# Convergence of Evolutionary Algorithms on the $n$-Dimensional Continuous Space

Alexandru Agapie, Mircea Agapie, *Member, IEEE,* Günter Rudolph, *Member, IEEE,* and Gheorghita Zbaganu

***Abstract*—Evolutionary algorithms (EAs) are random optimization methods inspired by genetics and natural selection, resembling simulated annealing. We develop a method that can be used to find a meaningful tradeoff between the difficulty of the analysis and the algorithms' efficiency. Since the case of a discrete search space has been studied extensively, we develop a new stochastic model for the continuous $n$-dimensional case. Our model uses renewal processes to find global convergence conditions. A second goal of the paper is the analytical estimation of the computation time of EA with uniform mutation inside the (hyper)-sphere of volume 1, minimizing a quadratic function.**

***Index Terms*—Computational complexity, evolutionary computation, optimization, stochastic processes.**

## I. Introduction

AN ASTUTE reader of the theoretical developments in evolutionary algorithms (EAs) might feel let down by either the excessive restrictiveness of the assumptions under which the analysis is performed, or by its extreme generality, which means little practical usefulness. A precise tradeoff is needed in order to find a meaningful middle ground, and we find that the difficulty of the task is inversely proportional to the algorithms' efficiency. Building on probability theory rather than statistical physics, we aim at introducing a new stochastic modeling for continuous EAs, one based on renewal processes.

Our second goal is to analytically estimate the hitting time[1] of the (1 + 1)EA with uniform mutation inside the (hyper)-sphere of volume 1, minimizing the well-known *SPHERE* function[2]

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

Our novel approach can be traced back to the following ideas.

1) Rudolph [26], [27], and Dorea [14] for the global convergence analysis of evolutionary algorithms based on stochastic processes.
2) Beyer [7], [8], for the geometrically oriented approach and the integral calculus of (one-step) progress rate on the SPHERE model.
3) Jägersküpper [19], [20], for the analysis of atypical (that is, nonnormal) mutation operators, the application of Wald's equation and, again, the geometrical and integral calculus of progress rate on SPHERE.
4) Doerr and Goldberg [13], Oliveto and Witt [24], and references [11], [12], [17], [18] for the similarities to *drift analysis*, a method that provided important insights into the computational complexity of discrete EAs over the last decade.

Rudolph [26], [27] and Dorea [14] completed the global convergence analysis for the discrete, respectively, continuous case, using the classification of states from Markov chains, but the problem of computation times remained open. The theory of evolution strategies (ES) developed by Beyer has been criticized for using several approximations, not all mathematically rigorous: "A first approximation called the *first-order approximation* amounts to considering averaged instead of random behaviors. In a second kind of approximation, fluctuations are modelled as Gaussian noises which can be a crude approximation of the actual behavior" [10, p.270]. Actually, approximations are inevitable due to the analytical intractability of the one-step progress rate

$$\phi_{1+1} = E\left[d - \sqrt{(d-x)^2 + h^2}\right] \tag{1}$$

in case of normal mutation. $E$ stands for the expected value operator, and ((1)) defines progress as difference between distances to origin of current point $S$, and the next point of the algorithm $C$. As seen in Fig. 1, $d$ measures distance from origin to $S$, while distance to $C$ is expressed as $\sqrt{(d-x)^2 + h^2}$ via Pythagora's theorem in the triangle

Manuscript received May 19, 2012; revised September 16, 2012; accepted March 25, 2013. Date of publication May 3, 2013; date of current version September 11, 2013. This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PNII-ID-PCCE-2011-0015, and by the German Academic Exchange Service, DAAD, under Grant A/10/05445/2010. This paper was recommended by Associate Editor K.-C. Tan.

A. Agapie is with the Applied Mathematics Department, Bucharest University of Economic Studies, Calea Dorobantilor 15-17, Bucharest 010552, Romania (e-mail: agapie@clicknet.ro).

M. Agapie is with the Engineering and Computer Science Department, Tarleton State University, Box T-0930 Stephenville, TX 76402 USA (e-mail: agapie@tarleton.edu).

G. Rudolph is with the Computer Science Department, TU Dortmund University, Dortmund 44221, Germany (e-mail: guenter.rudolph@tu-dortmund.de).

G. Zbaganu is with the Institute of Mathematical Statistics and Applied Mathematics, Bucharest 050711, Romania (e-mail: zbagang@csm.ro).

Digital Object Identifier 10.1109/TCYB.2013.2257748

[1]For hitting time we apply the *stopping time* definition from stochastic processes; $T_d$ will denote the random time needed by the algorithm to cover a distance $d$ in the $n$-dimensional continuous space—see Section III.

[2]To avoid confusion, we use uppercase when referring to the fitness function, and lowercase when referring to the mutation operator.

with hypotenuse $SC$, with $h$ and $x$ standing for the other two sides.

In addition to Beyer's work, Jägersküpper proved (using a modification of *Wald's equation*) the following lower bound on the expected runtime of the (1+1)EA minimizing the SPHERE:

*Theorem 1.1 ([19]):* Let the $(1 + 1)$ES minimize SPHERE using isotropic.[3] mutations, and let $\mathbf{c} \in \Re^n$ denote the current search point[4] Independently of the mutation adaptation used the expected number of steps necessary to obtain a spatial gain of $\Theta(|\mathbf{c}|)$ is in $\Omega(n)$.

Building on previous analyses of atypical probabilistic algorithms, like *ball walk* and *hit-and-run* [23], Jägersküpper generalized the previous result to a wider class of black-box optimization algorithms, the ones using random directions [20]. All the $(1 + 1)$EAs with isotropic mutation—including the classical normal-mutation ES, and the uniform mutation inside the sphere (ball walk) algorithm studied in this paper—fall into that class. Focusing on the expected value of the angle $\alpha = \angle OSC$ in Fig. 1—rather than on the expected location of the next individual, Jägersküpper derived a closed form of $E(\cos\alpha)$ and narrow upper and lower bounds for $E(\sin\alpha)$. His conclusion is the same as that of Theorem 1.1: the expected number of steps to halve the approximation error has a lower bound linear in $n$.

The contribution of our paper is twofold:

1) a general stochastic modelling based on renewal process, which holds for population-based EAs and many classes of fitness functions;
2) an original method of estimating the one-step progress of the (1+1)EA on the SPHERE, using the geometrical method of centroids.

Section II underlines the theoretical similarity between normal and uniform mutation, then Section III introduces the new stochastic model of continuous EAs, valid for any elitist algorithm of type $(\mu + \lambda)$. Step by step, classical theory of renewal processes (initially developed for queueing systems) is adapted to our algorithmic paradigm, culminating in a theorem describing the asymptotics of the expected first hitting time. This type modeling is quite general, holding for EAs with populations consisting of more than one individual, and various fitness landscapes. In Section IV, we show how the stochastic characterization of the hitting time can be read in terms of drift analysis. The focus of the paper then shifts to the direct computation of convergence time for a particular algorithm acting on a particular landscape. An integral calculus of hypersphere volumes and centroids is developed in Section V, and the results are used in Section VI, together with the previous stochastic analysis, to compute lower and upper bounds on the convergence time of $(1 + 1)$ EA with uniform mutation on the SPHERE. Section VII concludes the paper. To improve readability, some of the more technical proofs are omitted.[5]

[3]See Section II for the definition of isotropic mutation.

[4]$|\mathbf{c}|$ is replaced in our notation by $d$.

[5]A complete version of the paper—all proofs included—is available upon request to the first author via e-mail.

## II. Uniform Versus Normal Mutation

One could argue that a theoretical analysis employing only uniform mutation is too simplistic to be practical. To answer such criticism we point out that the first hitting time analysis based on renewal processes performed in this paper does not rely on the particular form of the mutation operator; it only requires the first two moments (expected value and variance) of the random variable (RV) one-step progress. Depending on the complexity of the fitness function and of the mutation distribution, the moments of this RV could be computed either analytically or numerically.

From a stochastic point of view, the two distributions—normal (Gaussian), and uniform inside the sphere—are very similar. According to [16, p.28], they both belong to the same class of spherical distributions, i.e. they both can be represented as

$$\mathbf{x} \stackrel{d}{=} r \cdot \mathbf{u}$$

where $\mathbf{x}$ is the $n$-dimensional spherical random vector, $\mathbf{u}$ is the random vector distributed uniformly *on* the unit sphere surface in $\Re^n$, $r = ||x||$ is a nonnegative scalar RV independent of $\mathbf{u}$, and $\stackrel{d}{=}$ denotes the fact that two random vectors have the same distribution. Thus, a mutation according to a multivariate normal distribution can be interpreted as a step in random direction $\mathbf{u}$ with random step length $r$ [28]. The same holds for the uniform mutation inside the sphere, with an only slightly different step length; for the normal mutation $\mathbf{x} = N(0, \sigma^2 I)$, an analytical derivation of the first two moments of $r$ yields [26, p.21]

$$E(r) \approx \sigma\sqrt{n - \frac{1}{2} + \frac{1}{16n}}, \quad Var(r) \approx \frac{\sigma^2}{2}\left(1 - \frac{1}{8n}\right)$$

indicating that the random step length actually does not vary strongly [28].

The concept of spherical distribution has quite a history in EA literature under a different name: isotropic mutation.

*Definition 1 ([19]):* For $m \in \Re^n$ let $|m|$ denote its length/$L_2$-norm and $\hat{m} := m/|m|$ the normalized vector. The random mutation vector $m$ is isotropically distributed iff $|m|$ is independent of $\hat{m}$ and $\hat{m}$ is uniformly distributed upon the unit hypersphere $U := \{x \in \Re \mid |x| = 1\}$.

It follows that the uniform mutation employed in this paper is isotropic, just like the normal mutation of classical EA and ES theory.

## III. Continuous EA is a Renewal Process

This part of our analysis relies heavily on stochastic processes and applies to any optimization problem and finite population size EA fulfilling the condition $H$ below.

The search space considered here is $\Re^n$. The simplest type of continuous algorithm one can imagine is the so-called $(1 + 1)$EA, with population consisting of only one individual. In each iteration a new individual (*offspring*) is produced via *mutation*, and the best of the old and new individuals is selected. Such preservation of the best individual found so far from a generation to another is called *elitism*. The mutation's

effect is modeled by a probability distribution, set to be *normal* in practice and most theoretical studies[6].

If, instead of generating *one* offspring per iteration, we generate $\lambda$ (with $\lambda > 1$) and then we select the best from $1+\lambda$ (old plus new), we call the resulting algorithm $(1+\lambda)$EA. Furthermore, if we keep $\mu$ individuals in the current population, $\mu > 1$, and generate $\lambda$ offspring at each iteration via mutation and other operations involving at least two different parents (*crossover*), and then select the best $\mu$ from the whole set of $\mu + \lambda$ individuals, we call the algorithm $(\mu + \lambda)$EA. As explained below, this is the natural framework for the renewal process modeling of continuous EAs.

For each $t = 0, 1, 2, \ldots$, let $P_t$ be the RV (best individual from) EA population at iteration $t$. Then $\{P_t\}_{t\geq 0}$ is a stochastic process on $\Re^n$. We also define the (1-D) distance to optimum $d : \Re^n \leftarrow \Re_+$, that is, to $0 := (0, \ldots, 0)$ since we are minimizing. As generally the case with probabilistic algorithms on continuous space, we say that convergence is achieved at iteration $t$ if the algorithm has entered an $\epsilon$-vicinity of 0 for some fixed $\epsilon$, $0 \leq d(P_t) < \epsilon$.

Define the stochastic process $\{X_t\}_{t\geq 1}$ given by

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

In our EA framework, $X_t$ stands for the (relative) progress of the algorithm in one step, namely from the $(t-1)$st iteration to the $t$th. Due to EA's elitism, $\{X_t\}_{t\geq 1}$ are *nonnegative* RVs, and we also assume they are independent. Each $X_t$ has a point mass (singular, or Dirac measure) at, zero accounting for the event where there is no improvement from $P_{t-1}$ to $P_t$, and a continuous part accounting for the actual progress toward the optimum. The latter is, for example, a truncated uniform or normal distribution. A second natural assumption is $P\{X_t = 0\} < 1$, or equivalently $P\{X_t > 0\} > 0$, for all $t$, otherwise the algorithm would not converge. This does not require a progress at each iteration, but only a strictly positive *probability* of such progress.

To provide first hitting times for the algorithms we study, we also require the following natural condition:

| | |
|---|---|
| $H$: | $\{X_t\}_{t\geq 1}$ are nonnegative, independent RVs 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 $t$. |

It is easy to see that $H$ is fulfilled by the $(1+1)$EA with many types of fitness functions; $H$ simply means fixed positive upper and lower bounds for the expected one-step progress, and a fixed positive upper bound for the variance of the one-step progress. Assuming the normal form $N(0, \sigma^2 I)$ for the mutation operator, elitism truncates the normal RV asymmetrically, and the truncation differs with the position of $P_{t-1}$ within the fitness landscape. Consider, for example, the $(1 + 1)$EA on the SPHERE. Then $X_t$ represents the truncation of $N(0, \sigma^2 I)$ inside the sphere centered in the origin with radius $d(P_{t-1})$. The closer we are from the origin, the more we truncate the normal variable, so the RVs $\{X_t\}_t$ are not identical. All that $H$ requires is that they be uniformly bounded with strictly positive margins.

Out of the three margins, the most difficult to achieve is undoubtedly $\mu_1$, the lower bound on expected value. Let us see how this margin can be obtained on the SPHERE: if we impose a fixed $\epsilon$-neighborhood of the origin, call it $V_\epsilon$, $\epsilon > 0$, as target zone (that is, the algorithm converges and stops if it enters this zone), then the truncation volume cannot be smaller than $V_\epsilon$. This provides the lower bound $\mu_1$.

On a particular fitness function, if $H$ holds for the $(1+1)$EA, it will also hold for the $(1+\lambda)$ algorithm. To see this, consider again the fragile margin $\mu_1$. Since the one-step progress is always larger for the $(1 + \lambda)$ algorithm than for the $(1 + 1)$ version—the chances for a bigger jump increase when we do $\lambda$ attempts instead of only one—the same holds for the expected value, thus the lower bound that works for the $(1+1)$ algorithm also works for $(1+\lambda)$. Similar reasoning carries the conclusion to $(\mu + \lambda)$EAs, but to make the proof rigorous in case of multiindividual algorithms, one needs *order statistics*, a technique often used in EA theory [4], [8], [26].

One should note that condition $H$ is flexible enough to cover the (adaptive) case of different mutation rates at different algorithmic iterations. For example, $H$ describes families of distributions that are all normal, or all uniform, but with parameters ranging within certain positive bounds.

On the other hand, the limitations imposed by the condition $H$ to the subsequent analysis are twofolded which are discussed as follows.

1) $X_t \geq 0$ implies we are considering only elitist EAs, leaving actually no room for the study of deceptive landscapes. Yet we expect that the assumption may be dropped in further analysis, on the expense of a deeper look into stochastic processes.
2) $E(X_t) \geq \mu_1 > 0$ allows adaptive mutation rates, but not mutation rates that are decreasing to zero. That precludes the accomplishment of $O(n)$ convergence time on the SPHERE in Theorem 6.2, but also recommends the RIDGE landscape [9] as a more suitable case study.

It is shown below that $H$ yields a stronger restriction on the progress probabilities than the already stated $P\{X_t > 0\} > 0$ for all $t$. We first need some general results from probability theory.

*Lemma 1:* If $X$ is a positive RV and $\alpha > 0$ such that $P\{X \geq \alpha\} = 0$, then $E(X) \leq \alpha \cdot P\{X < \alpha\}$.

*Lemma 2:* $H \Rightarrow$ there exist $\alpha, \beta > 0$ such that $P\{X_t \geq \alpha\} \geq \beta$ for all $t$.

By summing up the relative progress at each iteration we obtain $S_t$, the overall progress in $t$ iterations

$$S_t = X_1 + \cdots + X_t. \tag{3}$$

*Remark 1:* Under hypothesis $H$, the stochastic process $\{S_t\}_{t\geq 1}$ is an *atypical* discontinuous random walk, since RVs $X_t$ are independent but not identically distributed. One could speak in this case of an *inhomogeneous* random walk.

[6]For the second part of the paper, the estimation of hitting times, normal mutation will be replaced by uniform mutation inside the sphere.

From ((2)) we obtain

$$S_t = d(P_0) - d(P_1) + d(P_1) - \ldots - d(P_t) = \\ = d(P_0) - d(P_t) \qquad t = 1, 2, \ldots .$$

*Remark 2:* 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 $\delta$ 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 RV, 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 a certain positive threshold $d$[7]

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

Following [25], the process $\{T_d\}_{d>0}$ will be called a renewal process,[8] and it has 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 toward the optimum. After each renewal the process (the EA) starts over again.

### A. First Hitting Time

*Proposition 1:* Under hypothesis $H$, the first hitting time of the continuous EA is finite with probability 1.

*Proof:* Under $H$, the strong law of large numbers for independent *nonidentical* RVs[9] yields

$$\mu_1 \leq \frac{S_t}{t} \leq \mu_2 \qquad \text{with probability 1.} \qquad (4)$$

■

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.

*Definition 2:* An integer valued positive RV $T$ 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 results:

*Lemma 3:* $T_d$ defined as above is a stopping time for $\{X_t\}_{t\geq 1}$, for any $d > 0$.

The first hitting time of a distance $d$ from the starting point is greater than $t$ if and only if the $t$th iteration yields a point situated at distance less than or equal $d$:

*Lemma 4:*

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

According to [25], $E(T_d)$, the 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 $\epsilon$ 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 RVs $X_t$, $S_t$, and $T_d$ for continuous EAs:

| RV | Interpretation |
|---|---|
| $X_t$ | $1-D$ progress between the $(t-1)$st and the $t$th iteration |
| $S_t$ | overall progress up to the $t$th 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 [19], yet outside the context of renewal processes.

*Theorem 3.1 (Wald's equation, [25], p.38):* If $\{X_t\}_{t\geq 1}$ are independent and identically distributed RVs 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, we need some limit theorems from renewal theory. First we reformulate Wald's equation in terms of inequalities:

*Theorem 3.2 (Wald's Inequality):* If $\{X_t\}_{t\geq 1}$ are independent, nonnegative, $\mu_1 \leq E(X_t) \leq \mu_2$ for all $t$ and $T$ is a stopping time for $\{X_t\}_{t\geq 1}$, then

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

Apparently, one could conclude that whenever Wald's inequality is applied, $H$ may be replaced by the simplified hypothesis of the previous theorem. 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 also need $E(T) < \infty$ for our analysis, and that is proved under the continuous EA paradigm in proposition 2 below, relying strongly on Lemma 2, which in turn does not work unless all requirements in $H$ are fulfilled.

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

We need a simple result on the expected value of the geometric distribution:

*Lemma 5:* If the RVs $Z_1, Z_2, \ldots$ are independent and identically distributed (i.i.d.) as $Z = \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$, and the stopping time $M = \inf\{m \mid Z_1 + \cdots + Z_m = 1\}$, then $E(M) = 1/p$.

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

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

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

*Proposition 2:* Under hypothesis $H$, 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 $H$, the existence of $\alpha$ and $\beta$ is guaranteed by $H$ itself and Lemma 2. Define a related renewal process $\{\overline{X}_t\}_{t\geq 1}$ by truncating each $X_t$ to

$$\overline{X}_t = \begin{pmatrix} 0 & \alpha \\ 1-\beta & \beta \end{pmatrix}. \tag{5}$$

Note that $\overline{X}_t$ does not depend on $t$ anymore. Also, under hypothesis $H$, we have $\overline{X}_t \leq \overline{X}'_t$, where

$$\overline{X}'_t = \begin{cases} 0 & \text{if} \quad X_t < \alpha \\ \alpha & \text{if} \quad X_t \geq \alpha \end{cases}. \tag{6}$$

Let us now define $\overline{T}_d = \inf\{t \mid \overline{X}_1 + \cdots + \overline{X}_t > d\}$. For the related process, successive iterations can move the algorithm only along the lattice $d = t\alpha,\ t = 0, 1, 2 \ldots$. The number of iterations required for a success (a real jump of length $\alpha$) are independent RVs with mean $1/\beta$. To prove this, apply Lemma 5 for the RV $Z_t = \overline{X}_t/\alpha$ which registers 1 for a success and 0 for a stagnation of the EA from iteration $(t-1)$ to iteration $t$. We have

$$E(\overline{T}_d) \leq \frac{[d/\alpha]+1}{\beta} < \infty \tag{7}$$

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

### B. Expected Hitting Time

The expression $1/E(X_t)$ is often called the *progress rate* between the $(t-1)$st and the $t$th iteration. Following the general theory of renewal processes [25], 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 $H$ through proposition 2 .

*Theorem 3.3:* Under hypothesis $H$ we have, as $d \to \infty$

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

*Proof:* Under $H$, $E(T_d)$ is finite due to proposition 2 and Wald's *inequality* 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

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

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 $t = 1, 2, \ldots$

$$\overline{X}_t = \begin{cases} X_t & \text{if} \quad X_t \leq M \\ M & \text{if} \quad X_t > M. \end{cases} \tag{10}$$

Let $\overline{S}_t = \overline{X}_1 + \cdots + \overline{X}_t$ and $\overline{T}_d = \inf\{t \mid \overline{S}_t > d\}$. For each positive integer $M$ the process $\overline{X}_t$ is nonnegative. Since $0 \leq \overline{X}_t \leq X_t$, we have for each $M > 0$

$$0 \leq \mu_M^1 \leq E(\overline{X}_t) \leq \mu_2 \qquad \text{for all } t \tag{11}$$

where $\mu_M^1$ does not depend on $t$ and is defined for each $M$ by

$$\mu_M^1 = \min\{\mu_1, \inf_t E(\overline{X}_t)\}. \tag{12}$$

We show that $\mu_M^1 \to \mu_1$ as $M \to \infty$. Reasoning by contradiction, assume nonconvergence. As $\mu_M^1 \leq \mu_1$ always holds, for any $\epsilon > 0$, there are infinitely many $M_k$ outside the $\epsilon$-vicinity of $\mu_1$, the interval $(\mu_1 - \epsilon, \mu_1]$. According to the definition of $\mu_M^1$, we have for an arbitrary $\epsilon > 0$

$$E(\overline{X}_t) < \mu_1 - \epsilon \ \text{ for all } t \text{ and all } \{M_k\}_{k\geq 1}. \tag{13}$$

However, if we apply Lebesque's dominated convergence theorem[10] 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

$$E(\overline{X}_{t_0}) \to E(X_{t_0}) \qquad \text{as } M_k \to \infty, \tag{14}$$

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

$$E(\overline{X}_{t_o}) \geq \mu_1 - \epsilon \ \text{ for all } M_k \text{ with } k \geq k_0 \tag{15}$$

which obviously contradicts ((13)).

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

$$0 < \mu_M^1 \leq E(\overline{X}_t) \leq \mu_2 \ \text{ for all } t \text{ and } M \geq M_0. \tag{16}$$

It is easy to see that, for sufficiently large $M$

$$0 \leq a_t \leq \alpha < \mu_M^1 \leq E(\overline{X}_t) \leq \mu_2 \ \text{ for all } t \tag{17}$$

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

$$\mu_M^1 E(\overline{T}_d) \leq E(\overline{S}_{\overline{T}_d}) \leq \mu_2 E(\overline{T}_d). \tag{18}$$

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

$$\mu_M^1 E(T_d) \leq \mu_M^1 E(\overline{T}_d) \leq E(\overline{S}_{\overline{T}_d}) < d + M. \tag{19}$$

We obtain

$$\limsup_{d\to\infty} \frac{E(T_d)}{d} \leq \frac{1}{\mu_M^1} \tag{20}$$

and letting $M \to \infty$ yields

$$\limsup_{d\to\infty} \frac{E(T_d)}{d} \leq \frac{1}{\mu_1} \tag{21}$$

which together with ((9)) completes the proof. ■

As one can see from the proof of Theorem 3.3, the left-hand side of the inequality—the one giving a lower bound

[10]If $|X_n| \leq Z$ such that $E(Z) < \infty$ and $X_n \to X$ *in probability*, then $E(X_n) \to E(X)$—see [29, p.59].

on $E(T_d)$—is a simple consequence of Wald's inequality. 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 3.3 says that the expected average[11] hitting time is bounded, under hypothesis $H$, 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, optimizing 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.

The analysis performed so far on *continuous* EAs regarded as renewal processes is similar to the Markov chain analysis of *discrete* EAs performed in [2], [3], [26]; see [27] for the state-of-art in stochastic convergence for discrete EAs. It completes the theoretical analysis of convergence of discrete EAs, opening the door for particular estimations of local progress rates $\mu_1$ and $\mu_2$. We shall use some of the previous well-known results in the remainder of the paper. However, first let us digress and see how Theorem 3.3 extrapolates the drift theorems of discrete EAs.

## IV. Drift Analysis

Drift analysis is relatively old within probabilistic optimization theory, yet it was only recently introduced as a powerful tool in the study of convergence of evolutionary algorithms [11], [17], [18]. Applied exclusively to discrete EAs, drift analysis made obsolete the highly technical proof given in [15] for the hitting time of the $(1+1)$EA with linear (pseudo-Boolean) functions [13].

We start by reviewing the definition and main drift theorem 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 $T := \inf\{t \mid d(P_t) = 0\}$ be a RV, 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 RV, so will be $d(P_t)$ and also the so-called decrease in distance function $X_t$ given by

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

By definition, $E(X_t|T \geq 0)$ is called the (one step) mean drift. We also define

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

Then we have the drift theorem providing an upper bound for expected first hitting time [11, Lemma 3]:

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

*Theorem 4.1:* If $\Delta > 0$ then

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

We replace $M$ in the proof of Theorem 3.3 by the distance $d$ used in the continuous space model of Section III ($d$ may later grow to infinity). Similarly, $T$ is replaced by $T_d$ (the first hitting time of an area situated at distance $d$ from the initial population). The existence of $\Delta$ is postulated by hypothesis $H$: $\Delta = \mu_1$.

With these substitutions, the drift theorem is a corollary of Theorem 3.3, since we can use the right-hand side (upper bound) limit

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

A both sided formulation of the drift theorem is provided in [17, Lemmas 1 and 2].

*Theorem 4.2:* Let $\{P_t\}_{t>0}$ be a Markov process over a set of states $S$, and $d : S \to \Re_+$ a function that assigns to every state a nonnegative real number. Let the time to reach the optimum be $T := \min\{t > 0 \mid d(P_t) = 0\}$. If there exist constants $\mu_1 > 0$ and $\mu_2 > 0$ such that for any time $t > 0$ and state $P_t$ with $d(P_t) > 0$ the following condition holds:

$$\mu_1 \leq E[d(P_{t-1}) - d(P_t)|P_{t-1}] \leq \mu_2 \tag{22}$$

then the mean first hitting time satisfies

$$\frac{d(P_0)}{\mu_2} \leq E[T|P_0] \leq \frac{d(P_0)}{\mu_1}. \tag{23}$$

Or, again one-sided but with expected value of $d(P_0)$ in [24, Theorem 1]

$$E(T) \leq \frac{E[d(P_0)]}{\mu_1}. \tag{24}$$

The resemblance between theorems 3.3 and 4.2 is obvious, considering that:

1) the conditional expectation in ((22)) acknowledges the fact that we are measuring only one-step progress. In other words, we start with population $P_{t-1}$ in a fixed (yet arbitrary) point $x$, from which we perform a random move to $P_t$. So ((22)) leads actually to the $\mu_1 \leq E(X_t) \leq \mu_2$ inequality from our condition $H$;
2) as $d(P_0)$ is considered arbitrarily fixed at the beginning of the algorithm, $d(P_0) = E[d(P_0)]$ and it can be further replaced by the (constant) maximal distance to the optimum $d$ in both ((23)) and ((24)). That also allows for the substitution $E[T|P_0] = E(T_d)$ in ((23)).

In [13, Definition 1] 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 $P_t$

$$E[d(P_t)] \leq \left(1 - \frac{\delta}{n}\right) d(P_{t-1}). \tag{25}$$

This being a one-step condition, we can assume that $P_{t-1}$ is constant, and only $P_t$ is RV, which ensures $P_{t-1} = E(P_{t-1})$ and after insertion in ((25)) 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 $H$

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

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

Summing up, drift analysis provides, under the assumption of (strictly) positive bounds on the expected one-step progress of the algorithm toward the optimum, positive bounds on the expected hitting time of the algorithm—all for the discrete case.[12] Our renewal process analysis does the same, but for the continuous case.

## V. Hypersphere Volumes and Centroids

In order to apply the renewal process model developed in Section III to a practical EA acting on a particular fitness landscape, one has to first find good estimates for the lower and upper bounds imposed on the one-step progress by Condition $H$. The progress rate is considered the key problem in continuous EA theory, and a lot of effort has been spent in that direction—see the work of Beyer [7], [8], and Jägersküpper [19], [20].

Instead of considering the classical normal mutation from continuous EAs, we choose for the subsequent analysis uniform mutation, inside the $n$-dimensional sphere. As explained in Section II, the two are similar in many respects. Yet the benefits of using the uniform distribution are tremendous: the integral calculus of progress rates reduces to a geometrical calculus of volumes and centroids. For the SPHERE fitness landscape, we only need to find the volume and centroid of the intersection of two hyperspheres: one corresponding to mutation, the other one corresponding to the fitness function. For the RIDGE fitness landscape, we need the volume and centroid of the intersection between a hypersphere (mutation) and a hypercone (fitness).

This section presents the geometrical calculation of such volumes and centroids. In order to improve readability, we skip some of the technical details and present only the main results.

According to Li [21], the volume of an $n$-dimensional hypersphere (hereafter *n-sphere*) of radius $r$ is

$$V_n(r) = C_n r^n, \tag{26}$$

where the coefficient can be expressed for any $n$ in terms of Gamma functions, but for computational reasons we prefer this pair of formulas

$$C_n = \frac{2^{\frac{n+1}{2}} \pi^{\frac{n-1}{2}}}{n!!}, \qquad \text{when } n \text{ is odd, and} \tag{27}$$

$$C_n = \frac{\pi^{\frac{n}{2}}}{(\frac{n}{2})!}, \qquad \text{when } n \text{ is even.} \tag{28}$$

[12]There are actually two incipient attempts of extrapolating drift analysis to continuous space but so far only the 1-D and 2-D versions of the SPHERE have been considered [12], [18].

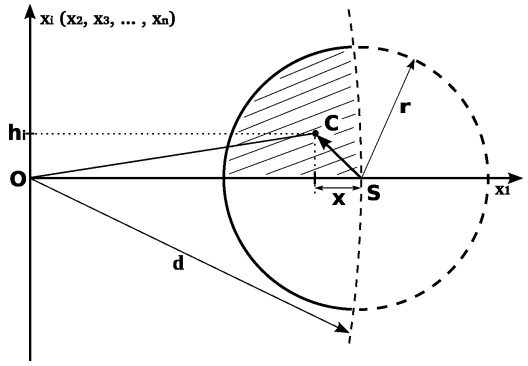


Fig. 1. Mutation sphere far away.

For geometric derivations involving two $n$-spheres, we take the progress axis $Ox_1$ to pass through the centers of both spheres. To simplify the notation, we refer to all other $n-1$ axes perpendicular to it ($Ox_2$,..., $Ox_n$) as $Ox_i$. The $x_1$-coordinate of the centroid of any $n$-volume is by definition

$$x = \frac{\int x_1 \, dV}{\int dV} = \frac{\int x_1 A(x_1) \, dx_1}{\int A(x_1) \, dx_1} \tag{29}$$

where $dV$ is the infinitesimal element of volume perpendicular to $Ox_1$, and $A(x_1)$ is the intersection between the volume and the (hyper)plane projecting at $x_1$. If the volume is $n$-dimensional, the area $A(x_1)$ is $(n-1)$-dimensional.

Following established ES literature, we denote by $h_i$ ($i = 2, ..., n$) the coordinates of the centroid on $Ox_i$, respectively. All $h_i$ are perpendicular to the progress axis, and, due to symmetry, all $h_i$ are equal. They are also best expressed as an integral over $x_1$

$$h_i = \frac{\int x_{iC}(x_1) A(x_1) \, dx_1}{\int A(x_1) \, dx_1}, \quad \forall \, i = 2, \ldots, n \tag{30}$$

where $x_{iC}(x_1)$ is the coordinate $i$ of the centroid of $A(x_1)$.

### A. Sphere "Far-Away": Centroid of a Quadrant

To evaluate the expected one-step progress of our EA when the starting point $S$ is far-away from the origin, we need the centroid of one (hyper)quadrant of the $n$-sphere centered at $S$, Fig. 1. Without loss of generality, we consider the positive quadrant—i.e., the first quadrant, the part of the sphere where all coordinates are positive.

Consider an $n$-sphere of radius $r = 1$, centered at the origin. Due to symmetry, its first quadrant has a centroid whose projections on all axes are equal, i.e., $x = h_i$, $\forall \; i = 2 \ldots n$. We shall evaluate only $x$, the projection on $Ox_1$. Also due to symmetry, $x$ is the same for all quadrants with positive coordinate $x_1$, so it is sufficient to calculate $x$ for the entire positive semi-$n$-sphere.

*Exact formulas:* In (29), the intersection $A(x_1)$ is an $(n-1)$-sphere of radius $\sqrt{1-x_1^2}$, according to Pythagora's theorem. With the notation from (26), we have

$$A(x_1) = V_{n-1}(\sqrt{1-x_1^2}).$$

The integration limits are 0 and 1. Substituting all this into (29), we have

$$x = \frac{\int_0^1 x_1 \, V_{n-1}(\sqrt{1-x_1^2})dx}{\frac{1}{2}V_n(1)}.$$

In the above expression, we use (26) and factor out the constant coefficients to obtain

$$x = \frac{2C_{n-1}}{C_n}\int_0^1 x_1 \, (1-x_1^2)^{\frac{n-1}{2}} dx_1.$$

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

$$x = \frac{2C_{n-1}}{(n+1)C_n}. \tag{31}$$

If we now allow an arbitrary radius $r$, the $n$-sphere is scaled by a factor of $r$. The centroid, being a first-order moment, also scales by $r$, so we simply multiply (31) by $r$. We change the variable on one hand to $n = 2k+1$, and on the other to $n = 2k$, and use the double factorial formulas. In the even and odd expressions obtained, we substitute, respectively, (27) and (28) to obtain the following *exact* expressions for the centroid of the semi-$n$-sphere:

$$x = h_i = \frac{(2k+1)!}{2^{2k+1}k(k!)^2}\, r, \qquad \text{when n} = 2k+1 \tag{32}$$

$$x = h_i = \frac{2^{2k+1}(k!)^2}{\pi(2k+1)(2k)!}\, r, \qquad \text{when n} = 2k. \tag{33}$$

The above formulas can be easily checked for the first three values of $n$: For $n = 1$, (32) gives $1/2$, which is indeed the centroid of the segment [0..1]. For $n = 2$, (33) gives $4/3\pi$, which is the centroid of the positive half-disc of radius 1. For $n = 3$, (32) gives $3/8$, which is the centroid of the positive half-sphere of radius 1.

*Limits for $n \to \infty$:* We apply Stirling's formula in (32) and (33) and find that, for $n$ both odd and even, the limit of all coordinates of the centroid are

$$x = h_i = \frac{r}{\sqrt{\pi k}}\left[1 + O\left(\frac{1}{k}\right)\right] = \\ = \frac{r\sqrt{2}}{\sqrt{\pi n}}\left[1 + O\left(\frac{1}{n}\right)\right] \approx \frac{r\sqrt{2}}{\sqrt{\pi n}}. \tag{34}$$

The following two cases are of particular importance.

A. If $r$ is kept constant as $n \to \infty$, $x$ and all $h_i$ approach the origin as $1/\sqrt{n}$. The *distance* from the center $S$ of the $n$-sphere to the centroid $C$ of the quadrant (Fig. 1) is the diagonal of an $n$-hypercube whose edge has this limit. Apply next Pythagora to obtain

*Theorem 5.1:* When $n \to \infty$, the distance $\|SC\|$ to the centroid of a quadrant of the semi-$n$-sphere of constant radius $r$ tends to a constant value

$$\lim_{n\to\infty} \|SC\| = r\sqrt{\frac{2}{\pi}}. \tag{35}$$

B. If, on the other hand, we keep the *volume* of the $n$-sphere constant as $n \to \infty$, the radius also changes, according to (26). 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 $C_n$ (27) and (28), 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

$$r = \sqrt{\frac{k}{\pi e}}\left[1 + O\left(1 + \frac{1}{k}\right)\right] = \\ = \sqrt{\frac{n}{2\pi e}}\left[1 + O\left(1 + \frac{1}{n}\right)\right] \approx \sqrt{\frac{n}{2\pi e}}. \tag{36}$$

We proved the following:

*Lemma 6:* When $n \to \infty$, the radius of an $n$-sphere of constant volume $V_n = 1$ also tends to $\infty$. The growth is approximately

$$\sqrt{\frac{n}{2\pi e}}$$

Substituting into (34), we obtain

*Proposition 3:* When $n \to \infty$, all coordinates of the centroid of the positive quadrant of the $n$-sphere of volume $V_n = 1$ tend to the same fixed value

$$\lim_{n\to\infty} x = \lim_{n\to\infty} h_i = \frac{1}{\pi\sqrt{e}} \approx 0.193 \qquad \forall\, i = 2, \ldots, n. \tag{37}$$

The assumption $V = 1$ can now be dropped in both results 6 and 3, since

$$\lim_{n\to\infty} V^{1/n} = 1.$$

It follows that

*Theorem 5.2:* When $n \to \infty$, all coordinates of the centroid of the positive quadrant of the $n$-sphere of constant volume $V_n = V$ tend to a fixed value which is independent of $V$:

$$\lim_{n\to\infty} x = \lim_{n\to\infty} h_i = \frac{1}{\pi\sqrt{e}} \approx 0.193 \qquad \forall\, i = 2, \ldots, n \tag{38}$$

When considering only one quadrant, the limit above applies individually to the coordinate $x$ and each coordinate $h$. The *distance* from the center $S$ of the $n$-sphere to the centroid $C$ of the quadrant (Fig. 1) is the diagonal of an $n$-hypercube whose edge tends to this limit. Applying Pythagora we obtain

*Theorem 5.3:* When $n \to \infty$, the distance $\|SC\|$ to the centroid of a quadrant of the semi-$n$-sphere of constant volume $V_n = V$ tends to infinity:

$$\|SC\| = \sqrt{n}\frac{1}{\pi\sqrt{e}}\left[1 + O\left(\frac{1}{n}\right)\right] \approx \sqrt{n}\cdot 0.193 \tag{39}$$

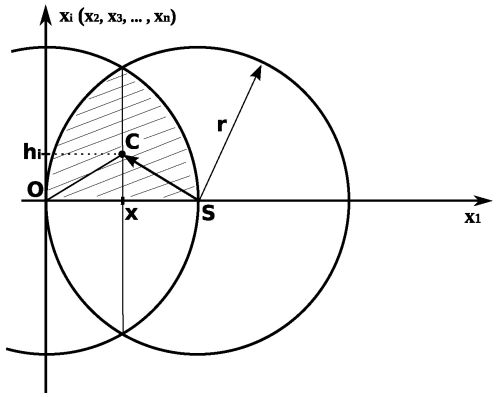


Fig. 2. Mutation sphere close-by.

### B. Sphere "Close-By": Centroid of a Cap Quadrant

To evaluate the average 1-step progress of our EA when the starting point $S$ is close to the origin, we consider an $n$-sphere centered at $S$, where $S$ is one radius away from the origin. When a successful mutation is generated, the new point is closer to the origin, therefore inside the $n$-sphere of same radius $r$, centered at the origin. The intersection of the two (equal) spheres is usually called a (symmetric) spherical lens, and it consists of two mirror-image, back-to-back spherical caps. We consider again the restriction of this volume to the positive quadrant. Due to symmetry, $x = r/2$, all $h_i$ are equal, and the two cap quadrants making up the lens quadrant have the same $h_i$, so it is sufficient to calculate $h_i$ for only one cap quadrant, Fig. 2.

A computation similar to the one in Section V-A yields the following.

*Theorem 5.4:* When $n \to \infty$ and the radius $r$ of the $n$-sphere are kept constant, the distance $\|SC\|$ to the centroid of the cap quadrant tends to a constant value

$$\lim_{n\to\infty} \|SC\| = r\sqrt{\frac{3+\pi}{2\pi}} \approx r \cdot 0.989. \tag{40}$$

The perpendicular coordinates all tend to zero

$$h_i = \frac{r}{\sqrt{n}}\sqrt{\frac{3}{2\pi}}\left[1 + O\left(\frac{1}{n}\right)\right] \qquad \forall\, i = 2, \ldots, n \tag{41}$$

*Theorem 5.5:* When $n \to \infty$, and the volume $V$ of the $n$-sphere is kept constant, the distance $\|SC\|$ to the centroid of the cap quadrant tends to infinity

$$\|SC\| = \sqrt{n}\frac{\sqrt{\pi+6}}{\pi\sqrt{8e}}\left[1 + O\left(\frac{1}{\sqrt{n}}\right)\right] \approx \sqrt{n} \cdot 0.206 \tag{42}$$

The perpendicular coordinates are all equal, tending to a fixed value which is independent of $V$

$$h_i = \frac{\sqrt{3}}{2\pi\sqrt{e}}\left[1 + O\left(\frac{1}{n}\right)\right] \approx 0.167 \qquad \forall\, i = 2, \ldots, n \tag{43}$$

The progress coordinate tends to infinity

$$x = \frac{r}{2} = \frac{\sqrt{n}}{\sqrt{8\pi e}}\left[1 + O\left(\frac{1}{n}\right)\right] \approx \sqrt{n} \cdot 0.121 \tag{44}$$

It also follows that the volume of the cap quadrant is given by the following.

*Theorem 5.6:* If both $n$-dimensional spheres in Fig. 2 have volume one, the volume of the cap quadrant is

$$\mathrm{V}_{cap} = \frac{C_{n-1}}{C_n}\, B_{3/4}\left(\frac{n+1}{2}, \frac{1}{2}\right) \tag{45}$$

which converges to zero quasi-exponentially, as $n \to \infty$

$$\mathrm{V}_{cap} = \left(\frac{3}{4}\right)^{\frac{n+1}{2}} \frac{4}{\sqrt{2\ \Pi\ n}} \approx 0.75^{\frac{n+1}{2}} \cdot \frac{1.596}{\sqrt{n}}. \tag{46}$$

## VI. Convergence Time Analysis

We are going to use the centroid formulas derived in Section V 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 $Ox_1$—namely $\mu_1$ and $\mu_2$ of Theorem 3.3. 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 always on $Ox_1$.

From a geometrical point of view, uniform mutation inside the sphere is more tractable than normal mutation. To see that, note the following simple fact.

*Remark 3:* The expected value of a uniform variable defined inside a figure of volume 1 is the *centroid* (center of mass) of the corresponding figure. In case of the elitist EA on SPHERE, not all of the mutation sphere is active for the next generation, the removed volume (probability) being charged to a single point, the current position of the algorithm. That corresponds to a truncated volume, and the expected value in this case is the centroid of the truncated figure, times the corresponding volume.

If we apply that to the case far-away, it yields a scaling factor $1/2$ between the coordinates of the centroid, and the coordinates of the expected progress $\phi_{1+1}$. In case close-by, the adjustment is more severe, by the factor tending exponentially to zero w.r.t. space dimension $n$ from ((46)). All these will be summed up in Lemma 7.

On the other hand, if the mutation sphere is no longer of volume 1—e.g., when mutation radius $r$ is decreased, one has to divide the uniform variable and consequently its expected value by the volume of the new sphere, in order for the nonunitary sphere to define a proper RV.

Beyond geometrical considerations, there is also an algorithmic modeling issue, that we address in the following.

When starting our analysis, the first choice was to consider for the expected one-step progress the centroid of the whole active area, the semihypersphere in Fig. 1, respectively the full lens in Fig. 2. That was wrong, since the centroid would then lie on $Ox_1$, ignoring completely the radial components $h_i$. In $n$ dimensions, these $n-1$ lateral components are by a factor of $\sqrt{n-1} \approx \sqrt{n}$ larger than the central component,[13] which means that we expect (informally) the mutant not to lie close to the line through the mutated individual and the optimum, but almost perpendicular to it. That is the reason for considering

[13]See ((38)) and ((39)) in Section V-A, or the deviation estimate $E(\cos\alpha) = 0.8/\sqrt{n}$ computed in [19].

in Section V only the upper hyperquadrant of the area when computing the centroid, yielding a representation of the lateral component (Theorems 5.3 and 5.5) in perfect agreement with the established ES literature.

We are now in the position of bringing together all the main results of the previous sections.

*Lemma 7:* If $d = n$ and $r = \sqrt{n/(2\pi e)}$, the expected one-step progress on SPHERE of the (1 + 1)EA with uniform mutation satisfies, for large $n$

$$\phi_{1+1}(\text{far-away}) > \phi_{1+1}(\text{close-by}). \tag{47}$$

*Proof:* Apply ((1)), then ((38)) and ((39)), and Remark 3 to obtain

$$\phi_{1+1}(\text{far-away}) = E\left[d - \sqrt{\left(d - \frac{x}{2}\right)^2 + \left(\frac{h}{2}\right)^2}\right] =$$

$$= n - \sqrt{n^2 - 2n\frac{x}{2} + \left(\frac{x}{2}\right)^2 + \left(\frac{||SC||}{2}\right)^2 - \left(\frac{x}{2}\right)^2} \approx$$

$$\approx n - \sqrt{n^2 - 0.183n} = \frac{0.183n}{n + \sqrt{n^2 - 0.183n}}.$$

A similar calculation for case close-by yields

$$\phi_{1+1}(\text{close-by}) = E\left[r - \sqrt{\left(r - x\mathbf{V}_{\text{cap}}\right)^2 + \left(h\mathbf{V}_{\text{cap}}\right)^2}\right] =$$

$$= \frac{2rx\mathbf{V}_{\text{cap}} - (||SC||\mathbf{V}_{\text{cap}})^2}{r + \sqrt{r^2 - 2rx\mathbf{V}_{\text{cap}} + (||SC||\mathbf{V}_{\text{cap}})^2}} \approx$$

$$\approx 7.93 \cdot 0.75^{\frac{n+1}{2}}.$$

One can easily see that for $n$ large enough—namely, $n > 30$—the following holds:

$$\frac{0.183n}{n + \sqrt{n^2 - 0.183n}} > 7.93 \cdot 0.75^{\frac{n+1}{2}} \tag{48}$$

since the left-hand side converges to 0.091, while the right-hand side converges to zero, as $n \to \infty$. ■

Rigorously speaking, Lemma 7 depends on the following conjecture:

*Conjecture 6.1:* The progress function ((1)) is monotone w.r.t. $d$ on $[r, \infty)$.

In order to prove the conjecture, one has to generalize the calculus of Section V-B to an *asymmetrical* lens, since the left sphere is now larger. Finding the centroid of the new figure analytically might be cumbersome, but there is also the variant of evaluating numerically the four $\beta$-integrals involved, for different values of $n$, as $n \to \infty$. Something similar was done in [8, p.56, Fig. 3.2], while analyzing the progress rate $\phi$ versus mutation strength $\sigma$ of the (1 + 1)ES on the SPHERE. Straightforward as it seems, the proof of Conjecture 6.1 is still in progress.

*Theorem 6.2:* Let the (1 + 1)EA with uniform mutation inside the sphere of (constant) radius $r = \sqrt{n/(2\pi e)}$ minimize the $n$-dimensional SPHERE function. Assume the algorithm starts at distance $n$ from optimum, with $n > 30$. Then the expected number of steps to reach distance $\Theta(\sqrt{n})$ from optimum is in $\Omega(n)$, and in $O(e^{c \cdot n})$.

*Proof:* We check hypothesis $H$ on RV 'one-step progress of the algorithm', confined to distance $[r, \infty)$ from optimum. The variance of bounded uniform distribution is obviously bounded, while the confinement on the expected value is obtained due to ((48)) and Conjecture 6.1 as

$$7.93 \cdot 0.75^{\frac{n+1}{2}} \le E(X_t) \le \frac{0.183n}{n + \sqrt{n^2 - 0.183n}}. \tag{49}$$

Then Theorem 3.3 ensures that, for $d = n$ large enough

$$\frac{n + \sqrt{n^2 - 0.183n}}{0.183n} \le \frac{E(T_{d-r})}{d - r} \le 0.126 \cdot 1.33^{\frac{n+1}{2}}$$

or, after substituting $d$, $r$, and multiplying the whole inequality by the denominator of the middle fraction

$$\frac{(n - 0.242\sqrt{n})\left(n + \sqrt{n^2 - 0.183n}\right)}{0.183n} \le$$

$$\le E(T_{n-0.242\sqrt{n}}) \le 0.126(n - 0.242\sqrt{n})1.33^{\frac{n+1}{2}}.$$

Recalling that $E(T_{n-0.242\sqrt{n}})$ represents the expected number of steps required by the (1+1)EA starting in $n$ to reach distance $0.242\sqrt{n}$ from optimum, the lower and upper bounds $n$ and $e^{c \cdot n}$ are computed from the left-, respectively right-hand side of the above inequality. ■

As pointed out in Section II, the uniform mutation inside the sphere is *isotropic*, thus Theorem 1.1 applies also to the algorithm of Theorem 6.2, providing expected runtime $\Omega(n)$ to reach a spatial gain of $n - \Theta(\sqrt{n}) = \Theta(n)$. So our result is in good agreement with established EA literature.

Compared to another result of Jägersküpper, showing that the *adaptive*, 1/5-rule (1 + 1)EA converges in $\Theta(n)$ [19], our exponential upper bound in Theorem 6.2 is due to the *constant* mutation rate in our case study.

## VII. Conclusion

For the theoretical analysis of EAs to be of practical relevance, it should provide, besides necessary and sufficient conditions for global convergence, hitting time estimations for specific algorithms on particular fitness landscapes. Both in discrete and continuous space, significant insight into the algorithm's behavior can be gained by mapping the successive EA populations onto the infinite sequence of RVs defining a stochastic process. So far, Markov processes and their variants seemed to be the perfect candidate, yet there are also other options; the present paper made the case for the renewal process, a stochastic model widely used in queueing theory.

We introduced a natural bounding assumption on the RV that models the one-step progress, $H$, corresponding to the practical case where the expected progress of the EA is different from one iteration to another (this is particularly good for modeling adaptive algorithms). Our approach then followed the general lines of renewal theory, up to an estimate of the first hitting time of a distance $d$ from the starting point of the algorithm. The estimate is simple and highly intuitive: the expected average number of iterations required per distance unit enters (when $d \to \infty$) the interval of inverted one-step progress bounds.

After finding good similarities with the drift analysis, we tested the new method on the (1+1)EA with constant, uniform mutation optimizing the SPHERE function. With extensive use

of the geometrical properties of the $n$-dimensional uniform distribution, we proved that the EA exhibits a convergence time in $\Omega(n)$ and $O(e^{c \cdot n})$.

We claim that renewal process modeling is powerful enough to cover complex algorithms and fitness landscapes. In order to obtain practical hitting times, one needs estimates of the upper and lower bounds for the one-step progress of the EA on particular problems. This is the main challenge for the theory of continuous EAs.

Besides the results on the SPHERE function presented in the paper, a similar study is under construction for the (1 + 1)EA with uniform mutation on the RIDGE problem—a natural extension of the SPHERE that breaks the symmetry in one direction [9].

## REFERENCES


[1] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*. New York: Dover, 1970.

[2] A. Agapie, "Modelling genetic algorithms: From Markov chains to dependence with complete connections," in *Proc. PPSN*, LNCS 1498, 1998, pp. 3–12.

[3] A. Agapie, "Theoretical analysis of mutation-adaptive evolutionary algorithms," *Evol. Comput.*, vol. 9, no. 2, pp. 127–146, 2001.

[4] A. Agapie, "Estimation of distribution algorithms on non-separable problems," *Int. J. Comput. Math.*, vol. 87, no. 3, pp. 491–508, 2010.

[5] A. Agapie, M. Agapie, and G. Zbaganu, "Evolutionary algorithms for continuous space optimization," *Int. J. Syst. Sci.*, vol. 44, no. 3, pp. 502–512, 2013.

[6] A. Auger, "Convergence results for the $(1, \lambda)$-SA-ES using the theory of $\phi$-irreducible Markov chains," *Theor. Comput. Sci.*, vol. 334, pp. 35–69, Apr. 2005.

[7] H.-G. Beyer, "Toward a theory of evolution strategies: Some asymptotical results from the $(1,^+ \lambda)$-theory," *Evol. Comput.*, vol. 1, no. 2, pp. 165–188, 1993.

[8] H.-G. Beyer, *The Theory of Evolution Strategies.* Heidelberg, Germany: Springer, 2001.

[9] H.-G. Beyer, "On the performance of $(1, \lambda)$-evolution strategies for the ridge function class," *IEEE Trans. Evol. Comput.*, vol. 5, no. 3, pp. 218–235, Jun. 2001.

[10] A. Bienvenu̇e and O. Francois, "Global convergence for evolution strategies in spherical problems: Some simple proofs and difficulties," *Theor. Comput. Sci.*, vol. 306, pp. 269–289, Sep. 2003.

[11] T. Chen, J. He, G. Sun, G. Chen, and X. Yao, "A new approach for analyzing average time complexity of population-based evolutionary algorithms on unimodal problems," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol. 39, no. 5, pp. 1092–1106, Oct. 2009.

[12] Y. Chen, X. Zou, and J. He, "Drift conditions for estimating the first hitting times of evolutionary algorithm," *Int. J. Comput. Math.*, vol. 88, no. 1, pp. 37–50, 2011.

[13] B. Doerr and L. A. Goldberg, "Adaptive drift analysis," in *Proc. PPSN*, LNCS 6238, 2010, pp. 32–41.

[14] C. C. Y. Dorea, "Stationary distribution of Markov chains in $\Re^d$ with application to global random optimization," *Bernoulli*, vol. 3, no. 4, pp. 415–427, 1997.

[15] S. Droste, T. Jansen, and I. Wegener, "On the analysis of the (1+1) evolutionary algorithm," *Theor. Comput. Sci.*, vol. 276, nos. 1–2, pp. 51–81, 2002.

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

[17] J. He and X. Yao, "A study of drift analysis for estimating computation time of evolutionary algorithms," *Natural Computing*, vol. 3, no. 1, pp. 21–35, 2004.

[18] J. He, X. Yao, and Q. Zhang, "To understand fitness landscapes in continuous space by using drift analysis," in *Proc. IEEE Congr. Evol. Comp.*, Jun. 2004, pp. 1248–1253.

[19] J. Jägersküpper, "Analysis of a simple evolutionary algorithm for minimisation in Euclidean spaces," in *Proc. ICALP*, LNCS 2719, 2003, pp. 1068–1079.

[20] J. Jägersküpper, "Lower bounds for hit-and-run direct search," in *Proc. SAGA*, LNCS 4665, 2007, pp. 118–129.

[21] S. Li, "Concise formulas for the area and volume of a hyperspherical cap," *Asian J. Math. and Stat.*, vol. 4, no. 1, pp. 66–70, 2011.

[22] J. Lopez and J. Sesma, "Asymptotic expansion of the incomplete gamma function for large values of the first parameter," *Integral Transforms Special Functions*, vol. 8, nos. 3–4, pp. 233–236, 1999.

[23] L. Lovasz and S. Vempala, "Hit-and-run from a corner" *SIAM J. Computing*, vol. 35, no. 4, pp. 985–1005, 2006.

[24] P. S. Oliveto and C. Witt, "Simplified drift analysis for proving lower bounds in evolutionary computation," *Algorithmica*, vol. 59, no. 3, pp. 369–386, 2010.

[25] S. Ross, *Applied Probability Models with Optimization Applications.* New York: Dover, 1992.

[26] G. Rudolph, *Convergence Properties of Evolutionary Algorithms.* Hamburg: Kovać, 1997.

[27] G. Rudolph, "Stochastic convergence," in *Handbook of Natural Computing*, G. Rozenberg, T. Bäck and J. Kok, Eds. Berlin, Germany: Springer, 2013, pp. 847–869.

[28] G. Rudolph, "Evolutionary strategies," in *Handbook of Natural Computing*, G. Rozenberg, T. Bäck and J. Kok, Eds. Berlin, Germany: Springer, 2013, pp. 673–698.

[29] D. Williams, *Probability with Martingales.* Cambridge, MA, USA: Cambridge Univ. Press, 1991.


**Alexandru Agapie** received the Lic. and Ph.D. degrees in mathematics from the University of Bucharest, Bucharest, Romania in 1993 and 2001, respectively.

He has been a Researcher with the National Institute for Microtechnology, Bucharest, and with the Fraunhofer Institute for Autonomous Intelligent Systems, Sankt Augustin, Germany. Currently, he is an Associate Professor with the Bucharest University of Economic Studies, Bucharest, Romania, and a Researcher with the Institute of Mathematical Statistics and Applied Mathematics, Bucharest, Romania. His current research interests include stochastic processes applied to evolutionary algorithms and to socioeconomic models.

**Mircea Agapie** (M'01) received the Electrical Engineer Diploma from the Politehnica University of Bucharest, Bucharest, Romania, in 1992, and the Ph.D. degree in computer science from the University of Missouri, Kansas City, USA, in 2000. Currently, he is an Associate Professor with the Department of Engineering and Computer Science at Tarleton State University, Stephenville, TX, USA.

His current research interests include artificial intelligence, robotics and applied probability.

**Günter Rudolph** (M'98) received the Dipl. and Ph.D. degrees in computer science from the University of Dortmund, Germany in 1991 and 1996, respectively.

He has been a Researcher with the Informatics Center Dortmund, and with the Product- and Software-Development, Parsytec AG, Aachen. Currently, he is a Professor with the Department of Computer Science at TU Dortmund, Dortmund, Germany. His current research interests include parallel computing, modeling, simulation, optimization in general, and evolutionary algorithms in particular.

**Gheorghita Zbaganu** received the Lic. and Ph.D. degrees in mathematics from the University of Bucharest, Bucharest, Romania in 1975 and 1986, respectively.

Currently, he is a Professor with the Department of Mathematics and Computer Sciences at the University of Bucharest, and a Researcher with the Institute of Mathematical Statistics and Applied Mathematics, Bucharest. His current research interests include probability theory, actuaries, mathematical statistics, information theory, and more specifically stochastic comparisons and portfolio theory.