Conditioning, Halting Criteria and Choosing λ

Olivier Teytaud

Equipe TAO (Inria), LRI, UMR 8623 (CNRS - Université Paris-Sud),
bât 490 Université Paris-Sud 91405 Orsay Cedex France
olivier.teytaud@inria.fr

Abstract. We show the convergence of $1+\lambda$-ES with standard step-size update-rules on a large family of fitness functions without any convexity assumption or quasi-convexity assumptions ([3,6]). The result provides a rule for choosing λ and shows the consistency of halting criteria based on thresholds on the step-size.

The family of functions under work is defined through a condition-number that generalizes usual condition-numbers in a manner that only depends on level-sets. We consider that the definition of this condition-number is the relevant one for evolutionary algorithms; in particular, global convergence results without convexity or quasi-convexity assumptions are proved when this condition-number is finite.

1 Introduction

We consider here a $1+\lambda_t$-ES algorithm as in Algorithm 1. We will, in a more general framework than state of the art papers (in spite of the fact that the functions are unimodal), show: (i) conditions under which the halting criterion ensure a good final output (Section 2); (ii) how to choose λ (Sections 3 and 4); (iii) the convergence of the algorithm (Section 5).

The state of the art contains convergence proofs on simple functions (e.g. the sphere function [4,1,2]), or more general lower bounds ([7,10]), or for simplified algorithms. In fact, the positive results are essentially convergence results for convex of quasi-convex fitness functions (*i.e.*, functions for which level sets are convex); this is not close to the practice of evolutionary algorithms, which can follow long non-convex valleys as in e.g. Rosenbrock's banana function. We here show our convergence on hypothesis which do not imply neither convexity nor quasi-convexity.

2 The Model and the Consistency of the Halting Criterion

Assume that the fitness is such that

$$\forall v \in \mathbb{R}, \texttt{fitness}^{-1}(v) = g(v)E_v \tag{1}$$

where $E_v \subset \mathbb{R}^d$ and where g is an increasing mapping $[0,\infty[\to [0,\infty[$ with $g(0)=0$. This implies that the $\inf \texttt{fitness} = 0$ and $\texttt{fitness}(0)=0$; as the

N. Monmarché et al. (Eds.): EA 2007, LNCS 4926, pp. 196–206, 2008.

Algorithm 1. $1+\lambda_t$-ES. The population size λ_t depends on t. The halting criterion depends on the mutation strength σ. The $N_{t,i}$ are usually, but not necessarily, independent Gaussians. λ_t will be chosen as in Equation 13 (quasi-random case, Section 3) or Equation 15 (random case, Section 4).

```
initialize x_1 ∈ R^d, σ_1 > σ_0, t = 1.
while σ_t ≥ σ_0 do
   Update λ_t (Equation 13 or 15).
   for i ∈ {1, ..., λ_t} do
      x(i) = x_t + σ_t N_{t,i}.
   end for
   x' = arg min_{x∈{x_t, x(1), ..., x(λ)}} fitness(x).
   if fitness(x') < fitness(x_t) then
      Acceptance for time step t: x_{t+1} = x'.
      Choose σ_{t+1} > σ_t.
   else
      Rejection for time step t: x_{t+1} = x_t.
      Choose σ_{t+1} < σ_t.
   end if
   t = t + 1.
end while
Output x'.
```

algorithm is translation-invariant (both in the fitness-space and in the domain) this does not reduce the generality. As the algorithm only uses comparisons, we can equivalently consider Equation 2 (*i.e.*, $g(v) = v$):

$$\forall v \in \mathbb{R},\ \texttt{fitness}^{-1}(v) = vE_v \tag{2}$$

$$\text{and we assume } \forall v \in \mathbb{R},\ E_v \subset \bigcup_{B^o(z,1)\subset \bigcup_{v'<v} v'E_{v'}} S(z,1) \tag{3}$$

where $B^o(x,r) = \{t; ||t-x|| < r\}$ and $S(x,r) = \{t; ||t-x|| = r\}$.The constant 1 is arbitrary, but we can rescale E_v; in fact, the hypothesis is that for some ϵ, a level-set vE_v is included in the union of all spheres of radius $v\epsilon$ enclosing areas of lower fitness. We let

$$C(\texttt{fitness}) = \inf_{(E_v)_{v\in[0,\infty[}\text{such that (2) and (3) hold}} \sup_v \sup_{e\in E_v} ||e||.$$

This equation is not simple. The family $(E_v)_{v\in[0,\infty[}$ is not uniquely determined by the fitness function; we consider the inf for all possible families $(E_v)_{v\geq 0}$ such that Equations 2 and 3 hold. There is also a supremum of $||e||$ for all $v \geq 0, e \in E_v$.

$C(\texttt{fitness})$ depends on the shape of the level-sets of the fitness, can be seen as a condition number, dedicated to comparison-based algorithms. For example, for the sphere-function, $E_v = E$ (independent of v) and $E = S$ (the unit sphere in $\mathbb{R}^d$) and $C(\texttt{fitness}) = 1$. This number is finite for many, many fitness-functions; mainly, the level-sets have to be connected. For example, the fitness function with level-sets as in Fig. 1 has a finite $C(\texttt{fitness})$. Another nice property is that this condition-number is finite for quadratic fitness functions and generalizes the classical condition-number of quadratic fitness functions. Yet another feature is illustrated by experiments in Fig. 2: an infinite $C(\texttt{fitness})$ can lead to premature convergence of $1+\lambda$-ES.

We claim:

Main lemma for the halting criterion. *Assume that eqs 2 and 3 hold.*

$$\text{If for all } t \in \mathbb{N}, \quad \epsilon S \subset \cup_{i\in\{1,\dots,\lambda_t\}} B^o(N_{t,i},\epsilon) \tag{4}$$

$$\text{and if } \sigma_T < \sigma_0, \text{ then} \quad \texttt{fitness}(x_T) \leq \epsilon\sigma_{T-1} \tag{5}$$

$$\text{and} \quad ||x_T|| \leq \epsilon\sigma_{T-1} C(\texttt{fitness}). \tag{6}$$

Proof: Assume that eqs 2 and 3 hold, and that for all t,

$$\epsilon S \subset \cup_{i\in\{1,\dots,\lambda_t\}} B(N_{t,i},\epsilon).$$

Then,

$$\text{Equation 3 leads to} \quad E_v \subset \cup_{B^o(z,1)\subset \cup_{v'<v} \frac{v'}{v} E_{v'}} S(z,1) \tag{7}$$

$$\text{which leads to} \quad e \in E_v \Rightarrow \exists f; ||f-e|| = 1 \wedge B^o(vf,v) \subset \cup_{v'<v} v' E_{v'}. \tag{8}$$

Assume that $\sigma_T < \sigma_0$, and that T is minimal with this condition (as $t \mapsto \texttt{fitness}(x_t)$ is non-increasing, there's no loss of generality in this assumption).

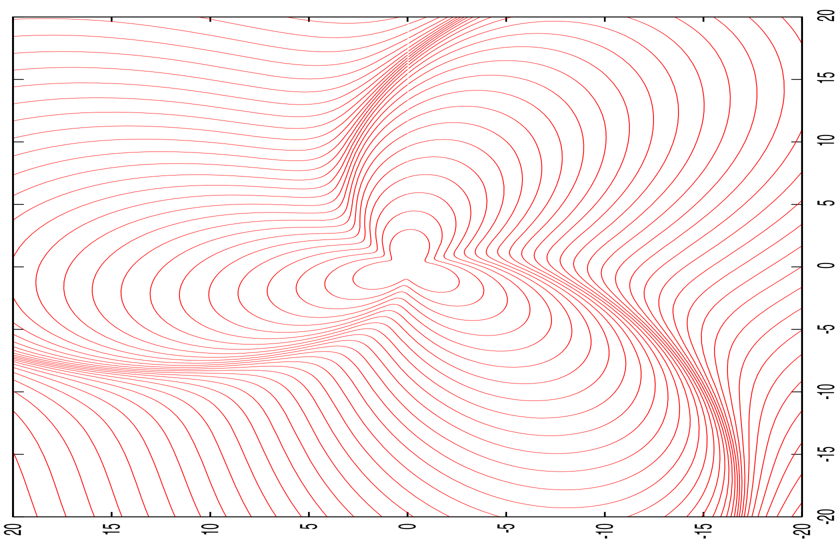

Fig. 1. An example of fitness-function (level sets are plotted) with finite $C(\texttt{fitness})$. The fitness is not convex; it is also not quasi-convex. Much more complicated examples can be defined; mainly, we need level sets which all contain a "wide" path to the optimum (at least with width scaling as the level set).

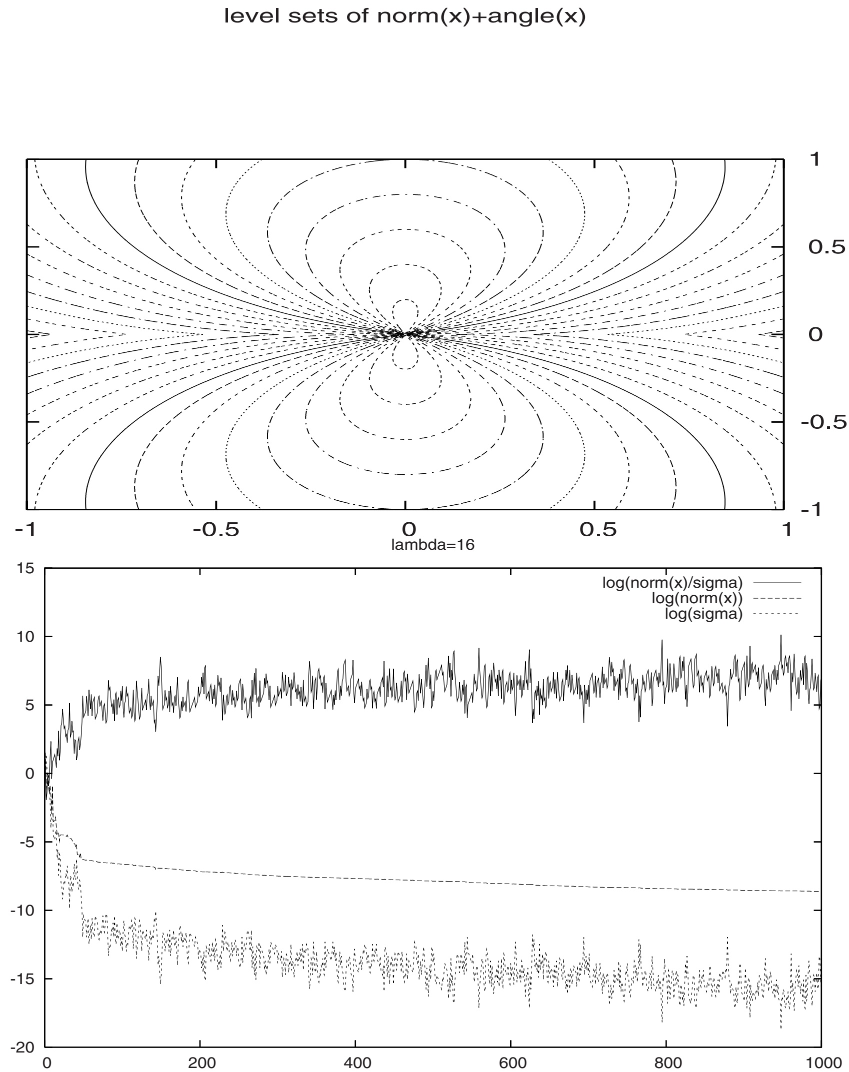

Fig. 2. Level sets of a simple function ($x \mapsto ||x|| + angle(x)^2$, with $angle(x)$ the angle between x and an axis) with infinite $C(\texttt{fitness})$ (top), and results of $1+\lambda$-ES with $\lambda = 16$ and one-fifth rule ([9,8]) on this function (bottom). We see that σ falls down, without convergence: this is a premature convergence which illustrates Corollary 1.

This implies that at $t = T - 1$, we have a reject; therefore,

$$\forall i \in \{1, \ldots, \lambda_t\}, \texttt{fitness}(x(i)) \geq \texttt{fitness}(x_t)). \tag{9}$$

Let $x = x_t$ for short. Equation 2 implies that

$$x = \texttt{fitness}(x)e \text{ for some } e \in E_{\texttt{fitness}(x)}. \tag{10}$$

Equation 9 leads to, for all $i \in \{1, \ldots, \lambda_t\}$,

$$
\begin{aligned}
& x + \sigma_t N_{t,i} \notin \cup_{v' < \texttt{fitness}(x)} v' E_{v'} \text{ and successively:} \\
& \texttt{fitness}(x) e + \sigma_t N_{t,i} \notin \cup_{v' < \texttt{fitness}(x)} v' E_{v'} \\
& \qquad \text{with } e \text{ as in Equation 10,} \\
& \texttt{fitness}(x) e + \sigma_t N_{t,i} \notin B^o(\texttt{fitness}(x) f, \texttt{fitness}(x)) \\
& \qquad \text{with } f \text{ as in Equation 8,} \\
& \sigma_t N_{t,i} \notin B^o(\underbrace{\texttt{fitness}(x) f - \texttt{fitness}(x) e}_{=r}, \texttt{fitness}(x)) \\
& \qquad \text{with } ||r|| = \texttt{fitness}(x) \text{ by Equation 8,} \\
& \sigma_t N_{t,i} \notin B^o(r, \texttt{fitness}(x)), \\
& N_{t,i} \notin B^o(r/\sigma_t, \texttt{fitness}(x)/\sigma_t), \\
& N_{t,i} \notin B^o(\delta, ||\delta||) \\
& \qquad \text{where } \delta = r/\sigma_t \text{ verifies } ||\delta|| = \texttt{fitness}(x)/\sigma_t.
\end{aligned}
$$

We assume, to get a contradiction, that

$$||\delta|| = \texttt{fitness}(x)/\sigma_t \geq \epsilon. \tag{11}$$

Then, Equation 11, together with $c > 1 \Rightarrow B(a, ||a||) \subset B(c.a, c||a||)$, implies

$$\forall i \in \{1, \ldots, \lambda_t\}, N_{t,i} \notin B^o(\epsilon \frac{1}{||\delta||} \delta, \epsilon).$$

This is a contradiction with the assumption that for all t, $\epsilon S \subset \cup_{i \in \{1,\ldots,\lambda_t\}} B(N_{t,i}, \epsilon)$. Therefore, Equation 11 does not hold. Hence, $\texttt{fitness}(x) < \epsilon\sigma_t = \epsilon\sigma_0$. This leads to Equation 5; Equation 5 and Equation 2 lead to Equation 6. □

3 Choosing λ in the Derandomized Setting

Equation 4 (recalled below, Equation 12, in the case $\epsilon = 1$) is the main assumption of the main lemma above:

$$S \subset \cup_{i \in \{1,\ldots,\lambda_t\}} B^o(N_{t,i}, 1). \tag{12}$$

We consider $\epsilon = 1$ as this hypothesis has moderate impact on the result; the results below are similar with other values of ϵ. We now study how to ensure Equation 12. A solution consists in using a minimal 1-cover of S; $\lambda_t = \lambda_{QR}$ where

$$\lambda_{QR} = \inf\{\lambda \in \mathbb{N}; d_1, \ldots, d_\lambda \in S^\lambda; S(0,1) \subset B^o(d_1, 1) \cup B^o(d_2, 1) \cup \cdots \cup B^o(d_\lambda, 1)\}. \tag{13}$$

It is known ([5]) that $\lambda_{QR} \geq c \cos(\phi_1)/\sin(\phi_1)^d d^{3/2} \ln\left(1 + d^2 \cos(\phi_1)\right)$, with $\phi_1 = \arg\cos(1/2)$. This leads to λ_{QR} of order roughly $1/\sin(\phi_1)^d$; the exponential dependency in d can not be removed.

Let's show that we can not ensure the halting criterion without at least λ_{QR} points, for any deterministic offspring ($N_{t,i} = N_{1,i}$ deterministically fixed).

Corollary 1 (lower bound on λ for deterministic offsprings). *If $\lambda < \lambda_{QR}$ and for any fixed $N_{t,1}, \ldots, N_{t,\lambda}$ independent of t, then there exists an update rule for σ (see Algorithm 1), σ_0, and a function `fitness` verifying eqs 2 and 3, such that $\sigma_T < \sigma_0$ and $||x_T|| > \sigma_{T-1}C($`fitness`$)$.*

Proof: We build a counter-example with $T = 2$, $\texttt{fitness}(x) = ||x||$, any update rule setting $\sigma_{t+1} = 0$ in case of rejection, $\sigma_1 = 1$. Then, for all $v > 0$, $E_v = S = S(0, 1)$.
We just have to choose $x_1 \in S$ such that for all $i \in \{1, \ldots, \lambda\}$,

$$x_1 + N_{t,i} \notin B^o(0, 1)$$

or equivalently, we need, for building the counter-example, an x such that for all $i \in \{1, \ldots, \lambda\}$,

$$N_{t,i} \notin B^o(-x_1, 1)$$

$$i.e., \ \ ||N_{t,i} + x|| \geq 1;$$

such an x_1 exists by equation 13, as soon as $\lambda < \lambda_{QR}$, as the $B^o(N_{t,i}, 1)$ can't cover S. □

We note $d_1^{QR}, \ldots, d_{\lambda_{QR}}^{QR}$ the points realizing Equation 13; these points are by definition a minimal covering of the sphere by open balls of radius ones with centers on the sphere.

4 Choosing λ_t in the Random Case

We now consider $N_{t,1}, \ldots, N_{t,\lambda_t}$ independently randomly uniformly drawn in $S(0, 1)$. The question is: for which values of λ_t do we ensure Equation 4 (or 12) with probability $1 - \delta$? We set

$$N = \inf\{\lambda \in \mathbb{N}; y_1, \ldots, y_\lambda \in S^\lambda; S(0,1) \subset B^o(y_1, \frac{1}{2}) \cup B^o(y_2, \frac{1}{2}) \cup \cdots \cup B^o(y_\lambda, \frac{1}{2})\}. \tag{14}$$

The formula of N is close to Equation 13 but with radius $\frac{1}{2}$ instead of 1. [5] shows that $N \leq c\cos(\phi_2)/\sin(\phi_2)^d d^{3/2} \ln\left(1 + d\cos(\phi_2)^2\right)$ with $\phi_2 = 2\arg\sin(1/4)$; roughly, N is of order $O(1/\sin(\phi_2)^d)$. It is not possible to get rid of the exponential dependency in d.

Theorem 2. *Assume that*

$$\lambda_t \geq N\left(\log(N) + \log(t^2) + \log(1/\delta) - \log(\pi^2/6)\right) \tag{15}$$

and that the $N_{t,i}$ are independently uniformly drawn on $S(0, 1)$. Then, Equation 12 holds with probability at least $1 - \delta$.

Before the proof of this result, let's show a simple corollary, based on theorem 2 and on the main lemma:

Corollary 3 for algorithm 1. Assume that eqs 2 and 3 hold, and that

$$\lambda_t \geq N\left(\log(N) + \log(t^2) + \log(1/\delta) - \log(\pi^2/6)\right)$$

with $N_{t,i}$ independent random variables uniform on S. Then, with probability at least $1-\delta$, $\sigma_T < \sigma_0 \Rightarrow ||x_T|| \leq \sigma_{T-1} C(\texttt{fitness})$.

Remark A. If the step-size adaptation rule is of the form $\sigma_{n+1} = \beta\sigma_n$ in case of rejection, then the result implies $\sigma_T < \sigma_0 \Rightarrow ||x_T|| \leq \sigma_0 C(\texttt{fitness})/\beta$.

Remark B: Gaussian mutations. We use spheres instead of Gaussians as it is more parsimonious (λ smaller) than in the case of Gaussians; however, the result is essentially the same with Gaussians. With just have to add a multiplicative factor in Equation 17 in the proof below (the factor is polynomial in d).

Proof of the corollary: Application of theorem 2 and of the main lemma. □

Let's now show theorem 2.

Proof of Theorem 2: Assume that

$$\lambda_t \geq N\left(\log(N) + \log(t^2) + \log(1/\delta) + \log(\sum_{i\geq 1} 1/i^2)\right).$$

This is equivalent to Equation 15. We note δ_t the probability of Equation 4 with $\epsilon = 1$, namely δ_t is the probability of

$$S \subset \cup_{i\in\{1,\ldots,\lambda_t\}} B^o(N_{t,i}, 1). \tag{16}$$

We let $y_1, \ldots, y_n$ be elements of S realizing Equation 14. We see that if

$$\forall i \in \{1,\ldots,N\}, \exists j \in \{1,\ldots,\lambda_t\} N_{t,j} \in B^o(y_i, \frac{1}{2}),$$

then Equation 16 holds.

Therefore, with μ the uniform measure,

$$\delta_t \leq \sum_i \pi_j \left(1 - P(N_{t,j} \in B^o(y_i, \frac{1}{2}))\right),$$

$$\delta_t \leq N(1-1/N)^{\lambda_t} \text{ as } \mu\left(B^o(y_i, \frac{1}{2}) \cap S\right) \geq \mu(S)/N, \tag{17}$$

$$\log(\delta_t) \leq \log(N) + \lambda_t \log(1-1/N) \leq \log(N) - \lambda_t/N.$$

Then,

$$\sum_{t\geq 1} \delta_t \leq N\exp(-\lambda_t/N)$$

$$\leq \sum_t \delta\left(\sum_{i\geq 1} 1/i^2\right)/(t^2)$$

$$\leq \delta \qquad \text{which is the expected result.}$$

□

5 Convergence Issues: $1+\lambda$-ES Almost Surely Halts

We have considered above the risk of raising the halting criterion before a good fitness value is met. This is meaningless, however, if we do not show that, after a finite time, the halting criterion will be met.

Theorem 4: almost sure convergence. *We assume that the update rules are as follows:*

- $\sigma_{t+1} = \min(\alpha\sigma_t, \sigma_{max})$ *in case of acceptance* ($\alpha > 1$)*;*
- $\sigma_{t+1} = \beta\sigma_t$ *in case of rejection* ($0 < \beta < 1$)*.*

We assume that Equations 2 and 3 hold for some $\epsilon > 0$*. We assume that the measure* $\mu([0,1[E_v)$ *of* $[0,1[E_v > G > 0$*. We assume that* $C(\texttt{fitness}) < \infty$ *and* $N_{t,i}$ *are independent standard multivariate Gaussians. We also assume that* $\lambda_t \leq Zt^\zeta$ *for some* $Z < \infty, \zeta < 1$*. Then, almost surely,* $\exists T > 0, \sigma_T < \sigma_0,$ *i.e., the algorithm halts.*

Proof
We note $T = \inf\{t; \sigma_t < \sigma_0\}$ (possibly, a priori, $T = \infty$). We first point out some simple useful facts about the $(\sigma_t)_{t\in\mathbb{N}}$:

1. $\forall t > 0, \sigma_t \leq \sigma_{max}$.
2. $\forall t < T, \sigma_t \geq \sigma_0$.
3. If rejection holds at all steps $t+1, \ldots, t+n_0$, with $n_0 \geq \log(\sigma_{max}/\sigma_0)/\log(1/\beta)$, then $T \leq t + n_0 + 1 < \infty$.

Now, some simple facts about the $(x_t)_{t\in\mathbb{N}}$:

1. $t \mapsto \texttt{fitness}(x_t)$ is non-increasing.
2. $||x_t|| \leq C\texttt{fitness}(x_t) \leq C\texttt{fitness}(x_0)$.
3. Thanks to $t \leq T \Rightarrow (\sigma_t \geq \sigma_0 \wedge ||x_t|| \leq C.\texttt{fitness}(x_t) \leq C.\texttt{fitness}(x_0)$,

$$\begin{aligned} &P(\texttt{fitness}(x_t) < \epsilon | x_{t-1}, \sigma_{t-1}) \\ &> P(x_t + \sigma_t N_{t,1} \in cE_c | x_{t-1}, \sigma_{t-1}) \\ &> P(N_{t,1} \in (cE_c - x_t)/\sigma_t | x_{t-1}, \sigma_{t-1}) \\ &> c^d \mu(E_c) d\left((||x_t|| + cC(\texttt{fitness})) / \sigma_t \right) \\ &> K\epsilon^d \end{aligned}$$

   for some $K > 0$ that only depends on d, Z, and σ_0.
4. The previous point implies that $P\left(\exists u < t; \texttt{fitness}(x_u) < \epsilon\right) > 1 - (1 - K\epsilon^d)^t$, and therefore if $d' < d$,

$$P\left(\exists u < t; \texttt{fitness}(x_u) < (1/t)^{1/d'}\right) > 1 - \left(1 - K/t^{d/d'}\right)^t \rightarrow 1 \text{ as } t \rightarrow \infty.$$

5. The previous points implies that if $d' > d$, then almost surely, there exists $t_0 < \infty$ such that

$$t \geq t_0 \Rightarrow \texttt{fitness}(x_t) < t^{-1/d'}. \quad (18)$$

Let's now consider the probability p_t of rejection at steps t, conditionally to x_t and σ_t, conditionally to $t \leq T$.

We point out that if $\forall i \leq \lambda_t, \sigma_t||N_{t,i}|| > ||x_t|| + \mathtt{fitness}(x_t)C$, then there is rejection (all x'_i have in that case norm $> C\mathtt{fitness}(x_t)$ and therefore have fitness $> \mathtt{fitness}(x_t)$). This implies that

$$\begin{aligned}
p_t &\geq 1 - (P(\sigma_t||N_{t,1}|| > ||x_t|| + C\mathtt{fitness}(x_t)))^{\lambda_t} \\
&\geq 1 - (P(\sigma_0||N_{t,1}|| > C\mathtt{fitness}(x_t) + C\mathtt{fitness}(x_t)))^{\lambda_t} \\
&\qquad\qquad \text{as } \sigma_t \geq \sigma_0 \text{ and } ||x_t|| \leq C\mathtt{fitness}(x_t) \\
&\geq 1 - (P(\sigma_0||N_{t,1}|| > 2C\mathtt{fitness}(x_t)))^{Zt^{\zeta}} \text{ as } \lambda_t \leq Zt^{\zeta} \\
&\geq 1 - \left(P\left(\sigma_0||N_{t,1}|| > 2Ct^{-1/d'}\right)\right)^{Zt^{\zeta}} \text{ if } t \geq t_0 \text{ thanks to Equation 18} \\
&\geq 1 - \left(1 - 1/t^{d/d'}\right)^{Zt^{\zeta}} \\
&\geq p_0 > 0 \text{ if we choose } d' \text{ s.t. } d < d' < d/\zeta.
\end{aligned}$$

The probability of rejection at all steps $t+1, \ldots, t+n_0$, conditionally to x_t and σ_t, is therefore at least $1-(1-p_0)^{n_0} > p > 0$. This quantity is lower bounded by a positive number; this implies that almost surely, such a sequence of rejections almost surely occurs, hence the expected result. □

6 Discussion: Derandomization, Halting Criteria, Robustness, Conditioning

Let's summarize our results about the $1+\lambda$-ES for fitness functions with not-too-bad conditioning in the sense of Equations 2, 3 and $C(\mathtt{fitness})$ with $\lambda_t = O(t^{\zeta})$ for some $\zeta < 1$, and with an update rule for σ as in Theorem 4. By Theorem 4, we know that the algorithm converges almost surely (*i.e.*, it halts after a finite number of time steps). By Corollaire 3, we know that if the population size verifies Equation 15, then with probability at least $1-\delta$, the algorithm stops close to the optimum - within distance $\sigma_0 C(\mathtt{fitness})\beta$. $C(f)$ quantifies the conditioning, and is finite also for many non-convex functions. Therefore, we have, for some λ_t logarithmic in t:

– global convergence with high probability;
– consistency of the halting criterion, *i.e.*no premature convergence.

A main strength of this result is that no convexity, no smoothness, no quasi-convexity is assumed and we have global convergence; see Fig. 1. As far as we know, there's no convergence proof of $1+\lambda$-ES that is not covered by the results in this paper. Another strength is that $C(\mathtt{fitness})$ appears as an important relevant criterion for evolutionary algorithms: it generalizes the usual conditioning (which is a local criterion), and:

– Fig. 2 shows that very simple functions with $C(\mathtt{fitness})$ lead to premature convergence;

- corollary 3 and Theorem 4 show that $C(\texttt{fitness})$ finite leads to both (i) convergence with high probability (ii) consistency of the halting criterion.

$C(\texttt{fitness})$ only depend on level sets, as well as the behavior of most evolutionary algorithms, and is finite for many fitness functions without convexity or quasi-convexity; mainly, it assumes that at each scale, the width of the path to the optimum scales as the diameter of the level set. We believe that the definition of $C(\texttt{fitness})$ is the main contribution of this paper.

A weakness is that we ensure convergence, and the efficiency of the halting criterion, but there's no convergence rate. However, evolutionary algorithms are more well known for robustness than for convergence rates. Moreover, a convergence rate can easily be derived under some slightly stronger assumptions.

Our results propose a rule for choosing λ_t as a function of t, δ, d (see Equations 14 and 15). This rule is reasonable for its dependency in t and δ (logarithmic dependency); the dependency in the dimension is prohibitively high, but it is a fact that evolutionary algorithms are not stable in front of large dimensionality.

We see in the results above that:

- the population size should scale as
  - $\log(t)$ (recall that $\log(t^2) = 2\log(t)$); population size should therefore increase with time (very slowly).
  - $\log(1/\delta)$; more robustness requires a bigger population size.
  - $N\log(N)$, which is exponential in d.
- we can compare the number of points required for avoiding too early convergence of the algorithm in the randomized and in the derandomized setting by comparing λ_{QR} (Equation 13) and λ_t (random case, Equation 15); in both cases, λ is exponential in d, but with a much better constant in the derandomized case. On the other hand, the convergence proof (theorem 3) only holds for the random case.

Acknowledgments. We thank Anne Auger for fruitful talks and Pierre Liardet for constructive remarks on a preliminar version of this paper. This work was supported in part by the Pascal Network of Excellence.

References

1. Auger, A.: Convergence results for (1,λ)-SA-ES using the theory of φ-irreducible markov chains. Theoretical Computer Science 334, 35–69 (2005)
2. Auger, A., Hansen, N.: Reconsidering the progress rate theory for evolution strategies in finite dimensions. In: Press, A. (ed.) Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2006), pp. 445–452 (2006)
3. Bäck, T.: Evolutionary Algorithms in theory and practice. Oxford University Press, York (1995)
4. Bienvenue, A., Francois, O.: Global convergence for evolution strategies in spherical problems: some simple proofs and difficulties. Theor. Comput. Sci. 306(1-3), 269–289 (2003)

5. Boroczky, K., Wintsche, G.: Covering the sphere by equal spherical balls. In: Discrete and Computational Geometry, pp. 237–253 (2003)
6. Eiben, A., Smith, J.: Introduction to Evolutionary Computing. Springer, Heidelberg (2003)
7. Jagerskupper, J., Witt, C.: Runtime analysis of a (mu+1)es for the sphere function. Technical Report ISSN 1433-3325, University of Dortmund (2005)
8. Kern, S., Müller, S., Hansen, N., Büche, D., Ocenasek, J., Koumoutsakos, P.: Learning Probability Distributions in Continuous Evolutionary Algorithms - A Comparative Review. Natural Computing 3, 77–112 (2004)
9. Rechenberg, I.: Evolutionstrategie: Optimierung Technischer Systeme nach Prinzipien des Biologischen Evolution. Fromman-Holzboog Verlag, Stuttgart (1973)
10. Teytaud, O., Gelly, S.: General lower bounds for evolutionary algorithms. In: Tenth International Conference on Parallel Problem Solving from Nature (PPSN 2006) (2006)