Lower Bounds for Evolution Strategies Using VC-Dimension

Olivier Teytaud[1] and Hervé Fournier[2]

[1] TAO (Inria), LRI, UMR 8623 (CNRS - Univ. Paris-Sud), Bât 490
Univ. Paris-Sud 91405 Orsay, France
teytaud@lri.fr
[2] Laboratoire PRiSM
CNRS UMR 8144 and Univ. Versailles St-Quentin en Yvelines
45 av. des États-Unis, 78035 Versailles, France
herve.fournier@prism.uvsq.fr

Abstract. We derive lower bounds for comparison-based or selection-based algorithms, improving existing results in the continuous setting, and extending them to non-trivial results in the discrete case. This is achieved by considering the VC-dimension of the level sets of the fitness functions; results are then obtained through the use of Sauer's lemma. In the special case of optimization of the sphere function, improved lower bounds are obtained by bounding the possible number of sign conditions realized by some systems of equations.

Keywords: Evolution Strategies, Convergence ratio, VC-dimension, Sign conditions.

1 Introduction

Evolution strategies (ES), defined by Rechenberg [15], are a family of optimization algorithms with nice robustness properties. Most ES use only comparisons between fitness values and not the fitness values themselves. This fact has been used in [18] in order to provide lower bounds that match some upper bounds known for evolutionary algorithms [8,2,16]. The optimality of this comparison-based principle for some robustness criterion was shown in [10] (see also [3,20,4]). In [18] is provided a new tool for proving lower bounds for evolutionary algorithms, but, as pointed out by the authors, some bounds are not tight and in particular: (i) the discrete case provides essentially trivial results; (ii) the bounds for the (μ, λ)-ES are far too large. In this work, we propose improved lower bounds for evolution strategies of type $(\mu \dot{+} \lambda)$-ES (i.e. upper bounds on the convergence ratios of these algorithms) in terms of the VC-dimension of level sets of the fitness functions. In the special case of optimization of the sphere function, improved upper bound on the convergence ratio of evolution strategies are presented; they are obtained by bounding the number of sign conditions realized by a system of equations. The paper is organized as follows. Basic definitions and terminology of evolution strategies we consider are described in Section 2.

G. Rudolph et al. (Eds.): PPSN X, LNCS 5199, pp. 102–111, 2008.

Lower bounds on $(\mu \dot{+} \lambda)$-ES based on the branching factor, obtained in [18], are recalled in Section 3. Improved lower bounds on $(\mu \dot{+} \lambda)$-ES in terms of the VC-dimension are presented in Section 4. At last, some questions are raised in Section 5.

Notations. In all the paper, $\log(x)$ denotes the logarithm with basis 2, i.e. $\log(2) = 1$. The set of integers $\{1, 2, \ldots, n\}$ is denoted by $[[1, n]]$.

2 Evolution Strategies of Type $(\mu \dot{+} \lambda)$

We define in this section $(\mu \dot{+} \lambda)$-algorithms – we refer to Beyer and Schwefel [6] for a comprehensive introduction to evolution strategies. The aim of a $(\mu \dot{+} \lambda)$-algorithm is to find the minimum of a function f (called the fitness function) defined over a domain D. This algorithm cannot evaluate the function f but has to work only with comparisons: given two points x and y, the algorithm has access to a black-box telling whether $f(x) < f(y)$, $f(x) = f(y)$ or $f(x) > f(y)$. Of course such an algorithm is not required to work for one fitness function but for a whole family of fitness functions. In the following we denote by $\mathcal{F}$ the set of fitness functions we consider. In the rest of the paper, we assume we never have a case of equality $f(x) = f(y)$ among the generated points. Let λ

Algorithm 1. SB-(μ, λ)-ES (resp. SB-$(\mu+\lambda)$-ES), i.e. evolution strategies based on selection, working on a fitness function f. The real number ω is a random seed, uniform in $[0, 1]$. We do not specify the generation of the offspring, because we work on the whole family of algorithms matching this framework.

Initialize $I_0 \in \mathcal{I}$, $S_{-1} = \emptyset$ and $n = 0$
while true **do**
  Generate an offspring O_n of λ distinct points: $O_n = \text{generate}(I_n, \omega)$.
  Selection: Use the fitness f in order to partition O_n (resp. $O_n \cup S_{n-1}$) in two sets S_n of cardinal $\min(\mu, \text{Card}(O_n))$ and R_n such that

$$x \in S_n \text{ and } y \in R_n \Rightarrow f(x) < f(y).$$

  We denote this by $S_n = \text{select}(O_n, f)$ (resp. $S_n = \text{select}(O_n \cup S_{n-1}, f)$).
  Update the internal state:

$$I_{n+1} = \text{update}(I_n, f, O_n) = \text{selectionUpdate}(I_n, S_n, R_n) \in \mathcal{I}.$$

  $x^{(f)}_{\omega, n+1} = \text{proposal}(I_n)$
  $n = n + 1$
end while

and μ bee two integers (subject to $\mu \leqslant \lambda$ in the (μ, λ) case). A SB-$(\mu \dot{+} \lambda)$-ES (Selection Based $(\mu \dot{+} \lambda)$-ES) is an algorithm working as follows. There is a set $\mathcal{I}$ of internal states and an initial state I_0. At each iteration, the algorithm follows these three successive steps. First generate a set of λ points, called the *offspring*. Then select only the μ best ones, i.e. the μ points with lowest fitness values;

Algorithm 2. (μ, λ)-ES (resp. $(\mu + \lambda)$-ES) based on full ranking, working on a fitness function f. The real number ω is a random seed, uniform in $[0, 1]$. Compared to Algorithm 1, S_n is now a vector of points, ordered with respect to their fitness values. This family of algorithms is more general than Algorithm 1, as we can use all the ranking information.

Initialize $I_0 \in \mathcal{I}$, $S_{-1} = \emptyset$ and $n = 0$
while true **do**
  Generate an offspring O_n of λ distinct points: $O_n = \text{generate}(I_n, \omega)$.
  Selection with ranking: Use the fitness f in order to partition O_n (resp. $O_n \cup S_{n-1}$) in a vector $S_n = (x'_1, \ldots, x'_{c_n})$ of cardinal $c_n = \min(\mu, \text{Card}(O_n))$ (resp. $c_n = \min(\mu, \text{Card}(O_n \cup S_{n-1}))$) and a set R_n such that

$$\forall i \in [[1, c_n]], \forall y \in R_n, f(x'_i) < f(y),$$

$$\text{and } \forall i \in [[1, c_n - 1]], f(x'_i) < f(x'_{i+1}).$$

  We denote this by $S_n = \text{select}(O_n, f)$ (resp. $S_n = \text{select}(O_n \cup S_{n-1}, f)$).
  Update the internal state:

$$I_{n+1} = \text{update}(I_n, f, O_n) = \text{fullRankUpdate}(I_n, S_n, R_n) \in \mathcal{I}.$$

  $x^{(f)}_{\omega,n+1} = \text{proposal}(I_n)$
  $n = n + 1$
end while

in the case of a SB-(μ, λ)-ES, points generated at previous stages are forgotten and this selection is performed only among the offspring, while an algorithm of type SB-$(\mu + \lambda)$-ES selects the μ best points among the offspring *and* the points selected at the previous step (hence these μ selected points are always the μ points with lowest fitness values found so far). At last the internal state is updated. General outlines of SB-(μ, λ)-algorithms (resp. SB-$(\mu + \lambda)$-algorithms) are summarized in Algorithm 1.

Algorithms with the "+" are usually termed *elitist*; this means that we always keep the best individuals. Algorithms with the "," are termed *non-elitist*. Elitist strategies are usually faster on easy fitness functions, but less robust; therefore, non-elitist strategies are usually prefered.

At last we would like to explain a generalization of SB-$(\mu \dagger \lambda)$-ES, called $(\mu \dagger \lambda)$-ES. Instead of just giving the best μ points (i.e. the μ points with the lowest fitness values), we can consider a selection procedure which returns the best μ points *ordered with respect to their fitness*. More precisely, given the points $(y_1, \ldots, y_p)$ (O_n in the case of (μ, λ)-ES or $O_n \cup S_{n-1}$ in the case of $(\mu + \lambda)$-ES), it returns μ distinct indices $(i_1, \ldots, i_\mu)$ such that $f(y_{i_1}) < \ldots < f(y_{i_\mu})$ and for all $j \notin \{i_1, \ldots, i_\mu\}$, $f(y_{i_\mu}) < f(y_j)$. We call *full ranking* this kind of "selection" [4,3,20]. The outline of these algorithms is summarized in Algorithm 2.

Note that both Algorithms 1 and 2 define a class of algorithms: in order to obtain an algorithm, one has to specify how generation of points is done, what is the set of internal states as well as the update function. We assume that all functions involved in these algorithms are measurable. A usual case is

retrieved when the offspring is randomly and independently drawn according to a Gaussian distribution, with parameters (mean, variance and covariances) depending on the internal state of the algorithm.

3 Branching Factor and Convergence Ratio

We consider a (possibly discrete) domain $D \subset \mathbb{R}^d$ and a norm $\|\cdot\|$ on $\mathbb{R}^d$. For $\varepsilon > 0$, we define $N(\varepsilon)$ to be the maximum integer n such that there exist n distinct points $x_1, \ldots, x_n \in D$ with $\|x_i - x_j\| \geqslant 2\varepsilon$ for all $i \neq j$. If each function $f \in \mathcal{F}$ has one and only one optimum f^*, for any given optimization algorithm as in Algorithm 2, and for $\varepsilon > 0$ and $\delta > 0$, we let $n_{\varepsilon,\delta}$ be the minimum number n of iterations such that with probability at least $1 - \delta$, an optimum is found at the n-th iteration within distance ε. I.e. $n_{\varepsilon,\delta}$ is minimal such that for all $n \geqslant n_{\varepsilon,\delta}$ and for all $f \in \mathcal{F}$,

$$\mathbb{P}_{w\in[0,1]}[\|x^{(f)}_{\omega,n} - f^*\| < \varepsilon] \geqslant 1 - \delta.$$

For an algorithm of type $(\mu \dot{+} \lambda)$-ES working over a set $\mathcal{F}$ of fitness functions, we define the *branching factor* of any algorithm as in Algorithm 2 as

$$K = \sup_{I\in\mathcal{I},O} \mathrm{Card}\{\mathrm{update}(I, f, O) \mid f \in \mathcal{F}\}.$$

Notice that in the case of selection based algorithms (any algorithm fitting Algorithm 1), we have

$$K \leqslant \sup_{O} \mathrm{Card}\{\mathrm{select}(O, f) \mid f \in \mathcal{F}\}$$

where the supremum holds for: (i) O any set of λ points in the case of SB-(μ, λ)-ES; (ii) O any set of $\lambda + \mu$ points in the case of SB-$(\mu + \lambda)$-ES. A similar remark holds in the case of full ranking $(\mu \dot{+} \lambda)$-ES, except that a bound on K is given by the possible number of choices of selected points together with their order (with respect to their fitness values). Let us recall the following result from Teytaud and Gelly [18] (restricted here to our purpose) relating the convergence ratio and the branching factor of a $(\mu \dot{+} \lambda)$-ES.

Theorem 1 (Lower bound on the convergence ratio of $(\mu \dot{+} \lambda)$-ES.). *Consider a (μ, λ)-ES or $(\mu + \lambda)$-ES as in Algorithm 2. Consider a set $\mathcal{F}$ of possible fitness functions on domain D, i.e. $\mathcal{F} \subset \mathbb{R}^D$, such that any fitness function $f \in \mathcal{F}$ has only one min-argument f^*, and such that $\{f^* \mid f \in \mathcal{F}\} = D$. Let $\varepsilon > 0$ and $\delta \in]0, 1[$. Let $L_n(\omega)$ be the number of different paths (when the function f runs over $\mathcal{F}$) followed by the algorithm on the random seed ω after n steps of computation; then*

$$\mathbb{E}_{\omega\in[0,1]}[L_{n_{\varepsilon,\delta}}(\omega)] \geqslant (1 - \delta)N(\varepsilon).$$

In particular, if K denotes the branching factor of the algorithm, then

$$n_{\varepsilon,\delta} \geqslant \left\lceil \frac{\log(1-\delta)}{\log(K)} + \frac{\log(N(\varepsilon))}{\log(K)} \right\rceil.$$

We can define the convergence ratio for both discrete and continuous domains thanks to the following unified definitions. We want a definition of convergence ratio which matches bounds of the form $O(1/d)$ established in [13]; therefore, we define the *convergence ratio* of an algorithm for precision ε as

$$\mathrm{CR}_\varepsilon = \frac{\log N(\varepsilon)}{d n_{\varepsilon,\frac{1}{2}}}.$$

We also define the *normalized convergence ratio* (normalized by the number of individuals generated per epoch) by

$$\mathrm{NCR}_\varepsilon = \frac{\log N(\varepsilon)}{d\lambda n_{\varepsilon,\frac{1}{2}}}.$$

The ratio CR_ε is relevant in the parallel setting (i.e. it is the convergence ratio when working on a parallel computer, with parallel evaluation of the offspring), while NCR_ε is relevant in the sequential setting, i.e. when individuals are evaluated sequentially.

Theorem 1 can be reformulated with these unified definitions of convergence ratios as follows. Consider a $(\mu \dot{+} \lambda)$-ES satisfying the hypothesis of Theorem 1. Let $\alpha(\varepsilon) = 1/(1 - 1/N(\varepsilon))$. Then

$$\mathrm{CR}_\varepsilon \leqslant \frac{\log K}{d} \cdot \alpha(\varepsilon) \text{ and } \mathrm{NCR}_\varepsilon \leqslant \frac{\log K}{d\lambda} \cdot \alpha(\varepsilon). \tag{1}$$

4 Sauer's Lemma and VC-Dimension

Teytaud and Gelly [18] applied the bounds obtained in Section 3 in the following way: the number of subsets of size μ of a set of λ points, is at most $\binom{\lambda}{\mu} \leqslant \binom{\lambda}{\lfloor \lambda/2 \rfloor} \leqslant (2^\lambda/\sqrt{2\pi\lambda})$ – see e.g. [7, p587] or [9] for these inequalities. This surely holds, but it is a worst case on possible selections: if the fitness funtions are "nice", many of these subsets cannot be realized. This is precisely quantified by Sauer's lemma in the theory of VC-dimension. In this section, we show how this allows to obtain more precise lower bounds on the convergence ratio of $(\mu \dot{+} \lambda)$-ES.

Given a function f defined over D and $r > 0$, let $O_{f,r} = \{x \in D \mid f(x) < r\}$. We define the *level sets* $L_\mathcal{F}$ of a set $\mathcal{F}$ of functions defined over the domain D as

$$L_\mathcal{F} = \{O_{f,r} \mid f \in \mathcal{F}, r > 0\}.$$

We now briefly recall the definition of VC-dimension and Sauer's lemma [19,17] – our presentation is based on [14]. A set system on a set A is a family $\mathcal{S}$ of subsets of A. For $B \subseteq A$, we define the restriction of $\mathcal{S}$ to B as $\mathcal{S}|_B = \{S \cap B \mid S \in \mathcal{S}\}$. The VC-dimension of the set system $\mathcal{S}$ defined over A is defined as $\sup\{|B| \mid \mathcal{S}|_B = 2^B\}$ where 2^B denotes the powerset of B; in other words, it is the size of the largest subset B of A such that any subset of B can be obtained by intersecting B with an element of $\mathcal{S}$. Given a set system $\mathcal{S}$ over A,

the shatter function $\pi_{\mathcal{S}}$ is defined by $\pi_{\mathcal{S}}(m) = \max\{|\mathcal{S}|_B| \mid B \subseteq A,\ |B| = m\}$; thus $\pi_{\mathcal{S}}(m)$ is the maximum number of different subsets of A which can be obtained by intersecting a single subset of size m of A with all elements of $\mathcal{S}$. We next recall Sauer's lemma which gives an upper bound on $\pi_{\mathcal{S}}$ in terms of the VC-dimension of $\mathcal{S}$.

Lemma 1 (Sauer's lemma). *For any set system $\mathcal{S}$ of VC-dimension d, then for all integer m, it holds that $\pi_{\mathcal{S}}(m) \leqslant \sum_{i=0}^{d} \binom{m}{i}$.*

At last, let us recall the following classical bound [7] which is valid whenever $d \geqslant 3$:

$$\sum_{i=0}^{d} \binom{m}{i} \leqslant \min\{m^d, 2^m\}. \tag{2}$$

Note that the trivial bound 2^m is tight when $m \leqslant d$. The interesting case happens when m is large with respect to the VC-dimension d: the bound becomes polynomial in m in this case. This element is central for the difference between the results in this paper and results in [18].

In the rest of the paper, we assume the VC-dimension of considered set systems is always at least 3 (however, the case of VC-dimension smaller than 3 can be handled in a similar way; the bound above has to be replaced with $\sum_{i=0}^{d} \binom{m}{i} \leqslant m^d + 1$).

4.1 Non-elitist Strategies

We first give an upper bound on the branching factor of a SB-(μ, λ)-ES in terms of the VC-dimension of level sets.

Lemma 2. *Consider a SB-(μ, λ)-ES as described in Algorithm 1. Let $V \geqslant 3$ be the VC-dimension of the level sets of the family $\mathcal{F}$ of fitness functions under consideration. Then the branching factor of this algorithm satisfies $K \leqslant \lambda^V$.*

Proof. Given a set of λ points $P = \{x_1, \ldots, x_\lambda\}$ in the domain D, and $f \in \mathcal{F}$, let us define $M_f(P)$ to be the subset Q of size μ of P correponding to the μ points of P with lowest fitness values with respect to f. Note that the branching factor satisfies

$$K \leqslant \max_{P \subset D,\ |P| = \lambda} |\{M_f(P) \mid f \in \mathcal{F}\}|.$$

Now remark that for any P, the set Q of the μ points of P with lowest value (with respect to the fitness function f) can be separated from $P \setminus Q$ by an element from the level sets: in other words, there exists $O \in L_{\mathcal{F}}$ such that $O \cap P = Q$. It follows that

$$|\{M_f(P) \mid f \in \mathcal{F}\}| \leqslant \pi_{L_{\mathcal{F}}}(\lambda).$$

If the VC-dimension of $L_{\mathcal{F}}$ is at most V, it follows from Sauer's lemma and the bound given in Equation 2 that $\pi_{L_{\mathcal{F}}}(\lambda) \leqslant \lambda^V$. Thus $K \leqslant \lambda^V$. □

Theorem 2 (SB-(μ,λ)-ES). *Consider a SB-(μ,λ)-ES (Algorithm 1) in a domain $D \subset \mathbb{R}^d$, such that $D = \{f^* \mid f \in \mathcal{F}\}$. Let $V \geqslant 3$ be the VC-dimension of the level sets of $\mathcal{F}$. The convergence ratio of this algorithm satisfies*

$$\mathrm{CR}_\varepsilon \leqslant \frac{V \log \lambda}{d} \cdot \alpha(\varepsilon),$$

where $\alpha(\varepsilon) = 1/(1 - 1/N(\varepsilon))$.

Proof. The result easily follows from the upper bound on the branching factor given in Lemma 2, and from Theorem 1 as stated in Equation 1. □

4.2 Non-elitist Strategies with Full Ranking

This subsection deals with algorithms of type full ranking (μ,λ)-ES. It is organized as follows:

- First we study to which extent lower bounds obtained for SB-(μ,λ)-ES are modified when we use the full ranking information and not only selection information (i.e. we move from Algorithm 1 to Algorithm 2);
- Although the bounds obtained in the general case do not forbid a linear speed-up in λ, we show that the speed-up is asymptotically at most logarithmic in the special case of the sphere function;
- At last, for the sphere function again, we remark that a convergence ratio $\mathrm{CR}_\varepsilon = \Theta(1)$ can be reached in the case $\lambda = 2d$; this is to be compared to the best convergence ratio $\mathrm{CR}_\varepsilon = \Theta(1/d)$ we are aware of for $\lambda = O(1)$.

Keeping the full ranking information. Consider the case of Algorithm 2 instead of Algorithm 1; we have a wider family of algorithms as we can use all the ranking information. There are evolutionary algorithms which use the full ranking information of the selected points and not only selection; for example, roulette-wheel with rank-based fitness assignment (stochastic sampling [4], rank-based fitness assignment [3,20]), weighted recombination [11,1] or BREDA [10]. In this case, an upper bound on the number of possible outcomes of the selection step (including the ranking of children) is obtained by multiplying by $\mu!$ the number of possible outcomes in the case of selection only. This gives $\mathrm{CR}_\varepsilon \leqslant \frac{V \log(\lambda) + \mu \log \mu}{d} \cdot \alpha(\varepsilon)$. However, we can say better in the case where μ is large with respect to the VC-dimension V of the level sets of the fitness functions. (Proof of the following theorem is omitted due to space limitations.)

Theorem 3 (Full ranking (μ,λ)-ES). *Consider a (μ,λ)-ES (Algorithm 2) in a domain $D \subset \mathbb{R}^d$, such that $D = \{f^* \mid f \in \mathcal{F}\}$. Let $V \geqslant 3$ be the VC-dimension of the level sets of $\mathcal{F}$. The convergence ratio of this algorithm satisfies*

$$\mathrm{CR}_\varepsilon \leqslant \frac{V\,(\log \lambda + 4\mu)}{d} \cdot \alpha(\varepsilon),$$

where $\alpha(\varepsilon) = 1/(1 - 1/N(\varepsilon))$.

The case of the sphere function: complexity bounds for λ large. For the sphere function and the Euclidean norm, we next give an upper bound on the convergence ratio of a selection-based algorithm using full ranking.

Proposition 1. *Let $d \geqslant 3$. Consider a (μ, λ)-ES, as in Algorithm 2, optimizing the sphere function in a domain $D \subset \mathbb{R}^d$. Then* $\mathrm{CR}_\varepsilon \leqslant 2\log(\lambda) \cdot \alpha(\varepsilon)$*, where* $\alpha(\varepsilon) = 1/(1 - 1/N(\varepsilon))$.

Proof. Given two distinct points p and q in $\mathbb{R}^d$, we denote by $H_{p,q}$ be the mediator hyperplane of p and q, i.e. $H_{p,q} = \{x \in \mathbb{R}^d \mid \|x - p\| = \|x - q\|\}$.

At each iteration of the algorithm, an offspring of λ points $\{x_1, \ldots, x_\lambda\}$ is generated and the algorithm receives the sequence of indices of the μ points with lowest fitness values, ordered with respect to their fitness values. Obviously the branching factor is maximal when $\mu = \lambda$, i.e. when the algorithm is given the full ordering of points with respect to their fitness values. This information corresponds to giving the sign $s_{i,j}$ of $f(x_i) - f(x_j)$ for each $1 \leqslant i < j \leqslant \lambda$; this sign is positive or negative since we assumed equality never occurs. The number of possible sign vectors $s = (s_{i,j})_{1 \leqslant i < j \leqslant \lambda}$ is exactly the number of cells of the arrangement of hyperplanes $\{H_{x_i,x_j} \mid 1 \leqslant i < j \leqslant \lambda\}$ in $\mathbb{R}^d$. But it is known that n hyperplanes in $\mathbb{R}^d$ define at most n^d cells – see chapter 6 of [14]. Since there are $\binom{\lambda}{2} \leqslant \lambda^2/2$ hyperplanes here, we obtain $K \leqslant \left(\lambda^2/2\right)^d$. Applying Equation 1 yields the announced bound on the convergence ratio. □

When ε tends towards 0 and as $N(\varepsilon) \to \infty$, this gives $\mathrm{CR}_\varepsilon \leqslant 2\log\lambda$; this shows that the upper bound given by Theorem 3 cannot be reached in this case.

The case of the sphere function: Fast convergence ratio with $\lambda = 2d$. We point out here that for the specific case of the sphere function, a convergence ratio $\mathrm{CR}_\varepsilon = \Theta(1)$ can be reached with $\lambda = 2d$ in the domain $[0,1]^d$ by some algorithm of type full ranking (μ, λ)-ES.

This convergence ratio is easily obtained with the following algorithm. Let e_i denote the vector $(0, \ldots, 0, 1, 0, \ldots, 0)$ with a unique 1 in position i. First split $[0,1]^d$ into the 2^d cells delimited by the d hyperplanes of equations $x_i = 1/2$; the full ranking of the $2d$ points $\{(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}) + \frac{\eta}{2} e_i \mid 1 \leqslant i \leqslant n,\ \eta \in \{-1, 1\}\}$ allows to decide in which of these cells the optimum lies; then the algorithm proceeds recursively. This is quite similar to the Hooke and Jeeves algorithm [12].

After n iterations, the point $x_n^{(f)}$ proposed by this algorithm satisfies $\|x_n^{(f)} - f^*\|_2 \leqslant \sqrt{d}/2^n$. Moreover, this distance is realized by some fitness functions. It follows that $n_{\varepsilon,\frac{1}{2}} = \log\frac{1}{\varepsilon} + \frac{1}{2}\log d$. On the other hand $\log(N(\varepsilon)) = \Theta(d \log\frac{1}{\varepsilon})$. Thus, we have obtained:

$$\text{For } \lambda = 2d: \ \mathrm{CR}_\varepsilon = \frac{\log N(\varepsilon)}{d\, n_{\varepsilon,\frac{1}{2}}} = \Theta(1). \tag{3}$$

4.3 Elitist Strategies

Results obtained in the case of (μ, λ) algorithms can be translated into the elitist setting. Bounds obtained in these cases are given in Figure 1 (Section 5). Proofs of these results are omitted due to space limitation.

5 Summary of Results

Let's apply the results obtained in the previous section to the simple framework of the domain $D = [0,1]^d$ with the Euclidean norm. Lower bounds obtained in this setting are summarized in Figure 1. Higher values mean better possible convergence ratios. However, it is not known when these convergence ratios can be achieved. Indeed, result marked with (*) in Figure 1 is improved in the special case of the sphere function in Section 4.2: this shows that at least in this case, general bounds on convergence ratio derived from VC-dimension are not tight. Discussion of these results follows.

	SB-(μ,λ)-ES	SB-$(\mu+\lambda)$-ES	Full ranking (μ,λ)-ES	Full ranking $(\mu+\lambda)$-ES	Full ranking $(\infty+\lambda)$-ES
CR	$\frac{V}{d}\log\lambda$	$\frac{V}{d}\log(\mu+\lambda)$	$\frac{V}{d}(\log(\lambda)+4\mu)$ (*)	$\frac{V}{d}(\log(\lambda+\mu)+4\mu)$	$\frac{4V\lambda}{d}$

Fig. 1. Upper bound on the convergence ratio in the case of Euclidean norm in the domain $[0,1]^d$, when the level sets of fitness functions have VC-dimension V

Asymptotic speed-up in the case of selection only, non-elitist. In the case of evolution strategies based on selection only (algorithms of type SB-(μ,λ)-ES), the linear speed-up of selection-based evolution strategies shown in [5] cannot be obtained for λ large enough. Asymptotically, the speed-up obtained with such an algorithm is at most logarithmic as shown in Theorem 2.

Selection based algorithms vs. full ranking. When moving from selection based algorithms of type SB-(μ,λ)-ES to full ranking (μ,λ)-ES, upper bounds on the convergence ratio obtained here in the general case do not forbid a strong improvement asymptotically; essentially, the speed-up that could be achieved moves from logarithmic to linear in λ.

However, we know from Proposition 1 that the speed-up is at most logarithmic for a full ranking (μ,λ)-ES in the special case of sphere function. This raises the following question: for which kind of fitness functions is it interesting to keep the full ranking information?

Acknowledgements. Many thanks to Anne Auger, Nikolaus Hansen and Fabien Teytaud for constructive talks. This work was partially supported by the Pascal Network of Excellence.

References

1. Arnold, D.V.: Optimal weighted recombination. In: Wright, A.H., Vose, M.D., De Jong, K.A., Schmitt, L.M. (eds.) FOGA 2005. LNCS, vol. 3469, pp. 215–237. Springer, Heidelberg (2005)
2. Auger, A.: Convergence results for (1,λ)-SA-ES using the theory of φ-irreducible Markov chains. Theoretical Computer Science 334(1-3), 35–69 (2005)

3. Bäck, T., Hoffmeister, F., Schwefel, H.-P.: Extended selection mechanisms in genetic algorithms. In: Belew, R.K., Booker, L.B. (eds.) Proceedings of the Fourth International Conference on Genetic Algorithms, pp. 92–99. Morgan Kaufmann Publishers, San Mateo (1991)
4. Baker, J.E.: Reducing bias and inefficiency in the selection algorithm. In: Proceedings of the Second International Conference on Genetic Algorithms on Genetic algorithms and their application, pp. 14–21. Lawrence Erlbaum Associates, Inc., Mahwah (1987)
5. Beyer, H.-G.: Toward a theory of evolution strategies: On the benefit of sex - the $(\mu/\mu, \lambda)$-theory. Evolutionary Computation 3(1), 81–111 (1995)
6. Beyer, H.-G., Schwefel, H.-P.: Evolution strategies: a comprehensive introduction. Natural Computing 1(1), 3–52 (2002)
7. Devroye, L., Györfi, L., Lugosi, G.: A probabilistic Theory of Pattern Recognition. Springer, Heidelberg (1997)
8. Droste, S.: Not all linear functions are equally difficult for the compact genetic algorithm. In: Proc. of the Genetic and Evolutionary Computation COnference (GECCO 2005), pp. 679–686 (2005)
9. Feller, W.: An introduction to Probability Theory and its Applications. Wiley, Chichester (1968)
10. Gelly, S., Ruette, S., Teytaud, O.: Comparison-based algorithms are robust and randomized algorithms are anytime. Evolutionary Computation Journal (MIT Press), Special issue on bridging Theory and Practice 15(4), 411–434 (2007)
11. Hansen, N., Ostermeier, A.: Completely derandomized self-adaptation in evolution strategies. Evolutionary Computation 9(2), 159–195 (2001)
12. Hooke, R., Jeeves, T.A.: "Direct search" solution of numerical and statistical problems. Journal of the ACM 8(2), 212–229 (1961)
13. Jägersküpper, J., Witt, C.: Rigorous runtime analysis of a $(\mu + 1)$ES for the sphere function. In: GECCO, pp. 849–856 (2005)
14. Matoušek, J.: Lectures on Discrete Geometry. Graduate Texts in Mathematics, vol. 212. Springer, Heidelberg (2002)
15. Rechenberg, I.: Evolutionstrategie: Optimierung Technischer Systeme nach Prinzipien des Biologischen Evolution. Fromman-Holzboog Verlag, Stuttgart (1973)
16. Rudolph, G.: Convergence rates of evolutionary algorithms for a class of convex objective functions. Control and Cybernetics 26(3), 375–390 (1997)
17. Sauer, N.: On the density of families of sets. Journal of Combinatorial Theory, Ser. A 13(1), 145–147 (1972)
18. Teytaud, O., Gelly, S.: General lower bounds for evolutionary algorithms. In: Proceedings of PPSN, pp. 21–31 (2006)
19. Vapnik, V.N., Chervonenkis, A.Ya.: On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and its Applications XVI(2), 264–280 (1971)
20. Whitley, D.: The GENITOR algorithm and selection pressure: Why rank-based allocation of reproductive trials is best. In: Schaffer, J.D. (ed.) Proceedings of the Third International Conference on Genetic Algorithms, pp. 116–121. Morgan Kaufmann, San Mateo (1989)