Evolutionary Algorithms for Continuous Space Optimization

Alexandru Agapie*  Mircea Agapie†  Gheorghita Zbaganu‡

Abstract

From a global viewpoint, Evolutionary Algorithms (EA) working on continuous search-spaces can be regarded as homogeneous Markov chains (MC) with discrete time and continuous state. We analyze from this viewpoint the (1+1)EA on the inclined plane fitness landscape, and derive a closed-form expression for the probability of occupancy of an arbitrary target zone, at an arbitrary iteration of the EA. For the hitting-time of an arbitrary target zone, we provide lower and upper bounds, as well as an asymptotic limit. Discretization leads to a MC with discrete time, whose simple structure is exploited to carry out efficient numerical investigations of the theoretical results obtained. The numerical results thoroughly confirm the theoretical ones, and also suggest various conjectures which go beyond the theory.

Keywords: Evolutionary algorithms with continuous state-space, Markov transition function, Discretization of Markov chains, Renewal process, Convolution of distributions, Hitting time, Non-commutative binomial.

1 Introduction

Since their appearance in their middle 1960s, evolutionary algorithms (EAs) proved successful in many applications, both on finite and continuous search-space, yet they still lack a unitary theory. The situation is more dramatic for continuous EAs, where only few rigorous attempts towards a theoretical analysis have been made. The most important are summarized in the following.

The martingale approach to random search (pioneered in [15]) can be criticized for assuming a certain positive success rate in each algorithmic iteration. This implies a local, rather than global, approach. The same is true for the evolution strategy theory developed in [5]. A more recent approach is due to Jägersküpper, who builds his computation time analysis on local concepts such

*corresponding author: Academy of Economic Studies, Dept. of Applied Mathematics, Calea Dorobantilor 15-17, Bucharest 010552, Romania, agapie@clicknet.ro
†Computer Science, Tarleton State University, Box T-0560 Stephenville, TX 76402, USA
‡Institute of Math. Statistics and Applied Mathematics, Casa Academiei Romane, Calea 13 Septembrie 13, Bucharest 050711, Romania
as “success probability” and “expected spatial gain of a step”. Starting from how the (1+1)EA minimizes the well-known SPHERE-function, [10] extrapolates the results to a broader class of functions, namely a family of positive definite quadratic forms. The main difference between the present work and that of Rapple, Beyer or Jägersküpner is that we approach the problem globally, by means of end-to-end stochastic processes. We “keep track” of all possible trajectories, which technically translates into multiple-integrals or convolution products. Of course, such a global analysis is not always tractable. But when it is, the results are more powerful, more transparent and more precise than those obtained through local analysis. The present paper deals with a case where the global analysis is tractable, and improves on the existing results by providing closed-form occupancy distributions and hitting times, as well as in-depth numerical validations and further explorations.

For finite search-space EAs, theoretical results on convergence rates and computation time have been around for some time [19]. A comprehensive review of the early investigations can be found in [18], while a complete characterization of an EA variant with respect to computation complexity was given by Droste, Jansen and Wegener [6]. As in the present paper, their algorithm does not use crossover, and the population is restricted to one individual - represented in their case as a binary string of fixed length n. A bitwise mutation operator flips each bit independently of the others with some probability \( p_m \), producing a new individual. Following the paradigm of natural selection, the new string replaces the old string if it is better (according to a given objective function), otherwise the old one stays. This very simple probabilistic algorithm is usually referred to in literature as (1+1)EA. A stochastic formalization of this algorithm will be given below, but for the individual moving in a continuous subset of \( \mathbb{R}^n \), instead of \( \{0,1\}^n \). The main problem considered in [6] was to estimate the expected running time \( E(T) \) of the (1+1)EA on different types of objective functions \( f : \{0,1\}^n \rightarrow \mathbb{R} \), where T is the random variable number of iterations until convergence. Droste et al. achieve their goal largely through a probabilistic approach, but their proofs rely strongly on combinatorics and computational complexity. They show that linear functions are always solvable in expected time \( O(n \cdot \log n) \), while there exist functions of order 2 with expected running time \( \Theta(n^n) \) - which is the worst case, considering that a simple enumeration of all strings in \( \{0,1\}^n \) would find the global optimal string in maximum \( 2^n \) iterations. Next, dealing with unimodal functions - a superclass of the linear functions - they derive lower bounds on the expected running time for two concrete objective functions, and show that this class has exponential running time in the worst case.

The fundamental difference between the algorithm analyzed in [6] and the one in the present paper is the search-space: \( \{0,1\}^n \) in their case, a continuous subspace of \( \mathbb{R}^n \) here. Although one would expect similarities in methodology and results, this is not the case, since the change of search-space implies huge changes in the formalization, making extrapolation from finite to continuous impossible.

Departing from [6] and the entire finite case literature, the present paper is
aimed at EAs optimizing on continuous space. The underlying mathematical theory is that of continuous state-space stochastic processes. The first step in this direction was taken by [18], which introduced the formalization and presented the first global convergence result. According to that, two sufficient conditions for the convergence of an EA acting on continuous space are: (i) elitism, that is, the best state found so far can not be lost from an iteration to another, and (ii) positively bounded probability for reaching the target zone in one iteration, from any point of the space. The original approach of Rudolph was also extrapolated to adaptive EAs, modeled by multi-order Markov chains [4]. Still, from a practical viewpoint, both analyses share the same deficit: they are too general; the two conditions above do not provide the EA user with important information, like convergence rates or computation time for particular algorithms.

Our work aims to fill this gap by developing a framework for the analysis of specific continuous EAs, with specific mutation operators, and specific fitness landscapes. The EA (1+1), mutation (uniform) and landscape (inclined plane) considered in this paper are so simple as to be regarded as purely theoretical, however the stochastic analysis we propose is powerful enough to allow extrapolation to more complex problems, as sketched in the final section. The first results, including the discrete-time continuous-space (DTCS) Markov model and the theoretical computation of n-iteration success probability, have been presented in [3]. Building upon those techniques, the aim of the current paper is twofold: First, we complete the analysis of the inclined plane landscape by providing a closed-form cumulative distribution of the $n$th iteration of the EA, and by deriving bounds and the asymptotic limit of the mean hitting time for an arbitrary target zone. Second, we show how these theoretical results can be retrieved numerically, by discretizing the DTCS Markov chain into a discrete-time discrete-state (DTDS) MC with a simple structure. Computationally, discretization replaces multiple convolution integrals with matrix products.

2 Preliminaries

To introduce the notation and basic concepts of stochastic processes, let $\mathbb{R}^k$ be the k-dimensional real space, $f: \mathbb{R}^k \rightarrow \mathbb{R}_+$ be a fitness function, and assume an optimization task of the form

$$\max_x \{f(x)\}. \quad (1)$$

The current (1+1)EA consists of a single individual, a point in $\mathbb{R}^k$ which is a feasible solution to problem (1). The transition between successive individuals is achieved by applying the evolutionary operators of mutation and selection. Mutation acts by generating randomly, according to some fixed distribution (normal, uniform etc.), another feasible point. Then (elitist) selection compares
the new individual against the old one, and keeps the best individual with respect to (1). A formal description of the algorithm is given in the following

Procedure (1+1)EA

1. Set $t = 0$ and set the initial point of the algorithm, $x_t = (0, \ldots, 0)$

2. repeat
   - (i) generate a new point $x$ in $\mathbb{R}^k$ according to a given distribution
   - (ii) if $f(x) > f(x_t)$ then $x_t := x$
   - (iii) $t := t + 1$

3. until $f(x_t) > M$, for some fixed $M > 0$

As usual in probability theory, $\mathbb{R}^k$ is endowed with:

i) The Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^k)$, generated by the class of all open sets in $\mathbb{R}^k$

ii) A set-function $P : \mathcal{B}(\mathbb{R}^k) \rightarrow [0, 1]$, called probability and satisfying the following properties: $P(\emptyset) = 0$; $P(\bigcup_{n \geq 1} A_n) = \sum_{n \geq 1} P(A_n)$, for $A_i \cap A_j = \emptyset \forall i, j$; and $P(\bar{A}) = 1 - P(A)$, for $\bar{A}$ the complement of $A$.

A function $\xi : \mathbb{R}^k \rightarrow \mathbb{R}$ is called a random variable if the preimage through $\xi$ of any open set from $\mathbb{R}$ is an open set of $\mathbb{R}^k$. The cumulative distribution function (CDF) associated to random variable $\xi$ is a function $F_\xi : \mathbb{R} \rightarrow [0, 1]$, continuous to the right, defined by $F_\xi(t) = P(\omega | \xi(\omega) \leq t)$.

Definition 2.1 Let $F_1$ and $F_2$ be two CDFs. The real function $F$ defined by

$$F(x) = (F_1 * F_2)(x) = \int_{-\infty}^{\infty} F_1(x - y)dF_2(y) \quad (2)$$

is also a CDF, called the convolution of $F_1$ and $F_2$.

The convolution is associative and can be generalized to more than two variables. There is an important connection between the convolution and the sum of independent random variables, cf. [16]:

Theorem 2.2 The CDF of a sum of $k$ independent random variables is the $k$-order convolution of the corresponding CDFs.

Let $\{z_n\}_{n \geq 1}$ be a sequence of independent identically distributed (i.i.d.) random variables, with common distribution $z$. One can define a general random walk by summing up consecutive terms $z_n$. Rigorously, one gets a special type of stochastic process, $\hat{X} = \{X_n\}_{n \geq 1}$, given by

$$X_0 \equiv 0,$$

$$X_n = X_0 + z_1 + \ldots + z_n \quad n \geq 1. \quad (3)$$
Let \( A \) be a probability function.

Then \( \hat{N} = \{ N(t) \}_{t \in [0, \infty)} \) is called the renewal process associated to \( \hat{X} \), and \( 1 + N(t) \) is a random variable of a special type, generally referred to as stopping time. For a general theory of renewal processes we recommend the monograph [8], or Chapter 3 of [17]. The following result will prove important in our analysis:

**Lemma 2.3 (Wald’s identity)** Let \( \{ z_n \}_{n \geq 1} \) be a sequence of i.i.d. random variables, with common distribution \( z \) and \( S \) a corresponding stopping time. If \( E(S) < \infty \) and \( E(z) < \infty \), then

\[
E(z_1 + \cdots + z_S) = E(S) \cdot E(z).
\]

In case of the (1+1)EA on the inclined plane landscape, if we consider \( z \) to incorporate the effect of both evolutionary operators (mutation+elitism) and project it onto the 'progress' axis, then \( X_n \) will give the position of the algorithm starting in zero after \( n \) iterations (again, projected on the 'progress' axis) and \( 1 + N(t) \) will be giving the (first) hitting time of some predefined target zone \( S_t \). The expected value of this random variable is of great interest.

\[
E[1 + N(t)] = 1 + E[N(t)], \quad \forall t \in [0, \infty).
\]

The asymptotics of \( E[N(t)] \) receives an elegant characterization in [17].

**Theorem 2.4** Let \( \hat{N} = \{ N(t) \}_{t \in [0, \infty)} \) be the renewal process associated to \( \hat{X} = \{ X_n \}_{n \geq 1} \), and assume that \( m = E(X_1) = E(z) < \infty \) and \( \sigma^2 = \text{VAR}(X_1) = \text{VAR}(z) < \infty \). Then

\[
E[N(t)] = \frac{t}{m} + \frac{\sigma^2 - m^2}{2m^2} + o(1), \quad \text{for } t \to \infty.
\]

The transition function is defined for the \( k \)-dimensional real space by the following, cf. [13].

**Definition 2.5** A function \( P : \mathbb{R}^k \times \mathcal{B}(\mathbb{R}^k) \to [0, 1] \) is said to be a transition function (kernel) if \( P(w, \cdot) \) is a probability measure on \( \mathcal{B}(\mathbb{R}^k) \) for all \( w \in \mathbb{R}^k \), and \( P(\cdot, A) \) is a random variable on \( \mathbb{R}^k \) for all \( A \in \mathcal{B}(\mathbb{R}^k) \).

The kernel \( P(w, A) \) is the transition probability from the point \( w \) to the set \( A \). In the present paper it will represent the probability for the (1+1)EA to move from one point \( w \) in the search-space to a (hyper-)volume \( A \) of the same space, in one iteration. For \( k = 2 \), \( A \) is simply a two-dimensional area of the plane. Similarly, we have the \( n \)-step transitions \( P^n(w, A) \), with arbitrary positive integer \( n \), which model \( n \) successive iterations of the EA. For continuous space, transition probabilities between two points are always zero: \( P(w, v) = 0 \), for any \( w, v \in \mathbb{R}^k \).

With respect to its first argument, \( P(\cdot, A) \) is a function from points of \( \mathbb{R}^k \) to probabilities in \( [0, 1] \), and it must be a random variable. With respect to its second argument, \( P(w, \cdot) \) is a function from sets of \( \mathcal{B}(\mathbb{R}^k) \) to \( [0, 1] \) which must be a probability function.
3 The Inclined Plane fitness landscape

In this section, we introduce the Markov kernel of the inclined plane model and state without proof the main result in [3]. We restrict the search to the real plane (\(k = 2\), with variables \(x\) and \(y\)) and consider the (1+1)EA with square uniform mutation starting at zero. For the fitness function, we orient the coordinate system so that the plane slopes in the direction of the \(x\)-axis, i.e. \(x = \infty\) corresponds to the optimum [20]. For this reason, the \(x\)-axis is also called progress axis in this paper. The simplest example of fitness function satisfying this requirement is \(f(x, y) = x\).

Departing from the traditional random walk, the kernel of the (1+1)EA is no longer continuous with respect to the Lebesgue measure \(d\). This is due to the elitism, which causes the associated probability measure to have an atom at zero (a discontinuity of the distribution function). Namely, the algorithm is allowed to move only to the right, any unsuccessful mutation (i.e. to the left) making the EA stagnate in its current state. From this point on we omit the \'(1+1)' tag of the EA, since this is the only type of algorithm we analyze.

The associated one-step kernel can be described as a sum of two measures, one singular (Dirac) and one continuous (note that here \(A \in B(\mathbb{R}^2)\)).

\[
P((x, y), A) = \frac{1}{2} \delta_{(x, y)}(A) + \mathbf{1}_{(x, x + \frac{1}{2}) \times (y - \frac{1}{2}, y + \frac{1}{2})} \cdot d(A) = (8)
\]

\[
= \frac{1}{2} \delta_{(x, y)}(A) + d \left( (x, x + \frac{1}{2}) \times (y - \frac{1}{2}, y + \frac{1}{2}) \cap A \right).
\]

We shall frequently use the 1-dimensional version of (8) corresponding to the progress along the \(x\)-axis, and also its density form:

\[
P(x, A) = \frac{1}{2} \delta_{x}(A) + d((x, x + \frac{1}{2}) \cap A) \quad A \in B(\mathbb{R}) (9)
\]

\[
P(x, du) = \frac{1}{2} \delta_{x}(u) + \mathbf{1}_{(x, x + \frac{1}{2})} \cdot du
\]

where the first term from (9) carries only the null set \(\{x\}\).

Note that the algorithm can still be seen as a special type of random walk. In terms of stochastic process (3), \(z\) is distributed here as half of the sum between a Dirac probability concentrated in zero and a uniform probability of length \(1/2\).

\[
z = \frac{1}{2} \left( \delta_{0} + U_{[0, \frac{1}{2}]} \right). \quad (10)
\]

On the other hand, the 2-step kernel \(P^2\) can be computed as

\[
P^2(0, A) = \int P(0, dx)P(x, A) = \frac{1}{2} P(0, A) + \int_{0^+} P(x, A)dx.
\]

The only discontinuity of this kernel is at zero. This is true not only for \(P^2\) but for any power \(P^n\) with \(n \geq 1\). The analysis in [3] concentrated on characterizing
the n-iteration progress of the EA along the x-axis, by calculating its probability to reach in n iterations the $S_n$ rectangle (figure 1).

\[
S_n = \left\{ (x, y) : \left\{ \begin{array}{c}
\frac{n-1}{2} \leq x < \frac{n}{2}, \\
|y| < \frac{n}{2}
\end{array} \right. \right\} \quad n \geq 1.
\]

Figure 1: Regions of progress for EA with inclined plane fitness landscape

The desired probability was calculated in closed form by the following

**Theorem 3.1** Let $k \geq 0$ and $(x, y) \in S_k$. Then, for all $n \geq 1$ we have

\[
P^n(0, S_n) = \frac{1}{n!} \frac{2^n}{2^n}
\]

The formula above still holds in the case of uniform mutation inside the square of area $r^2$, if we multiply the limits of the interval $S_n$ by $r$.

### 3.1 Cumulative Distribution Function

The CDF for the $k^{th}$ iteration of the EA can be computed in two steps, employing first the formula for the CDF of the sum of $k$ independent and identically distributed (i.i.d.) uniform distributions. This is an old problem, first studied by Lobatchewski, then often revisited [14, 16], even extrapolated to independent non-identically distributed variables. According to (10), we are interested in the CDF of a sum of $k$ i.i.d. uniform variables with basis $[0, 1/2]$. Using definition 2.1 and theorem 2.2 it is easy to prove the formula:

\[
F_k(s) = F^{sk}(s) = \frac{1}{k!} \cdot 2^k \left[ k^k - \sum_{n=0}^{k} \binom{k}{n} \left( s - \frac{1}{2} \right)^k + \ldots + (-1)^k \binom{k}{k} \left( s - \frac{k}{2} \right)^k \right]
\]
where the subscript '+' denotes the positive part of an expression, i.e. the expression itself if positive, zero otherwise. (The '+' operator is applied before the power \( k \).)

But according to (10), our variable \( z \) also has an atom at the origin. Fortunately, the effect of the Dirac measure is not too disruptive. We have the following

**Theorem 3.2** The CDF of the sum of \( n \) i.i.d. random variables with common distribution \( z \) given by (10) - denoted \( X_n \) in (3) - is

\[
F_{X_n}(s) = \sum_{k=0}^{n} \frac{1}{2^n} \binom{n}{k} F_k(s), \quad (12)
\]

where \( F_k \) is given by (11).

**Proof.**

We proceed by induction and consider first two terms in the sum. We need the CDF of \( z \), which is

\[
F_z = \frac{1}{2}(1_{[0,\infty)} + F_{[0,\frac{1}{2})}).
\]

Let us fix \( x > 0 \), arbitrarily (the interesting case). According to definition 2.1 and theorem 2.2,

\[
F_{X_2}(x) = F_z^2(x) =
\]

\[
= (F_z * F_z)(x) = \frac{1}{2^2} \int_{-\infty}^{\infty} \left(1_{[0,\infty)} + F_{[0,\frac{1}{2})}\right)(x-y) d\left(1_{[0,\infty)} + F_{[0,\frac{1}{2})}\right)(y) =
\]

\[
= \frac{1}{2^2} \left[ 1 + F_{[0,\frac{1}{2})}(x-y) \cdot \delta_0(y) + \int_{0}^{x} dF_{[0,\frac{1}{2})}(y) + \int_{-\infty}^{\infty} F_{[0,\frac{1}{2})}(x-y) dF_{[0,\frac{1}{2})}(y) \right] =
\]

\[
= \frac{1}{2^2} \left[ 1 + F_{[0,\frac{1}{2})}(x) + F_{[0,\frac{1}{2})}(x) - F_{[0,\frac{1}{2})}(0) + F_2(x) \right] =
\]

\[
= \frac{1}{2^2} \left[ 1 + 2F_{[0,\frac{1}{2})}(x) + F_2(x) \right].
\]

To complete the induction we need only assume that (12) holds for \( n = m \) and show that it then is true for \( n = m + 1 \). Using the same method for combining CDFs of \( mz \) and \( z \) as was used above for \( z \) and \( z \), this poses no difficulty. □

The next section presents an alternate way of computing the CDF numerically, through discretization of the continuous state space.

### 4 Discretized Markov chain and numerical investigations

The theoretical jump distribution we use in this study is continuous and uniform, with support \((-1/2 \ldots 1/2)\). Algorithms with uniform mutation sampling inside
the hypercube of volume one have also been used in so-called compact EAs [11, 12], belonging to the larger class of *Estimation of Distribution Algorithms* [2]. Such algorithms do not hold a full population of individuals, yet retain the whole information in distribution functions from which new individuals are generated. That makes them more suitable for real-world applications related to portable devices, e.g. for the control of commercial robots.

As a result of continuity, the Markov Chain (MC) governing the evolution of the individual in the evolutionary algorithm (EA) has continuous state. The time of the MC is discrete, so we have a discrete-time, continuous-state MC. Since the time of the MC is always discrete (iterations of the EA), we shall refer to it as the continuous-state MC.

For numerical purposes, we approximate the continuous distribution with a uniform discrete one \( f \). The probability of \( f \) is spread among the \( 2m + 1 \) equidistant points \( \{0, \pm 1/(2m + 1), \pm 2/(2m + 1), \ldots, \pm m/(2m + 1)\} \), and for normalization each point has probability \( 1/(2m + 1) \). Due to the elitism of our EA, the symmetrical distribution \( f \) is transformed into one with about half the mass at zero:

\[
g(x) = \begin{cases} 
  (m + 1)/(2m + 1) & \text{if } x = 0; \\
  1/(2m + 1) & \text{if } x \in \{1/(2m + 1) \ldots m/(2m + 1)\}; \\
  0 & \text{otherwise.}
\end{cases}
\]

In this way we obtain a discrete-time, discrete-state MC. We shall refer to it as the discrete-state MC. The transient and steady-state distributions of the discrete-state MC converge to the corresponding distributions of the continuous-state MC when \( m \to \infty \). The discrete-state MC is well suited to numerical investigations due to its simple structure. Its one-step transition matrix has block-bidiagonal form:

\[
P^1 = \begin{bmatrix}
  A & B & 0 & 0 & \cdots \\
  0 & A & B & 0 & \cdots \\
  0 & 0 & A & B & \cdots \\
  \vdots & \vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\]

Moreover, the matrices repeated on the main diagonal and first upper diagonal are, respectively, upper- and lower-triangular, with repeated values:

\[
A = \begin{bmatrix}
  \frac{m+1}{2m+1} & \frac{1}{2m+1} & \cdots & \frac{1}{2m+1} \\
  0 & \frac{2m+1}{2m+1} & \cdots & \frac{1}{2m+1} \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \cdots & \frac{m+1}{2m+1}
\end{bmatrix} \\
B = \begin{bmatrix}
  \frac{1}{2m+1} & 0 & \cdots & 0 \\
  \frac{2m+1}{2m+1} & \frac{1}{2m+1} & \cdots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  \frac{1}{2m+1} & \frac{1}{2m+1} & \cdots & \frac{1}{2m+1}
\end{bmatrix}
\]

The rows of concatenated \( A \) and \( B \) sum to one, since all transition matrices are stochastic. Both \( A \) and \( B \) are square matrices, of size equal to the discretization parameter \( m \).
All higher-order transition matrices of the discrete-state MC are powers of $P$. For illustration purposes, we present below the 2-step transition matrix:

$$P^2 = \begin{bmatrix} A^2 & AB + BA & B^2 & 0 & 0 & \cdots \\ 0 & A^2 & AB + BA & B^2 & 0 & \cdots \\ 0 & 0 & A^2 & AB + BA & B^2 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

From either probabilistic or purely algebraic reasons, it is clear that the k-step transition matrix $P^k$ has a block-multidiagonal structure, with identical blocks repeated along the main diagonal and upper diagonals. $P^k$ has a non-null main diagonal (consisting of blocks $A^k$) and $k$ non-null upper-diagonals (the uppermost of which consists of blocks $B^k$).

Due to the redundancy of all transition matrices, it is sufficient to calculate their first block-row. The first block-row of $P^k$ has $k + 1$ non-zero blocks. For illustration, we present below these blocks for $k = 1, 2, 3, 4$:

$$P^1 \rightarrow [A, B, 0, 0, 0, \cdots]$$
$$P^2 \rightarrow [A^2, AB + BA, B^2, 0, 0, 0, \cdots]$$
$$P^3 \rightarrow [A^3, A^2B + ABA + BA^2, AB^2 + BAB + B^2A, B^3, 0, 0, \cdots]$$
$$P^4 \rightarrow [A^4, A^3B + A^2BA + ABA^2 + BA^3, A^2B^2 + ABAB + BA^2B + AB^2A + BAB + B^2A^2, AB^3 + BAB^2 + B^2AB + B^3A, B^4, 0, \cdots]$$

Examination of the multiplication $P^k = P^{k-1}P$ reveals the familiar triangular arrangement of binomial coefficients of $(A + B)^k$, with the added twist of matrix multiplication not being commutative. The first block-row of $P^k$ can be obtained directly from the first block-row of $P^{k-1}$ through the following recursion:

$$(\text{first row of } P^k)_i = (\text{first row of } P^{k-1})_{i-1} \cdot B + (\text{first row of } P^{k-1})_i \cdot A$$

We implemented the above formula in a C program, and were able to compute in reasonable time (less than a day) powers of $P$ up to 40, for $m$ up to 1500 on a PC. Once any given $P^k$ has its first block-row computed, the probabilities corresponding to the EA starting at the origin are simply the first row of each block. To obtain the probability for the individual to be in interval $S_i$, we sum up the first row of block $i$. (In the notation of figure 1, the interval indices $i$ start at 1, so the last non-zero block of $P^k$ has index $k + 1$).

The plot in figure 2 presents occupancy distributions (probability mass functions - PMF) obtained from the discrete-state MC, with the same discretization parameter $m = 25$, for $n = 5, 15$ and 25 iterations of the EA. Also plotted are the exact values obtained in Maple from formula (12) for the same respective numbers of iterations. All the PMFs are unimodal, with the maximum occurring in the left-hand-side of their support. The decrease in the tails is sub-exponential. Since the mutation probabilities (9) are i.i.d. with finite variance, the Central Limit Theorem guarantees that the underlying distribution of the

\[\text{The program runs on one of the two cores of an Intel Pentium @ 2.2 GHz.}\]
sum (12) converges to normal. However, the occupancy distributions are two steps removed from a normal, first because it’s not clear how fast the convergence is, and second because of the discrete nature of the regions of progress $S_i$ from figure 1. Future work will attempt a better characterization of this family of distributions.

We note that, within the same PMF, the error is negative for the first few regions, i.e. the occupancy probabilities given by the discrete-state MC are smaller than the real ones. (This is hard to see on the plot, because of the logarithmic scaling.) Somewhere in the vicinity of the maximum, the errors change sign, i.e. the approximate probabilities become larger than the real ones. They stay larger until the end of the distribution, as can be seen in the plot. This behavior of the errors is just an artifact of the discretization process. It can be traced back to the definition of the matrices $A$ and $B$ at the beginning of the current section, where the probability of the point zero, $(m+1)/(2m+1)$, was assigned in $A$ to the first discretization interval, in effect shifting the entire distribution slightly to the right. Consequently, the first row of $B$ has $1/(m+1)$
in the first column, a small but non-zero probability which does not exist in the continuous case. Since the one-step transition distribution is shifted right slightly, some mass is taken away from the left regions and, due to normalization, it ends up in the right regions of the PMF.

To quantify the closeness between the exact PMF (of the continuous-state MC) and the approximate PMF (of the discrete-state MC), we define the error of the \textit{entire} approximate PMF to be the average of the absolute values of the pointwise relative errors:

\[
\text{Average PMF error} = \frac{1}{n} \sum_{k=1}^{n} \left| \frac{PMF_{\text{approx}}(k) - PMF_{\text{exact}}(k)}{PMF_{\text{exact}}(k)} \right| \ast 100 \quad (13)
\]

The use of relative errors (over absolute) is motivated by the convergence to zero of the tail probabilities. Absolute errors would assign practically all weight to only the left side of the distribution, whereas relative errors scale the weight of each error naturally. The multiplication by 100 denotes percent.

In figure 3 we plot the average PMF error for the same number of iterations of the EA \((n = 15)\) over a wide range of values of \(m\). The plot confirms that the distribution of the discrete-state MC converges to the one of the continuous-time MC as \(m\) goes to infinity. The same type of dependency was obtained for the entire range of iterations investigated \((n = 2 \ldots 40)\).

Figure 3: Convergence of approximate PMF to exact PMF of the 15-step EA, as the discretization parameter \(m\) increases.

To see how ‘expensive’ the approximation is, we set a threshold of 5% on the average PMF error, and found numerically for each \(n\) a corresponding smallest value of the discretization parameter \(m\) that gets the error under the threshold. The results are plotted in figure 4.

We performed least squares fitting with a variety of curves (line, parabola, cubic, quartic, various exponentials) and found the best match to be the following exponential (RMS of error 1.21):

\[
m(n) = -20.00 + 10.34 \cdot n^{1.35} \quad (14)
\]

Given the exponential form of the \(n\)-step transition matrix \(P^k\), this was expected.
Figure 4: Smallest discretization parameter $m$ that makes the average PMF error less than 5% (square markers) and linear least-squares regression (triangle markers).

5 First Hitting Time

We show now that the expected hitting time for the (1+1) EA of the target zone $S_n$ is $O(n)$. To this end, one needs upper and lower bounds for the expected value $E(N(t))$, bounds that are given in the following

Theorem 5.1 Let $\hat{N} = \{N(t)\}_{t \in [0, \infty)}$ be the renewal process associated to $\hat{X} = \{X_n\}_{n \geq 1}$, and assume further that $m = E(X_1) = E(z) < \infty$. Then

$$\frac{t}{m} < E[N(t)] + 1 < \frac{t + \frac{1}{2}}{m}, \quad \text{for all } t > 0. \quad (15)$$

Proof.

Let us fix $t$, an arbitrary point on the positive semi-axis. Then we can imagine the following sequence on the real line:

$$0 < X_1 < X_2 < \ldots < X_n < t < X_{n+1}.$$
using the obvious bound $z_n \leq 1/2$ satisfied for all $n$, and $1 + N(t)$ to be the stopping time associated to $\hat{X}$ and $t$ by (4), the previous ordering gives

$$X_{N(t)+1} - \frac{1}{2} \leq X_{N(t)+1} - z_{N(t)} \leq X_{N(t)} < t < X_{N(t)+1}$$

and finally

$$t < X_{N(t)+1} < t + \frac{1}{2}. \quad (16)$$

Wald’s identity (5) applied to the stopping time $N(t) + 1$

$$E \left[ X_{N(t)+1} - m(N(t) + 1) \right] = 0$$

and (16) lead to

$$E \left[ t - m(N(t) + 1) \right] < 0 < E \left[ t + \frac{1}{2} - m(N(t) + 1) \right]$$

then to

$$t - m(E[N(t)] + 1) < 0 < t + \frac{1}{2} - m(E[N(t)] + 1)$$

and finally, by reversing the inequalities, we obtain

$$m(E[N(t)] + 1) - t - \frac{1}{2} < 0 < m(E[N(t)] + 1) - t$$

which provides the conclusion. \qed

Now, the bounds for the (1+1) EA’s first hitting time are straightforward:

**Corollary 5.2** Let the (1+1) EA moving on the inclined plane according to equation (10), and $S_n = \{(x, y) : \frac{n-1}{2} \leq x < \frac{n}{2}, \; |y| < \frac{n}{2}\}$, with $n \geq 1$ arbitrarily fixed be the target zone. Then the expected first hitting time of $S_n$ is

$$HT(S_n) := E \left[ N \left( \frac{n-1}{2} \right) \right] + 1 \quad (17)$$

which satisfies

$$4n - 4 < HT(S_n) < 4n. \quad (18)$$

**Proof.**

A simple computation shows that the expected value $m = E(z)$ is in our case $1/8$. We substitute this value, along with $t = (n - 1)/2$, in (15), and the result follows. \qed

Moreover, from theorem 2.4 one can derive the asymptotics of the hitting time.
Corollary 5.3 The expected first hitting time of $S_n$ satisfies also

$$HT(S_n) = 4n - \frac{8}{3} + o(1), \quad \text{for } n \to \infty. \quad (19)$$

Proof. First compute the variance of $z$:

$$\sigma^2 = \frac{5}{192}.$$

Replace $\sigma^2$, $m = 1/8$ and $t = (n - 1)/2$ in the formula of theorem 2.4 and add one to the expression of $E[N(t)]$ in order to get the proper hitting time. □

Figure 5: Bounds, asymptotic limit and approximate values of hitting times for the intervals $S_i$.

A final numerical test was performed to confirm the hitting time bounds and asymptotic limit obtained theoretically in formulas (15) and (19). In order to calculate the hitting time for the discretized MC, a hitting time distribution (PMF) was calculated for each interval $S_i$. For each $S_i$, the hitting probabilities for iterations up to $i$ are of course zero, since the first iteration the individual can reach $S_i$ is $i + 1$. Then $HT(i)$ was calculated as a weighted sum, according to its corresponding definition for the discrete case - see e.g. [9].

15
As seen in figure 5, the approximate hitting time tracks the asymptotic limit very closely (all errors less than 1%) even for moderate values of \( m \) (400 in the figure) and all iteration numbers \( n \) examined (all the way up to \( n = 30 \)). This suggests that the bounds (15) are not tight, and, moreover, that the actual values of the hitting time may actually coincide with the limit for all indices \( i^! \) Future work will cast light on this conjecture.

6 Conclusions

For the theoretical analysis of Evolutionary Algorithms (EA) to be of practical relevance, it should provide, besides necessary and sufficient conditions for global convergence, case studies of specific algorithms presented with particular fitness landscapes. In continuous space, significant insight into the algorithm’s global behavior can be gained by analyzing the associated homogeneous Markov chain (MC). Based on manipulations of the MC’s transition function, we can derive various probabilistic measures of interest, like the probability distribution of occupancy of general target zone at arbitrary iterations, the distributions of hitting times, or the average hitting times for those target zones. This analysis is carried out in the present paper for the one-individual EA, with uniform mutation, on the inclined plane fitness landscape.

We derived a closed-form expression for the cumulative distribution function (CDF) of occupancy of a general family of intervals in the search-space. This exact formula was first reached theoretically, by using the convolution of independent random variables. An alternative, numerical approach, based on discretization and multiplication of the corresponding transition matrix confirms the theoretical formula, by generating approximate results which converge to those predicted by the formula. Numerical explorations suggest that the tails of these distributions are sub-exponential.

The formalism of renewal processes was then used to derive upper and lower bounds, together with an asymptotic limit for the hitting time of the intervals previously defined. Again, the (numerical) discretization method confirms the results with great precision. Moreover, it suggests that the theoretical bounds are not tight, and instead the hitting time is actually equal to the asymptotic limit for all target zones.

There are two reasons for performing the theoretical and numerical analyses in parallel: First, since a similar theoretical derivation had not been done before for EAs, we needed numerical confirmation. Second, as the agreement between the two methods proved to be so good, we intend to use the discretization for further analysis of the same model, and also for more complex algorithms and/or fitness landscapes (see below). From the work so far, it is clear that the numerical approximation “reaches” farther than the theoretical formulas: in the case of hitting times, the theory only gave bounds and limits, but numerically we obtained complete probability mass functions of the hitting times.

The results reported in the paper have brought up some interesting directions for future inquiry. One is the better characterization of the occupancy
distributions pictured in figure 2. It would be useful to know how fast they converge to the theoretical normal as the number of EA iterations increases. The other is the conjecture that the asymptotic limit for the hitting times is not a limit after all, but the actual exact value taken by the hitting times, even for low indices of the target zone. The numerical results obtained through discretization point in this direction, so maybe the theoretical bounds can be improved upon and even collapsed to the asymptote.

The work presented in this paper has dealt naturally with the simplest choices for all “ingredients” of evolutionary optimization: the algorithm is (1+1)EA, the mutation distribution is uniform, and the fitness landscape is linear. We are already applying the tools developed here (continuous and discrete MC analysis, renewal processes, stopping times) to more complex/realistic EAs, distributions and landscapes. Here is a short list of pointers:

i The Markov kernel of the algorithm with uniform mutation on the corridor landscape has been already analyzed in [3]. An extended computational time study based on renewal processes for the (1+1) EA with uniform mutation inside the sphere optimizing n-dimensional quadratic functions is the subject of a paper in progress [1].

ii The mutation distribution most often used in continuous EAs is normal, not uniform. Although the direct extension of our method to normal poses some analytical challenges, we’ve made advances with other well-known distributions with infinite support (Erlang and chi-square).

iii The discrete analysis performed in section 4 can still be performed for mutation distributions with infinite support, after truncation. This is a major direction for future work.

iv The first hitting time analysis based on renewal processes presented in section 5 does not rely on the particular form of the mutation operator, but only on the expected value and variance of the random variable one-step progress. Depending on the complexity of the fitness function, the moments of this random variable could be computed either analytically or numerically.

Acknowledgments

The first author acknowledges support from DAAD (German Academic Exchange Service), study and research grant A/10/05445, and from CNCSIS-UEFISCSU, project number 844 PNII - IDEI, code 1778/2008. The second author acknowledges financial support from an Organized Research Grant at Tarleton State University.
References


