technische universität
dortmund

Theoretical Analysis of Continuous Evolutionary Algorithms

Alexandru Agapie
Günter Rudolph

Algorithm Engineering Report
TR11-2-001
February 2011
ISSN 1864-4503

Faculty of Computer Science
Algorithm Engineering (Ls11)
44221 Dortmund / Germany
http://ls11-www.cs.uni-dortmund.de/

Theoretical Analysis of Continuous Evolutionary Algorithms

Alexandru Agapie* Günter Rudolph†

1 The Continuous EA is a Renewal Process

For each $t = 0, 1, 2, \ldots$, let P_t be the random variable '(best individual from) EA population at iteration t'. Then $\{P_t\}_{t\geq 0}$ is a stochastic process on $\Re^n$. We also define a distance $d : \Re^n \leftarrow \Re_0^+$, accounting for the (one-dimensional) distance to optimum, that is, to $0 := (0, \ldots, 0)$ since we are minimising. Distance d will also stand for our *drift function*. As generally the case with probabilistic algorithms on continuous space, we say convergence is achieved at iteration t if the algorithm has entered an ϵ-vicinity of 0 for some fixed ϵ, $0 \leq d(P_t) < \epsilon$. We also define the stochastic process $\{X_t\}_{t\geq 1}$ given by

$$X_t = d(P_{t-1}) - d(P_t) \qquad t = 1, 2, \ldots.$$

In our EA framework, X_t will stand for the (relative) progress of the algorithm in one step, namely from the $(t-1)$st iteration to the tth. Due to EA's elitism $\{X_t\}_{t\geq 1}$ are *non-negative* random variables (r.v.s), and we shall also assume they are independent. Each X_t is composed of a point mass (singular, or Dirac measure) in zero accounting for the event where there is no improvement from P_{t-1} to P_t, and a continuous part accounting for the real progress toward the optimum - a truncated uniform or normal distribution, e.g.. A second natural assumption is that $P\{X_t = 0\} < 1$, or equivalently $P\{X_t > 0\} > 0$, for all t, otherwise convergence of the algorithm would be precluded. That does not require a progress at each iteration, but only a strictly positive probability to that event, which is different. However, in order for the stochastic analysis to be consistent, the fulfilment of either one of the following hypotheses will be required.

*Academy of Economic Studies, Dept. Mathematics, Calea Dorobantilor 15-17, Bucharest 010552, Romania, agapie@clicknet.ro

†Computer Science XI, Technical University Dortmund 44227, Germany, guenter.rudolph@tu-dortmund.de

H_1:	$\{X_t\}_{t\geq 1}$ are non-negative, independent, identically distributed r.v.s with finite mean μ.
H_2:	$\{X_t\}_{t\geq 1}$ are non-negative, independent r.v.s and there exist constants $\mu_1, \mu_2, \sigma > 0$ such that $\mu_1 \leq E(X_t) \leq \mu_2$ and $Var(X_t) \leq \sigma^2$, for all t.

H_1 is well-known within the theory of stochastic processes, yet cumbersome to achieve when modelling continuous EAs on practical problems. H_2 is more flexible, allowing for different mutation rates and different success probabilities at different algorithmic iterations. For example, H_2 describes a family of distributions that are all normal, or all uniform, with the parameters ranging within certain positive bounds. One can easily see that, under supplementary assumption 'X_t has finite variance', the following implication holds:

$$H_1 \Rightarrow H_2$$

but not vice-versa.

It is shown below that both hypothesis yield a stronger confinement on the progress probabilities, than the already stated '$P\{X_t > 0\} > 0$ for all t'. We need first some general results from probability theory.

Lemma 1.1 *If X is a positive random variable and $\alpha > 0$ s.t. $P\{X \geq \alpha\} = 0$, then $E(X) \leq \alpha \cdot P\{X < \alpha\}$.*

Proof.
Let $M > \alpha$ and define the r.v.s

$$X_\alpha = \begin{cases} X & \text{if} \quad X \geq \alpha \\ \alpha & \text{if} \quad X < \alpha \end{cases} \qquad\qquad X_M = \begin{cases} \inf\{X, M\} & \text{if} \quad X \geq \alpha \\ \alpha & \text{if} \quad X < \alpha \end{cases}.$$

Then X_α, X_M are positive and

$$E(X_M) \;\leq\; M \cdot P\{X \geq \alpha\} + \alpha \cdot P\{X < \alpha\} = \alpha \cdot P\{X < \alpha\}.$$

Moreover, since $\{X_M\}_M$ is monotone increasing and $X_M \rightarrow X_\alpha$ as $M \rightarrow \infty$, Lebesque's Monotone Convergence theorem[1] - see e.g. [24] p.59 - ensures that $E(X_M) \rightarrow E(X_\alpha)$ as $M \rightarrow \infty$, thus

$$E(X_\alpha) \;\leq\; \alpha \cdot P\{X < \alpha\}$$

and conclusion yields after transporting $X \leq X_\alpha$ to expected values. □

Lemma 1.2 $H_2 \Rightarrow$ *there exist $\alpha, \beta > 0$ such that $P\{X_t \geq \alpha\} \geq \beta$ for all t.*

[1] If $0 \leq X_1 \leq X_2 \leq \ldots$ and $X_n \rightarrow X$ with probability 1, then $E(X_n) \rightarrow E(X)$.

Proof.
We show first that H_2 implies $P\{X_t \geq \alpha\} > 0$ for all $\alpha < \mu_1$ and all t. Let us fix $0 < \alpha < \mu_1$ arbitrarily. Suppose, ad absurdum, there is a t with $P\{X_t \geq \alpha\} = 0$. Then lemma 1.1 implies

$$E(X_t) \leq \alpha \cdot P\{X_t < \alpha\} = \alpha \cdot 1 < \mu_1$$

which contradicts H_2. So $P\{X_t \geq \alpha\} > 0$ holds for all t. We show next that the same inequality holds if we intercalate some $\beta > 0$. Actually, we are going to prove that for any non-negative r.v. X with $E(X) = \mu$ and $Var(X) = \sigma^2$,

$$P(X > \alpha) \geq \frac{(\mu - \alpha)^2}{\sigma^2 + \mu^2} \qquad \text{for any } \alpha < \mu. \tag{1}$$

The proof involves *conditional expectations* - see e.g. [24] p.83.

$$\begin{aligned} \mu =&E(X \cdot I_{X\leq\alpha}) + E(X \cdot I_{X>\alpha}) = \\ =&E(X|X \leq \alpha) \cdot P\{X \leq \alpha\} + E(X|X > \alpha) \cdot P\{X > \alpha\} \end{aligned}$$

where $I_{X\leq\alpha}$ is the indicator function, $I_{X\leq\alpha} = 1$ if $X \leq \alpha$ and zero otherwise. We have next

$$E(X|X > \alpha) \cdot P\{X > \alpha\} \geq \mu - \alpha P\{X \leq \alpha\} \Rightarrow$$
$$E(X|X > \alpha) \geq \frac{\mu - \alpha P\{X \leq \alpha\}}{P\{X > \alpha\}}$$

and if we apply the same decomposition to $E(X^2)$,

$$\begin{aligned} \sigma^2 + \mu^2 =&E(X^2) = E(X^2 \cdot I_{X\leq\alpha}) + E(X^2 \cdot I_{X>\alpha}) = \\ =&E(X^2|X \leq \alpha) \cdot P\{X \leq \alpha\} + E(X^2|X > \alpha) \cdot P\{X > \alpha\} \geq \\ \geq&E(X^2|X > \alpha) \cdot P\{X > \alpha\} \geq E(X|X > \alpha)^2 \cdot P\{X > \alpha\} \geq \\ \geq&\frac{(\mu - \alpha P\{X \leq \alpha\})^2}{P\{X > \alpha\}} \geq \frac{(\mu - \alpha)^2}{P\{X > \alpha\}} \end{aligned}$$

which proves (1). Now, if we have a family of non-negative r.v.s $\{X_t\}_{t\geq1}$ such that $\mu_1 \leq E(X_t) \leq \mu_2$ and $Var(X_t) \leq \sigma^2$ for all t (hypothesis H_2), the same reasoning yields, for an arbitrarily fixed α with $0 < \alpha < \mu_1$,

$$P(X_t > \alpha) \geq \frac{(\mu_1 - \alpha)^2}{\sigma^2 + \mu_2^2} =: \beta.$$

$\square$

A somehow different[2] proof is given in *Appendix A* for the case of normal mutations with uniformly bounded mean and variance.

One can easily see that, under H_1, the conclusion of lemma 1.2 is a direct consequence of $P\{X_t > 0\} > 0$ for all t. Note that lemma 1.2 holds also for a different version of hypothesis H_2, namely:

[2]Actually, the proof for normal mutations relies on the operation of *truncation*, which is a particular version of conditional expectation.

H_2':	$\{X_t\}_{t\geq 1}$ non-negative, independent, and there is r.v. Z with $E(Z) < \infty$ and $X_t \leq Z$ for all t, and constant μ_1 with $0 < \mu_1 \leq E(X_t)$ for all t.

Remark 1.3 *Hypothesis H_2' does* not *imply H_2, nor vice-versa. H_2' applies, e.g., to a familly of uniform r.v.s that are uniformly bounded.*

However, we preferred version H_2 over H_2' having in mind the typical normal mutation used in continuous EAs.

Lemma 1.4 *$H_2' \Rightarrow$ there exist $\alpha, \beta > 0$ such that $P\{X_t \geq \alpha\} \geq \beta > 0$ for all t.*

Proof.
The inequality in lemma 1.4 can be also written with a single constant, α

$$P\{X_t \geq \alpha\} \geq \alpha > 0 \qquad\qquad \text{for all } t. \tag{2}$$

Assume, ad absurdum, (2) does not stand. Then for any $\alpha > 0$, say $\alpha := 1/n$, there is an index t_n such that

$$P\left\{X_{t_n} \geq \frac{1}{n}\right\} < \frac{1}{n}.$$

If we let $n \to \infty$, we obtain that $X_{t_n} \to 0$ *in probability*, as $n \to \infty$. Than Lebesque's Dominated Convergence theorem[3] - see e.g. [24] p.59 - implies $E(X_{t_n}) \to 0$, which contradicts $0 < \mu_1 \leq E(X_t)$ from H_2'.

□

Let us return to defining the renewal process in case of the continuous EA optimisation. By summing up the relative progress at each iteration we obtain S_t, the (overall) progress in t iterations:

$$\begin{aligned} S_t = \sum_{k=1}^{t} X_k =& d(P_0) - d(P_1) + d(P_1) - d(P_2) + \ldots + d(P_{t-1}) - d(P_t) = \\ =& d(P_0) - d(P_t) \qquad\qquad t = 1, 2, \ldots. \end{aligned}$$

Remark 1.5 *By definition, S_t is bounded within the closed interval $[0,\ d(P_0)]$, for all $t \geq 1$. If we fix at the start of the algorithm a positive δ to designate the 'maximal distance to optimum', then we have*

$$0 \leq S_t \leq d(P_0) \leq \delta.$$

Let us now introduce another r.v., accounting for the EA's *first hitting time* of the area $[0,\ d(P_0) - d)$, or equivalently, for the overall progress to go beyond d - a certain positive threshold[4]:

$$T_d = inf\{t \mid d(P_t) < d(P_0) - d\} = inf\{t \mid S_t > d\}.$$

[3] If $|X_n| \leq Z$ such that $E(Z) < \infty$ and $X_n \to X$ *in probability*, then $E(X_n) \to E(X)$.

[4] In order to keep the notation simple, we shall use the same letter 'd' for denoting the distance function $d(\cdot)$, and a scalar $d > 0$.

According to [8, 17], the process $\{T_d\}_{d>0}$ will be called a *renewal process*[5] with the following interpretation: We say a renewal occurs at distance $d(P_0)-d$ from the optimum if $S_t = d$ for some iteration t. A *renewal* is actually a 'successful iteration', that is, an iteration that produced a strictly positive progress towards the optimum. After each renewal the process (the algorithm) starts over again.

2 First Hitting Time

From this point further, all results concerning the convergence of the renewal process associated to the continuous EA will be stated '*under hypotheses H_1/H_2*', meaning '*either under hypothesis H_1, or under H_2*'. Accordingly, we shall split each proof in two parts; as H_1 corresponds to the classical definition of a renewal process, the first part will be in general a simple adaptation of the corresponding result from [17].

Proposition 2.1 *Under hypothesis H_1/H_2, the first hitting time of the continuous EA is finite with probability 1.*

Proof.
The Strong Law of Large Numbers yields $S_t/t \to \mu$ with probability 1. Hence for any positive d, $S_t \leq d$ only finitely often and thus $T_d < \infty$ with probability 1.

Assuming now H_2, the Strong Law of Large Numbers for independent *non-identical* r.v.s[6] yields

$$\mu_1 \leq \frac{S_t}{t} \leq \mu_2 \qquad \text{with probability 1.}$$

Using the left hand side of the inequality for a fixed d yields $S_t \leq d$ only finitely often, and $T_d < \infty$ with probability 1. $\square$

Definition 2.2 *An integer valued positive random variable T is called a stopping time for the sequence $\{X_t\}_{t\geq 1}$ if the event $\{T = t\}$ is independent of $X_{t+1}, X_{t+2}, \ldots$ for all $t \geq 1$.*

We have the following simple result.

Lemma 2.3 *T_d defined as above is a stopping time for $\{X_t\}_{t\geq 1}$, for any $d > 0$.*

Proof.

$$\{T_d = t\} = \{S_t > d,\ S_{t-1} \leq d\} = \{\sum_{k=1}^{t} X_k > d,\ \sum_{k=1}^{t-1} X_k \leq d\}$$

[5] The continuous-time index t of a classical renewal process $\{N_t\}_{t\geq 0}$ in queueing theory is replaced in our paradigm by a continuous-distance index d.

[6] $\{X_t\}_{t\geq 1}$ independent s.t. $\sum_{1}^{\infty} \frac{Var(X_i)}{i^2} < \infty$, then $\frac{\sum_{1}^{t}[X_i - E(X_i)]}{t} \to 0$ with probability 1.

which is obviously independent of $X_{t+1}, X_{t+2}, \dots$. □

We also have the relationship that the first hitting time of a distance d from the starting point is greater than t if and only if the tth iteration yields a point situated at distance less than or equal d. Formally,

$$T_d > t \quad \Leftrightarrow \quad S_t \leq d.$$

According to [17], $E(T_d)$, the mean/expected value of T_d is called the *renewal function*, and much of classical renewal theory is concerned with determining its properties. In our EA framework, if we set $d := d(P_0) - \epsilon$ with some fixed positive ϵ defining the target-zone of the continuous space algorithm, then $T_d = inf\{t \mid d(P_t) < \epsilon\}$ is the first hitting time of the target-zone, and $E(T_d)$ the *expected (first) hitting time*. So determining the properties of the renewal function seems to be the principal goal of EA theory as well.

Table below summarises the intuitive interpretation of the random variables X_t, S_t and T_d under the continuous EA setting.

Random Variable	Interpretation
X_t	(one-dimensional) progress between the $(t-1)$st and the tth iteration
S_t	overall progress up to the tth iteration
T_d	(no. of iterations) first hitting time of a distance d from the starting point

The following theorem is crucial to the stochastic analysis of continuous EAs. Note that this result was also used in [12], yet outside the context of renewal processes.

Theorem 2.4 (Wald's Equation, [17] p.38) *If $\{X_t\}_{t\geq 1}$ are independent and identically distributed random variables having finite expectations $E(X)$, and T is a stopping time for $\{X_t\}_{t\geq 1}$ such that $E(T) < \infty$, then*

$$E\left(\sum_{t=1}^{T} X_t\right) = E(T) \cdot E(X).$$

□

When applied to the continuous EA paradigm, Wald's equation provides only a lower bound on the expected hitting time. In order to obtain both upper and lower bounds, the application of limit theorems from renewal processes is necessary.

A reformulation in terms of inequalities of Wald's equation is first required.

Theorem 2.5 (Wald's Inequation) *If $\{X_t\}_{t\geq 1}$ are independent, non-negative, $\mu_1 \leq E(X_t) \leq \mu_2$ for all t and T is a stopping time for $\{X_t\}_{t\geq 1}$, then*

$$\mu_1 E(T) \leq E\left(\sum_{t=1}^{T} X_t\right) \leq \mu_2 E(T).$$

Proof.
Letting

$$Y_t = \begin{cases} 1 & \text{if} \quad T \geq t \\ 0 & \text{if} \quad T < t \end{cases}$$

we have that

$$\sum_{t=1}^{T} X_t = \sum_{t=1}^{\infty} X_t Y_t.$$

Due to Lebesque's Monotone Convergence theorem (see proof of lemma 1.1) since all products $X_t Y_t$ are positive, one can interchange expectation and summation

$$E\left(\sum_{t=1}^{T} X_t\right) = \sum_{t=1}^{\infty} E(X_t Y_t). \tag{3}$$

Note that $Y_t = 1$ if and only if we have not stopped after succesively observing $X_1, \ldots, X_{t-1}$, therefore X_t is independent of Y_t for all t. Then

$$\begin{aligned} E\left(\sum_{t=1}^{T} X_t\right) &= \sum_{t=1}^{\infty} E(X_t) E(Y_t) \\ \mu_1 \sum_{t=1}^{\infty} E(Y_t) &\leq E\left(\sum_{t=1}^{T} X_t\right) \leq \mu_2 \sum_{t=1}^{\infty} E(Y_t) \\ \mu_1 \sum_{t=1}^{\infty} P\{T \geq t\} &\leq E\left(\sum_{t=1}^{T} X_t\right) \leq \mu_2 \sum_{t=1}^{\infty} P\{T \geq t\} \\ \mu_1 E(T) &\leq E\left(\sum_{t=1}^{T} X_t\right) \leq \mu_2 E(T). \end{aligned}$$

That $E(T) = \sum_{t=1}^{\infty} P\{T \geq t\}$ can be seen - for any non-negative integer r.v. T - as follows:

$$\begin{aligned} \sum_{t=1}^{\infty} P\{T \geq t\} &= \sum_{t=1}^{\infty} \sum_{k=t}^{\infty} P\{T = k\} = \sum_{k=1}^{\infty} P\{T = k\} \cdot \sum_{t=1}^{k} 1 = \\ &= \sum_{k=1}^{\infty} k \cdot P\{T = k\} = E(T). \end{aligned}$$

□

Note that the only confinements on $\{X_t\}_{t \geq 1}$ required by theorem 2.5 were '$X_t \geq 0$', and '$\mu_1 \leq E(X_t) \leq \mu_2$', for all t - hence a simplified version of H_2. Condition '$E(T) < \infty$', which appeared in Wald's equation, was no longer used in the inequation. Actually, if one follows the proof of theorem 2.5, she/he will observe that the only point where such condition could apply would be at interchanging expectation and summation in equation (3). Instead, we have used Lebesque's Monotone Convergence theorem, which does not require a condition like '$E(T) < \infty$' but only monotony of the partial sums - ensured by '$X_t \geq 0$ for all t'.

So, apparently, one could conclude that whenever Wald's inequation is applied, H_2 may be replaced by that simplified hypothesis. That is not the case, since $E(T)$ will designate the expected hitting time of an area at certain distance from the starting point of the algorithm, and if $E(T) = \infty$ there is no convergence at all. Hence we need also $E(T) < \infty$ for our analysis, and that is proved under the continuous EA paradigm in proposition 2.7 below, relying strongly on lemma 1.2, which in turn does not work unless all requirements in H_2 are fulfilled!

We show next that the result of proposition 2.1 holds also for the *expected* hitting time of the renewal process modelling the continuous EA. That is not trivial, since finiteness with probability 1 of a positive random variable does not imply finiteness of its expected value, see e.g. the Cauchy distribution.

First we need a simple result.

Lemma 2.6 *Let us consider a discrete random variable* $Z = \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$ *and* $Z_1, Z_2, \ldots$ *be independent, identically distributed as* Z. *Let also consider the stopping time* $M = inf\{m \mid Z_1 + \ldots + Z_m = 1\}$. *Then* $E(M) = 1/p$.

Proof.
For each positive integer k

$$P\{M = k\} = P\{Z_1 = 0, \ldots, Z_{k-1} = 0, Z_k = 1\} = p(1-p)^k$$

and thus the mean of discrete r.v. M,

$$M = \begin{pmatrix} 1 & 2 & \ldots & k & \ldots \\ p & p(1-p) & \ldots & p(1-p)^k & \ldots \end{pmatrix}$$

is computed as

$$E(M) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p \sum_{k=1}^{\infty} kp(1-p)^{k-1} \ \rightarrow \ \frac{1}{p}$$

where convergence comes from both-side derivation of the geometrical series $\sum_{k=1}^{\infty}(1-p)^k$. □

Proposition 2.7 *Under hypotheses* H_1/H_2, *the expected hitting time of the continuous EA is finite.*

Proof.
We need to prove that $E(T_d) < \infty$ for all $d > 0$. Under H_1, since $P\{X_t > 0\} > 0$, one can easily prove - ad absurdum - the existence of constants $\alpha, \beta > 0$ s.t.

$$P\{X_t > \alpha\} \geq \beta > 0 \qquad \text{for all } t. \tag{4}$$

Under H_2, the existence of α and β is guaranteed by lemma 1.2. Define a related renewal process $\{\overline{X}_t\}_{t\geq 1}$ by truncating each X_t to

$$\overline{X}_t = \begin{pmatrix} 0 & \alpha \\ 1-\beta & \beta \end{pmatrix}.$$

Note that $\overline{X}_t$ does not depend on t anymore. Also, under hypothesis H_1, $\overline{X}_t$ is the same with

$$\overline{X}'_t = \begin{cases} 0 & \text{if} \quad X_t < \alpha \\ \alpha & \text{if} \quad X_t \geq \alpha \end{cases}$$

but not under H_2, where we have only $\overline{X}_t \leq \overline{X}'_t$.

Let now define $\overline{T}_d = inf\{t \mid \overline{X}_1 + \ldots + \overline{X}_t > d\}$. For the related process, succesive iterations can move the algorithm only along the lattice $d = t\alpha,\ t = 0, 1, 2 \ldots$. Also, the number of iterations required for a *success* (a real jump of length α) are independent random variables with mean $1/\beta$.

To see that, apply lemma 2.6 for the r.v. $Z_t = \overline{X}_t/\alpha$ which registers 1 for a success and 0 for a stagnation of the EA from iteration $(t-1)$ to iteration t. Thus,

$$E(\overline{T}_d) \leq \frac{[d/\alpha]+1}{\beta} < \infty$$

and the rest follows since $\overline{X}_t \leq X_t$ holds under either H_1 or H_2, and that implies $\overline{T}_d \geq T_d$.

$\square$

3 Main Result

The expression $1/E(X_t)$ is often called the *progress rate* between the $(t-1)$st and the tth iteration. Following the general theory of renewal processes [8, 17], we prove next the highly intuitive result that the (expected) average number of iterations required per distance unit converges to the progress rate. As $E(T_d)$ represents the expected hitting time of an area situated at distance d from the starting point of the algorithm, the result below will provide estimates of the convergence time for continuous EAs.

We stress again that the estimates given below are meaningless without the assertion '$E(T_d) < \infty$ *for all* $d > 0$', ensured under H_1/H_2 through proposition 2.7 .

Theorem 3.1 *Under hypotheses H_1/H_2 we have, as $d \to \infty$*

$$\frac{E(T_d)}{d} \;\to\; \frac{1}{\mu},$$

respectively

$$\frac{1}{\mu_2} \;\leq\; \frac{E(T_d)}{d} \;\leq\; \frac{1}{\mu_1}.$$

Proof.
Assume first hypothesis H_1. As $E(T_d)$ is finite due to proposition 2.7, Wald's equation applied on $\{X_t\}_{t\geq 1}$ and T_d yields

$$E(S_{T_d}) = E(T_d) \cdot \mu. \tag{5}$$

Since $S_{T_d} > d$ we also have $E(T_d) \cdot \mu \geq d$ which yields

$$\liminf_{d\to\infty} \frac{E(T_d)}{d} \;\geq\; \frac{1}{\mu}. \tag{6}$$

To go the other way, let us fix a constant M and define a new renewal process $\{\overline{X}_t\}_{t\geq 1}$ by letting for each $t = 1, 2, \ldots$

$$\overline{X}_t = \begin{cases} X_t & \text{if} \quad X_t \leq M \\ M & \text{if} \quad X_t > M \end{cases} \tag{7}$$

Let $\overline{S}_t = \overline{X}_1 + \ldots + \overline{X}_t$ and $\overline{T}_d = inf\{t \mid \overline{S}_t > d\}$. Since $\overline{X}_t \leq M$ for all t,

$$\overline{S}_{T_d} \;\leq\; d + M$$

and since $\{\overline{X}_t\}_{t\geq 1}$ satisfies also H_1, $E(\overline{T}_d)$ is finite and Wald's equation yields

$$E(\overline{T}_d) \cdot \mu_M \;\leq\; d + M$$

where we denoted $\mu_M := E(\overline{X}_t)$. Thus

$$\limsup_{d\to\infty} \frac{E(\overline{T}_d)}{d} \;\leq\; \frac{1}{\mu_M}$$

and since $\overline{X}_t \leq X_t$ for all t, it follows that $\overline{T}_d \geq T_d$ and $E(\overline{T}_d) \geq E(T_d)$, thus

$$\limsup_{d\to\infty} \frac{E(T_d)}{d} \;\leq\; \frac{1}{\mu_M}.$$

Letting now $M \to \infty$ yields

$$\limsup_{d\to\infty} \frac{E(T_d)}{d} \;\leq\; \frac{1}{\mu}$$

which together with (6) completes the proof for this case.

Under H_2, $E(T_d)$ is finite due to proposition 2.7 and Wald's *inequation* yields

$$\mu_1 E(T_d) \leq E(S_{T_d}) \leq \mu_2 E(T_d). \tag{8}$$

Since $S_{T_d} > d$ we also have $\mu_2 E(T_d) \geq E(S_{T_d}) > d$ and thus

$$\liminf_{d\to\infty} \frac{E(T_d)}{d} \;\geq\; \frac{1}{\mu_2}. \tag{9}$$

On the other hand, for each *positive integer* M the process $\overline{X}_t$ defined by (7) is non-negative. Since $0 \leq \overline{X}_t \leq X_t$, we have for each $M > 0$

$$0 \leq \mu_M^1 \leq E(\overline{X}_t) \leq \mu_2 \qquad \text{for all } t \tag{10}$$

where μ_M^1 does *not* depend on t and is defined for each M by

$$\mu_M^1 = min\{\mu_1, \inf_t E(\overline{X}_t)\}.$$

We show that $\mu_M^1 \to \mu_1$ as $M \to \infty$. Assume, ad absurdum, the contrary. As $\mu_M^1 \leq \mu_1$ always holds, divergence implies that for any $\epsilon > 0$, there are infinitely many M_k outside the ϵ-vicinity of μ_1, the interval $(\mu_1 - \epsilon, \mu_1]$. According to the definition of μ_M^1, we have for an arbitrary $\epsilon > 0$

$$E(\overline{X}_t) < \mu_1 - \epsilon \qquad \text{for all } t \text{ and all } \{M_k\}_{k \geq 1}. \tag{11}$$

But if we apply Lebesque's Dominated Convergence theorem (see proof of lemma 1.4 and note that convergence *with probability* 1 implies convergence *in probability*) to a fixed t_0, $0 \leq \overline{X}_{t_0} \leq X_{t_0}$, $E(X_{t_0}) \leq \mu_2 < \infty$ and $\overline{X}_{t_0} \to X_{t_0}$, as $M_k \to \infty$, it yields

$$E(\overline{X}_{t_0}) \to E(X_{t_0}) \qquad \text{as } M_k \to \infty,$$

and combined with $\mu_1 \leq E(X_{t_0})$ provides an index k_0 such that

$$E(\overline{X}_{t_o}) \geq \mu_1 - \epsilon \qquad \text{for all } M_k \text{ with } k \geq k_0$$

which obviously contradicts (11)!

So $\mu_M^1 \to \mu_1$ from below and as $0 < \mu_1$, one can find a positive M_0 such that

$$0 < \mu_M^1 \leq E(\overline{X}_t) \leq \mu_2 \qquad \text{for all } t \text{ and all } M \geq M_0.$$

It is easy to see that, for sufficiently large M

$$0 \leq a_t \leq \alpha < \mu_M^1 \leq E(\overline{X}_t) \leq \mu_2 \qquad \text{for all } t$$

which means that $\{\overline{X}_t\}_{t \geq 1}$ also satisfies an H_2 condition, thus $E(\overline{T}_d) < \infty$ for all d and one can call Wald's inequation on $\{\overline{X}_t\}_{t \geq 1}$ and $\overline{T}_d$, yielding

$$\mu_M^1 E(\overline{T}_d) \leq E(\overline{S}_{\overline{T}_d}) \leq \mu_2 E(\overline{T}_d).$$

Return to the main proof to observe that $\overline{S}_{\overline{T}_d} \leq d + M$ together with $\overline{X}_t \leq X_t$ and $\overline{T}_d \geq T_d$ imply

$$\mu_M^1 E(T_d) \leq \mu_M^1 E(\overline{T}_d) \leq E(\overline{S}_{\overline{T}_d}) < d + M.$$

As in the first case,

$$\limsup_{d \to \infty} \frac{E(T_d)}{d} \leq \frac{1}{\mu_M^1}.$$

and letting $M \to \infty$ yields

$$\limsup_{d\to\infty} \frac{E(T_d)}{d} \;\leq\; \frac{1}{\mu_1}$$

which together with (9) completes the proof.

□

As one can see from the proof of theorem 3.1, the left hand side of the inequality - the one giving a lower bound on $E(T_d)$ - is a simple consequence of Wald's inequation. Most of the effort was concentrated on validating the upper bound of the expected hitting time - far more significant for computation time analysis.

Translated to our continuous EA paradigm, theorem 3.1 says that the expected *average*[7] hitting time:

i converges, under hypothesis H_1, to the inverse of the expected progress in one step, respectively

ii is bounded, under hypothesis H_2, by the inverse bounds of the expected progress in one step.

The estimates for the expected hitting time hold for a general $(1 + \lambda)$ EA, optimising an arbitrary fitness function defined on n-dimensional continuous space. The case of EA with constant parameters is obviously covered, but also the more practical situation where parameters are adapted (are allowed to vary) during the evolution - see section 5.

The analysis performed so far on *continuous* EAs regarded as renewal processes is similar to the Markov chain analysis of *discrete* EAs performed in [1, 2, 18, 20] - see [21] for an accurate state of art in stochastic convergence for discrete EAs. It closes the theoretical discussion on convergence of the algorithm, opening the door for particular estimations of local progress rates μ, respectively μ_1 and μ_2. As this calculus has a long history in EA theory, we shall use some of the previous results in the remaining sections of the paper.

But first let us digress and see how theorem 3.1 extrapolates the drift theorems of discrete EAs.

4 Drift Analysis

Drift analysis is relatively old within probability optimisation theory [9], yet it was only recently that it has been introduced as a powerful tool in studying convergence of evolutionary algorithms [10, 11, 16]. Exclusively devoted to discrete EAs, drift analysis made obsolete the highly technical proof given in [7] for hitting time in case of $(1 + 1)$ EA on linear (pseudo-Boolean) functions [6].

Following ..., we start by reviewing the definition and main drift theorem, as introduced for a *finite* space Z containing all possible EA populations.

Define *distance* $d := Z \to \Re_+$ with $d(P) = 0$ if and only if population P contains the optimal solution - so we are minimizing. Let $T = inf\{t \mid d(P_t) =$

[7]With respect to distance on the progress axis.

$0\}$ be a random variable, and consider a maximum distance of an arbitrary population to the optimum

$$M := max\{d(P)|P \in Z\}.$$

The fact that $M < \infty$ comes from the finiteness of the search space. As for each iteration t, the current population P_t is a random variable, so will be $d(P_t)$ and also the so-called *decrease in distance* function X_t (D_t in the original approach), given by

$$X_t := d(P_{t-1}) - d(P_t).$$

By definition, $E(X_t|T \geq 0)$ is called *drift*. We also introduce

$$\Delta := min\{E(X_t|T \geq t)|t \geq 1\}.$$

Theorem 4.1 (Drift Theorem - Upper Bound) *If* $\Delta > 0$ *then*

$$E(T) \leq \frac{M}{\Delta}$$

□

To see the resemblance note that, under the continuous space paradigm of Section 1, M is replaced by d (which may later grow to infinity), T is replaced by T_d (the first hitting time of an area situated at distance d from the initial population), while the existence of Δ is postulated by hypotheses H_1/H_2: $\Delta = \mu$ respectively $\Delta = \mu_1$.

With this substitution in mind, drift theorem is contained in theorem 3.1, as the right hand side (upper bound) limit:

$$\frac{E(T_d)}{d} \leq \frac{1}{\mu} \qquad \text{respectively} \qquad \frac{E(T_d)}{d} \leq \frac{1}{\mu_1}.$$

A similar formulation of the drift theorem comes from [10, 11].

Theorem 4.2 *Let* $\{X_t\}_{t>0}$ *be a Markov process over a set of states* S*, and* $g : S \to \Re_+$ *a function that assigns to every state a non-negative real number. Let the time to reach the optimum be* $T := min\{t > 0 \mid g(X_t) = 0\}$*. If there exists* $\delta > 0$ *such that at any time step* $t > 0$ *and at any state* X_t *with* $g(X_t) > 0$ *the following condition holds:*

$$E[g(X_{t-1}) - g(X_t)|g(X_{t-1}) > 0] \geq \delta \tag{12}$$

the

$$E[T|X_0, g(X_0) > 0] \leq \frac{g(X_0)}{\delta}. \tag{13}$$

□

Or, with a slightly different conclusion in [16]:

$$E(T) \leq \frac{E[g(X_0)]}{\delta}. \tag{14}$$

To see the resemblance, notice first that as we set 0 to be the minimum of the optimisation problem, the conditional probabilities in both (12)-(13) vanish. Second, $g(X_0)$ can be replaced by the (constant) maximal distance to the optimum, d in our renewal process setting. Denoting $\mu_1 := \delta$, and observing that expected value of a constant is the constant itself, inequations (13) and (14) read now the same as the right hand side of theorem 3.1

$$\frac{E(T)}{d} \leq \frac{1}{\mu_1}.$$

In [6] a somehow different condition on the drift function is imposed, namely, there is a constant $\delta > 0$ such that for all n and all populations P_t,

$$E[d(P_t)] \;\leq\; \left(1 - \frac{\delta}{n}\right) d(P_{t-1}). \tag{15}$$

But as this is a one-step condition, one can assume that P_{t-1} is constant, and only P_t is (a random) variable, which ensures $P_{t-1} = E(P_{t-1})$ and after insertion in (15) we get

$$E[d(P_t) - d(P_{t-1})] \;\leq\; -\frac{\delta \cdot d(P_{t-1})}{n}.$$

After introducing X_t and reversing the inequality we obtain

$$E(X_t) \;\geq\; \frac{\delta \cdot d(P_{t-1})}{n},$$

thus a more elaborated version of the lower bound in H_2

$$E(X_t) \;\geq\; \mu_1,$$

accounting also for the space-dimension n and for the current position P_{t-1}.

Summing up, drift analysis provides conditions that ensure the existence of a strictly positive lower bound on the expected one-step progress of the algorithm towards the optimum, yielding finite upper bounds on the expected hitting time of the algorithm - all on the discrete case. Our renewal process analysis did the same, but for the continuous case. Nota bene, lower bounds on the hitting time are also available under the new paradigm, provided the existence of a finite upper bound on the expected one-step progress.

5 Adaptive mutation

How can one apply theorem 3.1 to computing practical hitting times of continuous EAs? In general, estimates of the one-step expected progress could be derived either

a. *ANALITICALLY*, provided the optimisation problem, fitness function and evolutionary operators are manageable enough, or

b. *NUMERICALLY*, by running a single iteration of the algorithm for several times and/or from different points in the search space and then averaging the outcomes.

The first path resonates lauder - to a mathematical ear, at least - but so far only the smoothest functions (linear, quadratic) and simplest algorithms ((1+1) EA, mainly) exhibit close formulas for the expected one-step progress in the continuous case [5, 12, 13, 19, 20]. In turn, the numerical approach is far more general, its potential application varying from smooth to black-box optimisation problems, from $(1+1)$ EA to $(\mu+\lambda)$ EAs including all sort of evolutionary operators. However, we defer the experimental study to a future paper, and concentrate within this section on estimating *analytically* the hitting time of the $(1+1)$ EA with *uniform mutation inside the (hyper)-sphere of radius r* (r variable), minimising the well-known *SPHERE* function[8]

$$f : \Re^n \to \Re \qquad\qquad f(x) = f(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i^2 \, .$$

One can bound the uniform mutation both in mean and variance such that it satisfies hypothesis H_2. On the other hand, we claim that uniform mutation inside the sphere is more tractable than normal mutation, at least from a geometrical point of view.

To see that, note the following simple facts. First, the expected value of a uniform variable defined inside a figure of volume 1 is the *centroid* (center of mass) of the corresponding figure. If the figure of volume 1 is truncated - as the case with elitist EAs on SPHERE, where not all of the mutation sphere is active for next generation, the removed volume (probability) being charged to a single point, zero, the expected value will still be the centroid of the truncated figure. Second, if the mutation sphere is no longer of volume 1 - as it happens when we successively decrease mutation radius r, we need to divide the uniform 'variable' and consequently its expected value by the volume of the new sphere - call it $V_n(r)$ - in order for the non-unitary sphere to define a proper random variable.

We are also going to need the following geometrical results, the proof being deferred to Appendix B.

Proposition 5.1 *Let S_n be the n-dimensional sphere of volume* 1*, centered in* $0 = (0, \ldots, 0)$*. Consider the positive semi-sphere that is symmetric around the x_1 axis. Then the centroid A_n of the semi-sphere satisfies, as $n \to \infty$*

$$A_n \to A = \left(\frac{1}{\pi\sqrt{e}}, 0, \ldots, 0 \right) .$$

[8] In order to avoid confusion, we shall use capital letters when referring to the fitness function, and small letters when referring to the mutation operator.

□

Corollary 5.2 *If we multiply the radius of the sphere from proposition 5.1 by $r = r(n)$, the coordinates of the centroid will be multiplied by the same factor.*

□

Note that the limit value obtained for the position of the centroid along the x_1 axis is $1/\pi\sqrt{e} = 0.193$, in good concordance with 1/5, the well-known threshold value used for mutation adaptation in *evolutionary strategies* - see e.g. [22, 23]!

We are going to use the calculus of centroids for estimating the upper and lower bounds on the expected one-step progress of the $(1+1)$ EA with spherical mutation along the 'progress axis' Ox_1 [9]. As usually the case in adaptive EAs, we shall decompose the algorithm into different phases with respect to distance to the optimum, each phase keeping a fixed mutation radius, and progressively decrease the radius from one phase to the other. As in [12, 20], we are fixing the initial mutation radius to some carefully chosen optimal value. The particular mutation adaptation rule is made clear in the following.

Theorem 5.3 *Assume the $(1+1)$ EA with uniform mutation inside the sphere of radius r minimising the n-dimensional SPHERE function starts at distance d such that $d \gg \sqrt{n}$, and let k be fixed in $\Theta(\ln(d/\sqrt{n}))$. For all $t \geq 1$, phase t of the algorithm is defined by mutation radius $r_t := d/2^{tk}$, maximal distance to optimum $d/2^{(t-1)k}$ and minimal distance to optimum $d/2^{tk}$. Then the expected convergence time of the algorithm is in $\Omega(1)$ and in $O(\sqrt{n})$.*

Proof.
In a single phase of the algorithm, under constant mutation radius r, expected one-step progress increases the closer we get to the optimum. To see that, consider two extreme positions of the current EA: far away - at distance $d \gg r$ - Figure 1, and close-by - at distance r - Figure 2, respectively.

Assume for the moment that $r = R \approx \sqrt{n/2\pi e}$, the radius of the the n-dimensional sphere of volume 1. For large n, one can approximate the intersection of the two spheres in the first case by the semi-sphere of radius r, then proposition 5.1 provides the value of the centroid as $A = 1/\pi\sqrt{e}$. In the second case, the centroid of the intersection is $R/2$, due to symmetry of the figure.

Consider next the more general situation where $r \neq R$. The centroids will change, according to corollary 5.2, and respectively to symmetry of the figure, into

$$\begin{array}{ccc} A & \longrightarrow & A\dfrac{r}{R} = \dfrac{r\sqrt{2}}{\sqrt{n\pi}} \\ \dfrac{R}{2} & \longrightarrow & \dfrac{r}{2}. \end{array}$$

[9] Because of the symmetry of the SPHERE, we can assume without loss of generality that we rotate the axes at each iteration such that the current EA position lies on Ox_1.

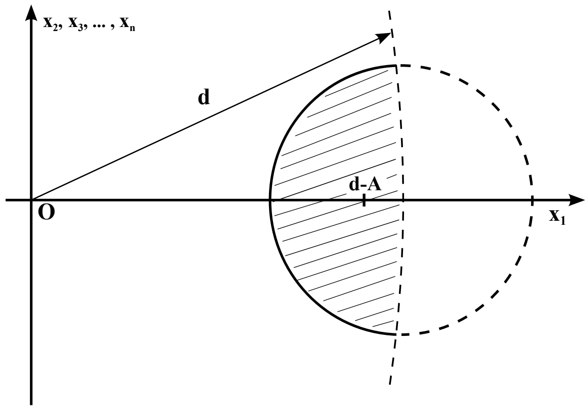

Figure 1: Mutation sphere far away

Accordingly, the expected values of the one-step progress will also change into

$$
\begin{aligned}
A &\quad\longrightarrow\quad A\frac{r}{RV_n(r)} = \frac{r\sqrt{2}}{\sqrt{n\pi}\,V_n(r)} \\
\frac{R}{2} &\quad\longrightarrow\quad \frac{r}{2V_n(r)}.
\end{aligned}
$$

Comparing the two extreme cases, one can easily see that for $n > 8/\pi \approx 2.5$, the value of the centroid far away is less than the value of the centroid close-by. The same holds for expected values, and hence the announced monotonic behavior - true for each algorithmic phase with constant mutation radius r. Summing up,

$$\frac{r\sqrt{2}}{\sqrt{n\pi}\,V_n(r)} \le E(X_t) \le \frac{r}{2V_n(r)}.$$

With this inequality in mind let us return to the original setting $r = r_t = d/2^{tk}$ and make $t = 1$, thus $r = r_1 = d/2^k$, k constant to be fixed later. For large d, we can use theorem 3.1 to estimate the expected hitting time of distance r_1, provided the algorithm starts at distance d:

$$
\begin{aligned}
\frac{2V_n(r_1)}{r_1} \;\le\; \frac{E\left(T_{d-\frac{d}{2^k}}\right)}{d-\frac{d}{2^k}} \;\le\; \frac{\sqrt{n\pi}\,V_n(r_1)}{r_1\sqrt{2}} \quad \Leftrightarrow \\
\frac{2dV_n(r_1)}{r_1}\left(1-\frac{1}{2^k}\right) \;\le E\left(T_{d-\frac{d}{2^k}}\right) \;\le\; \frac{d\sqrt{n\pi}\,V_n(r_1)}{r_1\sqrt{2}}\left(1-\frac{1}{2^k}\right)
\end{aligned}
$$

or, under the common assumption $1 \gg 1/d^k$, after removing the parentheses and substituting the value of r_1

$$2^{k+1}V_n(r_1) \;\le\; E\left(T_{d-\frac{d}{2^k}}\right) \;\le\; \sqrt{\frac{n\pi}{2}}\,2^kV_n(r_1). \tag{16}$$

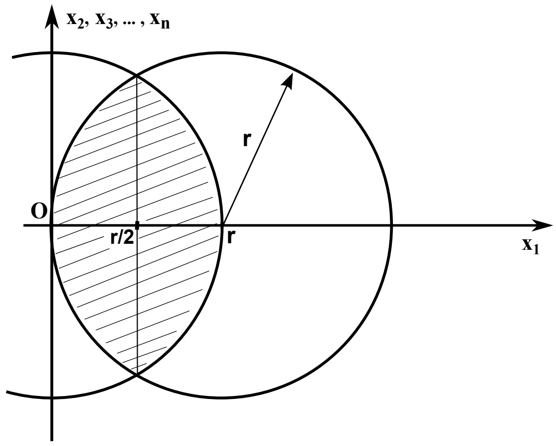

Figure 2: Mutation sphere close-by

At this point one can fix k such that $2^k V_n(r_1) = 1$, equivalent to $C_n d^n = 2^{(n-1)k}$ - see Appendix B. From relation (24) and Stirling's formula (27), we obtain the solution k of the exponential equation as

$$k \approx \frac{n}{(n-1)2\ln 2} \ln\left(\frac{d\sqrt{2\pi e}}{\sqrt{n}}\right).$$

The value found for k is in $\Theta(\ln(d/\sqrt{n}))$, while the prior confinement '2^k *large*' is equivalent to $d \gg \sqrt{n}$. Under fulfillment of these conditions we have $2^k V_n(r_1) = 1$, which simplifies inequality (16) to:

$$2 \;\leq\; E\left(T_{d-\frac{d}{2^k}}\right) \;\leq\; \sqrt{\frac{n\pi}{2}}. \tag{17}$$

Let us make now $t = 2$. Mutation radius is $r_2 = d/2^{2k}$ and a derivation similar to the one leading to (17) provides

$$2\frac{1}{2^{kn}} \;\leq\; E\left(T_{\frac{d}{2^k}-\frac{d}{2^{2k}}}\right) \;\leq\; \sqrt{\frac{n\pi}{2}}\frac{1}{2^{kn}} \tag{18}$$

and recursively, after t steps,

$$2\frac{1}{2^{tkn}} \;\leq\; E\left(T_{\frac{d}{2^{tk}}-\frac{d}{2^{(t+1)k}}}\right) \;\leq\; \sqrt{\frac{n\pi}{2}}\frac{1}{2^{tkn}}. \tag{19}$$

All we have to do now is sum up relations (17)-(19) and let $t \to \infty$. The middle term converges to $E(T_d)$, the expected hitting time of distance d from the starting point of the algorithm - recall that $E(T_d) < \infty$ according to proposition 2.7 - which is exactly the convergence time of our $(1+1)$ EA. As for the left and right hand terms, they each sum up to the geometrical series with ratio $1/2^{kn}$, which converges to $1/(1-1/2^{kn})$ as $t\to\infty$. Thus

$$2\frac{1}{1-\frac{1}{2^{kn}}} \;\leq\; E\left(T_d\right) \;\leq\; \sqrt{\frac{n\pi}{2}}\frac{1}{1-\frac{1}{2^{kn}}}.$$

By removing again the small term $1/2^{kn}$ from both sides we are left with

$$2 \;\leq\; E\left(T_d\right) \;\leq\; \sqrt{\frac{n\pi}{2}}$$

thus convergence time of the $(1+1)$ EA is in $\Omega(1)$ and in $O(\sqrt{n})$. □

Compared to the main results in [12, 13], where convergence time of the $(1+1)$ EA minimising SPHERE using *normal* mutation and the 1/5 adaptation rule is estimated to be in $\Omega(poly(n))$, one may find the result of theorem 5.3 surprising. We claim that the substantially better convergence time obtained in this section is not a consequence of the special mutation we used (uniform instead of normal), but of the more accurate theoretical modelling. The proof of this conjecture is deferred to a future paper.

Acknowledgements

This work was done while the first author was visiting the Technical University of Dortmund. Financial support from *DAAD - German Academic Exchange Service* under Grant A/10/05445, and scientific support from Chair Computer Science XI are gratefully acknowledged.

Appendix A: Direct proof of Lemma 1.2 - normal mutation

Assume that for each t, the continuous part of X_t is a truncated normal variable, with support $(0, b_t) \subset (0, \infty)$ - as is the case for the $(1+1)$ EA with normal mutation acting on the SPHERE. We also assume, under hypothesis H_2, the existence of constants $\mu_1, \mu_2, \sigma_1, \sigma_2$ such that $0 < \mu_1 \leq E(X_t) \leq \mu_2$ and $0 < \sigma_1^2 \leq Var(X_t) \leq \sigma_2^2$ for all t. Note that the existence of σ_1 was not required in H_2, yet working with mutation variance bounded from below is not atypical for practical EAs.

According to [14] p.156, e.g., a (doubly) truncated normal variable with support (a, b) has p.d.f.

$$f_{a,b,\mu,\sigma}(x) = f\left(\frac{x-\mu}{\sigma}\right) \cdot C_{a,b} \cdot I_{a,b}(x), \tag{20}$$

where f is the p.d.f. of the standard normal distribution, $C_{a,b}$ is a scaling factor accounting for the degree of truncation - see (21) below - and $I_{a,b}(x)$ is the indicator function, $I_{a,b}(x) = 1$ if $a < x < b$ and zero otherwise. This is the most common type of truncation, but unfortunately not the type occurring in modeling the continuous EA. In our case the continuous part of X_t is truncated from the normal distribution, yet not scaled up. Instead, the removed probability is charged onto a single point, the EA position at iteration $(t-1)$, so the p.d.f. of the continuous part of X_t is

$$f^*_{b_t,\mu_t,\sigma_t}(x) = f\left(\frac{x-\mu_t}{\sigma_t}\right) \cdot I_{0,b_t}(x).$$

Thus for any α with $0 < \alpha < \mu_1$

$$\begin{aligned} P\{X_t \geq \alpha\} &= \int_{\alpha}^{b_t} f^*_{b_t,\mu_t,\sigma_t}(x)dx = \int_{\alpha}^{b_t} f\left(\frac{x-\mu_t}{\sigma_t}\right)dx = \\ &= \frac{1}{C_{\alpha,b_t}} \int_{\alpha}^{b_t} f_{\alpha,b_t,\mu_t,\sigma_t}(x)dx = \frac{1}{C_{\alpha,b_t}} \end{aligned}$$

and all comes down to finding an upper bound for C_{α,b_t}, independent of t. According to [14], the scaling factor reads

$$C_{\alpha,b_t} = \frac{1/\sigma_t}{\Phi\left(\frac{b_t-\mu_t}{\sigma_t}\right) - \Phi\left(\frac{\alpha-\mu_t}{\sigma_t}\right)} \tag{21}$$

where Φ denotes the standard normal c.d.f.. We have next

$$\begin{aligned} &\frac{1}{\sigma_t} \leq \frac{1}{\sigma_1} \\ &\Phi\left(\frac{b_t-\mu_t}{\sigma_t}\right) \geq \Phi(0) = \frac{1}{2} \\ &\Phi\left(\frac{\alpha-\mu_t}{\sigma_t}\right) = 1 - \Phi\left(\frac{\mu_t-\alpha}{\sigma_t}\right) \end{aligned}$$

and hence

$$C_{\alpha,b_t} \leq \frac{1/\sigma_1}{\frac{1}{2} - 1 + \Phi\left(\frac{\mu_t-\alpha}{\sigma_t}\right)} \leq \frac{1/\sigma_1}{\Phi\left(\frac{\mu_1-\alpha}{\sigma_2}\right) - \frac{1}{2}} < \infty$$

which yields the constant $\beta > 0$ as the inverse upper bound of C_{α,b_t}

$$\beta = \sigma_1 \cdot \left[\Phi\left(\frac{\mu_1-\alpha}{\sigma_2}\right) - \frac{1}{2}\right].$$

Appendix B: Hypersphere Volumes and Centroids

According to Li [15], the volume of an n-dimensional hypersphere (hereafter *n-sphere*) of radius r is

$$V_n(r) = C_n r^n, \tag{22}$$

where the coefficient is given by one of two formulas:

$$C_n = \frac{2^{\frac{n+1}{2}} \pi^{\frac{n-1}{2}}}{n!!}, \text{ when n is odd, and} \tag{23}$$

$$C_n = \frac{\pi^{\frac{n}{2}}}{(\frac{n}{2})!}, \text{ when n is even.} \tag{24}$$

The function $n!!$ in (23) is the *double factorial*, defined as the product of *every other number* from n down to either 2 or 1 (depending on n's parity). The following formulas connect the double factorial with the regular factorial:

$$(2k+1)!! = \frac{(2k+1)!}{2^k k!} \tag{25}$$

$$(2k-1)!! = \frac{(2k)!}{2^k k!} \tag{26}$$

Note that a similar formula exists for the double factorial of an even number $2k$, but it is clear from (23) and (24) that we only need it for odd numbers. To capture the limiting behavior of the factorials, we use Stirling's formula

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \tag{27}$$

in the sense that the ratio of the LHS and RHS tends to 1 when $n \to \infty$. e is the base of the natural logarithms, $e \approx 2.718$. We shall also need the stronger form

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left[1 + O\left(\frac{1}{n}\right)\right], \tag{28}$$

which shows that the convergence is of the order $1/n$.

Without loss of generality, we take the axis Ox to pass through the centroids of our figures, so the coordinate of the centroid of any figure is given by

$$\overline{X_C} = \frac{\int x\,dV}{\int dV} = \frac{\int xA(x)\,dx}{\int dV}, \tag{29}$$

where dV is the infinitesimal element of (hyper)volume perpendicular to Ox, and $A(x)$ is the (hyper)area of the intersection between the figure and the (hyper)plane projecting at x.

Exact Formulas

Consider an n-sphere of radius $r = 1$, centered at the origin. In (29), the intersection $A(x)$ is an $(n-1)$-sphere of radius $\sqrt{1-x^2}$, according to Pythagora's theorem. In the notation from (22), we have $A(x) = V_{n-1}(\sqrt{1-x^2})$. We shall need the centroid of the positive semi-sphere, so the integration limits are 0 and 1. Substituting all this into (29), we have

$$\overline{X_C} = \frac{\int_0^1 x\,V_{n-1}(\sqrt{1-x^2})dx}{\frac{1}{2}V_n(1)}$$

In the above expression, we use (22) and factor out the constant coefficients to obtain:

$$\overline{X_C} = \frac{2C_{n-1}}{C_n} \int_0^1 x\,(1-x^2)^{\frac{n-1}{2}}\,dx$$

Integration by parts shows that the integral is $\frac{1}{n-1}$, which leads to

$$\overline{X_C} = \frac{2C_{n-1}}{(n+1)C_n} \tag{30}$$

We now substitute into (30) the expressions for the odd/even coefficients from (23) and (24) to obtain the following *exact* expressions for the centroid of the semi-hypersphere:

$$\overline{X_C} = \frac{n!!}{2^{\frac{n-1}{2}}(n-1)(\frac{n-1}{2})!}, \text{ when n is odd, and} \tag{31}$$

$$\overline{X_C} = \frac{2^{\frac{n+2}{2}}(\frac{n}{2})!}{\pi(n+1)(n-1)!!}, \text{ when n is even.} \tag{32}$$

The above formulas can be easily checked for the first three values of n: For $n = 1$, (31) gives 1/2, which is indeed the centroid of the segment [0..1]. For $n = 2$, (32) gives $4/3\pi$, which is the centroid of the positive half-disc of radius 1. For $n = 3$, (31) gives 3/8, which is the centroid of the positive semi-sphere of radius 1.

If we now remove the restriction $r = 1$, all the n-spheres are scaled by a factor of r. It is well-known that the centroid, being the first-order moment of the figure, also gets scaled by r, so we simply multiply (31) and (32) by r. We also make the change of variable $n = 2k + 1$ in (31) and use formula (25) to obtain

$$\overline{X_C} = \frac{(2k+1)!}{2^{2k+1}k(k!)^2}\,r, \text{ when n} = 2k+1 \tag{33}$$

Similarly, we make the change of variable $n = 2k$ in (32) and use formula (26) to obtain

$$\overline{X_C} = \frac{2^{2k+1}(k!)^2}{\pi(2k+1)(2k)!}\,r, \text{ when n} = 2k \tag{34}$$

Limits for the Centroid of a Semi-Hypersphere

We apply Stirling's formula (27) in (33) and (34) and find that, *for n both odd and even*, the limit of the centroid is

$$\lim_{n\to\infty} \overline{X_C} = \lim_{n\to\infty} \frac{1}{\sqrt{\pi n}}\,r = \lim_{n\to\infty} \sqrt{\frac{2}{\pi n}}\,r \tag{35}$$

Two cases are of particular importance:

i If r is kept constant as $n \to \infty$, the centroid approaches the origin as $1/\sqrt{n}$.

ii If, on the other hand, we keep the *volume* of the n-sphere constant as $n \to \infty$, the radius also changes, according to (22). If $V_n(r) = 1$, we have

$$r = \frac{1}{(C_n)^{\frac{1}{n}}}$$

As we did in the previous section, we use the odd-even expressions for C_n (23) and (24), we change the variable $n = 2k + 1$ for n odd and $n = 2k$ for n even, use the Stirling approximation and take the limit $n \to \infty$. After some work, we find that, *for n both odd and even*, the limit of r is

$$\lim_{n\to\infty} r = \lim_{n\to\infty} \sqrt{\frac{n}{\pi e}} = \lim_{n\to\infty} \sqrt{\frac{n}{2\pi e}}$$

We proved the following result:
When $n \to \infty$, the radius of an n-sphere of constant volume also tends to ∞, with the order of $\sqrt{n}$.
Substituting into (35), we obtain the result of proposition 5.1:
When $n \to \infty$, the centroid of a semi-n-sphere of volume 1 tends to a fixed value:

$$\lim_{n\to\infty} \overline{X_C} = \frac{1}{\pi\sqrt{e}} \approx 0.193 \qquad (36)$$

Interestingly, the assumption that $V = 1$ can also be dropped when $n \to \infty$, since the limit of $V^{1/n}$ is 1. It follows that the limit $\frac{1}{\pi\sqrt{e}}$ holds for any constant (with respect to n) volume V of the n-sphere!

Since the convergence in Stirling's formula is $O(1/n)$, the convergence of the centroid to the limit $\frac{1}{\pi\sqrt{e}}$ is also $O(1/n)$.

References

[1] Agapie, A.; Genetic Algorithms: Minimal Conditions for Convergence, in AE'97, Hao, J. K., Lutton, E., Ronald, E., Schoenauer, M. and Snyers, D. (Eds.), LNCS 1363, Springer, Berlin 183–193 (1998)

[2] Agapie, A.: Theoretical analysis of mutation-adaptive evolutionary algorithms. Evolutionary Computation **9**, 127–146 (2001)

[3] Agapie, A., Agapie, M.: Transition Functions for Evolutionary Algorithms on Continuous State-Space, Journal of Mathematical Modelling and Algorithms **6(2)**, 297–315 (2007)

[4] Agapie, A.: Estimation of Distribution Algorithms on Non-Separable Problems, International Journal of Computer Mathematics, **87(3)**, 491–508 (2010)

[5] Beyer, H.-G.: The Theory of Evolution Strategies. Springer, Heidelberg (2001)

[6] Doerr, B., Goldberg, L.A.: Adaptive drift Analysis, in PPSN XI, Schaefer, R., Cotta, C., Kolodziej, J. and Rudolph, G. (Eds.), LNCS 6238, Springer, Berlin 51–81 (2010)

[7] Droste, S., Jansen, T., Wegener, I.: On the analysis of the (1+1) evolutionary algorithm. Theor. Comput. Sci. **276(1–2)**, 51–81 (2002)

[8] Gut, A.: Stopped Random Walks: Limit Theorems and Applications. Springer, New York (2009)

[9] Hajek, B.: Hitting-time and occupation-time bounds implied by drift analysis with applications. Advances in Applied Probability **13**, 502–525 (1982)

[10] He, J., Yao, X.: Drift analysis and average time complexity of evolutionary algorithms. Artificial Intelligence **127**, 57–85 (2001)

[11] He, J., Yao, X.: A study of drift analysis for estimating computation time of evolutionary algorithms. Natural Computing **3**, 21–35 (2004)

[12] Jägersküpper, J.: Analysis of a simple evolutionary algorithm for minimization in Euclidean spaces. In Proc. of the 30th Int. Colloquium on Automata, Languages and Programming (ICALP), LNCS 2719,, Springer, Berlin, 1068–1079 (2003)

[13] Jägersküpper, J.: How the (1+1) ES using isotropic mutations minimizes positive definite quadratic forms. Theoretical Computer Science **361**, 38–56 (2006)

[14] Johnson, N., Kotz, S., Balakrishnan, N.: Continuous Univariate Distributions: Vol.1. Wiley, New York (1994)

[15] Li, S.: Concise Formulas for the Area and Volume of a Hyperspherical Cap. Asian Journal of Mathematics and Statistics **4(1)**, 66–70 (2011)

[16] Oliveto, P.S., Witt, C.: Simplified Drift Analysis for Proving Lower Bounds in Evolutionary Computation. Algorithmica, DOI: 10.1007/s00453-010-9387-z (2010)

[17] Ross, S.: Applied Probability Models with Optimization Applications. Dover, New York (1992)

[18] Rudolph, G.: Convergence Analysis of Canonical Genetic Algorithms. IEEE Trans. on Neural Networks, **5**, 98–101 (1994)

[19] Rudolph, G.: Convergence of Evolutionary Algorithms in General Search Spaces. In Proc. of the 3rd IEEE Conf. on Evolutionary Computation, Piscataway, NJ. IEEE Press (1996)

[20] Rudolph, G.: Convergence Properties of Evolutionary Algorithms. Kovać, Hamburg (1997)

[21] Rudolph, G.: Stochastic Convergence, in Handbook of Natural Computing, Rozenberg, G., Bck, T. and Kok, J. (Eds.) Springer, Berlin (2010)

[22] Rudolph, G.: Evolutionary Strategies, in Handbook of Natural Computing, Rozenberg, G., Bck, T. and Kok, J. (Eds.) Springer, Berlin (2010)

[23] Schwefel, H.-P.: Evolution and Optimum Seeking. John Wiley, New York (1995)

[24] Williams, D.: Probability with martingales. Cambrigde University Press (1991)