Impatience Mechanism in Saddles' Crossing

Iwona Karcz-Duleba

Institute of Computer Engineering, Control and Robotics, Wroclaw University of Technology,
Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland

Abstract: Evolutionary inspired heuristics suffer from a premature convergence at local optima and, consequently, a population diversity loss. Thus, breaking out of a local optimum trap and crossing saddles between optima in multimodal and multidimensional search spaces is an important issue in an evolutionary optimization algorithm. In this paper, an impatience mechanism coupled with a phenotypic model of evolution is studied. This mechanism diversifies a population and facilitates escaping from a local optima trap. An impatient population polarizes itself and evolves as a dipole centered around an averaged individual. The operator was modified by supplying it with an extra knowledge about a currently found optimum. In the case, behavior of a population is quite different – a significant diversification is observed but the population is not polarized and evolves as a single cluster. Both mechanisms allow to cross saddle relatively fast for a wide range of parameters of a bimodal multidimensional fitness function.

1 INTRODUCTION

An indisputable success of evolutionary inspired heuristics in optimization results from efficiency in dealing with complex optimization problems. However, the methods are not free from drawbacks. One of the most important among them is a search stoppage at local optima which usually results in loss of a population diversity and limited ability to leave the optima and finding other, possibly better, ones. Many strategies of regaining a population diversity and maintaining multiple solutions for multimodal fitness functions were proposed. Techniques of crowding (De-Jong, 1975; Mengshoel and Goldberg, 2008), fitness sharing (Goldberg and Richardson, 1987; Sareni and Krähenbühl, 1998) and spatially-structured populations (Tomassini, 2005; Dick and Whigham, 2006) were designed just for the task.

In many practical applications, it is more important to quickly find a better than a current best solution rather than to maintain multiple solutions in a long-lasting process. We can call this task a "local global optimization". While solving the task, crossing saddles between optima by a population actually being in a state of quasi-equilibrium around a local optimum becomes a key problem. Softing a selective pressure and taking a small population size may help in leaving a current optimum. Selections with a soft selective

pressure allow for survival also worse adapted individuals thus a mean population fitness decreases and a chance of a saddle crossing increases (Galar, 1989; Goldberg and Deb, 1991; Chorazyczewski and Galar, 1998), (Chen et al., 2010). Smaller populations are more mobile than bigger ones and they are also susceptible to a genetic drift what makes saddles crossing easier (Rogers and Prügel-Bennett, 1999; Karcz-Duleba, 2006). Other ideas proposed to facilitate saddle crossing are based on modification either a fitness function (Obuchowicz, 1997) or individual fitness values. Methods based on fitness sharing (Goldberg and Richardson, 1987) or exploiting a mechanism of impatience (Galar and Kopciuch, 1999) can be classified to the latter group. The method of fitness sharing is based on the concept of finite resources to encourage elements of the population to explore different regions of the fitness function. The impatience mechanism modifies fitness of impatient individuals within a population trapped at a local optimum and promotes those individuals located on periphery of a population cloud. The mechanism causes polarization of a population which evolves as a dipole centered at a population mean. In this paper a modification of the impatience operator is proposed. It relies on adding extra knowledge concerning location of a local optimum.

Those versions of the impatience mechanism will

be studied in details in this paper which is organized as follows. In Section 2 a model of phenotypic evolution with an operator of impatience and population polarization mechanisms are described. The operator of impatience with an extra knowledge about a local optimum is depicted in Section 3. In Section 4 results of simulation studies are presented. Section 5 concludes the paper.

2 OPERATOR OF IMPATIENCE

2.1 Model of Evolution with Impatience

An operator of impatience was first introduced into a model of a simple phenotypic asexual evolution (Galar, 1989) in the paper (Galar and Kopciuch, 1999). A population composed of m elements is described in an n-dimensional unbounded real search space R^n. The evolution is subordinated to a proportional selection and a Gaussian mutation with the standard deviation σ. The model is generational what means that the whole population is replaced in every generation. An operator of impatience replaces an original fitness of an individual by a fitness depending on the distance from a mean individual within a population.

More formally, let $x_j \in R^n$ is a position of the jth individual characterized by its fitness $q_j(x)$ and d_j denotes a distance of this individual from the population mean. A modified fitness of the j-th individual is given by

$$q_a(x_j) = (\frac{d_j}{d_A} + c)q(x_j) \qquad (1)$$

where d_A is an average distance of all individuals from the mean of the whole population

$$\overline{x} = \frac{1}{m}\sum_{i=1}^{m} x_i, \quad d_A = \frac{1}{m}\sum_{i=1}^{m} \|x_i - \overline{x}\| = \frac{1}{m}\sum_{i=1}^{m} d_i$$

and c is a constant parameter taken from interval $[0,1]$. Some past (Galar and Kopciuch, 1999) and preliminary results have revealed a negligible influence of parameter c on performance of the impatience mechanism, therefore its value will be set to $c = 1$ later on.

An impact of the operator of impatience on crossing saddles was studied in an adaptive landscape of the n-dimensional bimodal fitness function being a sum of two bell shaped functions with different heights forming a distinct saddle

$$q(x) = \exp(-a_1 \sum_{k=1}^{n} x_k^2) + h\exp(-a_2(x-1)^2 + a_2 \sum_{k=2}^{n} x_k^2)$$
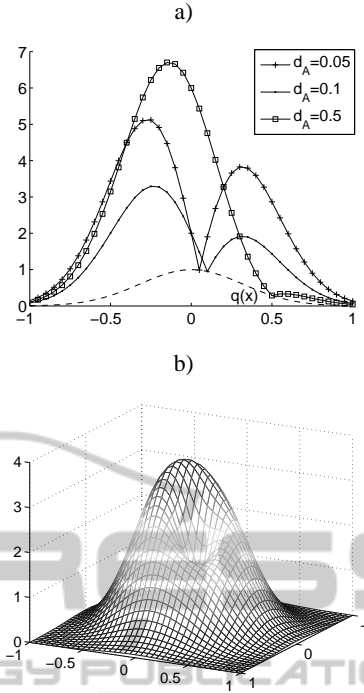$$(2)$$

Figure 1: a) A fitness transformation of one-dimensional local optimum hill q (denoted by a dotted line) by the operator of impatience (1) for three population mean locations: $d_A = 0.05, 0.1, 0.5$. b) a fitness transformation of two-dimensional local optimum hill q by operator of impatience (1), $d_A = 0.1$.

where $h \geq 1$. A local optimum with a height equal to one is located at $x_l = (0,0,\ldots,0)$ and the global optimum with the height $h = 2$, shifted in the first coordinate, is located at $x_g = (1,0,\ldots,0)$. Parameters a_1, a_2 influence slopes of both hills. In all experiments presented, a slope of the local hill was set to $a_1 = 5$ (by rescaling coordinates linearly any other case of quality function (2) can be transformed into the case considered). The global hill becomes slimmer and slimmer as parameter a_2 increases and, consequently, a saddle between both optima is wider and deeper (see also Fig. 14.a). The width Δ_s of a saddle was defined as the shortest distance between the top of the local hill and a point at the global hill with the same value as at the local optimum attains. Later on the parameter was scaled in units of standard deviation of mutation σ and depicted as Δ_s'.

The operator of impatience decreases quality of individuals close to the population mean and increases quality of those located at peripheries of the population, i.e. distant from the population mean.

In Fig. 1 exemplary transformations, according to Eq. (1), of one- and two-dimensional fitness function (2) is presented. As the operator of impatience is

a) after generation No. 30

b) after generation No. 40
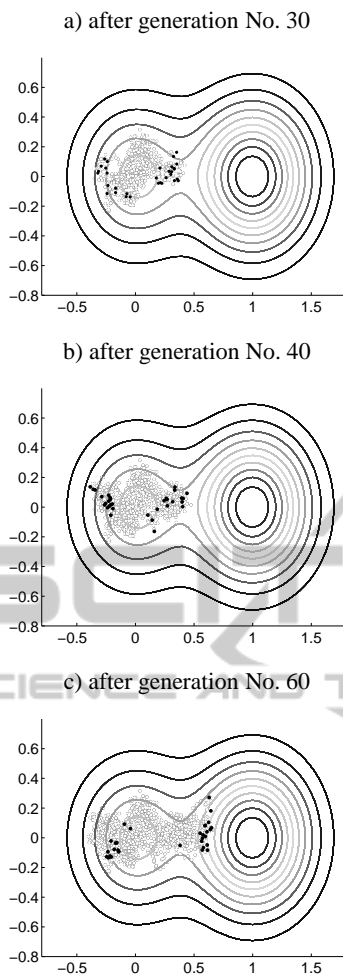
c) after generation No. 60

Figure 2: Stages of saddle crossing by a polarized population, $m = 32$, $n = 2$, $\sigma = 0.025$.

applied to the population placed around the local hill, only the first (local) component of Eq. (2) was visualized. Fitness values of individuals located far away from the population mean increases significantly and, what is more interesting, the function is not symmetrical.

2.2 Polarization of Population

The action of the operator of impatience is illustrated in Fig. 2. As mentioned before, the operator increases the effective adaptation of those individuals located on the boundary of a population cluster by increasing probability of their selection (cf. Fig. 1). Thus, the distant from the mean individuals reproduce more intensively and increase a population diversity. Too large dispersion of a population is counteracted by a decrease in fitness of far-away-from-optimum individuals subjected to selection. Consequently, a pop-

ulation is located on some kind of orbit around the population mean where both trends are in balance and the orbit acts as an attractor for a fluctuating, around an optimum, population. Additionally, some locations on the orbit are favored (two optima in Fig. 1.a). In effect, a population polarizes and spontaneously breaks into two clustered sub-populations spawning a kind of dipole (Fig. 2.a). The dipole fluctuates along the orbit. A number of individuals that belong to each sub-population changes dynamically.

When a local hill is separated from another optimum by a saddle, a dipole may be attracted by the saddle where the adaptation is slightly higher than on other parts of the orbit, and directed its axis along the saddle (Fig. 2.b). In result, a saddle can be crossed faster by this part of a dipole situated closer to a saddle. A saddle crossing is done by one of the sub-populations whereas the other one remains somewhere around a local hill. Remaining one of sub-populations on a local optimum may be explained when looking at Fig. 1.a and a function depicted to a mean of population $d_A = 0.5$. In the case, sub-population is about to cross a saddle which is $\Delta_s = 0.6$ wide. Transformed fitness of individuals on a local hill is very high comparing to individuals just crossing saddles, so they stay on a local optimum (Fig. 2.c). This sub-population slowly becomes extinct after saddle crossing as individuals on the global hill reproduce themselves more often.

The comparison of a phenotypic evolution with proportional selection and Gaussian mutation with and without impatience mechanism involved is presented in Fig. 3. Evolution of both populations was initialized at a local hill, starting with a uniform population of sub-optimal individuals. For a population treated by the operator of impatience, the operator was turned on after five generations, letting for a preliminary diversification of clones. Such a population imitates a state of quasi-stability around the local optimum. The polarized population is much widely distributed and crossed a saddle before a population without impatience.

In Figs. 4-5 some statistics of both populations from Fig. 3 are presented. The diversification of a population was defined as an average distance from the mean individual. For a population with the impatience operator, the diversification is higher than for the population without it, i.e. distances from a population mean for polarized population are larger than for non-polarized one (Fig. 4). Thus, the impatience operator may be classified into diversification preserving methods. An average fitness of the population without the impatience mechanism is changed slightly (about 10%) and remains at a similar level whereas the fit-
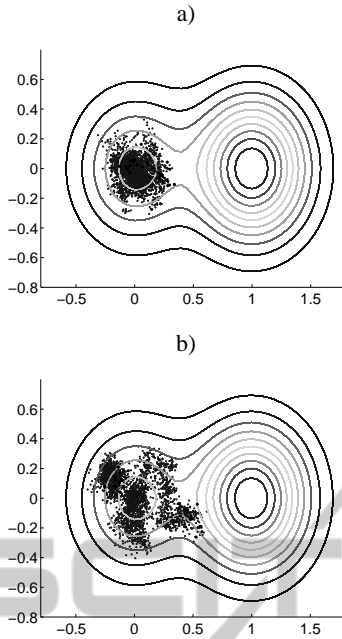
a)

b)

Figure 3: Evolution without (a) and with (b) operator of impatience; traces of all generations from No. 1 to No. 80 are presented; $m = 32$, $n = 2$, $\sigma = 0.025$.
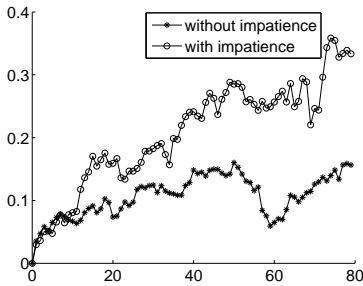

Figure 4: Diversification of populations - an average distance from the mean individual for populations from Fig. 3.

ness of the population with the impatience operator is greatly reduced (about 20%-30%), what may help in saddle crossing (Fig. 5).

3 IMPATIENCE WITH KNOWLEDGE

The operator of impatience is switched on when a population gets stuck at a quasi-equilibrium state around a currently found optimum. Thus, the value of the optimum either is known or can be reliably estimated. So, we decided to include the extra knowledge about the optimum placement into the operator of impatience and studied its influence on population dynamics.
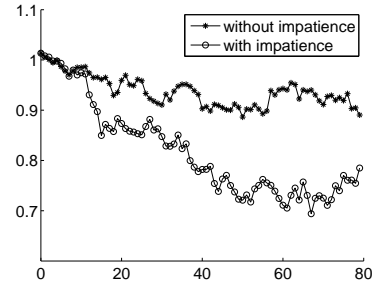

Figure 5: Average fitness of populations from Fig. 3.

Consequently, operator (1) is changed in such a way that the distance from the population mean is replaced with the distance from the (estimated) optimum. A modified fitness of the j-th individual located at distance d_{jo} from the local optimum is now described by equation

$$q_K(x_j) = \left(\frac{d_{jo}}{d_{Ao}} + c\right)q(x_j) \qquad (3)$$

where d_{Ao} denotes an average distance from the optimum for the whole population

$$d_{Ao} = \frac{1}{m}\sum_{i=1}^{m}\|x_i - x^*\| = \frac{1}{m}\sum_{i=1}^{m}d_{io}$$

and x^* is a placement of the local (possibly estimated) optimum.

While studying populations in a landscape of fitness function (2) with the local optimum placed at $x = 0$, it is possible to figure out (cf. Fig. 6) how an individual fitness, described by Eq. (3), is modified. Similarly to Fig. 6 only the first (local) component of Eq. (2) and, as the transformation Eq. (3) is symmetrical, only positive half-plane was visualized. The upper panel presents the modified one-dimensional local optimum for two values of average distance $d_{Ao} = 0.1, 0.2$. The bottom panel displays a two-dimensional local optimum hill transformation for $d_{Ao} = 0.1$. The operator of impatience with knowledge decreases a quality of near optimal individuals and increases of those located at some distance from the optimum. When a population is closer to a local optimum (i.e. the average distance d_{Ao} is getting small), a transformed quality of some distant individuals highly inceases (more than twice). When the population is more dispersed, d_{Ao} is relatively large, the transformed quality of an individual is smaller but still exceeds the local hill quality.

In Fig. 7 stages of saddle crossing by the population influenced by the operator of impatience with knowledge are presented. The main difference between the operator of impatience with and without knowledge is clearly visible: the population, for
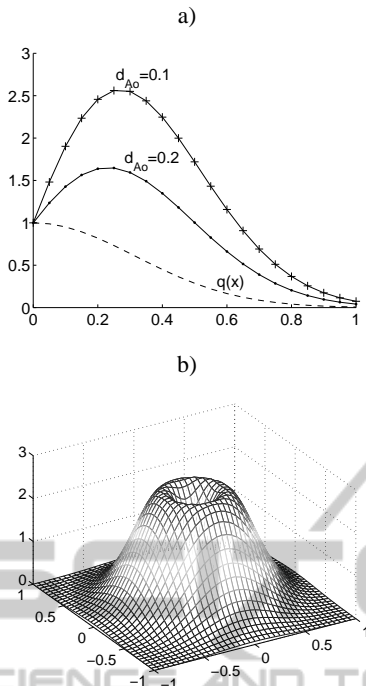
a)

b)

Figure 6: a) A fitness transformation of one-dimensional local optimum hill q (denoted by a dotted line) by the operator of impatience with knowledge (3) for two average distances: $d_{Ao} = 0.1, 0.2$. b) the fitness transformation of two-dimensional local optimum hill q by operator of impatience, $d_{Ao} = 0.1$.

which operator of impatience is calculated acording to Eq. (3) is widely diversified but it is not polarized (cf. Fig. 2). The difference can be explained as follows. Populations evolved without knowledge have got no information about a fitness function configuration. Their knowledge is restricted to information about an actual population itself. When operator with knowledge (3) is acting, population is located on an orbit around the local optimum. All locations on the orbit are equally likely (cf. Fig. 6). Additionally, existing in evolving populations tendency to concentrating individuals imply that population is not polarized and evolves as a cluster, but is more dispersed than population without impatience operator.

Informative characteristics (diversification, average fitness, average distances from the local optimum) of populations evolved with the operator of impatience with and without knowledge are depicted in Figs. 8-10. Both simulations were stopped after 60 generations, just after crossing a saddle by some individuals but not by the population mean (what can be seen in the average fitness plot c.f. Fig. 8). In all figures, the operator of impatience with knowledge was marked as wK (with Knowledge). The average fitness of both populations are similar: after turning
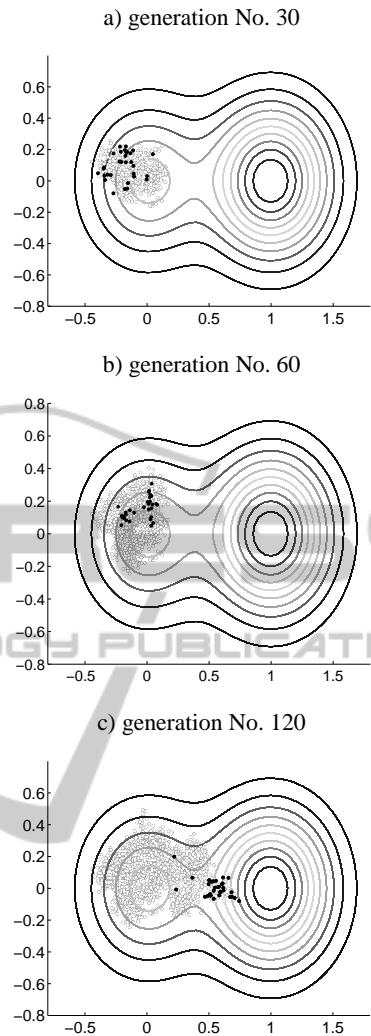

a) generation No. 30

b) generation No. 60

c) generation No. 120

Figure 7: Stages of saddle crossing by a population subjected to the operator of impatience with knowledge, $m = 32$, $n = 2$, $\sigma = 0.025$.
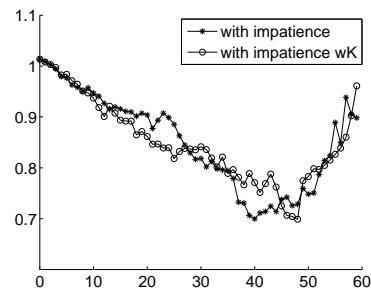

Figure 8: An average fitness of populations evolved with the operator of impatience with (wK) and without knowledge after 60 generations, $m = 32$, $n = 2$, $\sigma = 0.025$.
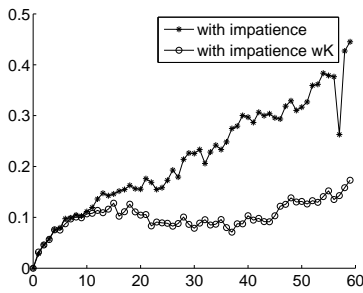
Figure 9: Diversification of population - an average distance of population individuals from the mean for whole populations evolving with the operator of impatience with (wK) and without extra knowledge after 60 generations, $m = 32$, $n = 2$, $\sigma = 0.025$.
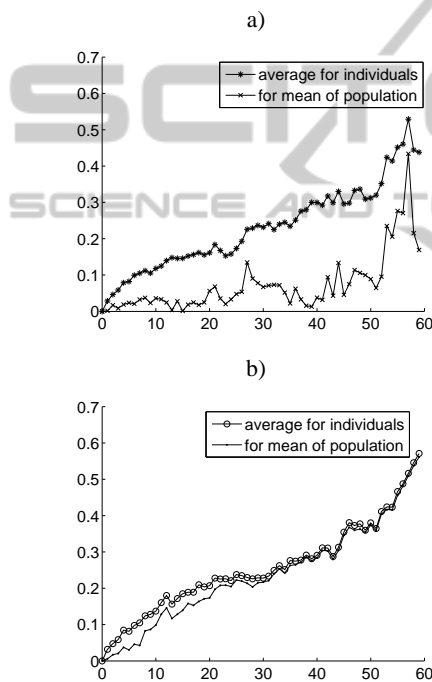
a)

b)

Figure 10: An average distance of population individuals from the local optimum and the distance of the population mean from the local optimum. Evolution with the operator of impatience: a) without knowledge, b) with knowledge (wK) after 60 generations, $m = 32$, $n = 2$, $\sigma = 0.025$.

on the operator fitness decreases about 20-30% and while crossing saddle fitness increases again (Fig. 8). In order to understand actions of both operators better, not only distances of individuals from the population mean but also distances from the local optimum were calculated. Populations evolved with the operator of impatience without knowledge is polarized, so it is more dispersed than the population with knowledge. The average distance of population individuals from the population mean for the operator of impatience without knowledge is growing during the pre-

sented evolution run and achieved quite a high value after crossing sadle as polarized sub-populations are located on both optima (Fig. 9). For the impatience operator with knowledge, the average distance remains on similar level during evolution on a local hill and increases only slightly while crossing a saddle. Distances from the local optimum for both operators were calculated for the mean individual and averaged for all population individuals (Fig. 10). The mean individual of population evolved with the operator of impatience without knowledge is located close to the optimum until population crossed saddle. This means that while population wanders around the local optimum, a center of the dipol is located very close to optimum. When population crossed a saddle, the distance to the local optimum inreases and the center of the dipol moves towards the global optimum. An average distance of individuals from the local optimum increases slowly as a population is diffused and one sub-population crossed a saddle. For the operator of impatience with knowledge both distances are very similar as a population is not polarized and evolves as whole.

4 EFFICIENCY OF CROSSING SADDLES: SIMULATION RESULTS

The efficiency of crossing a saddle by a population subjected to the operator of impatience with and without knowledge was studied for fitness function (2). The total number of fitness function evaluation needed to cross a saddle and averaged over 1000 runs was selected as a performance criterion. Computations were terminated when a hypotetical, average individual crossed a saddle, i.e. the average population fitness exceeded the value of the local optimum. Results were related to the phenotypic evolution with a proportional selection and the Gaussian mutation without using the impatience operator.

An influence of the following parameters on crossing saddles were tested:

- population size m (Fig. 11),
- search space dimension n (Fig. 12),
- saddle width Δ'_s (scaled in units of standard deviation of mutation σ) (Fig. 13),
- saddle depth (depending on a slope, a_2, of the global optimum) (Fig. 14).

If not varied, the values of other parameters $m = 16$, $n = 4$, $\sigma = 0.05$, $a_2 = 5$, $\Delta'_s = 12$ were set. Because the slope a_2 of the global optimum causes not only
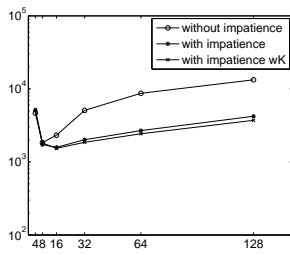
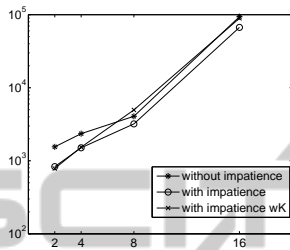Figure 11: Efficiency of saddle crossing - impact of population size m.

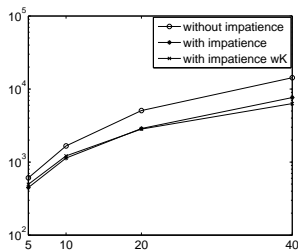Figure 12: Efficiency of saddle crossing - impact of a search space dimension n.

Figure 13: Efficiency of saddle crossing - impact of saddle width Δ'_s (scaled in units of standard deviation of mutation).

deepening but also widening of a saddle, for every a_2 an appropriate value of σ was calculated (Fig. 14.a, Tab. 1) to preserve a constant saddle width $\Delta'_s = 12$.

As it can be seen in Figs. 11-14, an evolution exploiting the impatience operator outperforms its impatience-free version in a wide range of parameters of the fitness function and the model. An extra knowledge concerning the location of the local optimum accelerates crossing saddles only slightly. The bigger benefit is obtained for large populations ($m = 32$–128). What is more interesting, increasing a population size does not increase significantly the number of fitness function calculations needed to cross a saddle (Fig. 11). Efficiency of the three evolution models

Table 1: Values of σ for different parameters a_2 to preserve a constant saddle width $\Delta'_s = 12$ (scaled in units of σ).

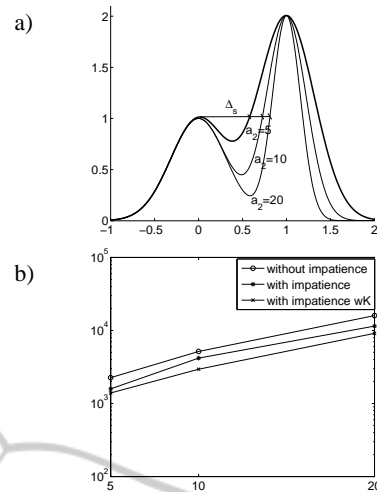| a_2	5	10	20
σ	0.05	0.063	0.071

Figure 14: a) Profile of fitness function (2) for different parameters a_2 with depicted unscaled value of Δ_s. b) Efficiency of saddle crossing - impact of a saddle depth (depending on the slope parameter a_2).

Table 2: Number of generations needed to find 2nd and 3rd optimum averaged over 100 runs and the number of unsuccessful runs; $m = 32$, $\sigma = 0.2$.

knowledge	2nd opt.	3rd opt.	unsuccess.
without	362	403	6
with	417	454	7

were similar for small search space dimensionallities $n = 2$–8 (Fig. 12), although the model without impatience is slightly worse. As the dimensionallity n grows, the efficiency of model with impatience and knowledge decreases and resembes that without impatience. The best performance displayed the model with impatience. Populations evolving with the impatience operator crossed efectivelly large and deep saddles (Figs. 13-14) and the efficiency of both versions of the operator are comparable.

Aforementioned examples are rather simple but recent results concerning more difficult quality functions are promising. In Fig. 15 the optimisation of multimodal Ursem 03 test function

$$f_U(x,y) = \sin(2.2\pi x + 0.5\pi)(2 - |y|)(3 - |x|)/4 + \\ + \sin(0.5\pi y^2 + 0.5\pi)(2 - |y|)(2 - |x|)/4$$

using the operator of impatience without knowledge is illustrated. The initial population was located on the first optimum and calculations were terminated after 5000 generations. In Table 2 average results of 100 runs are presented. In this case the impatience operator without knowledge is a little bit more efficient that that with the knowledge.
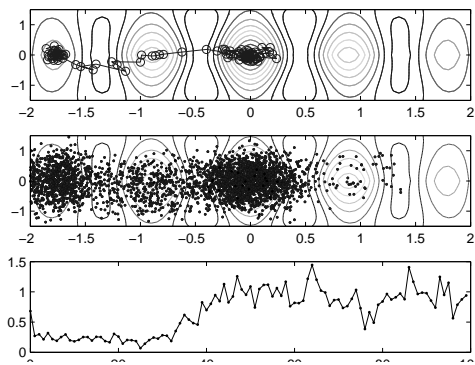
Figure 15: Ursem function optimization with the operator of impatience without knowledge. Upper panel: population mean location; middle panel: paths of all individuals; lower panel: average fitness; $m = 32$, $n = 2$, $\sigma = 0.2$.

5 CONCLUSIONS

In this paper a mechanism of impatience was checked as a method to preserve diversifiction of a population and thereby escaping a local optimum trap. Two versions of the impatience operator were examined: with and without extra knowledge concerning (estimated) position of a local optimum. Both versions increase diversity of population. However, when an impatience is related to a current population mean, the polarization of a population was observed (a population is divided into dipol-like sub-populations). When the impatience operator is related to a local optimum placement the polarization effect was not observed.

In contrary to other diversity preserving methods (fitness sharing and clearing), the impatience mechanism is parameter-free and increases a computational effort only slightly. The mechanism is related to a current population state and/or already explored parts of a search space, so it can be used in dynamical landscapes.

Actions of both versions of impatience operators were tested to check the efficiency of crossing a saddle between the local and the global optimum of bimodal multidimensional fitness functions. Both versions demonstrated high efficiency in crossing multidimensional, width and deep saddles. Preliminary results of applying the mechanisms to more complex test functions are promising and they will stimulate our future work.

REFERENCES

Chen, T., He, J., Chen, G., and Yao, X. (2010). Choosing selection pressure for wide-gap problems. *Theoretical Computer Science*, 411(6):926–934.

Chorazyczewski, A. and Galar, R. (1998). Visualization of evolutionary adaptation in R^n. In *Evolutionary Programming VII*, pages 657–668.

DeJong, K. (1975). *An analysis of the behavior a class of genetic adaptive systems*. PhD thesis, University of Michigan.

Dick, G. and Whigham, P. A. (2006). Spatially-structured evolutionary algorithms and sharing: Do they mix? In et al., T. W., editor, *SEAL 2006*, LNCS no. 4247, pages 457–464. Springer-Verlag.

Galar, R. (1989). Evolutionary search with soft selection. *Biological Cybernetics*, 60:357–364.

Galar, R. and Kopciuch, P. (1999). Impatience and polarization in evolutionary processes. In *Proc. of KAEiOG Conference*, pages 115–122. in Polish.

Goldberg, D. and Deb, K. (1991). A comparative analysis of selection schemes used in genetic algorithms. In *Foundations of Genetic Algorithms*, pages 69–93. Morgan Kaufmann.

Goldberg, D. and Richardson, J. (1987). Genetic algorithms with sharing for multi-modal function optimisation. In *Proc. of the 2nd Int. Conf. on Genetic Algorithms and Their Applications*, pages 41–49.

Karcz-Duleba, I. (2006). Dynamics of two-element populations in the space of population states. *IEEE Transactions on Evolutionary Computation*, 10(2):199–209.

Mengshoel, O. and Goldberg, D. (2008). The crowding approach to niching in genetic algorithms. *Evolutionary Computation*, 16(3):315–354.

Obuchowicz, A. (1997). The evolutionary search with soft selection and deterioration of the objective function. In *Proc. 6th Int. Conf. Intel. Inform. Systems*, pages 288–295.

Rogers, A. and Prügel-Bennett, A. (1999). Genetic drift in genetic algorithm selection schemes. *IEEE Transactions on Evolutionary Computation*, 3(4):298–303.

Sareni, B. and Krähenbühl, L. (1998). Fitness sharing and niching methods revisited. *IEEE Transactions on Evolutionary Computation*, 2(3):97–106.

Tomassini, M. (2005). *Spatially structured evolutionary algorithms*. Springer-Verlag.