# Chapter 2
# Subgradient Projection Algorithm

In this chapter we study the subgradient projection algorithm for minimization of convex and nonsmooth functions and for computing the saddle points of convex–concave functions, under the presence of computational errors. We show that our algorithms generate a good approximate solution, if computational errors are bounded from above by a small positive constant. Moreover, for a known computational error, we find out what an approximate solution can be obtained and how many iterates one needs for this.

## 2.1 Preliminaries

The subgradient projection algorithm is one of the most important tools in the optimization theory and its applications. See, for example, [1–3, 12, 30, 44, 51, 79, 89, 92, 95, 96, 105, 108, 109, 112] and the references mentioned therein.

In this chapter we use this method for constrained minimization problems in Hilbert spaces equipped with an inner product denoted by $\langle \cdot, \cdot \rangle$ which induces a complete norm $\| \cdot \|$. For every $z \in R^1$ denote by $\lfloor z \rfloor$ the largest integer which does not exceed $z$: $\lfloor z \rfloor = \max\{i \in R^1 : i \text{ is an integer and } i \le z\}$.

Let $X$ be a Hilbert space. For each $x \in X$ and each $r > 0$ set

$$B_X(x, r) = \{y \in X : \ \|x - y\| \le r\}.$$

For each $x \in X$ and each nonempty set $E \subset X$ set

$$d(x, E) = \inf\{\|x - y\| : \ y \in E\}.$$

Let $C$ be a nonempty closed convex subset of $X$, $U$ be an open convex subset of $X$ such that $C \subset U$ and let $f : U \to R^1$ be a convex function. Recall that for each $x \in U$,



A.J. Zaslavski, *Numerical Optimization with Computational Errors*, Springer Optimization and Its Applications 108, DOI 10.1007/978-3-319-30921-7_2

$$\partial f(x) = \{l \in X : f(y) - f(x) \geq \langle l, y - x\rangle \text{ for all } y \in U\}. \tag{2.1}$$

Suppose that there exist $L > 0$, $M_0 > 0$ such that

$$C \subset B_X(0, M_0), \tag{2.2}$$

$$|f(x) - f(y)| \leq L\|x - y\| \text{ for all } x, y \in U. \tag{2.3}$$

In view of (2.1) and (2.3), for each $x \in U$,

$$\emptyset \neq \partial f(x) \subset B_X(0, L). \tag{2.4}$$

It is easy to see that the following result is true.

**Lemma 2.1.** *Let $z, y_0, y_1 \in X$. Then*

$$\|z - y_0\|^2 - \|z - y_1\|^2 - \|y_0 - y_1\|^2 = 2\langle z - y_1, y_1 - y_0\rangle.$$

The next result is given in [13, 14].

**Lemma 2.2.** *Let $D$ be a nonempty closed convex subset of $X$. Then for each $x \in X$ there is a unique point $P_D(x) \in D$ satisfying*

$$\|x - P_D(x)\| = \inf\{\|x - y\| : y \in D\}.$$

*Moreover,*

$$\|P_D(x) - P_D(y)|| \leq \|x - y\| \textit{ for all } x, y \in X$$

*and for each $x \in X$ and each $z \in D$,*

$$\langle z - P_D(x), x - P_D(x)\rangle \leq 0,$$
$$\|z - P_D(x)\|^2 + \|x - P_D(x)\|^2 \leq \|z - x\|^2.$$

**Lemma 2.3.** *Let $A > 0$ and $n \geq 2$ be an integer. Then the minimization problem*

$$\sum_{i=1}^{n} a_i^2 \to \min$$

$$a = (a_1, \ldots, a_n) \in R^n \text{ and } \sum_{i=1}^{n} a_i = A$$

*has a unique solution $a^* = (a_1^*, \ldots, a_n^*)$ where $a_i^* = n^{-1}A$, $i = 1, \ldots, n$.*

*Proof.* Clearly, the minimization problem has a solution $a^* = (a_1^*, \dots, a_n^*) \in R^n$. Then

$$a_n^* = A - \sum_{i=1}^{n-1} a_i^*$$

and $(a_1^*, \dots, a_{n-1}^*)$ is a minimizer of the function

$$\phi(a_1, \dots, a_{n-1}) := \sum_{i=1}^{n-1} a_i^2 + \left(A - \sum_{i=1}^{n-1} a_i\right)^2, \ (a_1, \dots, a_{n-1}) \in R^{n-1}.$$

It is clear that for all $i = 1, \dots, n-1$,

$$0 = (\partial\phi/\partial a_i)(a_1^*, \dots, a_{n-1}^*) = 2a_i^* - 2\left(A - \sum_{i=1}^{n-1} a_i^*\right) = 2a_i^* - 2a_n^*.$$

Thus $a_i^* = a_n^*$ for all $i = 1, \dots, n-1$ and $a_i^* = n^{-1}A$ for all $i = 1, \dots, n$. Lemma 2.3 is proved.

## 2.2 A Convex Minimization Problem

Let $\delta \in (0, 1]$ and $\{a_k\}_{k=0}^{\infty} \subset (0, \infty)$.

Let us describe our algorithm.

**Subgradient Projection Algorithm**
**Initialization:** select an arbitrary $x_0 \in U$.

**Iterative step:** given a current iteration vector $x_t \in U$ calculate

$$\xi_t \in \partial f(x_t) + B_X(0, \delta)$$

and the next iteration vector $x_{t+1} \in U$ such that

$$\|x_{t+1} - P_C(x_t - a_t\xi_t)\| \le \delta.$$

In this chapter we prove the following result.

**Theorem 2.4.** *Let* $\delta \in (0, 1]$, $\{a_k\}_{k=0}^{\infty} \subset (0, \infty)$ *and let*

$$x_* \in C \tag{2.5}$$

*satisfies*

$$f(x_*) \le f(x) \text{ for all } x \in C. \tag{2.6}$$

*Assume that* $\{x_t\}_{t=0}^{\infty} \subset U$, $\{\xi_t\}_{t=0}^{\infty} \subset X$,

$$\|x_0\| \le M_0 + 1 \tag{2.7}$$

*and that for each integer* $t \ge 0$,

$$\xi_t \in \partial f(x_t) + B_X(0, \delta) \tag{2.8}$$

*and*

$$\|x_{t+1} - P_C(x_t - a_t\xi_t)\| \le \delta. \tag{2.9}$$

*Then for each natural number* $T$,

$$\begin{aligned}
&\sum_{t=0}^{T} a_t(f(x_t) - f(x_*)) \\
&\quad \le 2^{-1}\|x_* - x_0\|^2 + \delta(T+1)(4M_0+1) \\
&\qquad + \delta(2M_0+1)\sum_{t=0}^{T} a_t + 2^{-1}(L+1)^2\sum_{t=0}^{T} a_t^2.
\end{aligned} \tag{2.10}$$

*Moreover, for each natural number* $T$,

$$\begin{aligned}
&f\left(\left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} a_t x_t\right) - f(x_*),\ \min\{f(x_t) :\ t = 0, \dots, T\} - f(x_*) \\
&\le 2^{-1}\left(\sum_{t=0}^{T} a_t\right)^{-1} \|x_* - x_0\|^2 + \left(\sum_{t=0}^{T} a_t\right)^{-1} \delta(T+1)(4M_0+1) \\
&\quad + \delta(2M_0+1) + 2^{-1}\left(\sum_{t=0}^{T} a_t\right)^{-1} (L+1)^2 \sum_{t=0}^{T} a_t^2.
\end{aligned} \tag{2.11}$$

Theorem 2.4 is proved in Sect. 2.4.

We are interested in an optimal choice of $a_t$, $t = 0, 1, \dots$. Let $T$ be a natural number and $A_T = \sum_{t=0}^{T} a_t$ be given. By Theorem 2.4, in order to make the best choice of $a_t$, $t = 0, \dots, T$, we need to minimize the function

$$\phi(a_0, \dots, a_T) := 2^{-1}A_T^{-1}\|x_* - x_0\|^2 + A_T^{-1}\delta(T+1)(4M_0+1)$$

$$+\delta(2M_0+1)+2^{-1}A_T^{-1}(L+1)^2\sum_{t=0}^{T}a_t^2$$

on the set

$$\left\{a=(a_0,\dots,a_T)\in R^{T+1}:\ a_i\ge 0,\ i=0,\dots,T,\ \sum_{i=0}^{T}a_i=A_T\right\}.$$

By Lemma 2.3, this function has a unique minimizer $a^*=(a_0^*,\dots,a_T^*)$ where $a_i^*=(T+1)^{-1}A_T$, $i=0,\dots,T$. This is the best choice of $a_t$, $t=0,1,\dots,T$.

Theorem 2.4 implies the following result.

**Theorem 2.5.** *Let* $\delta\in(0,1]$, $a>0$ *and let* $x_*\in C$ *satisfies*

$$f(x_*)\le f(x)\ \textit{for all}\ x\in C.$$

*Assume that* $\{x_t\}_{t=0}^{\infty}\subset U$, $\{\xi_t\}_{t=0}^{\infty}\subset X$,

$$\|x_0\|\le M_0+1$$

*and that for each integer* $t\ge 0$,

$$\xi_t\in\partial f(x_t)+B_X(0,\delta)$$

*and*

$$\|x_{t+1}-P_C(x_t-a\xi_t)\|\le\delta.$$

*Then for each natural number* $T$,

$$\begin{aligned}&f\left((T+1)^{-1}\sum_{t=0}^{T}x_t\right)-f(x_*),\ \min\{f(x_t):\ t=0,\dots,T\}-f(x_*)\\&\le 2^{-1}(T+1)^{-1}a^{-1}(2M_0+1)^2+a^{-1}\delta(4M_0+1)\\&\quad+\delta(2M_0+1)+2^{-1}(L+1)^2a.\end{aligned}$$

Now we will find the best $a>0$. Since $T$ can be arbitrary large, we need to find a minimizer of the function

$$\phi(a):=a^{-1}\delta(4M_0+1)+2^{-1}(L+1)^2a,\ a\in(0,\infty).$$

Clearly, the minimizer $a$ satisfies

$$a^{-1}\delta(4M_0 + 1) = 2^{-1}(L + 1)^2 a$$

and

$$a = (2\delta(4M_0 + 1))^{1/2}(L + 1)^{-1}$$

and the minimal value of $\phi$ is

$$(2\delta(4M_0 + 1))^{1/2}(L + 1). \tag{2.12}$$

Theorem 2.5 implies the following result.

**Theorem 2.6.** *Let* $\delta \in (0, 1]$,

$$a = (2\delta(4M_0 + 1))^{1/2}(L + 1)^{-1},$$

$x_* \in C$ *satisfies*

$$f(x_*) \le f(x) \text{ for all } x \in C.$$

*Assume that* $\{x_t\}_{t=0}^{\infty} \subset U$, $\{\xi_t\}_{t=0}^{\infty} \subset X$,

$$\|x_0\| \le M_0 + 1$$

*and that for each integer* $t \ge 0$,

$$\xi_t \in \partial f(x_t) + B_X(0, \delta)$$

*and*

$$\|x_{t+1} - P_C(x_t - a\xi_t)\| \le \delta.$$

*Then for each natural number* $T$,

$$\begin{aligned}
&f\left((T + 1)^{-1}\sum_{t=0}^{T} x_t\right) - f(x_*),\ \min\{f(x_t) :\ t = 0, \dots, T\} - f(x_*) \\
&\quad \le 2^{-1}(T + 1)^{-1}(2M_0 + 1)^2(L + 1)(2\delta(4M_0 + 1))^{-1/2} + \delta(2M_0 + 1) \\
&\qquad + 2^{-1}(2\delta(4M_0 + 1))^{1/2}(L + 1) + \delta(4M_0 + 1)(L + 1)(2\delta(4M_0 + 1))^{-1/2}.
\end{aligned}$$

Now we can think about the best choice of $T$. It is clear that it should be at the same order as $\lfloor \delta^{-1} \rfloor$. Putting $T = \lfloor \delta^{-1} \rfloor$, we obtain that

$$f\left((T+1)^{-1}\sum_{t=0}^{T} x_t\right) - f(x_*),\ \min\{f(x_t):\ t=0,\dots,T\} - f(x_*)$$
$$\le 2^{-1}(2M_0+1)^2(L+1)(8M_0+2)^{-1/2}\delta^{1/2} + \delta(2M_0+1)$$
$$+ (8M_0+2)^{1/2}(L+1)\delta^{1/2} + (4M_0+1)(L+1)(8M_0+2)^{-1/2}\delta^{1/2}.$$

Note that in the theorems above $\delta$ is the computational error produced by our computer system.

In view of the inequality above, which has the right-hand side bounded by $c_1\delta^{1/2}$ with a constant $c_1 > 0$, we conclude that after $T = \lfloor \delta^{-1} \rfloor$ iterations we obtain a point $\xi \in U$ such that

$$B_X(\xi, \delta) \cap C \neq \emptyset$$

and

$$f(\xi) \le f(x_*) + c_1\delta^{1/2},$$

where the constant $c_1 > 0$ depends only on $L$ and $M_0$.

## 2.3 The Main Lemma

We use the notation and definitions introduced in Sect. 2.1.

**Lemma 2.7.** *Let $\delta \in (0,1]$, $a > 0$ and let*

$$z \in C. \tag{2.13}$$

*Assume that*

$$x \in U \cap B_X(0, M_0+1), \tag{2.14}$$

$$\xi \in \partial f(x) + B_X(0,\delta) \tag{2.15}$$

*and that*

$$u \in U \tag{2.16}$$

*satisfies*

$$\|u - P_C(x - a\xi)\| \le \delta. \tag{2.17}$$

*Then*

$$a(f(x)-f(z)) \le 2^{-1}\|z-x\|^2 - 2^{-1}\|z-u\|^2 \\ + \delta(4M_0+1+a(2M_0+1)) + 2^{-1}a^2(L+1)^2.$$

*Proof.* In view of (2.15), there exists

$$l \in \partial f(x) \tag{2.18}$$

such that

$$\|l-\xi\| \le \delta. \tag{2.19}$$

By Lemmas 2.1 and 2.2 and (2.13),

$$\begin{aligned} 0 &\le \langle z-P_C(x-a\xi), P_C(x-a\xi)-(x-a\xi)\rangle \\ &= \langle z-P_C(x-a\xi), P_C(x-a\xi)-x\rangle \\ &\quad +\langle a\xi, z-P_C(x-a\xi)\rangle \\ &= 2^{-1}[\|z-x\|^2-\|z-P_C(x-a\xi)\|^2-\|x-P_C(x-a\xi)\|^2] \\ &\quad +\langle a\xi, z-x\rangle + \langle a\xi, x-P_C(x-a\xi)\rangle. \end{aligned} \tag{2.20}$$

Clearly,

$$|\langle a\xi, x-P_C(x-a\xi)\rangle| \le 2^{-1}(\|a\xi\|^2+\|x-P_C(x-a\xi)\|^2). \tag{2.21}$$

It follows from (2.20) and (2.21) that

$$\begin{aligned} 0 &\le 2^{-1}[\|z-x\|^2-\|z-P_C(x-a\xi)\|^2-\|x-P_C(x-a\xi)\|^2] \\ &\quad +\langle a\xi, z-x\rangle + 2^{-1}a^2\|\xi\|^2 + 2^{-1}\|x-P_C(x-a\xi)\|^2 \\ &\le 2^{-1}\|z-x\|^2 - 2^{-1}\|z-P_C(x-a\xi)\|^2 + 2^{-1}a^2\|\xi\|^2 + \langle a\xi, z-x\rangle. \end{aligned} \tag{2.22}$$

Relations (2.2), (2.13), and (2.17) imply that

$$\begin{aligned} &|\|z-P_C(x-a\xi)\|^2-\|z-u\|^2| \\ &\quad = |\|z-P_C(x-a\xi)\|-\|z-u\||(\|z-P_C(x-a\xi)\|+\|z-u\|) \\ &\quad \le \|u-P_C(x-a\xi)\|(4M_0+1) \le (4M_0+1)\delta. \end{aligned} \tag{2.23}$$

By (2.2), (2.13), (2.14), and (2.19),

$$\langle a\xi, z-x\rangle = \langle al, z-x\rangle + \langle a(\xi-l), z-x\rangle$$

$$\leq \langle al, z - x \rangle + a\|\xi - l\|\|z - x\|$$
$$\leq \langle al, z - x \rangle + a\delta(2M_0 + 1). \tag{2.24}$$

It follows from (2.4), (2.18), (2.19), (2.22), (2.23), and (2.24) that

$$\begin{aligned} 0 &\leq 2^{-1}\|z - x\|^2 - 2^{-1}\|z - P_C(x - a\xi)\|^2 + 2^{-1}a^2\|\xi\|^2 + \langle a\xi, z - x \rangle \\ &\leq 2^{-1}\|z - x\|^2 - 2^{-1}\|z - u\|^2 + \delta(4M_0 + 1) + 2^{-1}a^2(L + 1)^2 \\ &\quad + \langle al, z - x \rangle + a\delta(2M_0 + 1). \end{aligned} \tag{2.25}$$

By (2.1), (2.18), and (2.25),

$$a(f(z) - f(x)) \geq \langle al, z - x \rangle$$

and

$$\begin{aligned} a(f(x) - f(z)) &\leq \langle al, x - z \rangle \\ &\leq 2^{-1}\|z - x\|^2 - 2^{-1}\|z - u\|^2 + \delta(4M_0 + 1) + 2^{-1}a^2(L + 1)^2 \\ &\quad + a\delta(2M_0 + 1). \end{aligned}$$

This completes the proof of Lemma 2.7.

## 2.4 Proof of Theorem 2.4

It is clear that

$$\|x_t\| \leq M_0 + 1, \ t = 0, 1, \ldots.$$

Let $t \geq 0$ be an integer. Applying Lemma 2.7 with

$$z = x_*, \ a = a_t, \ x = x_t, \ \xi = \xi_t, \ u = x_{t+1}$$

we obtain that

$$\begin{aligned} a_t(f(x_t) - f(x_*)) &\leq 2^{-1}\|x_* - x_t\|^2 - 2^{-1}\|x_* - x_{t+1}\|^2 \\ &\quad + \delta(4M_0 + 1 + a_t(2M_0 + 1)) + 2^{-1}a_t^2(L + 1)^2. \end{aligned} \tag{2.26}$$

By (2.26), for each natural number $T$,

$$\begin{aligned}
&\sum_{t=0}^{T} a_t(f(x_t) - f(x_*)) \\
&\quad \le \sum_{t=0}^{T}(2^{-1}\|x_* - x_t\|^2 - 2^{-1}\|x_* - x_{t+1}\|^2 \\
&\qquad + \delta(4M_0 + 1) + a_t(2M_0 + 1)\delta + 2^{-1}a_t^2(L + 1)^2) \\
&\quad \le 2^{-1}\|x_* - x_0\|^2 + \delta(T + 1)(4M_0 + 1) \\
&\qquad + \delta(2M_0 + 1)\sum_{t=0}^{T} a_t + 2^{-1}(L + 1)^2 \sum_{t=0}^{T} a_t^2.
\end{aligned}$$

Thus (2.10) is true. Evidently, (2.10) implies (2.11). Theorem 2.4 is proved.

## 2.5 Subgradient Algorithm on Unbounded Sets

We use the notation and definitions introduced in Sect. 2.1. Let $X$ be a Hilbert space with an inner product $\langle \cdot, \cdot \rangle$, $D$ be a nonempty closed convex subset of $X$, $V$ be an open convex subset of $X$ such that

$$D \subset V, \tag{2.27}$$

and $f : V \to R^1$ be a convex function which is Lipschitz on all bounded subsets of $V$. Set

$$D_{\min} = \{x \in D : \ f(x) \le f(y) \text{ for all } y \in D\}. \tag{2.28}$$

We suppose that

$$D_{\min} \neq \emptyset. \tag{2.29}$$

We will prove the following result.

**Theorem 2.8.** *Let $\delta \in (0, 1]$, $M > 0$ satisfy*

$$D_{\min} \cap B_X(0, M) \neq \emptyset, \tag{2.30}$$

$$M_0 \ge 4M + 4, \tag{2.31}$$

$L > 0$ *satisfy*

$$|f(v_1) - f(v_2)| \le L\|v_1 - v_2\| \text{ for all } v_1, v_2 \in V \cap B_X(0, M_0 + 2), \tag{2.32}$$

$$0 < \tau_0 \le \tau_1 \le (L+1)^{-1}, \tag{2.33}$$

$$\epsilon_0 = 2\tau_0^{-1}\delta(4M_0+1) + 2\delta(2M_0+1) + 2\tau_1(L+1)^2 \tag{2.34}$$

*and let*

$$n_0 = \lfloor \tau_0^{-1}(2M+2)^2\epsilon_0^{-1} \rfloor. \tag{2.35}$$

*Assume that* $\{x_t\}_{t=0}^{\infty} \subset V$, $\{\xi_t\}_{t=0}^{\infty} \subset X$,

$$\{a_t\}_{t=0}^{\infty} \subset [\tau_0, \tau_1], \tag{2.36}$$

$$\|x_0\| \le M \tag{2.37}$$

*and that for each integer* $t \ge 0$,

$$\xi_t \in \partial f(x_t) + B_X(0,\delta) \tag{2.38}$$

*and*

$$\|x_{t+1} - P_D(x_t - a_t\xi_t)\| \le \delta. \tag{2.39}$$

*Then there exists an integer* $q \in [1, n_0+1]$ *such that*

$$\|x_i\| \le 3M+2, \ i = 0, \dots, q$$

*and*

$$f(x_q) \le f(x) + \epsilon_0 \textit{ for all } x \in D.$$

We are interested in the best choice of $a_t$, $t = 0, 1, \dots$. Assume for simplicity that $\tau_1 = \tau_0$. In order to meet our goal we need to minimize the function

$$2\tau^{-1}\delta(4M_0+1) + 2(L+1)^2\tau, \ \tau \in (0,\infty).$$

This function has a minimizer

$$\tau = (\delta(4M_0+1))^{1/2}(L+1)^{-1},$$

the minimal value of $\epsilon_0$ is

$$2\delta(2M_0+1) + 4(\delta(4M_0+1))^{1/2}(L+1)$$

and $n_0 = \lfloor \Delta \rfloor$ where

$$\Delta = (2(\delta(4M_0+1))^{1/2}(L+1)^{-1})^{-1}(2M+2)^2(2\delta(2M_0+1)$$

$$
\begin{aligned}
&+ 4(\delta(4M_0+1))^{1/2}(L+1)^{-1}) \\
&\le \delta^{-1/2}(4M_0+1)^{-1/2}(L+1)(2M+2)^2(4L+4)^{-1}(4M_0+1)^{-1/2}\delta^{-1/2} \\
&= \delta^{-1}(4M_0+1)^{-1}(2M+2)^2 4^{-1}.
\end{aligned}
$$

Note that in the theorem above $\delta$ is the computational error produced by our computer system. In view of the inequality above, in order to obtain a good approximate solution we need $\lfloor c_1\delta^{-1}\rfloor + 1$ iterations, where

$$c_1 = 4^{-1}(4M_0+1)^{-1}(2M+1)^2.$$

As a result, we obtain a point $\xi \in V$ such that

$$B_X(\xi, \delta) \cap D \neq \emptyset$$

and

$$f(\xi) \le \inf\{f(x) :\ x \in D\} + c_2\delta^{1/2},$$

where the constant $c_2 > 0$ depends only on $L$ and $M_0$.

## 2.6 Proof of Theorem 2.8

By (2.30) there exists

$$z \in D_{\min} \cap B_X(0, M). \tag{2.40}$$

Assume that $T$ is a natural number and that

$$f(x_t) - f(z) > \epsilon_0,\ t = 1, \ldots, T. \tag{2.41}$$

Lemma 2.2, (2.36), (2.37), (2.39), and (2.40) imply that

$$
\begin{aligned}
\|x_1 - z\| &\le \|x_1 - P_D(x_0 - a_0\xi_0)\| + \|P_D(x_0 - a_0\xi_0) - z\| \\
&\le \delta + \|x_0 - z\| + a_0\|\xi_0\| \le 1 + 2M + \tau_1\|\xi_0\|.
\end{aligned}
\tag{2.42}
$$

In view of (2.32), (2.37), and (2.38),

$$
\begin{aligned}
&\xi_0 \in \partial f(x_0) + B_X(0,1) \subset B_X(0, L) + 1, \\
&\|\xi_0\| \le L + 1.
\end{aligned}
\tag{2.43}
$$

It follows from (2.33), (2.40), (2.42), and (2.43) that

$$\|x_1 - z\| \le 2M + 2, \tag{2.44}$$

$$\|x_1\| \le 3M + 2. \tag{2.45}$$

Set

$$U = V \cap \{v \in X : \ \|v\| < M_0 + 2\} \tag{2.46}$$

and

$$C = D \cap B_X(0, M_0). \tag{2.47}$$

By induction we show that for every integer $t \in [1, T]$,

$$\|x_t - z\| \le 2M + 2, \tag{2.48}$$

$$\begin{aligned} &f(x_t) - f(z) \\ &\quad \le (2\tau_0)^{-1}(\|z - x_t\|^2 - \|z - x_{t+1}\|^2) \\ &\qquad + \tau_0^{-1}\delta(4M_0 + 1) + \delta(2M_0 + 1) + 2^{-1}\tau_1(L + 1)^2. \end{aligned} \tag{2.49}$$

In view of (2.44), (2.48) holds for $t = 1$.

Assume that an integer $t \in [1, T]$ and that (2.48) holds. It follows from (2.31), (2.40), (2.46), (2.47), and (2.48) that

$$z \in C \subset B_X(0, M_0), \tag{2.50}$$

$$x_t \in U \cap B_X(0, M_0 + 1). \tag{2.51}$$

Relation (2.39) implies that $x_{t+1} \in V$ satisfies

$$\|x_{t+1} - P_D(x_t - a_t\xi_t)\| \le 1. \tag{2.52}$$

By (2.32), (2.38), and (2.51),

$$\xi_t \in \partial f(x_t) + B_X(0, 1) \subset B_X(0, L + 1). \tag{2.53}$$

It follows from (2.33), (2.36), (2.40), (2.48), (2.53), and Lemma 2.2 that

$$\begin{aligned} \|z - P_D(x_t - a_t\xi_t)\| &\le \|z - x_t + a_t\xi_t\| \\ &\le \|z - x_t\| + \|\xi_t\|a_t \le 2M + 3, \\ \|P_D(x_t - a_t\xi_t)\| &\le 3M + 3. \end{aligned} \tag{2.54}$$

In view of (2.47) and (2.54),

$$P_D(x_t - a_t\xi_t) \in C, \tag{2.55}$$

and

$$P_D(x_t - a_t\xi_t) = P_C(x_t - a_t\xi_t). \tag{2.56}$$

Relations (2.44), (2.52), and (2.54) imply that

$$\|x_{t+1}\| \le 3M + 4, \ x_{t+1} \in U. \tag{2.57}$$

By (2.32), (2.38), (2.39), (2.46), (2.47), (2.50), (2.51), (2.55), (2.56), (2.57), and Lemma 2.7 which holds with

$$x = x_t, \ a = a_t, \ \xi = \xi_t, \ u = x_{t+1},$$

we have

$$\begin{aligned} a_t(f(x_t) - f(z)) &\le 2^{-1}\|z - x_t\|^2 - 2^{-1}\|z - x_{t+1}\|^2 \\ &\quad + \delta(4M_0 + 1 + a_t(2M_0 + 1)) + 2^{-1}a_t^2(L+1)^2. \end{aligned}$$

The relation above, (2.34) and (2.36) imply that

$$\begin{aligned} f(x_t) - f(z) &\le (2\tau_0)^{-1}\|z - x_t\|^2 - (2\tau_0)^{-1}\|z - x_{t+1}\|^2 \\ &\quad + \tau_0^{-1}\delta(4M_0 + 1) + (2M_0 + 1)\delta + 2^{-1}\tau_1(L+1)^2. \end{aligned} \tag{2.58}$$

In view of (2.41), (2.58) and the inclusion $t \in [1, T]$,

$$\begin{aligned} &\|z - x_t\|^2 - \|z - x_{t+1}\|^2 \ge 0, \\ &\|z - x_{t+1}\| \le \|z - x_t\| \le 2M + 2. \end{aligned} \tag{2.59}$$

Therefore we assumed that (2.48) is true and showed that (2.58) and (2.59) hold. Hence by induction we showed that (2.49) holds for all $t = 1, \ldots, T$ and (2.48) holds for all $t = 1, \ldots, T + 1$.

It follows from (2.49) which holds for all $t = 1, \ldots, T$, (2.41) and (2.44) that

$$\begin{aligned} T\epsilon_0 &< T(\min\{f(x_t) : \ t = 1, \ldots, T\} - f(z)) \\ &\le \sum_{t=1}^{T}(f(x_t) - f(z)) \end{aligned}$$

$$\le (2\tau_0)^{-1}\sum_{t=1}^{T}(\|z-x_t\|^2-\|z-x_{t+1}\|^2)$$
$$+T\tau_0^{-1}\delta(4M_0+1)+T(2M_0+1)\delta+2^{-1}T\tau_1(L+1)^2$$
$$\le (2\tau_0)^{-1}(2M+2)^2+T\tau_0^{-1}\delta(4M_0+1)$$
$$+T(2M_0+1)\delta+2^{-1}T\tau_1(L+1)^2.$$

Together with (2.34) and (2.35) this implies that

$$\epsilon_0 < (2\tau_0 T)^{-1}(2M+2)^2+\tau_0^{-1}\delta(4M_0+1)$$
$$+(2M_0+1)\delta+2^{-1}\tau_1(L+1)^2,$$
$$2^{-1}\epsilon_0 < (2\tau_0 T)^{-1}(2M+2)^2,$$
$$T < \tau_0^{-1}(2M+2)^2\epsilon_0^{-1}\le n_0+1.$$

Thus we have shown that if an integer $T$ satisfies (2.41), then $T\le n_0$ and

$$\|z-x_t\|\le 2M+2,\ t=1,\dots,T+1,$$
$$\|x_t\|\le 3M+2,\ t=0,\dots,T+1.$$

This implies that there exists an integer $q\in[1,n_0+1]$ such that

$$\|x_t\|\le 3M+2,\ t=0,\dots,q$$

and

$$f(x_q)-f(z)\le\epsilon_0.$$

Theorem 2.8 is proved.

## 2.7 Zero-Sum Games with Two-Players

We use the notation and definitions introduced in Sect. 2.1.

Let $X$, $Y$ be Hilbert spaces, $C$ be a nonempty closed convex subset of $X$, $D$ be a nonempty closed convex subset of $Y$, $U$ be an open convex subset of $X$, and $V$ be an open convex subset of $Y$ such that

$$C\subset U,\ D\subset V \tag{2.60}$$

and let a function $f : U \times V \to R^1$ possess the following properties:

(i) for each $v \in V$, the function $f(\cdot, v) : U \to R^1$ is convex;
(ii) for each $u \in U$, the function $f(u, \cdot) : V \to R^1$ is concave.

Assume that a function $\phi : R^1 \to [0, \infty)$ is bounded on all bounded sets and positive numbers $M_0, L$ satisfy

$$C \subset B_X(0, M_0),\ D \subset B_Y(0, M_0), \tag{2.61}$$

$$|f(u, v_1) - f(u, v_2)| \le L\|v_1 - v_2\| \text{ for all } u \in U \text{ and all } v_1, v_2 \in V, \tag{2.62}$$

$$|f(u_1, v) - f(u_2, v)| \le L\|u_1 - u_2\| \text{ for all } v \in V \text{ and all } u_1, u_2 \in U. \tag{2.63}$$

Let

$$x_* \in C \text{ and } y_* \in D \tag{2.64}$$

satisfy

$$f(x_*, y) \le f(x_*, y_*) \le f(x, y_*) \tag{2.65}$$

for each $x \in C$ and each $y \in D$.

In the next section we prove the following result.

**Proposition 2.9.** *Let $T$ be a natural number, $\delta \in (0, 1]$, $\{a_t\}_{t=0}^T \subset (0, \infty)$ and let $\{b_t\}_{t=0}^T \subset (0, \infty)$. Assume that $\{x_t\}_{t=0}^{T+1} \subset U$, $\{y_t\}_{t=0}^{T+1} \subset V$, for each $t \in \{0, \dots, T+1\}$,*

$$B(x_t, \delta) \cap C \neq \emptyset,\ B(y_t, \delta) \cap D \neq \emptyset, \tag{2.66}$$

*for each $z \in C$ and each $t \in \{0, \dots, T\}$,*

$$a_t(f(x_t, y_t) - f(z, y_t)) \le \phi(\|z - x_t\|) - \phi(\|z - x_{t+1}\|) + b_t \tag{2.67}$$

*and that for each $v \in D$ and each $t \in \{0, \dots, T\}$,*

$$a_t(f(x_t, v) - f(x_t, y_t)) \le \phi(\|v - y_t\|) - \phi(\|v - y_{t+1}\|) + b_t. \tag{2.68}$$

*Let*

$$\hat{x}_T = \left(\sum_{i=0}^T a_i\right)^{-1} \sum_{t=0}^T a_t x_t,$$

$$\hat{y}_T = \left(\sum_{i=0}^T a_i\right)^{-1} \sum_{t=0}^T a_t y_t. \tag{2.69}$$

*Then*

$$B(\hat{x}_T, \delta) \cap C \neq \emptyset,\ B(\hat{y}_T, \delta) \cap D \neq \emptyset, \tag{2.70}$$

$$\begin{aligned}
&\left| \left( \sum_{t=0}^{T} a_t \right)^{-1} \sum_{t=0}^{T} a_t f(x_t, y_t) - f(x_*, y_*) \right| \\
&\quad \leq \left( \sum_{t=0}^{T} a_t \right)^{-1} \sum_{t=0}^{T} b_t + \left( \sum_{t=0}^{T} a_t \right)^{-1} \sup\{\phi(u) :\ u \in [0, 2M_0 + 1]\},
\end{aligned} \tag{2.71}$$

$$\begin{aligned}
&\left| f(\hat{x}_T, \hat{y}_T) - \left( \sum_{t=0}^{T} a_t \right)^{-1} \sum_{t=0}^{T} a_t f(x_t, y_t) \right| \\
&\quad \leq \left( \sum_{t=0}^{T} a_t \right)^{-1} \sum_{t=0}^{T} b_t + L\delta \\
&\qquad + \left( \sum_{t=0}^{T} a_t \right)^{-1} \sup\{\phi(u) :\ u \in [0, 2M_0 + 1]\},
\end{aligned} \tag{2.72}$$

*and for each $z \in C$ and each $v \in D$,*

$$\begin{aligned}
f(z, \hat{y}_T) \geq f(\hat{x}_T, \hat{y}_T) &- 2\left( \sum_{t=0}^{T} a_t \right)^{-1} \sup\{\phi(s) :\ s \in [0, 2M_0 + 1]\} \\
&- 2\left( \sum_{t=0}^{T} a_t \right)^{-1} \sum_{t=0}^{T} b_t - L\delta,
\end{aligned} \tag{2.73}$$

$$\begin{aligned}
f(\hat{x}_T, v) \leq f(\hat{x}_T, \hat{y}_T) &+ 2\left( \sum_{t=0}^{T} a_t \right)^{-1} \sup\{\phi(s) :\ s \in [0, 2M_0 + 1]\} \\
&+ 2\left( \sum_{t=0}^{T} a_t \right)^{-1} \sum_{t=0}^{T} b_t + L\delta.
\end{aligned} \tag{2.74}$$

**Corollary 2.10.** *Suppose that all the assumptions of Proposition 2.9 hold and that*

$$\tilde{x} \in C,\ \tilde{y} \in D$$

*satisfy*

$$\|\hat{x}_T - \tilde{x}\| \leq \delta,\ \|\hat{y}_T - \tilde{y}\| \leq \delta. \tag{2.75}$$

*Then*

$$|f(\tilde{x}, \tilde{y}) - f(\hat{x}_T, \hat{y}_T)| \le 2L\delta \tag{2.76}$$

*and for each $z \in C$ and each $v \in D$,*

$$\begin{aligned} f(z, \tilde{y}) \ge f(\tilde{x}, \tilde{y}) & \\ & - 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sup\{\phi(s) :\ s \in [0, 2M_0 + 1]\} \\ & - 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} b_t - 4L\delta, \\ f(\tilde{x}, v) \le f(\tilde{x}, \tilde{y}) & \\ & + 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sup\{\phi(s) :\ s \in [0, 2M_0 + 1]\} \\ & + 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} b_t + 4L\delta. \end{aligned}$$

*Proof.* In view of (2.62), (2.63), and (2.75),

$$\begin{aligned} & |f(\tilde{x}, \tilde{y}) - f(\hat{x}_T, \hat{y}_T)| \\ & \le |f(\tilde{x}, \tilde{y}) - f(\tilde{x}, \hat{y}_T)| + |f(\tilde{x}, \hat{y}_T) - f(\hat{x}_T, \hat{y}_T)| \\ & \le L\|\tilde{y} - \hat{y}_T\| + L\|\tilde{x} - \hat{x}_T\| \le 2L\delta \end{aligned}$$

and (2.76) holds.

Let $z \in C$ and $v \in D$. Relations (2.62), (2.63), and (2.75) imply that

$$|f(z, \tilde{y}) - f(z, \hat{y}_T)| \le L\delta,$$
$$|f(\tilde{x}, v) - f(\hat{x}_T, v)| \le L\delta.$$

By the relation above, (2.73), (2.74), and (2.75),

$$\begin{aligned} f(z, \tilde{y}) & \ge f(z, \hat{y}_T) - L\delta \\ & \ge f(\hat{x}_T, \hat{y}_T) - 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sup\{\phi(s) :\ s \in [0, 2M_0 + 1]\} \\ & \quad - 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} b_t - 2L\delta \end{aligned}$$

$$\geq f(\tilde{x},\tilde{y}) - 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sup\{\phi(s):\ s \in [0, 2M_0+1]\}$$

$$- 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} b_t - 4L\delta,$$

$$\begin{aligned}
f(\tilde{x}, v) &\leq f(\hat{x}_T, v) + L\delta \\
&\leq f(\hat{x}_T, \hat{y}_T) + 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sup\{\phi(s):\ s \in [0, 2M_0+1]\} \\
&\quad + 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} b_t + 2L\delta \\
&\leq f(\tilde{x},\tilde{y}) + 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sup\{\phi(s):\ s \in [0, 2M_0+1]\} \\
&\quad + 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} b_t + 4L\delta.
\end{aligned}$$

This completes the proof of Corollary 2.10.

## 2.8 Proof of Proposition 2.9

It is clear that (2.70) is true. Let $t \in \{0, \ldots, T\}$. By (2.65), (2.67), and (2.68),

$$\begin{aligned}
&a_t(f(x_t, y_t) - f(x_*, y_*)) \\
&\quad \leq a_t(f(x_t, y_t) - f(x_*, y_t)) \\
&\quad \leq \phi(\|x_* - x_t\|) - \phi(\|x_* - x_{t+1}\|) + b_t, \qquad (2.77)
\end{aligned}$$

$$\begin{aligned}
&a_t(f(x_*, y_*) - f(x_t, y_t)) \\
&\quad \leq a_t(f(x_t, y_*) - f(x_t, y_t)) \\
&\quad \leq \phi(\|y_* - y_t\|) - \phi(\|y_* - y_{t+1}\|) + b_t. \qquad (2.78)
\end{aligned}$$

In view of (2.77) and (2.78),

$$\sum_{t=0}^{T} a_t f(x_t, y_t) - \sum_{t=0}^{T} a_t f(x_*, y_*)$$
$$\leq \sum_{t=0}^{T} (\phi(\|x_* - x_t\|) - \phi(\|x_* - x_{t+1}\|)) + \sum_{t=0}^{T} b_t$$
$$\leq \phi(\|x_* - x_0\|) + \sum_{t=0}^{T} b_t, \tag{2.79}$$

$$\sum_{t=0}^{T} a_t f(x_*, y_*) - \sum_{t=0}^{T} a_t f(x_t, y_t)$$
$$\leq \sum_{t=0}^{T} (\phi(\|y_* - y_t\|) - \phi(\|y_* - y_{t+1}\|)) + \sum_{t=0}^{T} b_t$$
$$\leq \phi(\|y_* - y_0\|) + \sum_{t=0}^{T} b_t. \tag{2.80}$$

Relations (2.61), (2.64), (2.66), (2.79), and (2.80) imply that

$$\left| \left( \sum_{t=0}^{T} a_t \right)^{-1} \sum_{t=0}^{T} a_t f(x_t, y_t) - f(x_*, y_*) \right|$$
$$\leq \left( \sum_{t=0}^{T} a_t \right)^{-1} \sum_{t=0}^{T} b_t + \left( \sum_{t=0}^{T} a_t \right)^{-1} \sup\{\phi(s) : \ s \in [0, 2M_0 + 1]\}. \tag{2.81}$$

By (2.70), there exists

$$z_T \in C \tag{2.82}$$

such that

$$\|z_T - \hat{x}_T\| \leq \delta. \tag{2.83}$$

In view of (2.82), we apply (2.67) with $z = z_T$ and obtain that for all $t = 0, \dots, T$,

$$a_t (f(x_t, y_t) - f(z_T, y_t))$$
$$\leq \phi(\|z_T - x_t\|) - \phi(\|z_T - x_{t+1}\|) + b_t. \tag{2.84}$$

It follows from (2.63) and (2.83) that for all $t = 0, \dots, T$,

$$|f(z_T, y_t) - f(\hat{x}_T, y_t)| \le L\|z_T - \hat{x}_T\| \le L\delta. \tag{2.85}$$

By (2.84) and (2.85), for all $t = 0, \dots, T$,

$$\begin{aligned} & a_t(f(x_t, y_t) - f(\hat{x}_T, y_t)) \\ & \quad \le a_t(f(x_t, y_t) - f(z_T, y_t)) + a_t L\delta \\ & \quad \le \phi(\|z_T - x_t\|) - \phi(\|z_T - x_{t+1}\|) + b_t + a_t L\delta. \end{aligned} \tag{2.86}$$

Combined with (2.61), (2.66), and (2.82) this implies that

$$\begin{aligned} & \sum_{t=0}^{T} a_t f(x_t, y_t) - \sum_{t=0}^{T} a_t f(\hat{x}_T, y_t) \\ & \quad \le \sum_{t=0}^{T} (\phi(\|z_T - x_t\|) - \phi(\|z_T - x_{t+1}\|)) + \sum_{t=0}^{T} b_t + \sum_{t=0}^{T} a_t L\delta \\ & \quad \le \phi(\|z_T - x_0\|) + \sum_{t=0}^{T} b_t + \sum_{t=0}^{T} a_t L\delta \\ & \quad \le \sup\{\phi(s) : \ s \in [0, 2M_0 + 1]\} + \sum_{t=0}^{T} b_t + \sum_{t=0}^{T} a_t L\delta. \end{aligned} \tag{2.87}$$

Property (ii) and (2.69) imply that

$$\begin{aligned} \sum_{t=0}^{T} a_t f(\hat{x}_T, y_t) &= \left(\sum_{i=0}^{T} a_i\right) \sum_{t=0}^{T} \left( a_t \left(\sum_{i=0}^{T} a_i\right)^{-1} f(\hat{x}_T, y_t)\right) \\ &\le \left(\sum_{t=0}^{T} a_t\right) f(\hat{x}_T, \hat{y}_T). \end{aligned} \tag{2.88}$$

By (2.87) and (2.88),

$$\begin{aligned} & \sum_{t=0}^{T} a_t f(x_t, y_t) - \sum_{t=0}^{T} a_t f(\hat{x}_T, \hat{y}_T) \\ & \quad \le \sum_{t=0}^{T} a_t f(x_t, y_t) - \sum_{t=0}^{T} a_t f(\hat{x}_T, y_t) \\ & \quad \le \sup\{\phi(s) : \ s \in [0, 2M_0 + 1]\} + \sum_{t=0}^{T} b_t + \sum_{t=0}^{T} a_t L\delta. \end{aligned} \tag{2.89}$$

By (2.70), there exists

$$h_T \in D \tag{2.90}$$

such that

$$\|h_T - \hat{y}_T\| \le \delta. \tag{2.91}$$

In view of (2.90), we apply (2.68) with $v = h_T$ and obtain that for all $t = 0, \dots, T$,

$$\begin{aligned} &a_t(f(x_t, h_T) - f(x_t, y_t)) \\ &\quad \le \phi(\|h_T - y_t\|) - \phi(\|h_T - y_{t+1}\|) + b_t. \end{aligned} \tag{2.92}$$

It follows from (2.62) and (2.91) that for all $t = 0, \dots, T$,

$$|f(x_t, h_T) - f(x_t, \hat{y}_T)| \le L\|h_T - \hat{y}_T\| \le L\delta. \tag{2.93}$$

By (2.92) and (2.93), for all $t = 0, \dots, T$,

$$\begin{aligned} &a_t(f(x_t, \hat{y}_T) - f(x_t, y_t)) \\ &\quad \le a_t(f(x_t, h_T) - f(x_t, y_t)) + a_t L\delta \\ &\quad \le \phi(\|h_T - y_t\|) - \phi(\|h_T - y_{t+1}\|) + b_t + a_t L\delta. \end{aligned} \tag{2.94}$$

In view of (2.94),

$$\begin{aligned} &\sum_{t=0}^{T} a_t f(x_t, \hat{y}_T) - \sum_{t=0}^{T} a_t f(x_t, y_t) \\ &\quad \le \sum_{t=0}^{T} (\phi(\|h_T - y_t\|) - \phi(\|h_T - y_{t+1}\|)) + \sum_{t=0}^{T} b_t + \sum_{t=0}^{T} a_t L\delta. \end{aligned} \tag{2.95}$$

Property (i) and (2.69) imply that

$$\begin{aligned} \sum_{t=0}^{T} a_t f(x_t, \hat{y}_T) &= \left(\sum_{i=0}^{T} a_i\right) \sum_{t=0}^{T} \left(a_t \left(\sum_{i=0}^{T} a_i\right)^{-1} f(x_t, \hat{y}_T)\right) \\ &\ge \sum_{t=0}^{T} a_t f(\hat{x}_T, \hat{y}_T). \end{aligned} \tag{2.96}$$

By (2.61), (2.66), (2.90), (2.95), and (2.96),

$$\begin{aligned}
&\sum_{t=0}^{T} a_t f(\hat{x}_T, \hat{y}_T) - \sum_{t=0}^{T} a_t f(x_t, y_t) \\
&\quad \le \sum_{t=0}^{T} a_t f(x_t, \hat{y}_T) - \sum_{t=0}^{T} a_t f(x_t, y_t) \\
&\quad \le \phi(\|h_T - y_0\|) + \sum_{t=0}^{T} b_t + \sum_{t=0}^{T} a_t L\delta \\
&\quad \le \sup\{\phi(s) : \ s \in [0, 2M_0 + 1]\} + \sum_{t=0}^{T} b_t + \sum_{t=0}^{T} a_t L\delta. \qquad (2.97)
\end{aligned}$$

It follows from (2.89) and (2.97) that

$$\begin{aligned}
&\left|\sum_{t=0}^{T} a_t f(\hat{x}_T, \hat{y}_T) - \sum_{t=0}^{T} a_t f(x_t, y_t)\right| \\
&\quad \le \sup\{\phi(s) : \ s \in [0, 2M_0 + 1]\} + \sum_{t=0}^{T} b_t + \sum_{t=0}^{T} a_t L\delta.
\end{aligned}$$

This implies (2.72).

Let $z \in C$. By (2.67),

$$\begin{aligned}
&\sum_{t=0}^{T} a_t f(x_t, y_t) - f(z, y_t)) \\
&\quad \le \sum_{t=0}^{T} [\phi(\|z - x_t\|) - \phi(\|z - x_{t+1}\|)] + \sum_{t=0}^{T} b_t. \qquad (2.98)
\end{aligned}$$

By property (ii) and (2.69),

$$\begin{aligned}
\sum_{t=0}^{T} a_t f(z, y_t) &= \left(\sum_{i=0}^{T} a_i\right) \sum_{t=0}^{T} \left(a_t \left(\sum_{i=0}^{T} a_i\right)^{-1} f(z, y_t)\right) \\
&\le \left(\sum_{t=0}^{T} a_t\right) f(z, \hat{y}_T). \qquad (2.99)
\end{aligned}$$

In view of (2.98) and (2.99),

$$\begin{aligned}
&\sum_{t=0}^{T} a_t f(x_t, y_t) - \sum_{t=0}^{T} a_t f(z, \hat{y}_T) \\
&\quad \leq \sum_{t=0}^{T} a_t (f(x_t, y_t) - f(z, y_t)) \\
&\quad \leq \sum_{t=0}^{T} [\phi(\|z - x_t\|) - \phi(\|z - x_{t+1}\|)] + \sum_{t=0}^{T} b_t \\
&\quad \leq \phi(\|z - x_0\|) + \sum_{t=0}^{T} b_t. \qquad (2.100)
\end{aligned}$$

It follows from (2.61), (2.70), and (2.72) that

$$\begin{aligned}
f(z, \hat{y}_T) &\geq \left(\sum_{i=0}^{T} a_i\right)^{-1} \sum_{t=0}^{T} a_t f(x_t, y_t) \\
&\quad - \left(\sum_{i=0}^{T} a_i\right)^{-1} \sup\{\phi(s) : \ s \in [0, 2M_0 + 1]\} - \left(\sum_{i=0}^{T} a_i\right)^{-1} \sum_{t=0}^{T} b_t \\
&\geq f(\hat{x}_T, \hat{y}_T) - 2 \left(\sum_{t=0}^{T} a_t\right)^{-1} \sup\{\phi(s) : \ s \in [0, 2M_0 + 1]\} \\
&\quad - 2 \left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} b_t - L\delta
\end{aligned}$$

and (2.73) holds.

Let $v \in D$. By (2.68),

$$\begin{aligned}
&\sum_{t=0}^{T} a_t (f(x_t, v) - f(x_t, y_t)) \\
&\quad \leq \sum_{t=0}^{T} [\phi(\|v - y_t\|) - \phi(\|v - y_{t+1}\|)] + \sum_{t=0}^{T} b_t. \qquad (2.101)
\end{aligned}$$

By property (i) and (2.69),

$$\sum_{t=0}^{T} a_t f(x_t, v) = \left(\sum_{i=0}^{T} a_i\right) \sum_{t=0}^{T} \left( a_t \left(\sum_{i=0}^{T} a_i\right)^{-1} f(x_t, v) \right)$$

$$\geq \left(\sum_{t=0}^{T} a_t\right) f(\hat{x}_T, v). \tag{2.102}$$

In view of (2.101) and (2.102),

$$\sum_{t=0}^{T} a_t f(\hat{x}_T, v) - \sum_{t=0}^{T} a_t f(x_t, y_t)$$
$$\leq \phi(\|v - y_0\|) + \sum_{t=0}^{T} b_t.$$

Together with (2.61), (2.66), and (2.72) this implies that

$$f(\hat{x}_T, v) \leq \left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} a_t f(x_t, y_t)$$
$$+ \left(\sum_{t=0}^{T} a_t\right)^{-1} \sup\{\phi(s) :\ s \in [0, 2M_0 + 1]\} + \left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} b_t$$
$$\leq f(\hat{x}_T, \hat{y}_T) + 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sup\{\phi(s) :\ s \in [0, 2M_0 + 1]\}$$
$$+ 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} b_t + L\delta.$$

Therefore (2.74) holds. This completes the proof of Proposition 2.9.

## 2.9 Subgradient Algorithm for Zero-Sum Games

We use the notation and definitions introduced in Sect. 2.1.

Let $X, Y$ be Hilbert spaces, $C$ be a nonempty closed convex subset of $X$, $D$ be a nonempty closed convex subset of $Y$, $U$ be an open convex subset of $X$, and $V$ be an open convex subset of $Y$ such that

$$C \subset U,\ D \subset V. \tag{2.103}$$

For each concave function $g : V \to R^1$ and each $x \in V$ set

$$\partial g(x) = \{l \in Y :\ \langle l, y - x\rangle \geq g(y) - g(x) \text{ for all } y \in V\}. \tag{2.104}$$

Clearly, for each $x \in V$,

$$\partial g(x) = -(\partial(-g)(x)). \tag{2.105}$$

Suppose that there exist $L > 0$, $M_0 > 0$ such that

$$C \subset B_X(0, M_0),\ D \subset B_Y(0, M_0), \tag{2.106}$$

a function $f : U \times V \to R^1$ possesses the following properties:

(i) for each $v \in V$, the function $f(\cdot, v) : U \to R^1$ is convex;
(ii) for each $u \in U$, the function $f(u, \cdot) : V \to R^1$ is concave,

for each $v \in V$,

$$|f(u_1, v) - f(u_2, v)| \le L\|u_1 - u_2\| \text{ for all } u_1, u_2 \in U \tag{2.107}$$

and that for each $u \in U$,

$$|f(u, v_1) - f(u, v_2)| \le L\|v_1 - v_2\| \text{ for all } v_1, v_2 \in V. \tag{2.108}$$

For each $(\xi, \eta) \in U \times V$, set

$$\partial_x f(\xi, \eta) = \{l \in X : \ f(y, \eta) - f(\xi, \eta) \ge \langle l, y - \xi\rangle \text{ for all } y \in U\}, \tag{2.109}$$

$$\partial_y f(\xi, \eta) = \{l \in Y : \ \langle l, y - \eta\rangle \ge f(\xi, y) - f(\xi, \eta) \text{ for all } y \in V\}. \tag{2.110}$$

In view of properties (i) and (ii) and (2.107)–(2.110), for each $\xi \in U$ and each $\eta \in V$,

$$\emptyset \ne \partial_x f(\xi, \eta) \subset B_X(0, L), \tag{2.111}$$

$$\emptyset \ne \partial_y f(\xi, \eta) \subset B_Y(0, L). \tag{2.112}$$

Let

$$x_* \in C \text{ and } y_* \in D$$

satisfy

$$f(x_*, y) \le f(x_*, y_*) \le f(x, y_*) \tag{2.113}$$

for each $x \in C$ and each $y \in D$.

Let $\delta \in (0, 1]$ and $\{a_k\}_{k=0}^{\infty} \subset (0, \infty)$.

Let us describe our algorithm.

**Subgradient Projection Algorithm for Zero-Sum Games**
**Initialization:** select arbitrary $x_0 \in U$ and $y_0 \in V$.

**Iterative step:** given current iteration vectors $x_t \in U$ and $y_t \in V$ calculate

$$\xi_t \in \partial_x f(x_t, y_t) + B_X(0, \delta),$$
$$\eta_t \in \partial_y f(x_t, y_t) + B_Y(0, \delta)$$

and the next pair of iteration vectors $x_{t+1} \in U$, $y_{t+1} \in V$ such that

$$\|x_{t+1} - P_C(x_t - a_t\xi_t)\| \leq \delta,$$
$$\|y_{t+1} - P_D(y_t + a_t\eta_t)\| \leq \delta.$$

In this chapter we prove the following result.

**Theorem 2.11.** *Let $\delta \in (0, 1]$ and $\{a_k\}_{k=0}^{\infty} \subset (0, \infty)$. Assume that $\{x_t\}_{t=0}^{\infty} \subset U$, $\{y_t\}_{t=0}^{\infty} \subset V$, $\{\xi_t\}_{t=0}^{\infty} \subset X$, $\{\eta_t\}_{t=0}^{\infty} \subset Y$,*

$$B_X(x_0, \delta) \cap C \neq \emptyset, \; B_Y(y_0, \delta) \cap D \neq \emptyset \tag{2.114}$$

*and that for each integer $t \geq 0$,*

$$\xi_t \in \partial_x f(x_t, y_t) + B_X(0, \delta), \tag{2.115}$$
$$\eta_t \in \partial_y f(x_t, y_t) + B_Y(0, \delta), \tag{2.116}$$
$$\|x_{t+1} - P_C(x_t - a_t\xi_t)\| \leq \delta \tag{2.117}$$

*and that*

$$\|y_{t+1} - P_D(y_t + a_t\eta_t)\| \leq \delta. \tag{2.118}$$

*Let for each natural number $T$,*

$$\hat{x}_T = \left(\sum_{i=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} a_t x_t, \; \hat{y}_T = \left(\sum_{i=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} a_t y_t. \tag{2.119}$$

*Then for each natural number $T$,*

$$\left|\left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} a_t f(x_t, y_t) - f(x_*, y_*)\right|$$
$$\leq [2^{-1}(2M_0 + 1)^2 + \delta(T + 1)(4M_0 + 1)] \left(\sum_{t=0}^{T} a_t\right)^{-1}$$

$$+ \delta(2M_0+1) + 2^{-1}\left(\sum_{t=0}^{T} a_t\right)^{-1} (L+1)^2 \sum_{t=0}^{T} a_t^2, \tag{2.120}$$

$$\left| f(\hat{x}_T, \hat{y}_T) - \left(\sum_{t=0}^{T} a_t\right)^{-1} \sum_{t=0}^{T} a_t f(x_t, y_t) \right|$$
$$\le [2^{-1}(2M_0+1)^2 + \delta(T+1)(4M_0+1)] \left(\sum_{t=0}^{T} a_t\right)^{-1}$$
$$+ \delta(2M_0+1) + 2^{-1}\left(\sum_{t=0}^{T} a_t\right)^{-1} (L+1)^2 \sum_{t=0}^{T} a_t^2 + L\delta, \tag{2.121}$$

*and for each natural number $T$, each $z \in C$, and each $u \in D$,*

$$f(z, \hat{y}_T) \ge f(\hat{x}_T, \hat{y}_T)$$
$$-(2M_0+1)^2\left(\sum_{t=0}^{T} a_t\right)^{-1} - 2\left(\sum_{t=0}^{T} a_t\right)^{-1} (T+1)\delta(4M_0+1)$$
$$-2\delta(2M_0+1) - \left(\sum_{t=0}^{T} a_t\right)^{-1} (L+1)^2 \sum_{t=0}^{T} a_t^2 - L\delta,$$
$$f(\hat{x}_T, v) \le f(\hat{x}_T, \hat{y}_T)$$
$$+(2M_0+1)^2\left(\sum_{t=0}^{T} a_t\right)^{-1} + 2\left(\sum_{t=0}^{T} a_t\right)^{-1} \delta(T+1)(4M_0+1)$$
$$+2\delta(2M_0+1) + \left(\sum_{t=0}^{T} a_t\right)^{-1} (L+1)^2 \sum_{t=0}^{T} a_t^2 + L\delta.$$

We are interested in the optimal choice of $a_t$, $t = 0, 1, \ldots, T$. Let $T$ be a natural number and $A_T = \sum_{t=0}^{T} a_t$ be given. By Theorem 2.11, in order to make the best choice of $a_t$, $t = 0, \ldots, T$, we need to minimize the function $\sum_{t=0}^{T} a_t^2$ on the set

$$\left\{ a = (a_0, \ldots, a_T) \in R^{T+1} : \ a_i \ge 0, \ i = 0, \ldots, T, \ \sum_{i=0}^{T} a_i = A_T \right\}.$$

By Lemma 2.3, this function has a unique minimizer $a^* = (a_0^*, \ldots, a_T^*)$ where $a_i^* = (T+1)^{-1}A_T$, $i = 0, \ldots, T$ which is the best choice of $a_t$, $t = 0, 1, \ldots, T$.

Now we will find the best $a > 0$. Let $T$ be a natural number and $a_t = a$ for all $t = 0, \ldots, T$. We need to choose $a$ which is a minimizer of the function

$$\begin{aligned}\Psi_T(a) &= ((T+1)a)^{-1}[(2M_0+1)^2 + 2\delta(T+1)(4M_0+1)] \\ &\quad + 2\delta(2M_0+1) + a(L+1)^2 \\ &= (2M_0+1)^2((T+1)a)^{-1} + 2\delta(4M_0+1)a^{-1} + 2\delta(2M_0+1) + (L+1)^2 a.\end{aligned}$$

Since $T$ can be arbitrary large, we need to find a minimizer of the function

$$\phi(a) := 2a^{-1}\delta(4M_0+1) + (L+1)^2 a,\ a \in (0,\infty).$$

In Sect. 2.2 we have already shown that the minimizer is

$$a = (2\delta(4M_0+1))^{1/2}(L+1)^{-1}$$

and the minimal value of $\phi$ is

$$(8\delta(4M_0+1))^{1/2}(L+1).$$

Now our goal is to find the best integer $T > 0$ which gives us an appropriate value of $\Psi_T(a)$. Since in view of the inequalities above, this value is bounded from below by $c_0\delta^{1/2}$ with the constant $c_0$ depending on $L, M_0$, it is clear that in order to make the best choice of $T$, it should be at the same order as $\lfloor \delta^{-1} \rfloor$. For example, $T = \lfloor \delta^{-1} \rfloor$.

Note that in the theorem above $\delta$ is the computational error produced by our computer system. We obtain a good approximate solution after $T = \lfloor \delta^{-1} \rfloor$ iterations. Namely, we obtain a pair of points $\hat{x} \in U$, $\hat{y} \in V$ such that

$$B_X(\hat{x},\delta) \cap C \neq \emptyset,\ B_Y(\hat{y},\delta) \cap D \neq \emptyset$$

and for each $z \in C$ and each $v \in D$,

$$f(z,\hat{y}) \geq f(\hat{x},\hat{y}) - c\delta^{1/2},\ f(\hat{x},v) \leq f(\hat{x},\hat{y}) + c\delta^{1/2},$$

where the constant $c > 0$ depends only on $L$ and $M_0$.

## 2.10 Proof of Theorem 2.11

By (2.106), (2.114), (2.117), and (2.118), for all integers $t \geq 0$,

$$\|x_t\| \leq M_0 + 1,\ \|y_t\| \leq M_0 + 1. \tag{2.122}$$

Let $t \geq 0$ be an integer. Applying Lemma 2.7 with

$$a = a_t,\ x = x_t,\ f = f(\cdot, y_t),\ \xi = \xi_t,\ u = x_{t+1}$$

we obtain that for each $z \in C$,

$$a_t(f(x_t, y_t) - f(z, y_t)) \le 2^{-1}\|z - x_t\|^2 - 2^{-1}\|z - x_{t+1}\|^2 + \delta(4M_0 + 1 + a_t(2M_0 + 1)) + 2^{-1}a_t^2(L + 1)^2. \quad (2.123)$$

Applying Lemma 2.7 with

$$a = a_t,\ x = y_t,\ f = -f(x_t, \cdot),\ \xi = -\eta_t,\ u = y_{t+1}$$

we obtain that for each $v \in D$,

$$a_t(f(x_t, v) - f(x_t, y_t)) \le 2^{-1}\|v - y_t\|^2 - 2^{-1}\|v - y_{t+1}\|^2 + \delta(4M_0 + 1 + a_t(2M_0 + 1)) + 2^{-1}a_t^2(L + 1)^2. \quad (2.124)$$

For all integers $t \ge 0$ set

$$b_t = \delta(4M_0 + 1 + a_t(2M_0 + 1)) + 2^{-1}a_t^2(L + 1)^2$$

and define

$$\phi(s) = 2^{-1}s^2,\ s \in R^1.$$

It is easy to see that all the assumptions of Proposition 2.9 hold and it implies Theorem 2.11.

Springer

http://www.springer.com/978-3-319-30920-0

Numerical Optimization with Computational Errors
Zaslavski, A.J.
2016, IX, 304 p., Hardcover
ISBN: 978-3-319-30920-0