8

Energetic Selection Principle

In this chapter[1] I shall introduce a new energetic approach and, based on it, the principle of energetic selection, which can be applied to any population-based optimization algorithm including differential evolution. It consists in both decreasing the population size and the computation efforts according to an energetic barrier function that depends on the number of generations. The value of this function acts as an energetic filter, through which can pass only individuals with lower fitness. Furthermore, this approach allows us to initialize the population of a sufficient (large) size. This method leads to an improvement of algorithm convergence.

8.1 Energetic Approach

Perhaps this new energetic approach may be associated with the processes taking place in physics. As a matter of fact, it was inspired by sociology from a certain sociobiological phenomenon, the so-called phenomenon of dispergated genes, that was observed during World War II. As only a few people know this phenomenon, I prefer to make reference to physics because it is in some sense similar and, in addition, many of the researchers working in evolutionary computation possibly know well a simulated annealing algorithm.

Let there be a population $\mathbb{P}$ consisting of NP individuals. Let us define the *potential* of an individual as its cost function value $\varphi = f(ind)$. Such a potential shows the remoteness from the optimal solution $\varphi^* = f(ind^*)$, that is, some energetic distance (potential) that should be overcome to reach the optimum. Then, the population can be characterized by superior and inferior potentials $\varphi_{\max} = \max f(ind_i)$ and $\varphi_{\min} = \min f(ind_i)$. As the population

[1] Some material in this chapter originally appeared in [FJ04f]; this work was selected as the best paper of ICEIS 2004 to be republished in the book *Enterprise Information Systems VI* [FJ06].

evolves the individuals take more optimal energetic positions, the closest possible to the optimum level. So if $t \to \infty$ then $\varphi_{\max}(t) \to \varphi_{\min}(t) \to \varphi^*$, where t is an elementary evolution step. Approaching the optimum, apart from stagnation cases, can also be expressed by $\varphi_{\max} \to \varphi_{\min}$ or $(\varphi_{\max} - \varphi_{\min}) \to 0$. By introducing the potential difference of population $\triangle\varphi(t) = \varphi_{\max}(t) - \varphi_{\min}(t)$ the theoretical condition of optimality is represented as

$$\triangle\varphi(t) \to 0\,. \tag{8.1}$$

In other words, the optimum is achieved[2] when the potential difference is close to 0 or to some desired precision ε. The value $\triangle\varphi(t)$ is proportional to the algorithmic efforts, which are needed in order to find the optimal solution.

Thus, the *action* A done by the algorithm for passing from one state t_1 to another t_2 is

$$A(t_1, t_2) = \int_{t_1}^{t_2} \triangle\varphi(t)dt\,. \tag{8.2}$$

We introduce then the *potential energy* of population E_p that describes total computational expenses.

$$E_p = \int_0^{\infty} \triangle\varphi(t)dt\,. \tag{8.3}$$

Notice that (8.3) graphically represents the area S_p between two functions $\varphi_{\max}(t)$ and $\varphi_{\min}(t)$.

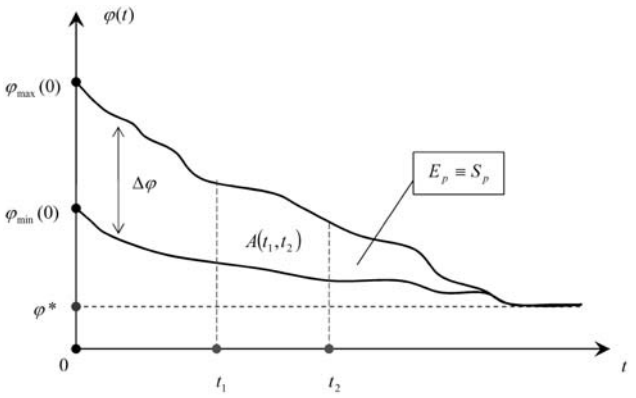

Fig. 8.1. Energetic approach.

[2] Recall that a population based algorithm is usually aimed at the global solution, so the cases where a local optimum is achieved are theoretically excluded from the context.

Let us recall that our purpose is to increase the speed of algorithm convergence. Logically, convergence is proportional to computational efforts. It is obvious that the smaller the potential energy E_p is, the fewer computational efforts are needed. Thus, by decreasing the potential energy $E_p \equiv S_p$ we augment the convergence rate of the algorithm. Hence, the convergence increase is transformed into a problem of potential energy minimization (or S_p minimization).

$$E_p^* = \min_{\triangle\varphi(t)} E_p(\triangle\varphi(t)) \,. \tag{8.4}$$

8.2 Energetic Selection Principle

8.2.1 Idea

Now we apply the above-introduced energetic approach to the DE algorithm. As an elementary evolution step t we choose a generation g.

In order to increase the convergence rate we minimize the potential energy of population E_p (Fig. 8.1). For that a supplementary procedure is introduced at the end of each generation g. The main idea is to replace the superior potential $\varphi_{\max}(g)$ by the so-called *energetic barrier* function $\beta(g)$. Such a function artificially underestimates the potential difference of generation $\triangle\varphi(g)$.

$$\begin{aligned} &\beta(g) - \varphi_{\min}(g) \le \varphi_{\max}(g) - \varphi_{\min}(g) \\ \Leftrightarrow \quad &\beta(g) \le \varphi_{\max}(g), \quad \forall g \in [1, g_{\max}] \,. \end{aligned} \tag{8.5}$$

From an algorithmic point of view this function $\beta(g)$ serves as an *energetic filter* for the individuals passing into the next generation. Thus, only the individuals with potentials less than the current energetic barrier value can participate in the next evolutionary cycle (Fig. 8.2).

In practice, it leads to the decrease of the population size NP by rejecting individuals such that:

$$f(ind) > \beta(g) \,. \tag{8.6}$$

8.2.2 Energetic Barriers

Here, I shall show you some examples of the energetic barrier function. At the beginning we outline the variables upon which this function should depend. First, this is the generation variable g, which provides a passage from one evolutionary cycle to the next. Second, it should be the superior potential $\varphi_{\max}(g)$ that presents the upper bound of the barrier function. And third, it should be the inferior potential $\varphi_{\min}(g)$ giving the lower bound of the barrier function (Fig. 8.3).

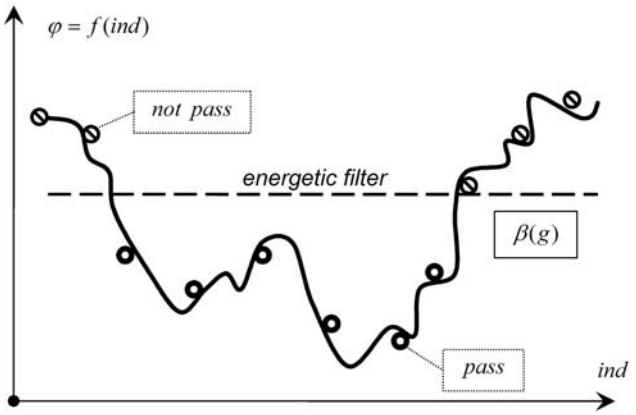

Fig. 8.2. Energetic filter.

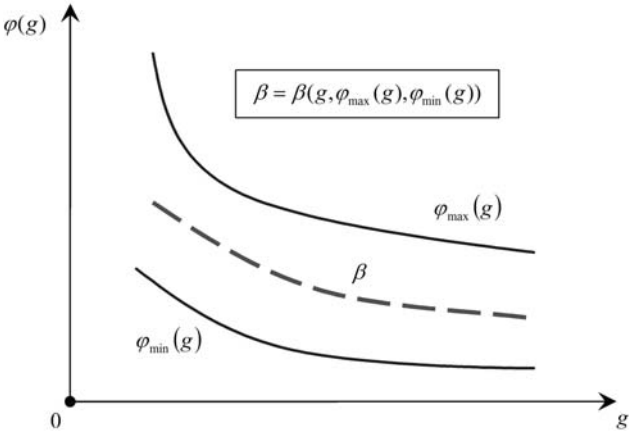

Fig. 8.3. Energetic barrier function.

Linear Energetic Barriers

The simplest example is the use of a proportional function. It is easy to obtain by multiplying either $\varphi_{\min}(g)$ or $\varphi_{\max}(g)$ with a constant K.

In the first case, the value $\varphi_{\min}(g)$ is always stored in the program as the current best value of the cost function. So, the energetic barrier looks like

$$\beta_1(g) = K \cdot \varphi_{\min}(g), \qquad K > 1\,. \tag{8.7}$$

The constant K is selected to satisfy the energetic barrier condition (8.5).

In the second case, a small procedure is necessary to find the superior potential (maximal cost function value of the population) $\varphi_{\max}(g)$. Here, the energetic barrier is

$$\beta_2(g) = K \cdot \varphi_{\max}(g), \qquad K < 1\,. \tag{8.8}$$

K should not be too small in order to provide a smooth decrease of the population size NP.

An advanced example would be a superposition of the potentials.

$$\beta_3(g) = K \cdot \varphi_{\min}(g) + (1 - K) \cdot \varphi_{\max}(g) \tag{8.9}$$

So, with $0 < K < 1$ the energetic barrier function is always found between the potential functions. Now, by adjusting K it is easier to get the smoothed reduction of the population without condition violation (8.5). Examples of the energetic barrier functions are shown in Fig. 8.4.

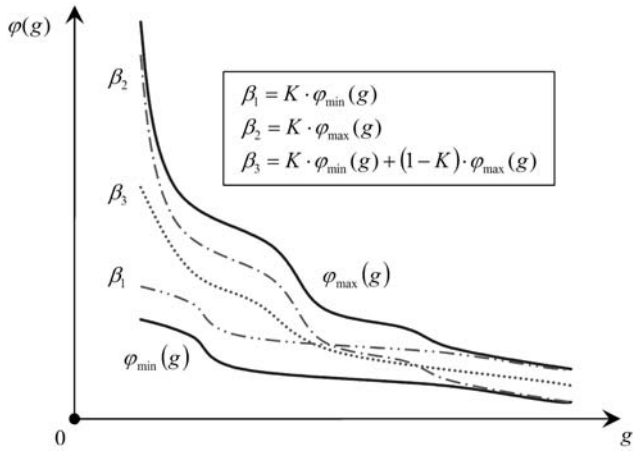

Fig. 8.4. Linear energetic barriers.

Nonlinear Energetic Barriers

As we can see, the main difficulty of using the linear barriers appears when we try to define the barrier function correctly in order to provide a desired dynamics of the population reduction. Taking into consideration that $\varphi_{\max} \to \varphi_{\min}$ when the algorithm converges locally, the ideal choice for the barrier function is a function that begins at a certain value between $\varphi_{\min}(0)$ and $\varphi_{\max}(0)$ and converges to $\varphi_{\max}(g_{\max})$.

Thereto, I propose an exponential function $K(g)$

$$K(g) = K_l + (K_h - K_l) \cdot e^{(-Tg/g_{\max})}\,. \tag{8.10}$$

This function, inspired by the color-temperature dependence from Bernoulli's law, smoothly converges from K_h to K_l. The constant T, so-called *temperature*, controls the convergence rate. The functional dependence on the temperature constant $K(T)$ is represented in Fig. 8.5.

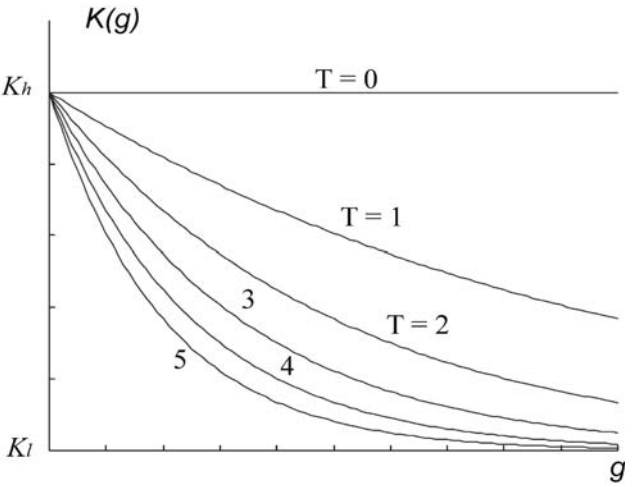

Fig. 8.5. Exponential function $K(g,T)$.

By substituting the constant K in (8.7)–(8.9) for the exponential function (8.10) we can supply the energetic barrier function with improved tuning (Fig. 8.6).

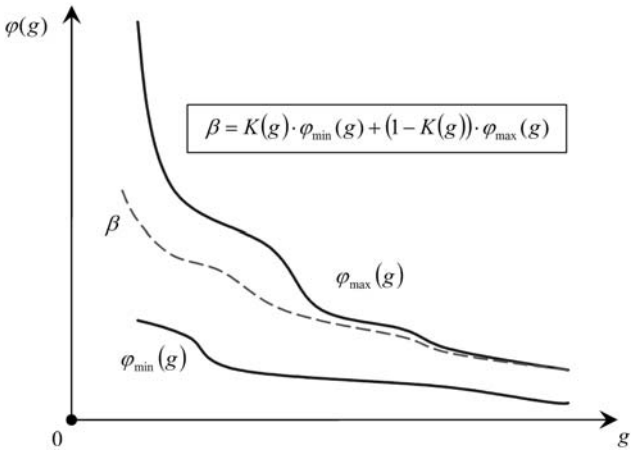

Fig. 8.6. Nonlinear energetic barrier.

8.2.3 Advantages

1. The *principle of energetic selection* permits us to initialize the population of a sufficiently large size. This fact leads to better (careful) exploration

of a search space during the initial generations as well as increasing the probability of finding the global optimum.
2. The *energetic barrier function* decreases the potential energy of the population and thereby increases the convergence.
3. The *double selection principle* is applied. The first one is a usual DE selection for each individual of a population. Here, there is no reduction of the population size. And the second one is a selection of the best individuals that pass in the next generation, according to the energetic barrier function. It leads to the reduction of the population size and consequently the number of function evaluations.

Practical Remarks

Notice that a considerable reduction of the population size occurs at the beginning of the evolutionary process. For more efficient exploitation of this fact a population should be initialized with a much larger size NP_0 than usual. Then, when the population shrinks to a certain size NP_f, it is necessary to stop the energetic selection procedure. This forced stopping is explained by possible stagnation and not so efficient search by a small size population. In fact, the first group of generations locates a set of promising zones. The selected individuals are conserved in order to make a thorough local search in these zones.

8.3 Comparison of Results

In order to test this approach I took three test functions (8.11) from a standard test suite (see Appendix C). The first two functions, Sphere f_1 and Rosenbrock's function f_2, are classical DeJong testbeds [DeJ75]. The third function, rotated ellipsoid f_3, is a quadratic nonseparable function.

$$\begin{aligned} f_1(X) &= \sum_{i=1}^{3} x_i^2 \\ f_2(X) &= 100(x_1^2 - x_2)^2 + (1 - x_1)^2 \\ f_3(X) &= \sum_{i=1}^{20} \left(\sum_{j=1}^{i} x_j \right)^2 . \end{aligned} \tag{8.11}$$

I fixed the differentiation F and crossover Cr constants to be the same for all functions. $F = 0.5$. $Cr = 0$ (there is no crossover in order to make the DE algorithm rotationally invariant; Appendix D). The stopping condition of the algorithm is a desirable precision of optimal solution VTR (*value to reach*). It is fixed for all tests as $VTR = 10^{-6}$. As usual, we count the number of

Table 8.1. Initial test data.

f_i	D	NP	NP_0	NP_f	K
1	3	30	90	25	0.50
2	2	40	120	28	0.75
3	20	200	600	176	0.15

function evaluations NFE needed to reach VTR. The initial data are shown in Table 8.1.

For DE with the energetic selection principle the initial population size was chosen three times larger than in the classical DE scheme: $NP_0 = 3 \cdot NP$. The forced stopping was applied if the current population became smaller than NP. Hence $NP_f \leq NP$. As an energetic barrier function the linear barrier $\beta_3(g)$ was selected (8.9). So, K is an adjusting parameter for barrier tuning, which was found empirically. D is the dimension of the test functions.

The average results of 10 runs for both the classical DE scheme and DE with the energetic selection principle are summarized in Table 8.2.

Table 8.2. Comparison of classical differential evolution (cl) and differential evolution with energetic selection principle (es).

f_i	NFE_{cl}	NFE_{es}	$\delta, \%$
1	1088.7	912.4	16,19
2	1072.9	915.3	14,69
3	106459.8	94955.6	10,81

The numbers of function evaluations (NFEs) were compared. It is considered that the NFE_{cl} value is equal to 100%, therefore the relative convergence amelioration percentagewise can be defined as

$$\delta = 1 - \frac{NFE_{es}}{NFE_{cl}} \,. \tag{8.12}$$

Thus, δ may be interpreted as the improvement of the algorithm's convergence.

Remark

I tested DE with a great range of other functions. The stability of results was observed. So, in order to demonstrate my contribution, here I have generated only 10 populations for each test function relying on statistical correctness.

Problems

8.1. What is the potential of an individual? potential difference? Give an explaining sketch.

8.2. Given a test function, the so-called Schubert's problem,

$$f(X) = \prod_{i=1}^{D} \left(\sum_{j=1}^{5} j \cos((j+1)x_i + j) \right) , \quad -10 \le x_i \le 10 .$$

Plot empirical curves for both superior and inferior potentials, consider one generation (iteration) as an elementary step of evolution. Calculate the action A done by the algorithm for the first 10 and last 10 generations. Estimate the operation efficiency of the algorithm at the beginning and the ending iterations. At which moment is the algorithm most efficient? Explain why.

8.3. Calculate the potential energy of the population. As a basis for this take the curves plotted in problem (8.2).

8.4. How are the potential energy and the algorithm's convergence related?

8.5. What is the energetic barrier? Explain, how does the energetic barrier influence the population?

8.6. On what parameters does the function defining the energetic barrier depend?

8.7. Which of the linear energetic barriers do you think is most efficient from a practical point of view?

8.8. What does the constant K influence?

8.9. In which cases should you use a nonlinear energetic barrier?

8.10. What is the constant T in (8.10) of Chapter 8 and what does it influence? Using the potential's curves from problem (8.2), plot functions of nonlinear energetic barriers for the constant $T = 0, 1, 3, 5$.

8.11. Solve, for example, the following test function, the so-called McCormick's problem,

$$f(X) = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1$$
$$-1.5 \le x_1 \le 4 , \quad -3 \le x_2 \le 3 ,$$

using linear and nonlinear energetic barriers. Compare the obtained results.

8.12. Implement "forced stopping" of the energetic selection procedure for problem (8.11).

8.13. What are the advantages of using the method of energetic selection? For what kind of problems (test functions) is this method more appropriate? Argue your suppositions.

8.14. for determining the promising zones using the population state. Implement it in your DE algorithm.

8.15. For the algorithm realized in problem (8.14) develop a technique which permits individuals to rapidly migrate from less promising to more promising zones. Estimate the efficiency of your algorithm on 2–3 multimodal test functions at your discretion.