# How does the Landscape of a Hopfield Associative Memory Look Like?

Akira Imada[1] and Keijiro Araki[2]

[1] Graduate School of Information Science
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, 630-01 Japan
akira-i@is.aist-nara.ac.jp

[2] Department of Computer Science and Computer Engineering
Graduate School of Information Science and Electrical Engineering
Kyusyu University
6-1 Kasuga-koen, Kasuga, Fukuoka, 816 Japan
araki@c.csce.kyusyu-u.ac.jp

**Abstract.** We apply evolutionary computations to Hopfield's neural network model of associative memory. In the model, some of the appropriate configurations of the synaptic weights give the network a function of associative memory. One of our goals is to obtain the distribution of these configurations in the synaptic weight space by exploring a fitness landscape defined on the space. In this paper, we use three types of walk on the fitness landscapes to have an insight into the geometry of it.

## 1 INTRODUCTION

Associative memory is a dynamical system which has a number of stable states with a domain of attraction around them [1]. If the system starts at any state in the domain, it will converge to the stable state. In 1982, Hopfield [2] proposed a fully connected neural network model of associative memory in which information is stored by being distributed among neurons and retrieved by being relaxed dynamically. The dynamical behavior of its neuron states strongly depends on synaptic strength between neurons. The synaptic strength from neuron $j$ to neuron $i$ is called *weight* and denoted as $w_{ij}$ in this paper. Hopfield used the Hebbian rule [3] to determine the synaptic weights, whilst we specify them with using evolutionary computations (see, e.g., [4]).

For the fully connected neural networks with $N$ neurons, all the possible combinations of $w_{ij}$ construct $N^2$-dimensional Euclidean space. Hereafter, we call this (synaptic) weight space. Every configuration of $w_{ij}$ is represented as a point in the weight space. Fitness values of the points are estimated according to the capability of the network that is determined by each configuration. Therefore, the task of our evolutionary computations is to locate the points of the highest fitness value. Our previous results show that the task to locate one of the optimum solutions becomes difficult as the number of patterns to be stored

increases. We conjecture that this phenomenon is due to the increased ruggedness of the fitness landscape.

This paper is a step towards analyzing statistically the fitness landscape of the Hopfield model defined on the synaptic weight space.

## 2 RELATED WORKS

Statistical properties of the landscapes defined by Kauffman's NK-model [5], Sherington & Kirkpatric's S-K model [6], or Derrida's random energy model [7] has been extensively studied as general cases (see, e.g., [8] and the references quoted therein). Or the landscapes, in particular, constructed as, e.g., the free energies of RNA folding into secondary structures [9]; or the chemical affinity of the antibody for the antigen in immune response [10] were analyzed statistically. They estimated, for example, the number of local minimum by sampling points along an evolutionary path (adaptive random walk) with using a concept such as *autocorrelation function* and/or *correlation length.*

The Hamiltonian energy of the Hopfield model also constructs fitness landscape on the configuration space of neuron states. When synaptic weights are selected randomly, the netwarks can be considered as the S-K model. Meta-stable states of the S-K model were analytically studied by Bray et al. [11], using correlations between the Hamiltonian energies of the meta-stable states. Thus they obtained the distribution of meta-stable states as a function of free energy. Or Derrida et al. [7] exactly solved the number of meta-stable states in one dimensional spin-glass chain, answering the question: if a spin configuration evolves in time, what stable state will it fall into?

All these landscapes were defined on the *discrete* configuration space of strings comprised of spin glasses, RNA molecules, whatsoever. On the other hand, we are exploring the fitness landscape defined on the *continuous* synaptic weight space, following Gardner [12] who discussed the *volume* of these solutions in the weight space. However, the number and distribution of the solution in the space are still open problem.

## 3 EXPERIMENT

The Hopfield model consists of $N$ neurons and $N^2$ synapses in which each neuron takes one of two states $\pm 1$. Then patterns

$$\xi^{\nu} = (\xi_1^{\nu}, \cdots, \xi_N^{\nu}), \qquad \nu = 1, \cdots, p$$

where $\xi_i^{\nu}$ takes the value of either 1 or $-1$, are to be memorized as equilibrium states. After one of these patterns is given to the network, neuron states are updated asynchronously, i.e., at most one neuron changes its state at a time, according to

$$s_i\,(t+1) = sgn\left(\sum_{j\neq i}^{N} w_{ij} s_j(t)\right),$$

where $s_i(t)$ is a state of $i$-th neuron at time $t$. The behaviors of the collective states of individual neurons are characterized by synaptic weights. Provided the synaptic weights are determined appropriately, networks store some number of patterns.

When one of stored patterns $\xi^\nu$ is given to the network as an initial state, possibly including a few errors, the state of neurons varies from time to time afterwards. In order for the network to function as an associative memory, these two states must be similar. The similarity as a function of time is defined by,

$$m^\nu(t) = \frac{1}{N}\sum_{i=1}^{N} \xi_i^\nu s_i^\nu(t).$$

This $m^\mu(t)$ is calculated for each stored pattern and averaged. We denote this as $\langle m(t)\rangle$. Then we define fitness value as a temporal average of this $\langle m(t)\rangle$. That is, the fitness value $f$ is

$$f = \frac{1}{t_0 \cdot p}\sum_{t=1}^{t_0}\sum_{\nu=1}^{p} m^\nu(t).$$

In this paper, $t_0$ is set to $2N$, twice the number of neurons. Note that the fitness 1 implies all the $p$ patterns are stored as fixed points, while fitness less than 1 includes many other cases.

To search configuration of $w_{ij}$ so that the network has higher fitness value, we used the following three schemes here: random walk, simple hill climbing and evolutionary programming (EP) [13]. In these schemes, we use the Gaussian random mutation to move a point on the weight space. To be more specific, each coordinate of a point in $N^2$-dimensional weight space, $x_i$, is given small perturbation using the Gaussian random variables with mean 0 and standard deviation $\sigma_i$, i.e.,

$$x_i^{\text{new}} = x_i^{\text{old}} + \sigma_i \cdot N_i(1,0).$$

*Random walk* uses this strategy to move a point to the next position. In *hill climbing*, on the other hand, the mutation is given $\lambda$ times to the current point. Then a point which obtains the best fitness value among the mutants is selected as the next point. In *EP*, $\mu$ points construct a population. Each of these $\mu$ points is mutated, which produces $\mu$ mutants. Each point has additional $N^2$ information $\sigma_i$ $(i = i, 2, \cdots, N^2)$, besides its coordinates. They are used as a standard deviation of a Gaussian random variable to be added to $i$-th coordinate of the point. After the mutation, these $\sigma_i$ are also modified as:

$$\sigma_i^{\text{new}} = \sigma_i^{\text{old}} + 0.01 \cdot \sigma_i^{\text{old}} \cdot N_i(1,0).$$

Now, we have the original $\mu$ points and their $\mu$ mutants. The fitness value of each of the $2\mu$ points are compared to those of $q$ points which are chosen randomly at every time of the comparison. Then the $2\mu$ points are ranked according

to the number of wins, and the best $\mu$ points survive ($q$-tournament selection). The cycle of reconstructing the new population with better fitness points and restarting the search is repeated until one of the global optima is found or a maximum number of generation (12000 here) has been reached.

Thus, the fitness landscapes defined on the weight space, are explored in the next section.

All the experiments in this paper were made on networks with 49 ($=N$) neurons.

# 4 RESULTS AND DISCUSSION

## 4.1 Down Hill Walk from a Peak

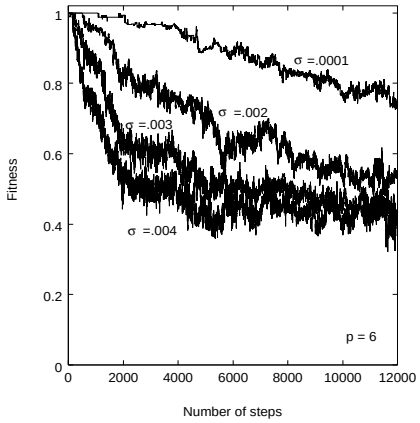


**Fig. 1. Fitness along a down hill walk**

In EP described above, all $\sigma_i$ of each point are *initialized* to $\sigma_0$. The tuning of this $\sigma_0$ is somewhat a difficult issue. To determine the appropriate value of $\sigma_0$, we experimented *down hill walk* from one of the optima in the landscape. The hill we walk down here is by the Hebbian $w_{ij}$, i.e.,

$$w_{ij} = \frac{1}{N}\sum_{\nu=1}^{p} \xi_i^\nu \xi_j^\nu \quad (i \neq j), \quad w_{ii} = 0.$$

The results are shown in Fig.1. The path with $\sigma_0 = 0.0001$ walks down the hill slowly, while the path with $\sigma_0 = 0.004$ makes it very quickly.

It is interesting to note that a path along this *down hill walk from a peak* could be used to study the ruggedness of the fitness landscape statistically, since it always includes one optimum.

### 4.2 Random Walk

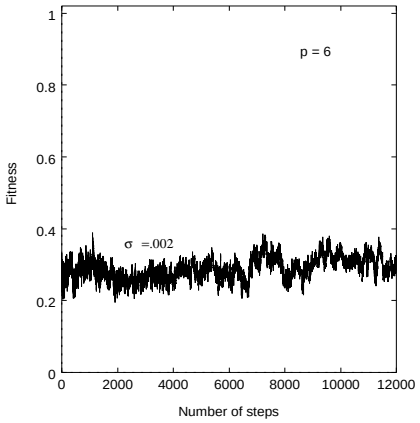


**Fig. 2. Fitness along a random walk**

With the value of standard deviation 0.02 for the Gaussian mutation, we experimented *random walk* through the landscape. An example is shown in Fig.2. We can see that the random walk does not approach to a peak. This suggests that the number of peaks in the landscape for $p = 6$ is infinitely small compared to the whole number of points in the space[3], and the basin of attraction of the peaks is probably very narrow. In fact, if we pick up the same number of points uniformly at random from the space, the distribution of the fitness values was very similar to a Gaussian distribution with a small variance, as shown in Fig.3. Hence in our problem, it is inappropriate to use points taken from the space at random or along a random walk for the study of fitness landscapes.

### 4.3 Hill Climbing

Then, how about a hill climbing with the Gaussian random mutations? Davis [14] reported that some problems were easier to solve by a hill climbing than by genetic algorithms. We experimented it with the standard deviation $\sigma_0 = 0.002$ and the number of mutants $\lambda = 200$. An example are shown in Fig.4. This time it seems to climb a hill, probably local, but the speed is extremely slow.

Whitley [15] claimed that test functions for evolutionary computations should be resistant to random search or simple hill climbing. In this sense, we might say our current problem of the Hopfield model could be a good test function for evolutionary computations.

[3] In the current problem, even if we restricted each weight value to the floating-point number from $\{-1.00, -0.99, -0.98, \cdots, 1.00\}$ the weight space is constructed with $201^N$ points.

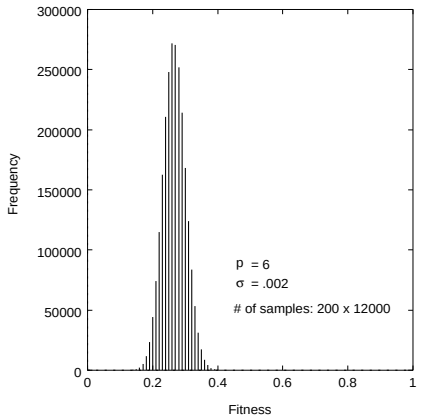


**Fig. 3. Fitness distribution**

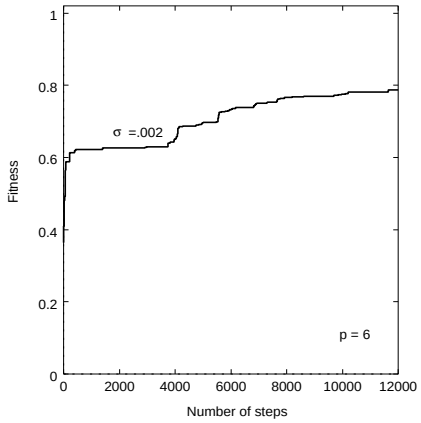


**Fig. 4. Fitness along a hill climbing**

### 4.4 Evolutionary Programming

Next, we experimented using EP with the initial standard deviation 0.002. The parameters $\mu$, $\lambda$ and $q$ were set to 200, 200 and 10, respectively. The coordinates of $\mu$ points are all initialized to zero. This means that all searches are started at the origin of the space. We repeated the EP run 30 times with different random number seed, and found that the two out of 30 runs succeeded in locating one of the optima. The example of the fastest convergence of the two are shown in Fig.5. Statistics of the other runs are shown in Table 1 which we describe later.

Then we proceed to the experiments with varying $\sigma_0$. A series of these experi-

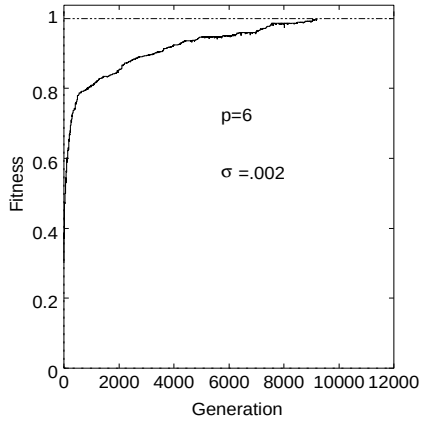


**Fig. 5. Fitness of the Best of Generation**

ments with varying $\sigma_0$ gives us information of fitness landscapes by telling us, e.g., the global optima are distributed uniformly from the region near the origin to the region far beyond. We should note that the schemes using recombination would not be used for this purpose due to the large step size by crossover operation. Only the *mutation and selection scheme* like EP allows us this kind of analysis.

Since we start the search at the origin of the weight space, the way a point explores the space can be observed in the time evolution of the distance of the point from the origin of the space. Distances of the optima from the origin found by EP are crucially dependent on the initial standard deviation $\sigma_0$. We studied this with the initial standard deviation varying from 0.0001 to 0.004, and show some results in Fig. 6. Each EP path shown in the figure eventually found the optimum which is located at certain distance from the origin. We can see that a path with larger $\sigma_0$ tends to explore further away from the origin and exploit the region.

It is interesting to notice that all the graphs have a similar profile. Although the one with $\sigma_0 = 0.0001$ looks different, the extended graph, as shown in the figure, is also similar to others.

We compared shapes of the graphs between random walk and EP runs. In Fig.7, we show a typical example of random walk with the Gaussian mutation of standard deviation 0.002 and of EP run with the initial standard deviation of 0.002. We can observe the significant difference in shape. This suggested that the EP is essentially different from random searches or simple hill climbing. This difference is due to the adaptive mechanism of Gaussian mutation of EP. The shape of the graphs reveals that EP suppresses the step sizes especially in early stages of evolution, though this is opposite to the usual adaptations which have been reported so far. And the EP path, with this size of stored patterns, keeps

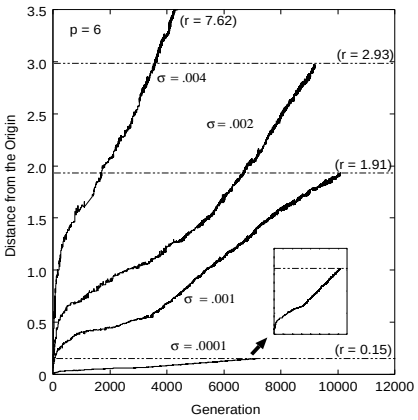


**Fig. 6. Temporal evolution of Distance from the origin**

**Table 1. Statistics of the best fitness in each generation and convergence speed (taken from successful runs)**

| number of patterns | avg of best-fit | std-dev of best-fit | fastest convergence (generation) | number of success |
|---|---|---|---|---|
| 0.0001 | .821935 | .067136 | 7171 | 3 |
| 0.001 | .984891 | .001642 | 10096 | 1 |
| 0.002 | .985462 | .001480 | 9199 | 2 |
| 0.004 | .987576 | .002138 | 10096 | 1 |

exploring until it locates one of the global optima without being trapped at a local optimum.

In contrast, EP with more patterns or simple hill climbing behaves differently. The pass is easily trapped in a local optimum. We show an example in Fig.8. A graph of EP run for $p = 12$ and hill-climbing for $p = 6$ indicate that they were trapped at a local optimum, while EP for $p = 6$ was not. The latter keeps searching *adaptively* even if it was unable to reach one of the global optima within 12000 generations.

In this sense, we can roughly estimate the density of local optima from the shape of graphs. All the other graphs resulted from the EP runs for $p = 6$, though not shown in the figure, have more or less a similar shape. And all the graphs resulted from the runs for $p = 12$ have another similar shape which is different from the shape for $p = 6$. These observations suggest us that the landscapes for

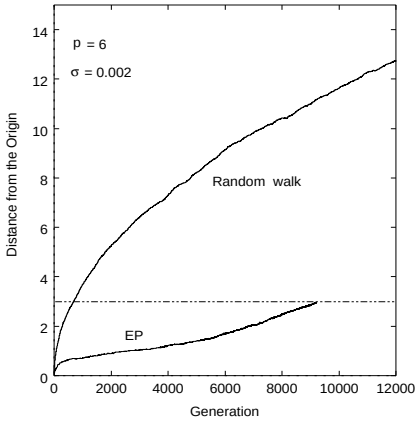


**Fig. 7. Temporal evolution of Distance from the origin**

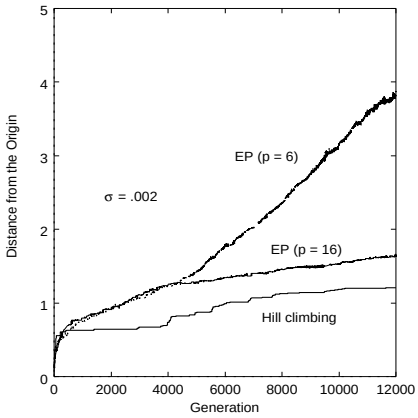


**Fig. 8. Temporal evolution of Distance from the origin**

$p = 6$ have some local optimum, since a simple hill-climber is easily trapped by them, but the number is much fewer than the ones for $p = 12$.

The way EP adapts the step size were also studied on the experiment with $p = 1$. All the runs in this case converge very rapidly. The statistics out of 30 runs in which all of them succeeded in locating one of the *global* optima. As shown in Fig.8, an evolutionary path which has a relatively large initial standard deviation $\sigma_0$ is also able to find one of the optima nearby the origin. The density of the global optima is much larger in $p = 1$ than $p = 6$. Therefore a path with

large $\sigma_0$ will quickly trapped at one of the *global* optima rather than keeping exploration further. We repeated EP run 30 times with $\sigma_0$ varying from 0.0001 to 0.008 for a fixed pattern. Statistics are shown in Table 2. This also suggests that the distribution is uniform since if it is not the case some of the runs will not converge.

**Table 2. Statistics of the number of steps to global optimum $s$ and the distance form the origin $r$ (out of 30 runs)**

| $\sigma_0$ | $s_{max}$ | $s_{avg}$ | $s_{min}$ | $s_{std}$ | $r_{max}$ | $r_{avg}$ | $r_{min}$ | $r_{std}$ | $n_{suc}$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.0001 | 419 | 146.1 | 59 | 12.4 | .045672 | .024304 | .018944 | .001009 | 30 |
| 0.001 | 540 | 157.8 | 68 | 18.0 | .878517 | .269580 | .175213 | .024233 | 30 |
| 0.002 | 540 | 161.1 | 68 | 18.0 | .979707 | .502780 | .350426 | .025278 | 30 |
| 0.004 | 540 | 160.9 | 68 | 18.4 | 1.959415 | 1.004425 | .700852 | .050692 | 30 |
| 0.008 | 540 | 157.8 | 49 | 18.4 | 3.918830 | 1.985039 | 1.284302 | .104130 | 30 |

Those experiments described above give us a rough image of the fitness landscape. Solutions in weight space are distributed uniformly and isotropically around the origin. Since EP paths with different random number seed is most provable to start the move towards the different directions. The basin of attraction of the global optima is larger than for $p = 6$, because we observe that the paths with large step size (e.g. 0.08) are also trapped at the global optima closer to the origin than the case for $p = 6$.

Weinberger [8] wrote that mean walk length of a point along a hill climbing path from a random point to a *local* optimum reflects the density of *local* optima. All the successful results described in this paper, however, are along a path from the origin to one of the *global* optima. Therefore, the mean walk length can be approximated by the distance from the origin, and it reflects the density of the *global* optima around the origin.

The rate of successful runs also reflects the density of the local optima. Hence, the comparison of the experiments for $p = 1$ and 6 gives us a suggestion that the density of the local optima around the origin is much higher in $p = 6$ than in $p = 1$.

Finally, we show the distribution of the distances from the origin of located optima for $p = 1$ in Fig. 9. We can observe that the density of the global optima is strongly dependent on the initial standard deviation $\sigma_0$.

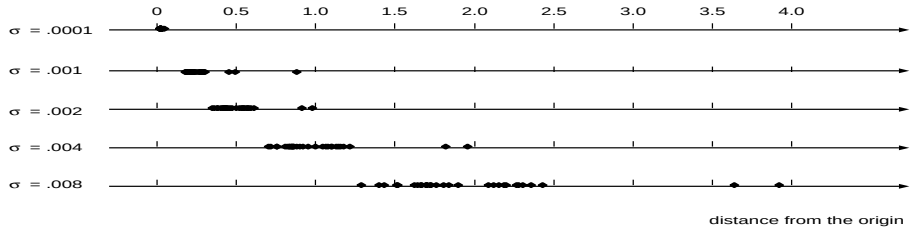


**Fig. 9. Distribution of the Euclidean distances from the origin of weight space of the located optima (p=1)**

## 5 CONCLUSION

We have analyzed evolutionary paths through the fitness landscape defined on weight space of fully connected neural networks. The paths starting at the origin of the space give us the following rough image of the landscapes. The global optima are distributed isotropically and uniformly around the origin; the number of the global optima is maximum in $p = 1$, then decreases as $p$ increases: the landscape with $p = 1$ has only a few local optima; the basin size of the globals optima is very narrow for $p = 6$; and so on.

The next step will be to analyze the paths using statistical properties such as *autocorrelation function* and *correlation length* of the points along a EP path.

## Acknowledgments

We thank Peter Davis at Advanced Telecommunication Research Institute (ATR) for providing us great insight into the dynamics of Hopfield neural networks.

## References

1. J. Komlós, and R. Paturi (1988) *"Convergence Results in an Associative Memory Model."* Neural Networks 1, pp239–250.
2. J. J. Hopfield (1982) *"Neural Networks and Physical Systems with Emergent Collective Computational Abilities."*, Proceedings of the National Academy of Sciences, USA, 79, pp2554-2558.
3. D. O. Hebb (1949) *The Organization of Behavior.*, Wiley.
4. A. Imada, and K. Araki (1997) *"Evolution of Hopfield Model of Associative Memory by the Breeder Genetic Algorithm."* Proceedings of the 7th International Conference on Genetic Algorithms, pp784–791.
5. S. A. Kauffman, and S. Levin (1987) *"Towards a general theory of adaptive walks on rugged landscapes."* Journal of Theoretical Biology, 128, pp11–45.
6. S. Kirkpatrick, and D. Sherrington (1978) *"Infinite-ranged models of Spin-glasses."* Physical Review, B17, pp4384-4403.

7. B. Derrida, and E. Gardner (1986) *"Metastable states of a Spin Glass Chain at 0 Temperature."* Journal of Physique 47, pp959–965.
8. E.Weinberger, and P.Stadler (1933) *"Why some fitness landscapes are fractal."* Journal of Theoretical Biology, 163, pp 255–275.
9. W. Fontana, and P. Schuster (1987) *"A Computer Model of Evolutionary Optimization."* Biophysical Chemistry, 26, pp123–147.
10. C. Macken, P. Hagan, and A. Perelson (1991) *"Evolutionary walks on rugged landscapes."* SIAM Journal of Applied Mathematics, 51, pp799–827.
11. A. Bray, and M. Moore (1980) *"Netastable states in spin glasses."* Journal of Physics, C: Solid State Physics, 13, ppL469–L476.
12. E. Gardner (1988) *"The Phase Space of Interactions in Neural Network Models."* J. Phys 21A, pp257-270.
13. D. B. Fogel (1992) *"An Analysis of Evolutionary Programming."* Proceedings of 1st Annual Conference on Evolutionary Programming, pp43–51.
14. L. Davis (1991) *Bit-Climbing, Representational Bias, and Test Suite Design.* Proceedings of the 4th International Conference on Genetic Algorithms, 18.
15. D. Whitley, K. Mathias, S. Rana, and J. Dzubera (1995) *Building Better Test Functions.* Proceedings of the 6th International Conference on Genetic Algorithms. pp239-247.

This article was processed using the LaTeX macro package with LLNCS style