15th Triennial World Congress, Barcelona, Spain

LEARNING AUTOMATA-BASED OPTIMIZATION IN A BINARY CODED SEARCH SPACE

K. Najim [*] **A.S. Poznyak** [**] **E. Ikonen** [***,1]

[*] *E.N.S.I.A.C.E.T., Toulouse, France*
[**] *CINVESTAV-IPN., Mexico D.F., Mexico*
[***] *University of Oulu, Finland*

Abstract: This paper presents an algorithm for optimization. This algorithm is based on a team of learning stochastic automata. Each automaton is characterized by two actions providing a binary output (0 or 1). The action of the team of automata consists of a digital number which represents the environment input. The probability distribution associated which each automaton is adjusted using a modified version of the Bush-Mosteller reinforcement scheme. This adaptation scheme uses a continuous environment response and a time-varying correction factor. A normalization procedure is used in order to preserve the probability measure. The asymptotic properties of this optimization algorithm are presented. A numerical example illustrates the feasibility and the performance of this optimization algorithm.

Keywords: asymptotic properties; discretization; genetic algorithms; learning algorithms; random searches

1. INTRODUCTION

Frequently, the information necessary for solving a problem (control, optimization, etc.) is not available or may be incomplete. It is then necessary to learn (acquire) additional information. Learning deals with the ability of systems to improve their response (performance in the sense of some criterion) based on past experience (Tsypkin, 1973). Learning models stem from diverse approaches, frequently grounded on heuristic intuitions and experiments. Learning automata are information processing systems whose architecture and behavior are inspired by the structure of biological systems (the organism is born with relatively little initial knowledge and learns actions that are appropriate through trial and error).

A learning automaton operates in a random environment (process to be controlled, function to be optimized, etc.) and adapts his probability in order to achieve the desired control (optimization) objective (learning goal). The main theoretical as well as practical results related to learning automata (Najim and Poznyak, 1994) (Poznyak and Najim, 1997) (Poznyak *et al.*, 2000) have been carried out in the last decade. The convergence and the estimation of the convergence rate of both binary and continuous reinforcement schemes have been carried on the basis of martingale theory and Lyapunov approach. The behavior of hierarchical structure of learning automata and learning automata with changing number of actions has also been analyzed and several theoretical results have been stated.

Learning automata have been used to solve engineering problems as well as problems stemmed from economy which are characterized by nonlinearity and a high level of uncertainty (Najim and Oppenheim, 1991). They have been used for process modelling and control, optimization, pattern recognition, image processing, signal processing, trajectory planning of robot manipulators, tele-

[1] Supported by the Academy of Finland, Project no 48545

phone and internet traffic routing, process navigation, neuro-fuzzy networks training, process synthesis, etc.

Binary coding is common with genetic algorithms. Howell (Howell, 2000) proposed a genetic learning automata (GLA) optimization algorithm, applied to probabilities of binary actions of a team of learning automata. The outputs of the team of automata form a binary number which constitutes the environment input. The continuous environment response is obtained from a realization of the function to be minimized, and a normalization procedure is then used to ensure the preservation of the probability measure. Each automaton in the team is provided with the same normalized environment response. In the GLA algorithm (Howell, 2000), the population consisted of strings of binary-action learning automata probabilities. At each generation, a set of sample vectors was generated using the probability distribution in the population. The sample vectors were evaluated, and a max-min normalization conducted within the current population. The probabilities in the strings of the population were then updated using a reinforcement scheme. He then further applied crossover and reordering operators, before proceeding to next generation.

In this paper, an optimization algorithm is developed based on a team of learning automata with two actions (one action is equal to 0 and the second one is equal to 1). Note that it can –loosely– be seen as a special case of Howell's approach (population of one probability string only, no genetic operators). It uses the Bush-Mosteller reinforcement scheme with a continuous input (continuous environment response) and a time-varying correction factor. In this paper, theoretical results concerning the asymptotic properties of the system are presented, as well as computer simulations illustrating the performance of the approach.

The remainder of this paper is organized as follows: The next section deals with the definition of a learning automaton and the presentation of the reinforcement scheme (adaptation mechanism) used in the adaptation procedure. The optimization problem is stated in section 3. Section 4 presents the asymptotic properties. A numerical example is presented in section 5. Some conclusions end this paper.

2. STOCHASTIC LEARNING AUTOMATA

A k-automaton $\left(k=\overline{1,N}\right)$ with binary output, belonging to a team with N participants, operating in a random environment (medium), is an adaptive discrete machine described by

$$\left\{\Xi, U^k, \left\{\xi_n^k\right\}, \left\{u_n^k\right\}, \left\{p_n^k\right\}, T^k\right\}$$

where:

(1) Ξ is the automaton input bounded set;

(2) U^k denotes the set $\left\{u^k(1)=1, u^k(2)=0\right\}$ of actions of the automata $\left(k=\overline{1,N}\right)$ (we consider a team of N automata), and $\left\{u_n^k\right\}$ is a sequence of binary automaton outputs (actions): $u_n^k=\{0;1\}$;

(3) $\left\{\xi_n^k\right\}$ is a sequence of automaton inputs (payoffs $\xi_n^k \in \Xi$) provided by the given mechanism in a binary (P-model environment) form;

(4) $p_n^k=\left[p_n^k(1), p_n^k(2)\right]^\mathsf{T}$ is the conditional probability distribution at time n :

$$p_n^k(i)=P\left\{\omega\in\Omega : u_n^k=u^k(i) \;/\; \mathsf{F}_{n-1}\right\}$$
$$\sum_{i=1}^{2} p_n^k(i)=1$$

where $\mathsf{F}_n=\sigma(u_1^k, p_1^k, \xi_1^k; ...; u_n^k, p_n^k, \xi_n^k)$ is the minimal σ-algebra generated by the corresponding events ($\mathsf{F}_n \subseteq \mathsf{F}$).

(5) $T^k=T_n^k$ represents the reinforcement scheme (updating scheme) which changes the probability vector p_n^k to p_{n+1}^k, that is,

$$p_{n+1}^k = p_n^k + \gamma_n^k T_n^k(p_n^k; \left\{\xi_t^k\right\}_{t=1,...,n}; \left\{u_t^k\right\}_{t=1,...,n}) \quad (1)$$

$p_1^k(i)>0$, $i=1,2$, where γ_n^k is a scalar correction factor and the vector $T_n^k(.)=\left[T_n^{k,1}(.)T_n^{k,2}(.)\right]^\mathsf{T}$ satisfies the following conditions (for preserving probability measure):

$$\sum_{i=1}^{N} T_n^{k,i}(\cdot)=0$$
$$p_n^k(i)+\gamma_n T_n^{k,i}(\cdot)\in[0,1]$$
$$\forall n,\; k=1,...,N$$

The environment establishes the relation between the actions of the automaton and the signals received at its input. It includes all external influences. The environment produces a random response whose statistics depend on the current stimulus or input.

3. OPTIMIZATION ALGORITHM

Several engineering problems require a multimodal functions optimization strategy. Usually, the function $f(x)$ to be optimized is not explicitly known: only samples of the disturbed values of $f(x)$ at various settings of x can be observed, complicating the application of the usual numerical optimization procedures.

Let us consider a real-valued scalar function $f(x)$, $x \in [x_{\min}, x_{\max}]$. We would like to find the value $x = x^*$ which minimizes this function, i.e.,

$$x^* = \arg \min_{x \in X = [x_{\min}, x_{\max}]} f(x) \tag{2}$$

There are almost no conditions concerning the function $f(x)$ (continuity, unimodality, differentiability, convexity, etc.) to be optimized. We are concerned with an ε-global optimization problem of multimodal and nondifferentiable functions.

The actions of the team of N stochastic automata form a binary string of length N:

$$u_n^1, u_n^2, ..., u_n^N$$

where $u_n^k = \{0; 1\}$. The quantized real value is given by

$$x_n = x(u_n) := x_{\min} + A \sum_{i=1}^{N} u_n^i 2^{i-1} \tag{3}$$

where $x_n \in X$, $X = \{x_{\min}, x_{\min}+A, x_{\min}+2A, ..., x_{\max}\}$ and $u_n := \left(u_n^1, u_n^2, ...u_n^N\right)^{\intercal}$. The resolution of the quantization is equal to

$$A = \frac{x_{\max} - x_{\min}}{2^N - 1} \tag{4}$$

Without loss of generality, we can assume that $x_{\min} = 0$. Let y_n be the observation of the function $f(x)$ at the point $x_n \in X$, i.e.,

$$y_n = f(x_n) + w_n \tag{5}$$

where w_n is the observation noise (disturbance) at time n. We assume that the observation noise is a conditionally zero mean random variable with finite variance, i.e.,

- (**H1**) The conditional mathematical expectations of the observation noise w_n are equal to zero for any time $n = 1, 2, ...$: $\mathbf{E}\{w_n / \mathcal{F}_{n-1}\} \overset{a.s.}{=} 0$, $\mathcal{F}_{n-1} := \sigma\left(x_s, w_s; s = \overline{1, n-1}\right)$, i.e. $\{w_n\}$ is a sequence containing martingale-differences.
- (**H2**) The conditional variances of the observation noises exist and are uniformly bounded: $\mathbf{E}\left\{w_n^2 / \mathcal{F}_{n-1}\right\} \overset{a.s.}{=} \sigma_n^2(i)$, $\max_i \sup_n \sigma_n^2(i) := \sigma^2 < \infty$.

The optimization algorithm operates as follows, see Fig. 1. At each time n, each automaton of the team selects randomly an action u_n^k $\left(k = \overline{1, N}\right)$. According to (3) and (4), these actions are in turn used to calculate the new value of the argument x_n, and then, the realization y_n of the function $f(x_n)$ is obtained. This realization is then normalized as follows:

$$\widehat{\xi}_n := \tag{6}$$

$$\left[s_n(i_1, ..., i_N) - \min_{j_1, \cdots, j_N} s_{n-1}(j_1, ..., j_N)\right]_+ \Bigg/$$

$$\max_{j_1, \cdots, j_N} \left[s_n(i_1, ..., i_N) - \min_{j_1, \cdots, j_N} s_{n-1}(j_1, ..., j_N)\right]_+ + 1$$

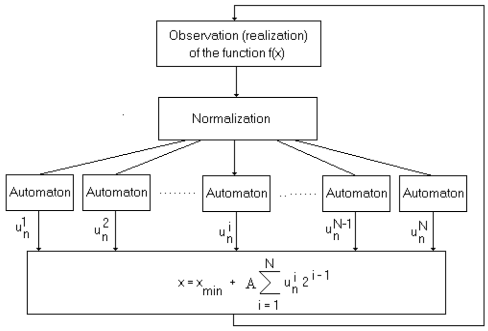

Fig. 1. Schematic diagram of the optimization algorithm.

where

$$s_n(i_1, ..., i_N) := \frac{\sum\limits_{t=1}^{n} y_t \prod\limits_{k=1}^{N} \chi(u_t^k = u^k(i_k))}{\sum\limits_{t=1}^{n} \prod\limits_{k=1}^{N} \chi(u_t^k = u^k(i_k))} \tag{7}$$

with

$$[x]_+ \; : = \begin{cases} x \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases}, \tag{8}$$

$$\chi(u_n^k = u^k(i)) := \begin{cases} 1 \text{ if } u_n^k = u^k(i) \\ 0 \text{ if } u_n^k \neq u^k(i) \end{cases}$$

The normalized environment response belongs to the unit segment, $\widehat{\xi}_n \in [0, 1]$. It is then used as the input of all the automata belonging to the team of automata, that is,

$$\widehat{\xi}_n^k = \widehat{\xi}_n, \; k = \overline{1, N}$$

($\widehat{\xi}_n^k$ represents the input of the k^{th} automaton). We are dealing with a random stationary environment where responses $\hat{\xi}_n^k$ are characterized (in view of (**H1**) and (**H2**)) by the following two properties.

Lemma 1. *Assume that assumptions (**H1**) and (**H2**) hold and suppose that the considered reinforcement scheme (1) generates the sequences $\{p_n^k\}$ such that for any index collection $(i_1, ..., i_N)$ the following "ergodic condition" fulfilled:*

$$\sum_{n=1}^{\infty} \prod_{k=1}^{N} p_n^k(i_k) \overset{a.s.}{=} \infty \tag{9}$$

Then, the normalized environment response $\hat{\xi}_n$ (6) possesses the following properties:

- *The number of selections of each action collection $(i_1, ..., i_N)$ is infinite, i.e.,*

$$\sum_{t=1}^{\infty} \prod_{k=1}^{N} \chi(u_t^k = u^k(i_k)) \overset{a.s.}{=} \infty \tag{10}$$

- *The random variable $s_n(i_1,...,i_N)$ (7) is asymptotically equal to the value of the function to be optimized for the corresponding point $x(u^1(i_1), u^2(i_2), ..., u^N(i_N))$ belonging to the finite set., i.e.,*

$$\begin{aligned} &s_n(i_1,...,i_N) = \\ &f(x(u^1(i_1), u^2(i_2), ..., u^N(i_N))) + o_\omega(1) \end{aligned} \tag{11}$$

- *For the selected actions $u_n^k = u^k(i_k)$ at time n, the normalized environment reaction $\hat{\xi}_n$ is asymptotically equal to $\Delta(i_1,...,i_N)$, i.e.,*

$$\hat{\xi}_n \overset{a.s.}{=} \Delta(i_1,...,i_N) + o_\omega(1) \in [0,1) \tag{12}$$

where

$$\begin{aligned} &\Delta(i_1,...,i_N) = \\ &[f(x(i_1,...,i_N)) - f(x(\alpha_1,...,\alpha_N))]/ \\ &\left[\max_{(i_1,...,i_N)} [f(x(i_1,...,i_N)) - f(x(\alpha_1,...,\alpha_N))] + 1\right] \end{aligned} \tag{13}$$

$\Delta(i_1,...,i_N) \in [0,1)$, $(\alpha_1,...,\alpha_N) :=$ arg $\min\limits_{(i_1,...,i_N)} f(x(i_1,...,i_N))$.

- *For the optimal action $u_n = u(x(\alpha_1,...,\alpha_N))$, the normalized environment reaction is asymptotically equal to 0, i.e.,*

$$\hat{\xi}_n \overset{a.s.}{=} o_\omega(1) \tag{14}$$

if $u_n = u(x(\alpha_1,...,\alpha_N))$.

Lemma 2. *If for some reinforcement scheme the following inequality holds*

$$\frac{1}{n}\sum_{t=1}^{n}\prod_{k=1}^{N} p_n^k(i_k) \geq O\left(\frac{1}{n^\tau}\right), \tau \in \left(0, \frac{1}{2}\right) \tag{15}$$

then, for any small positive ε this implies

$$\begin{aligned} &s_n(i_1,...,i_N) - f(x(i_1,...,i_N)) \\ &\overset{a.s.}{=} o_\omega\left(\frac{1}{n^{1/2-\tau-\varepsilon}}\right) \end{aligned} \tag{16}$$

and, as a result, for large enough $n \geq n_0(\omega)$, it follows

$$\hat{\xi}_n \overset{a.s.}{=} \Delta(i_1,...,i_N) + o_\omega\left(\frac{1}{n^{1/2-\tau-\varepsilon}}\right) \in [0,1) \tag{17}$$

$$\begin{aligned} &\mathbf{E}\left\{\left(\hat{\xi}_n - \Delta(i_1,...,i_N)\right)^2 \mid k = \overline{1,N}\right\} \\ &\overset{a.s.}{=} o\left(\frac{1}{n^{1-2\tau}}\right) \end{aligned} \tag{18}$$

with $u_n^k = u^k(i_k)$, *and*

$$\begin{aligned} &\mathbf{E}\left\{\hat{\xi}_n \mid k = \overline{1,N}\right\} \\ &= \Delta(i_1,...,i_N) + o\left(\frac{1}{n^{1/2-\tau}}\right) \end{aligned} \tag{19}$$

The proofs for the Lemmas are omitted here (see (Najim *et al.*, 2002)). They are based on the Borel-Cantelli Lemma, the strong law of large numbers and Lemma 4 in (Poznyak *et al.*, 2000) Appendix A.

Finally, the automaton input is used in connection with a modified version of the Bush-Mosteller reinforcement scheme (Najim and Poznyak, 1994) to adjust the probabilities distributions, i.e.,

$$\begin{cases} p_{n+1}^k(1) = p_n^k(1) + \Delta p_{n+1}^k(1) \\ p_{n+1}^k(2) = 1 - p_{n+1}^k(1) \end{cases} \tag{20}$$

$$\Delta p_{n+1}^k(1) = \gamma_n^k\left[u_n^k - p_n^k(1) + \widehat{\xi}_n^k(1 - 2u_n^k)\right]$$

where

$$\gamma_n^k \in [0,1], \quad \widehat{\xi}_n^i \in [0,1]$$

The original Bush-Mosteller reinforcement scheme uses a binary input (P-model environment) and a constant correction factor $\gamma_n^k = \gamma = const$.

The loss function Φ_n associated with each learning automaton is given by

$$\Phi_n^k = \frac{1}{n}\sum_{t=1}^{n}\widehat{\xi}_t^k = \frac{1}{n}\sum_{t=1}^{n}\widehat{\xi}_t \tag{21}$$

It is a useful quantity for judging the behavior of a learning automaton. We will show in the sequel that if a stochastic automaton minimizes its loss function then it automatically solves the corresponding unconstrained stochastic optimization problem on a discrete set.

Remark. *The accuracy of the approximation of the initial optimization problem on continuous set by the optimization problem on finite (discrete) set can be estimated for a wide enough class of Lipschitz functions (see section 1.5 of (Poznyak and Najim, 1997)) : to obtain an $\varepsilon-$approximation of the initial optimization problem it is enough to use the partition of the given compact set X with the diameter D into a subsets X_k $(k = 1,...,N)$ with diameters*

$$d_k \leq \frac{D}{N} \leq \frac{\varepsilon}{\max_{k=1,...,N} L_k^0}$$

(L_k^0 is the Lipschitz coefficient at X_k) and with the integer N, characterizing the number of such subsets, such a way that the following inequality:

$$N \geq \frac{D \max_{k=1,...,N} L_k^0}{\varepsilon}$$

should be satisfied.

The convergence as well the convergence rate of this optimization algorithm will be stated in the next section.

4. ASYMPTOTIC PROPERTIES

(H3) In this study, we assume that the correction factor γ_n^k in (20) is time-varying and is selected for any k according to the following rule:

$$\gamma_n^k = \frac{\gamma}{n+a},\ \gamma \in (0,1)\,,\ a > \gamma \qquad (22)$$

(H4) The initial probabilities are assumed to be strictly positive, i.e.,

$$p_1^k(i) > 0 \quad \forall k = 1, ..., N;\ i = 1, 2$$

The Bush-Mosteller scheme (20) can be rewritten in the following vector form:

$$\begin{aligned} &p_{n+1}^k = p_n^k \qquad (23)\\ &+\gamma_n^k \left[e(u_n^k) - p_n^k + \widehat{\xi}_n (\bar{e} - 2e(u_n^k)) \right] \end{aligned}$$

with $\gamma_n^k = \gamma_n := \frac{\gamma}{n+a}$, $\gamma \in (0, 1/N)$, $a > N\gamma$, $\widehat{\xi}_n \in [0, 1)$, $e(u_n^k) := \left(u_n^k, 1 - u_n^k\right)^\intercal$, $u_n^k = \{0; 1\}$ and $\bar{e} := (1, 1)^\intercal$.

Theorem. *For the Bush-Mosteller scheme (23), condition (9) is satisfied, and if the assumptions (**H1**) - (**H4**) hold, and the optimal action is single, i.e.,*

$$\min_{(i_1,...,i_N) \neq (\alpha_1,...,\alpha_N)} \Delta(i_1, ..., i_N) := \Delta^* > 0 \qquad (24)$$

then, the given collection of automata with binary actions selects asymptotically the global optimal point $x(\alpha_1, ..., \alpha_N)$ and the loss function Φ_n (21) tends to its minimal value equal to zero $\left(\min_{(i_1,...,i_N)} \Delta(i_1, ..., i_N) = 0 \right)$ with probability one, that is,

$$\Phi_n \overset{a.s.}{=} o_\omega\left(\frac{1}{n^{1/2 - N\gamma - \varepsilon}}\right)$$

The proof is obtained by selecting a Lyapunov function $W_n := (1 - \prod_{k=1}^N p_n^k(\alpha_k)) \,/\, \prod_{k=1}^N p_n^k(\alpha_k)$, finding an upper bound for $\mathbf{E}\{W_{n+1}/\mathcal{F}_n\}$ and using the Robbins-Siegmund theorem, hence showing that $W_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0$, which implies $\prod_{k=1}^N p_n^k(\alpha_k) \underset{n\to\infty}{\overset{a.s.}{\to}} 1$ and $\hat{\xi}_n \underset{n\to\infty}{\overset{a.s.}{\to}} 0$. The proof is omitted here as it is lengthy, see (Najim *et al.*, 2002); a similar treatment of $W_n := (1 - p(\alpha)) \,/\, p(\alpha)$ in the case of a single automaton can be found from Section 3.6.1 in (Poznyak and Najim, 1997). This theorem shows that, a team of learning binary automata using the Bush-Mosteller reinforcement scheme with the normalization procedure, described above, selects asymptotically the optimal actions.

The next corollary gives the estimation of the rate of optimization.

Corollary *(on convergence rate).* *Under the assumptions of this theorem it follows:*

$$W_n \overset{a.s.}{=} o\left(\frac{1}{n^\nu}\right)$$

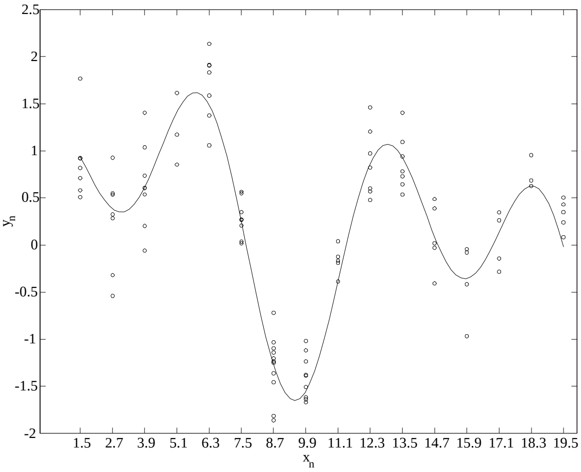

Fig. 2. Multimodal function. Solid line shows the true function f, noisy data is evaluated at points obtained using $N = 4$.

where

$$0 < \nu < \frac{1}{2} - 2N\gamma$$

So, for a given $\gamma \in (0, a/N)$ the order ν of the learning rate $n^{-\nu}$ decrease if the number N of automata team increase.

The proof is obtained using Lemma A.3-2 in (Poznyak and Najim, 1997) (again, see (Najim *et al.*, 2002)).

5. NUMERICAL EXAMPLE

In order to illustrate the feasibility and the performance of the algorithm presented above, let us consider the following optimization problem (minimization of a multimodal function):

$$\min_{x \in [1.5, 19.5]} f(x) \qquad (25)$$

where

$$f(x) = \cos x + \sin \frac{x}{2} + \frac{1}{2} \sin \frac{x}{4} \qquad (26)$$

and $w \sim N(0, 0.1)$. x is obtained from (3). The task is then to find the optimal combination of u^k's, $k = 1, 2, ..., N$. Figure 2 shows an example of the data for $N = 4$, $n = 100$. The minimum is found at $x^* = 9.9$.

Figure 3 (gray dots) shows the probabilities $p_t^k(1)$ in a typical simulation with $N = 4$, $\gamma = \frac{1}{2N}$, $a = 1$, $t_0 = 10000N$, $T = t_0 + 10000$, and

$$\gamma_n = \begin{cases} \gamma & \text{if } t < t_0 \\ \dfrac{\gamma}{(t - t_0) + a} & t_0 \leq t \leq T \end{cases} \qquad (27)$$

It can be seen that for the first 30 000 iterations, no apparent learning takes place. Taking the sample mean (solid line) shows, however, that

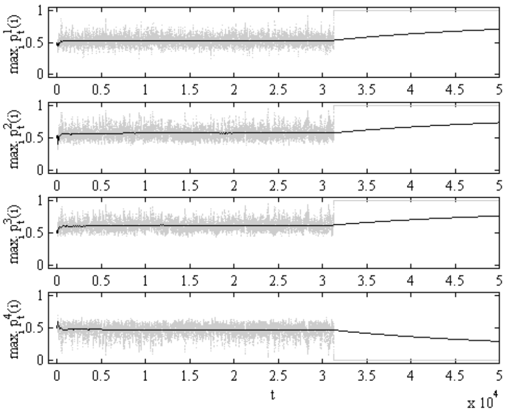

Fig. 3. Evolution of the probabilities p_t^k (1).

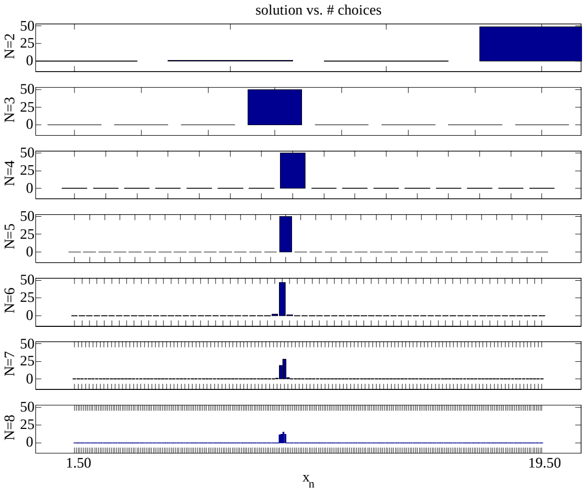

Fig. 4. Frequencies of the solutions found.

already after few hundred iterations the mean probabilities remain larger than 0.5 ($k = 1, 2, 3$) and less than 0.5 ($k = 4$). At iterations $t \approx 31275 - 31300$, the probabilities suddenly converge to 1's and 0's, and remain there until the iterations are terminated.

In order to have a better view of the practical viability of the approach, 50 repeated simulations were conducted using $N = 2, 3, ..., 8$ using a constant $\gamma_n = \gamma = 0.25$. Simulations were stopped when the probability for selecting a single action in the discretized search space was 99.99%, i.e., $\prod_{k=1}^{N} \max_{i=1,2} p_n^k(i) > 0.9999$. The results are summarized in Figs. 4–5. Fig 4 shows the number times a particular solution was obtained (out of the 50 test runs). For $N = 3$, 4 and 5, all 50 simulations converged to the optimum x^*. For $N = 6, 7$ and 8, the correct optimum was found in 47, 19 and 12 test runs, respectively, while the other solutions were found in the close neighborhood of x^*. Fig. 5 shows the histogram of the distribution of the number of iteration rounds required until convergence. As expected, the number of required iterations increases with N.

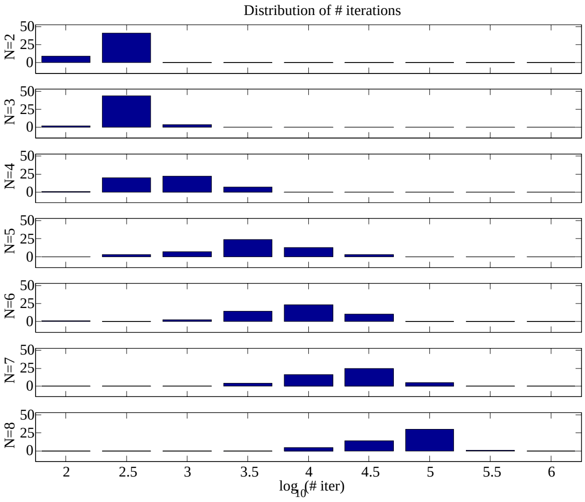

Fig. 5. Distribution of the number of iteration rounds.

6. CONCLUSIONS

This study demonstrates a potentially powerful tool for optimization purposes. The approach adopted here is based on a team of learning stochastic automata with a continuous input (environment response) and binary outputs. A modified version of the Bush-Mosteller reinforcement scheme is used for the adjustment of the probability distributions. The asymptotic properties of this random search optimization technique have been stated and a numerical example illustrates the feasibility and the performance of the optimization algorithm.

7. REFERENCES

Howell, M (2000). Genetic learning automata. Technical Report TT 2001. Department of Aeronautical and Automotive Engineering, Loughborough University.

Najim, K, A S Poznyak and E Ikonen (2002). Optimization technique based on a cooperative game of learning automata with binary outputs. *(submitted manuscript)*.

Najim, K. and A. S. Poznyak (1994). *Learning Automata Theory and Applications*. Pergamon Press. London.

Najim, K. and G. Oppenheim (1991). Learning systems: Theory and application. *IEE Proceedings-E* **138**(4), 183–192.

Poznyak, A. S. and K. Najim (1997). *Learning Automata and Stochastic Optimization*. Springer-Verlag. Berlin.

Poznyak, A.S., K. Najim and E. Gomez (2000). *Self-Learning Control of Finite Markov Chains*. Marcel Dekker. New York.

Tsypkin, Y. (1973). *Foundations of the Theory of Learning Systems*. Academic Press.