Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

A NEW PARTITION-BASED RANDOM SEARCH METHOD FOR DETERMINISTIC OPTIMIZATION PROBLEMS

Ziwei Lin
Shichang Du

Department of Industrial Engineering and Management
Shanghai Jiao Tong University
800 Dongchuan Road, Minhang
Shanghai, 200240, CHINA

Andrea Matta

Department of Mechanical Engineering
Politecnico di Milano
Via G. la Masa, 1
Milano, 20156, ITALY

ABSTRACT

The Nested Partition (NP) method is efficient in large-scale optimization problems. The most promising region is identified and partitioned iteratively. To guarantee the global convergence, a backtracking mechanism is introduced. Nevertheless, if inappropriate partitioning rules are used, lots of backtracking occur reducing largely the algorithm efficiency. A new partition-based random search method is developed in this paper. In the proposed method, all generated regions are stored for further partitioning and each region has a partition speed related to its posterior probability of being the most promising region. Promising regions have higher partition speeds while non-promising regions are partitioned slowly. The numerical results show that the proposed method finds the global optimum faster than the pure NP method if numerous high-quality local optima exist. It can also find all the identical global optima, if exist, in the studied case.

1 INTRODUCTION

Real-world optimization problems are very challenging because the scale of the problem is large and closed-form expressions of the objective function and/or constraints are normally not available (e.g., the objective function can only be estimated by simulation tools). The partition-based random search is a widely used heuristic framework for solving difficult large-scale optimization problems, e.g., Nested Partition (NP) (Shi and Ólafsson 2000). It partitions the feasible domain into sub-regions systematically and guides the search by deciding which region to concentrate the searching effort on, i.e., the most promising region. Mainly three components are involved in the NP method:

1. **Partitioning.** The current most promising region is partitioned into several sub-regions according to specific partitioning strategies.
2. **Sampling.** A limited budget is allocated to different regions based on specific budget allocation strategies. Then, solutions are randomly sampled from each region and evaluated.
3. **Most promising region.** The next most promising region is identified, according to the promising index, for further partitioning.

Different promising indexes can be used to identify the next most promising region in the NP method. The most commonly used indexes are the sample extreme value (the minimum or the maximum) and the sample mean. The promising index can also cooperate with mathematical programming bounds, probabilistic bounds, meta-model outputs or performance variability. After identifying and partitioning the most promising region, the NP method merges the remaining regions except the new-partitioned ones. The budget is usually allocated equally to all regions. Several budget allocation strategies are proposed to increase the probability of correctly selecting the next most promising region using Ranking and Selection algorithms (e.g., OCBA (Chen et al. 2000) and Indifference-zone (Hong and Nelson 2005)). For instance, an extreme value-based criterion is proposed by Chen et al. (2014) using the Weibull location parameter maximum likelihood estimation. Then, Optimal Sample Allocation Strategy is developed to allocate the sample budget to each region aiming at maximizing the probability of making the correct move. Linz et al. (2016) propose a budget allocation method to maximize the approximate probability of correctly selecting the next most promising region using the sample quantile of each region as the promising index.

The corresponding number of solutions are randomly sampled from each region after the budget is allocated. Uniform sampling is commonly used. In addition, biasing the sampling probability in one region based on the problem features, and/or incorporating heuristic algorithms in the sampling phase, can enhance the chance of making the correct move in practice (Shi and Olafsson 2009). For example, Simulated Annealing is used in the sampling phase in Luo and Yu (2007) to improve the efficiency of the NP method. In the NP method, the solutions are re-sampled in each iteration, i.e., the data sampled in the previous iterations are not used to calculate the promising index in the current iteration. This can mitigate the influence of the noise introduced by the sampling and increase the robustness of the algorithm. To maintain a global perspective, the NP method also samples solutions from the non-promising region, i.e., the complementary region, in each iteration and introduces a backtracking mechanism allowing the algorithm to escape the local optimum.

The partitioning rule is usually developed considering the features of the problem to be solved. Some intelligent partitioning rules are introduced by Shi and Olafsson (2009) for traveling salesman problems (TSP), beam angle selection problems, local pickup and delivery problems, etc. A good partitioning scheme is capable of gathering good solutions together, which can save searching effort and improve the efficiency of the NP method. Nevertheless, if an inappropriate partitioning rule is used, although the backtracking mechanism guarantees the global convergence of the NP method, a large amount of backtracking may occur, or a large size of budget is wasted to determine the next most promising region.

A new partition-based random search method is developed in this paper, which stores all the generated regions during the search and decides whether and when to further partition them based on observations. The main concept is that all regions will be further partitioned, at different speeds, until they are non-partitionable (e.g., singleton regions or acceptable precision regions in continuous problems). Promising regions are partitioned at a higher speed than non-promising regions. By modifying the partition speeds of regions, we control the speeds of reaching the non-partitionable size at different positions of the feasible domain, and furthermore control the searching effort. This idea is similar as TD-OCBA (Zhu et al. 2019), which controls the computational effort by changing the simulation speeds of competitive solutions in single-run simulation optimization problems.

In the proposed method, the quantile of the objective values in one region is considered as the promising index, which is similar to Linz et al. (2016). Linz et al. (2016) allocate budgets to determine which region has the optimal quantile and further partition the region. It belongs to the framework of the classic NP method. Nevertheless, in the proposed method, all regions will be eventually partitioned and determining the most promising region is not needed. If several similar promising regions are generated, we believe that an inappropriate partitioning is executed and there may be no difference among these regions. These regions are all partitioned into smaller sub-regions to get better performance estimates and make the decision in subsequent iterations. Therefore, neither a high budget is required to determine which one is the real most promising region to be further exploited, nor a lot of backtracking occur due to rough choices. This is the advantage of the proposed method compared to the classic NP method. Also, parallel partitioning of multiple promising regions makes the proposed method can be applied directly to handle multimodal optimization problems, in which multiple global optima or local optima with high quality are of interest. Currently, to deal with this kind of problems, evolutionary algorithms (e.g., Thomsen (2004), Li (2004)), or classic optimization techniques with multiple starts and multiple runs are used.

The proposed method is developed based on the budget allocation method proposed by Lin et al. (2018). A higher budget is allocated to the region with a higher approximate posterior probability of containing the optimal quantile based on previous observations. The partitioning is executed when the sample size in a region reaches a threshold. Like the NP method, the proposed method can also be applied combined with other meta-heuristic algorithms or intelligent partitioning rules. The proposed method is tested on two well-known functions: the Rastrigin function and the Himmelblau's function. The numerical results show that the proposed method has better performance than the pure NP method when a lot of high-quality local optima exist, and it is capable of finding all the identical global optima in the studied case.

The proposed method is described in detail in section 2 and tested on different functions in section 3. Finally, conclusions and future works are presented in section 4.

2 ALGORITHMS

A new partition-based random search method is proposed to solve minimization problems in which the objective function $y(\boldsymbol{x})$ is deterministic. The feasible domain is denoted as $\mathscr{D}$. A partitioning rule is pre-defined before applying the algorithm. In the proposed partition-based random search method, a budget allocation method aiming at minimizing the sample set quantile is used combined with the nested partitioning framework in order to bias the sampling probability among different regions.

2.1 The Pure Nested Partition (PNP)

The procedure of the pure NP method (PNP) is shown in Algorithm 1. Let the whole feasible domain $\mathscr{D}$ be the current most promising region. The user-defined parameter n_1^{NP} is the budget size in each new-partitioned region and n_2^{NP} is the budget size in the complementary region. At the first step, the current most promising region is partitioned into several disjoint sub-regions and the remaining regions are combined as the complementary region. A corresponding number of solutions are uniformly sampled from each region and evaluated at step 2. The best observation in a given region is chosen as the promising index of this region. Solutions are re-sampled in each iteration, which means that the solutions observed in the previous iterations are not taken into account in the calculation of the promising index in this iteration. This can mitigate the influence of the sampling noise. If the best solution in this iteration belongs to one of the new partitioned regions, this region becomes the next most promising region. If the best solution belongs to the complementary region, a backtracking occurs. Three backtracking rules are introduced by Shi and Olafsson (2009): *I*. backtrack to the last most promising region; *II*. backtrack to the whole feasible domain; *III*. backtrack to the latest most promising region that contains the current best solution (if the extreme value is chosen as the promising index). These steps are repeated until a stopping criterion is met.

Algorithm 1 Pure Nested Partition

Initialization. The whole feasible domain $\mathscr{D}$ is the current most promising region.

Step 1. Partition the current most promising region into several sub-regions. Combine the remaining regions as the complementary region.

Step 2. Randomly sample n_1^{NP} solutions from each new partitioned sub-region and n_2^{NP} solutions from the complementary region. Estimate the objective values at the new sampled solutions.

Step 3. Identify the next most promising region using the extreme value. If the next most promising region is the complementary region, backtrack.

Step 4. If the stopping criterion is met, stop. Otherwise, go to step 1.

2.2 Budget Allocation for Quantile Minimization

A budget allocation method is proposed by Lin et al. (2018) to generate initial designs with size N for minimization problems. Feasible solutions are clustered into K groups and the sampling probabilities are biased by allocating different budgets to different groups. The goal of the budget allocation is to minimize the α-quantile ($\alpha < 0.5$) of the objective values of all the sampled solutions. This budget allocation method is developed under the assumption that the objective values in one group are independently, identically and normally distributed. The budget is allocated to each group dynamically in order to let the budget size in group k approximately proportional to its posterior probability of having the optimal quantile, based on previous observations: $P(q_{\alpha,k} < q_{\alpha,i}, \forall i | \boldsymbol{Y}_i, \forall i)$, where $q_{\alpha,k}$ is the α-quantile of the objective values in group k and $\boldsymbol{Y}_i$ is the set of previous observations from group i. The numerical results show that this budget allocation method has good performance even if the normality assumption is not satisfied.

In the budget allocation method, a first stage sampling is performed by allocating a budget of size $n_{1,k}$ ($n_{1,k} \geq 2$) to group k to estimate the group performance. The group sample means $\hat{\mu}_k$ and group sample variances $\hat{\sigma}_k^2$ are calculated based on $n_{1,k}$ observations. The current best group $\hat{b}$ is defined as the group with the best estimated quantile under the given assumption: $\hat{b} = \arg\min_k\{\hat{\mu}_k + z_\alpha \hat{\sigma}_k\}$, where z_α is the α-quantile of the standard normal distribution. Denote the total budget size allocated to group k as n_k. The formulas of the group budget sizes are as follows:

$$\frac{n_k}{n_{\hat{b}}} = \frac{F(C_{k,\hat{b}}; n_{1,k}-1, n_{1,\hat{b}}-1)}{F(C_{\hat{b},k}; n_{1,\hat{b}}-1, n_{1,k}-1)}, \forall k \neq \hat{b}, \tag{1}$$

$$n_{\hat{b}} = N \Big/ \left(\sum_k \frac{F(C_{k,\hat{b}}; n_{1,k}-1, n_{1,\hat{b}}-1)}{F(C_{\hat{b},k}; n_{1,\hat{b}}-1, n_{1,k}-1)} \right), \tag{2}$$

where $F(\cdot; v_1, v_2)$ is the cumulative distribution function of the F-distribution with degrees of freedom v_1 and v_2, $\hat{\tau} = \min_k\{\hat{\mu}_k + z_\alpha \hat{\sigma}_k\}$,

$$C_{i,j} = \frac{1 + \frac{1}{\hat{c}_j^2} - \frac{1}{n_{1,j}}}{1 + \frac{1}{\hat{c}_i^2} - \frac{1}{n_{1,i}}}, \forall i, j \quad \text{and} \quad \hat{c}_k = \frac{\hat{\sigma}_k}{\hat{\mu}_k - \hat{\tau}}, \forall k.$$

Then, additional $\max(0, [n_k] - n_{1,k})$ solutions are sampled from group k at the second stage, where $[\cdot]$ indicates that the value is rounded to the nearest integer. This method can also be applied in multiple stages by adding a new budget at each stage.

2.3 The Proposed Partition-Based Random Search

The idea of controlling the partition speeds of different regions is realized taking advantage of the aforementioned budget allocation method. In this paper, the partition speed of a region is defined as the reciprocal of the iteration number between this region being generated and further partitioned into sub-regions. In the budget allocation method, the sample budget is allocated dynamically to each region trying to let the total budget size in a region proportional to its posterior probability of being the most promising region (the quantile is considered as the promising index). Let us further partition a region when the sample size in this region reaches a threshold n_{max} (the triggering condition). Regions with higher posterior probabilities will achieve this triggering condition faster than regions with lower posterior probabilities thanks to the feature of the budget allocation method, which can make promising regions have higher partition speeds than non-promising regions. In addition, if this triggering condition is met, which means more samples are available in this region, further partitioning the region to narrow down the region size can help to obtain more precise region performance estimates.

Differently from the PNP method, all the sampled data are reused in subsequent iterations due to the exponentially increasing number of regions. To mitigate the effect of the noise caused by the sampling, a modification on the budget allocation is introduced. Denote the partition depth of the generated region k as l_k. In this paper, the partition depth of a region refers to the minimum number of partition actions required to obtain this region. The $n_{1,k}$ in equation (1) and equation (2) are adjusted based on the partition depth:

$$n_{1,k} = \max\left(2, \left[\frac{l_k}{\max_i\{l_i\}} n'_{1,k}\right]\right), \forall k,$$

where $n'_{1,k}, \forall k$ are the number of observations in different regions and $[\cdot]$ indicates that the value is rounded to the nearest integer. This modification aims at forcing the method to sample also from non-promising but broad regions.

In the proposed method, the sample budget is allocated to regions using the modified budget allocation method iteratively. Once the sample size in one region reaches the threshold n_{max}, this region is further partitioned, according to a predefined partitioning rule. The procedure is repeated until a stopping criterion is met, such as the total budget is exhausted, the maximum partition depth is achieved (i.e., a non-partitionable region is obtained) or no better solutions are observed within a certain number of iterations. The detailed procedure of the proposed method is described as following.

Step 0 Initialization. Set user-defined parameters:

- $*$ N^{tot}: the total budget size;
- $*$ $\mathscr{P}$: the partitioning rule;
- $*$ α: the fraction of good solutions if the normality assumption holds, $\alpha < 0.5$;
- $*$ n_0: the base sample size in each region, $n_0 \geq 2$;
- $*$ Δ: the added budget size in each iteration;
- $*$ n_{max}: the sample size threshold, $n_{max} > n_0$.

Let the current iteration number $r = 0$, the set of current regions $\mathbb{S}^{(r)} = \{\Omega_1^{(r)}\}$, where $\Omega_1^{(r)} = \mathscr{D}$ is the whole feasible domain, and the size of the set $||\mathbb{S}^{(r)}|| = 1$. The sample size in each region is $n_k^{(r)} = 0, k = 1, \cdots, ||\mathbb{S}^{(r)}||$. The partition depth of each region is $l_k^{(r)} = 0, \forall k$.

Step 1 Partitioning. $r = r+1$ and $\mathbb{S}^{(r)} = \mathbb{S}^{(r-1)}$. For $k = 1, \cdots, ||\mathbb{S}^{(r-1)}||$, if the region $\Omega_k^{(r-1)}$ is partitionable and the sample size in this region reaches the sample size threshold, i.e., $n_k^{(r-1)} \geq n_{max}$, (or the original solution space has not been partitioned yet), partition this region $\Omega_k^{(r-1)}$ into several mutually exclusive new sub-regions: $\mathscr{P}(\Omega_k^{(r-1)}) = \{\Omega_1^{new}, \cdots, \Omega_{K(\mathscr{P})}^{new}\}$ where $\cup_i \Omega_i^{new} = \Omega_k^{(r-1)}$ and $\Omega_i^{new} \cap \Omega_j^{new} = \varnothing, \forall i \neq j$. The set of current regions is updated as $\mathbb{S}^{(r)} = \mathbb{S}^{(r)} / \{\Omega_k^{(r-1)}\} \cup \mathscr{P}(\Omega_k^{(r-1)})$. Update the sample size in each region $n_k^{(r)}$ and the partition depth $l_k^{(r)}$.

Step 2 First Stage Sampling. For $k = 1, \cdots, ||\mathbb{S}^{(r)}||$, uniformly sample $\tilde{n}_k = \max(0, n_0 - n_k^{(r)})$ additional solutions from the region $\Omega_k^{(r)}$ so that each region has a sample size not less than the base sample size. $n_k^{(r)} = n_k^{(r)} + \tilde{n}_k$.

Step 3 Additional Sampling. Calculate the sample means $\hat{\mu}_k$ and the sample variances $\hat{\sigma}_k$ in different regions. Modify the budget sizes in different regions according to their partition depth:

$$n_{1,k} = \max\left(2, \left[\frac{l_k^{(r)}}{\max_i\{l_i^{(r)}\}} n_k^{(r)}\right]\right), \forall k.$$

Allocate a new budget of size Δ to all regions, using equation (1) and equation (2) with $N = \sum_k n_{1,k} + \Delta$. Uniformly sample additional $\tilde{n}_k = \max(0, [n_k] - n_{1,k})$ solutions from the region $\Omega_k^{(r)}$. $n_k^{(r)} = n_k^{(r)} + \tilde{n}_k$.

Step 4 Stop Criteria. If the stopping criterion is met, stop the algorithm. Otherwise, go to step 1.

The budget allocation method requires the normality assumption on the observations in one region, which is usually not satisfied in practice. The estimated region quantile $\hat{\mu}_k + z_\alpha \hat{\sigma}_k$ is not the exact α-quantile of the objective values in this region. Therefore, in the proposed method, the α value is treated as an algorithm parameter to be calibrated, rather than the fraction of good solutions of interest.

According to equation (1) and equation (2), regions with lower group sample means $\hat{\mu}_k$, higher group sample variances $\hat{\sigma}_k$ and less observations $n_{1,k}$ will have a higher budget in the next iteration. It should be noticed that, in problems with discrete variables, a non-partitionable region is normally defined as a singleton region, in which the estimated sample variance $\hat{\sigma}_k$ is equal to zero. If this non-partitionable region k is not the current best group $\hat{b}$, then,

$$\hat{c}_k = 0, C_{k,\hat{b}} = 0, F(C_{k,\hat{b}}; n_{1,k} - 1, n_{1,\hat{b}} - 1) = 0 \quad \text{and} \quad n_k / n_{\hat{b}} = 0,$$

which means that no more budget will be allocated to this region. If this non-partitionable region is the current best group $\hat{b}$, then, $\hat{c}_k = 1/z_\alpha \neq 0$, which means that the proposed method will keep sampling from this region in subsequent iterations. In this case, additional actions can be taken to avoid repeat sampling from this non-partitionable region, such as increasing the value of $n_{\hat{b}}^{(r)}$ without really sampling from this region. Similar situations occur when the variance of one region is negligible (i.e., non-partitionable region) in problems with continuous domain.

3 NUMERICAL ANALYSIS

In this section, the proposed partition-based random search method is tested on the Rastrigin function and the Himmelblau's function. The Rastrigin function is characterized by a large number of high-quality local optima around the global optima, while the Himmelblau's function has four identical global optima. The PNP method described in Algorithm 1 is applied for comparison purposes as well as the pure random search (RS) method. The RS method is applied as a benchmark.

3.1 The Rastrigin Function

The proposed method is tested on the Rastrigin function (Rastrigin 1974):

$$f(x,y) = 20 + x^2 + y^2 - 10\cos(2\pi x) - 10\cos(2\pi y),$$

where the feasible domain is $x, y \in [-5.12, 5.12]$. The contour of the Rastrigin function is shown in Figure 1. Finding the minimum $f(0,0) = 0$ is difficult due to the large searching space and the large number of local optima.

Two partitioning rules are used. Partitioning rule A alternately partitions one region into two sub-regions evenly in the horizontal direction and the vertical direction. Partitioning rule B is the same as partitioning rule A, except three sub-regions are generated in each iteration. Define a non-partitionable region in this case as a region where all edges are less than 0.1% of their corresponding feasible ranges. Therefore, the maximal partition depth in partitioning rule A is 20 and in partitioning rule B is 14. In the studied case, intuitively, partitioning rule A partitions the feasible domain into two sub-regions, which are equivalent for the PNP method, in the first iteration. This may lead to a lot of backtracking during the optimization process, if the sample size in the complementary region n_2^{NP} is large. On the contrary, partitioning rule B seems to be a good partitioning rule for the PNP method, since a significant promising region is generated in each iteration.

The algorithm parameters of the PNP method and the proposed method are selected according to full factorial designs of experiment (DOE). The average value of the found optima among 500 replications is chosen as the response and the total budget size $N^{tot} = 2,000$. Table 1 and Table 2 show the DOE settings and the selected parameters for the PNP method and the proposed method, respectively.

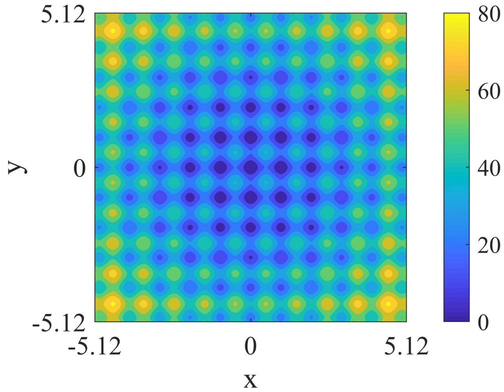

Figure 1: The contour of the Rastrigin function.

Table 1: The DOE setting (left) and the selected parameters (right) for the PNP method.

Parameters	Levels
n_1^{NP}	$20, 40, 60, 80, 100, 120, 140, 160$
n_2^{NP}	$20, 40, 60, 80, 100, 120, 140, 160$
Backtracking Rule	*I*, *II*, *III*

Partition Rule	A	B
n_1^{NP}	80	120
n_2^{NP}	20	100
Backtracking Rule	*I*	*III*

Table 2: The DOE setting (left) and the selected parameters (right) for the proposed method.

Parameters	Levels
α	$0.005, 0.01, 0.05, 0.1$
n_0	$5, 10, 15, 20$
Δ	$5, 10, 20, 30$
n_{max}	$30, 40, 50, 60$

Partition Rule	A	B
α	0.01	0.01
n_0	10	15
Δ	5	5
n_{max}	30	50

After selecting the algorithm parameters, 500 replications are executed with the total budget size increasing from $N^{tot} = 1{,}000$ to $N^{tot} = 5{,}000$. The optimal observation in one replication is collected and the boxplots of these values are presented in Figure 2 (partitioning rule *A* is adopted) and Figure 3 (partitioning rule *B* is adopted). The blue diamond, the red circle and the yellow triangle denote the means of the minimal observation obtained by the proposed method, the PNP method and the RS method, respectively. The first comment is that all methods' performance is improved as the total budget size increases, except for the PNP method using partitioning rule *A*, in which significant improvements cannot be observed after the total budget size reaches 3,000. This is because of the difficulty of escaping a local optimum with a low n_2^{NP} value. However, a high n_2^{NP} value will result in a lot of backtracking in the PNP method when partitioning rule *A* is used. When partitioning rule *B* is adopted, the performance of the PNP method is slightly improved compared to that using partitioning rule *A*. The proposed method dominates the other two methods. The global optimum is found with high probability as the total budget size reaches 2,000, if partitioning rule *A* is used, and 3,000, if partitioning rule *B* is used.

All the regions generated by the proposed method are presented in Figure 4, as well as all the most promising regions identified by the PNP method. The best replication (Figure 4 (a)) and the worst replication (Figure 4 (b)) among the 500 replications are selected. The total budget size $N^{tot} = 5{,}000$, partitioning rule A is used. The PNP method behaves more aggressively than the proposed method. This helps the PNP method to quickly direct to the optimum, if the real most promising region is identified correctly at each iteration (e.g., Figure 4 (a)). However, if multiple high-quality local optima exist, this feature also makes the PNP method more likely to be trapped in a local optimum if n_2^{NP} is small (e.g., Figure 4 (b)), or a lot of

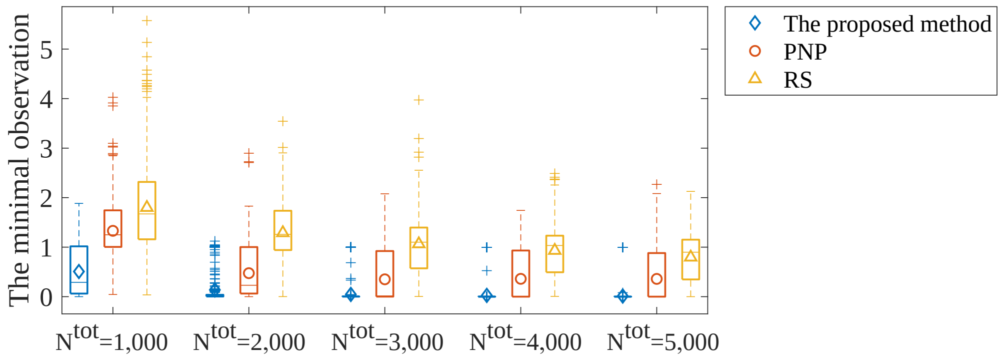

Figure 2: The boxplot of the minimal observation in each replication. 500 replications are executed for each experimental design. Partitioning rule A is adopted.

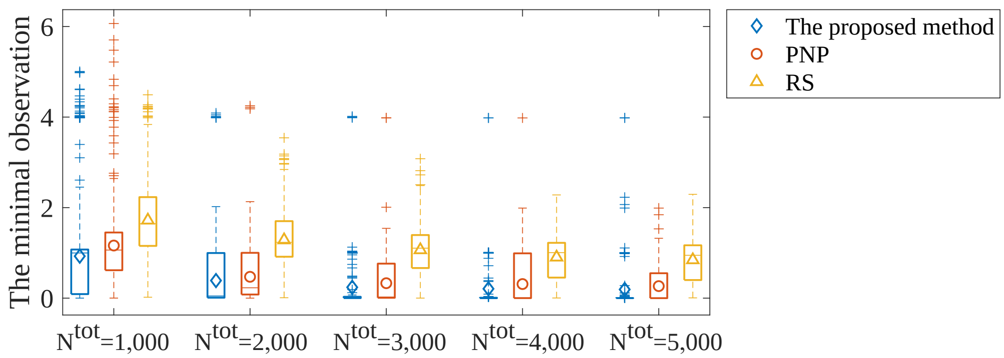

Figure 3: The boxplot of the minimal observation in each replication. 500 replications are executed for each experimental design. Partitioning rule B is adopted.

backtracking occur if n_2^{NP} is large. The proposed method is more conservative. It exploits also the regions containing the local optima with high quality (e.g., Figure 4 (a)), which makes the proposed method more robust than the PNP method. Conversely, if the total budget size is small, this feature may result in a poor performance, because more budget is used for exploration in earlier iterations. Under a certain probability, both methods miss the global optimum (e.g., Figure 4 (b)) due to the sampling noise, i.e., lack of knowledge about each region. In this problem, the rate of reaching the non-partitionable regions containing the global optimum are 98.6% and 50.2% for the proposed method and the PNP method, respectively.

Figure 5 shows the average partition depth, among 500 replications, at different positions, partitioned by the proposed method with total budget size $N^{tot} = 5,000$. The partition depth of one position represents the speed of reaching the non-partitionable size at this position. We can see that the region around the global optimum has the highest partition depth. The regions around the 9 high-quality local optima also have much higher partition depth than other regions.

3.2 The Himmelblau's Function

The Himmelblau's function (Himmelblau 1972):

$$f(x,y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$$

is considered in this section. The feasible domain is $x, y \in [-6, 6]$ and the contour is shown in Figure 6. The Himmelblau's function has four identical global optima $[x_1^*, y_1^*]^T = [3,2]^T$, $[x_2^*, y_2^*]^T \approx [-2.805, 3.131]^T$, $[x_3^*, y_3^*]^T \approx$

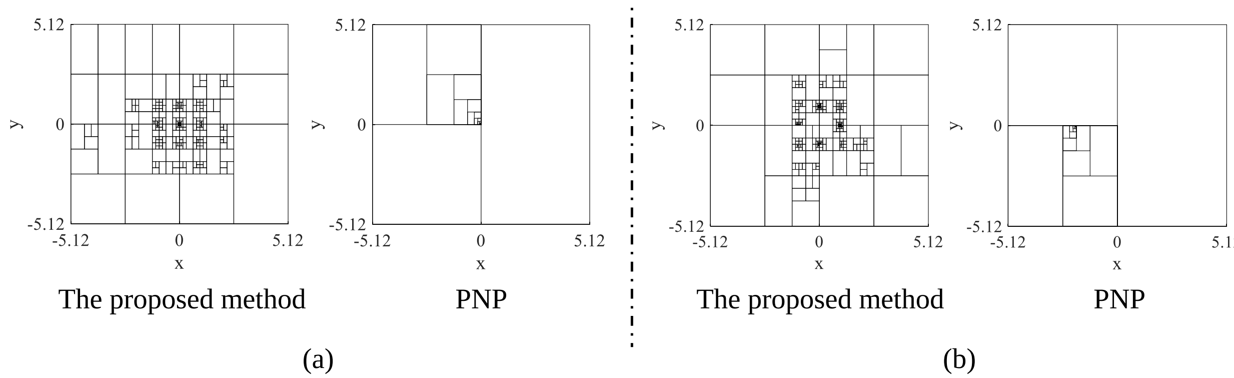

Figure 4: All the regions generated by the proposed method and all the most promising regions identified by the PNP method. The replication which has the best performance (a) and the worst performance (b) among the 500 replications are selected. The total budget size $N^{tot} = 5,000$ and partitioning rule A is applied.

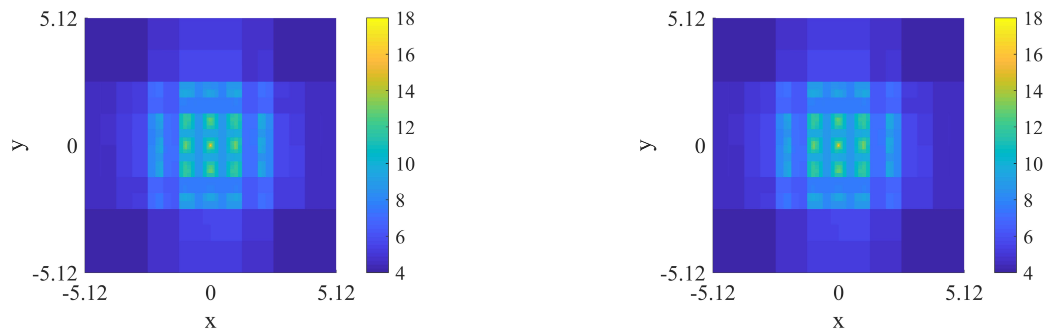

Figure 5: The average partition depth, among 500 replications, at different positions of the feasible domain. Partitioning rule A is used in the left figure and partitioning rule B is used in the right figure. The total budget size $N^{tot} = 5,000$.

$[-3.779, -3.283]^T, [x_4^*, y_4^*]^T \approx [3.584, -1.848]^T$ that $f(x_i^*, y_i^*) = 0, \forall i$. The positions of the four global optima are marked with red crosses in Figure 6. Let us set the acceptable half length $c = 0.006$ (the acceptable length is 0.1% of the variable range). If a point $[x, y]^T$ in the region that $x \in [x_i^* - c, x_i^* + c], y \in [y_i^* - c, y_i^* + c]$ is sampled (the objective values in this region are all less than 0.005), we say that the i-th global optimum is found.

The partitioning rule A described in the previous section is adopted in this case. The algorithm parameters of the PNP method and the proposed method are selected according to full factorial DOEs. The average value of the found optimum in each replication is chosen as the response and the total budget size $N^{tot} = 2,000$. The PNP method is applied with a single start. 500 replications are executed. Table 3 and Table 4 show the DOE settings and the selected parameters for the PNP method and the proposed method, respectively.

Figure 7 shows the average number of global optima found by different methods, as the total budget size increases. 500 replications are executed. Sampling points from the defined global optimum regions is difficult for the RS method because of the strict acceptable half length. The theoretical probability of sampling at least one point from one of the defined global optimal regions, by the RS method, is less than 20%, even if the total budget size is $N^{tot} = 50,000$. "PNP-k" indicates the results from the PNP method with k starts. When more starts are applied, the PNP method tends to find more of the defined global

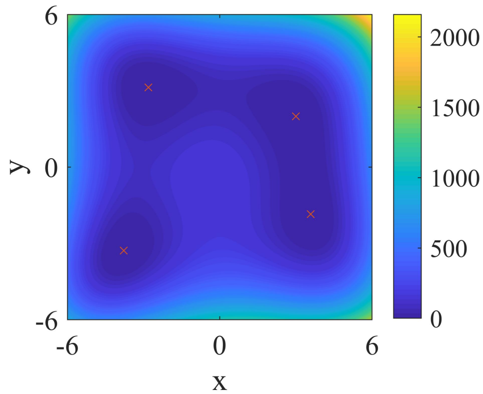

Figure 6: The contour of the Himmelblau's function and the four global optima.

Table 3: The DOE setting (left) and the selected parameters (right) for the PNP method.

Parameters	Levels
n_1^{NP}	$10, 20, 30, 40, 50, 60, 70, 80, 90, 100$
n_2^{NP}	$10, 20, 30, 40, 50, 60, 70, 80, 90, 100$
Backtracking Rule	*I*, *II*, *III*

Partition Rule	A
n_1^{NP}	20
n_2^{NP}	20
Backtracking Rule	*III*

Table 4: The DOE setting (left) and the selected parameters (right) for the proposed method.

Parameters	Levels
α	$0.005, 0.01, 0.05, 0.1$
n_0	$5, 10, 15, 20$
Δ	$5, 10, 20$
n_{max}	$20, 30, 40, 50$

Partition Rule	A
α	0.1
n_0	10
Δ	5
n_{max}	20

optima. If the PNP method has equal probability to find one of the defined optima by one start, the expected number of starts required to find all the four optima is 10.67. In this problem, the optima $[x_1^*, y_1^*]^T$ and $[x_4^*, y_4^*]^T$ have higher probabilities to be found by the PNP method than the optima $[x_2^*, y_2^*]^T$ and $[x_3^*, y_3^*]^T$, which means more starts may be needed. However, as the number of starts increases, the PNP method has poor performance when the budget size is small. This is because the budget used for exploitation is not enough since it is split to each start. Compared to the PNP method with multiple starts, the proposed method almost finds all the defined global optima with a much less budget size needed.

Figure 8 shows the average partition depth at different positions partitioned by the proposed method. The total budget size is equal to 4,000 and 500 replications are executed. As shown in Figure 8, the regions around the four global optima have much higher partition speeds than other regions.

4 CONCLUSION AND FUTURE WORK

This paper proposes a new partition-based random search method, in which identifying the most promising region is not required. The main concept is that all partitionable regions will be further partitioned as the search proceeds. The speed at which the region is partitioned is related to its approximate posterior probability of being the most promising region. Promising regions have higher partition speeds than non-promising ones. This allows multiple regions to be further exploited in parallel, if it is hard to determine which region is the real most promising region, and makes the algorithm less likely to be trapped at a

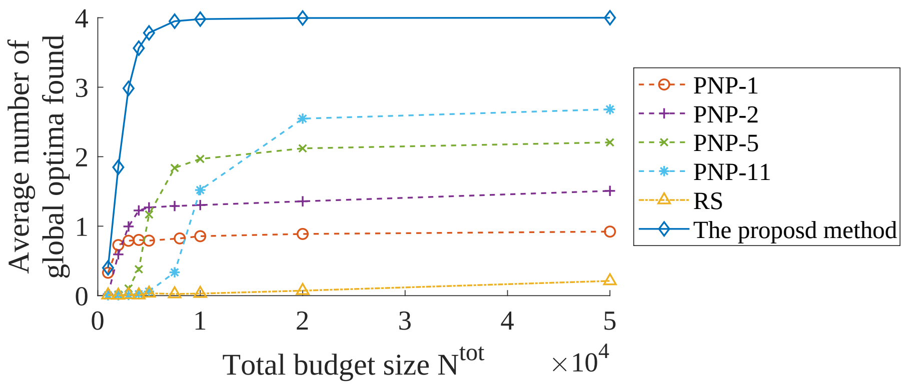

Figure 7: The average number of global optima found by different methods as the total budget size increases. 500 replications are executed. "PNP-k" indicates the results from the PNP method with k starts.

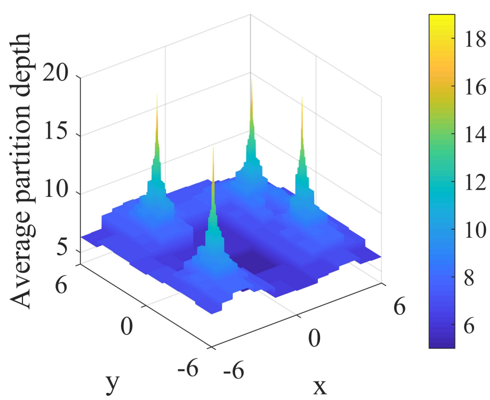

Figure 8: The average partition depth, among 500 replications, at different positions. The proposed method is used and $N^{tot} = 4,000$.

local optimum. Moreover, this feature enables the proposed method to deal with multimodal optimization problems, which is beyond the scope of the classic partition-based random search methods.

The proposed method is tested and compared to the pure NP method on the Rastrigin function, in which a large number of local optima exist, and the Himmelblau's function which has four identical global optima. The numerical results show that the proposed method finds the global optimum faster than the pure NP method, if a lot of high-quality local optima exist. The proposed method succeeds in finding all the global optima as the total budget size increases in the studied case. The numerical results also show that the proposed method controls the speed, at which each region is partitioned, as we expected.

Storing all generated regions helps the proposed method maintain a better global perspective. Nevertheless, as the search proceeds, a large amount of regions will be generated, which makes re-sampling not affordable. Although reusing the old data can increase the use of information, it reduces the method's ability to resist the noise introduced by sampling, which lowers the robustness of the method. It also requires a lot of computer memory.

The proposed method can be further improved in several directions. One is exploring alternative methods to increase the robustness of the proposed method to sampling noise. Currently it is done by modifying the number of observations in the budget allocation equations to increase the sampling probability in the regions with lower partition depth. Incorporating heuristic methods in the sampling phase can be another direction. Analogous to the NP method, a good sampling method may increase the efficiency of the proposed method. Also, adaptive parameters can be used. For example, the α value, which controls the trade-off between region means and region variances, can be varied during the search. The method can be extended to stochastic problems. In this case, another budget allocation strategy should be incorporated to decide the number of replications for each sampled solution.

REFERENCES

Chen, C.-H., J. Lin, E. Yücesan, and S. E. Chick. 2000. "Simulation Budget Allocation for Further Enhancing the Efficiency of Ordinal Optimization". *Discrete Event Dynamic Systems* 10(3):251–270.

Chen, W., S. Gao, C.-H. Chen, and L. Shi. 2014. "An Optimal Sample Allocation Strategy for Partition-Based Random Search". *IEEE Transactions on Automation Science and Engineering* 11(1):177–186.

Himmelblau, D. M. 1972. *Applied Nonlinear Programing*. New York, USA: McGraw-Hill, Inc.

Hong, L. J., and B. L. Nelson. 2005. "The Tradeoff Between Sampling and Switching: New Sequential Procedures for Indifference-Zone Selection". *IIE Transactions* 37(7):623–634.

Li, X. 2004, June. "Adaptively Choosing Neighbourhood Bests Using Species in A Particle Swarm Optimizer for Multimodal Function Optimization". In *Genetic and Evolutionary Computation Conference GECCO 2004*, edited by K. Deb, 105–116. Berlin, Heidelberg: Springer.

Lin, Z., A. Matta, and S. Du. 2018. "Initial Sampling Using Multi-Fidelity Information in Simulation Optimization of Manufacturing Systems". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 4212–4213. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Linz, D. D., H. Huang, and Z. B. Zabinsky. 2016. "A Quantile-Based Nested Partition Algorithm for Black-Box Functions on A Continuous Domain". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 590–601. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Luo, Y., and C. Yu. 2007. "An Improved Nested Partitions Algorithm Based on Simulated Annealing in Complex Decision Problem Optimization". In *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence. ICIC 2007.*, edited by D.-S. Huang, L. Heutte, and M. Loog, 572–583. Berlin, Heidelberg: Springer.

Rastrigin, L. A. 1974. *Extremal Control Systems*. Theoretical Foundations of Engineering Cybernetics Series. Moscow: Nauka.

Shi, L., and S. Ólafsson. 2000. "Nested Partitions Method for Global Optimization". *Operations Research* 48(3):390–407.

Shi, L., and S. Olafsson. 2009. *Nested Partitions Method, Theory and Applications*. New York, USA: Springer.

Thomsen, R. 2004. "Multimodal Optimization Using Crowding-Based Differential Evolution". In *Proceedings of the 2004 Congress on Evolutionary Computation*, Volume 2, 1382–1389. Portland, Oregon: Institute of Electrical and Electronics Engineers, Inc.

Zhu, Y., G. Pedrielli, and L. H. Lee. 2019. "TD-OCBA: Optimal Computing Budget Allocation and Time Dilation for Simulation Optimization of Manufacturing Systems". *IISE Transactions* 51(3):219–231.

AUTHOR BIOGRAPHIES

ZIWEI LIN is a joint PhD candidate of Department of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, China, and Department of Mechanical Engineering, Politecnico di Milano, Italy. Her thesis topic is system performance evaluation and optimization using multi-fidelity information. Her email address is linziwei@sjtu.edu.cn.

ANDREA MATTA is a professor at the Department of Mechanical Engineering at Politecnico di Milano, Italy. He currently teaches integrated manufacturing systems and manufacturing. His research area includes analysis and design of manufacturing and health care systems. He is Editor-in-Chief of Flexible Services and Manufacturing Journal. His email address is andrea.matta@polimi.it.

SHICHANG DU is a professor of Department of Industrial Engineering and Management, School of Mechanical Engineering, Shanghai Jiao Tong University, China. His research interest is about product quality modeling and control in complex manufacturing processes. His email address is lovbin@sjtu.edu.cn.