Math. Program., Ser. A 86: 41–50 (1999)

Digital Object Identifier (DOI) 10.1007/s101079900088

Stephen M. Robinson

Linear convergence of epsilon-subgradient descent methods for a class of convex functions⋆

Received March 29, 1996 / Revised version received March 5, 1999
Published online June 11, 1999

Abstract. This paper establishes a linear convergence rate for a class of epsilon-subgradient descent methods for minimizing certain convex functions on $\mathbb{R}^n$. Currently prominent methods belonging to this class include the resolvent (proximal point) method and the bundle method in proximal form (considered as a sequence of serious steps). Other methods, such as a variant of the proximal point method given by Correa and Lemaréchal, can also fit within this framework, depending on how they are implemented. The convex functions covered by the analysis are those whose conjugates have subdifferentials that are locally upper Lipschitzian at the origin, a property generalizing classical regularity conditions.

Key words. resolvent method – proximal point method – bundle method – bundle-trust region method – subgradient

1. Introduction

This paper deals with ϵ-subgradient descent methods for minimizing a convex function f on $\mathbb{R}^n$. The class of methods we consider consists of those treated by Correa and Lemaréchal in [3], with the additional restrictions that the minimizing set be nonempty, the stepsize parameters be bounded, and a condition for sufficient descent be enforced at each step. We give a precise description of this class in Sect. 2.

Currently prominent methods belonging to this class include the resolvent (proximal point) method and the bundle method in proximal form (considered as a sequence of serious steps). The resolvent method was treated by Rockafellar [14, 15] and by Brézis and Lions [1], and has since been the subject of much attention. Implementations of the proximal bundle method have been given recently by Zowe [19], Kiwiel [7], and Schramm and Zowe [16], building on a considerable amount of earlier work; see [6] for references. Certain other methods, such as the variant of the proximal point method given by Correa and Lemaréchal in Algorithm 3.3 of [3], can also fit into the class we consider, depending on how they are implemented.

We show that the methods we consider will converge with (at least) an R-linear rate in in the sense of Ortega and Rheinboldt [10], in the case when they are used to minimize

S.M. Robinson: Department of Industrial Engineering, University of Wisconsin–Madison, 1513 University Avenue, Madison, WI 53706-1572, USA, e-mail: smrobins@facstaff.wisc.edu

Mathematics Subject Classification (1991): 90C25, 49M45

⋆ This material is based upon work supported by the U. S. Army Research Office under Grants DAAH04-95-1-0149 and DAAG55-97-1-0324, and by the North Atlantic Treaty Organization under grant number CRG.950360. Preliminary research for this paper was conducted in part at the Institute for Mathematics and its Applications, Minneapolis, Minnesota, with funds provided by the National Science Foundation, and in part at the International Institute for Applied Systems Analysis, Laxenburg, Austria.

closed proper convex functions f on $\mathbb{R}^n$ that are of a special type: namely, those whose conjugates f^* have subdifferentials that are *locally upper Lipschitzian,* in the sense defined in [11], at the origin. The following definition gives the specific property we require. With one exception, throughout the paper we use an inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$ that induces a norm $\| \cdot \|$ by the relation $\|x\|^2 = \langle x, x \rangle$. The notation B stands for the unit ball of this norm. The exception is in Theorem 2, where we use similar conventions but in a real Hilbert space.

Definition 1. *A convex function f on $\mathbb{R}^n$ satisfies the* inverse growth condition *with modulus μ if there exist a neighborhood U of the origin in $\mathbb{R}^n$ and a constant μ such that for each $x^* \in U$,*

$$\partial f^*(x^*) \subset \partial f^*(0) + \mu \|x^*\| B.$$

This inverse growth condition has been employed by Luque [9] (who attributed it to Bertsekas), Zhang and Treiman [17], and Zhu [18].

For the problem of unconstrained minimization of a C^2 function, the standard second-order sufficient condition (that is, positive definiteness of the Hessian at a minimizer) implies that the function is strongly convex if restricted to a suitable neighborhood of the minimizer, that the conjugate of this restricted function is finite near the origin, and that the inverse growth condition holds. However, in the more general situation that we consider here the inverse growth condition may hold even if the minimizing set is not a singleton.

Rates of convergence for classes of methods overlapping those considered here have previously been given by Luque [9] for the inexact resolvent method, and by Zhu [18] for several classes of methods. The analysis of Luque requires that the subproblems in the resolvent method be solved to increasing accuracy as the iteration proceeds. This accuracy is measured in terms of the distance between the accepted point and the true solution of the subproblem, a quantity that is not observable. The analysis of Zhu does not require the measurement of unobservable quantities, but it does require boundedness of the minimizing set. It also provides convergence rates defined only in terms of the speed at which the function values approach the minimum of f. Convergence rates as usually defined (in terms of the successive iterates x_n) are not provided; in fact Zhu does not even prove that the sequence $\{x_n\}$ converges.

By contrast, for the class of methods considered here we do not require boundedness of the minimizing set. Further, we are able to show that the sequence of iterates converges to a minimizer, and to prove conventional R-linear convergence in terms of that sequence.

The rest of this paper is organized in two sections. Section 2 describes precisely the class of minimization methods we consider, and provides some useful information about their behavior, including convergence. Section 3 then shows that their rate of convergence is at least R-linear if the function being minimized has a nonempty minimizing set and satisfies the inverse growth condition.

2. Epsilon-subgradient descent methods

In this section we describe the class of minimization methods with which we are concerned, and we review some results about their behavior.

Let f be a closed proper convex function on $\mathbb{R}^n$, which we wish to minimize. The authors of [3] investigated a class of ϵ-subgradient descent methods for such minimization. These methods proceed by fixing a starting point $x_0 \in \mathbb{R}^n$ and then generating succeeding points by the formula

$$x_{n+1} = x_n - t_n d_n^*, \qquad d_n^* \in \partial_{\epsilon_n} f(x_n), \tag{1}$$

where t_n is a positive stepsize parameter, $\epsilon_n \geq 0$, and $\partial_{\epsilon_n} f(x_n)$ is the ϵ_n-subdifferential of f at x_n, defined for $\epsilon_n \geq 0$ by

$$\partial_{\epsilon_n} f(x_n) = \{x^* \mid \text{ for each } z \in \mathbb{R}^n, \ f(z) \geq f(x_n) + \langle x^*, z - x_n \rangle - \epsilon_n \}.$$

Thus, for $\epsilon_n = 0$ we have the ordinary subdifferential, whereas for positive ϵ_n we have a larger set. For more information about the ϵ-subdifferential, see §25 of [12].

We point out that in this method the stepsize parameters t_n are generally fixed in advance, but the tolerances ϵ_n are determined adaptively as each step is taken. This is somewhat different from the situation in smooth optimization, where one generally uses an adaptive method to determine the stepsizes.

In addition to requiring the function f to satisfy certain properties, we shall impose two requirements on the implementation of (1). They are stricter than those imposed in [3], but they will permit us to obtain the convergence rate results that we seek. One of these is that the sequence of stepsize parameters be bounded away from 0 and from ∞: namely, there are t_* and t^* such that for each n,

$$0 < t_* \leq t_n \leq t^* < \infty. \tag{2}$$

The other requirement is that at each step a sufficient descent be obtained: specifically, there is a constant $m \in (0, 1]$ such that for each n,

$$f(x_{n+1}) \leq f(x_n) + m \left(\langle d_n^*, x_{n+1} - x_n \rangle - \epsilon_n \right). \tag{3}$$

Note that because $d_n^* = -t_n^{-1}(x_{n+1} - x_n)$, the quantity in parentheses in (3) is non-positive, and in fact negative if $x_{n+1} \neq x_n$ or if $\epsilon_n > 0$, so that we are working with a descent method: that is, one that forces the function value at each successive step to be "sufficiently" smaller than its predecessor. Indeed, if $\epsilon_n = 0$ and if the subgradient is actually a gradient, this is a descent condition very familiar from the literature (for example, see ([4], p. 101). However, the ϵ-descent condition in the general form given here may seem somewhat strange. For that reason, we show next that some well known methods satisfy this condition.

The first of these methods is the resolvent, or proximal point, method in the form appropriate for minimization of f. This algorithm is specified by

$$x_{n+1} = (I + t_n \partial f)^{-1}(x_n);$$

that is, we obtain x_{n+1} by applying to x_n the resolvent J_{t_n} of the maximal monotone operator ∂f. To see that this is in the form (1), note that the algorithm specification implies that there is $d_n^* \in \partial f(x_{n+1})$ such that

$$x_n = x_{n+1} + t_n d_n^*,$$

which is a rearrangement of (1). Further, for each z we have

$$f(z) \geq f(x_{n+1}) + \langle d_n^*, z - x_{n+1} \rangle = f(x_n) + \langle d_n^*, z - x_n \rangle - \epsilon_n,$$

where

$$\epsilon_n = f(x_n) - f(x_{n+1}) - \langle d_n^*, x_n - x_{n+1} \rangle,$$

which is nonnegative because $d_n^* \in \partial f(x_{n+1})$. Therefore $d_n^* \in \partial_{\epsilon_n} f(x_n)$. Moreover, we have

$$f(x_{n+1}) = f(x_n) + \langle d_n^*, x_{n+1} - x_n \rangle - \epsilon_n,$$

so that (3) holds with $m = 1$.

The resolvent method is unfortunately not implementable except in special cases. Noting this, Correa and Lemaréchal gave a variant in Algorithm 3.3 of [3]. We are concerned here only with the outer loop of that variant, in which we fix two parameters, $\kappa > 1$ and $m \in (0, 1)$, and for some sequence of positive stepsizes $\{t_n\}$ we then require successive points of the sequence $\{x_n\}$ to satisfy the two conditions

$$\eta = \kappa \left[f(x_n) - f(x_{n+1}) - m t_n^{-1} \|x_{n+1} - x_n\|^2 \right] \tag{4}$$

and

$$t_n^{-1}(x_n - x_{n+1}) \in \partial_\eta f(x_n). \tag{5}$$

If the parameters m and κ are required to satisfy $\kappa m \leq 1$ and the stepsizes t_n are bounded away from zero and ∞, then this algorithm satisfies our conditions (1), (2), and (3). The second of these holds by choice of the t_n, and for the first part of (1) we simply define d_n^* to be the left-hand side of (5). Now we define ϵ_n to be $(\kappa m)^{-1}\eta$ and rewrite (4) as

$$\begin{aligned} f(x_{n+1}) &= f(x_n) - m t_n^{-1} \|x_{n+1} - x_n\|^2 - \kappa^{-1}\eta \\ &= f(x_n) + m \left[\langle d_n^*, x_{n+1} - x_n \rangle - \epsilon_n \right]. \end{aligned}$$

This shows that (3) holds. Further, as we required $\kappa m \leq 1$ we have $\epsilon_n \geq \eta$ and therefore $\partial_\eta f(x_n) \subset \partial_{\epsilon_n} f(x_n)$, so that (5) implies the second part of (1). Therefore (1), (2), and (3) all hold, so this algorithm fits within our class.

For practical minimization of nonsmooth convex functions a very effective tool is the well known bundle method, which as is pointed out in [3] can be regarded as a systematic way of approximating the iterations of the resolvent method. The method uses two kinds of steps, "serious steps," which as we shall see correspond to (1), and "null steps," which are used to prepare for the serious steps. Specifically, by means of a sequence of null steps the method builds up a piecewise affine minorant $\hat{f}$ of f. Then a resolvent step is taken, using $\hat{f}$ instead of f:

$$x_{n+1} = \left(I + t_n \partial \hat{f} \right)^{-1} (x_n), \tag{6}$$

and it is accepted if

$$f(x_n) - f(x_{n+1}) \geq m \left[f(x_n) - \hat{f}(x_{n+1}) \right]. \tag{7}$$

Now from (6) we see that

$$x_{n+1} = x_n - t_n d_n^*,$$

with $d_n^* \in \partial \hat{f}(x_{n+1})$. Then for each $z \in \mathbb{R}^n$ we have

$$f(z) \geq \hat{f}(z) \geq \hat{f}(x_{n+1}) + \langle d_n^*, z - x_{n+1} \rangle = f(x_n) + \langle d_n^*, z - x_n \rangle - \epsilon_n,$$

where we can write ϵ_n as

$$\epsilon_n = \left[f(x_n) - \hat{f}(x_n) \right] + \left[\hat{f}(x_n) - \hat{f}(x_{n+1}) - \langle d_n^*, x_n - x_{n+1} \rangle \right], \tag{8}$$

which must be nonnegative since $\hat{f}$ minorizes f and $d_n^* \in \partial \hat{f}(x_{n+1})$. In fact, $\hat{f}$ is typically constructed in such a way that $\hat{f}(x_n) = f(x_n)$, so the first term in square brackets is actually zero (this will be the case as long as a subgradient of f at x_n belongs to the bundle). In that case we have from the minorization property and (8)

$$f(x_n) - \hat{f}(x_{n+1}) \geq \hat{f}(x_n) - \hat{f}(x_{n+1}) = \langle d_n^*, x_n - x_{n+1} \rangle + \epsilon_n,$$

so that (7) yields

$$f(x_n) - f(x_{n+1}) \geq m \left[\langle d_n^*, x_n - x_{n+1} \rangle + \epsilon_n \right];$$

that is, (3) holds. Therefore the bundle method, if implemented with bounded t_n, fits within our class of methods.

Although our proof of R-linear convergence in Sect. 3 therefore applies to the bundle method, it must be noted that this analysis takes into account only the serious steps, whereas for each serious step a possibly large number of null steps may be required to build up an adequate approximation $\hat{f}$. Therefore our analysis does not provide a bound on the total work required to implement the bundle method. Such bounds have been investigated, for example, in a recent paper of Kiwiel [8], using a completely different approach from that employed here.

We have now seen that some well known methods fit into the class we analyze. In the analysis we use the following theorem, which summarizes the convergence properties of this class.

Theorem 1. *Let f be a lower semicontinuous proper convex function on $\mathbb{R}^n$, having a nonempty minimizing set X_*. Let x_0 be given and suppose the algorithm (1) is implemented in such a way that (2) and (3) hold. Then the sequence $\{x_n\}$ generated by (1) converges to a point $x_* \in X_*$, $\{f(x_n)\}$ converges to* min f, *and*

$$\sum_{n=0}^{\infty} \left(\|d_n^*\|^2 + \epsilon_n \right) < \infty. \tag{9}$$

In particular, the sequences $\{\epsilon_n\}$ and $\{\|d_n^*\|\}$ converge to zero.

Proof. Note that for each n we have $\langle d_n^*, x_{n+1} - x_n \rangle = -t_n \|d_n^*\|^2$. From (2) and (3) we obtain

$$m\left(t_* \|d_n^*\|^2 + \epsilon_n\right) \le m\left(t_n \|d_n^*\|^2 + \epsilon_n\right) \le f(x_n) - f(x_{n+1}),$$

so for each $k \ge 1$ we have

$$m \sum_{n=0}^{k-1} \left(t_* \|d_n^*\|^2 + \epsilon_n\right) \le f(x_0) - f(x_k) \le f(x_0) - \min f,$$

and consequently

$$m \sum_{n=0}^{\infty} \left(t_* \|d_n^*\|^2 + \epsilon_n\right) \le f(x_0) - \min f,$$

which establishes (9). The condition (2) shows that the sum of the t_n is infinite, so that Conditions (1.4) and (1.5) of [3] hold. Moreover, (3) shows that for each n

$$f(x_{n+1}) \le f(x_n) + m\left(\langle d_n^*, x_{n+1} - x_n \rangle - \epsilon_n\right) \le f(x_n) - m t_n \|d_n^*\|^2,$$

so that Condition (2.7) of [3] also holds. Then Proposition 2.2 of [3] shows that $\{f(x_n)\}$ converges to $\min f$ and that $\{x_n\}$ converges to some element x_* of X_*.

□

In this section we have specified the class of methods we are considering, and we have given two examples of concrete methods that belong to this class. Moreover, we have adapted from [3] a general convergence result applicable to this class. In the next section we present the main result of the paper, a proof that the convergence guaranteed by Theorem 1 will under additional conditions actually be at least R-linear.

3. Convergence-rate analysis

In order to prove the main result we need to use a tailored form of the well known Brøndsted-Rockafellar Theorem [2]. We give this next, along with a very simple proof. The technique of this proof is very similar to that given in Theorem 4.2.1 of [5], but this version gives slightly more information and it holds in any real Hilbert space. For a multifunction T we use the notation $(x, y) \in T$ to mean $y \in T(x)$.

Theorem 2. *Let H be a real Hilbert space and let f be a lower semicontinuous proper convex function on H. Suppose that $\epsilon \ge 0$ and that $(x_\epsilon, x_\epsilon^*) \in \partial_\epsilon f$. For each positive β there is a unique y_β with*

$$\left(x_\epsilon + \beta y_\beta, x_\epsilon^* - \beta^{-1} y_\beta\right) \in \partial f. \tag{10}$$

Further, $\|y_\beta\| \le \epsilon^{1/2}$.

Proof. Define a function g on H by

$$g(y) = (1/2)\|y - \beta x_\epsilon^*\|^2 + f(x_\epsilon + \beta y).$$

Then g is lower semicontinuous, proper, and strongly convex; its unique minimizer y_β then satisfies $0 \in \partial g(y_\beta)$, which upon rearrangement becomes (10); justification for the subdifferential computation can be found in, *e.g.*, Theorem 20, p. 56, of [13]. In turn, (10) implies

$$f(x_\epsilon) \geq f(x_\epsilon + \beta y_\beta) + \langle x_\epsilon^* - \beta^{-1} y_\beta, x_\epsilon - (x_\epsilon + \beta y_\beta)\rangle.$$

But the ϵ-subgradient inequality yields

$$f(x_\epsilon + \beta y_\beta) \geq f(x_\epsilon) + \langle x_\epsilon^*, (x_\epsilon + \beta y_\beta) - x_\epsilon\rangle - \epsilon,$$

and by combining these we obtain

$$0 \geq \langle x_\epsilon^* - \beta^{-1} y_\beta, -\beta y_\beta\rangle + \langle x_\epsilon^*, \beta y_\beta\rangle - \epsilon = \|y_\beta\|^2 - \epsilon,$$

which proves the assertion about $\|y_\beta\|$.

□

Here is the main theorem, which says that under the inverse growth condition the ϵ-subgradient descent method is at least R-linearly convergent.

Theorem 3. *Let f be a lower semicontinuous, proper convex function on $\mathbb{R}^n$ that satisfies the inverse growth condition with modulus $\mu > 0$. Assume that f has a nonempty minimizing set X_*, and that starting from some x_0 the ϵ-subgradient descent method (1) is implemented with (2) and (3) satisfied at each step.*

Then the sequence $\{x_n\}$ produced by (1) converges at least R-linearly to a limit $x_* \in X_*$.

Proof. Consider the step from x_n to x_{n+1}. From (1) we find that $d_n^* \in \partial_{\epsilon_n} f(x_n)$, and by applying Theorem 2 we conclude that there is a unique y with $\|y\| \leq \epsilon_n^{1/2}$ and with

$$\left(x_n + \mu^{1/2} y, d_n^* - \mu^{-1/2} y\right) \in \partial f.$$

For any k let u_k be the projection of x_k on the optimal set X_*. We have shown in Theorem 1 that $\|d_n^*\|$ and ϵ_n converge to zero. Therefore there is some N such that for $n \geq N$ the point $d_n^* - \mu^{-1/2} y$ will lie in the neighborhood U associated with the inverse growth condition. For such n the condition yields

$$\left\|(x_n + \mu^{1/2} y) - u_n\right\| \leq \mu \left\|d_n^* - \mu^{-1/2} y\right\|.$$

Therefore

$$\begin{aligned} \|x_n - u_n\| &\leq \left\|\left(x_n + \mu^{1/2} y\right) - u_n\right\| + \mu^{1/2}\|y\| \\ &\leq \mu \left\|d_n^* - \mu^{-1/2} y\right\| + \mu^{1/2}\epsilon_n^{1/2} \\ &\leq \mu\|d_n^*\| + 2\mu^{1/2}\epsilon_n^{1/2}. \end{aligned} \tag{11}$$

Next, let $f_* = \min f$; write ϕ_n for $f(x_n) - f_* = f(x_n) - f(u_n)$, and σ_n for μt_n^{-1}. Using the fact that $d_n^* \in \partial_{\epsilon_n} f(x_n)$, together with (11), we obtain

$$\begin{aligned}\phi_n &\le -\langle d_n^*, u_n - x_n\rangle + \epsilon_n \\ &\le \mu\|d_n^*\|^2 + 2\mu^{1/2}\|d_n^*\|\epsilon_n^{1/2} + \epsilon_n \\ &= \left(\mu^{1/2}\|d_n^*\| + \epsilon_n^{1/2}\right)^2 \\ &= \left(\sigma_n^{1/2} t_n^{1/2}\|d_n^*\| + \epsilon_n^{1/2}\right)^2 .\end{aligned} \tag{12}$$

For any real numbers α, β, and γ, the Schwarz inequality applied to $(1, \beta)$ and (α, γ) yields

$$|\alpha + \beta\gamma| \le \left(1 + \beta^2\right)^{1/2} \left(\alpha^2 + \gamma^2\right)^{1/2} .$$

Using this in (12) we obtain

$$\begin{aligned}\phi_n &\le \left[(1 + \sigma_n)^{1/2} \left(t_n\|d_n^*\|^2 + \epsilon_n\right)^{1/2}\right]^2 \\ &= (1 + \sigma_n)\left(t_n\|d_n^*\|^2 + \epsilon_n\right) .\end{aligned} \tag{13}$$

But from (1) and (3) we have

$$t_n\|d_n^*\|^2 + \epsilon_n \le m^{-1}[f(x_n) - f(x_{n+1})],$$

and we also have $f(x_n) - f(x_{n+1}) = \phi_n - \phi_{n+1}$. Therefore (13) yields

$$\phi_n \le (1 + \sigma_n)m^{-1}(\phi_n - \phi_{n+1}),$$

which, since $t_n \ge t_* > 0$, implies

$$\phi_{n+1} \le \theta^2\phi_n,$$

with

$$\theta = \left[1 - m\Big/\left(1 + \mu t_*^{-1}\right)\right]^{1/2} .$$

Therefore for fixed N and any $n \ge N$ we have

$$\phi_n \le \kappa\theta^{2n}, \tag{14}$$

with

$$\kappa = \theta^{-2N}\phi_N .$$

Now from Theorem 4.3 of [17] we find that for some $\gamma \ge 0$ and all z with $d(z, X_*)$ sufficiently small the inequality

$$f(z) \ge f_* + \gamma d(z, X_*)^2 \tag{15}$$

holds. We know that $d(x_n, X_*)$ converges to zero, so for all n at least as large as some $N' \geq N$ we have from (14) and (15)

$$\xi_n := d(x_n, X_*) \leq \gamma^{-1/2}\phi_n^{1/2} \leq \lambda\theta^n, \tag{16}$$

with

$$\lambda = \gamma^{-1/2}\theta^{-N}\phi_N^{1/2}.$$

Now let x_* be the unique limit of the sequence $\{x_n\}$, as established in Theorem 1. We have established a suitable bound on $d(x_n, X_*)$, but we need to find such a bound for the error $\|x_n - x_*\|$. The remainder of the proof does this.

From Equation (1.3) of [3] we have, for any $y \in \mathbb{R}^n$,

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 + t_n^2\|d_n^*\|^2 + 2t_n[f(y) - f(x_n) + \epsilon_n].$$

If we restrict our attention to points $y \in X_*$ we may simplify this to

$$\|x_{n+1} - y\|^2 \leq \|x_n - y\|^2 + 2t_n\left[t_n\|d_n^*\|^2 + \epsilon_n - \phi_n\right].$$

For $j > n \geq N'$ we then use the fact that $t_k \leq t^*$ for all k to obtain the upper bound

$$\begin{aligned}\|x_j - y\|^2 &\leq \|x_n - y\|^2 + 2t^*\sum_{k=n}^{j-1}\left[t_k\|d_k^*\|^2 + \epsilon_k - \phi_k\right]\\ &\leq \|x_n - y\|^2 + 2t^*\left(\sum_{k=n}^{j-1}\left[t_k\|d_k^*\|^2 + \epsilon_k\right] - \phi_n\right).\end{aligned}$$

The condition (3) gives

$$f(x_{k+1}) \leq f(x_k) + m(\langle d_k^*, x_{k+1} - x_k\rangle - \epsilon_k) = f(x_k) - m\left[t_k\|d_k^*\|^2 + \epsilon_k\right],$$

from which we conclude that

$$\sum_{k=n}^{j-1}\left[t_k\|d_k\|^2 + \epsilon_k\right] \leq m^{-1}[f(x_n) - f(x_j)] \leq m^{-1}\phi_n.$$

Therefore

$$\|x_j - y\|^2 \leq \|x_n - y\|^2 + 2t^*(m^{-1} - 1)\phi_n,$$

and by taking the limit as $j \to \infty$ we find that

$$\|x_* - y\|^2 \leq \|x_n - y\|^2 + 2t^*(m^{-1} - 1)\phi_n.$$

Now set $y = u_n$ to obtain

$$\|x_* - u_n\|^2 \leq \xi_n^2 + 2t^*(m^{-1} - 1)\phi_n.$$

The bounds (14) and (16) now yield, for $n \geq N'$,

$$\|x_* - u_n\| \leq \tau\theta^n,$$

with

$$\tau = \left(\lambda^2 + 2t^*(m^{-1} - 1)\kappa\right)^{1/2}.$$

Then we have, using (16) again,

$$\|x_n - x_*\| \leq \xi_n + \|x_* - u_n\| \leq (\lambda + \tau)\theta^n,$$

so that $\{x_n\}$ converges at least R-linearly to the limit x_*, as claimed.

□

References

1. Brézis, H., Lions, P.L. (1978): Produits infinis de résolvantes. Israel J. Math. **29**, 329–345
2. Brøndsted, A., Rockafellar, R.T. (1965): On the subdifferentiability of convex functions. Proc. Am. Math. Soc. **16**, 605–611
3. Correa, R., Lemaréchal, C. (1993): Convergence of some algorithms for convex minimization. Math. Program. **62**, 261–275
4. Gill, P.E., Murray, W., Wright, M.H. (1981): Practical Optimization. Academic Press, London
5. Hiriart-Urruty, J.-B., Lemaréchal, C. (1993): Convex Analysis and Minimization Algorithms II, volume 306 of Grundlehren der mathematische Wissenschaften. Springer, Berlin
6. Kiwiel, K.C. (1985): Methods of Descent for Nondifferentiable Optimization. Lect. Notes Math. **1133**. Springer, Berlin
7. Kiwiel, K.C. (1990): Proximity control in bundle methods for convex nondifferentiable optimization. Math. Program. **46**, 105–122
8. Kiwiel, K.C. (1999): The efficiency of proximal bundle methods. Revised edition, March 1999. Submitted to J. Optim. Theory Appl.
9. Luque, F.J. (1984): Asymptotic convergence analysis of the proximal point algorithm. SIAM J. Control Optim. **22**, 277–293
10. Ortega, J.M., Rheinboldt, W.C. (1970): Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York
11. Robinson, S.M. (1979): Generalized equations and their solutions, part I: Basic theory. Math. Program. Studies **10**, 128–141
12. Rockafellar, R.T. (1970): Convex Analysis. Princeton University Press, Princeton, NJ
13. Rockafellar, R.T. (1974): Conjugate Duality and Optimization. Number 16 in CBMS Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia, PA
14. Rockafellar, R.T. (1976): Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**, 97–116
15. Rockafellar, R.T. (1976): Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898
16. Schramm, H., Zowe, J. (1992): A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results. SIAM J. Optim. **2**, 121–152
17. Zhang, R., Treiman, J. (1995): Upper-Lipschitz multifunctions and inverse subdifferentials. Nonlinear Analysis: Theory, Methods, Appl. **24**, 273–286
18. Zhu, C. (1996): Aymptotic convergence analysis of some inexact proximal point algorithms for minimization. SIAM J. Optim. **6**, 626–637
19. Zowe, J. (1989): The BT-algorithm for minimizing a nonsmooth functional subject to linear constraints. In: Clarke, F.H., Demyanov, V., Giannessi, F., eds., Nonsmooth Optimization and Related Topics. Plenum Press, New York