B1) with the following assumptions.

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

(B1') $\varphi : E \rightarrow R$ is uniformly convex and has the Lipschitz continuous gradient map $\nabla \varphi : E \rightarrow E^*$.

Observe that the convexity of a function and the continuity of its gradient map imply lower semicontinuity (see, e.g., [1, Chapter I, Theorem 1.1]), so that (A2') and (B1') imply (A2) and (B1), respectively.

We start our considerations from the standard equivalence results for problem (2.6), which are modifications of those in [2, Chapter 2, Propositions 2.1 and 2.2].

LEMMA 3.1. *Suppose that (A1), (A2'), and (B1') are fulfilled. Then problem (2.6) is equivalent to each of the following variational inequalities:*

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

$$\langle \nabla f(x^\epsilon), x - x^\epsilon \rangle + \epsilon[\varphi(x) - \varphi(x^\epsilon)] \geq 0 \quad \forall x \in D, \tag{3.1}$$

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

$$\langle \nabla f(x^\epsilon) + \epsilon \nabla \varphi(x^\epsilon), x - x^\epsilon \rangle \geq 0 \quad \forall x \in D. \tag{3.2}$$

Proof. Obviously, (3.2) is the classical necessary and sufficient condition of optimality for (2.6) (see, e.g., [1, Chapter I, Theorem 0.4]). Utilizing the standard inequality for the differentiable convex function φ :

$$\varphi(x) - \varphi(x^\epsilon) \geq \langle \nabla \varphi(x^\epsilon), x - x^\epsilon \rangle \tag{3.3}$$

in (3.2) (see also (2.2)), we obtain the implication (3.2) \Rightarrow (3.1). Utilizing the same property with respect to f in (3.1), we obtain (3.1) \Rightarrow (2.6), but (2.6) \Leftrightarrow (3.2). Therefore, the assertion is true.

We introduce the auxiliary function

$$\Phi_\epsilon(x, y) = \langle \nabla f(x), y - x \rangle + \epsilon[\varphi(y) - \varphi(x)]. \tag{3.4}$$

Since $\Phi_\epsilon(x, \cdot)$ is clearly uniformly convex and lower semicontinuous under (A2') and (B1'), the optimization problem

$$\min_{y \in D} \Phi_\epsilon(x, y) \tag{3.5}$$

has a unique solution, which will be denoted by $y^\epsilon(x)$. Observe that the function Φ_ϵ in (3.5) can be in principal replaced with the simplified expression

$$\tilde{\Phi}_\epsilon(x, y) = \langle \nabla f(x), y \rangle + \epsilon\varphi(y). \tag{3.6}$$

Anyway, taking into account Lemma 3.1, we obtain immediately the fixed point characterization of the solution of (2.6). □

PROPOSITION 3.2. *If (A1), (A2'), and (B1') are fulfilled, then x^ϵ is a solution of (2.6) if and only if $x^\epsilon = y^\epsilon(x^\epsilon)$.*

Observe that

$$\psi_\epsilon(x) = -\min_{y \in D} \Phi_\epsilon(x, y) = -\Phi_\epsilon(x, y^\epsilon(x)) \tag{3.7}$$

can be regarded as the primal gap function for problem (3.1) or, equivalently, as the regularized gap function for problem (3.2) with the regularization term $\varepsilon[\varphi(x) - \varphi(x^\varepsilon) - \langle \nabla \varphi(x^\varepsilon), x - x^\varepsilon \rangle]$ (see [4, 7]). The idea of utilizing auxiliary terms in regularization and proximal point methods for constructing smooth gap functions was first suggested in [4] and called the nonlinear smoothing approach. This approach enables us to avoid including additional parameters and functions in descent methods and to propose very flexible computational schemes.

We now establish an error bound for the auxiliary problem (2.6).

PROPOSITION 3.3. *Suppose that (A1), (A2'), and (B1') are fulfilled. Then, for each $\varepsilon > 0$,*

$$2\varepsilon\theta(\|x - x^\varepsilon\|) \leq (L_f + 2\varepsilon L_\varphi)\|x - y^\varepsilon(x)\| \quad \forall x \in D, \tag{3.8}$$

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 $y = y^\varepsilon(x)$. Then, due to Lemma 3.1, we have

$$\langle \nabla f_\varepsilon(x^\varepsilon), y - x^\varepsilon \rangle \geq 0, \tag{3.9}$$

$$\langle \nabla f(x) + \varepsilon \nabla \varphi(y), x^\varepsilon - y \rangle \geq 0, \tag{3.10}$$

where the second inequality follows from the same optimality criterion applied to problem (3.5). Adding these inequalities gives

$$\langle \nabla f_\varepsilon(x^\varepsilon) - \nabla f_\varepsilon(x), y - x^\varepsilon \rangle + \varepsilon \langle \nabla \varphi(y) - \nabla \varphi(x), x^\varepsilon - y \rangle \geq 0. \tag{3.11}$$

Using (A2'), (B1'), and Proposition 2.1(iii), we have

$$\begin{aligned} 2\varepsilon\theta(\|x - x^\varepsilon\|)\|x - x^\varepsilon\| &\leq \langle \nabla f_\varepsilon(x) - \nabla f_\varepsilon(x^\varepsilon), x - x^\varepsilon \rangle \\ &= \langle \nabla f_\varepsilon(x) - \nabla f_\varepsilon(x^\varepsilon), x - y \rangle + \langle \nabla f_\varepsilon(x) - \nabla f_\varepsilon(x^\varepsilon), y - x^\varepsilon \rangle. \end{aligned} \tag{3.12}$$

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

$$\begin{aligned} 2\varepsilon\theta(\|x - x^\varepsilon\|)\|x - x^\varepsilon\| &\leq \langle \nabla f_\varepsilon(x) - \nabla f_\varepsilon(x^\varepsilon), x - y \rangle + \varepsilon \langle \nabla \varphi(y) - \nabla \varphi(x), x^\varepsilon - x \rangle \\ &\quad + \varepsilon \langle \nabla \varphi(y) - \nabla \varphi(x), x - y \rangle \\ &\leq \langle \nabla f(x) - \nabla f(x^\varepsilon), x - y \rangle + \varepsilon \langle \nabla \varphi(x) - \nabla \varphi(x^\varepsilon), x - y \rangle \\ &\quad + \varepsilon \langle \nabla \varphi(y) - \nabla \varphi(x), x^\varepsilon - x \rangle \\ &\leq L_f \|x - x^\varepsilon\| \|x - y\| + \varepsilon L_\varphi \|x - x^\varepsilon\| \|x - y\| \\ &\quad + \varepsilon L_\varphi \|y - x\| \|x^\varepsilon - x\|, \end{aligned} \tag{3.13}$$

hence

$$2\varepsilon\theta(\|x - x^\varepsilon\|) \leq (L_f + 2\varepsilon L_\varphi)\|x - y\|, \tag{3.14}$$

and the result follows. □

We now give the basic descent property which utilizes the direction $y^\epsilon(x) - x$.

PROPOSITION 3.4. *Suppose that (A1), (A2'), and (B1') are fulfilled. Then, for each $\epsilon > 0$,*

$$\begin{aligned} \langle \nabla f_\epsilon(x), y^\epsilon(x) - x \rangle &\leq -\epsilon \langle \nabla \varphi(y^\epsilon(x)) - \nabla \varphi(x), y^\epsilon(x) - x \rangle \\ &\leq -2\epsilon\theta(\|y^\epsilon(x) - x\|)\|y^\epsilon(x) - x\| \quad \forall x \in D. \end{aligned} \tag{3.15}$$

Proof. Fix $x \in D$ and again set, for brevity, $y = y^\epsilon(x)$. Then, writing the optimality condition for problem (3.5), we have

$$\langle \nabla f(x) + \epsilon \nabla \varphi(y), z - y \rangle \geq 0, \quad \forall z \in D, \tag{3.16}$$

(cf. (3.10)). Setting $z = x$ in this inequality gives

$$\langle \nabla f_\epsilon(x), x - y \rangle + \epsilon \langle \nabla \varphi(y) - \nabla \varphi(x), x - y \rangle \geq 0. \tag{3.17}$$

Thus the first inequality in (3.15) holds. The second inequality now follows from Proposition 2.1(iii).

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

PROPOSITION 3.5. *If (A1), (A2'), and (B1') are fulfilled, then $x \mapsto y^\epsilon(x)$ is continuous.*

Proof. Take arbitrary points $x', x'' \in D$ and set $y' = y^\epsilon(x')$, $y'' = y^\epsilon(x'')$. Using the optimality condition (3.16) gives

$$\begin{aligned} \langle \nabla f(x') + \epsilon \nabla \varphi(y'), y'' - y' \rangle &\geq 0, \\ \langle \nabla f(x'') + \epsilon \nabla \varphi(y''), y' - y'' \rangle &\geq 0. \end{aligned} \tag{3.18}$$

Adding these inequalities and taking into account (3.2), we obtain

$$\begin{aligned} \langle \nabla f(x') - \nabla f(x''), y'' - y' \rangle &\geq \epsilon \langle \nabla \varphi(y'') - \nabla \varphi(y'), y'' - y' \rangle \\ &\geq 2\epsilon\theta(\|y'' - y'\|)\|y'' - y'\|, \end{aligned} \tag{3.19}$$

hence

$$L_f \|x' - x''\| \geq 2\epsilon\theta(\|y' - y''\|), \tag{3.20}$$

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

4. Descent method for convex optimization

First we describe a descent algorithm for solving the auxiliary problem (2.6) for a fixed $\epsilon > 0$. This algorithm follows the general iteration scheme from [7].

Since the function φ may be chosen rather arbitrarily within rule (B1'), this algorithm admits in fact a wide variety of iteration schemes. That is, one can choose φ to be suitable for approximation of properties of the cost function f in (1.2), for solution of the auxiliary problem (3.5), and for computation of the gradient $\nabla \varphi$. These properties make Algorithm 4.1 very flexible in comparison with the usual gradient schemes.

Step 1. Choose a point $x^0 \in D$ and numbers $\alpha \in (0, 1)$ and $\beta \in (0, 1)$. Set $i = 0$.
 Step 2. Compute $y^i = y^\varepsilon(x^i)$ and set $d^i = y^i - x^i$.
 Step 3. Find m as the smallest nonnegative integer such that

$$f_\varepsilon(x^i + \beta^m d^i) \leq f_\varepsilon(x^i) - \alpha\beta^m \varepsilon \langle \nabla \varphi(y^i) - \nabla \varphi(x^i), d^i \rangle, \quad (4.1)$$

set $\lambda_i = \beta^m$, $x^{i+1} = x^i + \lambda_i d^i$, $i = i + 1$ and go to Step 2.

ALGORITHM 4.1

The next theorem presents a convergence result for Algorithm 4.1.

THEOREM 4.1. *Suppose that assumptions (A1), (A2'), and (B1') are satisfied. Then any sequence $\{x^i\}$, generated by Algorithm 4.1, converges strongly to the unique solution x^ε of problem (2.6) for a given $\varepsilon > 0$.*

Proof. It has been mentioned that (2.6) has a unique solution due to Proposition 2.2. By (A2') and (B1'), ∇f_ε is Lipschitz continuous with constant $L = L_f + \varepsilon L_\varphi$, 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:

$$f_\varepsilon(x^i + \lambda d^i) - f_\varepsilon(x^i) \leq \lambda \langle \nabla f_\varepsilon(x^i), d^i \rangle + 0.5\lambda^2 L \|d^i\|^2. \quad (4.2)$$

In view of (3.15), we have

$$f_\varepsilon(x^i + \lambda d^i) - f_\varepsilon(x^i) \leq -\lambda \varepsilon \langle \nabla \varphi(y^i) - \nabla \varphi(x^i), d^i \rangle + 0.5\lambda^2 L \|d^i\|^2. \quad (4.3)$$

If

$$\varepsilon \langle \nabla \varphi(y^i) - \nabla \varphi(x^i), d^i \rangle - 0.5\lambda L \|d^i\|^2 \geq \alpha \varepsilon \langle \nabla \varphi(y^i) - \nabla \varphi(x^i), d^i \rangle \quad (4.4)$$

is satisfied for a positive λ , then the line-search procedure in Algorithm 4.1 becomes implementable. However, (4.4) is equivalent to

$$\lambda \leq 2(1 - \alpha) \varepsilon \langle \nabla \varphi(y^i) - \nabla \varphi(x^i), d^i \rangle / (L \|d^i\|^2) \quad (4.5)$$

and (see (3.15)) we have

$$\langle \nabla \varphi(y^i) - \nabla \varphi(x^i), d^i \rangle \geq 2\theta (\|d^i\|) \|d^i\|. \quad (4.6)$$

So (4.4) is satisfied if

$$\lambda \leq 4(1 - \alpha) \varepsilon \theta (\|d^i\|) / (L \|d^i\|). \quad (4.7)$$

□

By (A1), (A2'), and (B1'), f_ε is uniformly convex, hence it is bounded from below on D and the level set $D_0 = \{x \in D \mid f_\varepsilon(x) \leq f_\varepsilon(x^0)\}$ is also bounded because of Proposition 2.1.

But the sequence $\{f_\varepsilon(x^i)\}$ is nonincreasing due to (4.1), hence, by Proposition 3.5,

$$\|d^i\| \leq C < +\infty \quad \text{for } i = 0, 1, \dots \tag{4.8}$$

Suppose that

$$\|d^i\| \geq l > 0 \quad \text{for } i = 0, 1, \dots \tag{4.9}$$

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

$$f_\varepsilon(x^i + \lambda d^i) - f_\varepsilon(x^i) \leq -2\alpha\lambda\varepsilon\theta(\|d^i\|)\|d^i\|, \tag{4.10}$$

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

$$f_\varepsilon(x^i) - f_\varepsilon(x^{i+1}) \geq 2\alpha\lambda''\varepsilon\theta(\|d^i\|)\|d^i\| \rightarrow 0 \tag{4.11}$$

as $i \rightarrow +\infty$, which is a contradiction. Therefore, there exists a subsequence $\{i_s\}$ such that $\|d^{i_s}\| \rightarrow 0$ as $i_s \rightarrow +\infty$, that is, $\|y^\varepsilon(x^{i_s}) - x^{i_s}\| \rightarrow 0$, and by (3.8), $\|x^{i_s} - x^\varepsilon\| \rightarrow 0$ as $i_s \rightarrow +\infty$. It means that x^ε is a strong limit point of $\{x^i\}$. Since $\{f_\varepsilon(x^i)\}$ is nonincreasing, it means that

$$\lim_{i \rightarrow \infty} f_\varepsilon(x^i) = f_\varepsilon(x^\varepsilon). \tag{4.12}$$

Applying now (2.5) with $\psi = f_\varepsilon$, we conclude that the whole sequence $\{x^i\}$ converges strongly to x^ε and the result follows.

Being based on the above result, we can approximate the solution x^ε of each perturbed problem (2.6) with any prescribed accuracy in a finite number of iterations. In order to construct such a combined regularization and descent method, we will make use of error bound (3.8). The corresponding solution method for the initial convex optimization problem (1.2) can be described as follows.

Method. Choose a point $z^0 \in D$, a number $\delta > 0$, and a positive sequence $\{\varepsilon_k\} \searrow 0$. For each $k = 1, 2, \dots$, we have a point $z^{k-1} \in D$; apply Algorithm 4.1 with $x^0 = z^{k-1}$, $\varepsilon = \varepsilon_k$ and construct a sequence $\{x^i\}$ until

$$\|d^i\| \leq \varepsilon^{1+\delta}. \tag{4.13}$$

Then we set $z^k = x^i$ and increase $k = k + 1$.

We now show that our method is fully implementable and generates a strongly convergent sequence.

THEOREM 4.2. *Suppose that assumptions (A1), (A2'), (A3), and (B1') are satisfied and that a sequence $\{z^k\}$ is generated by the method. Then,*

- (i) *each k th iteration of the method is finite;*
- (ii) *the sequence $\{z^k\}$ converges strongly to the point x_n^* , which is the unique solution of problem (2.9).*

Proof. First we note that the inequality (4.13) will be satisfied in a finite number of iterations, since $\|d^i\| \rightarrow 0$ as $i \rightarrow +\infty$ due to Theorem 4.1 and Propositions 3.2 and 3.5. Hence, assertion (i) of the theorem is true. Next, combining (4.13) and (3.8) yields

$$\begin{aligned} 2\varepsilon\theta(\|x^i - x^\varepsilon\|) &\leq (L_f + 2\varepsilon L_\varphi)\varepsilon^{1+\delta}, \\ \theta(\|z^k - x^{\varepsilon_k}\|) &\leq F\varepsilon_k^\delta, \quad 0 < F < +\infty \end{aligned} \tag{4.14}$$

since $\{\varepsilon_k\} \searrow 0$. Moreover, it follows that

$$\|z^k - x^{\varepsilon_k}\| \leq \mu_k \tag{4.15}$$

for some sequence $\{\mu_k\} \searrow 0$. We now see that all the assumptions of Proposition 2.3 are satisfied. Therefore, assertion (ii) of the theorem also holds. \square

5. Numerical experiments

As has been mentioned, a great number of applied problems in physics, mechanics, optimal control, economics, engineering are solved via the regularization approach; for example, see [9, 10, 11]. In this section, we present some preliminary results of numerical experiments that illustrate convergence properties of the combined regularization and descent method.

We chose a well-known ill-posed optimal control problem; see, for example, [10, page 162, Example 2]. The problem is to minimize the functional

$$f(x) = \int_0^1 u^2(t) dt, \tag{5.1}$$

subject to

$$\begin{aligned} \dot{u}(t) &= x(t), \quad t \in [0, 1], \quad u(0) = 0, \\ x(t) &\in X = \{x(t) \in L_2[0, 1] \mid |x(t)| \leq 1 \text{ a.e.}\}. \end{aligned} \tag{5.2}$$

It is easy to see that the cost functional is convex and differentiable, that the optimal control $x^*(\cdot) \equiv 0$ is determined uniquely, and that the minimal value $f^* = 0$. At the same time, the sequence $x_k(t) = \sin(2\pi kt)$ minimizes the cost functional, but

$$\|x^*(\cdot) - x_k(\cdot)\|_{L_2[0,1]} = 0.5, \tag{5.3}$$

that is, it is not norm converging to the solution. It means that the initial problem is ill-posed.

We apply the described combined regularization and descent method to solve this problem. We implement the method with double-precision arithmetic with piecewise constant approximation of the control function $x(t)$. We used the stopping criterion

$$\|x^k - z^k\| \leq \eta \tag{5.4}$$

in Algorithm 4.1. The total number of iterations for several values of the accuracy η and the number N of pieces of control approximation is given in Table 6.1.

TABLE 6.1. $\varepsilon = 1$.

N	$\eta = 0.1$	$\eta = 0.01$	$\eta = 0.001$	$\eta = 0.0001$	$\eta = 0.00001$
50	2	4	5	8	12
100	2	4	5	8	12

The results show that the convergence of the method is rather stable and rapid for ill-posed problems.

6. Conclusions

Thus, we have presented a general iterative scheme of implementable algorithms with strong convergence for convex optimization problems. Although this scheme is used for differentiable problems, the above approach may be viewed as a basis for nondifferentiable problems.

Acknowledgments

The author is grateful to Professor I. V. Konnov for his suggestion of the subject of this paper and for useful discussions of the results. The author is also grateful to referees whose suggestions improved the exposition of the paper.

References

- [1] J. Céa, *Optimisation: Théorie et algorithmes*, Méthodes Mathématiques de l'Informatique, vol. 2, Dunod, Paris, 1971.
- [2] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, Studies in Mathematics and Its Applications, vol. 1, North-Holland, Amsterdam; American Elsevier, New York, 1976.
- [3] V. G. Karmanov, *Mathematical Programming*, Nauka, Moscow, 1986.
- [4] I. V. Konnov, *Approximate methods for primal-dual mixed variational inequalities*, Russian Math. (Iz. VUZ) **44** (2000), no. 12, 55–66.
- [5] I. V. Konnov, S. Kum, and G. M. Lee, *On convergence of descent methods for variational inequalities in a Hilbert space*, Math. Methods Oper. Res. **55** (2002), no. 3, 371–382.
- [6] E. S. Levitin and B. T. Polyak, *Constrained minimization methods*, Comput. Math. Math. Phys. **6** (1966), 1–50.
- [7] M. Patriksson, *Cost approximation: a unified framework of descent algorithms for nonlinear programs*, SIAM J. Optim. **8** (1998), no. 2, 561–582.
- [8] B. T. Polyak, *Introduction to Optimization*, Translations Series in Mathematics and Engineering, Optimization Software, New York, 1987.
- [9] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, Scripta Series in Mathematics, V. H. Winston & Sons, Washington, D.C.; John Wiley & Sons, New York, 1977.
- [10] F. P. Vasil'yev, *Methods of Solution of Extremal Problems*, Nauka, Moscow, 1981.
- [11] V. V. Vasin and A. L. Ageev, *Ill-Posed Problems with a Priori Information*, Inverse and Ill-posed Problems Series, VSP, Utrecht, 1995.