論文

Optimization of PSO by Determining Parameters for Multiple Stages with DoE

Michio Innami

Particle Swarm Optimization (PSO) is one of the population-based metaheuristics. The algorithm is simple yet the convergence is rather fast. The parameter values have a large influence on the search performance. Many studies on PSO parameter tuning have been conducted so far. This paper proposes determining parameters separately in each stage into which the iterative process is divided, using Design of Experiments (DoE) methodology. The situation of the swarm is considered to be changing as the process proceeds. Hence, a certain effect is expected with this method. The exploration with the proposed method is performed on four test functions. Additionally the exploration with three other methods is also conducted for comparison. The proposed method yields good results. This study also reveals that the search performance improves when parameters, such as inertia weight, increase as the process proceeds.

Keywords: Metaheuristics, PSO, Search performance, Parameter tuning, DoE

1. Introduction

It is not practical finding values for all possible combinations to solve NP-hard combinatorial optimization problems. In this case, metaheuristic algorithms are very effective. Optimization problems, *i.e.*, minimization or maximization, can be solved with the methods almost regardless of problem classes. The solutions can be obtained in a practical time. However, the exact values are not guaranteed. The methods have been applied to various problems, *e.g.*, the shortest path problem, the work scheduling problem, the facility location problem, etc. Genetic algorithm (GA)[1] is one of the most famous algorithms. Another representative algorithm, Particle Swarm Optimization (PSO) was proposed by Kennedy and Eberhart in 1995[2]. PSO algorithm is simple, yet the convergence to the solution is rather fast. However, it tends to drop into local optima.

Shi and Eberhart introduced into the original PSO a coefficient named inertia weight[3], which is now used commonly. Clerc proposed a constriction factor[4], which is beyond the scope of this paper.

PSO parameters have significant influence on search performance. Therefore, parameter tuning has also been discussed. Shi and Eberhart proposed linearly decreasing inertia weight during the iterative process[5]. Van den Bergh presented optimal parameters in his PhD thesis[6]. Xin, Chen and Hai divided the iterative process into three stages, and allocated different inertia weight for each of them[7]. On the other hand, Dao, Abhary and Marian applied to GA Taguchi experimental design, Design of Experiments (DoE), for parameter tuning[8].

The author applied DoE to determine PSO parameters, which were constant during the iterative process[9]. Generally, it is important for metaheuristics to balance between the performance of converging solutions and maintaining diversity. Since the situation of the swarm changes as the search process proceeds, it is considered effective to adjust the balance by changing parameters during the process.

This study divides the iterative process into four stages, and tries to determine PSO parameters separately in each of them by using DoE, to improve search performance.

The rest of this paper is structured as follows. The next section focuses on the theory of PSO, while Section 3 refers to DoE. Section 4 explains problem setting. Section 5 discusses experimental results. Finally Section 6 concludes the paper.

2. PSO

PSO was inspired by the behavior of a flock of birds, a swarm of insects, a school of fish, etc.[10] The particles (individuals) that compose a swarm exist in the n-dimensional search space. The sets of components of their position vectors are candidate solutions. Initially the particles are distributed randomly in the search space. Then their positions are updated according to the following equations.

$$\boldsymbol{v}_i^{k+1} = w\boldsymbol{v}_i^k + c_1 r_1(\boldsymbol{pbest}_i - \boldsymbol{x}_i^k) + c_2 r_2(\boldsymbol{gbest} - \boldsymbol{x}_i^k) \quad (1)$$

$$\boldsymbol{x}_i^{k+1} = \boldsymbol{x}_i^k + \boldsymbol{v}_i^{k+1} \quad (2)$$

where $\boldsymbol{x}_i^k$, $\boldsymbol{v}_i^k$ are the position vector and the velocity vector of particle i in the k^{th} update, respectively. $\boldsymbol{pbest}_i$, $\boldsymbol{gbest}$ are the position vectors at which the fitness (function) values are the most excellent in the migration history of particle i and of the whole swarm, respectively. w is a coefficient named inertia weight, c_1 and c_2 are coefficients that indicate the influence on particle i of $\boldsymbol{pbest}_i$ and $\boldsymbol{gbest}$, respectively. r_1 and r_2 are uniform random numbers between 0 and 1.

After the update are repeated for predetermined times, the component set of the position vector of $\boldsymbol{gbest}$ is treated as quasi-optimal solution.

3. DoE

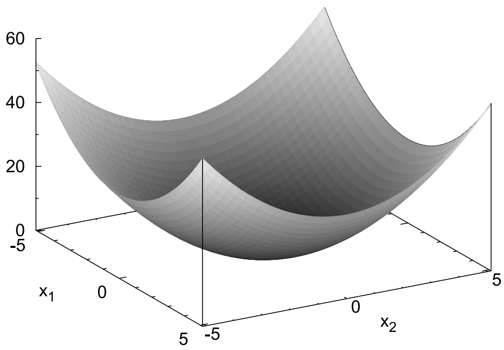
In an experiment, if the numbers of factors (parameters) and/or their levels are large, it is not practical to try all possible combinations. In this case, DoE proposes to experiment part of the combinations instead of all of them, with the orthogonal array experiment design[11]. An orthogonal array is a table which indicates the combination of factors. Various types are prepared corresponding to the numbers of factors and levels. The orthogonal array is generated based on a matrix called a latin square, which the same numbers are not exist in one row nor one column. The columns are orthogonal to each other. We calculate the mean value of each level of factors derived from the experiments executed according to the orthogonal array. Thus the best level of each factor is obtained. DoE extremely reduces total number of experiments.

4. Problem setting

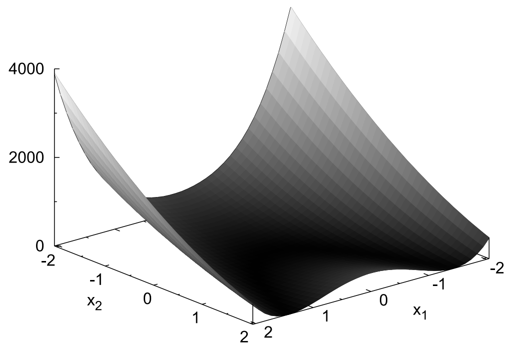
Test functions are used for evaluating the performance of search algorithms. For each function the domain is defined and the exact global minimum or maximum is known. The performance of search algorithm is indicated by the difference between the solution obtained by the algorithm and the exact global minimum/maximum. The test functions are classified into unimodal and multimodal. Multimodal functions are generally more difficult to find good solutions. This paper treats four functions as described below. They are all representative test functions and have a global minimum in the search space.

(1) Sphere function

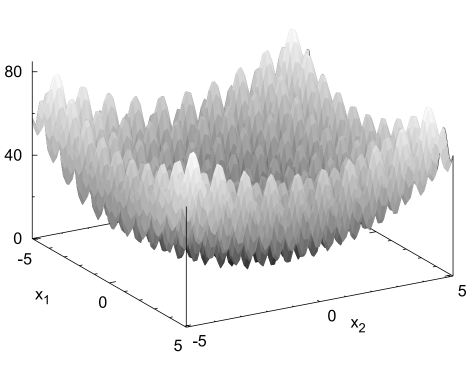
$$f_1 = \sum_{i=1}^{n} x_i^2$$



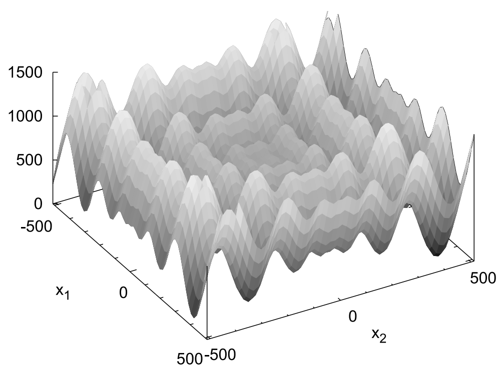
(1) Sphere



(2) Rosenbrock



(3) Rastrigin



(4) Schwefel

Fig. 1　Shapes of test functions ($n = 2$)

$$|x_i| \le 5.12 \quad (i = 1, \cdots, n)$$
$$\min f_1 = f_1(0, \cdots, 0) = 0$$

Unimodal function. No dependency between variables. It is relatively easy to obtain a good solution as a test function. Nevertheless as n get larger, the convergence becomes extremely worse.

(2) Rosenbrock function

$$f_2 = \sum_{i=1}^{n-1} \left\{ 100\left(x_i^2 - x_{i+1}\right)^2 + (1 - x_i)^2 \right\}$$
$$|x_i| \le 2.048 \quad (i = 1, \cdots, n)$$
$$\min f_2 = f_2(1, \cdots, 1) = 0$$

Unimodal function. Strong dependencies between variables. Global minimum exists in a narrow valley. Although the valley is rather easy to find, convergence to global minimum is difficult.

(3) Rastrigin function

$$f_3 = 10n + \sum_{i=1}^{n} \left\{ x_i^2 - 10\cos(2\pi x_i) \right\}$$
$$|x_i| \le 5.12 \quad (i = 1, \cdots, n)$$
$$\min f_3 = f_3(0, \cdots, 0) = 0$$

Multimodal function. No dependency between variables. The function has many local optima geometrically in a regular pattern. It has a big valley structure, of which the global optimum lies at the bottom.

(4) Schwefel function

$$f_4 = \sum_{i=1}^{n} \left(-x_i \sin\sqrt{|x_i|} \right) + 418.9828873n$$
$$|x_i| \le 512 \quad (i = 1, \cdots, n)$$
$$\min f_4 = f_4(420.9687, \cdots, 420.9687) = 0$$

Multimodal function. No dependency between variables. The global optimum exists close to the boundaries of the search space, around which there are no local optima. Hence the search algorithms easily trap into local optima. It is more difficult to get a good solution than Rastrigin function.

The shapes of above functions in the case of n = 2 are illustrated in Fig. 1.

Determining PSO parameters separately for multiple stages proposed in this paper is carried out as follows. Factors are three parameters: w, c_1 and c_2. Each of them has four levels as shown in Table 1. Swarm size is fixed to 20. Final iteration number is set to 5000 to terminate the exploration within a practical time. The dimension is set to n = 30 for all the test functions. With above setting, the convergence is expected to progress to a certain extent for observing properly.

Table 1 Factors and their levels

Factor	Parameter	Level			
		1	2	3	4
A	w	0.4	0.6	0.8	1.0
B	c_1	1.2	1.4	1.6	1.8
C	c_2	1.2	1.4	1.6	1.8

Table 2 Iteration numbers in each stage

Stage	Iteration number
I	1 - 10
II	11 - 100
III	101 - 1000
IV	1001 - 5000

Table 3 Orthogonal array L_{16} (4^3)

Experiment	Level		
	Factor A	Factor B	Factor C
1	1	1	1
2	1	2	2
3	1	3	3
4	1	4	4
5	2	1	2
6	2	2	1
7	2	3	4
8	2	4	3
9	3	1	3
10	3	2	4
11	3	3	1
12	3	4	2
13	4	1	4
14	4	2	3
15	4	3	2
16	4	4	1

The iterative process are divided into four stages as shown in Table 2. To determine the parameters by using DoE methodology, the combinations of parameters for the experiments in each stage are designated according to the orthogonal array L_{16} (4^3) as shown in Table 3.

Each experiment is executed five times for each stage. The optimal parameters are determined by the mean function values derived from the experiments. Then with the optimal parameters, five trials are executed again in the stage. The position and velocity vectors of the particles at the end of the stage are taken over to the next one, for the subsequent experiments. As a comparison, the explorations with the other three types of parameters are also carried out as follows.

(i) Recommended parameters by Van den Bergh[6]

Van den Bergh proposed in his PhD thesis these parameters.

$$w = 0.729844$$

$$c_1 = 1.496180$$

$$c_2 = 1.496180$$

(ii) Linearly decreasing inertia weight[5]

w is decreasing linearly from 0.9 to 0.4 during the iterative process. c_1 and c_2 are as follows.

$$c_1 = 2$$

$$c_2 = 2$$

(iii) Constant parameters determined by DoE

Constant parameters are determined throughout the iterative process. The factors and the levels are the same as Table 1. Every experiment is executed according to Table 3 for five times. The optimal parameters are determined from the mean value derived in every experiment.

In all cases, swarm size is 20, and the iterative process is repeated 5000 times same as the proposed method.

5. Results

Table 4 shows the optimal parameters derived from the experiments in every stage using DoE. In comparison, Table 5 shows the parameters, constant throughout the iterative process, obtained by DoE. With the above optimal parameters, the explorations are executed again 30 times for both cases. The experiments with the parameters recommended by Van den Bergh and with linearly-decreasing w are also carried out 30 times. Fig. 2 shows the transition of each function value of worst, mean and best cases with four types of parameters. In most cases, good results were obtained in the explorations using DoE for multiple stages. For the transition of the best values of Rosenbrock function, the parameters of Van den Berge and linearly-decreasing w brought the better results. The DoE method for a single stage is also effective to the best values of Schwefel function. The above results are considered to be obtained because the determination process with DoE is judged by the mean values derived in each experiment. Every transition data of mean function value shows the better results with the proposed method. Furthermore, there are less differences among best and worst data obtained by the proposed method than the other ones. Generally, w is recommended to be set large number at the beginning which is suitable for global search, and is decreased gradually as the iterative process progresses for facilitating local search. This study shows the opposite results, especially for multimodal functions. At least the results indicate that it leads to good solutions without narrowing the migration range of particles to maintain diversity even the process proceeds. Initially

Table 4　Optimal parameters derived from DoE (multiple stages)

Function	Iteration number	w	c_1	c_2
Sphere	1 - 10	0.4	1.4	1.2
	11 - 100	0.6	1.6	1.2
	101 - 1000	0.6	1.6	1.6
	1001 - 5000	0.6	1.6	1.8
Rosenbrock	1 - 10	0.4	1.6	1.2
	11 - 100	0.6	1.6	1.6
	101 - 1000	0.6	1.8	1.8
	1001 - 5000	0.6	1.8	1.8
Rastrigin	1 - 10	0.4	1.4	1.2
	11 - 100	0.4	1.2	1.2
	101 - 1000	0.8	1.8	1.8
	1001 - 5000	0.8	1.8	1.4
Schwefel	1 - 10	0.4	1.2	1.4
	11 - 100	0.6	1.6	1.6
	101 - 1000	0.8	1.4	1.8
	1001 - 5000	0.8	1.8	1.8

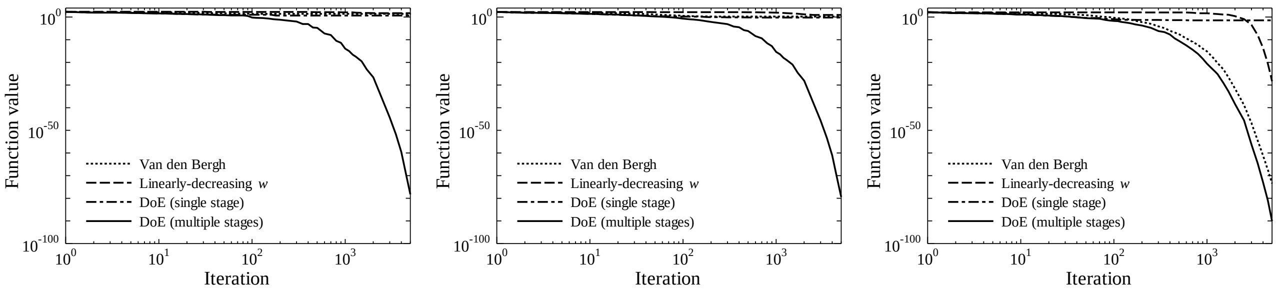
Table 5　Optimal parameters derived from DoE (single stage)

Function	w	c_1	c_2
Sphere	0.4	1.8	1.2
Rosenbrock	0.4	1.8	1.2
Rastrigin	0.8	1.8	1.2
Schwefel	0.8	1.8	1.4

particles are distributed randomly in the search space. That means the diversity of candidate solution is acquired then. Therefore, it is of little importance to get further diversity with large w at the first stage. As convergence progresses, larger w seems to be effective to maintain diversity. Especially for multimodal functions, PSO may trap into local optima. Hence larger w is considered to work effectively to get out of them. $\boldsymbol{pbest_i}$, ***gbest*** do not initially have enough information. As iteration proceeds, they grow reliable having more experience. Therefore, the larger c_1 and c_2 are considered to be effective at later stages.

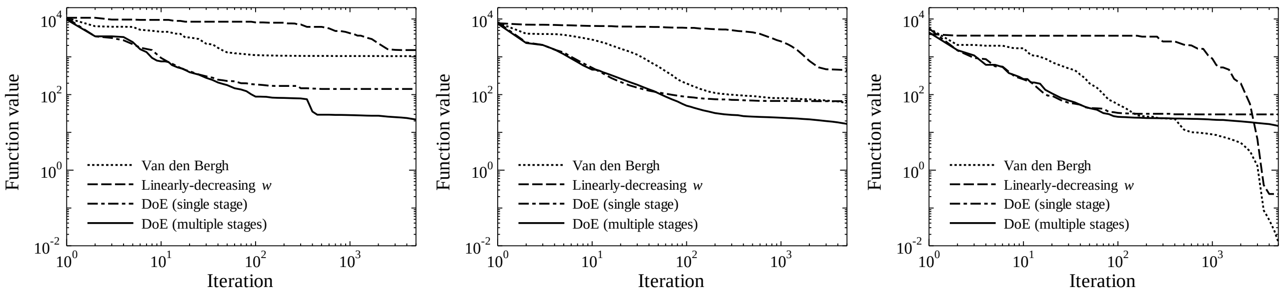
6. Conclusion

This study proposes determining PSO parameters separately for each stage into which the iterative process is divided, by using DoE methodology. The proposed method brought good results in most cases. For the test functions treated in this paper, it is revealed that increasing the values of w, c_1 and c_2 as the process proceeds leads to good results. In future work, the author would examine the further exploration in other functions. Since the exploration process proposed in this study is somewhat complicated, it is considered effective



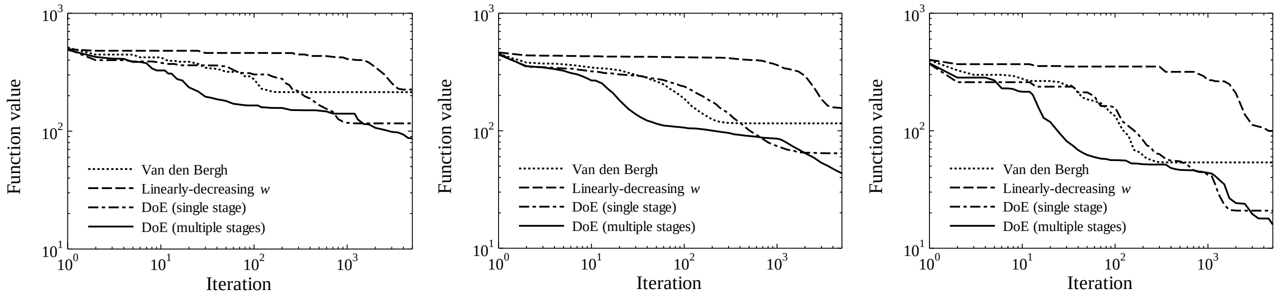
a. Worst b. Mean c. Best

(1) Sphere



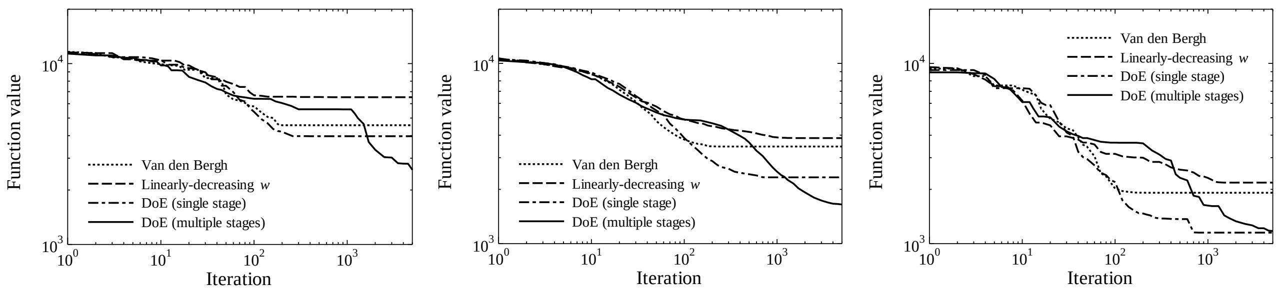
a. Worst b. Mean c. Best

(2) Rosenbrock



a. Worst b. Mean c. Best

(3) Rastrigin



a. Worst b. Mean c. Best

(4) Schwefel

Fig. 2 Transition of function values

to apply it to the functions that are hard to converge.

References

[1] J.H. Holland: “Adaptation in Natural and Artificial Systems”, University of Michigan Press, Ann Arbor, MI,. (1975).

[2] J. Kennedy, and R.C. Eberhart: “Particle Swarm Optimization”, Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ, pp. 1942-1948 (1995).

[3] Y. Shi, and R.C. Eberhart: “A Modified Particle Swarm Optimizer”, Proceedings of IEEE International Conference on Evolutionary Computation, Anchorage, Alaska, Piscataway, NJ, pp. 69-73 (1998).

[4] M. Clerc: “The Swarm and the Queen: Towards a Deterministic and Adaptive Particle Swarm Optimization”, Proceedings of the 1999 Congress on Evolutionary Computation, Vol. 3, pp. 1951-1957 (1999).

[5] Y. Shi, and R.C. Eberhart: “Empirical Study of Particle Swarm Optimization”, Proceedings of the 1999 Congress on Evolutionary Computation, Vol. 3, pp. 1945-1950 (1999).

[6] F. Van den Bergh: “An Analysis of Particle Swarm Optimizers”, PhD Thesis, University of Pretoria (2006).

[7] J. Xin, G. Chen, and Y. Hai: “A Particle Swarm Optimizer with Multi-Stage Linearly-Decreasing Inertia Weight”, 2009 International Joint Conference on Computational Sciences and Optimization, pp. 505-508 (2009).

[8] S.D. Dao, K. Abhary, and R. Marian: “Maximising Performance of Genetic Algorithm Solver in Matlab”, Engineering Letters, 24, pp. 75-83 (2016).

[9] M. Innami: “Application of Design of Experiments Methodology for Determining PSO Parameters”, Memoirs of Hokkaido Polytechnic College, No. 34, pp. 1-7 (2019). in press.

[10] C.W. Reynolds: “Flocks, Herds and Schools: A Distributed Behavioral Model”, Computer Graphics (ACM SIGGRAPH ‘87 Conference Proceedings), Vol. 21, No. 4, pp. 25-34 (1987).

[11] K. Yang, and B. El-Haik: “Design for Six Sigma: A Roadmap for Product Development”, NY, McGraw-Hill (2003).

（Received 2019/1/9，Accepted 2019/3/13）

*Michio Innami, Hokkaido Polytechnic College, 3-190 Zenibako, Otaru, Hokkaido 047-0292　email:Innami.Michio@jeed.or.jp