

Error Analysis for Designed Test and Numerical Integral by Using UED in Material Research

Maosheng Zheng*, Jie Yu, Haipeng Teng, Yi Wang

Abstract: In this article, the error analysis for designed test and definite integral in employing uniform experimental design is analogically developed on basis of midpoint rule in rectangle method for assessing definite integral through conducting discretized sampling approximately. It concluded that the discrete sampling-point by means of good lattice point in the evaluations of definite integral and maximum value of a function is promised with higher accuracy, and the predicted entire error of this uniform sampling point method for above problems decreases with the number of sampling points significantly.

Keywords: definite integral; error analysis; midpoint rule; rectangle method; uniform experimental design

1 INTRODUCTION

Experiment design is an essential topic to scientific and industrial developments. How do we arrange and design experiments to get effective results with less number of trials? Some approaches are proposed to answer this encountered question frequently, such as *Response Surface Methodology*, *Orthogonal Experimental Design*, and *Uniform Experimental Design*, etc. Better design could lead to results that are more effective.

In 1978, due to the need for missile designs, a special demand for total trial number being no greater than 50 with 18 factor levels of a five-factor experiment was faced to Prof. K. T. Fang [1], He and Prof. Y. Wang thus proposed a novel solution by employing number-theoretical methods [2]. They created a brand new experimental design methodology to conduct the design of the problem in that time, which was named as Uniform Experimental Design (UED). UED is attributed to number - theoretical method or quasi-Monte Carlo method. This novel methodology was employed to the design of missiles successfully, and got series of meaningful achievements in China due to its wide applications [1, 2].

Early in 1950s, Ulam and Von Neumann developed Monte Carlo method, which is a statistical stimulation. The main idea of their method is to convert an analysis problem into a probability problem and then employ a statistical simulation to deal with the problem to get a same solution. This seems an effective solution for some difficult analysis problems, including the approximate estimation of complicated definite integrals. The general idea of Monte Carlo method is the need of a set of stochastic numbers to enable to perform this statistical simulation. The precision of this method strongly depends on the independence and uniformity of stochastic numbers.

Almost in the same period of 1950s, deterministic methods were also proposed to deal with some difficult analysis problems by some mathematicians, which aimed to give solution by using uniformly distributed points in space instead of random numbers like that in Monte Carlo method [3], such as, Korobov put forward the concept of a point set, which is uniformly distributed. Since then Hua et al developed the good lattice point (GLP) method in 1960s for

evaluation of definite integral approximately, which is with low-discrepancy on basis of number theory [2, 3]. Therefore, this kind of method is called a number-theoretical method or quasi-Monte Carlo method naturally thereafter. UED can be seen as one of the successful applications of the number-theoretical method [1, 2]. This methodology was subsequently used in evaluations for approximation of multiple-integral successfully.

There are many beneficial features of GLP [4-6], besides the uniformity of distribution of sampling-point over the specific domain and good space-filling characteristic.

Currently, the UED finds its wide applications over the world, it spreads in designs of Chinese medicine, chemical reaction and missile, as well as Ford Motor Co. Ltd for its design of standard exercises and automotive as computer experiments for providing a support of the preliminary design of production [7].

1.1 Essential Characteristics of UED

The essential characteristics of UED involves [1, 2, 7]:

A) Homogenization

The distribution of specimen point is homogeneously scattered in the variable space; therefore it gains a surname of "space filling design" occasionally [1, 2, 7]. A number of "Uniform Design Table" (UDT) was specifically developed by Fang to arrange the distribution of the specimen point for UED [8], which is fully deterministic.

B) Entire Mean Model

UED intends to get an outcome consequence, which is with minimum deviation of the entire averaged value from the actual total averaged value through uniform distribution of specimen points.

C) Robustness

UED is expected to be used in many cases with robustness regardless of variation of model.

1.2 Basic Principle of UED

The fundamental principle of UED is as followings:

1) Entire Mean Model

The fundamental hypothesis is the existence of a deterministic relationship of the response g vs. the independent input variables $x_1, x_2, x_3, \dots, x_s$, the formula of the response can be expressed as,

$$g = G(x_1, x_2, x_3, \dots, x_s), x = \{x_1, x_2, x_3, \dots, x_s\} \in C^s. \quad (1)$$

The further hypothesis is that the entire averaged value of the response g on $C^s = [0, 1]^s$ is,

$$\overline{E(g)} = \int_{C^s} G(x_1, x_2, x_3, \dots, x_s) dx_1 dx_2 dx_3, \dots, dx_s. \quad (2)$$

Moreover, if one takes m sampling points $q_1, q_2, q_3, \dots, q_m$ on C^s to conduct an average value of g , then the averaged value of g over these m specimen points is,

$$\overline{g(D_m)} = \frac{1}{m} \sum_{i=1}^m G(q_i). \quad (3)$$

In Eq. (3), $D_m = \{q_1, q_2, q_3, \dots, q_m\}$ represents a design of such m specimen points.

Fang et al indicated that the deviation $\varepsilon = \overline{E(g)} - \overline{g(D_m)}$ of the specimen point set on C^s and D_m will be quite small provided the specimen points $q_1, q_2, q_3, \dots, q_m$ are uniformly distributed in the domain C^s .

2) Uniform Design Table

In order to provide an appropriate application for UED, a number of UDT as well as the "Utility Table" were conducted [8], with which the location of specimen points can be specifically determined with convenience.

3) Regression

In general, under condition of discretization with sampling points, an approximate expression for response $r' = R'(x_1, x_2, x_3, \dots, x_m)$ vs. the independent input variables can be regressed to reveal the resemble formation.

1.3 Aim of This Study

Until now, the estimation of the accuracy of applying each UDT to conduct actual problem is unclear though the discrepancy of each point set of UDT is provided by Fang in his book [5].

In this paper, the study of entire error of definite integral and maximum value by using sampling points from UDT is preliminarily conducted in the point of view of practical application.

2 ERROR ANALYSIS FOR APPLYING UED TO CONDUCT ACTUAL PROBLEMS

2.1 Distribution Rule of UED in Space

A uniform table $U_n(n^q)$ has n rows and q columns, and q is the Euler function for a positive integer n , $q = \varphi(n) = n \cdot (1 - 1/p_1) \cdot (1 - 1/p_2) \cdot \dots \cdot (1 - 1/p_v)$ for each n . According to number theory, it assumes that there is a unique prime decomposition $n = p_1^{r_1} \cdot p_2^{r_2} \cdot \dots \cdot p_v^{r_v}$ for each n [1, 2, 8], in which r_1, r_2, \dots, r_v express positive integers and p_1, p_2, \dots, p_v

indicate different primes. Furthermore, the number of independent factors at most is $t = \varphi(n)/2 + 1$, i.e., $s \leq t$ for each n , s is the actual number of the independent variables of the studied problem [1, 2, 8].

In the hyper cubic $[0, 1]^s$, the specimen point is homogeneously distributed. Especially, for one independent variable case, the design $X^* = [1/2n, 3/2n, 5/2n, \dots, (2n-3)/2n, (2n-1)/2n]^T$ is the unique design on $[0, 1]$ with low - discrepancy or under star discrepancy of $D^*(X^*) = 1/2n$ [1, 2, 8].

Obviously, the homogeneously spreading of the specimen point in the above design for one independent variable case is the same as that of the Midpoint Rule in rectangle method of definite integral [9-11].

Following midpoint rule, the definite integral $E(g) = \int_{x_0-\delta/2}^{x_0+\delta/2} g(x) dx$ around position x_0 with a subsection

δ is approximated by $I = g(x_0) \cdot \delta$, which is with the local error of $\varepsilon_M = \delta^3 \cdot g''(x_0)/24$ [9-11], where $g''(x_0)$ indicates the second derivative at location x_0 , in general ε_M is negligible as δ is sufficiently small.

According to mean value theorem of integral, one could always find a proper position ζ within $[x_0 - \delta/2, x_0 + \delta/2]$, such that the value of the function $g(\zeta)$ meets the demand of

$$g(\zeta) \cdot \delta = E(g) = \int_{x_0-\delta/2}^{x_0+\delta/2} g(x) dx. \quad \text{However, above discussion}$$

indicates that the value of the integral approximates to $I = g(x_0) \cdot \delta$ with the small local error of $\varepsilon_M = \delta^3 \cdot g''(x_0)/24$ [9-11], therefore, $g(\zeta)$ approximately equals to $g(x_0)$, i.e., $g(\zeta) \approx g(x_0)$.

Furthermore, the local error of the function $g(x)$ around x_0 within area of $[x_0 - \delta/2, x_0 + \delta/2]$ is about $g'(x_0) \cdot \delta/2$.

Moreover, considering the maximum value of the function $g(x)$ within its range of $[a, b]$, it supposes that if the function $g(x)$ within its range $[a, b]$ derives its maximum value $g(x_l)$ at a discrete point $x_l = a + (2l - 1)(b - a)/2n$, $l \in [1, 2, 3, \dots, n]$, then the realistic error of the actual maximum of this function $g_{\max}(x)$ from the nominal maximum $g(x_l)$ at a discrete point x_l is $E_{\text{actual}} = g_{\max}(x) - g(x_l)$, and it could be approximately estimated by,

$$E_{\text{est.}} = g'(x_l) \cdot \delta/2 \approx |g(x_{l+1}) - g(x_l)|/2. \quad (4)$$

Table 1 Distribution of 11 sampling points within domain [0, 1.5]

No.	Position	e^x
1	0.0682	1.0706
2	0.2045	1.2270
3	0.3409	1.4062
4	0.4773	1.6117
5	0.6136	1.8471
6	0.7500	2.1170
7	0.8864	2.4263
8	1.0227	2.7808
9	1.1591	3.1871
10	1.2955	3.6527
11	1.4318	4.1863

As an example, lets' conduct the error analysis for e^x within range of $[0, 1.5]$ by using 11 discrete uniform sampling points [12, 13]. As to such problem, following the

procedure of UED [1, 2, 8], the 11 sampling points are uniformly distributed within range of [0, 1.5], see Tab. 1.

The actual value of function e^x is 4.4817 at $x = 1.5$, while the maximum value of discretized function e^x in Tab. 1 is 4.1863. The actual error is $E_{\text{actual}} = 0.2954$, while the estimated value by using Eq. (4) is $E_{\text{est.}} = 0.2668$, which is close to E_{actual} .

2.2 Error Analysis of Applying UED for Maximum Value

In general, as to a s -dimensional problem, it assumes that the discretization is conducted for a function \bar{x}_p within its domain $[0, 1]^s$, and the maximum $g(\bar{x}_p)$ of this function $g(\bar{x})$ is at the discrete point \bar{x}_p due to this discretization, then the error of the maximum $g(\bar{x}_p)$ at the discrete point \bar{x}_p from the actual maximum $g_{\text{max}}(\bar{x})$ of the function $g(\bar{x})$ due to this discretization is $E_{\text{actual}} = g_{\text{max}}(\bar{x}) - g(\bar{x}_p)$, and it can be estimated by following equation,

$$E_{\text{est.}} \approx \frac{1}{2\gamma} \sum_{i=1}^{\gamma} |g(\bar{x}_{p+i}) - g(\bar{x}_p)|. \tag{5}$$

In Eq. (5), γ is the number of nearest neighbour of \bar{x}_p .

Thus, the error of the maximum $g(\bar{x}_p)$ at the discrete point \bar{x}_p from the actual maximum $g_{\text{max}}(\bar{x})$ of the function $g(\bar{x})$ corresponding to the discretization can be estimated by using Eq. (5) in principle.

2.3 Error Analysis of Applying UED for Integral

As to midpoint rule [9-11], the definite integral $\overline{E(g)} = \int_a^b g(x)dx$ can be approximated by a summation, i.e.,

$$g(D_n) = \frac{(b-a)}{n} \sum_{i=1}^n g[a + (i-1/2) \cdot (b-a)/n] \quad \text{in } 1\text{-dimension, while the entire error is,}$$

$$E_M = \frac{1}{24} \cdot \frac{(b-a)^3}{n^2} \cdot \frac{1}{n} \sum_{i=1}^n g''[a + (i-1/2) \cdot (b-a)/n] \approx \frac{(b-a)^2}{24n^2} \cdot \{g'[b - (b-a)/2n] - g'[a + (b-a)/2n]\}, \tag{6}$$

where $g''(a + (i-1/2) \cdot (b-a)/n)$ is the 2nd derivative at position " $a + (i-1/2) \cdot (b-a)/n$ ", and $g'[a + (b-a)/2n]$ expresses the 1st derivative at position " $a + (b-a)/2n$ ".

If the integrand $g(x)$ behaves a wavy form, the summarizing $\varepsilon_n = \sum_{i=1}^n g''[a + (i-1/2) \cdot (b-a)/n]$ is quite tiny; or else the summarizing ε_n remains for some monotone function like e^x . However even in latter case good precision could be obtained with 11 specimen points which are homogeneously distributed [12, 13]. Take the integral $\int_0^{1.5} e^x dx$ as an example, which is with the precise value of

$\overline{E(g)} = \int_0^{1.5} e^x dx = 3.4817$, while the summarizing of the discrete specimen points in manner of GLP is $\overline{g(D_{11})} = \frac{1.5}{11} \sum_{i=1}^{11} e^{[(i-0.5) \cdot 1.5/11]} = 3.4790$, the actual total error is 0.0027, while the predicted value of total error is of 0.0025, which is not far from the actual error.

For higher dimensions, the entire error could be estimated analogically by,

$$|E_M| \leq \frac{1}{24 \cdot s \cdot n} \sum_{i=1}^s (b_i - a_i)^3 \cdot M(l). \tag{7}$$

In Eq. (7), the term $M(l)$ indicates $\max|g''(x)|$ in l^{th} independent variable, $x_l \in [a_l, b_l]$.

Until now, the error estimation of applying UED for integral is estimated in principle.

3 APPLICATIONS

3.1 Error Analysis of Maximum Value of Functions in Their Domains by Using Discrete Uniform Sampling Points

Lets' study the function $f_1(x, y) = \ln(x + 2y)$ in the domain of $[1.4, 2.0] \times [1.0, 1.5]$ by using discrete uniform sampling points first. Assume that the actual maximum value of the function $f_1(x, y)$ within the domain is $f_{1\text{max}}(x, y)$, and the maximum value of the function $f_1(x, y)$ due to the discretization is at a discrete point \bar{x}_p and denoted by $f_1(\bar{x}_p)$.

Fig. 1 shows the variations of the actual error $E_{\text{actual}} = |f_{1\text{max}}(x, y) - f_1(\bar{x}_p)|$ vs. the number of sampling points n , together with the estimated error $E_{\text{est.}}$. It can be seen that the tendency of the variation of the actual error E_{actual} is the same as the those of the estimated error $E_{\text{est.}}$, which decreases significantly with the increase of the number of sampling points n .

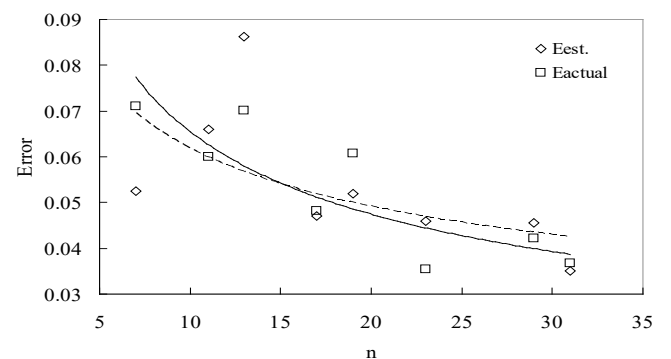


Figure 1 Comparison of the variations of the actual error E_{actual} and the estimated error $E_{\text{est.}}$ vs. number of sampling points n for $f_1(x, y) = \ln(x + 2y)$

Next, Lets' study the function $f_2(x, y) = 1 + 2x^2 + 2y^3$ in the domain of $[1.4, 2.0] \times [1.0, 1.5]$ by using discrete uniform sampling points. Fig. 2 shows the comparison of the variations of the actual error E_{actual} and the estimated error $E_{\text{est.}}$ vs. the number of sampling points n for $f_2(x, y)$. It can be seen again that the tendency of the variations of the actual error E_{actual} and the estimated error $E_{\text{est.}}$ is the same, which

decreases with the increase of the number of specimen points n obviously. The accuracy of the estimation varies with the exact detail of the function $f(x, y)$.

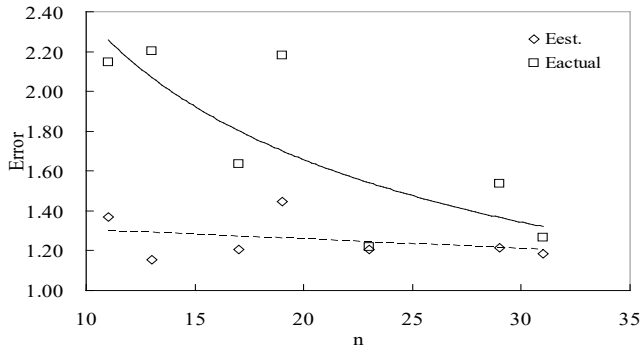


Figure 2 Comparison of the variations of the actual error E_{actual} and the estimated error E_{est} . vs. number of sampling points n for $f_2(x, y) = 1 + 2x^2 + 2y^3$

3.2 Error Analysis of Definite Integral in 2-D

Under condition of higher dimensions, Fang set up a number of UDTs and the corresponding utility tables specifically for appropriate utility of UED, which is based on GLP and number-theoretic methods [1, 2, 8].

Now, let's take a definite integral of a 2-dimensional problem as an example, i.e.,

$$H = \int_{x=1.4}^{2.0} dx \int_{y=1.0}^{1.5} H(x, y) dy \equiv \int_{x=1.4}^{2.0} dx \int_{y=1.0}^{1.5} \ln(x + 2y) dy. \quad (8)$$

The UDT $U_{17}(17^8)$ is tried to be used to conduct the definite integral, 17 sampling points are included [12, 13].

The accurate data of this integration is 0.429560 [14].

The specimen points of $U_{17}(17^8)$ within the integral range $[1.4, 2.0] \times [1.0, 1.5]$ are distributed uniformly and shown in Tab. 2, the values of function $H(x, y)$ at each discrete location are displayed in Tab. 2 as well. The symbols x_{10} and y_{20} in Tab. 2 show the nominal locations of the original table $U_{17}(17^8)$ within the domain $[1, 17] \times [1, 17]$ [2, 3].

Table 2 Positions of the specimen points within domain $[1.4, 2.0] \times [1.0, 1.5]$ by means of $U_{17}(17^8)$

No. = x_0	y_0	x	y	H
1	11	1.4176	1.3088	1.3951
2	5	1.4529	1.1324	1.3131
3	16	1.4882	1.4559	1.4816
4	10	1.5235	1.2794	1.4067
5	4	1.5588	1.1029	1.3257
6	15	1.5941	1.4265	1.4922
7	9	1.6294	1.25	1.4181
8	3	1.6647	1.0735	1.3381
9	14	1.7	1.3971	1.5028
10	8	1.7353	1.2206	1.4295
11	2	1.7706	1.0441	1.3504
12	13	1.8059	1.3676	1.5132
13	7	1.8412	1.1912	1.4407
14	1	1.8765	1.0147	1.3625
15	12	1.9118	1.3382	1.5235
16	6	1.9471	1.1617	1.4518
17	17	1.9824	1.4853	1.6000

In according with the method of UED [1, 2, 8], the integration H in the domain $[1.4, 2.0] \times [1.0, 1.5]$ becomes a summation in the discrete specimen points as

$$H \approx \frac{0.6 \times 0.5}{17} \sum_{i=1}^{17} H(x_i, y_i). \quad (9)$$

Summarizing Eq. (9) acquires a result, says 0.429613, which causes an error of 5.3×10^{-5} as respect to its actual result of 0.429560 [14], while the estimated error from the sampling points of the uniform design is 7.6×10^{-5} .

3.3 Comparative Assessment of Discrete Specimen Points by Means of GLP with Monte Carlo Simulation for Integral

$$F \equiv \int_0^6 f(x) dx = \int_0^6 (2 - x/3) dx$$

The accurate value of the definite integral of $F = \int_0^6 (2 - x/3) dx$ is 6.

Han and Ren once assessed this integral by employing Monte Carlo algorithm [15].

Now, let's try to re-study this integral by means of GLP with 11 discrete specimen points for comparison [15]. The locations of the specimen points and the discrete values of the function $f(x)$ are given in Tab. 3.

Table 3 Locations of the specimen points and the discrete data of the function

No.	x	$f(x)$
1	0.2727	1.9091
2	0.8182	1.7273
3	1.3636	1.5455
4	1.9091	1.3636
5	2.4545	1.1818
6	3	1
7	3.5455	0.8182
8	4.0909	0.6365
9	4.6364	0.4545
10	5.1818	0.2727
11	5.7273	0.0909

Following the procedure of UED, the definite integral F now becomes a summation of the 11 discrete specimen points within the integral domain $[0, 6]$. The summation gives a value of $\frac{6}{11} \sum_{i=1}^{11} f(x_i) = 6$, which is exactly and luckily same as that of the accurate value of this integral. However, the use of Monte Carlo simulation results in a varying error, which changes even up to 1000 random specimen points [15]. As was shown in [15], the Monte Carlo simulation gives an error of 0.1214 at 1000 specimen points, while the error of the simulated result reaches to 0.2908 with 100 specimen points [15]! This result reflects the merit of the assessment with discrete specimen points by means of GLP in evaluating definite integral and maximum value of a function once more.

4 CONCLUSION

Error analysis of designed test and definite integral by employing UED to conduct discrete specimen is analyzed in

this paper, the estimation of error is performed with the aid of midpoint rule in rectangle method for predicting definite integral analogically. The study indicates the decreasing tendency of the entire error of specimen point in predictions of definite integral and maximum value of a function vs. the number of specimen points, and the proper number of specimen points can be determined by the promised requirement of accuracy conversely.

Conflict Statement

There is no conflict of interest.

5 REFERENCES

- [1] Fang, K. T., Liu, M. Q., Qin, H., & Zhou, Y. D. (2018). *Theory and Application of Uniform Experimental Design*. Science Press, Beijing, China, Springer Nature, Singapore. <https://doi.org/10.1007/978-981-13-2041-5>
- [2] Fang, K-T. & Wang, Y. (1994). *Number-theoretic Methods in Statistics*. Chapman & Hall, London, UK. <https://doi.org/10.1007/978-1-4899-3095-8>
- [3] Hua, L-K. & Wang, Y. (1981). *Applications of Number Theory to Numerical Analysis*. Berlin, New York: Springer-Verlag & Beijing: Science Press.
- [4] Elsayah, A. M., Fang, K.-T., & Deng, Y. H. (2021). Some interesting behaviors of good lattice point sets. *Communications in Statistics - Simulation and Computation*, 50(11), 3650-3668. <https://doi.org/10.1080/03610918.2019.1628988>
- [5] Yang, L., Zhou, Y., & Liu, M.-Q. (2021). Maximin distance designs based on densest packings. *Metrika*, 84, 615-634. <https://doi.org/10.1007/s00184-020-00788-w>
- [6] Huang, Y. & Zhou, Y. (2022). Convergence of uniformity criteria and the application in numerical integration. *Mathematics*, 10, 3717. <https://doi.org/10.3390/math10193717>
- [7] Liu, M. Q., Lin, D. K. J., & Zhou, Y. (2020). The Contribution to Experimental Designs by Kai-Tai Fang. In *Contemporary Experimental Design, Multivariate Analysis and Data Mining, Festschrift in Honour of Professor Kai-Tai Fang*, eds. by J. Fan, J. Pan, Springer, 21-35, Cham, Switzerland. <https://doi.org/10.1007/978-3-030-46161-4>
- [8] Fang, K. T. (1994). *Uniform Design and Uniform Design Table*. Science Press, Beijing, China.
- [9] Holmes, M. H. (2016). *Analysis Introduction to Scientific Computing and Data Analysis*. Springer Nature Switzerland AG, Cham, Switzerland. <https://doi.org/10.1007/978-3-319-30256-0>
- [10] Izaac, J. & Wang, J. (2018). *Computational Quantum Mechanics*. Springer Nature Switzerland AG, Cham, Switzerland. <https://doi.org/10.1007/978-3-319-99930-2>
- [11] Stickler, B. A. & Schachinger, E. (2016). *Basic Concept in Computational Physics*, 2nd Ed. Springer Nature Switzerland AG, Cham, Switzerland. <https://doi.org/10.1007/978-3-319-27265-8>
- [12] Yu, J., Zheng, M., Teng, H., & Wang, Y. (2022). An efficient approach for calculating a definite integral with about a dozen of sampling points. *Vojnotehnicki Glasnik*, 70(2), 340-356. <https://doi.org/10.5937/vojtahg70-36029>
- [13] Zheng, M., Teng, H., Yu, J., Cui, Y., & Wang, Y. (2023). *Probability-Based Multi-objective Optimization for Material Selection*. Springer Science and Business Media LLC, Singapore. <https://doi.org/10.1007/978-981-19-3351-6>
- [14] Song, S. & Chen, S. (2004). Two Effective Quadrature

Schemes for Calculating Double Integration. *Journal of Zhengzhou University*, 36, 16-19.

- [15] Han, J. & Ren, W. (2007). Monte Carlo Integration and Quasi-Monte Carlo Integration. *Journal of Shanxi Normal University (Natural Science Ed.)*, 121(1), 13-17.

Authors' contacts:

Maosheng Zheng

(Corresponding author)

School of Chemical Engineering, Northwest University, Xi'an, 710069, China

E-mail: mszhengok@aliyun.com

Jie Yu

School of Life Science, Northwest University,

Xi'an, 710069, China

Haipeng Teng

School of Chemical Engineering, Northwest University,

Xi'an, 710069, China

Yi Wang

School of Chemical Engineering, Northwest University,

Xi'an, 710069, China