TAIWANESE JOURNAL OF MATHEMATICS
Vol. 14, No. 6, pp. 2307-2328, December 2010
This paper is available online at http://www.tjm.nsysu.edu.tw/

CONVERGENCE OF A PROXIMAL-LIKE ALGORITHM IN THE PRESENCE OF COMPUTATIONAL ERRORS

Alexander J. Zaslavski

Abstract. We study the convergence of a proximal-like minimization algorithm using Bregman functions. We extend the convergence results by Censor and Zenios (1992) and by Chen and Teboulle (1993) by showing that the convergence of the algorithm is preserved in the presence of computational errors.

1. Introduction

We consider the space R^n with the Euclidean norm $||\cdot||$ and the convex optimization problem

$$(P) \qquad \min\{f(x):\ x \in R^n\},$$

where $f : R^n \to R^1 \cup \{\infty\}$ is a convex lower semicontinuous bounded from below function which is not identically ∞. One method of solving (P) is to regularize the objective function by using the proximal mapping introduced by Moreau [15]. Given a real positive number λ, a proximal approximation of f is defined by

$$f_\lambda(x) = \inf\{f(u) + (2\lambda)^{-1}||x-u||^2 :\ u \in R^n\}.$$

As proved by Moreau [15], the function f_λ is a convex and differentiable, and when it is minimized it possesses the same set of minimizers and the same optimal value as problem (P). Using these properties, Martinet [14] introduced the proximal minimization algorithm for solving problem (P). The method is as follows: given an initial point $x^{(0)} \in R^n$, a sequence $\{x^{(k)}\}_{k=0}^\infty$ is generated by solving

$$x^{(k)} = \text{argmin}\{f(x) + (2\lambda_{k-1})^{-1}||x - x^{(k-1)}||^2\},$$

Received January 23, 2009, accepted April 25, 2009.
Communicated by B. S. Mordukhovich.
2000 *Mathematics Subject Classification*: 49M37, 90C25, 90C30.
Key words and phrases: Bregman functions, Proximal methods, Convex programming.

where $\{\lambda_k\}_{k=0}^{\infty}$ is a sequence of positive numbers. For further contributions on proximal methods see [1, 2, 6, 7, 9-13, 17, 18, 20] and the references mentioned therein.

In particular, Censor and Zenios [6] and Chen and Teboulle [7] studied the convergence of a method of the form

$$x^{(k)} = \text{argmin}\{f(x) + \lambda_{k-1}^{-1} D(x, x^{(k-1)})\},$$

with D being a Bregman's distance or D-function defined below. Auslender and Teboulle [1] used proximal methods for convex minimization problems on subsets of a finite-dimensional Euclidean apace which are the closure of open convex sets. A broad class of optimization algorithms based on Bregman distances in Banach spaces is unified around the notion of Bregman monotonicity by Bauschke, Borwein and Combettes [2]. Eckstein [10] used proximal methods for solving the variational inequality problem formed by a general set-valued maximal monotone operator. Inexact proximal methods were investigated by Solodov and Svaiter in [19] by using a unified framework.

In the present paper our goal is to show that the convergence of the algorithm by Chen and Teboulle [7] described above is preserved in the presence of computational errors. In this algorithm Chen and Teboulle [7] considered a sequence $\{x^{(k)}\}_{k=0}^{\infty}$ such that for any natural number k, $x^{(k)}$ is a solution of the auxiliary minimization problem and established that $f(x^{(k)})$ converges to the infimum of the function f and $x^{(k)}$ converges to the set of solutions of the problem (P) as $k \to \infty$.

It should be mentioned that in practice computations introduce numerical errors and if one uses methods in order to solve the auxiliary minimization problems these methods usually provide only approximate solutions of the problems. More precisely, the algorithm generates a sequence $\{x^{(k)}\}_{k=0}^{\infty}$ such that for any natural number k,

$$x^{(k)} + e^{(k)} = \text{argmin}\{f(x) + \lambda_{k-1}^{-1} D(x, x^{(k-1)})\},$$

where $e^{(k)}$ is a computational error at the iteration k. Clearly, it is very important from the view of practice to study the convergence of iterations of the algorithm in the presence of computational errors. Note that most results on proximal methods which exist in the literature do not take into account computational errors. Convergence results on proximal methods with computational errors were obtained in Eckstein [10] under assumption that the computational errors $e^{(k)}$ are summable. It means that $\sum_{k=1}^{\infty} ||e^{(k)}|| < \infty$. In the present paper our goal is to establish the convergence of the proximal algorithm for solving the problem (P) in the presence of the computational errors $e^{(k)}$ without assuming their summability. Actually we do not even assume that the computational errors tend to zero when k tends to infinity. All this makes our case more realistic from the point of view of practice.

More precisely, we show (Theorem 1.2) that for a given positive number ϵ there exists $\delta > 0$ such that if the computational errors satisfy $||e^{(k)}|| \leq \delta$ for all natural numbers k, then for all sufficiently large natural numbers k, $x^{(k)}$ belongs to an ϵ-neighborhood of the set of solutions of the problem (P).

Another type of approximation of solutions is considered in Theorem 1.3. In that theorem we establish that for a given positive number ϵ there exists $\delta > 0$ such that if a sequence $\{x^{(k)}\}_{k=0}^{\infty}$ satisfies for any integer $k \geq 0$

$$f(x^{(k+1)}) + \lambda_k^{-1} D_\psi(x^{(k+1)}, x^{(k)}) \leq \inf\{f(z) + \lambda_k^{-1} D_\psi(z, x^{(k)}) :\ z \in R^n\} + \delta,$$

then for all sufficiently large natural numbers k,

$$f(x^{(k)}) \leq \inf\{f(z) :\ z \in R^n\} + \epsilon.$$

Note that the type of approximation established by Theorem 1.2 implies the type of approximation obtained in Theorem 1.3 if the function f is continuous. In view of assumption (A1) which is posed below (see also Proposition 1.1) the type of approximation obtained in Theorem 1.3 implies the type of approximation in Theorem 1.2.

Put

$$f_* = \inf\{f(z) :\ z \in R^n\}. \tag{1.1}$$

Note that f_* is a finite number.

For each $x \in R^n$ and each $A \subset R^n$ put

$$\rho(x, A) = \inf\{||x - y|| :\ y \in A\}. \tag{1.2}$$

For each $x \in R^n$ and each $r > 0$ set

$$\begin{aligned} B(x, r) &= \{y \in R^n :\ ||x - y|| < r\}, \\ \bar{B}(x, r) &= \{y \in R^n :\ ||x - y|| \leq r\}. \end{aligned} \tag{1.3}$$

In the paper we assume that the following property holds.

(A1) There is $c_f > f_*$ such that the set $\{z \in R^n :\ f(z) < c_f\}$ is bounded.

Assumption (A1) easily implies the following result.

Proposition 1.1.

$$\lim_{||x|| \to \infty} f(x) = \infty.$$

In this paper we use the notations and definitions introduced in [16]. In particular, dom f, ran f and $\bar{C}$ denote the domain and range of f and the closure of the set C, respectively.

Given a differentiable function ψ, a measure of distance based on Bregman's distance [3] is defined by

$$D_\psi(x,y) = \psi(x) - \psi(y) - < x - y, \nabla\psi(y) >,$$

where $< \cdot, \cdot >$ denotes the inner product in R^n and $\nabla\psi$ is the gradient of ψ.

The function ψ is called a Bregman function if it satisfies the properties given in the definition below (see [5, 8]).

Let $S \subset R^n$ be a nonempty open set. Then $\psi : \bar{S} \to R^1$ is called a Bregman function with zone S if the following hold:

(i) ψ is continuously differentiable on S;

(ii) ψ is strictly convex and continuous on $\bar{S}$;

(iii) For every $\alpha \in R^1$ the partial level sets $L_1(y,\alpha) = \{x \in \bar{S} :\ D_\psi(x,y) \le \alpha\}$ and $L_2(x,\alpha) = \{y \in S :\ D_\psi(x,y) \le \alpha\}$ are bounded for every $y \in S$ and every $x \in \bar{S}$.

(iv) If $\{y^{(k)}\}_{k=1}^\infty \in S$ converges to y^*, then $D_\psi(y^*, y^{(k)}) \to 0$ as $k \to \infty$.

(v) If $\{x^{(k)}\}_{k=1}^\infty$ and $\{y^{(k)}\}_{k=1}^\infty$ are sequences such that $y^{(k)} \to y^* \in \bar{S}$, $\{x^{(k)}\}_{k=1}^\infty$ is bounded, and if $D_\psi(x^{(k)}, y^{(k)}) \to 0$ as $k \to \infty$, then $x^{(k)} \to y^*$ as $k \to \infty$.

$D_\psi(\cdot,\cdot)$ is not a distance (it might not be symmetric and might not satisfy the triangle inequality), but by the strict convexity of ψ it follows immediately that $D_\psi(x,y) \ge 0$ and is equal to zero if and only if $x = y$. With the special choice $S = R^n$ and $\psi(x) = (1/2)||x||^2$ one obtains $D_\psi(x,y) = (1/2)||x-y||^2$. Another important example useful in applications [4] is obtained by choosing the entropy kernel.

In [4] Butnariu, Byrne and Censor showed that $\psi : \bar{S} \to R^1$ is a Bregman function if and only if the property (ii) holds, ψ is differentiable on S, for each $x \in \bar{S}$ and each $\alpha > 0$ the level set $L_2(x,\alpha)$ is bounded and if the following property holds:

if $\{x^{(k)}\}_{k=1}^\infty \subset S$ and $x^* = \lim_{k\to\infty} x^{(k)} \in \bar{S}\setminus S$, then $\lim_{k\to\infty} < \nabla g(x^{(k)}), x^* - x^{(k)} >= 0$.

In this paper we assume that $S \subset R^n$ is a nonempty open set, $\bar{S}$ is its closure and $\psi : \bar{S} \to R^1$ is a Bregman function with zone S such that ran $\nabla\psi = R^n$.

For any $\Delta \ge 1$ put

$$S_\Delta = \{z \in S :\ ||z|| \le \Delta \text{ and } B(z, \Delta^{-1}) \subset S\}. \tag{1.4}$$

Clearly,

$$\cup\{S_\Delta :\ \Delta \in [1,\infty)\} = S.$$

The following simple auxiliary result will be proved in Section 2.

Proposition 1.2. *Let $\Delta \geq 1$. Then S_Δ is a closed subset of R^n.*

Corollary 1.1. *Let $\Delta \geq 1$. Then*

$$\sup\{|\psi(z)| : z \in S_\Delta\} < \infty, \ \sup\{|\nabla\psi(z)| : \ z \in S_\Delta\} < \infty$$

and for each $u \in R^n$, $\sup\{D_\psi(u, z) : \ z \in S_\Delta\} < \infty$.

In the sequel we assume that

$$\text{dom}(f) \subset S. \tag{1.5}$$

The following proposition will be proved in Section 2.

Proposition 1.3. $\lim_{\Delta\to\infty} \inf\{f(z) : \ z \in R^n \setminus S_\Delta\} = \infty$.

Set

$$X_* = \{x \in R^n : \ f(x) = f_*\}. \tag{1.6}$$

By (1.1), (A1) and the lower semicontinuity of f, X_* is a nonempty closed convex bounded set.

It is not difficult to see that the following proposition holds.

Proposition 1.4. *For each $\epsilon > 0$ there is $\delta > 0$ such that if $x \in R^n$ satisfies $f(x) \leq f_* + \delta$, then $\rho(x, X_*) \leq \epsilon$.*

For each $x \in S$ and each $\lambda > 0$ set

$$T_\lambda x = \text{argmin}_{z\in R^n}\{f(z) + \lambda^{-1} D_\psi(z, x)\}. \tag{1.7}$$

Note that for all $x \in S$ and $\lambda > 0$, $T_\lambda x$ is well-defined and

$$T_\lambda x \in \text{dom}(f) \subset S. \tag{1.8}$$

The following convergence result was established in [7, Theorem 3.4].

Theorem 1.1. *Let $x^{(0)} \in S$, $\lambda_k > 0$, $k = 0, 1, \ldots$ satisfy $\sum_{k=0}^{\infty} \lambda_k = \infty$ and let for any natural number k, $x^{(k)} = T_{\lambda_{k-1}} x^{(k-1)}$. Then for each $u \in \bar{S}$ and each natural number k*

$$f(x^{(k)}) - f(u) \leq \left(\sum_{i=0}^{k-1} \lambda_i\right)^{-1} D_\psi(u, x^{(0)}).$$

Theorem 1.1 is a generalization of a convergence result of [6]. Here we extend Theorem 1.1 and show that the convergence of the algorithm is preserved in the presence of computational errors. Namely we prove the following two results.

Theorem 1.2. *Let $\bar\lambda_0 > 0$, $\epsilon > 0$ and $\Delta_0 \ge 1$. Then there exists a natural number n_0 such that for each $\bar\lambda_1 > \bar\lambda_0$ there exists $\delta > 0$ such that the following assertion holds.*

For each sequence $\{\lambda_k\}_{k=0}^{\infty} \subset [\bar\lambda_0, \bar\lambda_1]$ and for each sequence $\{x^{(k)}\}_{k=0}^{\infty} \subset S$ satisfying

$$x^{(0)} \in S_{\Delta_0} \text{ and } ||x^{(k+1)} - T_{\lambda_k} x^{(k)}|| \le \delta \text{ for all integers } k \ge 0$$

the inequality $\rho(x^{(k)}, X_*) \le \epsilon$ holds for all integers $k \ge n_0$.

Theorem 1.3. *Let $0 < \bar\lambda_0 < \bar\lambda_1$, $\epsilon_0 > 0$ and $\Delta_0 \ge 1$. Then there exist a natural number n_0 and $\delta > 0$ such that the following assertion holds.*

For each sequence $\{\lambda_k\}_{k=0}^{\infty} \subset [\bar\lambda_0, \bar\lambda_1]$ and for each sequence $\{x^{(k)}\}_{k=0}^{\infty} \subset S$ satisfying

$$x^{(0)} \in S_{\Delta_0}$$

and

$$f(x^{(k+1)}) + \lambda_k^{-1} D_\psi(x^{(k+1)}, x^{(k)}) \le f(T_{\lambda_k} x^{(k)}) + \lambda_k^{-1} D_\psi(T_{\lambda_k} x^{(k)}, x^{(k)}) + \delta$$

for all integers $k \ge 0$ the inequality $f(x^{(k)}) \le f_* + \epsilon_0$ holds for all integers $k \ge n_0$.

The paper is organized as follows. In Section 2 we prove Propositions 1.2 and 1.3 and a lemma which shows the continuity of the mapping $(\lambda, x) \to T_\lambda x$. Section 3 contains auxiliary results for Theorem 1.2 which is proved in Section 4. Auxiliary results for Theorem 1.3 are proved in Section 5 while Theorem 1.3 is proved in Section 6.

2. Proofs of Propositions 1.2 and 1.3 and a Continuity Lemma

Proof of Proposition 1.2. Assume that $x^{(k)} \in S_\Delta$, $k = 1, 2, \dots$ satisfies

$$\lim_{k\to\infty} x^{(k)} = x \text{ and } y \in B(x, \Delta^{-1}). \tag{2.1}$$

Then

$$||x - y|| < \Delta^{-1} \tag{2.2}$$

and there is an integer $k \ge 1$ such that

$$||x^{(k)} - x|| < (\Delta^{-1} - ||x - y||)2^{-1}. \tag{2.3}$$

It is easy to see that

$$(2.4)\ \ ||y-x^{(k)}|| \le |y-x|| + ||x-x^{(k)}|| < ||x-y|| + (\Delta^{-1} - ||x-y||)2^{-1} < \Delta^{-1}.$$

Since $x^{(k)} \in S_\Delta$ we obtain that $y \in S$. Since this inclusion holds for any $y \in B(x, \Delta^{-1})$ we conclude that $x \in S_\Delta$ and S_Δ is closed. Proposition 1.2 is proved.

Proof of Proposition 1.3. Assume that the proposition does not hold. Then there exist sequences $\{\Delta_i\}_{i=1}^\infty \subset (0, \infty)$ and $\{z_i\}_{i=1}^\infty \subset R^n$ such that

$$(2.5)\qquad \begin{aligned} &\lim_{i\to\infty} \Delta_i = \infty,\ z_i \in R^n \setminus S_{\Delta_i},\ i = 1, 2 \ldots, \\ &\sup\{f(z_i) :\ i = 1, 2, \ldots\} < \infty. \end{aligned}$$

In view of (2.5) and Proposition 1.1

$$(2.6)\qquad \sup\{||z_i|| :\ i = 1, 2, \ldots\} < \infty.$$

By (1.4), (2.5) and (2.6) for all sufficiently large natural numbers i

$$(2.7)\qquad B(z_i, \Delta_i^{-1}) \setminus S \ne \emptyset.$$

Extracting a subsequence and re-indexing if necessary we may assume that there exists

$$(2.8)\qquad z = \lim_{i\to\infty} z_i.$$

By (2.8), (2.5), (1.5) and lower semicontinuity of f

$$f(z) < \infty \text{ and } z \in \mathrm{dom}(f) \subset S.$$

There is $\kappa > 0$ such that

$$B(z, \kappa) \subset S.$$

Together with (2.5) and (2.8) this implies that for all sufficiently large natural numbers i

$$B(z_i, \Delta_i^{-1}) \subset S.$$

This contradicts (2.7). The contradiction we have reached proves Proposition 1.3.

Fix

$$(2.9)\qquad u_* \in X_*.$$

Lemma 2.1. *Let* $\Delta_0 \ge 1$, $0 < \bar\lambda_0 < \bar\lambda_1$,

$$(2.10)\qquad \{x^{(k)}\}_{k=0}^\infty \subset S_{\Delta_0},\ \{\lambda_k\}_{k=0}^\infty \subset [\bar\lambda_0, \lambda_1],$$

$$x = \lim_{k\to\infty} x^{(k)},\ \lambda = \lim_{k\to\infty} \lambda_k.$$

Then

$$T_\lambda x = \lim_{k\to\infty} T_{\lambda_k} x^{(k)}. \tag{2.11}$$

Proof. We may assume without loss of generality that

$$u_* \in S_{\Delta_0}. \tag{2.12}$$

By Proposition 1.2 and (2.10)

$$x \in S_{\Delta_0}.$$

By the definition of T_{λ_k}, $x^{(k)}$, (2.10) and (1.7) for each integer $k \geq 0$,

$$\begin{aligned} &f(T_{\lambda_k} x^{(k)}) + \lambda_k^{-1} D_\psi(T_{\lambda_k} x^{(k)}, x^{(k)}) \leq f(u_*) + \lambda_k^{-1} D_\psi(u_*, x^{(k)}) \\ &\leq f(u_*) + \bar{\lambda}_0^{-1}(\psi(u_*) - \psi(x^{(k)}) - < u_* - x^{(k)}, \nabla\psi(x^{(k)}) >) \end{aligned} \tag{2.13}$$

and

$$f(T_\lambda x) + \lambda^{-1} D_\psi(T_\lambda x, x) \leq f(u_*) + \lambda^{-1} D_\psi(u_*, x)$$

$$\leq f(u_*) + \bar{\lambda}_0^{-1}(\psi(u_*) - \psi(x) - < u_* - x, \nabla\psi(x) >). \tag{2.14}$$

Put

$$\begin{aligned} M_0 &= |f(u_*)| + \bar{\lambda}_0^{-1}(2\sup\{|\psi(z)| :\ z \in S_{\Delta_0}\} \\ &+ 2\Delta_0 \sup\{|\nabla\psi(z)| :\ z \in S_{\Delta_0}\}). \end{aligned} \tag{2.15}$$

Clearly, M_0 is finite. It follows from (2.13), (2.14), (2.15), (2.10) and (2.12) that

$$f(T_{\lambda_k} x^{(k)}) \leq M_0,\ k = 0, 1, \dots \text{ and } f(T_\lambda x) \leq M_0. \tag{2.16}$$

In view of (2.16) and Proposition 1.3 there is $\Delta_1 \geq \Delta_0$ such that

$$T_{\lambda_k} x^{(k)} \in S_{\Delta_1},\ k = 0, 1, \dots \text{ and } T_\lambda x \in S_{\Delta_1}. \tag{2.17}$$

We show that (2.11) holds. Assume the contrary. Then in view of (2.17) extracting a subsequence and re-indexing if necessary we may assume that the sequence $\{T_{\lambda_k} x^{(k)}\}_{k=1}^\infty$ converges and

$$T_\lambda x \neq \lim_{k\to\infty} T_{\lambda_k} x^{(k)}.$$

This implies that

$$
\begin{aligned}
&f(T_\lambda x)+\lambda^{-1}D_\psi(T_\lambda x,x)<f(\lim_{k\to\infty}T_{\lambda_k}x^{(k)})\\
&+\lambda^{-1}D_\psi(\lim_{k\to\infty}T_{\lambda_k}x^{(k)},x)-\beta,
\end{aligned}
\tag{2.18}
$$

where $\beta>0$ is a constant. Since f is lower semicontinuous we have

$$
f(\lim_{k\to\infty}T_{\lambda_k}x^{(k)})\le\liminf_{k\to\infty}f(T_{\lambda_k}x^{(k)}).
\tag{2.19}
$$

By (2.10), (2.17) and Proposition 1.2,

$$
\lambda^{-1}D_\psi(\lim_{k\to\infty}T_{\lambda_k}x^{(k)},x)=\lim_{k\to\infty}\lambda_k^{-1}D_\psi(T_{\lambda_k}x^{(k)},x^{(k)}).
\tag{2.20}
$$

Relations (2.19) and (2.20) imply that

$$
\begin{aligned}
&f(\lim_{k\to\infty}T_{\lambda_k}x^{(k)})+\lambda^{-1}D_\psi(\lim_{k\to\infty}T_{\lambda_k}x^{(k)},x)\\
&\le\liminf_{k\to\infty}(f(T_{\lambda_k}x^{(k)})+\lambda_k^{-1}D_\psi(T_{\lambda_k}x^{(k)},x^{(k)})).
\end{aligned}
\tag{2.21}
$$

By (2.18) for all sufficiently large natural numbers k

$$
f(T_\lambda x)+\lambda^{-1}D_\psi(T_\lambda x,x)<f(T_{\lambda_k}x^{(k)})+\lambda_k^{-1}D_\psi(T_{\lambda_k}x^{(k)},x^{(k)})-\beta/2.
\tag{2.22}
$$

By (2.10) and (2.17)

$$
f(T_\lambda x)+\lambda^{-1}D_\psi(T_\lambda x,x)=\lim_{k\to\infty}[f(T_\lambda x)+\lambda_k^{-1}D_\psi(T_\lambda x,x^{(k)})].
\tag{2.23}
$$

In view of (2.22) and (2.23) for all sufficiently large natural numbers k,

$$
\begin{aligned}
&f(T_\lambda x)+(\lambda_k)^{-1}D_\psi(T_\lambda x,x^{(k)})\le f(T_\lambda x)+\lambda^{-1}D_\psi(T_\lambda x,x)+\beta/4\\
&<f(T_{\lambda_k}x^{(k)})+\lambda_k^{-1}D_\psi(T_{\lambda_k}x^{(k)},x^{(k)})-\beta/4.
\end{aligned}
$$

This contradicts the definition of $T_{\lambda_k}x^{(k)}$.

The contradiction we have reached proves (2.11) and Lemma 2.1 itself.

3. Auxiliary Results for Theorem 1.2

Fix

$$
u_*\in X_*.
\tag{3.1}
$$

Lemma 3.1. *Let $\Delta_1\ge 1$ and $\bar\lambda_0>0$. Then there exist $M>0$ and $\Delta_2\ge 1$ such that if $\lambda\ge\bar\lambda_0$, $x\in S_{\Delta_1}$ and if $y\in R^n$ satisfies $||y-T_\lambda x||\le(4\Delta_2)^{-1}$, then*

$$f(T_\lambda x) \le M \text{ and } y \in S_{\Delta_2}.$$

Proof. Put

$$M = |f(u_*)| + \bar\lambda_0^{-1}[|\psi(u_*)| + \sup\{|\psi(z)| : \ z \in S_{\Delta_1}\} \tag{3.2}$$
$$+(|u_*| + \Delta_1)\sup\{|\nabla\psi(z)| : \ z \in S_{\Delta_1}\}].$$

By Proposition 1.3 there exists $\Delta_2 \ge 2$ such that

$$\text{if } z \in R^n \text{ satisfies } f(z) \le M \text{ then } z \in S_{\Delta_2/2}. \tag{3.3}$$

Let

$$\lambda \ge \bar\lambda_0, \ x \in S_{\Delta_1}, \ y \in R^n, \ ||y - T_\lambda x|| \le (4\Delta_2)^{-1}. \tag{3.4}$$

In view of (1.7) and (3.4)

$$f(T_\lambda x) \le f(T_\lambda x) + \lambda^{-1} D_\psi(T_\lambda x, x) \le f(u_*) + \lambda^{-1} D_\psi(u_*, x)$$
$$\le f(u_*) + \bar\lambda_0^{-1}[\psi(u_*) - \psi(x) - < u_* - x, \nabla\psi(x) >]$$
$$\le f(u_*) + \lambda_0^{-1}[|\psi(u_*)| + \sup\{|\psi(z)| : \ z \in S_{\Delta_1}\}$$
$$+(|u_*| + \Delta_1)\sup\{|\nabla\psi(z)| : \ z \in S_{\Delta_1}\}] = M.$$

Combined with (3.3) this implies that $T_\lambda x \in S_{\Delta_2/2}$. Together with (3.4) this implies that $y \in S_{\Delta_2}$. Lemma 3.1 is proved.

Lemma 3.2. *Let* $0 < \bar\lambda_0 > \bar\lambda_1$, $\Delta_0 \ge 1$, $\epsilon > 0$ *and let* m *be a natural number. Then there exists* $\delta > 0$ *such that for each sequence* $\{\lambda_k\}_{k=0}^{m-1} \subset [\bar\lambda_0, \bar\lambda_1]$ *and each sequence* $\{x^{(k)}\}_{k=0}^{m} \subset S$ *which satisfies*

$$x^{(0)} \in S_{\Delta_0}, \tag{3.5}$$
$$||x^{(k+1)} - T_{\lambda_k} x^{(k)}|| \le \delta, \ k = 0, 1, \dots, m-1$$

there exist a sequence $\{\lambda'_k\}_{k=0}^{m-1} \subset [\bar\lambda_0, \bar\lambda_1]$ *and a sequence* $\{y^{(k)}\}_{k=0}^{m} \subset S$ *which satisfies*

$$y^{(0)} \in S_{\Delta_0}, \tag{3.6}$$
$$y^{(k+1)} = T_{\lambda'_k} y^{(k)}, \ k = 0, 1, \dots, m-1,$$

$$||y^{(k)} - x^{(k)}|| \le \epsilon, \ k = 0, \dots, m. \tag{3.7}$$

Proof. First we define by induction sequences $\Delta_i \geq 1$, $i = 0, \dots, m$ and $\delta_i > 0$, $i = 1, \dots, m$. Note that $\Delta_0 \geq 1$ is given.

Assume that an integer k satisfies $0 \leq k < m$, $\Delta_0, \dots, \Delta_k \geq 1$ are defined and $\delta_i > 0$ are defined for all integers i satisfying $1 \leq i \leq k$.

By Lemma 3.1 there exist $\Delta_{k+1} \geq 1$ and $\delta_{k+1} \in (0,1)$ such that the following property holds:

(P1) If $\lambda \geq \bar{\lambda}_0$, $x \in S_{\Delta_k}$ and if $y \in R^n$ satisfies $||y - T_\lambda x|| \leq \delta_{k+1}$, then $f(T_\lambda x) \leq \Delta_{k+1}$ and $y \in S_{\Delta_{k+1}}$

Thus by induction we have defined $\Delta_i \geq 1$, $i = 0, \dots, m$ and $\delta_i > 0$, $i = 1, \dots, m$.

Put

$$\bar{\Delta} = \max\{\Delta_i : \ i = 0, \dots, m\}, \ \bar{\delta} = \min\{\delta_i : \ i = 1, \dots, m\}. \tag{3.8}$$

It follows from (3.8), (P1), the choice of Δ_i, $i = 0, \dots, m$ and δ_i, $i = 1, \dots, m$ that the following property holds:

(P2) If $\{\lambda_k\}_{k=0}^{m-1} \subset [\bar{\lambda}_0, \infty)$ and if $\{x^{(k)}\}_{k=0}^m \subset S$ satisfy $x^{(0)} \subset S_{\Delta_0}$ and

$$||x^{(k+1)} - T_{\lambda_k} x^{(k)}|| \leq \bar{\delta}, \ k = 0, \dots, m-1,$$

then

$$x^{(k)} \in S_{\bar{\Delta}}, \ k = 0, 1, \dots, m,$$
$$f(T_{\lambda_k} x^{(k)}) \leq \bar{\Delta}, \ k = 0, \dots, m-1.$$

Assume that the assertion of the lemma does not holds. Then for each natural number q there exist

$$\lambda_{k,q} \in [\bar{\lambda}_0, \bar{\lambda}_1], \ k = 0, \dots, m-1 \tag{3.9}$$

and a sequence $\{x^{(k,q)}\}_{k=0}^m \subset S$ such that

$$\begin{gathered} x^{(0,q)} \in S_{\Delta_0}, \\ ||x^{(k+1,q)} - T_{\lambda_{k,q}} x^{(k,q)}|| \leq \bar{\delta} q^{-1}, \ k = 0, \dots, m-1 \end{gathered} \tag{3.10}$$

and that the following property holds:

(P3) For each sequence $\{\lambda'_k\}_{k=0}^{m-1} \subset [\bar{\lambda}_0, \bar{\lambda}_1]$ and each sequence $\{y^{(k)}\}_{k=0}^m \subset S$ which satisfies

$$y^{(0)} \in S_{\Delta_0}, \ y^{(k+1)} = T_{\lambda'_k} y^{(k)}, \ k = 0, \dots, m-1$$

the following inequality holds:

$$\max\{||y^{(k)} - x^{(k,q)}|| : \ k = 0, \dots, m\} > \epsilon.$$

By (P2), (3.9) and (3.10) for all natural numbers q

$$x^{(k,q)} \in S_{\bar{\Delta}}, \ k = 0, \dots, m, \qquad f(T_{\lambda_{k,q}} x^{(k,q)}) \le \bar{\Delta}, \ k = 0, \dots, m-1. \tag{3.11}$$

Extracting a subsequence ad re-indexing if necessary we may assume without loss of generality that for any $k = 0, \dots, m$ there is

$$x^{(k)} = \lim_{q \to \infty} x^{(k,q)} \tag{3.12}$$

and for any $k = 0, \dots, m-1$ there exists

$$\lambda_k = \lim_{q \to \infty} \lambda_{k,q}. \tag{3.13}$$

It follows from (3.11), (3.12), (3.13), (3.9), (3.10) and Proposition 1.2 that

$$x^{(k)} \in S_{\bar{\Delta}}, \ k = 0, \dots, m, \ x^{(0)} \in S_{\Delta_0}, \qquad \lambda_k \in [\bar{\lambda}_0, \bar{\lambda}_1], \ k = 0, \dots, m-1. \tag{3.14}$$

By (3.12), (3.10), (3.13), (3.9), (3.11) and Lemma 2.1 for all $k = 0, \dots, m-1$,

$$x^{(k+1)} = \lim_{q \to \infty} x^{(k+1,q)} = \lim_{q \to \infty} T_{\lambda_{k,q}} x^{(k,q)} = T_{\lambda_k} x^{(k)}. \tag{3.15}$$

By (3.12) for all sufficiently large natural numbers q

$$||x^{(k,q)} - x^{(k)}|| \le \epsilon/4, \ k = 0, \dots, m.$$

This contradicts the property (P3). The contradiction we have reached proves Lemma 3.2.

4. Proof of Theorem 1.2

Fix

$$u_* \in X_*. \tag{4.1}$$

We may assume without loss of generality that

$$\epsilon < 2^{-1}, \ \{z \in R^n : \ \rho(z, X_*) \le \epsilon\} \subset S_{\Delta_0}. \tag{4.2}$$

By Proposition 1.4 there exists $\epsilon_0 \in (0, \epsilon/4)$ such that

$$\text{if } x \in R^n \text{ satisfies } f(x) \leq f_* + 2\epsilon_0, \text{ then } \rho(x, X_*) \leq \epsilon/4. \tag{4.3}$$

Put

$$M_0 = \sup\{|\psi(u_*)| + |\psi(z)| : \ z \in S_{\Delta_0}\} + \sup\{||\nabla\psi(z)|| : \ z \in S_{\Delta_0}\}. \tag{4.4}$$

Choose a natural number $n_0 \geq 4$ such that

$$n_0^{-1}(\min\{1, \bar{\lambda}_0\})^{-1} M_0(1 + |u_*| + \Delta_0) < \epsilon_0/4. \tag{4.5}$$

Assume that $\bar{\lambda}_1 > \bar{\lambda}_0$. By Lemma 3.2 there exists $\delta \in (0, \epsilon_0/4)$ such that the following property holds:

(P4) For each sequence $\{\mu_k\}_{k=0}^{4n_0-1} \subset [\bar{\lambda}_0, \bar{\lambda}_1]$ and each sequence $\{y^{(k)}\}_{k=0}^{4n_0} \subset S$ which satisfies

$$y^{(0)} \in S_{\Delta_0}, \ ||y^{(k+1)} - T_{\mu_k} y^{(k)}|| \leq \delta, \ k = 0, \dots, 4n_0 - 1$$

there exist a sequence $\{\mu'_k\}_{k=0}^{4n_0-1} \subset [\bar{\lambda}_0, \bar{\lambda}_1]$ and a sequence $\{\tilde{y}^{(k)}\}_{k=0}^{4n_0} \subset S$ which satisfies

$$\tilde{y}^{(0)} \in S_{\Delta_0}, \ \tilde{y}^{(k+1)} = T_{\mu'_k} \tilde{y}^{(k)}, \ k = 0, \dots, 4n_0 - 1,$$

$$||y^{(k)} - \tilde{y}^{(k)}|| \leq \epsilon/4, \ k = 0, \dots, 4n_0.$$

Assume that

$$\lambda_k \in [\bar{\lambda}_0, \bar{\lambda}_1] \text{ for all integers } k \geq 0, \tag{4.6}$$

$$\begin{gathered} x^{(k)} \in S, \ k = 0, 1, \dots, \\ x^{(0)} \in S_{\Delta_0}, \ ||x^{(k+1)} - T_{\lambda_k} x^{(k)}|| \leq \delta, \ k = 0, 1, \dots. \end{gathered} \tag{4.7}$$

By (P4), (4.6) and (4.7) there exist sequences

$$\{\tilde{\lambda}_k\}_{k=0}^{4n_0-1} \subset [\bar{\lambda}_0, \bar{\lambda}_1] \text{ and } \{\tilde{x}^{(k)}\}_{k=0}^{4n_0-1} \subset S \tag{4.8}$$

such that

$$\tilde{x}^{(0)} \in S_{\Delta_0}, \ \tilde{x}^{(k+1)} = T_{\tilde{\lambda}_k} \tilde{x}^{(k)}, \ k = 0, \dots, 4n_0 - 1, \tag{4.9}$$

$$||x^{(k)} - \tilde{x}^{(k)}|| \leq \epsilon/4, \ k = 0, 1, \dots, 4n_0 - 1. \tag{4.10}$$

By Theorem 1.1, (4.5), (4.8), (4.9) and (4.4) for each integer $k \in [n_0, 4n_0]$,

$$f(\tilde{x}^{(k)}) - f(u_*) \leq \left(\sum_{i=0}^{k-1} \tilde{\lambda}_i \right)^{-1} D_\psi(u_*, \tilde{x}^{(0)})$$
$$\leq n_0^{-1} \bar{\lambda}_0^{-1} [\psi(u_*) - \psi(\tilde{x}^{(0)}) - < u_* - \tilde{x}^{(0)}, \nabla \psi(\tilde{x}^{(0)}) >]$$
$$\leq n_0^{-1} \bar{\lambda}_0^{-1} [M_0 + (|u_*| + \Delta_0) M_0] < \epsilon_0/4.$$

Together with (4.3) this implies that

$$\rho(\tilde{x}^{(k)}, X_*) \leq \epsilon/4, \ k = n_0, \ldots, 4n_0 - 1. \tag{4.11}$$

By (4.11) and (4.10) for all $k = n_0, \ldots, 4n_0$,

$$\rho(x^{(k)}, X_*) \leq \rho(\tilde{x}^{(k)}, X_*) + ||\tilde{x}^{(k)} - x^{(k)}|| \leq \epsilon/2. \tag{4.12}$$

(Note that (4.12) holds for any sequence $\{\lambda_k\}_{k=0}^\infty \subset [\bar{\lambda}_0, \bar{\lambda}_1]$ and any sequence $\{x^{(k)}\}_{k=0}^\infty \subset S$ satisfying (4.7)).

Now we show that

$$\rho(x^{(k)}, X_*) \leq \epsilon \text{ for all integers } k \geq n_0. \tag{4.13}$$

Assume the contrary. Then there is an integer $j > n_0$ for which

$$\rho(x^{(j)}, X_*) > \epsilon. \tag{4.14}$$

By (4.12) and (4.14),

$$j > 4n_0. \tag{4.15}$$

We may assume without loss of generality that

$$\rho(x^{(i)}, X_*) \leq \epsilon \text{ for all integers } i \in [4n_0, j-1]. \tag{4.16}$$

Set

$$\tilde{\lambda}_i = \lambda_{i-2n_0+j}, \ i = 0, 1, \ldots, \ \bar{x}^{(i)} = x^{(i-2n_0+j)}, \ i = 0, 1, \ldots. \tag{4.17}$$

By (4.6), (4.7) and (4.17),

$$\begin{aligned} &\{\tilde{\lambda}_k\}_{k=0}^\infty \in [\bar{\lambda}_0, \bar{\lambda}_1], \ \bar{x}^{(k)} \in S, \ k = 0, 1, \ldots, \ ||\bar{x}^{(k+1)} - T_{\tilde{\lambda}_k} \bar{x}^{(k)}|| \\ &\leq \delta, \ k = 0, 1, \ldots. \end{aligned} \tag{4.18}$$

In view of (4.2), (4.15), (4.16) and (4.17),

$$\rho(\bar{x}_0, X_*) \leq \epsilon \text{ and } \bar{x}^{(0)} \in S_{\Delta_0}. \tag{4.19}$$

It follows from (4.17), (4.18), (4.19) and the note after the inequality (4.12) that for $k = n_0, \ldots, 4n_0$,

$$\epsilon \geq \rho(\bar{x}^{(k)}, X_*) = \rho(x^{(k-2n_0+j)}, X_*). \tag{4.20}$$

By (4.20), $\rho(x^{(j)}, X_*) = \rho(\bar{x}^{2n_0}, X_*) \leq \epsilon$. This contradicts (4.14). The contradiction we have reached proves (4.13). This completes the proof of Theorem 1.2.

5. Auxiliary Results for Theorem 1.3

Fix

$$u_* \in X_*. \tag{5.1}$$

Lemma 5.1. *Let $\Delta_0 \geq 1$, $0 < \bar{\lambda}_0 < \bar{\lambda}_1$ and $\epsilon > 0$. Then there exists $\delta > 0$ such that for each $x \in S_{\Delta_0}$, each $\lambda \in [\bar{\lambda}_0, \bar{\lambda}_1]$ and each $y \in S$ satisfying*

$$f(y) + \lambda^{-1} D_\psi(y, x) \leq f(T_\lambda x) + \lambda^{-1} D_\psi(T_\lambda x, x) + \delta$$

the inequality $||T_\lambda x - y|| \leq \epsilon$ holds.

Proof. Assume the contrary. Then for each natural number k there are

$$x^{(k)} \in S_{\Delta_0},\ \lambda_k \in [\bar{\lambda}_0, \bar{\lambda}_1],\ y^{(k)} \in S, \tag{5.2}$$

satisfying

$$f(y^{(k)}) + \lambda_k^{-1} D_\psi(y^{(k)}, x^{(k)}) \leq f(T_{\lambda_k} x^{(k)}) + \lambda_k^{-1} D_\psi(T_{\lambda_k} x^{(k)}, x^{(k)}) + 1/k, \tag{5.3}$$

$$||T_{\lambda_k} x^{(k)} - y^{(k)}|| > \epsilon. \tag{5.4}$$

By (5.3), for $k = 1, 2, \ldots$

$$\max\{f(y^{(k)}),\ f(T_{\lambda_k} x^{(k)})\} \leq 1 + f(u_*) + \bar{\lambda}_0^{-1} D_\psi(u_*, x^{(k)}). \tag{5.5}$$

Extracting a subsequence ad re-indexing if necessary we may assume without loss of generality that there exist

$$\bar{x} = \lim_{k \to \infty} x^{(k)},\ \bar{y} = \lim_{k \to \infty} y^{(k)},\ \bar{\lambda} = \lim_{k \to \infty} \lambda_k. \tag{5.6}$$

By (5.2), (5.4), (5.6) and Lemma 2.1,

$$T_{\bar{\lambda}}\bar{x} = \lim_{k\to\infty} T_{\lambda_k}x^{(k)} \text{ and } ||\bar{y} - T_{\bar{\lambda}}\bar{x}|| \geq \epsilon. \tag{5.7}$$

In view of (5.2), (5.3), (5.5) and (5.6),

$$\begin{aligned} f(\bar{y}) + \bar{\lambda}^{-1}D_\psi(\bar{y}, \bar{x}) &\leq \liminf_{k\to\infty} f(y^{(k)}) + \lim_{k\to\infty} \lambda_k^{-1}D_\psi(y^{(k)}, x^{(k)}) \\ &\leq \liminf_{k\to\infty}[f(T_{\lambda_k}x^{(k)}) + \lambda_k^{-1}D_\psi(T_{\lambda_k}x^{(k)}, x^{(k)})]. \end{aligned} \tag{5.8}$$

By (5.7)

$$f(T_{\bar{\lambda}}\bar{x}) + \bar{\lambda}^{-1}D_\psi(T_{\bar{\lambda}}\bar{x}, \bar{x}) < f(\bar{y}) + \bar{\lambda}^{-1}D_\psi(\bar{y}, \bar{x}). \tag{5.9}$$

Relations (5.8) and (5.9) imply that there is $\beta > 0$ such that for all sufficiently large natural numbers k

$$f(T_{\bar{\lambda}}\bar{x}) + \bar{\lambda}^{-1}D_\psi(T_{\bar{\lambda}}\bar{x}, \bar{x}) < f(T_{\lambda_k}x^{(k)}) + \lambda_k^{-1}D_\psi(T_{\lambda_k}x^{(k)}, x^{(k)}) - \beta. \tag{5.10}$$

It follows from (5.2) and (5.6) that

$$\bar{\lambda}^{-1}D_\psi(T_{\bar{\lambda}}\bar{x}, \bar{x}) = \lim_{k\to\infty} \lambda_k^{-1}D_\psi(T_{\bar{\lambda}}\bar{x}, x^{(k)}). \tag{5.11}$$

By (5.10) and (5.11) for all sufficiently large natural numbers k

$$\begin{aligned} f(T_{\bar{\lambda}}\bar{x}) + \lambda_k^{-1}D_\psi(T_{\bar{\lambda}}\bar{x}, x^{(k)}) &\leq f(T_{\bar{\lambda}}\bar{x}) + \bar{\lambda}^{-1}D_\psi(T_{\bar{\lambda}}\bar{x}, \bar{x}) + \beta/2 \\ &< f(T_{\lambda_k}x^{(k)}) + \lambda_k^{-1}D_\psi(T_{\lambda_k}x^{(k)}x^{(k)}) - \beta/2. \end{aligned}$$

This contradicts the definition of $T_{\lambda_k}x^{(k)}$. The contradiction we have reached proves Lemma 5.1.

Lemma 5.2. *Let* $\Delta_0 \geq 1$, $0 < \bar{\lambda}_0 < \bar{\lambda}_1$, $\epsilon > 0$. *Then there exists* $\delta > 0$ *such that for each* $\lambda \in [\bar{\lambda}_0, \bar{\lambda}_1]$ *and each* $y_1, y_2 \in S_{\Delta_0}$ *satisfying* $||y_1 - y_2|| \leq \delta$ *the following inequality holds:*

$$|f(T_\lambda y_1) + \lambda^{-1}D_\psi(T_\lambda y_1, y_1) - (f(T_\lambda y_2) + \lambda^{-1}D_\psi(T_\lambda y_2, y_2))| \leq \epsilon.$$

Proof. We may assume without loss of generality that $\bar{\lambda}_0 < 1$. By Lemma 3.1 there is $\Delta_1 > 1 + \Delta_0$ such that

$$\text{if } x \in S_{\Delta_0},\ \lambda \in [\bar{\lambda}_0, \bar{\lambda}_1],\ \text{then } T_\lambda x \in S_{\Delta_1}. \tag{5.)13}$$

Choose a number M_0 such that

$$M_0 > ||\nabla\psi(z)||,\ z \in S_{\Delta_0}. \tag{5.14}$$

There is $\epsilon_0 > 0$ such that

(5.15) if $z_1, z_2 \in S_{\Delta_1}$ satisfy $||z_1 - z_2|| \le \epsilon_0,$ then $||\psi(z_1) - \psi(z_2)|| \le \bar{\lambda}_0\epsilon/4.$

Since the mapping $(\lambda, y) \to T_\lambda y$, $\lambda \in [\bar{\lambda}_0, \bar{\lambda}_1]$, $y \in S_{\Delta_0}$ is continuous (see Lemma 2.1) there is

$$\delta \in (0, (8(M_0 + 1))^{-1}\epsilon\bar{\lambda}_0)$$

such that the following property holds:

(P5) if $y_1, y_2 \in S_{\Delta_0}$ satisfy $||y_1 - y_2|| \le \delta$ and if $\lambda \in [\bar{\lambda}_0, \bar{\lambda}_1]$, then

$$||T_\lambda y_1 - T_\lambda y_2|| \le \min\{\epsilon_0,\ (\epsilon/4)(M + 1)^{-1}\},$$

$$|\psi(y_1) - \psi(y_2)| \le \bar{\lambda}_0(\epsilon/8),$$

$$||\nabla\psi(y_1) - \nabla\psi(y_2)|| \le (\Delta_0 + \Delta_1)^{-1}\bar{\lambda}_0(\epsilon/8).$$

Let

$$\lambda \in [\bar{\lambda}_0, \bar{\lambda}_1],\ y_1, y_2 \in S_{\Delta_0},\ ||y_1 - y_2|| \le \delta. \tag{5.16}$$

By (5.13) and (5.16),

$$T_\lambda y_1,\ T_\lambda y_2 \in S_{\Delta_1}. \tag{5.17}$$

By the definition of $T_\lambda y_i$, $i = 1, 2$ (see (1.7))

$$\begin{aligned} f(T_\lambda y_1) + \lambda^{-1}D_\psi(T_\lambda y_1, y_1) &\le f(T_\lambda y_2) + \lambda^{-1}D_\psi(T_\lambda y_2, y_1) \\ = f(T_\lambda y_2) + \lambda^{-1}D_\psi(T_\lambda y_2, y_2) &+ \lambda^{-1}[D_\psi(T_\lambda y_2, y_1) - D_\psi(T_\lambda y_2, y_2)]. \end{aligned} \tag{5.18}$$

In view of (5.14), (5.16), (5.17), (P5) and the choice of δ,

$$D_\psi(T_\lambda y_2, y_1) - D_\psi(T_\lambda y_2, y_2) = \psi(T_\lambda y_2) - \psi(y_1) - < T_\lambda y_2 - y_1, \nabla\psi(y_1) >$$

$$-(\psi(T_\lambda y_2) - \psi(y_2) - < T_\lambda y_2 - y_2, \nabla\psi(y_2) >)$$

$$= \psi(y_2) - \psi(y_1) + < T_\lambda y_2 - T_\lambda y_1, \nabla\psi(y_2) - \nabla\psi(y_1) > - < y_2 - y_1, \nabla\psi(y_2) >$$

$$\le |\psi(y_2) - \psi(y_1)| + (\Delta_0 + \Delta_1)|\nabla\psi(y_2) - \nabla\psi(y_1)| + \delta||\nabla\psi(y_2)||$$

$$\le \bar{\lambda}_0(\epsilon/8) + \bar{\lambda}_0(\epsilon/8) + M_0\delta < \epsilon\bar{\lambda}_0.$$

Together with (5.18) this implies that

$$f(T_\lambda y_1) + \lambda^{-1}D_\psi(T_\lambda y_1, y_1) \le f(T_\lambda y_2) + \lambda^{-1}D_\psi(T_\lambda y_2, y_2) + \epsilon.$$

Lemma 5.2 is proved.

Lemma 5.2 implies the following result.

Lemma 5.3. *Let $\Delta_0 \geq 1$ be such that $X_* \subset S_{\Delta_0}$, let $0 < \bar{\lambda}_0 < \bar{\lambda}_1$ and $\epsilon > 0$. Then there exists $\delta > 0$ such that for each $\lambda \in [\bar{\lambda}_0, \bar{\lambda}_1]$ and each $y \in S_{\Delta_0}$ satisfying $\rho(y, X_*) \leq \delta$ the following inequality holds:*

$$|f(T_\lambda y) + \lambda^{-1} D_\psi(T_\lambda y, y) - f_*)| \leq \epsilon.$$

6. Proof of Theorem 1.3

By Theorem 1.2 for each $\Delta \geq 1$ and each $\epsilon > 0$ there exist a natural number $l(\Delta, \epsilon)$ and $\delta(\Delta, \epsilon) > 0$ such that the following property holds:

(P6) For each sequence $\{\lambda_k\}_{k=0}^{\infty} \subset [\bar{\lambda}_0, \bar{\lambda}_1]$ and each sequence $\{x^{(k)}\}_{k=0}^{\infty} \subset S$ satisfying

$$x^{(0)} \in S_\Delta, \ ||x^{(k+1)} - T_{\lambda_k} x^{(k)}|| \leq \delta(\Delta, \epsilon), \ k = 0, 1, \ldots$$

the inequality $\rho(x^{(k)}, X_*) \leq \epsilon$ holds for all integers $k \geq l(\Delta, \epsilon)$.

We may assume without loss of generality that

$$\Delta_0 \geq 4, \ X_* \subset S_{\Delta_0/2}. \tag{6.1}$$

Assume that

$$\{\lambda_k\}_{k=0}^{\infty} \subset [\bar{\lambda}_0, \bar{\lambda}_1], \ \{x^{(k)}\}_{k=0}^{\infty} \subset S, \tag{6.2}$$

$$x^{(0)} \in S_{\Delta_0} \text{ and } ||x^{(k+1)} - T_{\lambda_k} x^{(k)}|| \leq \delta(\Delta_0, (4\Delta_0)^{-1}), \ k = 0, 1 \ldots. \tag{6.3}$$

By (6.2), (6.3) and (P6) for each integers $k \geq l(\Delta_0, (4\Delta_0)^{-1})$,

$$\rho(x^{(k)}, X_*) \leq (4\Delta_0)^{-1}. \tag{6.4}$$

In view of (6.1) and (6.4) for each integer $k \geq l(\Delta_0, (4\Delta_0)^{-1})$,

$$x^{(k)} \in S_{\Delta_0}. \tag{6.5}$$

Thus we have shown that the following property holds:

(P7) For each sequence $\{\lambda_k\}_{k=0}^{\infty} \subset [\bar{\lambda}_0, \bar{\lambda}_1]$ and each sequence $\{x^{(k)}\}_{k=0}^{\infty} \subset S$ satisfying (6.3) the inclusion $x^{(k)} \in S_{\Delta_0}$ holds for each integer $k \geq l(\Delta_0, (4\Delta_0)^{-1})$.

By Lemma 3.1 there exist

$$\bar{\Delta} > \Delta_0,\ \bar{\delta} < \delta(\Delta_0, (4\Delta_0)^{-1}) \tag{6.6}$$

such that the following property holds:

For each sequence $\{\lambda_k\}_{k=0}^{\infty} \subset [\bar{\lambda}_0, \lambda_1]$ and each sequence $\{x^{(k)}\}_{k=0}^{\infty} \subset S$ satisfying

$$x^{(0)} \in S_{\Delta_0} \text{ and } ||x^{(k+1)} - T_{\lambda_k} x^{(k)}|| \le \bar{\delta},\ k = 0, 1, \ldots$$

the inclusion $x^{(k)} \in S_{\bar{\Delta}}$ holds for $k = 0, 1, \ldots,\ l(\Delta_0, (4\Delta_0)^{-1})$.

By the property above, (6.6) and (P7) the following property holds:

(P8) For each sequence $\{\lambda_k\}_{k=0}^{\infty} \subset [\bar{\lambda}_0, \bar{\lambda}_1]$ and each sequence $\{x^{(k)}\}_{k=0}^{\infty} \subset S$ satisfying (6.7) the inclusion $x^{(k)} \in S_{\bar{\Delta}}$ holds for all integers $k \ge 0$.

Lemma 5.3, (6.1) and (6.6) imply that there is $\epsilon_1 \in (0, 4^{-1}\bar{\delta})$ such that the following property holds:

(P9) If $\lambda \in [\bar{\lambda}_0, \bar{\lambda}_1]$ and if $y \in S_{\bar{\Delta}}$ satisfies $\rho(y, X_*) \le \epsilon_1$, then

$$f(T_\lambda y) + \lambda^{-1} D_\psi(T_\lambda y, y) \le f_* + \epsilon_0/4.$$

By Lemma 5.1 there exists

$$\delta \in (0, \epsilon_0/4)$$

such that the following property holds:

(P10) If $x \in S_{\bar{\Delta}}$, $\lambda \in [\bar{\lambda}_0, \bar{\lambda}_1]$ and if $y \in S$ satisfies

$$f(y) + \lambda^{-1} D_\psi(y, x) \le f(T_\lambda x) + \lambda^{-1} D_\psi(T_\lambda x, x) + \delta,$$

then

$$||T_\lambda x - y|| \le \epsilon_2 := \min\{\epsilon_1/4, \delta(\bar{\Delta}, \epsilon_1/4)\}.$$

Assume that

$$\{\lambda_k\}_{k=0}^{\infty} \subset [\bar{\lambda}_0, \bar{\lambda}_1],\ \{x^{(k)}\}_{k=0}^{\infty} \subset S,\ x^{(0)} \in S_{\Delta_0} \tag{6.8}$$

and

$$f(x^{(k+1)}) + \lambda_k^{-1} D_\psi(x^{(k+1)}, x^{(k)}) \le f(T_{\lambda_k} x^{(k)}) + \lambda_k^{-1} D_\psi(T_{\lambda_k} x^{(k)}, x^{(k)}) + \delta \tag{6.9}$$

for all integers $k \ge 0$. By induction we show that

$$x^{(i)} \in S_{\bar{\Delta}} \text{ and } ||x^{(i+1)} - T_{\lambda_i} x^{(i)}|| \le \epsilon_2 \text{ for all integers } i \ge 0. \tag{6.10}$$

Assume that an integer $k \geq 0$, for each integer $i = 0, \dots, k$

$$x^{(i)} \in S_{\bar{\Delta}} \tag{6.11}$$

and that for each integer i satisfying $0 \leq i < k$,

$$||T_{\lambda_i} x^{(i)} - x^{(i+1)}|| \leq \epsilon_2. \tag{6.12}$$

(Note that for $k = 0$ this assumption holds).

By (P10), (6.11), (6.8), (6.9) and the choice of ϵ_2 and ϵ_1,

$$||T_{\lambda_k} x^{(k)} - x^{(k+1)}|| \leq \epsilon_2 < \bar{\delta}/4.$$

Thus (6.12) holds for $i = 0, \dots, k$:

$$||x^{(i+1)} - T_{\lambda_i} x^{(i)}|| \leq \bar{\delta},\ i = 0, \dots, k.$$

By the relation above, (6.8) and (P8),

$$x^{(i)} \in S_{\bar{\Delta}},\ i = 0, 1 \dots, k+1.$$

Therefore the assumption we made for k also holds for $k+1$ holds. This implies (6.10).

By (P6), (6.6), (6.8), (6.10) and the choice of ϵ_2 (see (P10)), for all integers $k \geq l(\bar{\Delta}, \epsilon_1/4)$,

$$\rho(x^{(k)}, X_*) \leq \epsilon_1/4. \tag{6.13}$$

In view of (P9), (6.8), (6.10) and (6.13), for all integers $k \geq l(\bar{\Delta}, \epsilon_1/4)$,

$$f(T_{\lambda_k} x^{(k)}) + \lambda_k^{-1} D_\psi(T_{\lambda_k} x^{(k)}, x^{(k)}) \leq f_* + \epsilon_0/4.$$

Combined with (6.9) this implies that for all integers $k \geq l(\bar{\Delta}, \epsilon_1/4)$,

$$f(x^{(k+1)}) \leq f(x^{(k+1)}) + \lambda_k^{-1} D_\psi(x^{(k+1)}, x^{(k)}) \leq f_* + \epsilon_0/4 + \delta < f_* + \epsilon_0.$$

Theorem 1.3 is proved.

References

1. A. Auslender and M. Teboulle, Interior gradient and proximal methods for convex and conic optimization, *SIAM J. Optim.*, **16** (2006), 697-725.

2. H. H. Bauschke, J. M. Borwein and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.*, **42** (2003), 596-636.

3. L. Bregman, The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming, *USSR Comput. Math. and Math. Phys.*, **7** (1967), 200-217.

4. D. Butnariu, C. Byrne and Y. Censor, Redundant axioms in the definition of Bregman functions, *Journal of Convex Analysis*, **10** (2003), 245-254.

5. Y. Censor and A. Lent, An interval row action method for interval convex programming, *J. Optim. Theory Appl.*, **34** (1981), 321-353.

6. Y. Censor and S. A. Zenios, The proximal minimization algorithm with D-functions, *J. Optim. Theory Appl.*, **73** (1992), 451-464.

7. G. Chen and M. Teboulle, Convergence analysis of a proximal-like minimization algorithm using Bregman functions, *SIAM J. Optim.*, **3** (1993), 538-543.

8. A. DePierro and A. Iusem, A relaxed version of Bregman's method for convex programming, *J. Optim. Theory Appl.*, **5** (1986), 421-440.

9. J. Eckstein, Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Math. Oper. Res.*, **18** (1993), 202-226.

10. J. Eckstein, Approximate iterations in Bregman-function-based proximal algorithms, *Math. Programming*, **83** (1998), 113-123.

11. P. P. B. Eggermont, Multiplicative iterative algorithms for convex programming, *Linear Algebra Appl.*, **130** (1990), 25-42.

12. O. Guler, On the convergence of the proximal point algorithm for convex minimization, *SIAM J. Control Optim.*, **29** (1991), 403-419.

13. B. Lemaire, The proximal algorithm, in: *International Series of Numerical Mathematics*, J. P. Penot, ed., Birkhauser-Verlag Basel, Switzerland, **87** (1989), 73-87.

14. B. Martinet, Pertubation des methodes d'optimisation: Application, *RAIRO Anal. Numer.*, **12** (1978), 153-171.

15. J. J. Moreau, Proximite et dualite dans un espace Hilbertien, *Bull. Soc. Math. France*, **93** (1965), 273-299.

16. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ., 1970.

17. R. T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Math. Oper. Res.*, **1** (1976), 97-116.

18. R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.*, **14** (1976), 877-898.

19. M. V. Solodov and B. F. Svaiter, A unified framework for some inexact proximal point algorithms, *Numerical Functional Analysis and Optimization*, **22** (2001), 1013-1035.

20. M. Teboulle, Entropic proximal mappings with applications to nonlinear programming, *Math. Oper. Res.*, **17** (1992), 670-690.

Alexander J. Zaslavski
Technion-Israel Institute of Technology,
Haifa, 32000,
Israel
E-mail: ajzasl@techunix.technion.ac.il