Eur. Phys. J. B **21**, 145–152 (2001)

THE EUROPEAN PHYSICAL JOURNAL B
EDP Sciences

Stochastic dynamics of adaptive evolutionary search

A study on population-based incremental learning

E. Galic[a]

Institute of Physics, University of Wuerzburg, Am Hubland, 97074 Wuerzburg, Germany

Received 23 October 2000 and Received in final form 22 February 2001

Abstract. In this paper the stochastic dynamics of adaptive evolutionary search, as performed by the optimization algorithm Population–Based Incremental Learning, is analyzed with physicists' methods for stochastic processes. The master equation of the process is approximated by van Kampen's small fluctuations assumption. It results in an elegant formalism which allows for an understanding of the macroscopic behaviour of the algorithm together with its fluctuations. We consider the search process to be adaptive since the algorithm iteratively reduces its mutation rate while approaching an optimum. On the one hand, it is this feature which allows the algorithm to quickly converge towards an optimum. On the other hand it results in the possibility to get trapped by a local optimum only. To arrive at a detailed understanding we discuss the influence of fluctuations, as caused by mutation, on this behaviour. We study the algorithm for rather small sytem sizes in order to gain an intuitive understanding of the algorithm's performance.

PACS. 89.20.Ff Computer science and technology – 87.23.Kg Dynamics of evolution – 05.10.Gg Stochastic analysis methods (Fokker-Planck, Langevin, etc.)

1 Introduction

Population-Based Incremental Learning (PBIL) [1] was introduced as both, a general model for evolutionary optimization techniques, like the Genetic Algorithm [2], as well as an alternative to them, since it results in an elegant and easy to implement procedure for binary coded search problems. It is due to this fact that in recent years PBIL successfully has been applied to real world optimization tasks [3,4]. On the other hand theoretical results describing the dynamics of the algorithm are rare.

This work is thought of being a step towards an understanding of the algorithm, by application of physicists' methods for the analysis of stochastics processes. The analysis is done by use of earlier developed methods, namely van Kampen's linear noise approximation of the master equation of a stochastic process [5]. The formalism was previously utilized for the analysis of online and reinforcement learning schemes in neural networks, such as perceptrons or self-organizing maps, like in [6–9] and references therein.

The deepest understanding of PBIL, so far, can be found in [10], where convergence for a linear search problem has been proved and the behaviour for a non-linear function has been illustrated. The analysis given there is based on studies of the average of PBIL's iteration equation solely. The study given in the work here confirms their findings and considerably extends the older analysis in several directions. Firstly, by considering fluctuations within van Kampen's approximation, one can show that PBIL's behavior is only correctly described by the average of its system variable locally and when fluctuations behave sufficiently harmless (to be defined later). Secondly, the dynamics of the average behavior, together with the dynamics of the superimposed fluctuations is calculated exactly. Finally, the influence of PBIL's parameters, namely learning rate and population size, on the dynamics is discussed.

2 Population-based incremental learning

Let us consider binary search spaces $\{0,1\}^L$, where L is the dimension of the search space and at the same time the length of the genomes or individuals. Furthermore, let us consider individuals $\mathbf{x} \in \{0,1\}^L$ and populations $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, where N is the number of individuals within a population. The individuals are the corners of the L-dimensional hypercube and represent potential solutions of the search problem. To each individual $\mathbf{x}$ a fitness or evaluation value $f(\mathbf{x}) \in \Re$ is deterministically assigned. We consider minimization problems, consequently the goal of search is to find the set of solutions with minimal evaluation value.

PBIL is based on the idea that the individuals of a population may be replaced by parameters determining their probability density. These parameters are stored in a

[a] e-mail: elvisgalic@hotmail.com

probability vector $\mathbf{p}$, specifying the probabilities for each site to be equal 1. From this probability vector in each iteration step, or generation, a population of individuals is drawn. The population then is evaluated and the best individual is selected (Winner–takes–all–selection). In a final step, the probability vector is shifted with a learning rate η into the direction of the selected individual.

Let us explain the algorithm discussed here: Starting from some initial probability vector $\mathbf{p}_0 \in]0,1[^L$, the algorithm is executed for each iteration step or generation $i = 0, 1, \ldots, I$ as follows:

1. Sample a population $\mathbf{X}_i$ of N individuals from $\mathbf{p}_i$ for each site independently
2. Choose the individual $\mathbf{b}_i \in \mathbf{X}_i$ with minimal evaluation value from the population
3. Update $\mathbf{p}_i$ in direction of the best individual $\mathbf{b}_i$
4. This yields the new probability vector $\mathbf{p}_{i+1}$. Go to Step 1

As update rule, a moving average with exponential forgetting in generation i is used. According to this rule, $\mathbf{p}$ is shifted with a learning rate $\eta \in]0,1[$ towards the selected individual $\mathbf{b}$

$$\Delta\mathbf{p}_i = \mathbf{p}_{i+1} - \mathbf{p}_i = \eta(\mathbf{b}_i - \mathbf{p}_i)\ . \tag{1}$$

The process caused by the update rule is stochastic and Markovian, since the system state $\mathbf{p}_i$ and the population $\mathbf{X}_i$ are random variables which only depend on the preceding system state. PBIL's search takes place inside the binary hypercube. Because of the particular form of the update rule the process is restricted to the interval $]0,1[^L$, thus has natural boundary conditions. The selection of the best individual $\mathbf{b}_i$ depends on the current system state $\mathbf{p}_i$, the size of the population N and the particular evaluation values of the current population $\mathbf{X}_i$.

The main difference between PBIL and the Genetic Algorithm is the representation of the population. Additionally, PBIL does not explicitly mutate individuals. However, caused by the generation or sampling process of the individuals from the probability vector, mutation is implicitly realized, since a variety of different individuals might be generated from a single probability vector. Furthermore, since the probability vector eventually will converge towards one of the corners of the cube, the variety of the population will decrease. This might be interpreted as an adaptive mutation rate (or adaptive cooling), which reduces itself while approaching an optimum.

For a first illustration on how PBIL's search takes place, three randomly chosen searches starting at $\{0.5, 0.5\}$ are shown for $N = 2$, $\eta = 0.05$ and specific problems of sizes $L = 2$ as given by [10]. In Figure 1 search for a linear search problem of the form

$$f(\mathbf{x}) = x_1 + x_2\ , \tag{2}$$

is shown. The search space as defined by the evaluation function has one global optimum, given by the individual $\mathbf{x} = \{0, 0\}$. The figure illustrates how all three searches converge towards the global optimum, however with a considerable amount of fluctuations. In Figure 2 search for a nonlinear search problem of the form

$$f(\mathbf{x}) = 3x_1 + 2x_2 - 4x_1x_2\ , \tag{3}$$

is shown. The search problem has one global optimimum, given by individual $\mathbf{x} = \{0, 0\}$, and one local one, as given by individual $\mathbf{x} = \{1, 1\}$. The figure illustrates how two searches converge towards the global optimum and one ends up in the local one. It is the goal of this work to describe the tendencies and fluctuations as illustrated in the simulations.

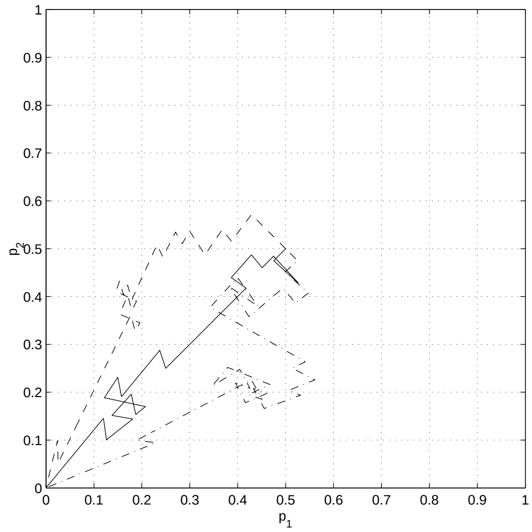

Fig. 1. Three randomly chosen searches for the 2–dimensional linear search problem with a unique optimum, the corner $\{0, 0\}$, are shown. All searches converge towards the global optimum, however with a considerable amount of variation.

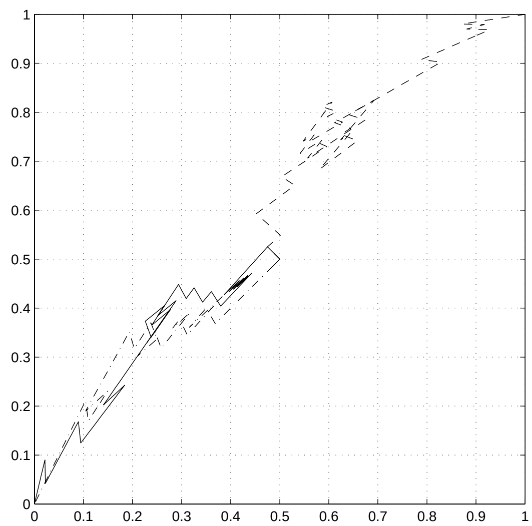

Fig. 2. Three randomly chosen searches for the 2–dimensional nonlinear search problem with a global optimum, the corner $\{0, 0\}$, and a local optimum, the corner $\{1, 1\}$, are shown. Two searches converge towards the global optimum, while a single one converges towards the local one. Again a considerable amount of variation is observed.

3 The Fokker-Planck equation

Within the analysis the focus is on the probability density $w_i(\mathbf{p})$ of finding the system in state $\mathbf{p}$ at generation i. The dynamics of this probability density formally is written

$$w_{i+1}(\mathbf{p}) = \int t(\mathbf{p}|\mathbf{p}')w_i(\mathbf{p}')\mathrm{d}\mathbf{p}' , \tag{4}$$

with $t(\mathbf{p}|\mathbf{p}')$ being the transition probability to generate $\mathbf{p}$ from $\mathbf{p}'$ in one generation. It is given by

$$t(\mathbf{p}|\mathbf{p}') = \sum_{\{\mathbf{b}\}} \phi(\mathbf{b}|\mathbf{p}')\delta(\mathbf{p} - \mathbf{p}' - \eta(\mathbf{b} - \mathbf{p})), \tag{5}$$

where $\delta(...)$ is the Dirac–Delta function, the sum runs over all possible events $\{\mathbf{b}\}$ leading to an individual to be generated and selected for the update (Eq. (1)) as a best one $\mathbf{b}$ and $\phi(\mathbf{b}|\mathbf{p}')$ is the probability that a random vector $\mathbf{b}$ is chosen for the update given the system state $\mathbf{p}'$.

In [11] it has been shown that such a discrete time Markov process can be converted into a continuous one by considering the time intervals Δt between successive discrete iteration steps to be randomly sampled from a probability density of the form $\rho(\Delta t) = \tau^{-1}\mathrm{e}^{-\Delta t/\tau}$. Since the probability

$$\pi(i,t) = \frac{1}{i!}\left(\frac{t}{\tau}\right)^i \mathrm{e}^{-\frac{t}{\tau}}, \tag{6}$$

that after time t there have been i iteration steps, then obeys a Poisson process, one can define the probability $W(\mathbf{p},t)$ that the system is in state $\mathbf{p}$ at some continuous time t as

$$W(\mathbf{p},t) = \sum_{i=0}^{\infty} \pi(i,t)\, w_i(\mathbf{p}). \tag{7}$$

Differentiation of the equation results in a continuous time Master equation of the form

$$\frac{\partial W(\mathbf{p},t)}{\partial t} = \int \Big(T(\mathbf{p}|\mathbf{p}')W(\mathbf{p}',t) - T(\mathbf{p}'|\mathbf{p})W(\mathbf{p},t)\Big)\mathrm{d}\mathbf{p}', \tag{8}$$

with $T(\mathbf{p}|\mathbf{p}') = t(\mathbf{p}|\mathbf{p}')/\tau$ being the transition probability per a unit time.

In the next step a small fluctuations expansion in the learning rate η is performed according to van Kampen's system size expansion [5]. It assumes that for small learning rates, the stochastic process can be approximated by a deterministic trajectory $\langle\mathbf{p}\rangle_t$ superimposed by fluctuations $\zeta(t)$. Formally, it can be written in the form

$$\mathbf{p}(t) = \langle\mathbf{p}\rangle_t + \sqrt{\eta}\zeta(t), \tag{9}$$

where the average, as denoted by $\langle...\rangle_t$, is taken over all possible states at time t.

Introducing equation (9) into equation (8) and performing the so called Kramers–Moyal expansion, a series expansion of the Master equation for small learning rates η, one finds, when truncating the series after second order in $\sqrt{\eta}$, for the time evolution of the probability density $W(...) = W(\langle\mathbf{p}\rangle_t + \sqrt{\eta}\zeta(t), t)$ the Fokker-Planck equation

$$\frac{\tau}{\eta}\frac{\partial W(...)}{\partial t} = -\sum_k A_{kl}(\langle\mathbf{p}(t)\rangle)\frac{\partial \zeta_l W(...)}{\partial \zeta_k} + \frac{1}{2}\sum_{kl} B_{kl}(\langle\mathbf{p}(t)\rangle)\frac{\partial^2 W(...)}{\partial\zeta_k\,\partial\zeta_l} , \tag{10}$$

which solution is given by a multivariate Gaussian and its moments are determined by

$$A_{kl}(\langle\mathbf{p}(t)\rangle) = \frac{\partial}{\partial\langle p_l\rangle}\sum_{\{\mathbf{b}\}}(b_k - p_k)\phi(\mathbf{b}|\mathbf{p}) \tag{11}$$

and

$$B_{kl}(\langle\mathbf{p}(t)\rangle) = \sum_{\{\mathbf{b}\}}(b_k - p_k)(b_l - p_l)\phi(\mathbf{b}|\mathbf{p}) . \tag{12}$$

To make the assumption hold, fluctuations are not allowed to increase in time to the order of magnitude of the deterministic or macroscopic dynamics

$$\frac{\tau}{\eta}\frac{\mathrm{d}\langle\mathbf{p}\rangle}{\mathrm{d}t} = \sum_{\{\mathbf{b}\}}(\mathbf{b} - \mathbf{p})\phi(\mathbf{b}|\mathbf{p}) . \tag{13}$$

It becomes clear from equation (10) that the condition that ensures the self-consistency of the small fluctuations assumption requires that the eigenvalues of the matrix $\mathbf{A}$ have a negative real part for all values of $\langle\mathbf{p}\rangle$.

The dynamics of the average and covariance of the fluctuations, namely $f_k(t) = \langle\zeta_k\rangle_t$ and $F_{kl}(t) = \langle\zeta_k\zeta_l\rangle_t - \langle\zeta_k\rangle_t\langle\zeta_l\rangle_t$, can be derived from the Fokker–Planck equation also. They are given by

$$\frac{t}{\tau}\frac{\mathrm{d}\mathbf{f}(t)}{\mathrm{d}t} = \mathbf{A}(\langle\mathbf{p}(t)\rangle)\;\mathbf{f}(t) \tag{14}$$

and

$$\frac{t}{\tau}\frac{\mathrm{d}\mathbf{F}(t)}{\mathrm{d}t} = \mathbf{A}(\langle\mathbf{p}(t)\rangle)\;\mathbf{F}(t) + \mathbf{F}(t)\;\mathbf{A}(\langle\mathbf{p}(t)\rangle)^T + \mathbf{B}(\langle\mathbf{p}(t)\rangle). \tag{15}$$

4 Analysis

4.1 Selection

For the analysis the probability of each individual to be generated and selected from the population as the best one $\mathbf{b}$, as it appears in equation (5), has to be specified. For Winner–takes–all selection it is given by [10]

$$\phi(\mathbf{b}|\mathbf{p}) = \upsilon(\mathbf{b})\sum_{n=0}^{N-1}\upsilon(f(\mathbf{x}) > f(\mathbf{b}))^n \times \upsilon(f(\mathbf{x}) \geq f(\mathbf{b}))^{N-1-n}, \tag{16}$$

where the sum runs over the individuals of the population. The term $\upsilon(\mathbf{b})$ reflects the probability of the best individual to be generated from the probability vector, whereas the sum over the remaining probabilities reflects the selection process where the best individual, the individual with smallest evaluation value $f(\mathbf{b})$ within the population, is selected.

Fortunately the single probabilities appearing in equation (16) can be expressed in terms of the components of $\mathbf{p}$. They are given by

$$\upsilon(\mathbf{b}) = \prod_{l=1}^{L} p_l^{b_l}(1 - p_l^{1-b_l}), \tag{17}$$

$$\upsilon(f(\mathbf{x}) > f(\mathbf{b})) = \sum_{\mathbf{x}\in\{0,1\}^L, f(\mathbf{x})>f(\mathbf{b})} \upsilon(\mathbf{x}) \tag{18}$$

and

$$\upsilon(f(\mathbf{x}) \geq f(\mathbf{b})) = \sum_{\mathbf{x}\in\{0,1\}^L, f(\mathbf{x})\geq f(\mathbf{b})} \upsilon(\mathbf{x}). \tag{19}$$

Equation (16) reflects the Winner–takes–all–selection process where the best individual is found in trial $n+1$, after the first n have not been any better then the best one found, and the following $N-1-n$ individuals have at least been as good, but not better then it. From equation (16) it becomes clear that it is selection which will introduce a coupling between the components of the search space into the dynamics of search.

4.2 Macroscopic dynamics

We begin the analysis by calculating the deterministic function (Eq. (13)) for the linear search problem (Eq. (2)). For the moment we fix population size at $N = 2$ and focus on the 2–dimensional examples. Let us write down the probabilities from equation (16) for this case first. They are given by

$$\phi(\{0,0\}|\mathbf{p}) = (1-p_1)(1-p_2) \times (1 + p_1 + p_2 - p_1 p_2), \tag{20}$$

$$\phi(\{0,1\}|\mathbf{p}) = (1-p_1)p_2(p_1+p_2), \tag{21}$$

$$\phi(\{1,0\}|\mathbf{p}) = p_1(1-p_2)(p_1+p_2) \tag{22}$$

and

$$\phi(\{1,1\}|\mathbf{p}) = (p_1 p_2)^2. \tag{23}$$

The deterministic equation for the linear search problem then, for each component, is given by

$$\frac{\tau}{\eta}\frac{\mathrm{d}\langle p_1\rangle}{\mathrm{d}t} = (\phi_{00}+\phi_{01})(0-p_1) + (\phi_{10}+\phi_{11})(1-p_1) \tag{24}$$

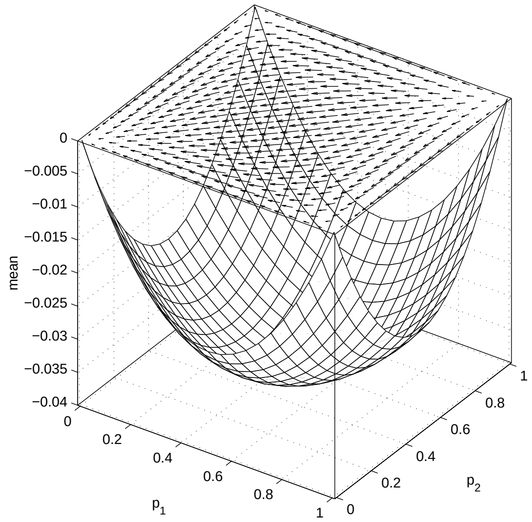

Fig. 3. A vector plot, as indicated by the arrows, and the corresponding hodogram are shown for the macroscopic dynamics of the linear search problem. It can be observed that the macroscopic dynamics indicates convergence towards the global optimum, the corner $\{0,0\}$, since only this corner shows to be an asymptotically stable solution.

and

$$\frac{\tau}{\eta}\frac{\mathrm{d}\langle p_2\rangle}{\mathrm{d}t} = (\phi_{00}+\phi_{10})(0-p_2) + (\phi_{01}+\phi_{11})(1-p_2), \tag{25}$$

with $\phi_{...} = \phi(\{...\}|\mathbf{p})$. The set of equations is plotted in Figure 3. The figure shows a hodogram, which is a plot of the deterministic function *versus* the state variable $\langle\mathbf{p}\rangle$, together with a vector plot of the corresponding dynamics. Although time is not explicitly shown in the hodogram, one knows that $\langle\mathbf{p}\rangle$ has to increase for positive values of $\langle\mathbf{p}\rangle$ and to decrease for negative ones. This flow is indicated by the arrows of the vector plot. Additionally one knows that the stationary solutions of the deterministic function are given by the roots of $\mathrm{d}\langle\mathbf{p}\rangle/\mathrm{d}t = 0$. At the zeros (dots) of the deterministic function both, magnitude and direction vanishes, indicating the fixed points of the deterministic dynamics. It can be seen that there exist 4 fixed points, but only one attracting one, which can be identified as the global optimimum, the corner $\{0,0\}$. It is known that a stochastic process has a high probability to be found in the neighborhood of an attracting fixed point of the deterministic equation. Since the remaining fixed points show to be unstable, we may conclude that the attracting one is unique and that the stochastic process will converge towards the global optimum.

In a next step let us consider the time evolution of the macroscopic variable. It is instructive to rewrite equations (24, 25) in a simpler form, expressing it fully in terms of the components. This yields

$$\frac{\tau}{\eta}\frac{\mathrm{d}\langle p_1\rangle}{\mathrm{d}t} = (-1+p_1)p_1(1-p_2+p_2^2) \tag{26}$$

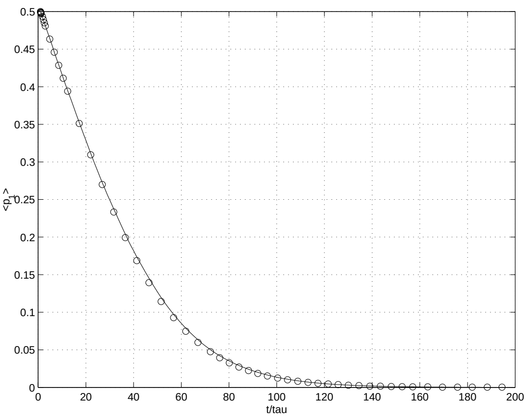

Fig. 4. The solution of the equations for the macroscopic dynamics (o) for the linear search problem, together with an average over 5000 searches (-) are shown for the first component of the probability vector $\mathbf{p}$. The results show to be in good agreement.

and

$$\frac{\tau}{\eta}\frac{\mathrm{d}\langle p_2\rangle}{\mathrm{d}t} = (-1+p_2)p_2(1-p_1+p_1^2). \qquad (27)$$

The solution of the set of coupled and nonlinear differential equations is plotted for the first component in Figure 4, together with a simulated average over 5000 evolution processes. The theoretical and the experimental result show to be in good agreement, although the correspondence is slightly better during the beginning and at the end of search nearer to the global optimum. This finding is a result of the appearance of larger fluctuations in some parts of the search space, causing the macroscopic dynamics to break down (we will discuss this in more detail later).

For the nonlinear search problem (Eq. (3)) the probabilities for receiving an individual for the update are given by

$$\phi(\{0,0\}|\mathbf{p}) = (1-p_1)(1-p_2) \times(1+p_1+p_2-p_1p_2), \qquad (28)$$

$$\phi(\{0,1\}|\mathbf{p}) = (1-p_1)p_2 \times((1-p_1)p_2+2p_1(1-p_2)), \qquad (29)$$

$$\phi(\{1,0\}|\mathbf{p}) = (p_1(1-p_2))^2 \qquad (30)$$

and

$$\phi(\{1,1\}|\mathbf{p}) = p_1p_2(p_1p_2 +2(1-p_1)p_2+2p_1(1-p_2)). \qquad (31)$$

The deterministic equation is constructed in the same way as for the linear search problem and again is plotted in form of a hodogram and vector plot in Figure 5. Clearly one can observe the existence of two stable fixed points of the macroscopic dynamics, namely the corners $\{0,0\}$ and $\{1,1\}$. The remaining corners and the fixed point appearing inside the plane show to be unstable. This result indicates that, regarding the macroscopic dynamics, it depends on initial conditions towards which corner search converges, towards the global or towards the local optimum. However, as we will see while discussing fluctuations, the expansion is only satisfied in the neighborhood of the two attracting fixed points, thus the macroscopic dynamics picture only valid locally.

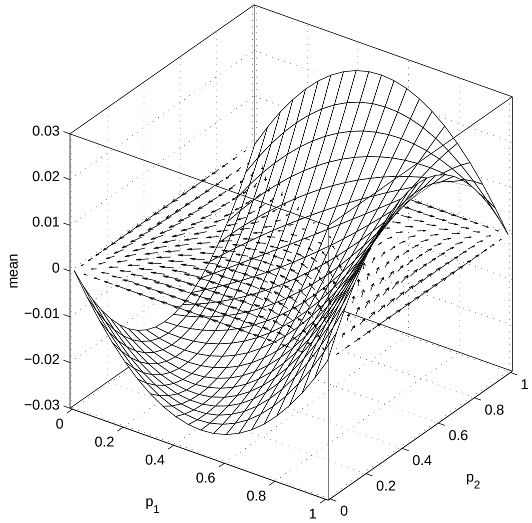

Fig. 5. A vector plot, as indicated by the arrows, and the corresponding hodogram are shown for the macroscopic dynamics of the nonlinear search problem. It can be observed that the macroscopic dynamics indicates convergence towards both optima, towards the corner $\{0,0\}$ (global optimum) and towards the corner $\{1,1\}$ (local optimum). These corners show to be asymptotically stable solutions, while the other corners and the fixed point appearing inside the search space show to be unstable.

4.3 Fluctuations

As previously mentioned, a full understanding of the dynamical properties of the stochastic process cannot be gained without a proper understanding of the properties of its fluctuations. Doing so we analyze the matrix $\mathbf{A}$ as given in equation (11). Taking the corresponding derivatives, we arrive for the linear search problem at

$$\mathbf{A} = \left\{ \begin{array}{ll} (-1+2p_1)(1-p_2+p_2^2) & (-1+p_1)p_1(-1+2p_2) \\ (-1+p_2)p_2(-1+2p_1) & (-1+2p_2)(1-p_1+p_1^2) \end{array} \right\}. \qquad (32)$$

The real eigenvalues of the matrix are plotted in Figure 6. It can be seen that even for the linear search problem there exist areas in the search space where the eigenvalues are well above 0, indicating that fluctuations cause the assumption to break down. In this case however, since

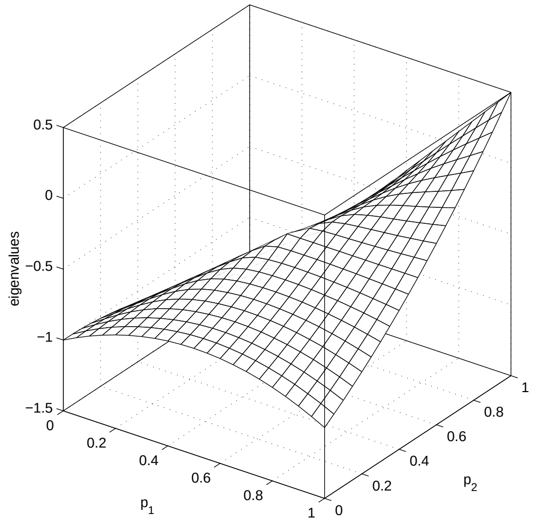

Fig. 6. The eigenvalues of the matrix **A**, justifying the small fluctuations assumption, are shown for the linear search problem. It can be observed that fluctuations dominate (eigenvalues above zero) near the corner $\{1, 1\}$, causing the expansion to be slightly inaccurate globally.

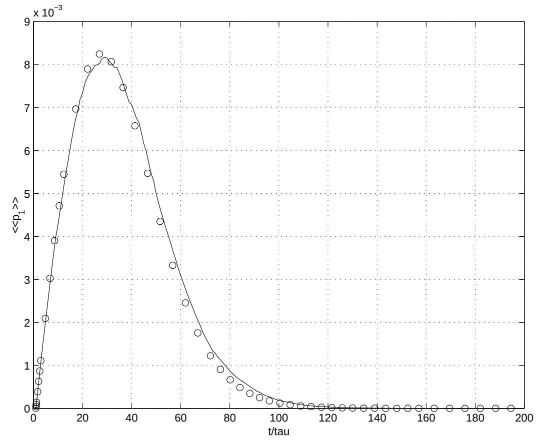

Fig. 7. The solution of the equations for the variance of the fluctuations (trace of the covariance matrix) (o) for the linear search problem, together with the variance obtained from 5000 simulated searches (-) are shown for the first component of the probability vector **p**. The results show to be in good agreement. The processes start with zero variance (all searches have been initialized at $\{0.5, 0.5\}$). It increases during search to decrease to zero variance while search approaches the optimum.

there exists a unique stable fixed point of the macroscopic dynamics, this result means that fluctuations may temporarily increase, but decrease as the process approaches the fixed point. This situation changes dramatically in the case of the nonlinear search problem as we will discuss later.

The time evolution of the fluctuations, in terms of the covariance, is given by equation (15). The solution for the linear search problem (Eq. (2)), in terms of the trace of the matrix **F**, which represents the variance of the process, is plotted in Figure 7, together with the variance from 5000 simulation runs. Again the results show to be in good agreement, although, as in the case of the deterministic function, the agreement is better during the start and as search approaches the optimum. One clearly sees that starting with zero variance (all search paths started at $\{0.5, 0.5\}$), variance temporarily increases, to become zero while search approaches the optimum.

For the nonlinear search problem, we follow the same arguments as for the linear search problem. The matrix **A** from equation (11) now becomes

$$\mathbf{A} = \left\{ \begin{array}{ll} (-1+2p_1)(1-2p_2^2) & 2(1-p_2)p_2 \\ 4(1-p_1)p_1p_2 & -1+2p_1+2p_2-4p_1p_2 \end{array} \right\}. \quad (33)$$

The real eigenvalues are plotted in Figure 8. It becomes clear that vast fluctuations are dominating a large portion of the search space, causing the macroscopic dynamics to break down globally. Furthermore one may conclude that the dependence of PBIL's dynamics from initial conditions is reduced, since fluctuations still allow search to escape a local attraction area.

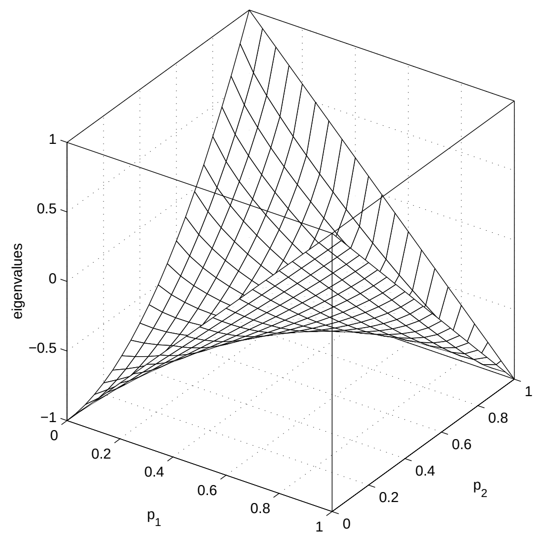

Fig. 8. The eigenvalues of the matrix **A**, justifying the small fluctuations assumption, are shown for the nonlinear search problem. It can be observed that vast fluctuations (eigenvalues well above zero) dominate a large portion of the search space, causing the expansion to break down globally.

4.4 Population size

Let us take a look at the influence of population size. The parameter N determining population size appears in the probability in equation (16). Through equations (10) and (13) it propagates its influence on both, the macroscopic dynamics as well as on the fluctuations.

Let us first study the influence of population size on the macroscopic dynamics (Eq. (13)) for the nonlinear search problem. Vector plots for population sizes from $N = 2, ..., 4$ are shown in Figure 9. It becomes clear that an

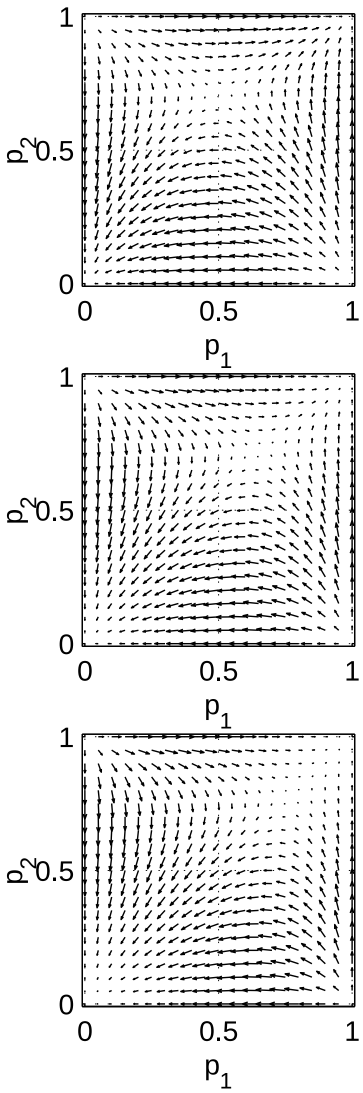

Fig. 9. A vector plot of the macroscopic dynamics is shown for the nonlinear search problem for population sizes $N = 2, ..., 4$ (top to bottom). One can observe that with increasing population size the unstable fixed point inside the search space is shifted towards the local optimum, thus increasing the domain of attraction of the global one.

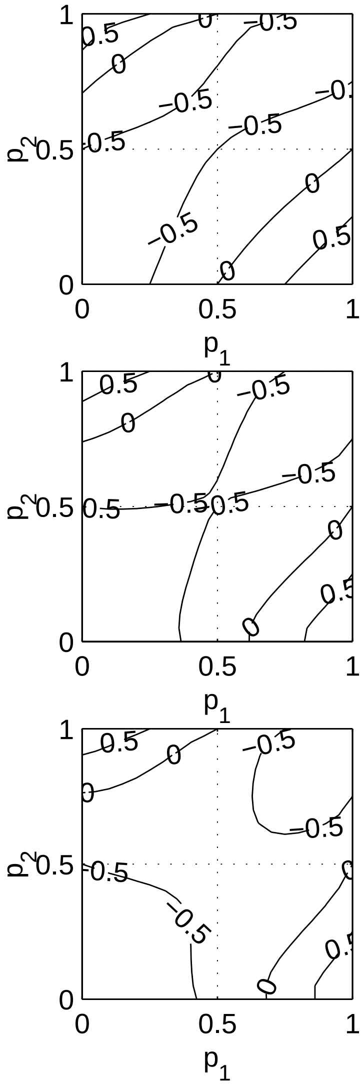

Fig. 10. The eigenvalues of the matrix $\mathbf{A}$ of the nonlinear search problem are shown in form of a contour plot for population sizes $N = 2, ..., 4$ (top to bottom). One can observe that with increasing population size the influence of fluctuations decreases, since the border separating positive from negative values moves further to the corners.

increasing population size causes the unstable fixed point inside the search space to move towards the corner of the local optimum, thus increasing the domain of attraction of the global one. Although the here analyzed population sizes are large compared to the dimension of the search space we can expect this phenomenon to hold also for higher population to search space size ratios. On the other hand, as mentioned, population size does influence the size of fluctuations, too. In Figure 10 the real eigenvalues of the matrix $\mathbf{A}$ of equation (11) for the nonlinear search problem (Eq. (3)) have been plotted for population sizes $N = 2, ..., 4$. One can see that an increase in the population size causes the positive eigenvalues to move in outer regions of the search space, thus indicating a weaker influence of fluctuations inside the plane.

5 Summary

It has been demonstrated that already for the most simple linear search problem, the macroscopic picture of PBIL's dynamics may break down, since fluctuations may temporarily grow in time. Nevertheless, since there only one asymptotically stable fixed point of the macroscopic dynamics exists for this case, PBIL will converge towards this global optimum. Furthermore, it has been shown that already in this simple case the time dependence of the macroscopic dynamics has to be described by a set of coupled and nonlinear differential equations. This effect is due to the influence of selection. Fluctuations are shown to grow temporarily in time, to decrease while search approaches an optimum.

For a nonlinear search function it has been shown that vast fluctuations exist in large portions of the search space, causing the macroscopic picture to be valid locally only. As a consequence, fluctuations allow search to escape the domain of attraction of local optima (as given by the macroscopic dynamics), thus reducing the influence on initial conditions.

Finally, it has been shown how population size determines the dynamics of search. Increasing population size is shown to alter the macroscopic dynamics of search by increasing the field of attraction of the global optimal solution and to reduce the influence of fluctuations.

Although the analysis has been performed for small search space sizes, it has allowed for an intuitive understanding of the search process and revealed several interesting properties. In the context of real world optimization tasks we may confirm that this version of PBIL remains a local search algorithm from the point of view of the macroscopic dynamics. However, since there exist vast fluctuations during middle stages of search, the algorithm shows capable of exploiting a large portion of the search space. When approaching an optimum these fluctuations iteratively decrease. In conclusion, one can expect PBIL to work well when good solutions of the search problem tend to be loosely spread over neighboring corners. In this situation PBIL will be able to approach promising areas of the binary hypercube quickly from inside, while still being capable of exploiting alternative solutions. Finally, from the picture developed here, one should prefer a setting of parameters with moderate learning rates and large population sizes, since this is shown to allow a good exploitation of the search space while at the same time increasing the domain of attraction of the global optimum.

References

1. S. Baluja, R. Caruana, *Proceedings of the International Conference on Machine Learning* (Morgan Kaufmann, San Mateo (CA), 1995), p. 38.
2. D.E. Goldberg, *Genetic Algorithms* (Addison–Wesley, 1989).
3. J.R. Greene, *Proceedings of International Conference on Evolutionary Computation and its Applications* (Russian Academy of Science, 1996), p. 65.
4. E. Galic, M. Hohfeld, *Proceedings of Parallel Problem Solving from Nature* (Springer, Berlin, 1996), p. 740.
5. N.G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier Science Publishers, Amsterdam, 1992).
6. T.M. Heskes, B. Kappen, Phys. Rev. A **44**, 2718 (1991).
7. H. Ritter, K. Schulten, Biol. Cybern. **60**, 59 (1988).
8. P.C. Bresslof, Network J. **6**, 289 (1995).
9. G. Radons, J. Phys. A **26**, 3455 (1993).
10. M. Hohfeld, G. Rudolph, *Proceedings of International Conference on Evolutionary Computation* (IEEE Press, New York 1997) p. 1.
11. D. Bedeux, K. Lakatos–Lindenberg, K.E. Shuler, J. Math. Phys. **12**, 2116 (1971).