

Population diversity in punctuated equilibrium and its relation to saddle crossing

Jarosław Arabas Institute of Electronic Systems
Warsaw University of Technology, Warsaw, Poland
email: jarabas@elka.pw.edu.pl

Abstract—

I. INTRODUCTION

In this contribution we discuss the ability of a generational Evolutionary Algorithm (EA) in R^n to perform task of finding attraction basins of local maxima of the fitness function, in a hope that one of these maxima is the global one. It has been reported by many authors, e.g. [4], that in consecutive EA generations, individuals form relatively compact clusters, whose middlepoints drift in the search space. Moreover, dynamics of the cluster midpoint contains relatively long periods of small changes which chaotically oscillate around a certain “stable” position (*stasis* period), and relatively short periods of rapid changes of the population midpoint, whose direction stays similar at least for few generations. Period of change is identified with the *saddle crossing* phenomenon [2], when the population changes its position by going from the attraction basin on one local maximum to the other, and on the way, poorly fit points around the border of these two attraction basins are generated and reproduced.

We approach to characterize the saddle crossing ability by the analysis of the pattern in which points are generated in the stasis period. EA generates individuals in a random way, so we characterize the probability distribution of generating an offspring individual when the population contents is known; we call this probability distribution a *sampling distribution*. Analytical formulas for the sampling distribution have been given by Qi and Palmieri [9], who considered proportional selection, uniform crossover and Gaussian mutation, and were extended by Nomura [8] for different version of linear crossover. In the aforementioned results, the authors made the assumption about the infinite size of the population. We do not make such assumption, since we are not much concerned about the distribution of individuals in a single population but rather about the set of points that are generated in multiple generations of the EA in the period of stasis. We provide approximate formulas for the sampling distribution in the stasis period assuming that the fitness function is correlated Gaussian in R^n . The approximation follows from the assumption, that the influence of a single individual’s fitness on the total fitness of the population is negligible. We consider the case when no crossover is applied, and when the crossover is uniform or arithmetic. Thus we enhance the results by Karcz-Duleba [3] who considered infinite population size EA model in R , without crossover. We also investigate to what degree the

actual selection probability of an individual will be different from the approximate selection probability we have used in the formulas. We show that for small populations, current covariance matrix of the sampling distribution may significantly differ from the approximated one, and the difference leads to wider distribution of points generated by the EA.

For the considered Gaussian fitness function, the sampling distribution is usually Gaussian (except for one type of arithmetic crossover), and we characterize this function not only with the covariance matrix, but also with the value of *spread*, which is defined as the product of eigenvalues of the covariance matrix. The intuition of the spread is the area of the set which will be densely covered with points when the EA is in the stasis period. Thus, spread is related to the population diversity [5]. The difference between popular diversity definitions and spread lies in the fact, that the diversity measures are based on distances between individuals in the population, whereas spread characterizes the probability distribution of the random variable, whose μ -fold realization is the population contents.

Finally we consider a fitness function which is sum of two Gaussian functions, one of them defining a local maximum, and another one — a global one. We attempt to characterize ability of saddle crossing assuming the “stable” sampling distribution corresponding to one of the Gaussian terms, and analyzing the chance to generate a point from the attraction basin of the maximum defined by the other Gaussian term. We perform experimental analysis of the saddle crossing process and we show the influence of the population size on the EA effectiveness in finding global maximum.

A. Evolutionary Algorithm

We consider a generational Evolutionary Algorithm (EA) depicted in Fig.1. Chromosomes are vectors from R^n , $n < \infty$, and $q : R^n \rightarrow R$ is the fitness function to be maximized without any additional constraints. In each generation, individuals are reproduced, crossed over and mutated, and the new population of mutants becomes the base population for the next generation. We assume that crossover takes two parents and produces a single offspring.

Symbol μ stands for the size of the population \mathbf{P}^t . We denote the i -th element of the population \mathbf{P}^t and \mathbf{O}^t by P_i^t and O_i^t , respectively.

In this contribution we consider only a fitness-proportional

initialize(\mathbf{P}^0)

while not stop condition satisfied

for each $i \in 1, \dots, \mu$

generate random value $c \sim U(0, 1)$

if $c < p_c$

$j :=$ select point number from $(1, \dots, \mu)$

$k :=$ select point number from $(1, \dots, \mu)$

$\mathbf{z} :=$ crossover (P_j^t, P_k^t)

$O_i^t :=$ mutate (\mathbf{z})

else

$j :=$ select point number from $(1, \dots, \mu)$

$O_i^t :=$ mutate (P_j^t)

$\mathbf{P}^{t+1} := \mathbf{O}^t$

Fig. 1. Outline of the Evolutionary Algorithm under consideration

selection where selection probabilities are defined as

$$P_s(i, t) = \frac{q(P_i^t)}{\sum_{j=1, \dots, \mu} q(P_j^t)} \quad (1)$$

Mutation consists in adding to the mutated chromosome \mathbf{x} a random correction vector \mathbf{d} , which results in a new point \mathbf{y}

$$\mathbf{y} = \mathbf{x} + \mathbf{d} \quad (2)$$

We assume that the vector \mathbf{d} is an n -dimensional normal random variable, so the probability density function (p.d.f.) of the mutation vector is

$$f_m(\mathbf{d}) = G_{\mathbf{0}, C_m}(\mathbf{d}) \quad (3)$$

where

$$G_{\mathbf{m}, C}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |C|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T C^{-1}(\mathbf{x} - \mathbf{m})\right) \quad (4)$$

stands for the Gaussian function — p.d.f. of the normal distribution with the expectation vector \mathbf{m} and the covariance matrix C .

Note that if the mutation distribution is given by (3), it is always possible to normalize the space of chromosomes. Observe that when we decompose the covariance matrix C_m into the diagonal matrix of eigenvalues L_m and the matrix of eigenvectors E_m such that

$$L_m = E_m^{-1} C_m E_m \quad (5)$$

then in the space of vectors

$$\mathbf{y} = E_m \mathbf{x} \quad (6)$$

mutation is Gaussian with zero mean and covariance I (where I is the unit matrix).

B. Trace of the generated points

After an EA is started with some initial population \mathbf{P}^0 , it generates points according to rules given in Fig. 1. When we plot all points that have been generated by the single run of the EA, we will observe a characteristic pattern — an EA trace. Let us draw the EA trace for an example fitness function. Trace of points generated in the single EA run is depicted in Fig. 2. Observe that points tend to form a number of

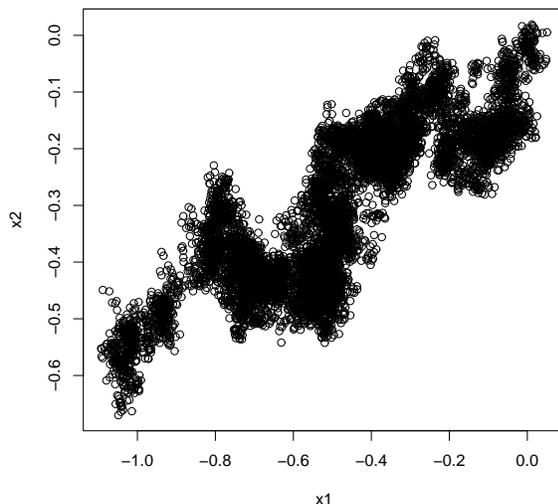


Fig. 2. Trace of the EA - set of points generated in a single EA run

relatively compact clusters. This characteristic pattern is in fact observed regardless of the fitness function (and appears even for constant functions), but its shape details, e.g., compactness of clusters, is fitness function dependent.

We can observe that clusters of points contained in populations in consecutive generations tend to remain in stable positions for long periods of time. Incidentally, a long distance shift of the population cluster may also be observed. Thus we can explain that the clusters in the EA trace are the effect of slow shifts and incidental jumps of a cluster of points which represents population contents. In the following text we attempt to investigate phenomena which lead to cluster formation and we also try to explain the way in which the cluster moves in the search space.

II. SAMPLING DISTRIBUTION

EA generates points in a randomized way, and we can describe the EA action by defining the p.d.f. of the sampling distribution which defines possible position of each point in the population \mathbf{O}^t ; denote the p.d.f. of this distribution by $f_s(t)$. In other words, population \mathbf{O}^t is a μ -fold realization of a random variable whose p.d.f. is $f_s(t)$. In this section we attempt to define $f_s(t)$ for certain popular EA types.

A. No crossover

If there is no crossover, each point in \mathbf{O}^t results from mutating a point selected from \mathbf{P}^t .

$$f_s(\mathbf{x}, t) = \int_{R^n} E \{ P_s(\mathbf{z}, t) f_m(\mathbf{x} - \mathbf{z}) \} d\mathbf{z} \quad (7)$$

where $P_s(\mathbf{z}, t)$ is the probability that a point \mathbf{z} will be selected to reproduce when the population contents is \mathbf{P}^t , f_m is the p.d.f. of the mutation distribution, and the symbol $E\{\cdot\}$ means the expected value over all possible contents of the population

\mathbf{P}^t . Observe that each points from \mathbf{P}^t is a realization of a random variable whose p.d.f. is $f_s(t-1)$.

If the population size increases, and the average fitness value of points from the population \mathbf{P}^t is quite high, the selection probability becomes well approximated by

$$\tilde{P}_s(\mathbf{z}, t) = \frac{q(\mathbf{z})}{\mu E_{f_s(t-1)}\{q(\mathbf{x})\}} \quad (8)$$

where $E_{f_s(t-1)}\{q(\mathbf{x})\}$ is the expected value of the fitness of points generated from the sampling distribution whose p.d.f. is $f_s(t-1)$.

Putting \tilde{P}_s (8) instead of P_s in (7) we get an approximate formula for f_s

$$f_s(\mathbf{x}, t) \approx \int_{R^n} \frac{q(\mathbf{z})f_s(\mathbf{z}, t-1)}{\sum_{i=1, \dots, \mu} q(P_i^t)} f_m(\mathbf{x} - \mathbf{z}) d\mathbf{z} \quad (9)$$

and

$$f_s(\mathbf{x}, t) \approx A^t \int_{R^n} f_q(\mathbf{z}, t) f_m(\mathbf{x} - \mathbf{z}) d\mathbf{z} \quad (10)$$

where

$$f_q(\mathbf{x}, t) = q(\mathbf{x})f_s(\mathbf{x}, t-1) \quad (11)$$

and

$$A^t = \frac{1}{\mu E_{f_s(t-1)}\{q(\mathbf{x})\}} \quad (12)$$

Thus we obtain a recursive formula that allows to predict future possible sampling distributions.

a) Uncorrelated Gaussian fitness: Let us assume that the fitness function is uncorrelated Gaussian; with \mathbf{M} and \mathbf{V} denote its mean vector and variance vector, respectively.

Note that the product of two uncorrelated Gaussian functions is also an uncorrelated Gaussian function; the same holds for the convolution. Thus $f_s(t)$ is uncorrelated Gaussian. With \mathbf{m}_t and V_t denote the mean vector and the variance vector of $f_s(t)$. Note that, since we deal with uncorrelated Gaussian functions, each dimension can be considered separately. In the i -th dimension, application of the formula (10) leads to

$$f_s(x_i, t) = A^t [G_{m_{t-1,i}, v_{t-1,i}}(x_i) G_{M_i, V_i}(x_i)] * G_{0,1}(x_i) \quad (13)$$

where ‘*’ is the symbol of convolution.

In the subsequent transformations we omit indexes t and i for better readability.

$$G_{m,v}(x)G_{M,V}(x) = \quad (14)$$

$$\frac{1}{(2\pi)\sqrt{vV}} \exp\left(-\frac{(d-m)^2}{2v} - \frac{(d-M)^2}{2V}\right) = \quad (15)$$

$$\exp\left(-\frac{d^2 - 2dm + m^2}{2Vv} + \frac{(d^2 - 2dM + M^2)v}{2Vv}\right) = \quad (16)$$

$$\exp\left(-\frac{d^2(V+v) - 2d(Vm + Mv) + Vm^2 + M^2v}{2Vv}\right) = \quad (17)$$

$$G_{\left(\frac{Vm+Mv}{V+v}, \frac{Vv}{V+v}\right)}(x) \exp\left(-\frac{M-m}{2(V+v)}\right) \quad (18)$$

Thus we can see that the first term of the convolution is a Gaussian function with the mean equal to $\frac{V_i m_{t-1,i} + M_i v_{t-1,i}}{V_i + v_{t-1,i}}$ and the variance equal to $\frac{V_i v_{t-1,i}}{V_i + v_{t-1,i}}$. After convolution we get

$$f_s(x_i, t) = B^t G_{m_{t,i}, v_{t,i}}(x_i) \quad (19)$$

where

$$B^t = A^t \exp\left(-\frac{M_i - m_{t-1,i}}{2(V_i + v_{t-1,i})}\right) \quad (20)$$

and

$$m_{t,i} = \frac{V_i m_{t-1,i} + M_i v_{t-1,i}}{V_i + v_{t-1,i}} \quad (21)$$

$$v_{t,i} = 1 + \frac{V_i v_{t-1,i}}{V_i + v_{t-1,i}} \quad (22)$$

In Fig. 3 we plot example curves for $m_{t,i}$ and $v_{t,i}$ for $M_i = 0$ and $V_i = 5$. It can be readily seen that these curves stabilize quite fast in values $m_{t+1,i} = m_{t,i} = m_i$, $v_{t+1,i} = v_{t,i} = v_i$. So the stable mean and variance values of the sampling distribution are

$$m_{s,i} = M_i \quad (23)$$

$$v_{s,i} = \frac{1 + \sqrt{1 + 4V_i}}{2} \quad (24)$$

b) Correlated Gaussian fitness: If the fitness function is correlated Gaussian with the mean vector \mathbf{M} and the covariance matrix C_q then the stable sampling distribution will be also correlated Gaussian with mean values equal to \mathbf{M} . The correlation matrix C_s of this stable sampling distribution can be computed after performing the decomposition of the matrix C_q into eigenvectors and eigenvalues:

$$L_q = E_q^{-1} C_q E_q \quad (25)$$

where L_q is the diagonal matrix of eigenvalues, and E_q is the matrix of eigenvectors — a new base of the search space where the fitness function becomes uncorrelated Gaussian. In that space, matrix L_q is the covariance matrix. We can then use the formula (24) to obtain the variance matrix L'_s of the stable sampling distribution:

$$L'_{s,ij} = \begin{cases} (1 + \sqrt{1 + 4L_{q,ii}})/2 & \text{when } i = j \\ 0 & \text{otherwise} \end{cases} \quad (26)$$

So, the covariance matrix C_s of the stable sampling distribution can be obtained by the retransformation of the matrix L'_s into the original coordinate system:

$$C_s = E_q L'_s \quad (27)$$

B. Crossover

When crossover is used, two points which have been selected to reproduce undergo recombination and yield a point which is further modified by the mutation. If $p_c = 1$ then the p.d.f. of the sampling distribution is defined as

$$f_s(\mathbf{x}, t) = \int_{R^n} f_x(\mathbf{z}, t) f_m(\mathbf{x} - \mathbf{z}) d\mathbf{z} \quad (28)$$

where f_x is the p.d.f. of generating a point as a result of crossover. Different crossover methods are characterized by different functions f_x .

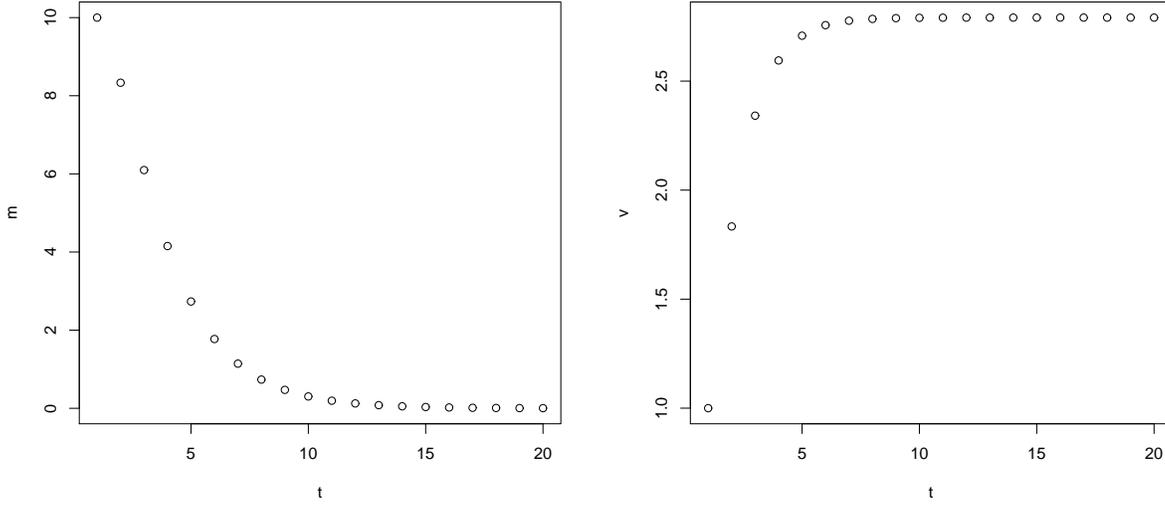


Fig. 3. Evolution of mean and variance of the sampling distribution according to equations (23) and (24)

C. Arithmetic crossover

We assume that the arithmetic crossover consists in generating point $\mathbf{z} = (\mathbf{x} + \mathbf{y})/2$, where \mathbf{x} and \mathbf{y} are parental chromosomes. Thus the p.d.f. of generating a point as a result of crossover is

$$f_x(\mathbf{x}, t) = \int_{R^n} E \{P_s(\mathbf{z}/2, t)P_s(\mathbf{x} - \mathbf{z}/2, t)\} d\mathbf{z} \quad (29)$$

c) *Uncorrelated Gaussian fitness*: Assume first that fitness function is uncorrelated Gaussian, and its mean and variance vector are \mathbf{M} and \mathbf{V} . Since the mutation p.d.f. is uncorrelated Gaussian either, we expect that $f_s(t)$ is Gaussian again. The formula for $f_x(t)$ is

$$f_x(\mathbf{x}, t) = \int_{R^n} E \{f_q(\mathbf{z}/2, t)f_q(\mathbf{x} - \mathbf{z}/2, t)\} d\mathbf{z} \quad (30)$$

Assuming large enough population, (30) can be rewritten as:

$$f_x(\mathbf{x}, t) \approx A^t \int_{R^n} f_q(\mathbf{z}/2, t)f_q(\mathbf{x} - \mathbf{z}/2, t)d\mathbf{z} \quad (31)$$

where A^t is a constant which makes (31) a proper definition of the p.d.f. Thus $f_x(t)$ is uncorrelated Gaussian with mean vector \mathbf{m}_x and variance vector \mathbf{v}_x given by

$$m_{x,i} = \frac{V_i m_{t-1,i} + M_i v_{t-1,i}}{V_i + v_{t-1,i}} \quad (32)$$

$$v_{x,i} = V_i v_{t-1,i} / (2(V_i + v_{t-1,i})) \quad (33)$$

Putting (32) and (33) into (28) we obtain

$$m_{t,i} = \frac{V_i m_{t-1,i} + M_i v_{t-1,i}}{V_i + v_{t-1,i}} \quad (34)$$

$$v_{t,i} = 1 + \frac{V_i v_{t-1,i}}{2(V_i + v_{t-1,i})} \quad (35)$$

With t increasing, mean and variance of the sampling distribution tend to stabilize at level

$$m_{s,i} = M_i \quad (36)$$

$$v_{s,i} = \frac{\sqrt{(V_i - 2)^2 + 16V_i} - (V_i - 2)}{4} \quad (37)$$

When the fitness function is correlated Gaussian, correlation matrix of the stable sampling distribution can be obtained using the eigenvector based method discussed for the no crossover case (in paragraph III.A.c).

D. Generalized arithmetic crossover

A more generalized case of the arithmetic crossover consists in generating point $\mathbf{z} = a\mathbf{x} + (1 - a)\mathbf{y}$, where \mathbf{x} and \mathbf{y} are parental chromosomes, and a is a the averaging factor whose value is a random variable with a certain distribution. Typically, a is drawn from a uniform distribution from the range (0,1), or, in a more general case, from a range $(-b, 1+b)$, where b is a user-defined parameter.

Assuming large enough population we obtain approximate formula for f_x :

$$f_x(\mathbf{x}, t) \approx A^t \int_{R^n} \int_{-b}^{1+b} f_q(a\mathbf{z}, t)f_q(\mathbf{x} - (1 - a)\mathbf{z}, 1)dad\mathbf{z} \quad (38)$$

which can be combined with (28) to obtain the p.d.f. of the sampling distribution.

If f_t and q are uncorrelated Gaussian functions then

$$f_x(\mathbf{x}, t) \approx A^t \int_{-b}^{1+b} G_{\mathbf{m}_a, v_a}(\mathbf{x})da \quad (39)$$

where A^t is a constant which allows to use (39) as a proper

definition of the p.d.f., and

$$m_{a,i} = \frac{V_i m_{t-1,i} + M_i v_{t-1,i}}{V_i + v_{t-1,i}} \quad (40)$$

$$v_{a,i} = (2(a - 1/2)^2 + 1/2) \frac{V_i v_{t-1,i}}{V_i + v_{t-1,i}} \quad (41)$$

Observe that the value of $m_{a,i}$ is independent of a , so the integral (39) defines a mixture of an infinite number of Gaussian functions whose expected value is identical, and the variance is a function of a . It seems that (39) has no compact analytical form, and in particular, it is not a Gaussian function any more. Still we can operate on moments of the distribution whose p.d.f. is given by (39). Observe that if a mixture of Gaussian functions with identical expectation is considered, the expected value remains identical to the mixture elements, and the variance is the weighted average of variances. Therefore we get

$$m_{x,i} = \frac{V_i m_{t-1,i} + M_i v_{t-1,i}}{V_i + v_{t-1,i}} \quad (42)$$

$$v_{x,i} = \int_{-b}^{1+b} (2(a - 1/2)^2 + 1/2) \frac{V_i v_{t-1,i}}{V_i + v_{t-1,i}} da = \quad (43)$$

$$2(b^3 + (b + 1)^3) \frac{V_i v_{t-1,i}}{3(V_i + v_{t-1,i})} \quad (44)$$

where \mathbf{m}_{t-1} and \mathbf{v}_{t-1} are vectors of mean and variance of the sampling distribution when the population contents is \mathbf{P}^{t-1} .

Combining (42) and (43) with (28) we obtain

$$m_{t,i} = m_{x,i} \quad (45)$$

$$v_{t,i} = 1 + \frac{2(b^3 + (b + 1)^3) V_i v_{t-1,i}}{3(V_i + v_{t-1,i})} \quad (46)$$

With t increasing, mean and variance of the sampling distribution tend to stabilize at level

$$m_{s,i} = M_i \quad (47)$$

$$v_{s,i} = \frac{h + \sqrt{h^2 + 4V_i}}{2} \quad (48)$$

$$h = \left(\frac{2}{3} (b^3 + (1 + b)^3) - 1 \right) V_i + 1 \quad (49)$$

In a typical case, when $b = 0$, formula for $v_{s,i}$ is simplified to

$$v_{s,i} = \frac{\sqrt{(V_i - 3)^2 + 36V_i} - (V_i - 3)}{6} \quad (50)$$

When the fitness function is correlated Gaussian, correlation matrix of the stable sampling distribution can be obtained using the eigenvector based method introduced for the no crossover case.

To better understand the role of the arithmetic crossover, let us study Fig. 4 in which stable levels of the sampling distribution variance $v_{s,i}$, obtained for an uncorrelated Gaussian fitness function, are plotted for various values of the fitness function variance V_i . For the sake of comparison, stable variance for the EA with no crossover is also provided.

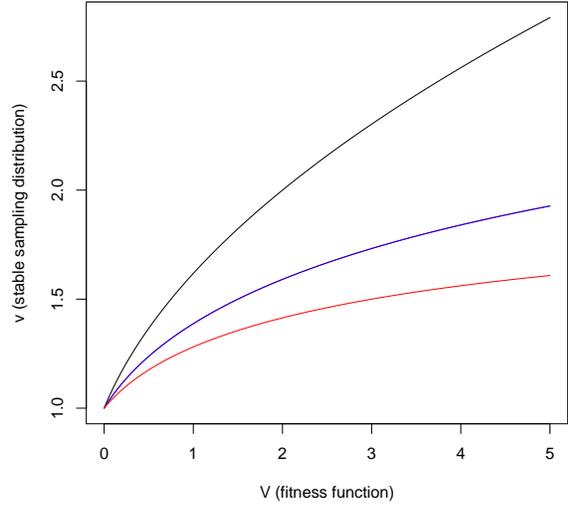


Fig. 4. Variance of the stable distribution vs. variance of the fitness function for the EA with Gaussian mutation and no crossover (black line), arithmetic crossover, $p_c = 1$ (red line), and generalized arithmetic crossover, $p_c = 1, b = 0$ (blue line)

E. Uniform crossover

Another popular method to perform crossover is to exchange genes between parents. Let us analyze results of such crossover method on the distribution f_x . Note that a point $\mathbf{z} = (z_1, z_2, \dots, z_n)$ resulting from the crossover of parents $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)$ is a combination of n genes that are taken from parents, and for each i , either $z_i = x_i$ (with probability 0.5) or $z_i = y_i$. If both points under crossover are generated from the same random variable whose p.d.f. is $f_s(t - 1)$, then the p.d.f. of the joint distribution of points generated by the crossover can be obtained by multiplying marginal distributions of the product $f_q(\mathbf{x}, t)$. With $f|_{\mathbf{l}(k)}$ denote the p.d.f. of the $(n - k)$ -dimensional marginal distribution, when the p.d.f. of the joint distribution is f . The marginal distribution is characterized by the vector $\mathbf{l}(k) \in V(n, k)$ where $V(n, k)$ is the set of k element variations without repetitions over the set $\{1, \dots, n\}$. Thus

$$f|_{\mathbf{l}(k)}(\mathbf{x}) = \int_{\mathbb{R}^k} f(\mathbf{x}) dx_{l(k)_1} \cdots dx_{l(k)_k} \quad (51)$$

In addition, define the complement vector $c(\mathbf{l}(k)) = \{1, \dots, n\} \setminus \mathbf{l}(k)$ as a vector of indexes that are missing in the vector $\mathbf{l}(k)$.

We can define f_x in the following way

$$f_x(\mathbf{x}, t) \approx A^t ((1 - p)^n + p^n) f_q(\mathbf{x}, t) + \quad (52)$$

$$\sum_{k=1}^{\lfloor n/2 \rfloor} \sum_{\mathbf{l}(k) \in V(n, k)} (p^k (1 - p)^{n-k} + p^{n-k} (1 - p)^k) \quad (53)$$

$$f_q|_{\mathbf{l}(k)}(\mathbf{x}, t) f_q|_{c(\mathbf{l}(k))}(\mathbf{x}, t) + \quad (54)$$

$$\frac{1}{2} \text{odd}(n) \sum_{\mathbf{l}(\nu) \in V(n, \nu)} (p^\nu (1 - p)^{n-\nu} + p^{n-\nu} (1 - p)^\nu) \quad (55)$$

$$f_q|_{\mathbf{l}(\nu)}(\mathbf{x}, t) f_q|_{c(\mathbf{l}(\nu))}(\mathbf{x}, t) \quad (56)$$

TABLE I

COVARIANCE MATRIX OF STABLE DISTRIBUTION FOR VARIOUS VERSIONS OF CROSSOVER

crossover type	stable covariance matrix	spread
no crossover	[2.929, 0.592; 0.592, 1.448]	3.891
arithmetic	[1.607, 0.172; 0.172, 1.178]	1.864
generalized arithmetic, $b = 0$	[1.947, 0.277; 0.277, 1.255]	2.368
uniform	[2.964, 0.296; 0.296, 1.533]	4.457

where $\nu = \lceil n/2 \rceil$, and $\text{odd}(n)$ returns 1 if n is odd, and 0 when n is even.

Observe that when the fitness function is uncorrelated Gaussian then $f_q(\mathbf{x}, t)$ is also uncorrelated Gaussian and $f_q(\mathbf{x}, t) = f_q|_{1(k)}(\mathbf{x}, t)f_q|_{c(1(k))}(\mathbf{x}, t)$ for all $k = 1, \dots, n - 1$. Then

$$f_x(\mathbf{x}, t) \approx \int_{R^n} E\{f_q(\mathbf{z}, t)\} d\mathbf{z} \quad (57)$$

and formulas (23) and (24) will apply for the mean and the marginal variance vectors.

If the fitness function is correlated Gaussian, then $f_q(\mathbf{x}, t)$ becomes correlated Gaussian, which may result in a correlated Gaussian $f_s(\mathbf{x}, t)$. Compact formulas for the evolution of covariance matrix are hard to define in a general case, however it is quite straightforward to observe that formula (56) will produce a mixture of Gaussian functions such that, at least for some of them, axes of ellipsoids of the sampling distribution p.d.f. will not be parallel to axes of ellipsoids of identical values of the fitness function. Note also that the mean vector of a stable sampling distribution will be equal to the position of the fitness maximum.

F. Discussion

Consider an example fitness function which is Gaussian with mean vector $\mathbf{M} = \mathbf{0}$ and the covariance matrix $\mathbf{C} = [6, 2; 2, 1]$. In Tab. I, covariance matrices of stable sampling distributions are provided. In Fig. 5, a graphical presentation of the covariance matrices from Tab. I is given. We plot equal value lines of the p.d.f. of the stable sampling distribution characterized by covariance matrices. Inside ellipses encircled with these lines, probability of generating a point is greater than 0.994. To indicate the shape of the fitness function, we plot one of its isolines. All ellipses, except for one corresponding to the uniform crossover, have their axes parallel to axes of the objective function isoline. Area of the ellipses is different and is smallest for the standard arithmetical crossover. To better characterize covariance matrices, define the spread of the covariance matrix as the product of its eigenvalues. Thus, spread is proportional to the area of ellipse which contains 99.4% of points that will be generated according to the p.d.f. of the sampling distribution. Spread values are given in the third column of Tab. I.

So we can say that if an arithmetic crossover scheme is used which consists in weighted averaging of parents, then populations are more compact than in case when no crossover is used. On the other hand, application of the uniform crossover results in partially decorrelated sampling distribution, whose which in the considered example results in the stable distribution whose spread is greater than for the case with no crossover.

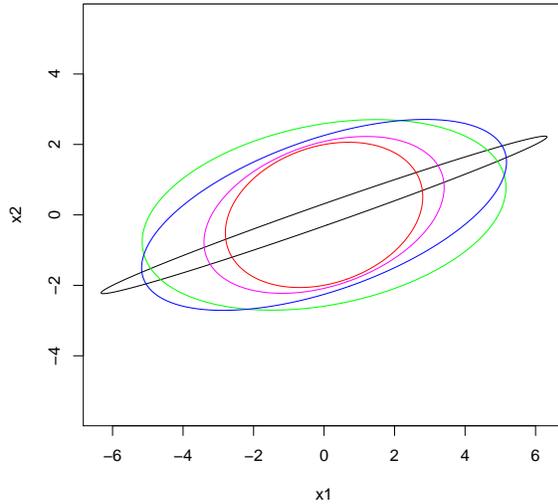


Fig. 5. Lines of equal p.d.f. value which encircle areas where probability of generating a point is greater than 0.994 for various EA versions: no crossover (blue), arithmetic crossover (red), generalized arithmetic crossover (magenta) and uniform crossover (green); in all cases, Gaussian mutation is applied; for the sake of comparison, one of the isolines of the fitness function is plotted

III. WHEN THE POPULATION IS SMALLER THAN INFINITE?

In the preceding section we considered equations that define mean vector and covariance matrix of a stable sampling distribution provided that the fitness function is Gaussian and the population is large enough to neglect the influence of a single individual's fitness on the sum of fitness used in the denominator of the expression (8). In this section we investigate the stable sampling distribution of individuals when the population is small and the aforementioned approximation is bold.

A. Stable distribution of large populations

In statistics, sample size of the size 100 is usually big enough to perform statistical analysis. We assumed that the population size $\mu = 100$ can be regarded “practically infinite” and we tested if the empirical distribution of points generated by the EA was consistent with the distribution predicted with the use of the infinite population size analysis.

We use the Gaussian fitness function in R^2 with zero mean vector and the covariance matrix equal to $[6, 2; 2, 1]$ — just the same function that was used to compare spread of the population under various crossover methods. We initialized base population with the uniform distribution in the box $[-5, 5] \times [-5, 5]$, and we let the evolution last for 500 generations. Then we took chromosomes from generations number 101 up to 500 (which makes 40000 chromosomes in total) and we used all these chromosomes to estimate mean vector and covariance matrix for each version of the EA. The results are presented in Tab. II.

A very good consistency can be observed between the covariance matrix values that have been predicted using the-

TABLE II

COVARIANCE MATRICES OF STABLE DISTRIBUTIONS FOR DIFFERENT VALUES OF CROSSOVER; MATRIX COEFFICIENTS WERE ESTIMATED FROM THE EXPERIMENTAL DATA

crossover type	stable covariance matrix
no crossover	[2.964, 0.605; 0.605, 1.453]
arithmetic	[1.624, 0.182; 0.182, 1.192]
generalized arithmetic, $b = 0$	[1.98, 0.287; 0.287, 1.254]
uniform	[2.798, 0.267; 0.267, 1.440]

oretical considerations and estimated from the experimental data.

B. Stable distribution of small populations

In equation (8) we assume that the influence of the individual's fitness on the sum in the denominator is negligible. In fact, if we assume that all individuals are generated from the random variable whose p.d.f. is $f_s(t-1)$, then the selection probability of an individual \mathbf{z} equals

$$P_s(\mathbf{z}, t) = \frac{q(\mathbf{z})}{q(\mathbf{z}) + (\mu - 1) \text{avg}_{f_s(t-1), \mu-1} \{q(\mathbf{x})\}} \quad (58)$$

where $\text{avg}_{f_s(t-1), \mu-1} \{q(\mathbf{x})\}$ is the average value of $\mu - 1$ points generated with the distribution whose p.d.f. is $f_s(t-1)$. Note that the selection probability of an individual is a random variable in contrast to its approximation, which is a deterministic value. Relation between the actual selection probability and its approximation $\tilde{P}_s(\mathbf{z}, t)$ (cf. (8)) is given by

$$P_s(\mathbf{z}, t) = \frac{\tilde{P}_s(\mathbf{z}, t)}{\frac{q(\mathbf{z})}{\mu E_{f_s(t-1)} \{q(\mathbf{x})\}} + \left(1 - \frac{1}{\mu}\right) \frac{\text{avg}_{f_s(t-1), \mu} \{q(\mathbf{x})\}}{E_{f_s(t-1)} \{q(\mathbf{x})\}}} \quad (59)$$

For small populations, the discrepancy between the actual and approximated selection probability may become quite large due to the term $1 - 1/\mu$ in the denominator of (59). Term $\frac{q(\mathbf{z})}{\mu E_{f_s(t-1)} \{q(\mathbf{x})\}}$ in the denominator of (59) is responsible for reducing the selection probability for best fit individuals, and increasing the selection probability of poorly fit ones. This will effect in reducing the selection pressure, and the degree of reduction is inversely proportional to the population size μ . Thus we conclude that for small populations, spread of the sampling distribution should be greater than for large populations.

We illustrate the aforementioned effects for the Gaussian fitness function with zero mean and covariance matrix [6, 2; 2, 1]. We consider EA without crossover. If we assume that the sampling distribution is a stable one (computed with the use of the approximated formula for the selection probability), we can use the stable p.d.f. in the formula (59). Thus we can observe what will happen with the selection probability and in the same time we can predict whether the approximation of the stable distribution will be close to the current one (in that case, we would see that the proportion $P_s(\mathbf{z}, t)/\tilde{P}_s(\mathbf{z}, t)$ is almost one) or the approximation is too bold to believe it.

We estimated selection probabilities by making simulations of the population contents with the use of the approximation to the "stable" distribution. Fitness of the first individual was changed in the range [0, ..., 0.1125] with a step 0.00125, and for

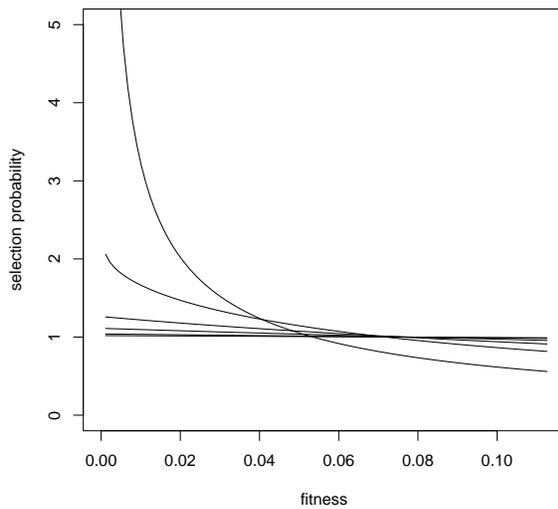


Fig. 6. Proportion of $P_s(\mathbf{x})/\tilde{P}_s(\mathbf{x})$ as a function of the individual's fitness for the population size $\mu \in \{2, 5, 10, 20, 50, 100\}$; curvature of the line grows with the population size decreasing

TABLE III
COVARIANCE MATRIX C OF THE STABLE SAMPLING DISTRIBUTION ESTIMATED FROM EXPERIMENTS FOR EA WITHOUT CROSSOVER, FOR DIFFERENT POPULATION SIZE μ

μ	C
2	[21.02, 7.04; 7.04, 4.63]
5	[4.55, 1.27; 1.27, 1.77]
10	[3.17, 0.67; 0.67, 1.53]
20	[2.96, 0.61; 0.61, 1.43]
50	[2.93, 0.58; 0.58, 1.43]
100	[2.94, 0.59; 0.59, 1.45]

each value from that range, 10000 simulations of the remaining population members were made. The results are plotted in Fig. 6.

To illustrate the effect of reducing the selection pressure and increasing sampling distribution spread for small populations, we ran EA with no crossover for several different values of the population size μ . Each EA run generated 500 populations, and we computed covariance matrix using last 400 generations. Obtained results are presented in Tab. III.

So we conclude that for the fitness function under testing, even for rather moderate population size $\mu = 20$, covariance matrix of the sampling distribution is well approximated by the covariance matrix obtained with the use of infinite population size model. In addition we observe that smaller population size indeed results in wider spread of the sampling distribution, so, paradoxically, increase of the explorative power of the EA can be obtained by reducing the population size down to even less than 10 individuals.

IV. ABILITY TO FIND GLOBAL MAXIMUM

If the fitness functions has many local maxima and the goal is to search for the global maximum then the spread of the sampling distribution plays the key role in attaining the global maximization task. Points from the current population usually

form a cluster whose position in the search space tends to stabilize for a number of generations, i.e., the mean value of the point position oscillates around a certain “stable” value. This tendency appears when the fitness function is locally concave, and stabilization may take long when the population is little diversified so that probability of generating points inside this locally concave area is high.

If the fitness function is symmetrical about the local maximum in its neighborhood, then the “stable” value of the population mean is simply the position of the local maximum.

A way in which the population may change its position is to go from the attraction basin of the current local maximum to the attraction basin of another local maximum in the neighborhood of the current one, so that the “stable” value of the mean position of points changes its position to the vicinity of another local maximum. This phenomenon is known as “saddle crossing”, although it is not always the case that the transition of the population goes through the saddle point.

Saddle crossing is more possible when the probability is increased of generating a point outside the attraction basin of the current local maximum. Moreover, if the aforementioned probability is too low, saddle crossing will not be observed in practice and the neighborhood of a current local maximum will trap the population for good.

When a realistic fitness function is maximized, a good strategy is to keep the saddle crossing ability on a reasonably high level which in turn implies keeping the sampling distribution with wide enough spread. However if this spread were too wide, it would result in poor ability to find a local maximum. So, sampling distribution spread should be a good compromise between the ability to stabilize for a number of generations (exploitation) and the ability to cross saddles (exploration). The problem with attaining this compromise is its asymmetry — if the spread is too high, EA behaves like a blind search but is still able to find an approximation to the global maximum, but when the spread is too small, EA loses its global optimization properties and reduces to a rather ineffective stochastic local optimizer.

A. Sampling distribution spread and the saddle crossing ability

Let us illustrate the above considerations with an example fitness function in R^2 given by the sum of two Gaussian functions

$$q(\mathbf{x}) = G_{m_1, C_1}(\mathbf{x}) + G_{m_2, C_2}(\mathbf{x}) \quad (60)$$

In our example we assume $\mathbf{m}_1 = [-2, 2]$, $C_1 = [6, 2; 2, 1]$, $\mathbf{m}_2 = [2, -2]$, $C_2 = [1, -1; -1, 4]$. Contour plot of the fitness function is given in Fig. 7. Global maximum is located in the vicinity of the point \mathbf{m}_1 , and the local maximum is observed nearby \mathbf{m}_2 . The saddle point is approximately $[0.8, 1.6]$.

We ran the EA with the proportional selection, no crossover, and Gaussian mutation. Population size was equal to 100, a unit correlation matrix of the mutation was used, and the algorithm was stopped after 500 generations. Initial population was generated as clones of the point $[2, -2]$ where the fitness function takes its local maximum. Thus, EA needed

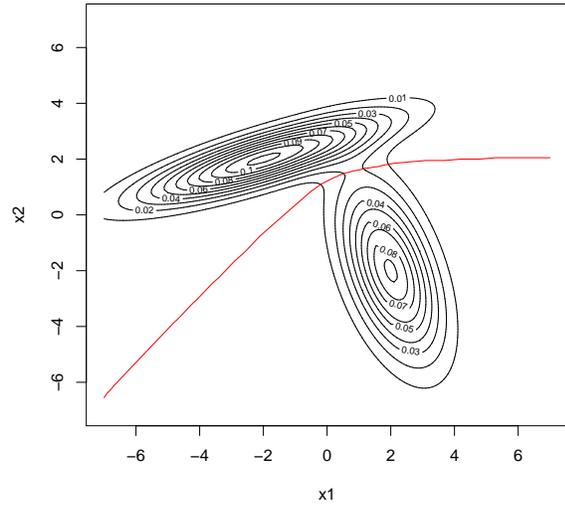


Fig. 7. Contour plot of the fitness function under consideration and the borderline between the attraction basins of the global and the local maximum

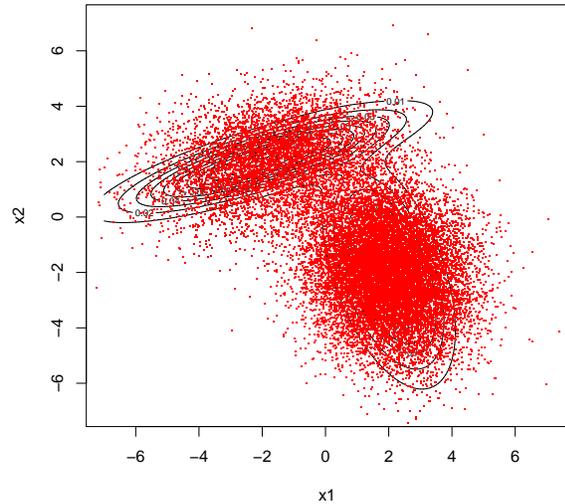


Fig. 8. All points generated by a single EA run vs. contour lines of the fitness function; see text for the details

to generate points outside the attraction basin of the local maximum in order to approach to the global one. The set of all points that have been generated by the EA is depicted in Fig. 8 It can be observed that generated points indeed form clusters whose centers are located about both maxima of the fitness function. Deeper insight into the population shows that the population periodically stays about either global or local maximum of the fitness function. This effect is indicated e.g. by the plot in which coordinates of mean values of points contained by consecutive populations are plotted as a function of the generation number (cf. Fig. 9). In the considered EA run, population is closer to the local maximum in generations 1-

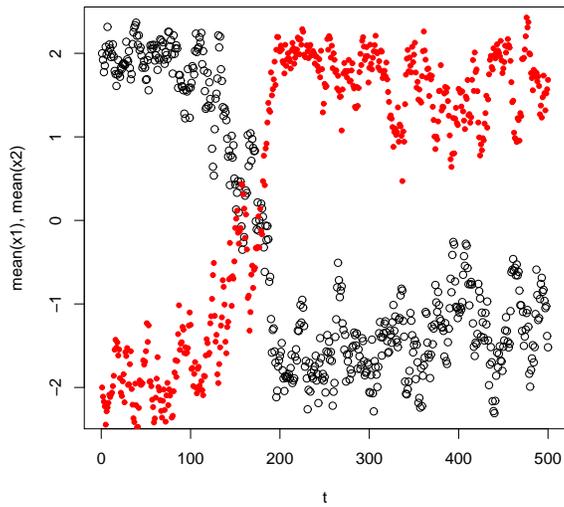


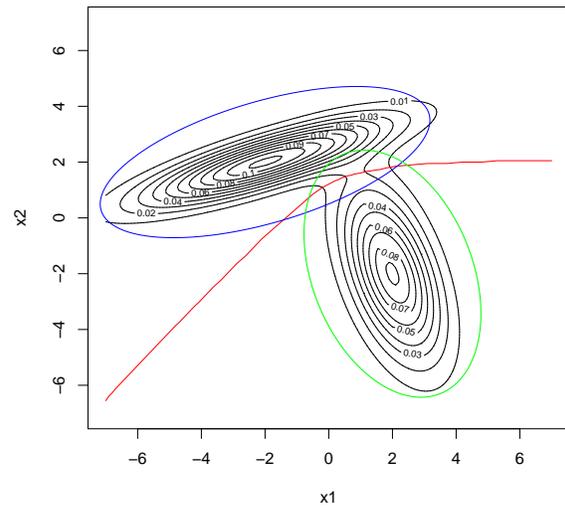
Fig. 9. Coordinates of the point being the mean of all points contained in generation number t : black circles – coordinate x_1 and red bullets – coordinate x_2 ; plot was obtained for the run whose trace of points is given in Fig. 8

100. A clear attraction by the global maximum can be observed in generations 200-500. Occasionally, population drifts further from the maximum around which it has stabilized. One of such drifts resulted in transition between the local and global maximum, which took place in generations 130-180.

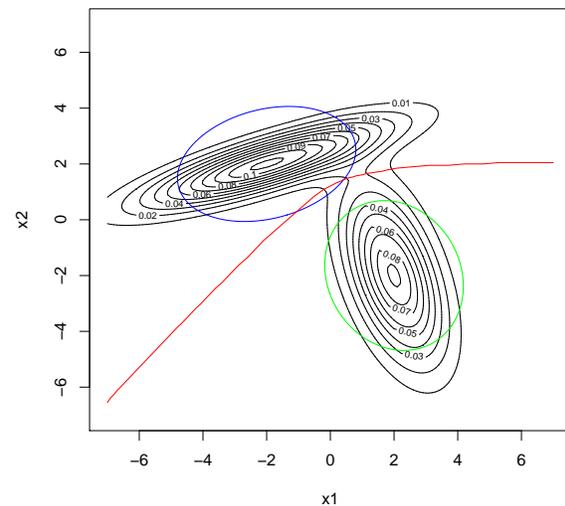
In each period when the population is attracted by a global or local maximum for a time longer than just few generations, the population can be treated as if it has been generated from a random variable with the p.d.f. which does not vary over time. This p.d.f. can be derived using formulas (23) and (24) where the fitness function is reduced to a single Gaussian term of the sum (60) which introduces either the global or the local maximum.

Consider EA for which we observed transition of the population between the attraction basins of both minima. In Fig 10a), ellipses associated with the local or the global maximum are plotted. Each ellipse is the isoline of the stable p.d.f. which encloses points that generated with probability 0.994 (a 2-dimensional version of the “ 3σ rule”). Observe that both ellipses encircle area contained by the attraction basin of the other maximum of the fitness function. Thus, it is possible for the population to perform the saddle crossing.

In the other case, when no such overlapping takes place, saddle crossing is practically impossible (although theoretically is still expected with a very small possibility). To confirm this intuition, we ran the EA with identical settings as previously, but using the arithmetical crossover which is performed with probability $p_c = 1$. Lines of equal p.d.f. value which encircle points accessible with probability 0.994 are plotted in Fig. 10b). It appears that the ellipse associated with the local maximum does not overlap the attraction basin of the global maximum, so saddle crossing can hardly be expected. Indeed, in 500 independent runs, when the population was started from clones of point [2,-2], we observed no case when



a)



b)

Fig. 10. Lines of equal p.d.f. that encircle sets where points would be generated with probability 0.994 if the fitness function were equal to a single Gaussian term of the sum (60); a) $p_c = 0$, b) arithmetic crossover, $p_c = 1$

the mean point of the population was contained in the 4th quarter of the coordinate system, which would indicate that the saddle crossing took place. In addition, we started EA in the global maximum to see if the opposite direction of the saddle crossing is possible, and again the negative result was obtained, although in that case some points from the attraction basin of the local maximum are contained in the “0.994 ellipse” associated with the global one.

For both cases under comparison, i.e., with and without arithmetical crossover, we computed covariance matrix of stable distributions derived for an infinite population, assuming that the fitness function contains only a single Gaussian term, associated with the local or the global maximum. We also estimated the covariance matrix taking points generated by the EA from periods where the population seemed to be stabilized

TABLE IV

COVARIANCE MATRIX VALUES ESTIMATED FROM THE EXPERIMENTAL DATA WHEN THE POPULATION REMAINED QUASI-STABLE ABOUT A SINGLE LOCAL MAXIMUM OF THE FITNESS FUNCTION (60); ‘L’ AND ‘G’ STAND FOR PERIODS WHEN POPULATIONS ARE STABILIZED AROUND THE LOCAL OR THE GLOBAL MAXIMUM, RESPECTIVELY.

p_c		theor.	estim.
0	l	[1.71, -0.53; -0.53, 3.13]	[1.66, -0.39; -0.39, 2.76]
0	g	[2.93, 0.59; 0.59, 1.44]	[3.85, -1.26; -1.26, 3.34]
1	l	[1.25, -0.10; -0.10, 1.54]	[1.27, -0.10; -0.10, 1.60]
1	g	[1.61, 0.17; 0.17, 1.18]	[1.66, 0.14; 0.14, 1.20]

around only one maximum. In Tab. IV we compare these covariance matrix values. For the comparison we took results from a single EA run. It can be observed that the estimated results are much better consistent with predicted for a single Gaussian term when no saddle crossing took place. It seems that the population diversity is too small to “see” anything else but the local concavity of the fitness function. This definitely confirms the intuitions formulated in the previous text and in Fig. 10b). On the other hand, when the population diversity becomes sufficiently large, sampling distribution spread has much less in common with the distribution predicted using unimodal, concave local approximation of the fitness function.

B. Small populations and the saddle crossing ability

As we already mentioned before, prediction of the spread of the sampling distribution based on the infinite population size model becomes underestimated when the population size decreases. This can have an interesting consequence for the saddle crossing ability, which can be increased by reducing the population size.

To observe this effect we ran the EA with Gaussian mutation defined by the unit covariance matrix, arithmetic crossover with probability from the set $p_c \in \{0, 0.1, \dots, 1\}$, and the population size from the range $\mu \in [2, 300]$. The algorithm was initialized with clones of the local maximum of the fitness function, and the algorithm was run for 200 generations. For each value of the population size μ , we performed 100 independent runs of the EA and in each run, we computed the fitness value of the best point found in that run. The average of the best point found in 100 independent runs can be treated as an indication of the saddle crossing ability of the EA, and in the same time, as an indicator of the EA robustness against premature convergence. In Fig. 11 we provide obtained results.

We can see a surprising pattern — average fitness of the best point tends to decrease with μ increasing, it takes its minimum for moderate size populations, and then increases. When the crossover probability is small, large populations ($\mu > 50$) perform slightly better than very small ones ($\mu < 5$). With p_c increasing, even population with hundreds of points does not perform as good as a very small one. It seems that EA which processes small populations is much less sensitive to improper settings of parameters resulting in too small spread of the sampling distribution for very large populations. It should be stressed that all algorithms under comparison were stopped after 200 generations, so e.g. when $p_c = 1$, the algorithm with $\mu = 2$, after examining 400 points, obtained better result than

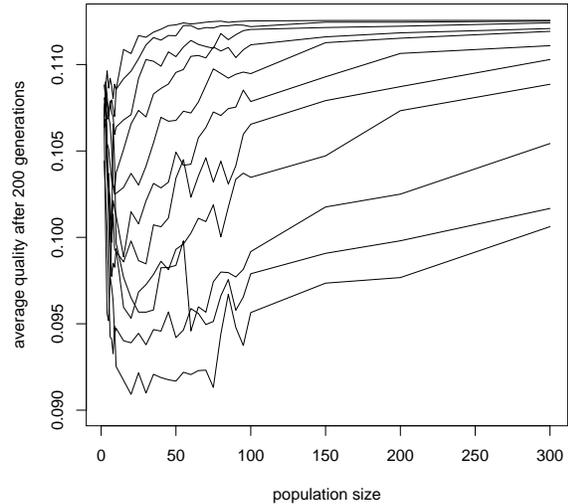


Fig. 11. Average fitness of the best point obtained for different population size values μ , for the crossover probability p_c changed gradually from 0 (upper line) to 1 (lower line)

the algorithm with $\mu = 300$ which used 60000 points. Thus, if the stop criterion were set on the number of fitness function evaluations rather than the number of generations, then the superiority of the small population over large population would be much better visible.

A look at the set of points generated in a single EA run indicates a substantial difference in EA behavior for small and large populations (see Fig. 12).

In the case of a small population, the population is more flexible and it is quite probable that the saddle crossing will appear, i.e., all individuals from the population will settle down in the attraction basin of the global maximum. Certainly, this settlement will be not for good, and one can expect that if the simulation is long, the population will oscillate many times between the local and the global maximum (cf. Fig. 12a).

When the population size grows, it becomes less flexible, and the saddle crossing is less probable. This does not exclude possibility for a point to approach the whole neighborhood of the global maximum, but rather rarely the population will cross the saddle. Points which are distant to the local maximum can be interpreted as results of macromutations — in Fig. 12b), only a few of them were luckily located nearby the point $[-2, 2]$, but they did not succeed to attract the whole population.

V. CONCLUSIONS

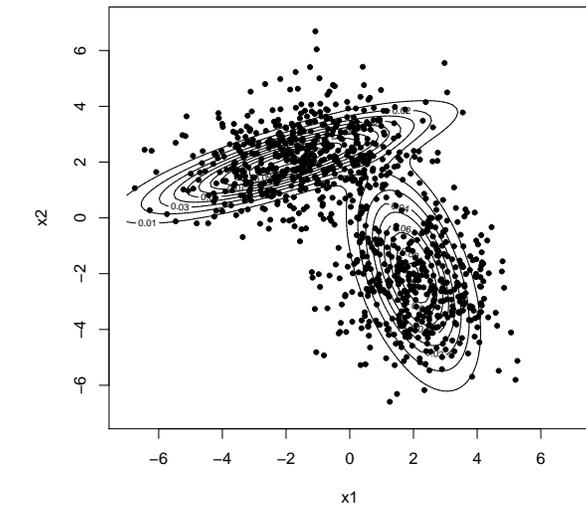
We provided a discussion of the saddle crossing property of an evolutionary algorithm. Rather than concentrating on the diversity of current population, we considered covariance matrix of the sampling distribution which defines the location of points generated by the EA.

We showed that when a Gaussian fitness function is considered, it is possible to compute an approximate form of the stable distribution which stays in a good agreement with

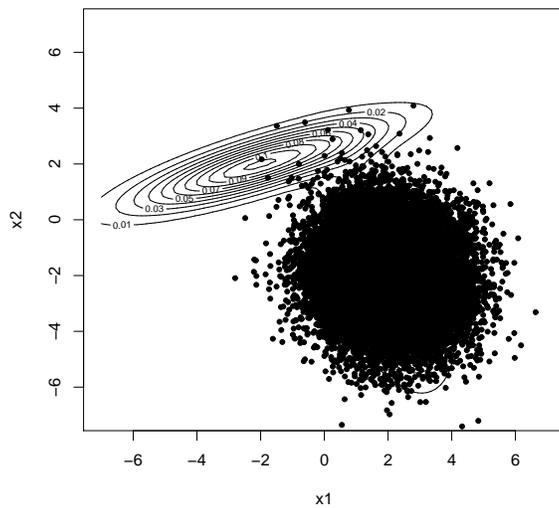
tion is small, ability to cross the saddle becomes less sensitive on the spread of the sampling distribution which allows more freedom in choosing the crossover and mutation method.

REFERENCES

- [1] Thomas Bäck, David Fogel, Zbigniew Michalewicz (eds.), *Handbook of Evolutionary Computation*, Oxford University Press, New York, 1996.
- [2] R.Galar, "Evolutionary search with soft selection", *Biological Cybernetics*, vol. 60, 1989, pp. 357–364.
- [3] I. Karcz-Duleba, "Dynamics of infinite populations evolving in a landscape of uni- and bimodal fitness functions", *IEEE Trans. on Evolutionary Computation*, vol. 5, 2001, pp. 398–409.
- [4] R. Galar, A. Chorazyczewski, "Evolutionary dynamics in space of states", in *Congress on Evolutionary Computation*, 2001, pp. 1366–1373.
- [5] R.Morrison, K. de Jong, "Measurement of Population Diversity", in *Artificial Evolution 2001*.
- [6] T. Nomura, "An analysis on crossovers for real number chromosomes in an infinite population size", in *Int. Joint Conf. on Artificial Intelligence*, 1997
- [7] X. Qi, F. Palmieri, "Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space. Part II: Analysis of the diversification role of crossover", *IEEE Trans. on Neural Networks*, vol. 5, pp. 120-128, 1994



a)



b)

Fig. 12. Sets of points generated in 200 generations of an EA with arithmetic crossover, $p_c = 0.8$; population size was set to a) $\mu = 5$ and b) $\mu = 100$

the experimental evidence. We also gave arguments that the estimated covariance matrix of the sampling distribution is a lower bound of the covariance matrix observed in the experimental data, and the approximation becomes quite good even for moderate population size.

When the fitness function is the sum of Gaussian functions, then it is possible to compute, for each Gaussian term of the sum, the p.d.f. of the stable sampling distribution as if the fitness function equal to that single Gaussian term. The integral of the aforementioned p.d.f. over the attraction basin of another local maximum is the probability to cross the saddle by a single point. Thus, if the integral is approximately zero, saddle crossing will not appear in practice.

We indicate that a possible way of coping with this problem is to reduce the population size. Moreover, when the popula-