AVERAGED SUBGRADIENT METHODS FOR CONSTRAINED CONVEX OPTIMIZATION AND NASH EQUILIBRIA COMPUTATION

DAN BUTNARIU and ELENA RESMERITA*

Department of Mathematics, University of Haifa
31905 Haifa, Israel

Abstract. We consider the problem of finding minima of convex functions under convex inequality constraints as well as the problem of finding Nash equilibria in n-person constant sum games. We prove that both problems can be solved by algorithms whose basic principles consist of representing the original problems as infinite systems of convex inequalities which, in turn, can be approached by outer projection techniques. Experiments showing how one of these algorithms behaves in test cases are presented and, in context, we describe a numerical method for computing subgradients of convex functions.

Keywords: Constrained optimization problem, Non-cooperative n-person game, Nash equilibrium, Subgradient.

Mathematics Subject Classifications 2000: 90C25, 90C30, 65D30, 65D25

1. Introduction

In this paper we consider the problem of finding minima of convex, not necessarily differentiable, functions over subsets of $\mathbb{R}^n$ defined by convex, not necessarily differentiable, constraints as well as the problem of finding Nash equilibria in non-cooperative n-person games. Our aim is to show that both problems can be solved by what we call *averaged subgradient methods*. The basic idea underlying these methods is to represent the original problem as problems of finding points in the intersection of an infinite family of closed convex sets and to apply outer projection algorithms for solving the latter.

The optimization problem we consider is the following:

$$\left\{\begin{array}{c} \min\ f(x) \\ \text{such that} \\ f_i(x) \leq 0,\ i \in I, \end{array}\right. \tag{1.1}$$

where I is a nonempty finite set and the functions $f : \mathbb{R}^n \to \mathbb{R}$ and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i \in I$, are convex. We presume that the following conditions are satisfied:

(i) The optimization problem (1.1) has (optimal) solutions;

(ii) The set of feasible solutions

$$S = \{x \in \mathbb{R}^n :\ f_i(x) \leq 0,\ i \in I\}$$

*Both authors gratefully acknowledge the support of the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities.

has nonempty interior.

Condition (ii) is a Slater type requirement and its only purpose in that follows is help to simplify the presentation of the method we propose for solving problem (1.1). A careful analysis of the method itself and of the arguments involved in the proof of its convergence shows that, if condition (ii) does not hold but $S \neq \emptyset$, then the method and the convergence argument can be reproduced without essential modifications inside the affine hull of S, aff(S), by replacing the functions f and f_i, with their respective restrictions to aff(S). In this context, one can use the nonemptiness of the relative interior of S which results from the nonemptiness of S instead of condition (ii).

The method we propose for solving the optimization problem (1.1) is described in Section 2 and named the *averaged subgradient method for optimization* ($ASMO$, for short). It belongs to the wide class of iterative procedures known under the generical name of *subgradient methods.* Such procedures were developed and studied from various viewpoints since the early sixties when Goldstein [15], Levitin and Polyak [20], [21], Polyak [29], Rosen [32] have emphasized the convergence properties and the applicability potential of the most fundamental of them. Many subgradient methods for constrained convex optimization problems are presented, discussed, compared and/or classified in surveys like that of Y.I. Lyubich and G.D. Maistrovskii [22] and, more recently, in monographs like [3], [5], [17], [30]. The $ASMO$ falls in the category of "one-step methods" (in the sense given to this term in [20, p. 12]), i.e., the iterate x^{k+1} is of the form $x^k + d^k$, where d^k is a linear combination of subgradients of the functions f and f_i. This distinguishes the $ASMO$ from the subgradient projection methods for solving problem (1.1): by contrast to the ASMO, in subgradient projection methods determining x^{k+1} requires the additional step of projecting a vector $x^k + d^k$ onto the closed convex set S - see, for instance, [3, p. 339]. From this point of view, the oldest relative of the ASMO we are aware of is the algorithm presented by Polyak in [31, p. 595]. However, one should note that Polyak's method is essentially sequential while the ASMO shares with the subgradient projection methods like those described in [18] and in [25] (see also the references therein) the feature of being inherently parallel in the sense that computations at each iterative step can be distributed among independent processors[1]. The closest relative of the $ASMO$ seems to be the procedure proposed in [30, p. 293] whose basic idea is to reduce the problem (1.1) to a global optimization problem. By contrast to $ASMO$, this procedure requires either to determinate estimates of the Lagrange multipliers occurring in the Karush-Kuhn-Tucker conditions associated to (1.1) or to repeatedly minimize convex functions "with strongly elongated level lines [which] are hard to minimize" (cf. [30, p. 294]), and, therefore, can not be efficiently implemented.

The second problem we consider in this paper is that of finding Nash equilibria in a non-cooperative n-person game $\mathbb{G} = (W_1, ..., W_n; F_1, ..., F_n)$ given such that

(iii) The sets of individual strategies are solid bounded sets of the form

$$W_i = \left\{ x \in \mathbb{R}^{r(i)} : g_{i,j}(x) \leq 0,\ j \in J_i \right\}, \tag{1.2}$$

[1]Using the notations in Section 2, one can see that ASMO calculations for finding x^{k+1} can be distributed among $|I_k| + 2$ parallel processors: $|I_k| + 1$ processors are needed for computing functional values and subgradients at x^k and one for the required multiple integral. An updating system should aggregate x^{k+1} from its components and transmit it back to the parallel processors for performing the next iterative step.

where the sets J_i are finite and the functions $g_{i,j} : \mathbb{R}^{r(i)} \to \mathbb{R}$ are convex;

(iv) The payoff functions $F_i : \prod_{i=1}^n \mathbb{R}^{r(i)} \to \mathbb{R}$ are continuous and, for each $y^i \in W_i$, the function $(x^1, ..., x^{i-1}, x^{i+1}, ..., x^n) \to F_i(x^1, ..., x^{i-1}, y^i, x^{i+1}, ..., x^n)$ is convex on $\prod_{j\neq i} \mathbb{R}^{r(j)}$;

(v) For any $x \in \prod_{i=1}^n \mathbb{R}^{r(i)}$ we have $\sum_{i=1}^n F_i(x) = c$, where c is a constant;

(vi) The game $\mathbb{G}$ has a Nash equilibrium, that is, there exists a vector $\bar{x} = (\bar{x}^1, ..., \bar{x}^n) \in W = \prod_{i=1}^n W_i$ such that

$$F_i(\bar{x}) \geq F_i(\bar{x}^1, ..., \bar{x}^{i-1}, x^i, \bar{x}^{i+1}, ..., \bar{x}^n), \tag{1.3}$$

for all $x^i \in W_i$ and for each $i \in \{1, ..., n\}$.

A classical result, due to J. Nash [24], guarantees existence of equilibria in $\mathbb{G}$, when

$$W_i = \left\{ x \in \mathbb{R}_+^{r(i)} : \sum_{j=1}^{r(i)} x_j \leq 1 \right\}, \tag{1.4}$$

(which, obviously, can be represented in the form (1.2)) and all the functions F_i are multilinear. If, for each i, the set W_i given by (1.2) is compact and if, for any vector $(x^1, ..., x^{i-1}, x^{i+1}, ..., x^n) \in \prod_{j\neq i} W_j$, the function $y^i \to F_i(x^1, ..., x^{i-1}, y^i, x^{i+1}, ..., x^n)$ is concave on W_i, then existence of equilibria in $\mathbb{G}$ is ensured by a result in [26]. More general Nash equilibrium existence criteria can be found in [33] and in the references therein.

The existing methods of computing Nash equilibria we are aware of use a variety of techniques. For matrix and bimatrix games (i.e., games with $n = 2$ and bilinear payoff functions F_i) linear and quadratic optimization methods are available (see, for instance, [28]). In more general situations one can represent Nash equilibria as fixed points of specific operators on W and, then, geometric and combinatorial methods of computing fixed points, as those summarized in [34], are applicable. Alternatively, one may use dynamic systems related techniques like those recently described in [2] and [13]. The *averaged subgradient method for computing Nash equilibria* ($ASME$, for short) presented in Section 3 is essentially based on the fact observed in [6], that Nash equilibria can be characterized as solutions of a system of convex inequalities of the form (3.9). Similarly to the $ASMO$, the $ASME$ is a "one-step" procedure (see (3.7)). By contrast to the subgradient projection methods for computing Nash equilibria like those presented in [13, Formula 4], the $ASME$ does not involve determining projections onto the sets W_i.

Most of the iterative optimization and Nash equilibria computation methods we know about proceed by reducing the original problem to a sequence of optimization subproblems which, hopefully, can be effectively solved with exogenous optimization techniques. The *ASMO* and *ASME* open an alternative algorithmic pathway: the subproblems they require solving consist of determining multiple integrals (see (2.1) and (3.3)). Readily available multiple-integration techniques are presented in [11], [12] and [19]. As in the case of the exogenous optimization procedures involved in other algorithms invented for the same purposes, the effectiveness of the multiple-integration methods when applied in the context of *ASMO* and *ASME* depends on the nature of the problem data. The function φ involved in both algorithms, *ASMO* and *ASME*, is essentially a weight function. When the *ASMO* is applied to optimization problems with bounded feasibility sets and in all applications of *ASME*, one can choose $\varphi \equiv 1$. For application of *ASMO* to problems with unbounded feasibility sets one can choose, for instance, $\varphi(x) = \exp\left(-\sum_{i=1}^{n}|x_i|\right)$. However, in both algorithms *ASMO* and *ASME* it is desirable to fit the choice of φ to the problem data in order to make the integration procedure easier to perform.

The basic theoretical idea underlying $ASMO$ as well as $ASME$ consists of reducing the problem to an infinite system of convex inequalities. It was shown in [9, Section 2.4] that solving the infinite systems of convex inequalities like those discussed here can be also done by using Bregman projection type algorithms and, in particular, the algorithm proposed in [8]. It should be noted that, by contrast to $ASMO$ and $ASME$, that approach involves determining projections onto sets with sometimes complicated geometry. In this respect, $ASMO$ and $ASME$ are procedures similar to the outer projection methods discussed in [14].

A natural question is whether $ASMO$ and $ASME$ behave well in practical computations. In order to answer this question we have done some computational experiments by running $ASMO$ on standard test problems available in the literature as well as on problems we have formulated for verifying some qualitative features of the algorithm. Several phenomena which we consistently observed in our experiments are listed in Section 4 together with several illustrative experimental results. It seems that these observations point to behavioral characteristics of $ASMO$, but we did not succeed to prove that mathematically. In our experiments we considered "smooth" as well as "nonsmooth" (in the sense that not all involved functions are differentiable) problems. In nonsmooth cases there is need of a numerical method of computing subgradients. Such a method, which we believe to be new, is presented in the Appendix.

2. Convergence analysis of the averaged subgradient method for optimization

In this section we consider the optimization problem (1.1) under the assumptions (i) and (ii). The *Averaged Subgradient Method for Optimization* ($ASMO$) is the following iterative procedure for producing sequences in $\mathbb{R}^n$:

Step 0 : Choose a *weight vector* $p := (p_i)_{i\in I} \in \mathbb{R}^I$ such that $p_i > 0$ for all $i \in I$ and $\sum_{i\in I} p_i = 1$, fix an *initial point* $x^0 \in \mathbb{R}^n$, choose a continuous function $\varphi : \mathbb{R}^n \to (0, +\infty)$ such that $\rho := \int_S \varphi(y)dy$ is finite and positive and define the probability measure

$$\mu_\varphi(A) = \frac{1}{\rho}\int_A \varphi(y)dy \tag{2.1}$$

for any set A in the σ-algebra $\mathcal{B}_S$ of all Lebesgue measurable subsets of S. Put $k = 0$ and go to Step $(k+1)$.

Step $(\mathbf{k+1})$: (a) Determine a vector $\gamma^k \in \partial f(x^k)$, let $\theta_k = 0$ if $\gamma^k = 0$, and

$$\theta_k = \frac{1}{\|\gamma^k\|^2}\int_S \max\left\{0, f(x^k) - f(y)\right\} d\mu_\varphi(y), \tag{2.2}$$

otherwise.

(b) Let

$$I_k := \left\{i \in I : f_i(x^k) > 0\right\}, \tag{2.3}$$

and, for each $i \in I_k$, determine a vector $\gamma^{k,i} \in \partial f_i(x^k)$ and put

$$\theta_{k,i} = \frac{f_i(x^k)}{\|\gamma^{k,i}\|^2}. \tag{2.4}$$

[We will show later that $\gamma^{k,i}$ cannot be zero when $i \in I_k$, because $0 \notin \partial f_i(x^k)$ in this case.]

Define

$$x^{k+1} = x^k - \frac{1}{2}\sum_{i \in I_k} p_i \theta_{k,i}\gamma^{k,i} - \frac{1}{2}\theta_k \gamma^k, \tag{2.5}$$

let $k := k+1$ and go to Step $(k+1)$.

The behavior of the sequences generated by $ASMO$ is described by the following result:

Theorem 1. *If the conditions* (i) *and* (ii) *are satisfied, then for any initial point* $x^0 \in \mathbb{R}^n$, *the sequence* $\{x^k\}_{k\in\mathbb{N}}$ *generated by the averaged subgradient method is well defined and converges to a solution of problem* (1.1).

Proof: We start our proof with several observations. Let Ω be the disjoint union of I and S, provided with the σ-algebra $\mathcal{A}$ of the disjoint unions of sets which belong to 2^I and $\mathcal{B}_S$. Define the measure $\mu : \mathcal{A} \to [0,1]$ by

$$\mu(A) = \frac{1}{2}[\bar{\mu}(A \cap I) + \mu_\varphi(A \cap S)],$$

where $\bar{\mu} : 2^I \to [0,1]$ is given by $\bar{\mu}(\{a_i\}) = p_i$, for all $i \in I$ and $\mu_\varphi : \mathcal{B}_S \to [0,1]$ is the probability measure given at (2.1). Define the function $G : \Omega \times \mathbb{R}^n \to \mathbb{R}$ by

$$G(\omega, x) = \begin{cases} f(x) - f(\omega), & \text{if } \omega \in S \\ \\ f_\omega(x), & \text{if } \omega \in I. \end{cases} \tag{2.6}$$

Clearly, solving problem (1.1) is equivalent to finding solutions of the infinite system of inequalities

$$G(\omega, x) \le 0, \ \omega \in \Omega. \tag{2.7}$$

Solving (2.7) means finding a point in the intersection of the closed convex sets

$$Q_\omega := \{x \in \mathbb{R}^n : G(\omega, x) \le 0\}. \tag{2.8}$$

According to (i), there exists at least one solution z of (2.7) and therefore $z \in \cap_{\omega\in\Omega} Q_\omega$.

Now, we prove that whenever x^k is well defined, the numbers θ_k and $\theta_{k,i}$ exist. Suppose that x^k is well defined and observe that, if $i \in I_k$, then $\gamma^{k,i} \neq 0$ because otherwise, for any optimal solution $\bar{y}$ of problem (1.1), we get

$$f_i(\bar{y}) - f_i(x^k) \ge \langle \gamma^{k,i}, \bar{y} - x^k\rangle = 0,$$

showing that $f_i(x^k) \le f_i(\bar{y}) \le 0$, i.e., $i \notin I_k$, a contradiction.

Denote by $P_{k,\omega}(x^k)$ the metric projection of x^k onto the half space

$$H_k(\omega) = \{y \in \mathbb{R}^n : \langle \Gamma_k(\omega), y - x^k\rangle \le -G(\omega, x^k)\}. \tag{2.9}$$

where

$$\Gamma_k(\omega) = \begin{cases} \gamma^k, & \text{if } \omega \in S \\ \gamma^{k,i}, & \text{if } \omega \in I. \end{cases}$$

Note that $\Gamma_k(\omega) \in \partial G(\omega, \cdot)(x^k)$ for all $\omega \in \Omega$ and $k \in \mathbb{N}$. The remainder of our proof is based on the following:

Claim 1. The sequence $\{x^k\}_{k\in\mathbb{N}}$ is well defined, bounded and for any optimal solution z of (1.1) it has

$$\|z - x^{k+1}\| \leq \|z - x^k\|, \; k \in \mathbb{N}. \tag{2.10}$$

Moreover, the following limit exists and we have

$$\lim_{k\to\infty} \int_\Omega \|P_{k,\omega}(x^k) - x^k\|^2 d\mu(\omega) = 0. \tag{2.11}$$

In order to prove this claim, note that for any $\omega \in \Omega$ we have $z \in Q_\omega \subseteq H_k(\omega)$ and thus, $P_{k,\omega}(z) = z$, and this implies

$$\|z - P_{k,\omega}(x^k)\| = \|P_{k,\omega}(z) - P_{k,\omega}(x^k)\| \leq \|z - x^k\|. \tag{2.12}$$

Observe that, if x^k is well defined, then

$$\begin{aligned} \int_\Omega P_{k,\omega}(x^k) d\mu(\omega) &= x^k + \int_{I\cup S} \min\left\{0, -\frac{G(\omega, x^k)}{\|\Gamma_k(\omega)\|^2}\right\} \Gamma_k(\omega) d\mu(\omega) \\ &= x^k + \frac{1}{2}\sum_{\omega\in I} p_\omega \min\left\{0, -\frac{f_\omega(x^k)}{\|\Gamma_k(\omega)\|^2}\right\} \Gamma_k(\omega) \\ &\quad + \frac{1}{2}\int_S \min\left\{0, \frac{f(\omega) - f(x^k)}{\|\Gamma_k(\omega)\|^2}\right\} \Gamma_k(\omega) d\mu_\varphi(\omega) \\ &= x^k - \frac{1}{2}\sum_{i\in I_k} p_i \frac{f_i(x^k)}{\|\gamma^{k,i}\|^2}\gamma^{k,i} \qquad (2.13) \\ &\quad - \frac{1}{2}\int_S \max\left\{0, \frac{f(x^k) - f(\omega)}{\|\gamma^k\|^2}\right\} \gamma^k d\mu_\varphi(\omega) = x^{k+1}, \end{aligned}$$

where the last integral exists due to the integrability of the function $\omega \to P_{k,\omega}(x^k)$ (which results from [8, Corollary 3.2] and [10, Lemma 2.29(iii)]) and the last equality follows from (2.5). By induction, this shows that $\{x^k\}_{k\in\mathbb{N}}$ is well defined. Taking this into account, integrating (2.12) and applying Jensen's inequality, we get

$$\|z - x^k\| \geq \int_\Omega \|z - P_{k,\omega}(x^k)\| d\mu(\omega) \geq \|z - \int_\Omega P_{k,\omega}(x^k) d\mu(\omega)\| = \|z - x^{k+1}\| \tag{2.14}$$

which proves (2.10). It was repeatedly shown (see, e.g., [35, pp. 239-240]) that

$$\|z - P_{k,\omega}(x^k)\|^2 + \|P_{k,\omega}(x^k) - x^k\|^2 \leq \|z - x^k\|^2.$$

Integrating this expression and applying Jensen's inequality, we obtain

$$\int_\Omega \|P_{k,\omega}(x^k) - x^k\|^2 d\mu(\omega) \leq \|z - x^k\|^2 - \|z - x^{k+1}\|^2,$$

which implies (2.11) because, according to (2.10), the sequence $\{\|z - x^k\|\}_{k\in\mathbb{N}}$ is convergent, as being nonincreasing and bounded.

Now, we establish the following fact:

Claim 2. If x^* is an accumulation point of the sequence $\{x^k\}_{k\in\mathbb{N}}$, then x^* is a solution of (2.7).

To prove that, we first show that $G(\omega, x^*) \leq 0,$ for almost all $(\omega \in \Omega)$. Observe that the sets $\bigcup_{i\in I,\, k\in\mathbb{N}} \partial f_i(x^k)$ and $\cup_{k\in\mathbb{N}}\partial f(x^k)$ are bounded because f_i, $i \in I$, and f are bounded on bounded sets (see, for instance, Proposition 1.1.11 in [9]) as being continuous on $\mathbb{R}^n$ and because the sequence $\{x^k\}_{k\in\mathbb{N}}$ is bounded (cf. Claim 1). By consequence, there exists some positive number m^* such that $\|\Gamma_k(\omega)\| \leq m^*$ for any $k \in \mathbb{N}$ and $\omega \in \Omega$. Since $P_{k,\omega}(x^k)$ belongs to the hyperplane

$$\{y \in \mathbb{R}^n : \langle \Gamma_k(\omega), y - x^k\rangle + G(\omega, x^k) = 0\},$$

for each $\omega \in \Omega$ and for any $k \in \mathbb{N}$, we have

$$\begin{aligned} G(\omega, x^k) &\leq |\langle \Gamma_k(\omega), x^k - P_{k,\omega}(x^k)\rangle| \\ &\leq \|\Gamma_k(\omega)\| \cdot \|x^k - P_{k,\omega}(x^k)\| \\ &\leq m^*\|x^k - P_{k,\omega}(x^k)\|. \end{aligned} \tag{2.15}$$

Let $\{x^{i_k}\}_{k\in\mathbb{N}}$ be a subsequence of $\{x^k\}_{k\in\mathbb{N}}$ which converges to x^*. According to (2.11) we have

$$\lim_{k\to\infty} \int_\Omega \|P_{i_k,\omega}(x^{i_k}) - x^{i_k}\|^2 d\mu(\omega) = 0.$$

This combined with Theorem 38.7 in [23] and Proposition 24.18 in [27] implies that there exists a subsequence $\{x^{j_k}\}_{k\in\mathbb{N}}$ of the sequence $\{x^{i_k}\}_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} \|x^{j_k} - P_{j_k,\omega}(x^{j_k})\| = 0,\ \mu - \text{a.e.}$$

Therefore, using (2.15), we get

$$G(\omega, x^*) = \lim_{k\to\infty} G(\omega, x^{j_k}) \leq 0,\ \mu - \text{a.e..} \tag{2.16}$$

Now, suppose that there exists $\omega_0 \in \Omega$ such that $G(\omega_0, x^*) > 0$. If $\omega_0 \in I$, then

$$\mu(\{\omega \in \Omega : G(\omega, x^*) > 0\}) \geq \mu(\{\omega_0\}) = \frac{1}{2}\bar{\mu}(\{\omega_0\}) = \frac{1}{2}p_{i_0} > 0,$$

and this contradicts (2.16). If $\omega_0 \in S$, the continuity of $G(\cdot, x^*)$ over S implies that $G(\omega, x^*) > 0$ for any ω in some open ball U centered in ω_0. Then

$$\mu(\{\omega \in \Omega : G(\omega, x^*) > 0\}) \geq \frac{1}{2}\mu_\varphi(U \cap S) = \frac{1}{2\rho} \int_{U\cap S} \varphi(y)dy > 0,$$

where the last inequality holds because φ is positive on $\mathbb{R}^n$ and the Lebesgue measure of $int(U \cap S)$ is positive since $int(S) \neq \emptyset$, thus contradicting (2.16). This proves Claim 2.

For completing the proof of the theorem, suppose that x^* is one of the accumulation points of $\{x^k\}_{k\in\mathbb{N}}$. Then, according to Claim 2, x^* is a solution of problem (1.1). Therefore, Claim 1 implies that the sequence $\{\|x^* - x^k\|\}_{k\in\mathbb{N}}$ is convergent. If $\{x^{j_k}\}_{k\in\mathbb{N}}$ is a subsequence of $\{x^k\}_{k\in\mathbb{N}}$ converging to x^*, then we have

$$\lim_{k\to\infty} \|x^* - x^k\| = \lim_{k\to\infty} \|x^* - x^{j_k}\| = 0,$$

showing that the whole sequence $\{x^k\}_{k\in\mathbb{N}}$ converges to x^*. $\square$

3. Convergence analysis of the averaged subgradient method for computing Nash equilibria

The problem we consider in this section is that of finding a Nash equilibrium in a non-cooperative n-person game $\mathbb{G} = (W_1, ..., W_n;\ F_1, ..., F_n)$ where the (individual) strategy sets W_i are given by (1.2) and the payoff functions F_i satisfy the conditions (iv)-(vi). An idea going back to [26], and which was further developed in [6], is to reduce finding Nash equilibria of $\mathbb{G}$ to determining a solution of the system of inequalities

$$F(y, x) \leq 0,\ \ y \in W, \tag{3.1}$$

where $F : \prod_{i=1}^{n} \mathbb{R}^{r(i)} \times \prod_{i=1}^{n} \mathbb{R}^{r(i)} \to \mathbb{R}$ is defined by

$$F(y, x) = \sum_{i=1}^{n} \left[F_i(x^1, ..., x^{i-1}, y^i, x^{i+1}, ..., x^n) - F_i(x) \right]. \tag{3.2}$$

In fact, it is obvious that any solution of (3.1) with F given by (3.2) is a Nash equilibrium of $\mathbb{G}$ and conversely. The *Averaged Subgradient Method for Equilibria* $(ASME)$, presented below, is the following method for solving (3.1).

Step 0 : Let $m = \sum_{i=1}^{n} r(i)$ and $M = \cup_{i \in \{1,...,n\}} (\{i\} \times J_i)$. Choose a *weight vector* $p := (p_{ij})_{(i,j)\in M} \in \mathbb{R}^M$ such that $p_{ij} > 0$ for all $(i, j) \in M$ and $\sum_{(i,j)\in M} p_{ij} = 1$. Fix an *initial point* $z^0 = (z^{0,1}, ..., z^{0,n}) \in \prod_{i=1}^{n} \mathbb{R}^{r(i)} \simeq \mathbb{R}^m$, choose a continuous function $\varphi : \mathbb{R}^m \to (0, +\infty)$ such that $\varrho := \int_W \varphi(y)dy$ is positive and define the probability measure

$$\xi_\varphi(A) = \frac{1}{\varrho} \int_A \varphi(y)dy \tag{3.3}$$

whenever A belongs to the σ-algebra $\mathcal{B}_W$ of all Lebesgue measurable subsets of W. Put $k = 0$ and go to Step $(k + 1)$.

Step $(\mathbf{k+1})$: (I) Let $z^k = (z^{k,1}, ..., z^{k,n})$ be the iterate determined at Step k. Find a measurable selector[2] $y \to \Gamma_k(y) : W \to \mathbb{R}^m$ of the point-to-set mapping $y \to \partial F(y, \cdot)(z^k)$ and compute

$$v^k = \int_W \frac{1}{\|\Gamma_k(y)\|^2} \max\left\{0, F(y, z^k)\right\} \Gamma_k(y) d\xi_\varphi(y). \tag{3.4}$$

(II) Let

$$M_k := \left\{(i, j) \in M : g_{i,j}(z^{k,i}) > 0\right\}, \tag{3.5}$$

and, for each $(i, j) \in M_k$, determine a vector $\gamma_k^{i,j} \in \partial g_{i,j}(z^{k,i})$ and put

$$\alpha_{ij}^{(k)} = \frac{g_{i,j}(z^{k,i})}{\|\Gamma_k(i, j)\|^2}, \tag{3.6}$$

where $\Gamma_k(i, j) \in \mathbb{R}^m$ is the vector such that, for each integer $q \in \{1, 2, ..., m\}$, the q-th coordinate of $\Gamma_k(i, j)$ is exactly the coordinate $q - \sum_{l=0}^{i-1} r(l)$ of $\gamma_k^{i,j}$ (with

[2] Its existence is guaranteed by [10, Lemma 2.2(i)]. If all the functions F_i are differentiable, then, for each $y \in W$, we necessarily have $\Gamma_k(y) \equiv \sum_{i=1}^{n} \zeta^{k,i}(y)$, where $\zeta^{k,i}(y)$ is the gradient of the function $x \to F_i(x^1, ..., x^{i-1}, y^i, x^{i+1}, ..., x^n)$ at the point z^k.

$r(0) = 0$) when $1 + \sum_{l=0}^{i-1} r(l) \leq q \leq \sum_{l=0}^{i} r(l)$, and zero, otherwise. Define

$$z^{k+1} = z^k - \frac{1}{2} \sum_{(i,j)\in M_k} p_{ij}\alpha_{ij}^{(k)} \Gamma_k(i,j) - \frac{1}{2}v^k, \tag{3.7}$$

put $k := k+1$ and go to Step $(k+1)$.

The asymptotic behavior of *ASME* is described by the following result:

Theorem 2. *If conditions (iii)-(vi) are satisfied, then for any initial point $z^0 \in \mathbb{R}^m$, the sequence $\{z^k\}_{k\in\mathbb{N}}$ generated by the ASME is well defined and converges to a Nash equilibrium for the game $\mathbb{G}$.*

Proof: Let Ω be the disjoint union of the sets M and W, provided with the σ-algebra $\mathcal{A}$ of the disjoint unions of sets which belong to 2^M and $\mathcal{B}_W$. We endow the measurable space $(\Omega, \mathcal{A})$ with the probability measure $\xi : \mathcal{A} \to [0,1]$ defined by

$$\xi(A) = \frac{1}{2}[\bar{\xi}(A\cap M) + \xi_\varphi(A\cap W)],$$

where $\bar{\xi} : 2^M \to [0,1]$ is given by $\bar{\xi}(\{a_{ij}\}) = p_{ij}$, for all $(i,j) \in M$, and $\xi_\varphi : \mathcal{B}_W \to [0,1]$ is the probability measure given at (3.3). Define the function $G : \Omega \times \mathbb{R}^m \to \mathbb{R}$ by

$$G(\omega, z) = \begin{cases} F(\omega, z), & \text{if } \omega \in W, \\ g_{i,j}(z^{k,i}), & \text{if } \omega = (i,j) \in M. \end{cases} \tag{3.8}$$

The system (3.1) is equivalent to

$$G(\omega, z) \leq 0,\ \omega \in \Omega. \tag{3.9}$$

For each $\omega \in \Omega$, the function $G(\omega, \cdot)$ is convex and continuous because F_i and $g_{i,j}$ are continuous and conditions (iii)-(v) hold.

Suppose that z^k is well defined and let $\bar{z}$ be a Nash equilibrium for $\mathbb{G}$ (which exists due to (vi)). If we suppose that $\Gamma_k(y) = 0$ for some $y \in W$ for which $F(y, z^k) > 0$, then $F(y, z^k) \leq F(y, \bar{z}) \leq 0$ and this is a contradiction. Hence, $\Gamma_k(y)$ cannot be zero in (3.4). Analogously, the subgradients $\Gamma_k(i,j) \neq 0$ when $(i,j) \in M_k$. Consider the half-space

$$H_k(\omega) = \{y \in \mathbb{R}^m : \langle \Gamma_k(\omega), y - z^k\rangle \leq -G(\omega, z^k)\}.$$

Let $P_{k,\omega}(z^k)$ be the metric projection of z^k onto $H_k(\omega)$. An argument similar to that involved in the proof of (2.13) shows that the following integral exists and we have

$$\begin{aligned}
\int_\Omega P_{k,\omega}(z^k) d\xi(\omega) &= z^k + \int_{M\cup W} \min\left\{0, -\frac{G(\omega, z^k)}{\|\Gamma_k(\omega)\|^2}\right\} \Gamma_k(\omega) d\xi(\omega) \\
&= z^k + \frac{1}{2}\int_W \min\left\{0, -\frac{F(\omega, z^k)}{\|\Gamma_k(\omega)\|^2}\right\} \Gamma_k(\omega) d\xi_\varphi(\omega) \\
&\quad + \frac{1}{2}\int_M \min\left\{0, -\frac{G(\omega, z^k)}{\|\Gamma_k(\omega)\|^2}\right\} \Gamma_k(\omega) d\bar{\xi}(\omega)
\end{aligned}$$

$$= \quad z^k - \frac{1}{2}\int_W \max\left\{0, \frac{F(\omega, z^k)}{\|\Gamma_k(\omega)\|^2}\right\}\Gamma_k(\omega)d\xi_\varphi(\omega)$$
$$- \quad \frac{1}{2}\sum_{(i,j)\in M_k} p_{ij}\alpha_{ij}^{(k)}\Gamma_k(i,j) = z^{k+1}.$$

Therefore, z^{k+1} is well defined. By induction, the whole sequence $\{z^k\}_{k\in\mathbb{N}}$ is well defined. Analogously to (2.14) one deduces

$$(3.10)\quad \|\bar{z}-z^k\| \geq \int_\Omega \|\bar{z}-P_{k,\omega}(z^k)\|d\mu(\omega) \geq \|\bar{z}-\int_\Omega P_{k,\omega}(x^k)d\mu(\omega)\| = \|\bar{z}-z^{k+1}\|$$

for all $k \in \mathbb{N}$, showing that the sequence $\{\|\bar{z} - z^k\|\}_{k\in\mathbb{N}}$ converges nonincreasingly, whenever $\bar{z}$ is a Nash equilibrium in $\mathbb{G}$. The payoff functions F_i, $1 \leq i \leq n$ are continuous and, therefore, they have a common upper bound on the compact set W. According to (3.2), this implies that the functions $F(y,\cdot)$, $y \in W$, are uniformly bounded on $W \times W$. One can reproduce without essential modifications the argument involved in the "only if" part of the proof of Proposition 1.1.11 in [9] in order to show that whenever a family of functions $F(y,\cdot)$, $y \in W$, is uniformly bounded on bounded sets, the family of point-to-set mappings $\partial F(y,\cdot)$, $y \in W$, is uniformly bounded on bounded sets. Consequently, the sets $\cup_{(k,y)\in\mathbb{N}\times W}\partial F(y,\cdot)(z^k)$ and, similarly, $\cup_{k\in\mathbb{N},(i,j)\in M}\partial g_{i,j}(z^k)$ are bounded because $\{z^k\}_{k\in\mathbb{N}}$ is bounded. This implies that the set $\{\Gamma_k(\omega),\ k \in \mathbb{N},\ \omega \in \Omega\}$ is bounded too. Now, one can proceed exactly as in the proof of Claim 1 in Section 2 in order to show that any accumulation point z^* of the bounded sequence $\{z^k\}_{k\in\mathbb{N}}$ is a solution of the system (3.9). Combining this fact and the monotonicity of $\{\|z^* - z^k\|\}_{k\in\mathbb{N}}$ which results from (3.10), one obtains that $\{z^k\}_{k\in\mathbb{N}}$ converges to a solution of (3.9), that is, to a Nash equilibrium of $\mathbb{G}$. □

4. Computational experiments with *ASMO*

The aim of this section is to present and discuss the results of several numerical experiments we did with *ASMO* and to emphasize phenomena which, because of their consistent occurrence, seem to point towards behavioral characteristics of this algorithm. *ASMO* was run on test problems extracted from bibliographic sources as well as on problems we have formulated in order to test the algorithm in special situations. The experiments selected for presentation below were chosen because they allow graphical visualization of patterns of behavior we repeatedly observed in the applications of *ASMO* to problems of equal or larger dimension. In these experiments, the integrals required by *ASMO* for determining the coefficient θ_k were approximated by Riemann sums. These sums are reasonably good estimates for double and triple integrals and allow good control of the error size thus making the observation of the behavior of ASMO in various situations possible. The runs were done on a personal computer with a Pentium III processor at 700 Mhz and we have performed 500 algorithmic steps in each case. The number on the line marked k in Figures 1-5 below are step counters. Meaningful applications of *ASMO* and/or *ASME* in dimensions larger than ten seem to be possible only by use of parallel computing because of the highly demanding calculations involved in the integration procedures contained in those algorithms.

Problem 1 is taken from [1, p. 70]:

$$\text{(P1)} \qquad \begin{cases} \min f(x_1, x_2) = x_1^2 - 2x_1 + x_2 \\ \textit{such that} \\ 3 - x_1 - x_2 \leq 0 \\ x_1^2 - x_2 + 1 \leq 0 \\ x_2 - 4 \leq 0 \\ -x_1 \leq 0 \\ -x_2 \leq 0 \end{cases}$$

Results of $ASMO$ applied to this problem are shown in Figure 1 and Figure 2.

Problem 2 was extracted from [16, p. 45]:

$$\text{(P2)} \qquad \begin{cases} \min f(x_1, x_2) = (x_1 - 2)^2 + (x_2 - 1)^2 \\ \textit{such that} \\ x_1 + x_2 - 2 \leq 0 \\ x_1^2 - x_2 \leq 0 \end{cases}$$

Figure 3 presents the result of an experiment with this problem. Figure 4 summarizes experimental results with *Problem 3* which is taken from [16, p. 58]:

$$\text{(P3)} \qquad \begin{cases} \min f(x_1, x_2, x_3) = 9 - 8x_1 - 6x_2 - 4x_3 + 2x_1^2 + \\ \qquad\qquad 2x_2^2 + x_3^2 + 2x_1 x_2 + 2x_1 x_3 \\ \textit{such that} \\ x_1 + x_2 + 2x_3 - 3 \leq 0 \\ -x_1 \leq 0 \\ -x_2 \leq 0 \\ -x_3 \leq 0 \end{cases}$$

Problems 1, 2 and 3 involve only differentiable functions in the objective as well as in the restrictions. The following *Problem 4* has nondifferentiable objective function:

$$\text{(P4)} \qquad \begin{cases} \min f(x_1, x_2) = \max\{x_1^2 + x_2^2,\ 2\} \\ \textit{such that} \\ (x_1 - 2)^2 + x_2^2 - 4 \leq 0 \\ -x_1 \leq 0 \\ -x_2 \leq 0 \end{cases}$$

Also, by contrast to (P1)-(P3) which have unique optimal solutions (their objective functions are strictly convex), problem (P4) has infinitely many optimal solutions; its set of optimal solutions is $\{(x_1, x_2) \in [0, \sqrt{2}] \times [0, \sqrt{2}] : (x_1 - 2)^2 + x_2^2 - 4 \leq 0\}$. In all cases considered here, computations of (sub)gradients can be easily done at any point. For problems in which more complicated functions occur, one can take advantage of the procedure for determining subgradients presented in the Appendix below.

Analyzing the results of these and the other numerical experiments we made with $ASMO$, we would like to emphasize several aspects of its behavior:

(i) Theorem 1 shows that $ASMO$ generated sequences $\{x^k\}_{k\in\mathbb{N}}$ converge and the corresponding sequence of distances $\{\|x^k - x^*\|\}_{k\in\mathbb{N}}$ to their limits x^* are nonincreasing. Figures 1(b), 2(b), 3(b) and 4(a) illustrate how these distances decrease along the computational process in the cases of problems (P1)-(P3). These as well as all other experiments we have done seem to show that the distances $\{\|x^k - x^*\|\}_{k\in\mathbb{N}}$

decrease geometrically, i.e., it seems that one can find some number $q \in (0,1)$ (which may, eventually, be close to 1) depending on the initial point x^0 and on the problem, such that

$$\limsup_{k\to\infty} \frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|} \leq q, \;\; k \in \mathbb{N}.$$

We do not have a mathematical proof of that.

(ii) There is no a priori condition for the choice of the initial point x^0 in $ASMO$. However, the choice of the initial point considerably affects the behavior of the algorithm. This can be clearly seen by comparing Figure 1 and Figure 2. These figures report results of running the algorithm twice on problem (P1) with two different initial points. Note that, in spite of choosing initial points having almost the same distance to the optimal solution, the choice of the initial point in Figure 1 leads to a faster convergence. This seems to be due to the relative position of the initial point to the set of optima and leads to the following question whose answer we do not know: Can one find rules of choosing initial points which will guarantee faster convergence of the $ASMO$ for some classes of problems?

(iii) The behavior of the sequence $\{f(x^k)\}_{k\in\mathbb{N}}$ in $ASMO$ seems atypical for subgradient methods (compare to [20, p. 18]). In $ASMO$, $\{f(x^k)\}_{k\in\mathbb{N}}$ is not necessarily monotonically decreasing, as clearly show Figure 2(d) and Figure 3(d).

(iv) The computational errors accumulated along the iterative process affect the behavior of $ASMO$ mostly when x^k is close to the optimum. Such a phenomenon can be observed in Figure 1(c), where, after 400 steps and being in the range of 10^{-12} from the solution, the algorithm behaves chaotically. This phenomenon is common to all subgradient type methods we know about.

The accuracy of $ASMO$ depends on the accuracy of the multiple integration procedures used in its implementation. For problems in dimension larger than three, where Riemann sums stop providing reasonably good approximations of multiple integrals, more efficient integration procedures should be used (e.g., the Monte Carlo type methods described in [11, 12, 19]). The improved efficiency of these methods is mostly due to the reduction of the number of points at which the integrand has to be evaluated in the integration process.

According to [12], existing computational experience with numerical procedures for computation of multiple integrals in use today would work efficiently when the number of variables in the integrand is in the range of ten. Fortunately, many optimization problems occurring in physics and engineering (see, for instance, [4] and [7]) are involving up to four variables and, at least for such problems, $ASMO$ seems to be well fitted.

(v) A question for which we are indebted to Prof. R. Schultz is whether we can replace the actual integral based coefficients in the algorithm by some estimates of them which are easier to compute, but still preserve the convergence of the procedure. This question naturally leads to another one: How stable is the $ASMO$ with respect to the errors occurring in the process of determining the coefficients θ_k and $\theta_{k,i}$? We do not have mathematical answers to these questions yet. What we can say in this sense comes from computational experiments we have performed. $ASMO$ seems to be quite robust under errors in determining the coefficients θ_k and $\theta_{k,i}$. Figure 6 shows how the distance of the iterate x^{500} to x^* varies when the integration needed in order to compute θ_k in $ASMO$ is done with different errors. As mentioned above, we approximated the integrals by Riemann sums and, in this case, the maximal integration error is proportional with the length s of the grid division. Figure 6 presents the results of five hundred runs of the $ASMO$ applied to (P2), at each stage increasing s by 10^{-3} from $s = 10^{-3}$ to $s = 10^{-1}$. The behavior of $\|x^* - x^{500}\|$ displayed in the diagram seems to confirm repeated experimental observations that $ASMO$ is quite stable under integration errors: The precision with which an iterate x^k approximates a solution stays between reasonable limits for as long as the integration error is kept between reasonable limits too.

5. Appendix: A numerical method for computing subgradients

Let $h : \mathbb{R}^n \to \mathbb{R}$ be a convex function and fix $x \in \mathbb{R}^n$. Consider the following procedure: Choose $\gamma^0 \in \mathbb{R}^n$ and let

$$\gamma^{k+1} = \gamma^k + \int_0^1 \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} \chi_k(x, \eta(t)) |J_n(t)| dt,$$

where for $\eta \in \mathbb{R}^n$,

$$\chi_k(x, \eta) = \min \left\{ \frac{h(x+\eta) - h(x) - \langle \gamma^k, \eta \rangle}{\|\eta\|^2} \eta, \, 0 \right\}$$

and, when $t = (t_1, ..., t_n)$, then $\eta(t)$ is given by

$$\begin{array}{lrcl} (5.1) & \eta_1(t) & = & t_1 \sin t_2 \sin t_3 \cdots \sin t_{n-1} \sin t_n, \\ & \eta_2(t) & = & t_1 \sin t_2 \sin t_3 \cdots \sin t_{n-1} \cos t_n, \\ & \eta_3(t) & = & t_1 \sin t_2 \cdots \sin t_{n-2} \cos t_{n-1}, \\ & & \vdots & \\ & \eta_{n-1}(t) & = & t_1 \sin t_2 \cos t_3, \\ & \eta_n(t) & = & t_1 \cos t_2, \end{array}$$

while

$$J_n(t) = (-1)^{n(n-1)/2} t_1^{n-1} \sin^{n-2} t_2 \cdots \sin t_{n-1}.$$

The following result shows that the procedure presented here produces vectors γ^k which converge to a subgradient of h.

Proposition. *If $h : \mathbb{R}^n \to \mathbb{R}$ is convex, then no matter how γ^0 is chosen in $\mathbb{R}^n$, $\{\gamma^k\}_{k \in \mathbb{N}}$ converges and its limit is a subgradient of h at x.*

Proof: Consider the infinite system of inequalities

$$(5.2) \qquad \langle \gamma, y \rangle \le h(x+y) - h(x), \;\; y \in B(0,1).$$

Any solution γ of this system is a subgradient of h because if $z \in \mathbb{R}^n$, then $tz \in B(0,1)$ for $t > 0$ sufficiently small and, by (5.2), we obtain

$$\langle \gamma, z \rangle \le \lim_{t \searrow 0} \frac{h(x+tz) - h(x)}{t} \le h(x+z) - h(x).$$

Consequently, using Theorem 2.1 in [10], we deduce that the sequence

$$\gamma^{k+1} = \gamma^k + \int_{B(0,1)} \chi_k(x, \omega) d\omega$$

converges to a subgradient of h at x. Changing the variables in the last integral according to (5.1), we get the conclusion. □

Acknowledgements

We wish to thank Professors M. Fukushima, R. Schultz and R. Tichatschke for interesting comments concerning the topics discussed in this paper. Also, we are grateful to the referees for comments which helped to improve an earlier version of this paper.

References

[1] N. Andrei (1999). *Advanced Mathematical Programming: Theory, Computational Methods, Applications* (Romanian), Editura Tehnica, Bucuresti.

[2] A.S. Antipin (2000). From optima to equilibria, in *Dynamics of Non-homogeneous Systems*, (Yu.S. Popkov, ed.), Vol. 3, Russian Academy of Sciences, Editorial URSS, Moscow, 35-64.

[3] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty (1993). *Nonlinear Programming: Theory and Algorithms*, 2-nd edition, John Wiley & Sons, New York.

[4] A.D. Belegundu and T.R. Chandrupatla (1999). *Optimization: Concepts and Applications in Engineering*, Prentice Hall, New Jersey.

[5] D.P. Bertsekas (1999). *Nonlinear Programming*, 2-nd edition, Athena Scientific, Belmont, Mass.

[6] E. Blum and W. Oettli (1994). From optimization and variational inequalities to equilibrium problems, *The Mathematics Student*, **63**, 123-145.

[7] P. Brousse (1988). *Optimization in Mechanics: Problems and Methods*, North Holland, Amsterdam.

[8] D. Butnariu (1995). The expected-projection methods: Its behavior and applications to linear operator equations and convex optimization, *Journal of Applied Analysis*, **1**, 95-108.

[9] D. Butnariu and A.N. Iusem (2000). *Totally Convex Functions for Fixed Points Computation and Infinite Dimensional Optimization*, Kluwer Academic Publishers, Dordrecht.

[10] D. Butnariu and E. Resmerita (2001). The outer Bregman projection method for stochastic feasibility problems in Banach spaces, 69-86, in *Inherently Parallel Algorithms for Feasibility and Optimization and their Applications*, (D. Butnariu, Y. Censor and S. Reich, eds.), Elsevier, Amsterdam.

[11] P.J. Davis and P. Rabinowitz (1984). *Methods of Numerical Integration*, 2-nd edition, Academic Press, Orlando.

[12] M. Evans and T. B. Swartz (2000). *Approximating Integrals via Monte Carlo and Deterministic Methods*, Clarendon Press, Oxford.

[13] S.D. Flåm (2001). Approaching equilibria in parallel, 267-278, in *Inherently Parallel Algorithms for Feasibility and Optimization and their Applications*, (D. Butnariu, Y. Censor and S. Reich, eds.), Elsevier, Amsterdam.

[14] S.D. Flåm and J. Zowe (1990). Relaxed outer projections, weighted averages and convex feasibility, *BIT*, **30**, 289-300.

[15] A.A. Goldstein (1964). Convex programming in Hilbert space, *Bulletin of the American Mathematical Society*, **70**(5), 709-710.

[16] W. Hock and K. Schittkowski (1981). *Test Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems, vol. 187, Springer Verlag, New York.

[17] K.C. Kiwiel (1985). *Methods of Descent for Nondifferentiable Optimization*, Springer, Berlin.

[18] K.C. Kiwiel and P.O. Lindberg (2001). Parallel subgradient methods for convex optimization, 335-344, in *Inherently Parallel Algorithms for Feasibility and Optimization and their Applications*, (D. Butnariu, Y. Censor and S. Reich, eds.), Elsevier, Amsterdam.

[19] A.R. Krommer and C.W. Ueberhuber (1988). *Computational Integration*, SIAM, Philadelphia.

[20] E.S. Levitin and B.T. Polyak (1966). Constrained minimization methods, *USSR Computational Mathematics and Mathematical Physics*, **6**(5), 1-50.

[21] E.S. Levitin and B.T. Polyak (1966). Convergence of minimizing sequences in conditional extremum problems, *Soviet Math. Doklady*, **7**, 764-767.

[22] Yu.I. Lyubich and G.D. Maistrovskii (1970). General theory of relaxation processes for convex functions, *Uspehi Mat. Nauk,* **1**, 57-112.

[23] M.E. Munroe (1970). *Measure and Integration*, Addison-Wesley Publishing Company, Reading, Mass.

[24] J. Nash (1950). Equilibrium points in bimatrix games, *Proceedings of the National Academy of Sciences USA*, **36**, 48-49.

[25] A. Nedić, D.P. Bertsekas and V.S. Borkar (2001). Distributed asynchronous incremental subgradient methods, 381-407, in *Inherently Parallel Algorithms for Feasibility and Optimization and their Applications*, (D. Butnariu, Y. Censor and S. Reich, eds.), Elsevier, Amsterdam.

[26] H. Nikaido and K. Isoda (1955). Note on noncooperative convex games, *Pacific Journal of Mathematics*, **5**, 807-815.

[27] K.R. Parthasarathy (1967). *Probability Measures on Metric Spaces*, Academic Press, New York.

[28] T. Parthasarathy and T.E.S. Raghavan (1971). *Some Topics in Two Person Games*, American Elsevier Publishing Co., New York.

[29] B.T. Polyak (1963). Gradient methods for minimizing functionals (Russian), *Zh. Vychisl. Mat. i Mat. Fiz.*, **3**, 643-653.

[30] B.T. Polyak (1987). *Introduction to Optimization*, Optimization Software, New York.

[31] B.T. Polyak (1967). A general method of solving extremum problems, *Soviet Math. Doklady*, **8**(3), 593-597.

[32] J.B. Rosen (1960). The gradient projection method for nonlinear programming I, *SIAM Journal of Applied Mathematics*, **8**(2), 181-217.

[33] A. Stefanescu (2000). *Competitive Models in Game Theory and Economic Analysis*, Editura Universitatii din Bucuresti, Bucuresti.

[34] M. Todd (1976). *The Computation of Fixed Points and Applications*, Springer-Verlag, Berlin.

[35] E.H. Zarantonello (1971), Projections on convex sets in Hilbert space and spectral theory, in *Contributions to Nonlinear Functional Analysis* (E.H. Zarantonello, ed.), Academic Press, New York, 239-343.

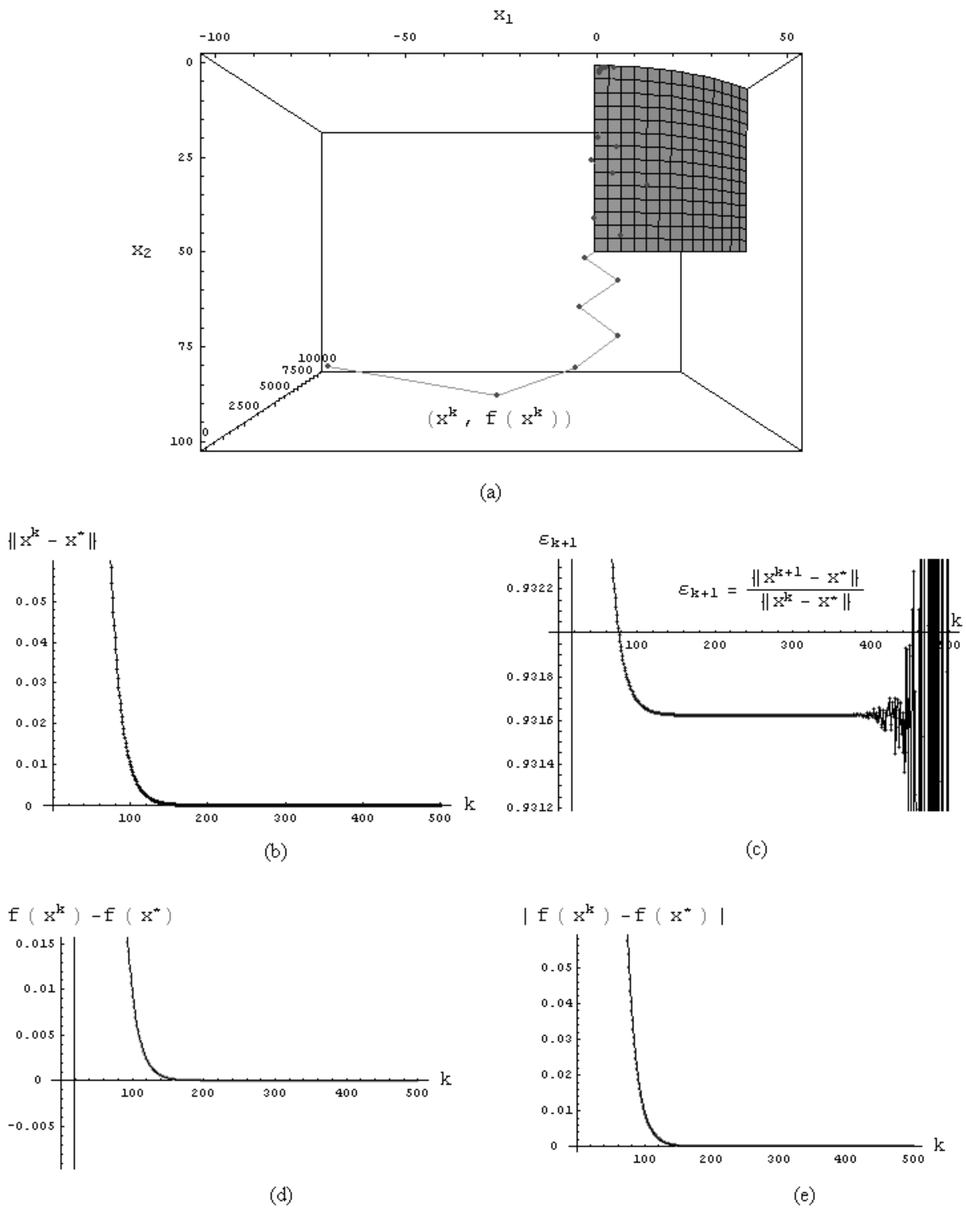

FIGURE 1. Problem 1 with initial point $(-100, 100)$ and $x^* = (1, 2)$

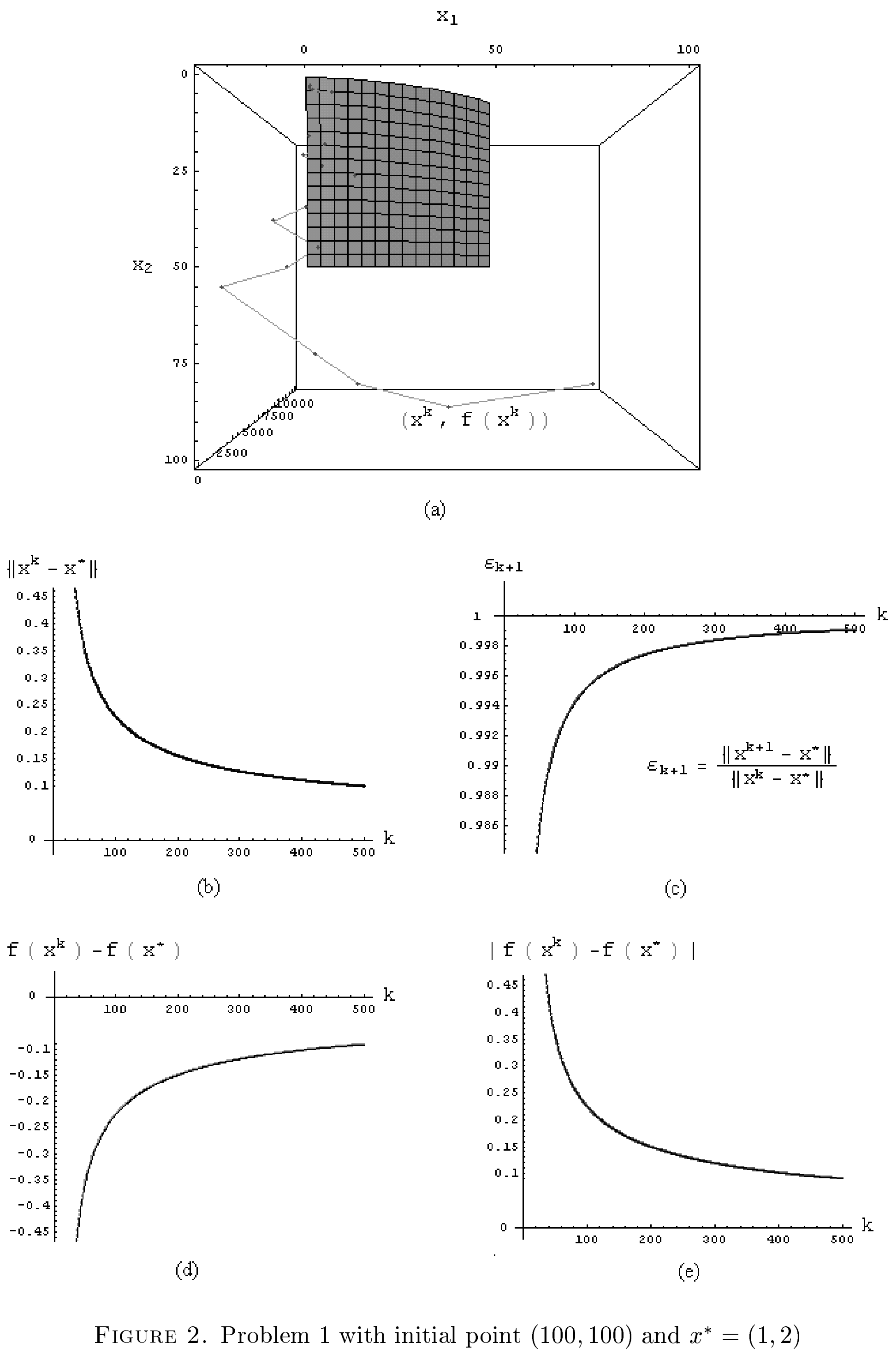

FIGURE 2. Problem 1 with initial point $(100, 100)$ and $x^* = (1, 2)$

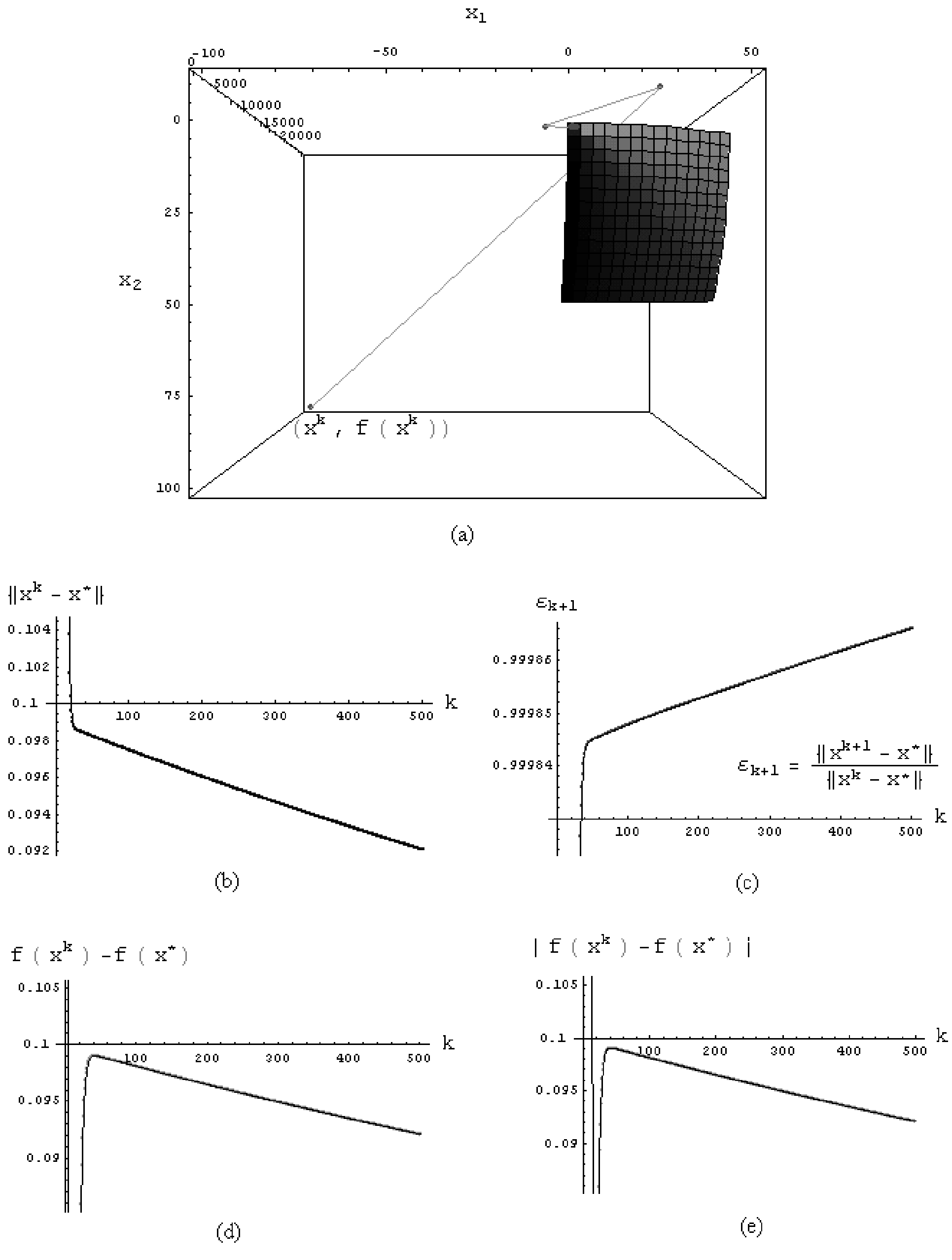

FIGURE 3. Problem 2 with initial point $(-100, 100)$ and $x^* = (1,1)$

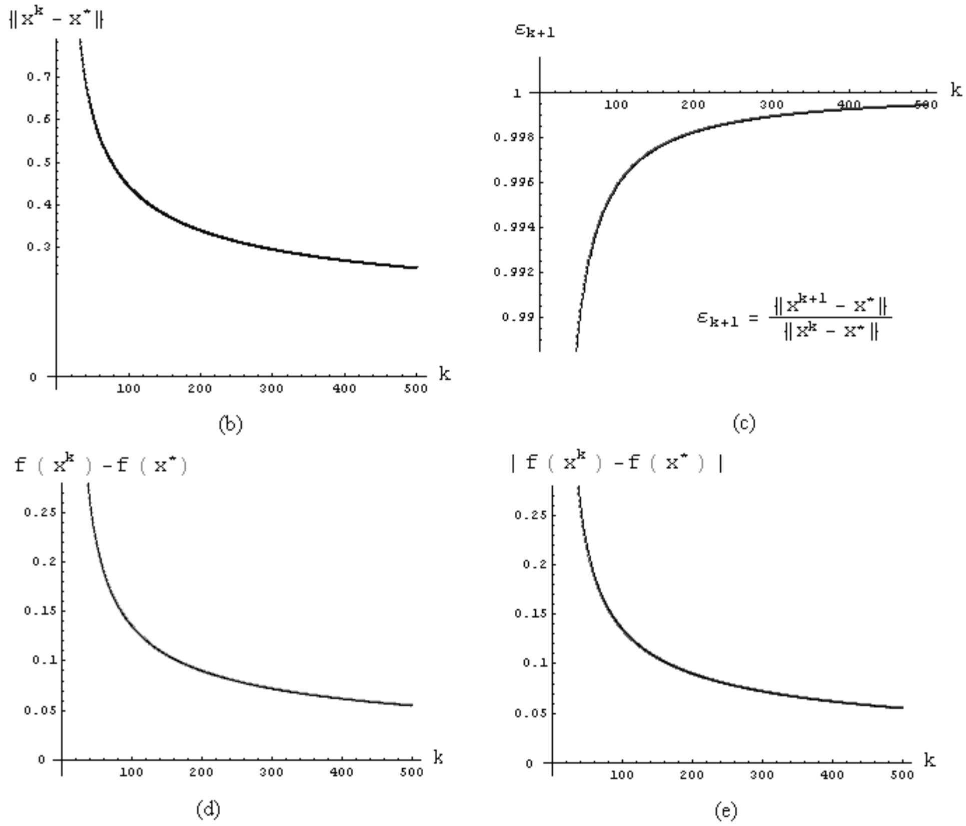

FIGURE 4. Problem 3 with initial point $(-100, 100, 100)$ and $x^* = (4/3, 7/9, 4/9)$

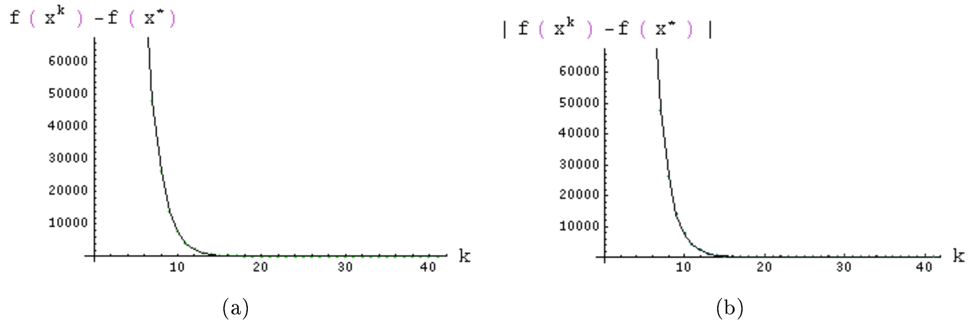

FIGURE 5. Problem 4 with initial point $(1000, -1000)$

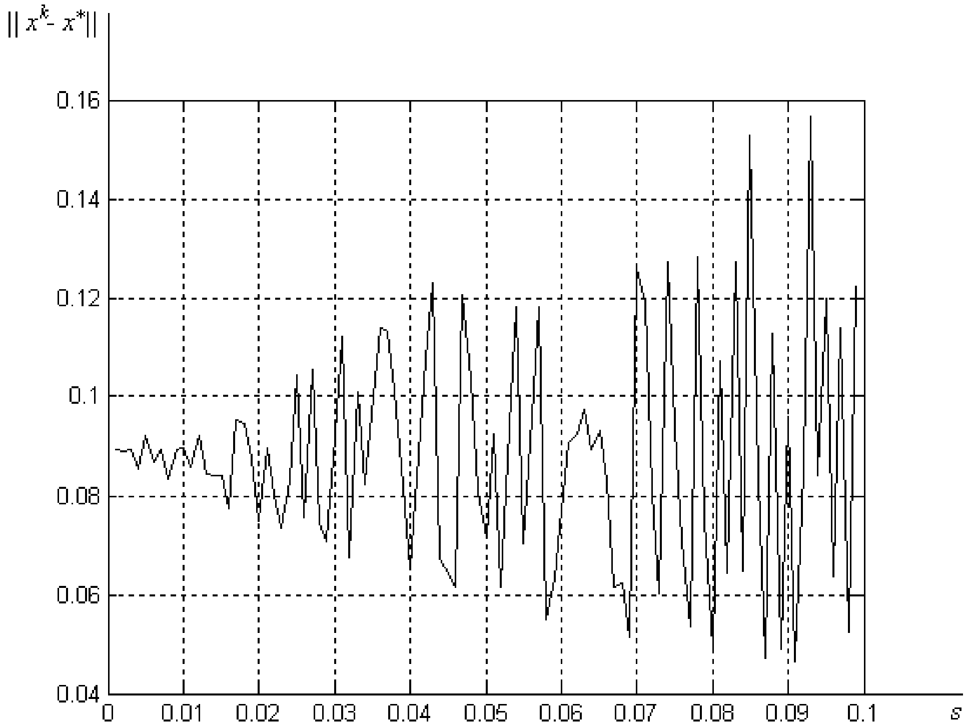

FIGURE 6. The impact of the integration errors upon $||x^k - x^*||$