ANALIZA SYSTEMOWA W FINANSACH I ZARZĄDZANIU

Wybrane problemy
Tom 2

Pod redakcją

Macieja KRAWCZAKA i Jerzego HOŁUBCA

Warszawa 2000

WYŻSZA SZKOŁA
INFORMATYKI STOSOWANEJ
I ZARZĄDZANIA

ANALIZA SYSTEMOWA
W FINANSACH I ZARZĄDZANIU

Wybrane problemy
Tom 2

Pod redakcją
Macieja KRAWCZAKA i Jerzego HOŁUBCA

Warszawa 2000

Wykaz opiniodawców artykułów zamieszczonych w tomie:

doc dr hab. Dariusz **GĄTAREK**

prof. dr hab. Jakub **GUTENBAUM**

prof. dr hab. Jerzy **HOŁUBIEC**

doc. dr hab. Marek **LIBURA**

prof. dr hab. Stanisław **PIASECKI**

prof. dr hab. Andrzej **STRASZAK**

doc. dr hab. Sławomir **WIERZCHOŃ**

doc dr. hab. Leszek **ZAREMBA**

IMPACT OF INITIALIZATION METHODS
OF START POPULATION ON CONVERGENCE
OF EVOLUTIONARY STRATEGIES

Marcin Słomka

Extra-mural Doctoral Course SRI PAS

Considering basic methods of generating pseudo random numbers I'm going to focus on linear congruential generators and low discrepancy sequences in order to analyse its application in problem of creating a start population of evolutionary strategies (ES). The main goal of this article is to analyse the impact of initial phase of evolutionary strategies on speed of convergence of optimization process.

1. Introduction

Genetic Algorithms(GA) [4][5][11]and Evolutionary Strategies [5][9][10] are currently the main representatives of the class of algorithms inspired by principles of biological evolution. Growing popularity of this class of algorithms arises from its simplicity and effectiveness when applied to a global optimization problem. The basic principles of evolution as search heuristic can be summarised as follows: the search points or candidate solutions are interpreted as individuals. Since there is a population of individuals, evolution is a multi-point search. The optimization criterion has to be one dimensional and is called the fitness of individual. Constraints can be embedded in the fitness function as a additional penalty terms. New candidate solutions called offsprings are created from members of current population (parents). The most popular operators are mutation and crossing. Offspring produced by mutation has almost all features of its parent but small variations whereas crossing produces an individual with mixed properties of its parents. The weakest individuals of population are replaced with the best (e.g. with the highest fitness) offsprings. Vast majority of articles concerning ES focus exclusively on ways of improvement of optimization by introducing better and more intelligent genetic operators [1][5]. Here I try to stress the importance of initial stage of algorithm. Logical premises indicate that populations with properly chosen predecessors with diversified genotypes will have better adaptation abilities

than populations with narrower stakes of genetic material. The rest of the article exclusively deals with ES since it is proved that ES are better optimizers then GA in a subclass of problems with continuous search space though drawn conclusions should also be valid for GA.

2. Evolutionary Strategies

Every heuristic method for searching global optimum of difficult optimization problems has to handle the dilemma between exploration and exploitation. The expected behaviour of ideal algorithm would be to explore the search space at the beginning and then having found a promising area with potential solutions exploit it till given level of accuracy is achieved. Higher priority of exploitation bears danger of getting stuck in a poor local minima whereas dominance of explorative features takes too much computational time and is too greedy with respect to every resource of computer systems. Unfortunately the balance between those two factors if very difficult to keep. Evolutionary algorithms avoid local minima by paralleling the search for solution space. Explorative character is assured by a cross-over operator. Exploitation is achieved by selection of fitter individuals and a mutation parameter which is here a part of chromosome submitted to evolutionary process. The ES algorithm can be described as follows:

```
begin
  t = 0
  initialise Pop(t)
  evaluate Pop(t)
  while (not termination_condition) do
    begin
      Offs(t) = reproduce Pop(t)
      Cross Offs(t)
      Mutate Offs(t)
      Pop(t+1) = select_best(Pop(t) + Offs(t))
      t = t+1
    end
end
```

Algorithm begins with a initialisation phase where a start population (of size μ) of randomly chosen solutions is generated. Evaluation is a process of calculating fitness functions for each individual in the population.

New population of size λ is created during reproduction phase. Individuals of a new created population of offspring are submitted to cross-over and mutation and finally a set of best μ individuals are selected from a joined populations of parents and offspring. This algorithm is an elitary model because fitness function of the best individual in the next generation is never worse then a fitness of the best from previous generation. Since the most important operators of this algorithm are cross-over and mutation little if any attention was paid on a initialisation phase. It turns out that well chosen method of initialisation implies significant improvement of convergence. The next section presents the most popular methods of generating random numbers and explains the underlying idea of so-called low discrepancy sequences.

3. Random and quasi-random numbers

Researchers often face a problem of simulating some physical process described by stochastic variables with different probability distributions. To complete experiment of this kind a generator of "randomness" is necessary. Obviously such generators are extensively used in optimization with ES as well. They drive a process of "noising" a parameters of mutation (based on Brownian motion) and help to implement the cross-over operator. The most popular generators are described below.

3. 1 Linear congruential generators
The most widely known random number generators are *linear congruential generators*. They work according to the recurrence formula:

$$X_{n+1} = (aX_n + c) \bmod m$$

Here m is called the *modulus*, and a and c are positive integers called the *multiplier* and the *increment* respectively. If m, a and c are properly chosen (see Table 1), then the period will be of maximal length, i.e., of length m. In that case, all possible integers between 0 and $m - 1$ occur at some point regardless of value of initial seed. Majority of the random number generators in available software is of this kind but unfortunately few of them have properly chosen modulus, multiplier and increment. To have a LCG with maximal period - m, a and c have to satisfy a few conditions. For more details concerning the choice of these three parameters see [6] [8]. Linear congruential generator considered in experiment was implemented with

following values $a = 1103515245$, $c = 12345$, $m = 2^{32}$ (default generator provided with Borland C++ 3.1 library).

a	c	m
$2^2 * 2^{37} + 1$	0	2^{35}
69069	1	2^{32}
16807	0	$2^{31} - 1$
630360016	0	$2^{31} - 1$
410092949	0	2^{32}
742938285	0	$2^{31} - 1$
40692	0	$2^{31} - 249$
1099087573	0	2^{32}

Table 1: Well picked parameters for LCG generator
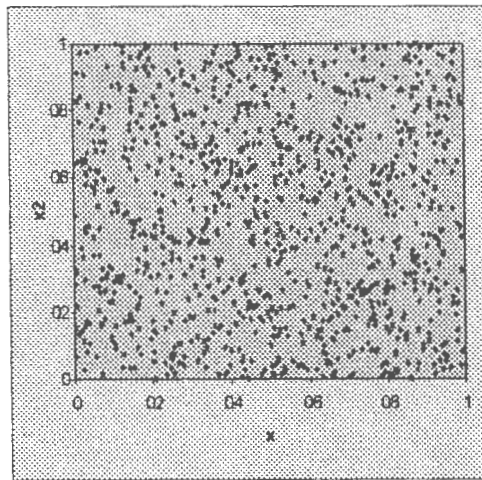$a = 1103515245$, $c = 12345$, $m = 2^{32}$

Figure 1: Two dimensional projection of random sequence
generated by LCG $a = 1103515245, c = 12345, m = 2^{32}$

3. 2 Low discrepancy sequences
For some complex problems performance of the basic Monte Carlo approach may be rather unsatisfactory because the error is $O(\frac{1}{\sqrt{N}})$, where

211

N is a size of sample. We can sometimes improve convergence by using preselected deterministic points to evaluate the integral. The accuracy of this approach depends on the extent of how evenly the deterministic points are dispersed throughout the domain of integration. Discrepancy measures the extent to which points are evenly dispersed across given region: the more evenly the points are dispersed the lower discrepancy. Low discrepancy sequences are often called "quasi - random" . The formal definition of LDS [2] [3] [8][12] is as follows:

> *For a given dimension d sequence of points*
> $\{x_0, x_1, x_2, ...\}$ *is called LDS in I^d hipercube when*
> *following criterion is satisfied*

$$D_N{}^*(x_1, x_2, ..., x_N) = O\left(\frac{(\log N)^d}{N}\right) \quad,$$

where D_N^* is a measure of uniform distribution of points defined as follows:

$$D_N^*(x_1, x_2, ..., x_N) = \sup_{\substack{0 \le v_i \le 1 \\ i=1,...,d}} \left| \frac{\left|\{x_1, x_2, ..., x_N\} \bigcap \prod_{i=1}^{d}[0, v_i)\right|}{N} - \prod_{i=1}^{d} v_i \right|.$$

One of the crucial theorems concerning LDS gives the estimation of upper bounds of D_N^* statistics:

$$D_N^*(x_1, x_2, ..., x_N) \le c_d \frac{(\ln N)^d}{N} + O\left(\frac{(\ln N)^{d-1}}{N}\right)$$

where c_d is usually small number. The table below comprises the values of c_d ratio for given dimensions and different types of LDS.

Dimension	2	3	4	5	6	7	13	∞
Halton	0,65	0,81	1,25	2,62	6,13	17,3	90580	∞
Sobol	1,04	1,00	1,44	1,66	3,20	5,28	647	∞
Faure	0,39	0,12	0,099	0,024	0,018	0,0041	0,00001	0

Table 2: Values of c_d ratio for LDS.

Values in Table 2 imply that with the growth of dimension the Faure's series are the best choice for Monte Carlo simulations since its upper bound of D_N^* statistic tends to zero. Experiments conducted by Fox and Bratley [2],[3] prove that there are no significant differences between results for Faure's [12] and Sobol's [8][12] series whereas Halton's [8][12] ones are much worse. The most important features of LDS which make them perfect for simulations are:

❏ Simplicity to create a single point
❏ Uniform distribution of points in sense of D_N^* (for every subset of subsequent points of reasonable size)
❏ Tendency to fill up regions which were not sampled yet.

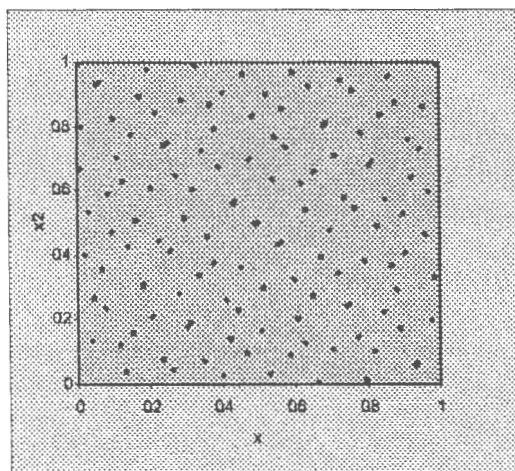The last feature is shown below (for Sobol's series)
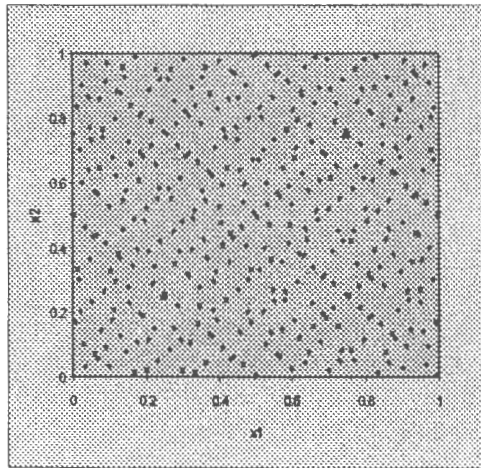

Figure 2: Points 1 – 128
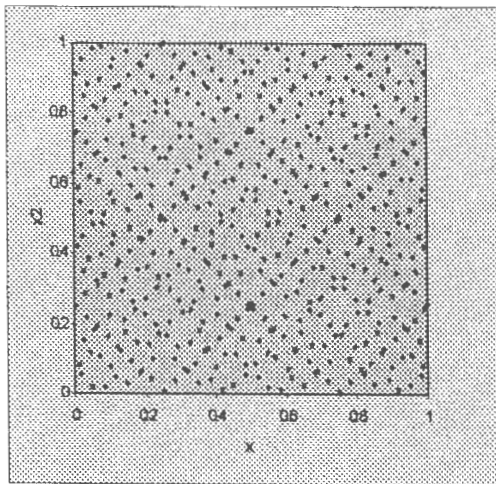
Figure 3: Points 128 – 512
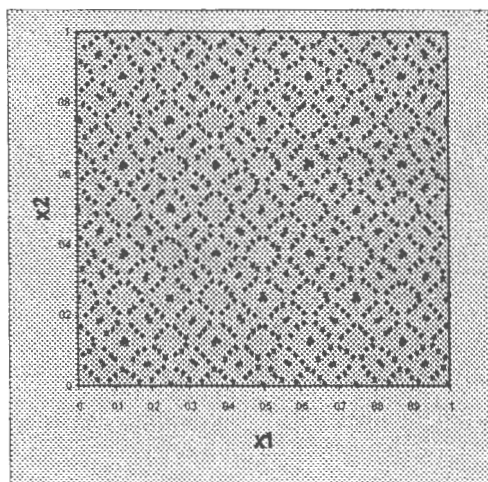

Figure 4: Points 512 – 1024

Figure 5: Points 1 – 1024

As we can see on figures above subsequent points of LDS have some knowledge about the regions which were already sampled and generate new ones exclusively where there are still some gaps.

4. Experimental results

Methods described in section above were used for generating start population of individuals for elitary evolutionary strategy. Additionally a Certesian grid was tested. Considered function was Rosenbrock's one given by formula:

$$f(x) = \sum_{i=1}^{n} 100(x_i^2 - x_{i+1})^2 + (1 - x_i)^2 \quad i \quad x \in R^n.$$

This function has a form of parabolic pipe with one global minima $x_M = (1,1,...,1)^T$ (Fig. 6) with $f(x_M) = 0$.
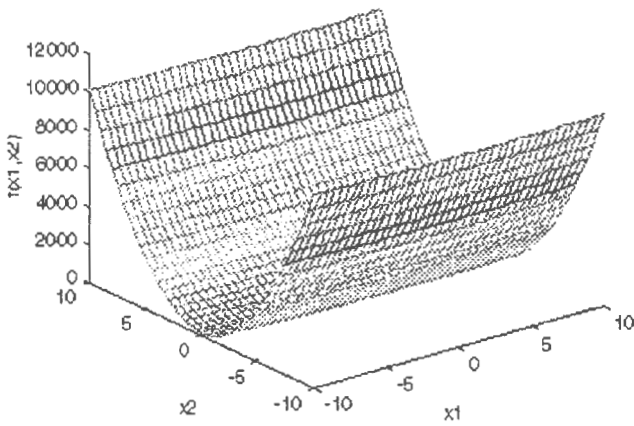
Figure 6: Rosenbrock's function.

For each method of initialisation 50 independent runs of algorithm was done. Control parameters are given in Table 2.

Parameter	Value
Population size (μ)	49
Offspring size(λ)	160
x_i	[-3000,3000]
Number of generations	300
Probability of mutation p_m	1
Probability of cross-over p_k	0.7
Stop criterion	10^{-8}
Dimension of search space n	2

Table 2: Control parameters for experiment.

The percentage of successful runs is shown in table 3. As we can see uniform distribution is definitely the worst with only 40% of successful runs. The best ones are Sobol's series and grid method with 54% of success. A successful test is the one in which a stop criterion was satisfied by the best individual in population e.g. fitness value of found solution was not greater than 10^{-8} .

	Sobol	Grid	Uniform Dist.	Halton	Faure
Percentage of Success	54%	54%	40%	44%	48%

Table 3: Successful runs of optimization process

Figure 7 shows a convergence curves for each initialisation method. Single curve describes the dependency between value of Rosenbrock's function on age of the population. A single sample of this curve was generated as a arithmetic average of fitness value calculated for the best individuals of population for given age but only for successful tests (satisfying stop criterion in whichever generation).
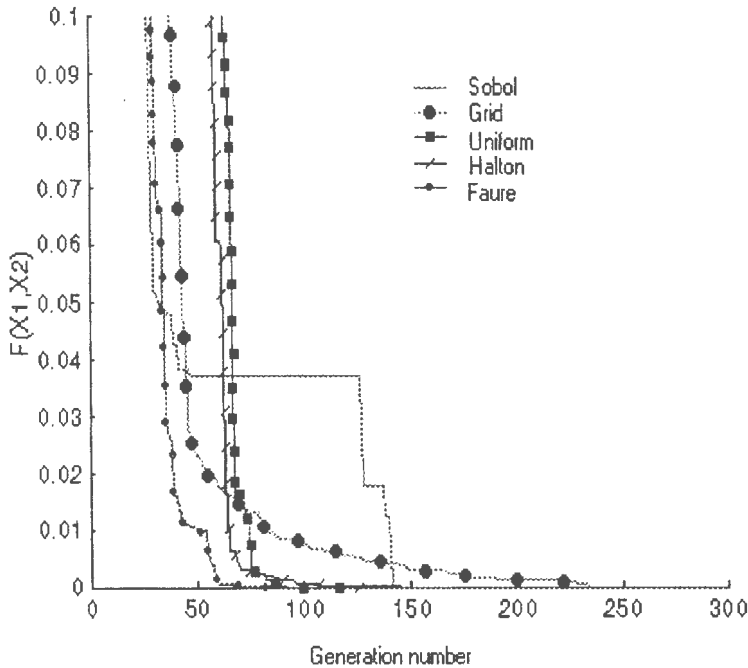

Figure 7: Convergence curves

The results imply that Faure's series guarantees the quickest way of finding optimum. In 50th generation the accuracy for this method was already less than 0.01. The interesting case is Sobol's series which converged quickest with accuracy 0.1 but later could not find a better solution throughout 100 generations. Such a shape of curve is due to "outlier" test which was not eliminated by author.

217

Tables 4 and 5 comprises final results of experiments presenting mean (m) and standard deviation (σ) of crucial variables. The most important is column AGE which explains at what age on average tests reached stop criterion. Other columns show the average value of input variables and average fitness function for best individuals.

	AGE		x_1		x_2		$F_{best}(x_1,x_2)$	
	m	σ	m	σ	m	σ	m	σ
SOBOL	181,24	123,91	0,95	0,25	0,97	0,27	0,06308230	0,23940900
GRID	194,38	117,53	1,16	2,27	6,43	19,9	5,70932860	20,7632099
UNIFORM	210,24	116,27	1,43	5,71	34,07	113,2	32,2139692	113,144852
HALTON	208,08	116,16	0,60	4,26	18,17	49,49	17,9734603	53,2574038
FAURE	152,90	122,95	0,91	0,32	0,91	0,31	0,10143222	0,30261207

Table 4: Convergence for all tests

	AGE		x_1		x_2		$F_{best}(x_1,x_2)$	
	m	σ	m	σ	m	σ	m	σ
SOBOL	79,22	73,58	1	0	1	0	0,000000005	0,000000003
GRID	103,55	86,09	1	0	1	0	0,000000006	0,000000003
UNIFORM	74,1	48,5	1	0	1	0	0,000000005	0,000000003
HALTON	89,82	72,93	1	0	1	0	0,000000005	0,000000003
FAURE	62,13	47,78	1	0	1	0	0,000000005	0,000000003

Table 5: Convergence for successful tests

In both cases the best results are obtained for Faure's and Sobol's series. The results for all tests show that there is significant difference of almost 30 generations between best series (Faure's) and second in order (Sobol's) however Sobol's series find local minima with better accuracy (difference between those two is approximately 0.06). Successful tests seem to support the dominance of Faure's series still in smaller extent – the difference of ages between tests for Sobol's and Faure series is 17 but the results for latter seem to be more stable.

5. Summary

Obtained results prove that the initial phase of evolutionary strategies has a significant impact on convergence of optimization process. Diversification of start population in the right way implicates better explorative character of algorithm in earlier stages of optimisation, which is easy to state analysing results for LDS. Faure's and Sobol's series produce

the best results not only in context of finding global minima but also provide the best accuracy of found solutions. The properties of Faure's series indicate that it should stress its dominance with growth of dimensionality of search space and further research ought to be conducted in this direction.

References

[1] Arabas J., Rzewuski M., Szreter M.: *Poszukiwania efektywniejszych operatorów genetycznych dla strategii ewolucyjnych,* KAEiOG'97 (in polish).

[2] Bratley P., Fox B.L.: *Alghoritm 659: Implementing Sobol's uasirandom Sequence Generator,* ACM Transactions on Mathematical Software, March 1988.

[3] Fox L. B.: *Algorithm 647Implementation and Relative Efficiency of Quasirandom Sequence Generators,* ACM Transaction on Mathematical Software, December 1986

[4] Goldberg D. : *Algorytmy genetyczne i ich zastosowania,* WNT, Warszawa, 1996 .

[5] Michalewicz Z.: *Genetic Algorithms + Data Structures =Evolution Programs,* Springer Verlag, 1992.

[6] Wieczorkowski R. , Zielinski R. : *Komputerowe generatory liczb losowych,* WNT (in polish).

[7] *Algorytmy Ewolucyjne,* Materiały I Krajowej Konferencji, Politechnika Warszawska, Warszawa (in polish).

[8] *Numerical Recipes in C ,* Oxford University Press, 1996

[9] Schwefel H. P. *Evolution and Optimum Seeking,* Wiley

[10] Bäck T,Hoffmeister F., Schwefel H. P. *A: Survey of Evolution Strategies*

[11] Holland J.H: *Adaptation in Natural and Artificial Systems,* University of Michigan Press, 1975.

[12] Dupire B: *MONTE CARLO Methodologies and Applications for Pricing and Risk Managment,*RISK Books, 1998

WYŻSZA SZKOŁA INFORMATYKI STOSOWANEJ I ZARZĄDZANIA

pod auspicjami
Polskiej Akademii Nauk

ZAŁOŻYCIELEM
Wyższej Szkoły Informatyki Stosowanej i Zarządzania
jest
FUNDACJA KRZEWIENIA NAUK SYSTEMOWYCH
powołana z inicjatywy
**Prezesa
POLSKIEJ AKADEMII NAUK**

FUNDATOREM
Fundacji Krzewienia Nauk Systemowych
jest
POLSKA AKADEMIA NAUK

ORGANEM
sprawującym nadzór jest
MINISTERSTWO EDUKACJI NARODOWEJ

Wyższa Szkoła Informatyki Stosowanej i Zarządzania
prowadzi studia wyższe na kierunkach:
**INFORMATYKA
ZARZĄDZANIE I MARKETING**

SIEDZIBA
**Instytut Badań Systemowych
Polskiej Akademii Nauk
ul. Newelska 6, 01-447 Warszawa**