Stochastic Modeling of Continuous Evolutionary Algorithms

Alexandru Agapie* Günter Rudolph† Gheorghita Zbaganu‡

Abstract

Evolutionary algorithms (EAs) acting on continuous space require a more sophisticated modeling than their discrete counterparts. Sharing with classical theory of evolution strategies only the interest for expected one-step progress, the proposed modeling is based on stochastic *renewal processes*. The new paradigm allows for global convergence results, as well as first hitting time computation for particular algorithms on particular fitness functions.

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 $P\{X_t = 0\} < 1$, or equivalently $P\{X_t > 0\} > 0$, for all t,

*Academy of Economic Studies, and Institute of Mathematical Statistics and Applied Mathematics, Bucharest, Romania, agapie@clicknet.ro

†Computer Science XI, Technical University Dortmund, Germany

‡Institute of Mathematical Statistics and Applied Mathematics,Bucharest, Romania

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, the conditions above are not sufficient for a consistent stochastic analysis, providing first hitting times for the algorithms in study. The fulfilment of either one of the following natural hypotheses will be also required.

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 $E(X) \leq \alpha \cdot P\{X < \alpha\}$.*

□

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

□

A somehow different 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:

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.*

□

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:

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

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[1]:

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

According to [7, 13], the process $\{T_d\}_{d>0}$ will be called a *renewal process*[2] 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 [13].

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

[2] 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.

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

□

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 $d > 0$.*

□

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 [13], $E(T_d)$, 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 summarizes 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 [9], yet outside the context of renewal processes.

Theorem 2.4 (Wald's Equation, [13] 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*

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

□

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 (**??**). 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 $E(M) = 1/p$.*

□

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

□

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 [7, 13], 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

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

□

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*[3] 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 4.

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, 14] - see [15] 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.

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

4 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 appeals more to the mathematician 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 [3, 9, 10, 14]. 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[4]

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

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.

[4] In order to avoid confusion, we shall use uppercase when referring to the fitness function, and lowercase when referring to the mutation operator.

Proposition 4.1 *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 $n \to \infty$*

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

□

Corollary 4.2 *If we multiply the radius of the sphere from proposition 4.1 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 *evolution strategies* - see e.g. [16]!

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 [5]. 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 [9, 14], 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 4.3 *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 $O(\sqrt{n})$ and in $\Omega(1)$.*

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 4.1 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 4.2, and respectively to symmetry of the figure,

[5] 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.

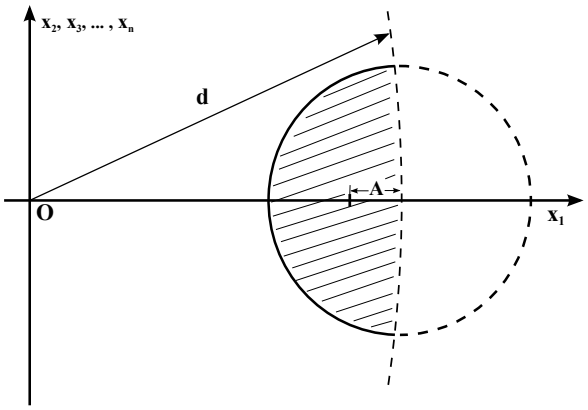

Figure 1: Mutation sphere far away

into

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

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

$$\begin{array}{lcl} A & \longrightarrow & A\dfrac{r}{RV_n(r)} = \dfrac{r\sqrt{2}}{\sqrt{n\pi}\,V_n(r)} \\ \dfrac{R}{2} & \longrightarrow & \dfrac{r}{2V_n(r)}. \end{array}$$

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,

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

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:

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

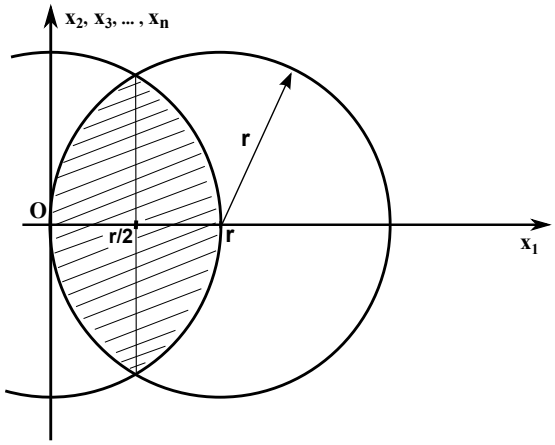

Figure 2: Mutation sphere close-by

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

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

At this point one can fix k such that $2^kV_n(r_1) = 1$, equivalent to $C_nd^n = 2^{(n-1)k}$ - see Appendix B. From relation (**??**) and Stirling's formula (**??**), 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 $2^kV_n(r_1) = 1$, which simplifies inequality (1) to:

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

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

$$2\frac{1}{2^{kn}} \;\leq\; E\left(T_{\frac{d}{2^k}-\frac{d}{2^{2k}}}\right) \;\leq\; \sqrt{\frac{n\pi}{2}}\frac{1}{2^{kn}} \tag{3}$$

and recursively, after t steps,

$$2\frac{1}{2^{tkn}} \;\leq\; E\left(T_{\frac{d}{2^{tk}}-\frac{d}{2^{(t+1)k}}}\right) \;\leq\; \sqrt{\frac{n\pi}{2}}\frac{1}{2^{tkn}}. \tag{4}$$

All we have to do now is sum up relations (2)-(4) 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 [9, 10], where convergence time of the $(1+1)$ EA minimising SPHERE using *normal* mutation and the 1/5 adaptation rule is estimated to be in $\Theta(poly(n))$, one may find the result of theorem 4.3 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 modeling.

Acknowledgement

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.

References

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

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

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

[4] Bienvenüe, A., Francois, O.: Global convergence for evolution strategies in spherical problems: some simple proofs and difficulties. Theor. Comput. Sci. **306**, 269–289 (2003)

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

[6] Fang, K.-T., Kotz, S., Ng, K.-W.: Symmetric Multivariate and Related Distributions. Chapman and Hall, London (1990)

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

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

[9] 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)

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

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

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

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

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

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

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