Estimating the transition kernel of continuous EAs. Case of inclined plane model*

Alexandru Agapie
Institute for Mathematical Statistics and Applied Mathematics,
Romanian Academy
agapie@rdslink.ro

Abstract

A static parameter EA working on continuous search space may be regarded as general homogeneous Markov chain. The finite case problem of determining the transition matrix turns into the problem of finding (approximating) a transition probability function (kernel). This function should define the probabilities of moving from the starting point to any set of positive measure in the space. Considering a $(1+1)$ EA with square uniform mutation in the two-dimensional real space, we analyze the transition kernel for the inclined plane model.

1 Introduction

A good introduction to Evolutionary Algorithms (EAs) theory could be provided by the analysis of random walk. Unfortunately, even the most exhaustive monographs on random walk confine to the case of discrete search space only - see, e.g. (Spitzer, 1976) - which is of little help when considering the continuous case. On the other hand, one of EAs' relatives, Simulated Annealing, is known to have theoretical convergence results, even for the continuous search space (Haario and Saksman, 1991). Yet, there are two major impediments making that approach inappropiate for extrapolation. First, the proof from (Haario and Saksman, 1991), involving a double chain convergence procedure (considering a chain of homogeneous Markov chains, each of them converging to a stationary distribution, gathering in a string that converges to a distribution concentrated on the optimal points) does not imply convergence in probability for the practical algorithm (even in infinite time). Second, that result relies strongly on the special selection form given by the Boltzman distribution, which is not used in EAs.

On the other hand, both the Robins-Monroe family of algorithms usually referred to as stochastic approximation (Wasan, 1969), and the martingale approach to random search algorithms performed in (Rapple, 1989) could be criticized for assuming a certain positive success rate at each iteration, which corresponds to local rather than global behaviour.

Depart from the quoted references, we intend to develop an EA theory based on its associated transition probability function. To start with, we consider the two-dimensional real space and fix $0 =: (0,0)$ as the starting point of the algorithm. Let $\{z_i\}_{i\geq 1}$ be a sequence of independent identically distributed random variables, distributed uniformly inside the square of area one and center zero. Then a static-parameter

[0]Work supported by the *Collaborative Research Center "Computational Intelligence" (SFB 531)*, Dortmund University.

(1,1) EA in the plane (which coincides with a random walk) is defined by the random variable Y_n, for all $n \geq 1$:

$$Y_n = 0 + \sum_{i=1}^{n} z_i = \sum_{i=1}^{n} z_i \tag{1.1}$$

Obviously, the sequence $\{Y_n\}_{n \geq 1}$ defines a Markov chain in the plane. We need next the notion of *transition probability function* (Nummelin, 1984). For the purpose of the subsequent analysis, the case of two-dimensional real space and associated Borel σ-algebra of sets is sufficient. We shall also use its one-dimensional version, with a straightforward understanding.

Definition 1.1 *A function $P : R^2 \times \mathcal{B}(R^2) \to [0,1]$ is said to be a* transition probability function (kernel) *if $P(w, \bullet)$ is a probability on $\mathcal{B}(R^2)$ for all $w \in R^2$, and $P(\cdot, A)$ is a random variable on R^2 for all $A \in \mathcal{B}(R^2)$.*

As for the optimization task, we consider the inclined plane model, describing the simplest dependence between the objective function and the variables of an n-dimensional real space. Following (Schwefel, 1995), we orient the coordinate system so that the plane only slopes in the direction of x_1 axis (thus $x_1 = \infty$ would correspond to the minimum) and consider the origin as the starting point of the (1+1) EA.

2 (1+1) EA on the inclined plane

In this section we consider the (1+1) EA with square uniform mutation starting in zero, as in Section 1. Depart from random walk, the transition kernel of the one-step (1+1) EA is no longer continuous with respect to the Lebesgue measure on the plane. This comes from the *elitist* property - as formalized for the continuous case by (Rudolph, 1997) - which makes the associated probability law to have an *atom* at zero (a point of discontinuity for the distribution function). Namely, the algorithm is allowed to move only to the right (see also figure 1), any unsuccessful (that is, to the left) *mutation* making the EA stagnate in the initial point 0.

The associated one-step transition kernel can be described as a sum of two measures, one singular (also called *Dirac*) and one continuous (note that d stands for the Lebesgue measure, both on the plane and on the real line, as previously)

$$\begin{aligned} P((x,y),A) &= \frac{1}{2}\delta_(x,y)(A) + 1_{(x,x+\frac{1}{2})\times(y-\frac{1}{2},y+\frac{1}{2})} \cdot d(A) \\ &= \frac{1}{2}\delta_(x,y)(A) + d((x,x+\frac{1}{2}) \times (y-\frac{1}{2},y+\frac{1}{2}) \cap A), \qquad \text{for all } A \in \mathcal{B}(R^2) \end{aligned} \tag{2.1}$$

We shall frequently use the one-dimensional version of (2.1) corresponding to the progress along the x-axis and also its 'density-like' expression, namely

$$P(x,A) = \frac{1}{2}\delta_x(A) + d((x,x+\frac{1}{2}) \cap A), \qquad \text{for all } A \in \mathcal{B}(R) \tag{2.2}$$

$$P(x,du) = \frac{1}{2}\delta_x(u) + 1_{(x,x+\frac{1}{2}]} \cdot du \tag{2.3}$$

where the first term from (2.2) 'carries' only the (Lebesgue) null set $\{x\}$.

One should notice the fact that for all fixed $x, y \in R$, formulas (2.1)-(2.2) define probability measures with respect to A. Yet, it is not obvious that $P(\cdot, A)$ is a measurable function, for any Borel set A. This is made clear by the following lemma. For simplicity we consider just one dimension and a set A of the form $A = (a, b)$. For two dimensions, one should notice that movements are independent with respect to the x and y axis, which makes the contribution of the y-term to be multiplicative.

Lemma 2.1 *Let $[a, b)$ be a fixed interval of the real line. The restriction to the x-axis of the transition kernel corresponding to the (1+1) EA with uniform mutation of 'radius' $1/2$ is a measurable function, given by*

$$P(x, [a, b)) = \begin{cases} 0 & , x < a - \frac{1}{2} \\ min\{x + \frac{1}{2}, b\} - a & , x \in [a - \frac{1}{2}, a) \\ \frac{1}{2} + min\{x + \frac{1}{2}, b\} - x & , x \in [a, b) \\ 0 & , x \geq b. \end{cases} \tag{2.4}$$

Proof.
On each branch in (2.4) we have the probability for the one-step EA to move from x to $[a, b)$, calculated in general as the Lebesgue measure of the intersection between $[a, b)$ and the $1/2$-radius interval centered in x, formula (2.2). The only case to watch out for is $x \in [a, b)$, where the *non-Lebesgue* factor $1/2$ is added, as the probability of remaining in x due to elitism. (We assume progress direction to the right.) Next, as each branch function is measurable with respect to x, so is $P(x, [a, b))$. And as $\mathcal{B}(R)$ is generated by the intervals of type $[a, b)$, $P(\cdot, A)$ is measurable for any Borel set A. $\square$

The following result shows that the product measure $d \cdot P$ is *absolutely continuous* with respect to d, which will be used in characterizing the n-step transition functions P^n.

Lemma 2.2 *Let $A \in \mathcal{B}(R)$ be a set with $d(A) = 0$. Then $\int P(x, A)dx = 0$.*

Proof.

$$\int P(x, A)dx = \int_A P(x, A)dx + \int_{A^C} P(x, A)dx$$

The function under the second integral sign from the sum is always zero, for $P(x, \bullet)$ restricted to $\mathcal{B}(R \setminus \{x\})$ is absolutely continuous with respect to Lebesgue measure - as stated by formula (2.2) - thus it will *not* carry A with $d(A) = 0$. So the corresponding integral will be zero and, keeping in mind that for any x , $P(x, \bullet)$ is a probability, the remaining term reads

$$\int_A P(x, A)dx \leq \int_A 1dx = d(A) = 0$$

$\square$

Now we can consider the n-step transition kernel P^n. Assume as usual that the algorithm starts in zero, and furthermore restrict the search to the real line. Applying the *chain rule* we get the following characterization.

$$P^2(0, A) = \int P(0, dx)P(x, A) = \frac{1}{2}P(0, A) + \int_{0+}^{1/2} P(x, A)dx \qquad \text{for all } A \in \mathcal{B}(R)$$

The only discontinuity point of this kernel is zero. This stands not only for P^2 but for any power P^n with $n \geq 1$, as stated by the following.

Proposition 2.3 *Let P be the kernel of a one-dimensional (1+1) EA and $A \in \mathcal{B}((0,\infty))$ with $d(A) = 0$. Then, for any $n \geq 1$ we have:*

$$P^n(0, A) = 0 \quad \text{(i)}$$

$$\int P^n(x, A)dx = 0. \quad \text{(ii)}$$

Proof.
The case $n = 1$ corresponds to definition (2.2) (i), respectively to lemma 2.2 (ii). Next, applying the chain rule to P^n and assuming by induction that the statement holds for $n-1$, we get

$$P^n(0, A) = \int P(0, dx)P^{n-1}(x, A) = \frac{1}{2}P^{n-1}(0, A) + \int_{0+}^{1/2} P^{n-1}(x, A)dx$$

$$\leq 0 + \int P^{n-1}(x, A)dx = 0$$

□

Returning to the inclined plane model, we shall derive in the following an exact formula for the n-step progress of the (1+1) EA along the x-axis by calculating the probability of reaching in n iterations the S_n rectangle, $S_n = \{(x, y) : \frac{n-1}{2} \leq x < \frac{n}{2}, |y| < \frac{n}{2}\}$ for all $n \geq 1$. A representation of the progress regions $\{S_n\}$ for $n = 1, 4$ is given in figure 1.

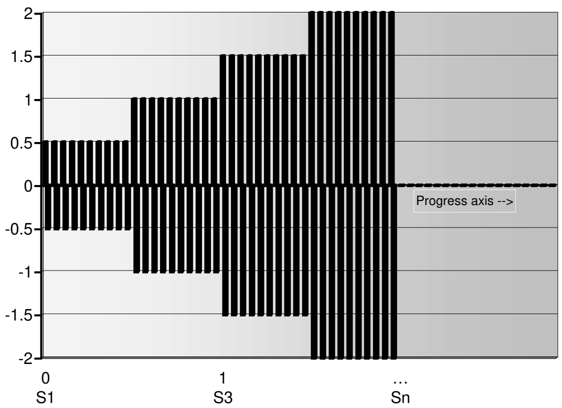

Figure 1: Regions of progress for the $(1 + 1)$ EA on the inclined plane model

To this end, we start with a rather intuitive result, stating that the transition kernel P is invariant to translations along the progress axis. Let $S_n^1 = [\frac{n-1}{2}, \frac{n}{2})$ be the x-projection of S_n, for all n.

Lemma 2.4 *For all $n \geq 1$, $k < n$, $m \geq 1$ and $x \geq \frac{k}{2}$ we have*

$$P^m(x, S_n^1) = P^m(x - \frac{k}{2}, S_{n-k}^1)$$

Proof.
We fix arbitrary n and $k < n$ and proceed with induction after m. For $n = 1$ formula (2.4) provides

$$P(x, S_n^1) = P(x, [\frac{n-1}{2}, \frac{n}{2})) = \begin{cases} 0 & , x \leq \frac{n-2}{2} \\ x + \frac{1}{2} - \frac{n-1}{2} & , x \in (\frac{n-2}{2}, \frac{n-1}{2}) \\ \frac{1}{2} + \frac{n}{2} - x & , x \in [\frac{n-1}{2}, \frac{n}{2}) \\ 0 & , x \geq \frac{n}{2}. \end{cases}$$

$$= \begin{cases} 0 & , x - \frac{k}{2} \leq \frac{n-k-2}{2} \\ (x - \frac{k}{2}) + \frac{1}{2} - \frac{n-k-1}{2} & , x - \frac{k}{2} \in (\frac{n-k-2}{2}, \frac{n-k-1}{2}) \\ \frac{1}{2} + \frac{n-k}{2} - (x - \frac{k}{2}) & , x - \frac{k}{2} \in [\frac{n-k-1}{2}, \frac{n-k}{2}) \\ 0 & , x - \frac{k}{2} \geq \frac{n-k}{2}. \end{cases}$$

$$= P(x - \frac{k}{2}, [\frac{n-k-1}{2}, \frac{n-k}{2})) = P(x - \frac{k}{2}, S_{n-k}^1).$$

Assume now that the property holds up to $m - 1$, and calculate

$$P^m(x, S_m^1) = \int P(x, du) P^{m-1}(u, S_n^1) = \int P(x, du) P^{m-1}(u - \frac{k}{2}, S_{n-k}^1)$$

$$= \int P(x - \frac{k}{2}, du) P^{m-1}(u - \frac{k}{2}, S_{n-k}^1) = P^m(x - \frac{k}{2}, S_{n-k}^1)$$

where the last equalities come from the induction hypotheses, cases $m - 1$ and 1. $\square$

The following is essential for characterizing the n-step transitions.

Proposition 2.5 *Let $x \in S_1^1$. For all $n \geq 1$ we have*

$$P^n(x, S_{n+1}^1) = \frac{1}{n!} x^n$$

Proof.
We proceed by induction. For $n = 1$ we apply formula (2.4) with $[a, b) = [\frac{1}{2}, 1)$ and get

$$P(x, [\frac{1}{2}, 1)) = x + \frac{1}{2} - \frac{1}{2} = x$$

Next, assume the condition holds for $n-1$ and calculate

$$P^n(x, S^1_{n+1}) = \int P(x, du) P^{n-1}(u, S^1_{n+1}) =$$

$$= \frac{1}{2} P^{n-1}(x, S^1_{n+1}) + \int_x^{x+\frac{1}{2}} P^{n-1}(u, S^1_{n+1}) du =$$

where the first term is zero because S^1_{n+1} is *not* accesible from $x < \frac{1}{2}$

$$= \int_x^{\frac{1}{2}} P^{n-1}(u, S^1_{n+1}) du + \int_{\frac{1}{2}}^{x+\frac{1}{2}} P^{n-1}(u, S^1_{n+1}) du =$$

where first term is again zero, then lemma 2.4, the change $u - \frac{1}{2} = v$ and induction provide

$$= \int_{\frac{1}{2}}^{x+\frac{1}{2}} P^{n-1}(u - \frac{1}{2}, S^1_n) du = \int_0^x P^{n-1}(v, S^1_n) dv = \int_0^x \frac{1}{(n-1)!} v^{n-1} dv = \frac{1}{n!} x^n$$

□

A simple result makes the step to the two-dimensional case, namely

Lemma 2.6 *Let $k \geq 0$ and $(x, y) \in S_k$. Then, for all $n \geq 1$ we have*

$$P^n((x, y), S_{n+k}) = P^n(x, S^1_{n+k})$$

Proof.
Induction after n. For $n = 1$ formula (2.1) provides

$$P((x, y), S_{k+1}) = \frac{1}{2} \delta_(x, y)(S_{k+1}) + d((x, x + \frac{1}{2}) \times (y - \frac{1}{2}, y + \frac{1}{2}) \cap S_{k+1})$$

$$= \frac{1}{2} \delta_(x)(S^1_{k+1}) + 1 \cdot d((x, x + \frac{1}{2}) \cap S^1_{k+1}) = P(x, S^1_{k+1})$$

Assume the equality holds for $n-1$ and calculate

$$P^n((x, y), S_{n+k}) = \int P((x, y), dw) P^{n-1}(w, S_{n+k})$$

$$= \frac{1}{2} \delta_(x, y)(w) P^{n-1}((x, y), S_{n+k}) + \int_x^{x+\frac{1}{2}} du \int_{y-\frac{1}{2}}^{y+\frac{1}{2}} P^{n-1}((u, v), S_{n+k}) dv$$

$$= \frac{1}{2} P^{n-1}(x, S^1_{n+k}) + \int_x^{x+\frac{1}{2}} du \int_{y-\frac{1}{2}}^{y+\frac{1}{2}} P^{n-1}(u, S^1_{n+k}) dv$$

$$= \frac{1}{2} P^{n-1}(x, S^1_{n+k}) + 1 \cdot \int_x^{x+\frac{1}{2}} du P^{n-1}((u, v), S_{n+k})$$

$$= \int P(x, du) P^{n-1}(u, S_{n+k}) = P^n(x, S_{n+k})$$

$\square$

Now we can state the main result for the inclined plane model.

Theorem 2.7 *Let $k \geq 0$ and $(x, y) \in S_k$. Then, for all $n \geq 1$ we have*

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

Proof.
According to lemma 2.6 it is sufficient to prove the one-dimensional result. We have

$$P^n(0, S_n^1) = \int P(0, dx) P^{n-1}(x, S_n) = \int_0^{\frac{1}{2}} P(0, dx) P^{n-1}(x, S_n)$$

$$= \int_0^{\frac{1}{2}} P(0, dx) \frac{x^{n-1}}{(n-1)!} = \int_0^{\frac{1}{2}} \frac{x^{n-1}}{(n-1)!} dx = \frac{1}{n!\, 2^n}$$

where in the last line proposition 2.5 and the fact that S_n^1 is not attainable from 0 in $n-1$ steps were essential. $\square$

The following corollary generalizes the analysis to the case of $(1+1)$ EA with r-size mutation.

Corollary 2.8 *If mutation is uniformly distributed inside the square of area r^2*

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

where in this case $S_n^r = \{(x, y) : \frac{r(n-1)}{2} \leq x < \frac{nr}{2}, |y| < \frac{nr}{2}\}$

Proof.
If one carries the same reasoning leading to theorem 2.7 but integrating from 0 to $r/2$ along the x-axis, respectively from $-r/2$ to $r/2$ along the y-axis, and divide by a factor of r^2 each time P appears in the calculation, the conclusion is straightforward.

References

Haario, H. and Saksman, E. (1991). Simulated Annealing Process in General State Space. *Advances in Applied Probability 23*, 866–893.

Rapple, G. (1989). On linear convergence of a class of random search algotithms, in *Zeitschrift für angewandte Mathematik und Mechanik (ZAMM) 69*(1), 37–45.

Rudolph, G. (1997). *Convergence Properties of Evolutionary Algorithms*, Verlag Dr. Kovać, Hamburg.

Nummelin, E. (1984). *General Irreducible Markov chains and non-negative operators*, Cambridge University Press, Cambridge.

Schwefel, H.-P. (1997). *Evolution and Optimum Seeking*, Wiley, New York.

Spitzer, F. (1976). *Principles of Random Walk*, Springer, New York.

Wasan, M.T. (1969). *Stochastic Approximation*, Cambridge University Press, Cambridge.