Quasi-Stability of Real Coded Finite Populations

Jarosław Arabas and Rafał Biedrzycki

Institute of Electronic Systems, Warsaw University of Technology, Poland {jarabas,rbiedrzy}@elka.pw.edu.pl

Abstract. This contribution analyzes dynamics of mean and variance of real chromosomes in consecutive populations of an Evolutionary Algorithm with selection and mutation. Quasi-stable state is characterized with an area in which population mean and variance will remain roughly unchanged for many generations. Size of the area can be indirectly estimated from the infinite population analysis and is influenced by the population size, selection type and parameter, and the mutation variance. The paper gives formulas that define this influence and illustrates them with numerical examples.

Keywords: Population diversity, response to selection, quasi-stability.

1 Introduction

Quasi-stability of populations is a state of an Evolutionary Algorithm (EA) when populations fluctuate in the same area of the search space for many consecutive generations. In effect, although the actual mean and variance of position of chromosomes are variable, they fluctuate around characteristic values which do not change. Expected values of population mean and variance have been analytically derived in [1] assuming that the fitness function is Gaussian (for fitness proportionate selection) or odd and concave (for other selection schemes).

Dynamics of the EA is usually put in the context of convergence or expected hitting time. Here another perspective is taken, where more attention is paid to the population diversity. Several authors have already taken this perspective before in the area of real coded EAs. For example, Beyer and Deb [2] analyzed the dynamics of the population mean and variance assuming a "flat fitness" model (identical selection probabilities) and various crossover schemes, without mutation. They identified a risk for an EA with identical selection probabilities that the population variance may collapse or blow up if the crossover parameters are improperly set and they came up with admissible levels of these parameters. They have also performed similar analysis for the Evolution Strategy for the flat fitness. Works of Arnold have gone in a similar direction. He considered the dynamics of central moments of the population distribution for the Evolution Strategy, assuming a random fitness and mixtures of a deterministic fitness function and noise [3,4].

An important role in the analysis of EA dynamics in real spaces plays the infinite population model introduced by Qi and Palmieri [5]. They considered the

T. Bartz-Beielstein et al. (Eds.): PPSN XIII 2014, LNCS 8672, pp. 872–881, 2014.

[©] Springer International Publishing Switzerland 2014

dynamics of populations assuming that they are infinite. This allowed to apply deterministic equations to transform probability density functions of consecutive populations. Their results have been adopted by Karcz-Duleba to analyze the dynamics of populations for functions with a single optimum and with two local optima [6].

In this paper we use results of the population diversity analysis based on infinite population model which have been published in [1] for various types of selection, with and without crossover, with and without elitism. We show that the quasi-stable behavior of an finite population EA can be well explained by this model.

2 Subject of Analysis

We consider a simple EA (see Fig. 1) which combines selection and mutation to maximize a fitness function $q: R \to R$. The algorithm in the *t*-th generation is characterized by the base population P^t which contains μ individuals. Each individual P_i^t is a real number. In each generation, a population R^t of reproduced individuals is created by selecting with replacement individuals from the population P^t . Each reproduced individual R_i^t undergoes Gaussian mutation with variance v_m and yields an offspring which becomes the *i*-th member of the base population for the next generation. The algorithm is stopped after reaching a specified number of generations T.

```
 \begin{split} &\text{initialize}(P^0) \\ &\text{evaluate}(P^0) \\ &\text{for } t = 1 \text{ to } T \text{ do} \\ &\text{for } i = 1 \text{ to } \mu \text{ do} \\ & R_i^t \leftarrow P_j^t \text{ where } j \leftarrow \text{select}(P^t) \\ & P_i^{t+1} \leftarrow R_i^t + M_i^t \text{ where } M_i^t \sim N(0, v_m) \\ &\text{ end for} \\ &\text{ evaluate}(P^{t+1}) \\ &\text{end for} \end{split}
```

Fig. 1. Outline of the EA under consideration

Empirical moments of the population P^t , the mean $\overline{P^t}$ and the sample variance $s^2(P^t)$, are defined as

$$\bar{P^t} = \frac{1}{\mu} \sum_{i=1}^{\mu} P_i^t \qquad s^2(P^t) = \frac{1}{\mu - 1} \sum_{i=1}^{\mu} (P_i^t - \bar{P^t})^2 \tag{1}$$

Population P^t can be characterized by its state $u_t = [\bar{P}^t, s^2(P^t)]$ which is defined as a pair of mean and sample variance of chromosomes contained by P^t . Thus a single EA run generates a trajectory of states u_t . The infinite population models have been used to analyze expected values of empirical moments of populations [1,6]. Infinite population size facilitates the analysis since the transformation state u_t to u_{t+1} is deterministic according to equations

$$m_P^{t+1} = m_R^t, \qquad v_P^{t+1} = v_R^t + v_m$$
 (2)

where m_R^t and v_R^t are the expected value and variance of reproduced chromosomes

$$m_R^t = \int_{-\infty}^{\infty} x f_R^t(x) dx, \qquad v_R^t = \int_{-\infty}^{\infty} (x - m_R^t)^2 f_R^t(x) dx$$
 (3)

and f_R^t is the p.d.f. of reproduced points provided that P^t is infinite and distributed with expectation m_P^t and variance v_P^t . If the fitness function is unimodal and even then iteration of equations (2) yields stable values m_P and v_P which have been derived in [1] for various types of selection, with and without crossover, with and without elitism. In general, distribution of f_P^t may be different than normal. Yet, as it has been shown in [1], values of m_P and v_P can be analytically derived with a small error assuming normality of f_P^t .

3 Finite Populations Generated with Stable Expectation and Variance

For a considered EA, populations will never settle down in any position, since state u_{t+1} relates to u_t in a stochastic fashion. Nevertheless, it may be usually observed that populations will fluctuate for many generations in a certain area. In effect, mean values of the population mean and variance in consecutive populations will agree with values predicted from the infinite population model, as it has been shown in [1]. On the other hand, if the population size is finite and variance of chromosomes is bounded, then stabilization of populations will be observed for functions that are not necessarily unimodal — it is only necessary that fitness function is unimodal in sufficiently large neighborhood of local maximum which is covered by all possible locations of populations. Size of this neighborhood can be predicted by analyzing the distribution of mean and variance of finite size populations which are generated by the EA, whose stable expectation and variance have been predicted with the infinite population model.

The population P^t is modeled as a vector of variates of random variables $P_1, \ldots P_{\mu}$ which generate chromosomes $P_1^t, \ldots P_{\mu}^t$. Since each chromosome is generated according to the same procedure, variables $P_1, \ldots P_{\mu}$ are identically distributed. A stable state $u_P = [m_P, v_P]$ of an infinite population model is assumed. This means that all populations are approximated as if they were generated with the normal distribution with expectation m_P and variance v_P .

Empirical moments of P^t , defined by equation (1), are estimators of theoretical moments. Since the contents of population is random, empirical moments themselves are random variables. It holds

$$E[\bar{P^t}] = m_P, \qquad E[s^2(P^t)] = v_P \tag{4}$$

where $E[\cdot]$ denotes the expected value. Further on, it is assumed that $m_P = 0$ to simplify notation without generality loss. If the chromosomes were independent and normally distributed then the population mean would be normally distributed with variance $V[\bar{P}^t] = m_P/\mu$ and the population variance would be chi-square distributed with variance $V[s^2(P^t)] = 2v_P/(\mu - 1)$.

Chromosomes contained by the population P^t result from selection with replacement and mutation of chromosomes from P^{t-1} . For this reason, it is possible that a pair of chromosomes $P_i^t, P_j^t \in P^t$ was generated by mutation of the same parent reproduced from the population P^{t-1} . Therefore random variables $P_1, \ldots P_{\mu}$ which generate populations are not independent. Variance of the population mean is then defined as

$$V[\bar{P}^{t}] = E\left[\frac{1}{\mu^{2}}\sum_{i=1}^{\mu}\sum_{j=1}^{\mu}P_{i}^{t}P_{j}^{t}\right] = E\left[\frac{1}{\mu^{2}}\sum_{i=1}^{\mu}(P_{i}^{t})^{2} + \frac{2}{\mu^{2}}\sum_{i=1}^{\mu-1}\sum_{j=i}^{\mu}P_{i}^{t}P_{j}^{t}\right]$$
$$= \frac{1}{\mu}E[(P^{t})^{2}] + \frac{\mu-1}{\mu}E[P_{i}^{t}P_{j}^{t}|i\neq j] = \frac{1}{\mu}v_{P} + \frac{\mu-1}{\mu}r_{P}^{t}$$
(5)

where $E[\cdot|\cdot]$ is the conditional expected value and r_P^t is the covariance coefficient between any pair of variables P_i^t, P_j^t .

In a pair of populations P^t and P^{t+1} , any point P_i^{t+1} will result from mutation of a reproduced point, hence $P_i^{t+1} = R_k^t + M_i^t$, where M_i^t is the *i*-th mutation in generation *t* which is independent of all other mutations; therefore

$$r_P^{t+1} = E\left[(R_i^t + M_i^t)(R_j^t + M_j^t)\right] = E\left[R_i^t R_j^t\right]$$
(6)

To derive value of r_P^{t+1} consider a pair of distinct chromosomes P_i^{t+1} and P_j^{t+1} . Observe that two scenarios are possible. In the first scenario, which holds with probability a, both chromosomes will be mutants of the same reproduced chromosomes R_k^t . In the second scenario, their parents will be distinct:

$$r_P^{t+1} = aE\left[(R^t)^2\right] + (1-a)E\left[R_i^t R_j^t | i \neq j\right], \qquad a = \sum_{k=1}^{\mu} (p_s(k))^2 \qquad (7)$$

Symbol $p_s(k)$ stands for the selection probability of the k-th chromosome. Value of $E\left[(R^t)^2\right]$ can be computed by observing that each chromosome from R^t results from sampling with replacement from P^t . Hence the variance of points which are sampled from P^t for reproduction can be derived from the weighted variance formula:

$$E\left[(R^{t})^{2}\right] = E\left[\sum_{k=1}^{\mu} (p_{s}(k))^{2} (P_{k}^{t})^{2}\right] = aE[(P^{t})^{2}]$$
(8)

Similar observations can be made for the covariance of points sampled from R^t :

$$E\left[R_{i}^{t}R_{j}^{t}|i\neq j\right] = E\left[2\sum_{k=1}^{\mu-1}\sum_{l=k}^{\mu}(p_{s}(k)p_{s}(l)P_{k}^{t}P_{l}^{t})\right] = bE\left[P_{k}^{t}P_{l}^{t}|k\neq l\right]$$
(9)

$$b = 2\sum_{k=1}^{\mu-1} \sum_{l=k}^{\mu} p_s(k) p_s(l)$$
(10)

Values of $E[(P^t)^2]$ and $E[P_k^t P_l^t | k \neq l]$ are the weighted sample variance and covariance values with weights that are equal selection probabilities, hence

$$E[(P^t)^2] = (1-a)v_P^t, \quad E[P_k^t P_l^t | k \neq l] = (1-a)r_P^t$$
(11)

where v_P^t and r_P^t are the (theoretical) variance and covariance of a random variable which generates the population P^t . Then (7) can be rewritten as

$$r_P^{t+1} = a^2 (1-a) v_P^t + b(1-a)^2 r_P^t$$
(12)

In the stable state it holds $r_P^{t+1} = r_P^t = r_P$ which allows to define r_P as

$$r_P = \frac{a^2(1-a)}{1-b(1-a)^2} v_P \tag{13}$$

Equation (5) can be transformed to

$$V[\bar{P^t}] = \frac{1}{\mu'} v_P, \qquad \mu' = \frac{1 - b(1 - a)^2}{1 - b(1 - a)^2 + a^2(1 - a)(\mu - 1)}$$
(14)

Value of μ' will be called the *effective population size* since it leads to formula for the variance of the population mean as if the chromosomes were independent. Then the empirical variance is chi-square distributed with the μ' degrees of freedom and its variance equals approximately

$$V[s^2(P^t)] \approx \frac{2(V[X])^2}{\mu' - 1}$$
 (15)

Example. Above considerations are validated with the following experiment for the fitness function $q(x) = \exp(-x^2)$ and the EA with binary tournament selection, population size $\mu = 100$ and Gaussian mutation with variance $v_m = 1$. The EA was run 100 times, each run took $T = 10^4$ generations. Populations were initialized with standardized normal distribution. For each run, the series of states $u_t = [\bar{P}^t, s^2(P^t)]$, which describe dynamics of populations, was characterized with their mean and empirical variance:

$$\bar{u} = \left[\overline{\bar{P}^t}, \overline{s^2(P^t)}\right], \qquad s^2(u) = \left[s^2(\bar{P}^t), s^2(s^2(P^t))\right] \tag{16}$$

For the binary tournament selection, values of a and b can be computed observing that $p_s(i) = (2\mu - 2i + 1)/(\mu^2)$ — see [1]:

$$a = \sum_{i=1}^{\mu} \left(\frac{2\mu - 2i + 1}{\mu^2}\right)^2 = \frac{4\mu^2 - 1}{3\mu^3}, \qquad b = \frac{3\mu^4 - 8\mu^3 + 11\mu - 6\mu^2}{3\mu^4}$$

which yields the approximate effective population size $\mu' = \frac{3}{4}\mu$.

Fig. 2 contains box-and-whisker plots of mean and variance of population empirical moments. Expected values of moments that have been predicted in previous section are given in corresponding plots. Mean relative error of prediction variance of moments equals 0.97% for $V[(\bar{P^t})]$ and 5.58% for $V[s^2(P^t)]$. If independence of chromosomes in populations were assumed then the corresponding error values would equal 24% for $V[(\bar{P^t})]$ and 21% for $V[s^2(P^t)]$.



Fig. 2. Box and whisker plots of mean and variance values of population mean and variance

4 Quasi-Stability of Finite Populations

For a realistic EA it is impossible to expect its stability in a strict sense since each state is affected by mutation. For this reason it is impossible to find a combination of population mean and variance values which will not be changed in the next generation. The quasi-stability discussed here is defined as follows.

Consider a set of EA states $U(u^*, p)$, which is characterized by a state u^* , and a p-value p. The set $U(u^*, p)$ is quasi-stable when the condition is satisfied

$$\forall u_t \in U(u^*, p) \quad E[u_{t+1}] \in U(u^*, p) \tag{17}$$

where $E[u_{t+1}] = [E[\bar{P}^{t+1}], E[s^2(P^{t+1})]]$ is the vector of expected values of the next state. In other words, if the EA population is characterized by the state $u_t \in U(u^*, p)$ then the next state should not tend to leave the set $U(u^*, p)$.

Bounds on Quasi-Stable Population State. Equations (4), (14) and (15) can be used to define the set $U(u^*, p)$ by defining ranges of values of population mean and variance where their actual values can be found with certain probability. If the population distribution is normal then values of \bar{P}^t are normally distributed and $s^2(P^t)$ is chi-square distributed with $\mu' - 1$ degrees of freedom. For each population state $[\bar{P}^t, s^2(P^t)] \in U(u^*, p)$ it holds

$$m^* - \alpha(v^*, \mu', p) \le \bar{P^t} \le m^* + \alpha(v^*, \mu', p)$$
 (18)

$$\beta(v^*, \mu', p) \le s^2(P^t) \le \gamma(v^*, \mu', p)$$
(19)

where p is the probability of observing values of population mean or variance outside the set $U(u^*, p)$. Values of α, β, γ are defined as upper and lower quantiles of normal and chi-square distributions for the probability p/2:

$$\alpha(v,\mu,p) = \sqrt{\frac{v}{\mu}} \cdot Q_n \left(1 - \frac{1}{2}p\right) \tag{20}$$

$$\beta(v,\mu,p) = \frac{v \cdot Q_c \left(\frac{1}{2}p,\mu-1\right)}{\mu-1}$$
(21)

$$\gamma(v,\mu,p) = \frac{v \cdot Q_c \left(1 - \frac{1}{2}p,\mu - 1\right)}{\mu - 1}$$
(22)

where $Q_n, Q_c(\cdot, k)$ represent the quantile generation functions for the normal and chi-square distribution with k degrees of freedom, respectively.

Testing Quasi-Stability for Finite Populations. For finite populations, if the population state $u_t = [\bar{P}^t, s^2(P^t)]$ is known then it is possible to predict expected values of the next population state, $E[u_{t+1}]$, without knowing the exact contents of P^t . The prediction is based on an assumption that P^{t+1} contains points which are generated randomly with expectation \bar{P}^t and variance $s^2(P^t)$:

$$E[\bar{P}^{t+1}] = \bar{P}^t + \sqrt{s^2(P^t)} \cdot \phi(u_t) \tag{23}$$

$$E[s^{2}(P^{t+1})] = s^{2}(P^{t}) \cdot \theta(u_{t}) + v_{m}$$
(24)

Symbol $\phi(u)$ is called the Response to Selection in Mean (RSM) and indicates the expected change of the population mean in effect of selection

$$\phi(u) = \frac{1}{\sqrt{v}} \int_{-\infty}^{\infty} (x - m)c(x)dx$$
(25)

where c(x) stands for the p.d.f. of chromosomes reproduced from a population which is normally distributed with expectation m, variance v and density $g_{m,v}(x)$. Symbol $\theta(u)$ denotes the Response to Selection in Variance (RSV) which is the degree of the population variance reduction after selection

$$\theta(u) = \frac{1}{v} \int_{-\infty}^{\infty} (x - m)^2 c(x) dx$$
(26)

Note that RSM and RSV depend on selection since c(x) depends on it.

Response to selection, which has been originally defined as the expected change in fitness of individuals before and after selection, is a concept which has been introduced to the evolutionary computation field by Muehlenbein [7] who adopted it from the breeding science in order to define the Breeder Genetic Algorithm. Here, instead of analyzing fitness of individuals, we concentrate on changes of mean and variance of chromosomes processed by the EA to analyze selection effects on the level of genotypes rather than fitness.

Equations (14), (18) and (19) allow for formulating the test of quasi-stability of a set of states. An area $U(u^*, p)$ is quasi-stable with mean m^* and variance v^* when for each population P characterized by the state $u = (\bar{P}, s^2(P)) \in U(u^*, p)$ it holds

$$-\alpha(v^*, \mu', p) \le \bar{P} - m^* + \phi(u) \cdot \sqrt{v} \le \alpha(v^*, \mu', p)$$
(27)

$$\beta(v^*, \mu', p) \le \theta(u) \cdot s^2(P) + v_m \le \gamma(v^*, \mu', p)) \tag{28}$$

Example. Consider a fitness function which is defined as a composition of two Gaussian hills¹— see Fig. 3.

$$q(x) = 5\exp(-x^2/2) + 4\exp(-(x-9)^2/8)$$
(29)



Fig. 3. Plot of the Galar function

Consider binary tournament selection. From [1] it follows that c(x) is given by

$$c(x) = \int_{q(y) < q(x)} g_{m,v}(y) dy \cdot g_{m,v}(x)$$
(30)

which allows to define α, β and γ .

Typical dynamics of the population mean and variance of single run of an EA with $\mu = 100$ chromosomes which optimizes the Galar functions is illustrated in Fig. 4 for few characteristic values of the mutation variance v_m . In all cases the population was initialized with clones of the point 0 and the EA was run for 10^4 generations. In each plot a point represents state u_t of a population P^t and a rectangle is the set $U(u_P^*, p)$ which is a candidate for the quasi-stability area whose limits have been computed putting the stable variance prediction v_P

¹ This function was introduced by R.Galar and discussed e.g. in [8].

yielded by the infinite population model [1] into formulas (27), (28). Value m_P^* has been determined by solving the equation $\phi(u) = 0$ with respect to m, assuming that $v = v_P$.

When $v_m = 0.01$, states of populations stay, except in a few observations, in the quasi-stability rectangle that relates to the global maximum $m_P \approx 0$. For $v_m = 1.18$ the population state mean makes incidental "excursions" from this quasi-stability area towards states characterized with larger mean and variance, but then it returns.

When $v_m = 1.3$, the population state remains for a number of generations in the quasi-stability area around $m_P \approx 0$. Then the population state shifts towards the quasi-stability area corresponding to the second local maximum of the fitness at $m_P \approx 9$. There it remains stable for the rest of the simulation. Note that in this case, populations changed their position despite of the fact that the first quasi-stability area corresponds with the global maximum of the fitness function. This is an illustration of the "survival of the flattest" effect [9] which consists in preference to chromosomes whose fitness values are little sensitive to changes of their position in the chromosomes space.

For $v_m = 10$ a single quasi-stable area exists and then $m_P \approx 7$. Note that this value differs significantly from position of any local maximum. Note that although the assumption about unimodality of the fitness function is not satisfied, the population variance values, which have been predicted in [1], yield correct stability margins and properly explain quasi-stability of populations.



Fig. 4. Plots of the states of populations which were started at 0 and were run for 10^4 generations for the following mutation variance values: a) $v_m = 0.01$, b) $v_m = 1.18$, c) $v_m = 1.3$, d) $v_m = 10$; rectangles represent quasi-stability areas for p-value $p = 10^{-5}$

5 Closing Remarks

In many engineering applications stability of a system is usually a desired feature. For the EA, quasi-stability is a mixed blessing. On one hand, quasi-stability allows for better exploitation of area $U(u_P, p)$ which may contain some local optimum nearby its middlepoint. On the other hand, if populations are quasi-stable for a very low p-value in an area that relates to an optimum of the fitness function, then it is hardly possible for the EA to switch to some other quasi-stability area that relates to another local optimum. Such quasi-stability is undesired since the resulting EA will be a poor global optimizer. This suggests that a method that tracks populations to detect quasi-stability and to break it may be considered as a yet another adaptation method.

It is interesting how the presented results will generalize in real space with many dimensions. Intuitively, response to selection functions are equivalents of derivatives and they may be generalized to a for of a "gradient" by computing the RSM and RSV values for each dimension separately. Then the quasi-stability margins would be defined for each dimension separately, provided that a proper infinite population model to predict the stable variance is developed. These directions of research define the scope of future work.

References

- Arabas, J.: Approximating the genetic diversity of populations in the quasiequilibrium state. IEEE Transactions on Evolutionary Computation 16(5), 632–644 (2012)
- Beyer, H.G., Deb, K.: On self-adaptive features in real-parameter evolutionary algorithms. IEEE Transactions on Evolutionary Computation 5(3), 250–270 (2001)
- 3. Arnold, D.V., Beyer, H.G.: On the benefits of populations for noisy optimization. Evolutionary Computation 11(2), 111–127 (2003)
- 4. Arnold, D.V.: Noisy Optimization with Evolution Strategies. Kluwer Academic Publishers (2002)
- Qi, X., Palmieri, F.: Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space part I: Basic properties of selection and mutation. IEEE Transactions on Neural Networks 5(1), 102–119 (1994)
- Karcz-Dulęba, I.: Dynamics of infinite populations evolving in a landscape of uni and bimodal fitness functions. IEEE Transactions on Evolutionary Computation 5(4), 398–409 (2001)
- Muehlenbein, H., Schlierkamp-Voosen, D.: Predictive models for the Breeder Genetic Algorithm – I. continuous parameter optimization. Evolutionary Computation 1, 25–49 (1993)
- Chorazyczewski, A., Galar, R.: Visualization of evolutionary adaptation in Rⁿ. In: Porto, V.W., Waagen, D. (eds.) EP 1998. LNCS, vol. 1447, pp. 659–668. Springer, Heidelberg (1998)
- Wilke, C., et al.: Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature 412(6844), 331–333 (2001)