ELSEVIER International Journal of Approximate Reasoning 31 (2002) 313–340

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

Mathematical modelling of $\mathrm{UMDA_c}$ algorithm with tournament selection. Behaviour on linear and quadratic functions

C. González, J.A. Lozano, P. Larrañaga *

Department of Computer Science and Artificial Intelligence, Intelligent Systems Group, University of the Basque Country, M.de Lardizabal Pasalekua, 1, 20009 Donostia-San Sebastian, Spain

Received 1 January 2002; accepted 1 July 2002

Abstract

This paper presents a theoretical study of the behaviour of the univariate marginal distribution algorithm for continuous domains ($\mathrm{UMDA_c}$) in dimension n. To this end, the algorithm with tournament selection is modelled mathematically, assuming an infinite number of tournaments.

The mathematical model is then used to study the algorithm's behaviour in the minimization of linear functions $L(\mathbf{x}) = a_0 + \sum_{i=1}^{n} a_i x_i$ and quadratic function $Q(\mathbf{x}) = \sum_{i=1}^{n} x_i^2$, with $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $a_i \in \mathbb{R}$, $i = 0, 1, \ldots, n$. Linear functions are used to model the algorithm when far from the optimum, while quadratic function is used to analyze the algorithm when near the optimum.

The analysis shows that the algorithm performs poorly in the linear function $L_1(\mathbf{x}) = \sum_{i=1}^{n} x_i$. In the case of quadratic function $Q(\mathbf{x})$ the algorithm's behaviour was analyzed for certain particular dimensions. After taking into account some simplifications we can conclude that when the algorithm starts near the optimum, $\mathrm{UMDA_c}$ is able to reach it. Moreover the speed of convergence to the optimum decreases as the dimension increases.

Keywords: $\mathrm{UMDA_c}$; Linear functions; Quadratic functions; Tournament selection; Mathematical modeling; EDAs

* Corresponding author. Tel.: +34-943-018045; fax: +34-943-219306.

E-mail addresses: ccpgomoc@si.ehu.es (C. González), ccploalj@si.ehu.es (J.A. Lozano), ccplamup@si.ehu.es (P. Larrañaga).

PII: S0888-613X(02)00092-0

1. Introduction

Estimation of distribution algorithms (EDAs) are a new and promising paradigm for evolutionary computation [11,15]. EDAs emerge as a generalization of genetic algorithms (GAs), for the purpose of overcoming the two main problems: poor performance in certain deceptive problems and the difficulty of mathematically modelling a huge number of algorithm variants.

Introduced by Mühlenbein and Paaß [15], EDAs constitute an example of stochastic heuristics based on populations of individuals, each of which encodes a possible solution of the optimization problem. These populations evolve in successive generations as the search progresses, organized in the same way as most evolutionary computation heuristics. In contrast to GAs, which consider the crossover and mutation operators as essential tools to generate new populations, EDAs replace those operators by estimating and sampling the joint probability distribution of the selected individuals.

However, the bottleneck of this new heuristic lies in estimating the joint probability distribution associated with the database containing the selected individuals. To avoid this problem, several authors have proposed different algorithms where simplified assumptions concerning the conditional dependencies between the variables of the joint probability distribution are made. A review of the different approaches in the combinatorial and numerical fields can be found in [8–10,16].

During recent years much effort has been devoted to creating new EDAs and EDA applications. However this development has not been accompanied by mathematical analysis. There are very few works devoted to a mathematical modelling of EDAs in the literature.

Reviewing the literature, we can distinguish between papers that analyze EDAs in discrete domains and those that consider continuous domains.

In discrete domains the most general results are given by González et al. [5] and by Mühlenbein et al. [14]. In the first paper the authors not only unify most of the theoretical results found in the discrete EDA literature, but present a new general convergence theorem for these algorithms. In the second paper an EDA that uses Boltzmann selection is introduced: Boltzmann estimation of distribution algorithm (BEDA). Furthermore the convergence for infinite populations of a general BEDA is shown.

There are other works that analyze particular instances of discrete EDAs. In [13], it is shown that univariate marginal distribution algorithm (UMDA) with infinite population and proportional selection can only reach local optima. In addition, there are papers that analyze the population based incremental learning (PBIL) algorithm. Höhfeld and Rudolph [7] study the behaviour of PBIL in linear functions. González et al. [6] model this algorithm by means of Markov chains to show that the convergence of PBIL, applied to the OneMax function in two dimensions, has a strong dependence on the initial parameters.

In [4], the authors associate a discrete dynamical system with PBIL, and demonstrate that the algorithm follows the iterates of that discrete dynamical system, concluding that all the points of the search space are fixed points of the dynamical system, and that the local optimum points coincide with the stable fixed points. Berny [2] shows that the PBIL algorithm can be derived from a gradient dynamical system. Furthermore he carries out a stability analysis of the cited system, showing that PBIL can only converge to local solutions.

Finally we mention papers devoted to a mathematical analysis of EDAs in continuous domains. In [1] the population based incremental learning algorithm for continuous domains (PBIL_c) is examined, carrying out an analysis for real continuous functions similar to the analysis made in [2]. However, in this case the author does not offer stability results. In [17] the factorized distribution algorithm is theoretically analyzed and convergence results are given.

The purpose of this paper is to contribute a mathematical analysis of a continuous EDA, the UMDA_c. The UMDA is the simplest version of EDAs. The discrete version was introduced by Mühlenbein [12], while the first continuous version was given by Larrañaga et al. [8,10]. This algorithm does not take into account dependencies among the variables, therefore it is assumed that the n-dimensional joint probability density factorizes as a product of n independent univariate marginal densities.

It was of particular interest to see how the UMDA_c algorithm with tournament selection performs. To this end we mathematically modelled the application of this algorithm to the minimization of two kinds of functions. First n-dimensional linear functions were used to model the algorithm when far from the optimum. Next quadratic function was used to analyze the algorithm when near the optimum.

The remainder of this paper is organized as follows: Section 2 introduces in detail UMDA_c with tournament selection. Section 3 is devoted to the mathematical modelling of the algorithm. Section 4 analyzes the modelling of linear functions, while Section 5 analyzes the case of quadratic functions. Finally, we draw conclusions in Section 6.

2. The UMDA_c algorithm with tournament selection

This section describes in detail how the UMDA_c algorithm with tournament selection works.

The algorithm works as follows. At each step t, an n-dimensional random variable $\mathbf{X}^t = (X_1^t, \ldots, X_n^t)$ is maintained. In the literature related to UMDA_c it is usual to assume that the joint probability distribution of $\mathbf{X}^t$ follows an n-dimensional normal distribution which is factorized by a product of n unidimensional and independent normal densities. This assumption will be made here. Therefore each component of $\mathbf{X}^t$ is distributed as a unidimensional normal, that is $X_i^t \rightsquigarrow \mathcal{N}(\mu_i^t, \sigma_i^t)$, where $f_{\mathcal{N}(\mu_i^t,\sigma_i^t)}(x_i) = (1/\sqrt{2\pi}\sigma_i^t)\mathrm{e}^{-(x_i-\mu_i^t)^2/2(\sigma_i^t)^2}$ with $i = 1, \ldots, n$. In other words $f_{\mathcal{N}(\mu_i^t,\sigma_i^t)}(x_i)$ denotes the density function of a unidimensional normal with mean μ_i^t and standard deviation σ_i^t in point x_i.

Drawing the above n-dimensional random variable, two individuals are obtained, and the better one is selected, i.e. a tournament selection is made. This process is repeated N times, obtaining the population of selected individuals, after which this population is used to obtain the means and standard deviations of the random variable $\mathbf{X}^{t+1}$. These parameters are estimated by using their corresponding maximum likelihood estimators. In this way the new unidimensional distributions at step $t+1$ are achieved. Fig. 1 shows a pseudocode for this algorithm for the minimization of function $G(\mathbf{x})$.

Our objective is to learn how the density functions change with time. This enables us to know how μ_i^t and σ_i^t evolve when t increases.

Obtain randomly the parameters of a normal probability distribution for each variable

while no convergence **do**

  begin

    for $(j = 1;\ j \leq N;\ j{+}{+})$

      begin

        Drawing $\mathbf{X}^t$ obtain 2 individuals:

          $\mathbf{x}_{1,j}^t = (x_{1,j}^{1,t}, \ldots, x_{1,j}^{n,t})$

          $\mathbf{x}_{2,j}^t = (x_{2,j}^{1,t}, \ldots, x_{2,j}^{n,t})$

        Evaluate $\mathbf{x}_{1,j}^t$, $\mathbf{x}_{2,j}^t$

        Select the better one:

          $\mathbf{x}_{(1:2),j}^t = argmin_{\mathbf{x}}\{G(\mathbf{x}_{1,j}^t), G(\mathbf{x}_{2,j}^t)\}$

      end

    for $(i = 1;\ i \leq n;\ i{+}{+})$

      begin

        Estimate the parameters of the new density functions:

$$\mu_i^{t+1} = \frac{\sum_{j=1}^{N} x_{(1:2),j}^{i,t}}{N}$$

$$\sigma_i^{t+1} = \sqrt{\frac{\sum_{j=1}^{N} (x_{(1:2),j}^{i,t} - \mu_i^{t+1})^2}{N}}$$

      end

  end

Fig. 1. Pseudocode for $\mathrm{UMDA_c}$ with tournament selection.

3. Mathematical modelling

To model the UMDA$_c$ algorithm with tournament selection for continuous optimization problems with n variables, we take a case in which at each step an infinite number of tournaments is made. The mathematical model will depend on the function being optimized. At first we try to model the algorithm as generally as possible, hence we assume that the function to minimize is

$$G : \mathbb{R}^n \to \mathbb{R}. \tag{1}$$

However, as we will see later, at some point it will be necessary to particularize the function we are analyzing.

As noted above, we assume that at each step t, each variable follows a unidimensional normal distribution, and has associated the following density function:

$$f_{X_i^t}(x_i) = f_{\mathcal{N}(\mu_i^t,\sigma_i^t)}(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i^t} \mathrm{e}^{-(x_i-\mu_i^t)^2/2(\sigma_i^t)^2}, \tag{2}$$

where $f_{X_i^t}(x_i)$ denotes the density function of the random variable X_i^t. As we are working with UMDA$_c$ these variables are independent. Hence at each step t we have an n-dimensional random variable $\mathbf{X}^t = (X_1^t, \ldots, X_n^t)$ following the density $f_{\mathbf{X}^t}(\mathbf{x})$ with

$$f_{\mathbf{X}^t}(\mathbf{x}) = \prod_{i=1}^{n} f_{\mathcal{N}(\mu_i^t,\sigma_i^t)}(x_i). \tag{3}$$

To simplify notation, each density function associated with each variable X_i^t will henceforth be denoted by

$$f_i^t(x_i) = f_{X_i^t}(x_i) = f_{\mathcal{N}(\mu_i^t,\sigma_i^t)}(x_i), \quad \text{with } i = 1, 2, \ldots, n. \tag{4}$$

Likewise, its associated distribution function will be denoted by

$$F_i^t(x_i) = \int_{-\infty}^{x_i} f_i^t(s)\,\mathrm{d}s, \quad \text{with } i = 1, 2, \ldots, n. \tag{5}$$

We use the usual notation not only in the case of the standard normal density:

$$\phi(x) = f_{\mathcal{N}(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}, \tag{6}$$

but also in the case of its associated distribution function:

$$\Phi(x) = \int_{-\infty}^{x} \phi(s)\,\mathrm{d}s. \tag{7}$$

At each step t the random variable $\mathbf{X}_{(1:2)}^t$ is considered, i.e. the random variable of the better of two variables $\mathbf{X}^t$. Thus

$$\mathsf{E}[\mathbf{X}^t_{(1:2)}] = (\mathsf{E}[X^t_{(1:2),1}], \ldots, \mathsf{E}[X^t_{(1:2),n}])$$
$$\mathsf{Var}[\mathbf{X}^t_{(1:2)}] = (\mathsf{Var}[X^t_{(1:2),1}], \ldots, \mathsf{Var}[X^t_{(1:2),n}]),$$

after which, the new distributions at time $t+1$ are obtained. Hence each X_i^{t+1} obeys $\mathcal{N}(\mu_i^{t+1}, \sigma_i^{t+1})$, with $\mu_i^{t+1} = \mathsf{E}[X^t_{(1:2),i}]$ and $\sigma_i^{t+1} = (\mathsf{Var}[X^t_{(1:2),i}])^{1/2}$.

As we want to model the behaviour of the algorithm, we need to know explicitly the expressions of μ_i^{t+1} and σ_i^{t+1} given μ_i^t, σ_i^t, for $i = 1, \ldots, n$. Then we can use these expressions to analyze the sequences $\{\mu_i^t\}_t$ and $\{\sigma_i^t\}_t$ with $t \in \mathbb{N}$, and to study how they evolve when the number of iterates increases. In other words, we study the limits:

$$\lim_{t\to\infty} \mu_i^t, \tag{8}$$

$$\lim_{t\to\infty} \sigma_i^t. \tag{9}$$

In order to calculate μ_i^{t+1} and σ_i^{t+1}, we have to obtain the density function associated with the best individual of each tournament. We denote this density function by $f^t_{(1:2)}(x_1, \ldots, x_n)$. Notice that the previous density will depend on G, the objective function that we are considering.

3.1. Calculation of $f^t_{(1:2)}(x_1, \ldots, x_n)$

In order to calculate the density function $f^t_{(1:2)}(x_1, \ldots, x_n)$ we proceed as follows. First we obtain its associated distribution function, $F^t_{(1:2)}(x_1, \ldots, x_n)$. Then we derive this distribution function and obtain the density function $f^t_{(1:2)}(x_1, \ldots, x_n)$.

Let $\mathbf{X}_1^t = (X^t_{1,1}, \ldots, X^t_{1,n})$ be the random variable associated with the first individual obtained in the tournament at step t, and $\mathbf{X}_2^t = (X^t_{2,1}, \ldots, X^t_{2,n})$ the random variable corresponding to the second individual. Thus, the distribution function $F^t_{(1:2)}(x_1, \ldots, x_n)$ is

$$F^t_{(1:2)}(x_1, \ldots, x_n) = P((X^t_{(1:2),1}, \ldots, X^t_{(1:2),n}) \leqslant (x_1, \ldots, x_n)). \tag{10}$$

To make the calculus easier, we express the random variable associated with the best individual in each tournament as a sum of random variables:

$$\mathbf{X}^t_{(1:2)} = \mathbf{X}^t_1 \cdot I_{\{G(\mathbf{X}^t_1) \leqslant G(\mathbf{X}^t_2)\}} + \mathbf{X}^t_2 \cdot I_{\{G(\mathbf{X}^t_1) > G(\mathbf{X}^t_2)\}}, \tag{11}$$

where the random variable I_A denotes the characteristic function of the event A, hence

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

The event described in (10) is written as the following union of events:

$$(X^t_{(1:2),1},\dots,X^t_{(1:2),n}) \leqslant (x_1,\dots,x_n) = \{U^t_1 \cap U^t_2\} \cup \{V^t_1 \cap V^t_2\}, \tag{13}$$

where

$$\begin{aligned}
U^t_1 &= \{(X^t_{1,1},\dots,X^t_{1,n}) \leqslant (x_1,\dots,x_n)\}\\
U^t_2 &= \{G(X^t_{1,1},\dots,X^t_{1,n}) \leqslant G(X^t_{2,1},\dots,X^t_{2,n})\}\\
V^t_1 &= \{(X^t_{2,1},\dots,X^t_{2,n}) \leqslant (x_1,\dots,x_n)\}\\
V^t_2 &= \{G(X^t_{1,1},\dots,X^t_{1,n}) > G(X^t_{2,1},\dots,X^t_{2,n})\}.
\end{aligned}$$

We denote by U^t the event $U^t_1 \cap U^t_2$ and by V^t the event $V^t_1 \cap V^t_2$, and since

$$\{(X^t_{(1:2),1},\dots,X^t_{(1:2),n}) \leqslant (x_1,\dots,x_n)\} = \{U^t\} \cup \{V^t\}.$$

Hence, given that we have disjoint events, we can state that

$$F^t_{(1:2)}(x_1,\dots,x_n) = P(U^t) + P(V^t). \tag{14}$$

Taking into account that $P(U^t) = P(V^t)$ (G is a continuous function), it is enough to obtain $P(U^t)$. In order to do so we find the conditional probability $P(U^t|X^t_{1,1} = x_{1,1},\dots,X^t_{1,n} = x_{1,n})$ and then we integrate over the rest of the variables:

$$\begin{aligned}
&P(U^t|X^t_{1,1} = x_{1,1},\dots,X^t_{1,n} = x_{1,n})\\
&\quad = \begin{cases} P(G(X^t_{2,1},\dots,X^t_{2,n}) \geqslant G(x_{1,1},\dots,x_{1,n})) & \text{if} \quad x_{1,1} \leqslant x_1,\dots,x_{1,n} \leqslant x_n \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

To simplify the notation we write:

$$P(G(X^t_1,\dots,X^t_n) \geqslant G(x_1,\dots,x_n)) = A^t(G(\mathbf{x})). \tag{15}$$

Therefore

$$\begin{aligned}
P(U^t) &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} P(U^t|X^t_{1,1} = x_{1,1},\dots,X^t_{1,n} = x_{1,n})\\
&\qquad \cdot f^t_1(x_{1,1}) \dots f^t_n(x_{1,n})\,\mathrm{d}x_{1,1} \dots \mathrm{d}x_{1,n}\\
&= \int_{-\infty}^{x_{1,1}} \dots \int_{-\infty}^{x_{1,n}} A^t(G(\mathbf{x}_1)) f^t_1(x_{1,1}) \dots f^t_n(x_{1,n})\,\mathrm{d}x_{1,1} \dots \mathrm{d}x_{1,n}.
\end{aligned}$$

Hence, the distribution function of $\mathbf{X}^t_{(1:2)}$ is

$$\begin{aligned}
F^t_{(1:2)}(x_1,\dots,x_n) &= P((X^t_{(1:2),1},\dots,X^t_{(1:2),n}) \leqslant (x_1,\dots,x_n))\\
&= 2\int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} A^t(G(\mathbf{x})) \cdot f^t_1(x_1) \dots f^t_n(x_n)\,\mathrm{d}x_1 \dots \mathrm{d}x_n.
\end{aligned}$$

Deriving the above expression we obtain the density function as

$$f^t_{(1:2)}(x_1,\ldots,x_n) = 2A^t(G(\mathbf{x}))\prod_{i=1}^{n} f^t_i(x_i). \tag{16}$$

3.2. Calculation of μ_i^{t+1} and σ_i^{t+1}

To obtain μ_i^{t+1} and σ_i^{t+1} we must first calculate each marginal density function $f^t_{(1:2),i}(x_i)$ with $i = 1,\ldots,n$:

$$\mu_i^{t+1} = \int_{-\infty}^{\infty} x_i f^t_{(1:2),i}(x_i)\,\mathrm{d}x_i, \tag{17}$$

$$(\sigma_i^{t+1})^2 = \int_{-\infty}^{\infty} x_i^2 f^t_{(1:2),i}(x_i)\,\mathrm{d}x_i - (\mu_i^{t+1})^2. \tag{18}$$

The marginal densities can be expressed as follows:

$$\begin{aligned} f^t_{(1:2),i}(x_i) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f^t_{(1:2)}(x_1,\ldots,x_n)\,\mathrm{d}x_1\ldots\mathrm{d}x_{i-1}\,\mathrm{d}x_{i+1}\ldots\mathrm{d}x_n \\ &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} 2A^t(G(\mathbf{x}))\prod_{j=1}^{n} f^t_j(x_j)\,\mathrm{d}x_1\ldots\mathrm{d}x_{i-1}\,\mathrm{d}x_{i+1}\ldots\mathrm{d}x_n \\ &= 2f^t_i(x_i)h^t_i(x_i), \end{aligned} \tag{19}$$

where

$$h^t_i(x_i) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} A^t(G(\mathbf{x}))\prod_{\substack{j=1\\ j\neq i}}^{n} f^t_j(x_j)\,\mathrm{d}x_1\ldots\mathrm{d}x_{i-1}\,\mathrm{d}x_{i+1}\ldots\mathrm{d}x_n. \tag{20}$$

As can be seen in (19) the calculations of μ_i^{t+1} and σ_i^{t+1} are closely related to the objective function G. For this reason, our analysis will now focus on the following two cases:

- The case of linear functions.
- The case of quadratic functions.

4. Linear functions

We shall start by studying the simplest case, where the function $L(\mathbf{x})$ under consideration is

$$\begin{aligned} L :\ & \mathbb{R}^n \to \mathbb{R} \\ & \mathbf{x} \mapsto a_0 + \sum_{i=1}^{n} a_i x_i. \end{aligned} \tag{21}$$

This will help us to see how the algorithm performs far from the optimum.

4.1. Calculation of μ_i^{t+1}

As the calculations are analogous for each component, they will only be given for the first component:

$$\mathsf{E}[X^t_{(1:2),1}] = \mu_1^{t+1} = \int_{-\infty}^{\infty} x_1 f^t_{(1:2),1}(x_1)\,\mathrm{d}x_1 = \int_{-\infty}^{\infty} 2x_1 f_1^t(x_1) h_1^t(x_1)\,\mathrm{d}x_1. \qquad (22)$$

To simplify the notation, in the following calculations the superscript corresponding to the step is left out. It is, however, written in the final expression of μ_i^{t+1} and σ_i^{t+1}.

First we need to know the value of $A(L(\mathbf{x}))$

$$A(L(\mathbf{x})) = A\left(a_0 + \sum_{j=1}^n a_j x_j\right) = P\left(a_0 + \sum_{j=1}^n a_j X_j \geqslant a_0 + \sum_{j=1}^n a_j x_j\right), \qquad (23)$$

since each X_i is a random variable with density function $f_{\mathcal{N}(\mu_i,\sigma_i)}(x_i)$, we know that the random variable $T = \sum_{j=1}^n a_j X_j$ has density function:

$$\mathcal{N}\left(\sum_{j=1}^n a_j \mu_j, \sqrt{\sum_{j=1}^n a_j^2 \sigma_j^2}\right). \qquad (24)$$

Therefore

$$\begin{aligned} A(L(\mathbf{x})) &= P\left(\frac{T - \sum_{j=1}^n a_j \mu_j}{\sqrt{\sum_{j=1}^n a_j^2 \sigma_j^2}} \geqslant \frac{\sum_{j=1}^n a_j (x_j - \mu_j)}{\sqrt{\sum_{j=1}^n a_j^2 \sigma_j^2}}\right) \\ &= 1 - \Phi\left(\frac{\sum_{j=1}^n a_j (x_j - \mu_j)}{\sqrt{\sum_{j=1}^n a_j^2 \sigma_j^2}}\right). \end{aligned} \qquad (25)$$

Now we can calculate $h_1(x_1)$

$$h_1(x_1) = \int_{-\infty}^{\infty} f_n(x_n) \ldots \int_{-\infty}^{\infty} f_k(x_k) \ldots \int_{-\infty}^{\infty} f_2(x_2) A(L(\mathbf{x}))\,\mathrm{d}x_2 \ldots \mathrm{d}x_k \ldots \mathrm{d}x_n \qquad (26)$$

using the following notation:

$$g_k(x_1, x_{k+1}, x_{k+2}, \ldots, x_n) = \int_{-\infty}^{\infty} f_k(x_k) \ldots \int_{-\infty}^{\infty} f_2(x_2) A(L(\mathbf{x}))\,\mathrm{d}x_2 \ldots \mathrm{d}x_k \qquad (27)$$

with $k = 2, \ldots, n$, we know that

$$h_1(x_1) = g_n(x_1). \qquad (28)$$

We are going to prove by induction on k that

$$g_k(x_1, x_{k+1}, x_{k+2}, \ldots, x_n) = 1 - \Phi\left(\frac{\sum_{\substack{i=1 \\ i\neq 2,\ldots,k}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}\right). \tag{29}$$

First this is demonstrated when $k = 2$, after which we use as inductive hypothesis case k and demonstrate that (29) is fulfilled in case $k + 1$.

Before proving Eq. (29), we must first take into account the following results borrowed from [3], which will help us to make the calculations:

$$I_0(a, b) = \int_{-\infty}^{\infty} \mathrm{e}^{-s^2/2}\Phi(as + b)\,\mathrm{d}s = \sqrt{2\pi}\Phi\left(\frac{b}{\sqrt{1 + a^2}}\right). \tag{30}$$

$$I_1(a, b) = \int_{-\infty}^{\infty} s\mathrm{e}^{-s^2/2}\Phi(as + b)\,\mathrm{d}s = \frac{a}{\sqrt{1 + a^2}} \exp\left(\frac{-1}{2}\frac{b^2}{1 + a^2}\right). \tag{31}$$

Now we verify that (29) is satisfied when $k = 2$

$$\begin{aligned} & g_2(x_1, x_3, x_4, \ldots, x_n) \\ & \quad = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-(x_2-\mu_2)^2/2\sigma_2^2}\left(1 - \Phi\left(\frac{\sum_{i=1}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right)\right)\mathrm{d}x_2. \end{aligned} \tag{32}$$

Making the transformation of variable $(x_2 - \mu_2)/\sigma_2 = s$ we find that

$$\begin{aligned} & g_2(x_1, x_3, x_4, \ldots, x_n) \\ & \quad = 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-s^2/2}\Phi\left(\frac{a_2\sigma_2 s + a_2\mu_2 + \sum_{\substack{i=1 \\ i\neq 2}}^{n} a_i x_i - \sum_{i=1}^{n} a_i\mu_i}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right)\mathrm{d}s \\ & \quad = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \mathrm{e}^{-s^2/2}\Phi\left(\frac{a_2\sigma_2 s}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}} + \frac{\sum_{\substack{i=1 \\ i\neq 2}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right)\mathrm{d}s. \end{aligned} \tag{33}$$

Taking into account the result (30):

$$\begin{aligned} g_2(x_1, x_3, x_4, \ldots, x_n) &= 1 - \frac{1}{\sqrt{2\pi}}I_0\left(\frac{a_2\sigma_2}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}, \frac{\sum_{\substack{i=1 \\ i\neq 2}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}}\right) \\ &= 1 - \Phi\left(\frac{\sum_{\substack{i=1 \\ i\neq 2}}^{n} a_i(x_i - \mu_i)}{\sqrt{a_2^2\sigma_2^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}\right). \end{aligned} \tag{34}$$

By the inductive hypothesis, we assume that (29) is true for k, let us now see whether it is true for $k + 1$:

$$
\begin{aligned}
&g_{k+1}(x_1, x_{k+2}, x_{k+3}, \ldots, x_n) \\
&\quad = \int_{-\infty}^{\infty} f_{k+1}(x_{k+1}) g_k(x_1, x_{k+1}, x_{k+2}, \ldots, x_n)\, \mathrm{d}x_{k+1} \\
&\quad = 1 - \int_{-\infty}^{\infty} f_{k+1}(x_{k+1}) \Phi\left(\frac{\sum_{\substack{i=1 \\ i \neq 2,\ldots,k}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}} \right) \mathrm{d}x_{k+1}. \qquad (35)
\end{aligned}
$$

Taking into account the change of variable $(x_{k+1} - \mu_{k+1})/\sigma_{k+1} = s$, we find that

$$
\begin{aligned}
&g_{k+1}(x_1, x_{k+2}, x_{k+3}, \ldots, x_n) \\
&\quad = 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-s^2/2} \Phi\left(\frac{a_{k+1}\sigma_{k+1}s + \sum_{\substack{i=1 \\ i \neq 2,\ldots,k+1}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}} \right) \mathrm{d}s.
\end{aligned}
$$

By again using the result (30), we find that

$$
\begin{aligned}
&g_{k+1}(x_1, x_{k+2}, x_{k+3}, \ldots, x_n) \\
&\quad = 1 - \frac{1}{\sqrt{2\pi}} I_0\left(\frac{a_{k+1}\sigma_{k+1}}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}, \frac{\sum_{\substack{i=1 \\ i \neq 2,\ldots,k+1}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}} \right).
\end{aligned}
$$

Substituting the corresponding value of I_0 in this point

$$
\begin{aligned}
g_{k+1}(x_1, x_{k+2}, x_{k+3}, \ldots, x_n) &= 1 - \Phi\left(\frac{\dfrac{\sum_{\substack{i=1 \\ i \neq 2,\ldots,k+1}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}}{\sqrt{1 + \dfrac{a_{k+1}^2\sigma_{k+1}^2}{\sum_{i=2}^{k} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}}} \right) \\
&= 1 - \Phi\left(\frac{\sum_{\substack{i=1 \\ i \neq 2,\ldots,k+1}}^{n} a_i(x_i - \mu_i)}{\sqrt{\sum_{i=2}^{k+1} a_i^2\sigma_i^2 + \sum_{i=1}^{n} a_i^2\sigma_i^2}} \right). \qquad (36)
\end{aligned}
$$

So we have proven that (29) is fulfilled, therefore

$$
h_1(x_1) = g_n(x_1) = 1 - \Phi\left(\frac{a_1(x_1 - \mu_1)}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^{n} a_i^2\sigma_i^2}} \right). \qquad (37)
$$

Now we can calculate μ_1^{t+1}

$$\mu_1^{t+1} = \int_{-\infty}^{\infty} 2x_1 f_1(x_1) h_1(x_1)\,\mathrm{d}x_1$$
$$= \int_{-\infty}^{\infty} 2\frac{1}{\sqrt{2\pi}\sigma_1} x_1 \mathrm{e}^{-(x_1-\mu_1)^2/2\sigma_1^2}\left(1 - \Phi\left(\frac{a_1(x_1-\mu_1)}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)\right)\mathrm{d}x_1.$$

Using the new variable $s = (x_1 - \mu_1)/\sigma_1$, we obtain:

$$\begin{aligned}\mu_1^{t+1} &= \frac{2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma_1 s + \mu_1)\mathrm{e}^{-s^2/2}\left(1 - \Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)\right)\mathrm{d}s \\ &= \frac{2}{\sqrt{2\pi}}\left[\sigma_1\int_{-\infty}^{\infty} s\mathrm{e}^{-s^2/2}\,\mathrm{d}s + \mu_1\int_{-\infty}^{\infty}\mathrm{e}^{-s^2/2}\,\mathrm{d}s\right. \\ &\quad - \sigma_1\int_{-\infty}^{\infty} s\mathrm{e}^{-s^2/2}\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)\mathrm{d}s \\ &\quad \left. - \mu_1\int_{-\infty}^{\infty}\mathrm{e}^{-s^2/2}\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)\mathrm{d}s\right].\end{aligned} \tag{38}$$

Eqs. (30) and (31) help us to express the above integrals as follows:

$$\begin{aligned}\mu_1^{t+1} &= \frac{2}{\sqrt{2\pi}}\left[\sigma_1 \cdot 0 + \mu_1 \cdot \sqrt{2\pi} - \sigma_1 \cdot I_1\left(\frac{a_1\sigma_1}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}, 0\right)\right. \\ &\quad \left. - \mu_1 \cdot I_0\left(\frac{a_1\sigma_1}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}, 0\right)\right].\end{aligned}$$

Using the results (30) and (31):

$$\begin{aligned}\mu_1^{t+1} &= \frac{2}{\sqrt{2\pi}}\left[\mu_1\sqrt{2\pi} - \sigma_1\frac{a_1\sigma_1}{\sqrt{2a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}} - \mu_1\frac{\sqrt{2\pi}}{2}\right] \\ &= \mu_1 - \frac{a_1\sigma_1^2}{\sqrt{\pi}\sqrt{\sum_{i=1}^n a_i^2\sigma_i^2}}.\end{aligned} \tag{39}$$

Summarizing

$$\mu_1^{t+1} = \mu_1^t - \frac{a_1(\sigma_1^t)^2}{\sqrt{\pi}\sqrt{\sum_{i=1}^n a_i^2(\sigma_i^t)^2}}. \tag{40}$$

The expression for the expectation in any component i is obtained analogously:

$$\mu_i^{t+1} = \mu_i^t - \frac{a_i(\sigma_i^t)^2}{\sqrt{\pi}\sqrt{\sum_{j=1}^n a_j^2(\sigma_j^t)^2}}. \tag{41}$$

4.2. Calculation of σ_i^{t+1}

As done before in calculating μ_i^{t+1}, we only make the calculations for $i = 1$, after which we generalize the result:

$$(\sigma_1^{t+1})^2 = \mathsf{Var}[X_{(1:2),1}^{t+1}] = \mathsf{E}[(X_{(1:2),1}^{t+1})^2] - (\mathsf{E}[X_{(1:2),1}^{t+1}])^2. \tag{42}$$

We start by obtaining $\mathsf{E}[(X_{(1:2),1}^{t})^2]$

$$\begin{aligned}\mathsf{E}[(X_{(1:2),1}^{t+1})^2] &= \int_{-\infty}^{\infty} 2x_1^2 f_1(x_1)h_1(x_1)\,\mathrm{d}x_1 \\ &= \int_{-\infty}^{\infty} 2\frac{1}{\sqrt{2\pi}\sigma_1} x_1^2 \mathrm{e}^{-(x_1-\mu_1)^2/2\sigma_1^2} \\ &\quad \cdot \left(1 - \Phi\left(\frac{a_1(x_1-\mu_1)}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)\right)\mathrm{d}x_1.\end{aligned}$$

Making the change of variable $(x_1 - \mu_1)/\sigma_1 = s$, we have

$$\begin{aligned}&\mathsf{E}[(X_{(1:2),1}^{t+1})^2] \\ &= \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma_1 s + \mu_1)^2 \mathrm{e}^{-s^2/2}\left(1 - \Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)\right)\mathrm{d}s \\ &= \frac{2}{\sqrt{2\pi}}\left[\sigma_1^2\int_{-\infty}^{\infty} s^2\mathrm{e}^{-s^2/2}\,\mathrm{d}s + 2\sigma_1\mu_1\int_{-\infty}^{\infty} s\mathrm{e}^{-s^2/2}\,\mathrm{d}s + \mu_1^2\int_{-\infty}^{\infty}\mathrm{e}^{-s^2/2}\,\mathrm{d}s\right. \\ &\qquad - \sigma_1^2\int_{-\infty}^{\infty} s^2\mathrm{e}^{-s^2/2}\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)\mathrm{d}s \\ &\qquad - 2\sigma_1\mu_1\int_{-\infty}^{\infty} s\mathrm{e}^{-s^2/2}\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)\mathrm{d}s \\ &\qquad \left.- \mu_1^2\int_{-\infty}^{\infty}\mathrm{e}^{-s^2/2}\Phi\left(\frac{a_1\sigma_1 s}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}\right)\mathrm{d}s\right].\end{aligned} \tag{43}$$

Taking into account expressions (30), (31) and

$$
\begin{aligned}
I_2(a,b) &= \int_{-\infty}^{\infty} s^2 \mathrm{e}^{-s^2/2} \Phi(as+b)\,\mathrm{d}s \\
&= \sqrt{2\pi}\Phi\left(\frac{b}{\sqrt{1+a^2}}\right) - \frac{a^2 b}{(1+a^2)\sqrt{1+a^2}} \exp\left(\frac{-1}{2}\frac{b^2}{1+a^2}\right),
\end{aligned}
\tag{44}
$$

the integrals in (43) can be written as follows:

$$
\begin{aligned}
\mathsf{E}[(X^{t+1}_{(1:2),1})^2] = \frac{2}{\sqrt{2\pi}} \Bigg[& \sigma_1^2 \cdot \sqrt{2\pi} + 2\sigma_1\mu_1 \cdot 0 + \mu_1^2 \cdot \sqrt{2\pi} \\
& - \sigma_1^2 \cdot I_2\left(\frac{a_1\sigma_1}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}, 0\right) \\
& - 2\sigma_1\mu_1 \cdot I_1\left(\frac{a_1\sigma_1}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}, 0\right) \\
& - \mu_1^2 \cdot I_0\left(\frac{a_1\sigma_1}{\sqrt{a_1^2\sigma_1^2 + 2\sum_{i=2}^n a_i^2\sigma_i^2}}, 0\right)\Bigg].
\end{aligned}
$$

Using (30), (31) and (44) we obtain:

$$
\begin{aligned}
& \mathsf{E}[(X^{t+1}_{(1:2),1})^2] \\
&\quad = \frac{2}{\sqrt{2\pi}}\left[\sigma_1^2\sqrt{2\pi} + \mu_1^2\sqrt{2\pi} - \sigma_1^2\frac{\sqrt{2\pi}}{2} - 2\sigma_1\mu_1\frac{a_1\sigma_1}{\sqrt{2\sum_{i=1}^n a_i^2\sigma_i^2}} - \mu_1^2\frac{\sqrt{2\pi}}{2}\right] \\
&\quad = \sigma_1^2 + \mu_1^2 - \frac{2a_1\mu_1\sigma_1^2}{\sqrt{\pi}\sqrt{\sum_{i=1}^n a_i^2\sigma_i^2}}.
\end{aligned}
\tag{45}
$$

Now the superscripts corresponding to the step are written in order to obtain the full expression:

$$
\mathsf{E}[(X^{t+1}_{(1:2),1})^2] = (\sigma_1^t)^2 + (\mu_1^t)^2 - \frac{2a_1\mu_1^t(\sigma_1^t)^2}{\sqrt{\pi}\sqrt{\sum_{i=1}^n a_i^2(\sigma_i^t)^2}}.
\tag{46}
$$

Finally

$$
\sigma_1^{t+1} = \sigma_1^t\sqrt{1 - \frac{a_1^2(\sigma_1^t)^2}{\pi\sum_{i=1}^n a_i^2(\sigma_i^t)^2}}.
\tag{47}
$$

An analogous expression for any component i can be given:

$$
\sigma_i^{t+1} = \sigma_i^t\sqrt{1 - \frac{a_i^2(\sigma_i^t)^2}{\pi\sum_{j=1}^n a_j^2(\sigma_j^t)^2}}.
\tag{48}
$$

4.3. Analyzing the algorithm's behaviour

Having obtained the expressions of μ_i^{t+1} and σ_i^{t+1}, we now try to predict the algorithm's behaviour when t increases. This is done by analyzing each sequence of means $\{\mu_i^t\}_t$ and each sequence of standard deviations $\{\sigma_i^t\}_t$ with $t \in \mathbb{N}$.

To prove that the algorithm performs properly we must show that

$$\begin{aligned} &\text{when } a_i > 0 \Rightarrow \mu_i^t \to -\infty, \quad \text{as } t \to \infty \\ &\text{when } a_i < 0 \Rightarrow \mu_i^t \to +\infty, \quad \text{as } t \to \infty, \end{aligned} \tag{49}$$

because if so, the algorithm would improve at each step, minimizing unboundedly the value of the objective function.

Unfortunately means sequences $\{\mu_i^t\}_t$ with $t \in \mathbb{N}$ are difficult to study when standard deviations σ_i^t are not equal in each component. However, we can state that the improvement at each step and in each component can be written as follows:

$$|\mu_i^{t+1} - \mu_i^t| = \left| \frac{a_i(\sigma_i^t)^2}{\sqrt{\pi}\sqrt{\sum_{j=1}^n a_j^2(\sigma_j^t)^2}} \right|. \tag{50}$$

Hence, given that sequences $\{\sigma_i^t\}_t$ decrease for all i (see Eq. (48), with $a_i > 0$), the improvement in each component decreases when t increases.

Given the difficulty of analyzing the sequences $\{\mu_i^t\}_t$ and $\{\sigma_i^t\}_t$ with $t \in \mathbb{N}$, we are going to study a particular case, in which the function to optimize is

$$L_1(\mathbf{x}) = \sum_{i=1}^n x_i, \tag{51}$$

and the sequence of standard deviations meets the condition:

$$\sigma_i^t = \sigma^t, \quad \text{with } i = 1, \ldots, n. \tag{52}$$

First of all we study the sequence of standard deviations $\{\sigma^t\}_t$ with $t \in \mathbb{N}$. Given that

$$\sigma^{t+1} = \sigma^t \left(\frac{n\pi - 1}{n\pi} \right)^{1/2}, \tag{53}$$

we can write σ^{t+1} as a function of σ^0. Therefore, solving the recurrence, the sequence of standard deviations can be written as follows:

$$\sigma^{t+1} = \sigma^0 \left(\frac{n\pi - 1}{n\pi} \right)^{(t+1)/2}. \tag{54}$$

The above expression helps us to analyze the means sequence $\{\mu^t\}_t$ with $t \in \mathbb{N}$. After substituting this expression in Eq. (41) we obtain:

$$\mu^t = \mu^{t-1} - \frac{\sigma^0}{\sqrt{n\pi}} \left(\frac{n\pi - 1}{n\pi} \right)^{(t-1)/2}. \tag{55}$$

We can also express μ^t in terms of μ^0 and σ^0:

$$\mu^t = \mu^0 - \frac{\sigma^0}{\sqrt{n\pi}} \left(1 + \left(\frac{n\pi - 1}{n\pi} \right)^{1/2} + \cdots + \left(\frac{n\pi - 1}{n\pi} \right)^{(t-1)/2} \right)$$

which yields

$$\mu^t = \mu^0 - \frac{\sigma^0}{\sqrt{n\pi}} \cdot \frac{\left(\frac{n\pi-1}{n\pi} \right)^{t/2} - 1}{\left(\frac{n\pi-1}{n\pi} \right)^{1/2} - 1}. \tag{56}$$

This new form of writing μ^t makes it easier to analyze the means sequence. This sequence decreases and has a finite limit.

$$\begin{aligned} \lim_{t\to\infty} \mu^t &= \lim_{t\to\infty} \left(\mu^0 - \frac{\sigma^0}{\sqrt{n\pi}} \cdot \frac{\left(\frac{n\pi-1}{n\pi} \right)^{t/2} - 1}{\left(\frac{n\pi-1}{n\pi} \right)^{1/2} - 1} \right) \\ &= \mu^0 + \frac{\sigma^0}{\sqrt{n\pi}} \frac{1}{\left(\frac{n\pi-1}{n\pi} \right)^{1/2} - 1} = \mu^0 + \frac{\sigma^0}{\sqrt{n\pi - 1} - \sqrt{n\pi}}. \end{aligned} \tag{57}$$

Therefore, although the mean values decrease at each step, this decrease is not unbounded. This fact implies poor algorithm performance, leading us to conclude that this algorithm does not work as expected when we are far from the optimum and the number of tournaments at each step is infinite.

To see how the algorithm performs with a finite number of tournaments, we carried out a number of experiments. Having chosen the number of tournaments, we ran the algorithm with the linear function:

$$L_1(\mathbf{x}) = \sum_{i=1}^{2} x_i, \quad \text{with } \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2. \tag{58}$$

The initial density functions used were

$$f_{X_i^0}(x_i) = f_{\mathcal{N}(1,2)}(x_i), \quad \text{with } i = 1, 2. \tag{59}$$

We ran the algorithm 50 times for different number of tournaments (10, 50, 100, 1000, 10,000). After which we calculated the average value of the mean at each generation. Here we only show the results for the mean values in the first component (the values for the second are analogous). The results can be seen in Fig. 2.

The experiments show that a low number of tournaments does not guarantee an unbounded decrease in the mean values. In fact, as can be seen in

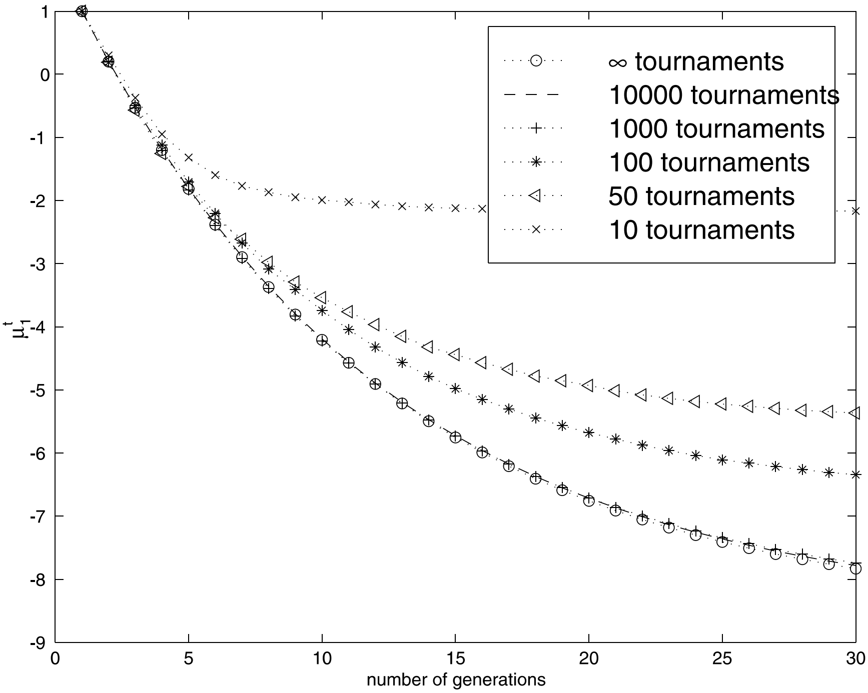

Fig. 2. Values of μ_1^t for different numbers of tournaments.

Fig. 2, the mean values block for a low number of generations when the number of tournaments is small.

The experiment shows that the algorithm performs worse the smaller the number of tournaments made.

5. Quadratic function

This section deals with our analysis of the case in which the function considered is

$$\begin{aligned} Q : \mathbb{R}^n &\to \mathbb{R} \\ \mathbf{x} &\mapsto \sum_{i=1}^{n} x_i^2. \end{aligned} \tag{60}$$

This function is used in the literature to study the algorithm's behaviour near the optimum.

We attempted to make a similar analysis to the one made in the linear case. However, during the study of this function, problems arose in calculating some integrals. These problems forced us to make certain simplifications in order to obtain as much information as possible concerning the algorithm's behaviour.

5.1. Calculation of μ_i^{t+1}

As in the previous case we make the calculation for the first component:

$$\mu_1^{t+1} = \int_{-\infty}^{\infty} 2x_1 f_1(x_1) h_1(x_1)\,\mathrm{d}x_1. \tag{61}$$

First we need to know the value of $A(Q(\mathbf{x}))$

$$A(Q(\mathbf{x})) = A\left(\sum_{j=1}^{n} x_j^2\right) = P\left(\sum_{j=1}^{n} (X_j)^2 \geqslant \sum_{j=1}^{n} x_j^2\right). \tag{62}$$

Since each X_i is a random variable with density function $f_{\mathcal{N}(\mu_i,\sigma_i)}(x_i)$, we know that

$$\begin{aligned} A(Q(\mathbf{x})) &= A\left(\sum_{j=1}^{n} x_j^2\right) \\ &= \int \cdots \int_{\mathscr{D}} \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sigma_1 \ldots \sigma_n} \\ &\quad \cdot \mathrm{e}^{-(u_1-\mu_1)^2/2\sigma_1^2} \cdots \mathrm{e}^{-(u_n-\mu_n)^2/2\sigma_n^2}\,\mathrm{d}u_1 \ldots \mathrm{d}u_n. \end{aligned}$$

where $\mathscr{D} = \{u_1^2 + \cdots + u_n^2 \geqslant x_1^2 + \cdots + x_n^2\}$.

Here we encounter the first problem: to solve the above integral. To do so we make the following simplification: assuming that each X_i is a random variable distributed as a normal with mean $\mu_i = 0$ and deviation $\sigma_i = \sigma$, in other words with density function $f_{\mathcal{N}(0,\sigma)}(x_i)$. This is the only case where we find the above integrals solvable.

Using these kind of density functions implies studying the algorithm's behaviour when the density functions are centered on the optimum, so we are really very near the optimum. But here we wonder: is the optimum actually reached? If the answer is yes, what is the speed of convergence as the dimension increases?

In order to answer these questions we carry out the following analysis:

1. To ensure that at each step we do not move away from the optimum, we have to demonstrate that

$$\mu_i^0 = 0 \Rightarrow \mu_i^t = 0 \quad \forall t. \tag{63}$$

2. To prove that the optimum is reached we have to see whether

$$\sigma^t \to 0 \quad \text{as } t \to \infty. \tag{64}$$

3. We study the speed of convergence as dimension increases. This study allows us to compare the difficulty in approaching the optimum as dimension increases.

5.2. Calculation of μ^{t+1} and σ^{t+1}

First of all, as in the linear case, in order to calculate μ^{t+1} and σ^{t+1} we need to find the expression of $A^t(Q(\mathbf{x}))$:

$$A^t(Q(\mathbf{x})) = A^t\left(\sum_{j=1}^{n} x_j^2\right) = \int \cdots \int_{\mathscr{D}} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \mathrm{e}^{-(u_1^2+\cdots+u_n^2)/2\sigma^2} \, \mathrm{d}u_1 \ldots \mathrm{d}u_n,$$

where $\mathscr{D} = \{u_1^2 + \cdots + u_n^2 \geqslant x_1^2 + \cdots + x_n^2\}$. Taking into account the change of variable $u_i/\sigma = t_i$, with $i = 1, \ldots, n$, we obtain:

$$A^t(Q(\mathbf{x})) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \int \cdots \int_{\mathscr{D}^*} \mathrm{e}^{-(t_1^2+\cdots+t_n^2)/2} \, \mathrm{d}t_1 \ldots \mathrm{d}t_n,$$

where $\mathscr{D}^* = \{t_1^2 + \cdots + t_n^2 \geqslant Q(\mathbf{x})/\sigma^2\}$. This integral can be solved thanks to the generalization to n dimensions of the spherical change of variable in dimension three. This change can be seen in detail in the Appendix A. Here we give the essential information:

- The variables $(t_1, \ldots, t_n)$ are changed to the variables $(\rho, \alpha_{n-1}, \alpha_{n-2}, \ldots, \alpha_2, \alpha_1)$.
- The range of variation of each new variable is

$$0 \leqslant \rho \leqslant \infty, \tag{65}$$

$$-\pi/2 \leqslant \alpha_{n-1}, \alpha_{n-2}, \ldots, \alpha_2 \leqslant \pi/2, \tag{66}$$

$$0 \leqslant \alpha_1 \leqslant 2\pi. \tag{67}$$

- The Jacobian of the transformation is

$$|J_n| = \rho^{n-1} \cdot \cos\alpha_2 \cdot \cos^2\alpha_3 \cdots \cos^{n-3}\alpha_{n-2} \cdot \cos^{n-2}\alpha_{n-1}. \tag{68}$$

Using this change the integral is modified to

$$\begin{aligned} A^t(Q(\mathbf{x})) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \int_0^{2\pi} \mathrm{d}\alpha_1 \int_{-\pi/2}^{\pi/2} \cos\alpha_2 \,\mathrm{d}\alpha_2 \int_{-\pi/2}^{\pi/2} \cos^2\alpha_3 \,\mathrm{d}\alpha_3 \cdots \\ &\quad \times \int_{-\pi/2}^{\pi/2} \cos^{n-3}\alpha_{n-2} \,\mathrm{d}\alpha_{n-2} \int_{-\pi/2}^{\pi/2} \cos^{n-2}\alpha_{n-1} \,\mathrm{d}\alpha_{n-1} \\ &\quad \times \int_{\sqrt{Q(\mathbf{x})}/\sigma}^{\infty} \rho^{n-1} \mathrm{e}^{-\rho^2/2} \,\mathrm{d}\rho \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \underbrace{2\pi \prod_{i=1}^{n-2} \int_{-\pi/2}^{\pi/2} \cos^i \beta \,\mathrm{d}\beta}_{S(n)} \int_{\sqrt{Q(\mathbf{x})}/\sigma}^{\infty} \rho^{n-1} \mathrm{e}^{-\rho^2/2} \,\mathrm{d}\rho, \end{aligned} \tag{69}$$

where $S(n) = 2\pi(\sqrt{\pi})^{n-2}/\Gamma(n/2)$ is the constant associated with the spherical change of variable in n dimensions (see the Appendix A). Therefore, substituting the value of $S(n)$ above

$$A^t(Q(\mathbf{x})) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{2\pi(\sqrt{\pi})^{n-2}}{\Gamma\left(\frac{n}{2}\right)} \int_{\sqrt{Q(\mathbf{x})}/\sigma}^{\infty} \rho^{n-1} \mathrm{e}^{-\rho^2/2} \,\mathrm{d}\rho. \tag{70}$$

Let I_n denote indefinite integral $\int \rho^{n-1} \mathrm{e}^{-\rho^2/2} \,\mathrm{d}\rho$, and $I_n(u, v)$ denote definite integral $I_n|_u^v$. The integral $I_n(\sqrt{Q(\mathbf{x})}/\sigma, \infty)$ has different values when n is odd or even. When n is odd $I_1(\sqrt{Q(\mathbf{x})}/\sigma, \infty)$ is an incomplete Gamma function (it has no explicit expression), meaning that from here we only work with even dimension. We emphasize this fact writing $A^t_{2n}(Q(\mathbf{x}))$. Therefore

$$\begin{aligned} A^t_{2n}(Q(\mathbf{x})) &= \left(\frac{1}{\sqrt{2\pi}}\right)^{2n} \frac{2\pi(\sqrt{\pi})^{2n-2}}{(n-1)!} \int_{\sqrt{Q(\mathbf{x})}/\sigma}^{\infty} \rho^{2n-1} \mathrm{e}^{-\rho^2/2} \,\mathrm{d}\rho \\ &= \frac{1}{2^{n-1}(n-1)!} \cdot I_{2n}(\sqrt{Q(\mathbf{x})}/\sigma, \infty). \end{aligned} \tag{71}$$

In order to solve integral I_{2n} we write:

$$u = \rho^{2n-2} \Rightarrow \mathrm{d}u = (2n-2)\rho^{2n-3} \,\mathrm{d}\rho, \tag{72}$$

$$\mathrm{d}v = \rho \mathrm{e}^{-\rho^2/2} \,\mathrm{d}\rho \Rightarrow v = -\mathrm{e}^{-\rho^2/2}, \tag{73}$$

so that

$$\begin{aligned} I_{2n} &= -\rho^{2n-2} \mathrm{e}^{-\rho^2/2} + (2n-2) \int \rho^{2n-3} \mathrm{e}^{-\rho^2/2} \,\mathrm{d}\rho \\ &= -\rho^{2n-2} \mathrm{e}^{-\rho^2/2} + (2n-2) \cdot I_{2n-2}. \end{aligned} \tag{74}$$

substituting the expressions of I_{2n-j}, with $j = 2, \ldots, 2n-2$:

$$\begin{aligned}
I_{2n} &= -\rho^{2n-2}\mathrm{e}^{-\rho^2/2} - (2n-2)\rho^{2n-4}\mathrm{e}^{-\rho^2/2} - (2n-2)(2n-4) \\
&\quad \cdot \rho^{2n-6}\mathrm{e}^{-\rho^2/2} - \cdots - (2n-2)(2n-4)\cdots 2\cdot 1\mathrm{e}^{-\rho^2/2} \\
&= -\mathrm{e}^{-\rho^2/2}\left[\rho^{2n-2} + \sum_{j=2}^{n}\rho^{2n-2j}(2n-2)(2n-4)\cdots(2n-2(j-1))\right] \\
&= -\mathrm{e}^{-\rho^2/2}\left[\rho^{2n-2} + \sum_{j=2}^{n}\rho^{2(n-j)}\cdot 2^{j-1}\cdot\frac{n!}{n(n-j)!}\right]. \qquad (75)
\end{aligned}$$

Taking into account equation (75), integral $A^t_{2n}(Q(\mathbf{x}))$ can be expressed as follows:

$$\begin{aligned}
&A^t_{2n}(Q(\mathbf{x})) \\
&\quad = \frac{1}{2^{n-1}(n-1)!}\cdot\left[-\mathrm{e}^{-\rho^2/2}\left(\rho^{2n-2} + \sum_{j=2}^{n}\rho^{2(n-j)}\cdot 2^{j-1}\cdot\frac{n!}{n(n-j)!}\right)\right]\Bigg|_{\sqrt{Q(\mathbf{x})}/\sigma}^{\infty} \\
&\quad = \frac{2^{1-n}}{(n-1)!}\cdot\mathrm{e}^{-Q(\mathbf{x})/2\sigma^2}\left[\left(\frac{Q(\mathbf{x})}{\sigma^2}\right)^{n-1} + \sum_{j=2}^{n}\left(\frac{Q(\mathbf{x})}{\sigma^2}\right)^{n-j}\cdot 2^{j-1}\cdot\frac{n!}{n(n-j)!}\right] \\
&\quad = \frac{2^{1-n}}{(n-1)!}\cdot\frac{\mathrm{e}^{-Q(\mathbf{x})/2\sigma^2}}{\sigma^{2(n-1)}}\left[(Q(\mathbf{x}))^{n-1} + \sum_{j=2}^{n}(Q(\mathbf{x}))^{n-j}\cdot\sigma^{2(j-1)}\cdot 2^{j-1}\cdot\frac{n!}{n(n-j)!}\right].
\end{aligned} \qquad (76)$$

The next step is to calculate integral $h_1(x_1)$:

$$\begin{aligned}
h_1(x_1) &= \int\cdots\int_{\mathbb{R}^{2n-1}}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{2n-1}\mathrm{e}^{-(x_2^2+\cdots+x_{2n}^2)/2\sigma^2}A^t_{2n}(Q(\mathbf{x}))\,\mathrm{d}x_2\ldots\mathrm{d}x_{2n} \\
&= \int\cdots\int_{\mathbb{R}^{2n-1}}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{2n-1}\mathrm{e}^{-(x_2^2+\cdots+x_{2n}^2)/2\sigma^2}\cdot\frac{2^{1-n}}{(n-1)!} \\
&\quad \cdot\frac{\mathrm{e}^{-(x_1^2+x_2^2+\cdots+x_{2n}^2)/2\sigma^2}}{\sigma^{2(n-1)}}\cdot\left[(Q(\mathbf{x}))^{n-1} + \sum_{j=2}^{n}(Q(\mathbf{x}))^{n-j}\cdot\sigma^{2(j-1)}\cdot 2^{j-1}\right. \\
&\quad \left.\cdot\frac{n!}{n(n-j)!}\right]\mathrm{d}x_2\ldots\mathrm{d}x_{2n}.
\end{aligned}$$

Changing the variable, $x_i/\sigma = t_i$, and taking into account that $(Q(\mathbf{x}))^k = (x_1^2 + \sigma^2(t_2^2 + \cdots + t_{2n}^2))^k$, the expression of $h_1(x_1)$ can be written as

$$h_1(x_1) = \left(\frac{1}{\sqrt{2\pi}}\right)^{2n-1} \cdot \frac{2^{1-n}}{\sigma^{2(n-1)}(n-1)!} \int \cdots \int_{\mathbb{R}^{2n-1}} \mathrm{e}^{-(t_2^2+\cdots+t_{2n}^2)} \cdot \mathrm{e}^{-x_1^2/2\sigma^2}$$
$$\cdot \left[\left(x_1^2 + \sigma^2(t_2^2 + \cdots + t_{2n}^2)\right)^{n-1} + \sum_{j=2}^{n} \left(x_1^2 + \sigma^2(t_2^2 + \cdots + t_{2n}^2)\right)^{n-j} \right.$$
$$\left. \cdot\ \sigma^{2(j-1)} \cdot 2^{j-1} \cdot \frac{n!}{n(n-j!)} \right] \mathrm{d}t_2 \ldots \mathrm{d}t_{2n}.$$

Making again the generalization to $2n-1$ dimensions of the spherical change of variable in dimension three:

$$h_1(x_1) = \left(\frac{1}{\sqrt{2\pi}}\right)^{2n-1} \cdot \frac{2^{1-n}}{\sigma^{2(n-1)}(n-1)!} \frac{2\pi(\sqrt{\pi})^{2n-3}}{\Gamma\left(\frac{2n-1}{2}\right)}$$
$$\cdot \int_0^{\infty} \rho^{2n-2} \mathrm{e}^{-\rho^2} \mathrm{e}^{-x_1^2/2\sigma^2} \left[\left(x_1^2 + \sigma^2\rho^2\right)^{n-1} \right.$$
$$\left. + \sum_{j=2}^{n} \left(x_1^2 + \sigma^2\rho^2\right)^{n-j} \cdot \sigma^{2(j-1)} \cdot 2^{j-1} \cdot \frac{n!}{n(n-j)!} \right] \mathrm{d}\rho$$
$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{2n-1} \cdot \frac{2^{1-n}}{\sigma^{2(n-1)}(n-1)!} \frac{2\pi(\sqrt{\pi})^{2n-3}}{\Gamma\left(\frac{2n-1}{2}\right)}$$
$$\cdot\ \mathrm{e}^{-x_1^2/2\sigma^2} \left[\int_0^{\infty} \rho^{2n-2} \mathrm{e}^{-\rho^2} \left(x_1^2 + \sigma^2\rho^2\right)^{n-1} \right.$$
$$\left. + \sum_{j=2}^{n} \sigma^{2(j-1)} \cdot 2^{j-1} \cdot \frac{n!}{n(n-j)!} \int_0^{\infty} \rho^{2n-2} \mathrm{e}^{-\rho^2} \left(x_1^2 + \sigma^2\rho^2\right)^{n-j} \mathrm{d}\rho \right]. \tag{77}$$

Unfortunately to compute integrals $\int_0^{\infty} \rho^{2n-2} \mathrm{e}^{-\rho^2} (x_1^2 + \sigma^2\rho^2)^{n-j}\, \mathrm{d}\rho$, with $j = 1, \ldots, 2n$ is not an easy task, hence to find an explicit general expression for $h_1(x_1)$ is difficult. Therefore we have solved this integral for some finite cases (by parts). It allows us to give a general idea concerning the algorithm's performance as the problem dimension increases.

5.3. Analyzing the algorithm's behaviour

As explained above we solve $h_1(x_1)$ in some finite cases. Although our analysis will not be so general as in the linear case, however we do obtain some information about the algorithm's behaviour near the optimum.

The finite cases are

$$2n = 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 40,\\ 50, 60, 70, 80, 90, 100, 150, 200, 300, 400, 500, 600. \tag{78}$$

After obtaining the values for $h_1(x_1)$, for the above cases, we substitute them in the expression of μ^{t+1}, obtaining $\mu^{t+1} = 0$ for every t. The results of substituting these values in the expression of σ^{t+1} are summarized in Table 1.

Table 1
Values of σ^{t+1} for some finite cases

$2n$	σ^{t+1}
2	0.7071 σ^t
4	0.7906 σ^t
6	0.8291 σ^t
8	0.8524 σ^t
10	0.8683 σ^t
12	0.8800 σ^t
14	0.8891 σ^t
16	0.8964 σ^t
18	0.9025 σ^t
20	0.9076 σ^t
22	0.9120 σ^t
24	0.9159 σ^t
26	0.9192 σ^t
28	0.9223 σ^t
30	0.9249 σ^t
40	0.9352 σ^t
50	0.9422 σ^t
60	0.9473 σ^t
70	0.9513 σ^t
80	0.9545 σ^t
90	0.9571 σ^t
100	0.9594 σ^t
150	0.9669 σ^t
200	0.9714 σ^t
300	0.9767 σ^t
400	0.9799 σ^t
500	0.9820 σ^t
600	0.9836 σ^t

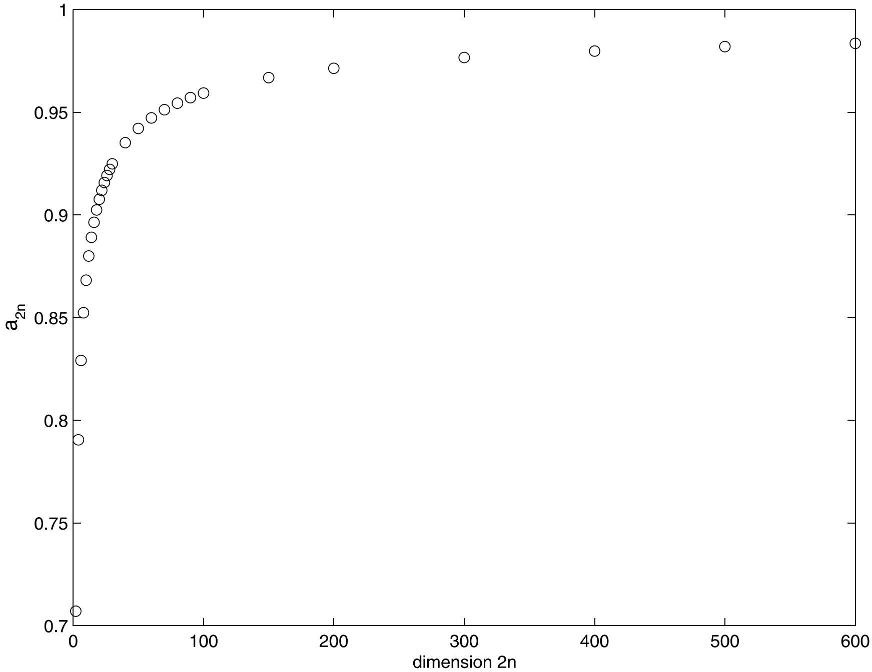

Fig. 3. Factor of decrease of σ^{t+1}.

As can be seen in Table 1, for these finite cases we can write $\sigma^{t+1} = a_{2n}\sigma^t$. The factor of decrease a_{2n} is represented in Fig. 3.

Having the $\{a_{2n}\}$ data we consider convenient to find a formula that approximates it, in other words, to "fit" a curve through the points in $\{a_{2n}\}$ data. It allows us to estimate the speed of convergence. We find the following least-squares fit to data:

$$g(n) = 1 - \frac{0.4}{\sqrt{n}}. \tag{79}$$

Fig. 4 shows that $g(n)$ fit properly the $\{a_{2n}\}$ data.

The results indicate that

1. The value of $\sigma^t \to 0$ as $t \to \infty$ in the analyzed dimensions, therefore the algorithm reaches the optimum.
2. Due to $g(n)$ seems to fit properly the points in $\{a_{2n}\}$ data, the speed of convergence decreases with the dimension as $O(1/\sqrt{n})$.

Therefore we can conclude that in the finite cases studied, the algorithm reaches the optimum, but the speed of convergence decreases as the dimension of the problem increases.

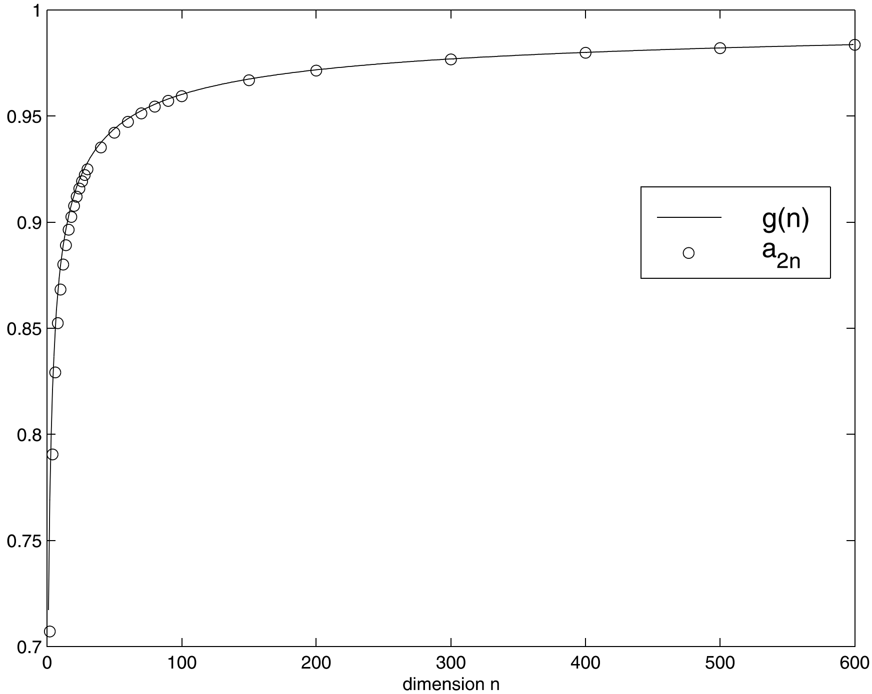

Fig. 4. Fitting $\{a_{2n}\}$ values.

6. Conclusions

This work is one of the few that deal with mathematical modelling of EDAs. We have modelled the $\mathrm{UMDA_c}$ algorithm with tournament selection applied to linear and quadratic functions when an infinite number of tournaments is performed.

Based on this modelling we have analyzed its behaviour in n-dimensional linear functions and in an n-dimensional quadratic function. In the case of linear functions we conclude that the algorithm does not work correctly in linear function $L_1(\mathbf{x}) = \sum_{i=1}^{n} x_i$, with $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. After doing certain assumptions in the case of quadratic function $Q(\mathbf{x}) = \sum_{i=1}^{n} x_i^2$, we have proved for some finite dimensions that the algorithm reaches the optimum. Moreover, the speed of convergence is slower when the dimension increases.

The obtained results are closely related to the distributions chosen (unidimensional normals). It will be helpful to study the behaviour of the algorithm when other distributions are used.

Now our main objective is to arrive at an analogous model and analysis for the UMDA algorithm in the discrete case.

Acknowledgements

We want to thank Professor J. de la Cal for his help on this project.

This work was supported by the University of the Basque Country and by the Ministry of Science and Technology under grants 9/UPV/EHU 00140.226-12084/2000 and TIC2001-2973-C05-03 respectively. Also C. González has a grant from the UPV-EHU.

Appendix A

This section explains in detail the generalization to n dimensions of the spherical change of variable in dimension three.

The first step is to solve the problem for $n = 4$. In spherical coordinates the position of a point $P(x_1, x_2, x_3, x_4)$ in the space is determined by four numbers ρ, α_1, α_2, α_3, where

- ρ is the distance from point P to the origin.
- α_3 is the angle formed by the vector $\overline{OP}$ and its projection (denoted by $\bar{r}_1$) upon the plane $\overline{OX_1X_2X_3}$.
- α_2 is the angle formed by the projection of $\bar{r}_1$ (denoted by $\bar{r}_2$) upon the plain $\overline{OX_1X_2}$.
- α_1 is the angle formed by axis X_1 and $\bar{r}_2$.

Taking these facts into account, we can write the old coordinates depending on the new ones:

$$x_4 = \rho \sin \alpha_3, \tag{A.1}$$

$$x_3 = \rho \cos \alpha_3 \sin \alpha_2, \tag{A.2}$$

$$x_2 = \rho \cos \alpha_3 \cos \alpha_2 \sin \alpha_1, \tag{A.3}$$

$$x_1 = \rho \cos \alpha_3 \cos \alpha_2 \cos \alpha_1. \tag{A.4}$$

For any point $P(x_1, x_2, x_3, x_4) \in \mathbb{R}^4$, each new variable varies in

$$0 \leqslant \rho \leqslant \infty, \tag{A.5}$$

$$-\pi/2 \leqslant \alpha_2, \alpha_3 \leqslant \pi/2, \tag{A.6}$$

$$0 \leqslant \alpha_1 \leqslant 2\pi. \tag{A.7}$$

Jacobian of the change J_4 is

$$J_4 = \rho \cos \alpha_2 \cos^2 \alpha_3. \tag{A.8}$$

Therefore after making this change of variable in integral:

$$\int\int\int\int_{\mathbb{R}^4} f(x_1, x_2, x_3, x_4)\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3\,\mathrm{d}x_4, \tag{A.9}$$

we obtain

$$\int_0^{2\pi}\int_{-\pi/2}^{\pi/2}\int_{-\pi/2}^{\pi/2}\int_0^{\infty} f(\rho, \alpha_3, \alpha_2, \alpha_1)\cdot\rho\cos\alpha_2\cos^2\alpha_3\,\mathrm{d}\alpha_1\,\mathrm{d}\alpha_2\,\mathrm{d}\alpha_3\,\mathrm{d}\rho. \tag{A.10}$$

Taking into account the above arguments, we can say that using the generalization to n dimensions of the spherical change of variable in dimension four, integral

$$I_n^S = \int\cdots\int_{\mathbb{R}^n} f(x_1, \ldots, x_n)\,\mathrm{d}x_1\ldots\mathrm{d}x_n, \tag{A.11}$$

changes to

$$\begin{aligned} I_n^S = {} & \int_0^{2\pi}\int_{-\pi/2}^{\pi/2}\cdots\int_{-\pi/2}^{\pi/2}\int_0^{\infty} f(\rho, \alpha_{n-1}, \ldots, \alpha_1)\rho^{n-1}\cos\alpha_2\cdot\cos^2\alpha_3 \\ & \cdot\cos^{n-3}\alpha_{n-2}\cdots\cos^{n-2}\alpha_{n-1}\,\mathrm{d}\alpha_1\,\mathrm{d}\alpha_2\,\mathrm{d}\alpha_3\ldots\mathrm{d}\alpha_{n-2}\,\mathrm{d}\alpha_{n-3}\,\mathrm{d}\alpha_{n-1}\,\mathrm{d}\rho. \end{aligned}$$

In the integrals we have to solve in this article using this change of variable, function f only depends on ρ. Therefore, the above integral can be written as follows:

$$\begin{aligned} I_n^S = {} & \int_0^{2\pi}\mathrm{d}\alpha_1\int_{-\pi/2}^{\pi/2}\cos\alpha_2\,\mathrm{d}\alpha_2\int_{-\pi/2}^{\pi/2}\cos^2\alpha_3\,\mathrm{d}\alpha_3\cdots \\ & \times\int_{-\pi/2}^{\pi/2}\cos^{n-3}\alpha_{n-2}\,\mathrm{d}\alpha_{n-2}\int_{-\pi/2}^{\pi/2}\cos^{n-2}\alpha_{n-1}\,\mathrm{d}\alpha_{n-1}\int_0^{\infty}\rho^{n-1}f(\rho)\,\mathrm{d}\rho \\ = {} & 2\pi\prod_{i=1}^{n-2}\int_{-\pi/2}^{\pi/2}\cos^i\beta\,\mathrm{d}\beta\int_0^{\infty}\rho^{n-1}f(\rho)\,\mathrm{d}\rho. \end{aligned} \tag{A.12}$$

To solve this integral it will be useful to calculate the constant:

$$S(n) = 2\pi\prod_{i=1}^{n-2}\int_{-\pi/2}^{\pi/2}\cos^i\beta\,\mathrm{d}\beta = 2\pi\prod_{i=1}^{n-2}\sqrt{\pi}\frac{\Gamma\left(\frac{i+1}{2}\right)}{\Gamma\left(1+\frac{i}{2}\right)}, \tag{A.13}$$

where Γ is the Gamma function. Simplifying

$$S(n) = \frac{2\pi(\sqrt{\pi})^{n-2}}{\Gamma\left(\frac{n}{2}\right)}. \tag{A.14}$$

References

[1] A. Berny, An adaptative scheme for real function optimization acting as a selection operator, in: X. Yao, (Ed.), Proceedings of the First IEEE Symposium on Combinations of Evolutionary Computation and Neural Networks, 2000.

[2] A. Berny, Selection and reinforcement learning for combinatorial optimization, in: M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J.J. Merelo, H.-P. Schwefel (Eds.), Proceedings of Parallel Problem Solving from Nature, PPSN-VI, Springer-Verlag, Berlin, Heidelberg, 2000, pp. 601–610.

[3] H.G. Beyer, The Theory of Evolution Strategies, Springer, Berlin, 2001.

[4] C. González, J.A. Lozano, P. Larrañaga, Analyzing the population based incremental learning algorithm by means of discrete dynamical systems, Complex Syst. 12 (4) (2000) 465–479.

[5] C. González, J.A. Lozano, P. Larrañaga, Mathematical modelling of discrete estimation of distribution algorithms, in: P. Larrañaga, J.A. Lozano (Eds.), Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation, Kluwer Academic Publishers, Dordrecht, 2001, pp. 147–163.

[6] C. González, J.A. Lozano, P. Larrañaga, The convergence behavior of the PBIL algorithm: a preliminary approach, in: V. Kurková, N.C. Steele, R. Neruda, M. Kárný (Eds.), Proceedings of the Fifth International Conference on Artificial Neural Networks and Genetic Algorithms, ICANNGA'2001, 2001, pp. 228–231.

[7] M. Höhfeld, G. Rudolph, Towards a theory of population-based incremental learning, in: Proceedings of the Fourth IEEE Conference on Evolutionary Computation, IEEE Press, NY, 1997, pp. 1–5.

[8] P. Larrañaga, R. Etxeberria, J.A. Lozano, J.M. Peña, Optimization by learning and simulation of Bayesian and Gaussian networks, Technical Report KZZA-IK-4-99, Department of Computer Science and Artificial Intelligence, University of the Basque Country, 1999.

[9] P. Larrañaga, R. Etxeberria, J.A. Lozano, J.M. Peña, Combinatorial optimization by learning and simulation of Bayesian networks, in: C. Boutilier, M. Goldszmidt (Eds.), Proceedings of Uncertainty in Artificial Intelligence, UAI-2000, Morgan Kaufmann, 2000, pp. 343–352.

[10] P. Larrañaga, R. Etxeberria, J.A. Lozano, J.M. Peña, Optimization in continuous domains by learning and simulation of Gaussian networks, in: A.S. Wu (Ed.), Proceedings of Genetic and Evolutionary Computation Conference, Workshop Program, 2000, pp. 201–204.

[11] P. Larrañaga, J.A. Lozano, Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation, Kluwer Academic Publishers, Kluwer Academic Publishers, Dordrecht, 2001.

[12] H. Mühlenbein, The equation for response to selection and its use for prediction, Evol. Comput. 5 (1998) 303–346.

[13] H. Mühlenbein, T. Mahnig, Application of Wright's equation in evolutionary computation, in: Proceedings of Advances in Fuzzy Systems and Evolutionary Computation, WSES, 2001, pp. 265–270.

[14] H. Mühlenbein, T. Mahnig, A. Ochoa, Schemata, distributions and graphical models in evolutionary optimization, J. Heuristics 5 (1999) 215–247.

[15] H. Mühlenbein, G. Paaß, From recombination of genes to the estimation of distributions I. Binary parameters, in: H.-M. Voigt, W. Ebeling, I. Rechenberg, H.-P. Schwefel (Eds.), Parallel Problem Solving from Nature, PPSN-IV, 1996, pp. 178–187.

[16] M. Pelikan, D.E. Goldberg, F. Lobo, A survey of optimization by building and using probabilistic models. Technical Report IlliGAL, 99018, Illinois Genetic Algorithms Laboratory, 1999.

[17] Q. Zhang, H. Mühlenbein, On global convergence of FDA with proporcionate selection, in: Proceedings of Second Symposium on Artificial Intelligence, CIMAF'99, 1999, pp. 340–343.