Chapter 7
Construction of Uniform Designs on Arbitrary Domains by Inverse Rosenblatt Transformation

Mei Zhang, Aijun Zhang, and Yongdao Zhou

Abstract The uniform design proposed by Fang [6] and Wang and Fang [17] has become an important class of designs for both traditional industrial experiments and modern computer experiments. There exist established theory and methods for constructing uniform designs on hypercube domains, while the uniform design construction on arbitrary domains remains a challenging problem. In this paper, we propose a deterministic construction method through inverse Rosenblatt transformation, as a general approach to convert the uniformly designed points from the unit hypercubes to arbitrary domains. To evaluate the constructed designs, we employ the central composite discrepancy as a uniformity measure suitable for irregular domains. The proposed method is demonstrated with a class of flexible regions, constrained and manifold domains, and the geographical domain with very irregular boundary. The new construction results are shown competitive to traditional stochastic representation and acceptance-rejection methods.

7.1 Introduction

The uniform design of experiments has generated a great amount of research papers and impact cases ever since it was first proposed by Fang [6] and Wang and Fang [17]. It has been successfully used for both traditional industrial experiments and modern computer experiments; see the monographs [7, 9].

M. Zhang
College of Mathematics, Sichuan University, Chengdu 610064, China
e-mail: zhangmei150320@163.com

A. Zhang (✉)
Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China
e-mail: ajzhang@hku.hk

Y. Zhou
School of Statistics and Data Science & LPMC, Nankai University, Tianjin 300071, China
e-mail: ydzhou@nankai.edu.cn

J. Fan and J. Pan (eds.), *Contemporary Experimental Design, Multivariate Analysis and Data Mining*,
https://doi.org/10.1007/978-3-030-46161-4_7

To construct uniform designs on the unit hypercube $C^s = [0, 1]^s$, there exist both theoretical approaches and numerical optimization methods; see the latest book of Fang et al. [8] for a complete treatment. Among these methods, the classical good lattice point (GLP) method based on number theory is simple yet effective, and it is also widely used in quasi-Monte Carlo sampling [14]. For the GLP method with respect to the classical star-discrepancy criterion, Fang and Wang [9] (Appendix A) provides a catalogue of generating vectors up to 18 dimensions. For two-dimensional uniform designs in particular, it is well-known that the GLPs generated through Fibonacci numbers enjoy the low star-discrepancy properties. However, it is not clear whether such Fibonacci designs also enjoy the low centered-ℓ_2 discrepancy (CD2) properties.

It is of our interest to construct the uniform designs on arbitrary experimental domains, including regular and irregular regions. For regular regions such as ball, sphere and simplex, Fang and Wang [9] suggested the inverse transformation method through establishing a non-trivial analytic stochastic representation (SR) for the random vector uniformly distributed on each regular region, and then generating the uniform designs by inversely mapping from the unit hypercubes. For mixture experiments with single-factor constraints, Fang and Yang [10] proposed a conditional distribution method that also takes a non-trivial SR form. This method is further applied by Tian et al. [16] for generating uniform designs on tetragon and convex polyhedrons. For an irregular region $\mathscr{X} \subset \mathbb{R}^s$ that does not have the explicit SR form, one usually resorts to the acceptance-rejection (AR) method that first generates uniform designs on a superset hypercube $\mathscr{C} \supseteq \mathscr{X}$, then retains only the design points within the region of interest. Such AR method was suggested by Borkowski and Piepel [2] for mixture experiments with complex multi-factor constraints. However, the AR method is less efficient especially when $\mathscr{X}$ is much smaller than $\mathscr{C}$, and the resulting design in $\mathscr{X}$ sometimes has poor uniformity.

The construction of uniform designs on arbitrary domains remains a challenging problem. Numerically, one may use the stochastic optimization methods to directly search for the design points according to a certain uniformity criterion. Chuang and Hung [5] proposed a central composite discrepancy (CCD) to measure uniformity with regard to an arbitrary domain $\mathscr{X}$, then used a switching algorithm to search for the best design over a set of pre-specified points. Also based on the CCD criterion, Lin et al. [13] applied the threshold accepting algorithm to optimize the U-type designs on flexible regions, and Chen et al. [3] developed the discrete particle swarm optimization algorithm with GPU acceleration. Other space-filling criteria can also be used to directly measure the uniformity on irregular regions, e.g. the maximin distance used by Chen et al. [4].

In this paper, we study a deterministic method based on the inverse Rosenblatt transformation (IRT) for constructing uniform designs on arbitrary domains. It can be viewed as a special kind of inverse method, as is remarked by Fang and Wang [9, p. 54]. To distinguish the IRT from the existing inverse method based on the specific analytical SR, we call the latter as the SR method in this paper. Unlike the SR methods reviewed above, the IRT method does not necessarily require the analytical forms of conditional distribution functions, which can be easily approximated for a uniform experimental domain with irregular boundary. The IRT method includes the conditional distribution method by Fang and Yang [10] as a special case for restricted mixtures. We demonstrate how the proposed method can be used to construct uniform designs on a class of flexible regions, constrained and manifold domains, and the irregular domains such as geographical maps. Among the uniformity criteria, we employ the aforementioned CCD to evaluate the constructed designs on general domains. For regular regions, the construction results by the IRT method are compared with both SR and AR methods; while for irregular regions, they are compared with the AR method.

The rest of this paper is organized as follows. In Sect. 7.2 we propose the IRT method based on the marginal and conditional distributions subject to permutation, and illustrate it through a synthetic example. Section 7.3 presents the construction results on a variety of regular and irregular domains. Some concluding remarks are given in Sect. 7.4. In the Appendix, we provide a brief review of the GLP method for constructing uniform designs on hypercube domains, which are used by the proposed IRT method. We show that the leave-one-out Fibonacci designs achieve the minimum centered ℓ_2-discrepancy for the mixed GLP method.

7.2 Inverse Rosenblatt Transformation Method

The Rosenblatt transformation [15] is a general mapping of multivariate random variables with a continuous distribution to the uniform distribution on unit hypercubes. It can be used as a tool for construction of multivariate distributions and goodness-of-fit testing; see e.g. Justel et al. [12] and Arnold et al. [1].

Let $X \in \mathscr{X} \subseteq \mathbb{R}^s$ be a random vector with joint density

$$f(x_1, \ldots, x_s) = f_1(x_1) f_{2|1}(x_2|x_1) \cdots f_{s|1,\ldots,s-1}(x_s|x_1, \ldots, x_{s-1}). \tag{7.1}$$

Denote F_1 as the marginal cumulative distribution function (CDF) of the first component X_1, and by $F_{2|1}, \ldots, F_{s|1,\ldots,s-1}$ the consecutive conditional CDFs. Then, the Rosenblatt transformation (RT) is defined by

$$\begin{cases} U_1 = F_1(X_1), \\ U_j = F_{j|1,\ldots,j-1}(X_j|X_1, \ldots, X_{j-1}), \ j = 2, \ldots, s. \end{cases} \tag{7.2}$$

It is clear that (a) $U_1, \ldots, U_s$ are independent Uniform[0, 1] random variables, (b) $(X_1, \ldots, X_s) \to (U_1, \ldots, U_s)$ is one-to-one from $\mathscr{X}$ to C^s, and (c) the Jacobian of RT corresponds to the joint density function f. Note that RT is not permutation-invariant and there exist $s!$ kinds of different forms according to reordering $(X_1, \ldots, X_s)$. For each permutation $(i_1, \ldots, i_s)$, denote $T_{(i_1,\ldots,i_s)}$ as the RT from the permutated $(X_{i_1}, \ldots, X_{i_s})$ to $(U_1, \ldots, U_s)$.

Given the uniform distribution on C^s, the inverse $T^{-1}_{(i_1,\ldots,i_s)}$ maps $\boldsymbol{u} = (U_1, \ldots, U_s)$ back to $\boldsymbol{x} = (X_{i_1}, \ldots, X_{i_s}) \in \mathscr{X}$. More specifically, for a given observation $(u_1, \ldots, u_s) \in C^s$, the inverse Rosenblatt transformation (IRT) can be expressed by

$$\begin{cases} x_{i_1} = Q_{i_1}(u_1), \\ x_{i_j} = Q_{i_j|i_1,\ldots,i_{j-1}}(u_j|X_{i_1} = x_{i_1}, \ldots, X_{i_{j-1}} = x_{i_{j-1}}), \ j = 2, \ldots, s \end{cases} \tag{7.3}$$

based on the quantile functions (i.e. inverse CDFs). To illustrate how the IRT works, we consider the following example with two disjoint rectangular regions, where the uniform distribution is assumed on each sub-region.

Example 7.1 (**Two-rectangle Domain**) Consider the uniform distribution on the domain $\mathscr{X}$ with two disjoint rectangles, as shown in Fig. 7.1. It is clear the marginal and conditional CDFs for either permutation (x_1, x_2) or (x_2, x_1) are of the piecewise linear forms, as depicted in Fig. 7.2. Suppose we are given a point $\boldsymbol{u} = (0.2, 0.75) \in C^2$, then it is straightforward to apply the IRT in (7.3), we can

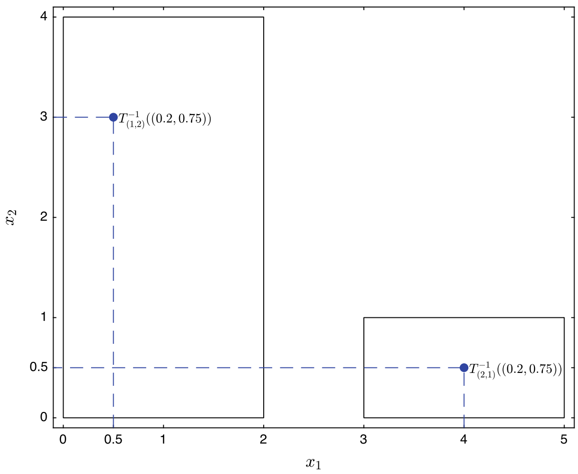

Fig. 7.1 Experimental domain $\mathscr{X}$ formed with two disjoint rectangles with the vertices ((0, 0), (2, 0), (2, 4), (0, 4)) and ((3, 0), (5, 0), (5, 1), (0, 1)), respectively. The two points at (0.5, 3) and (4, 0.5) are converted from $(0.2, 0.75) \in C^2$ through $T^{-1}_{(1,2)}$ and $T^{-1}_{(2,1)}$, respectively

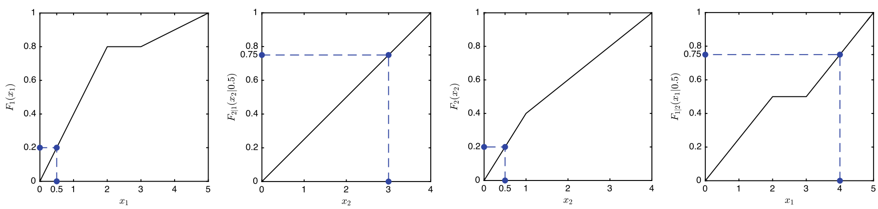

Fig. 7.2 The marginal and conditional CDFs used for obtaining the IRT points on the two-rectangle domain in Fig. 7.1. The point $(0.2, 0.75) \in C^2$ is taken as an example for performing the IRT

obtain the corresponding points in $\mathscr{X}$ through both $T^{-1}_{(1,2)}$ and $T^{-1}_{(2,1)}$, as shown in Fig. 7.1.

Algorithm 3 Inverse Rosenblatt Transformation Method

Input: An arbitrary domain $\mathscr{X}$ with closed boundary, and an n-run uniform design $\{\boldsymbol{u}_i\}_{i=1}^n$ on C^s.

1: Choose a permutation $(i_1, \ldots, i_s)$ of $(1, \ldots, s)$, then find the corresponding $T_{(i_1,\ldots,i_s)}$ based on uniform distribution within in the given boundary.

2: Use IRT to convert $\{\boldsymbol{u}_i\}_{i=1}^n$ to the domain $\mathscr{X}$:

$$\boldsymbol{x}_i = T^{-1}_{(i_1,\ldots,i_s)}(\boldsymbol{u}_i), \quad i = 1, \ldots, n.$$

3: Evaluate the CCD criterion (7.4) for the resulted $X = \{\boldsymbol{x}_i\}_{i=1}^n$.

4: Repeat Steps 1–3 for all $s!$ permutations. Output the best design X^* with the lowest CCD score.

The two-rectangle domain example motivates us to develop a practical IRT method for constructing uniform designs on arbitrary experimental domains, as presented in Algorithm 3. For any domain with uniform distribution and closed boundary, in the first step, we can find its RT's $T_{(i_1,\ldots,i_s)}$ subject to a different permutation. The marginal and conditional CDFs can be either derived analytically by (multiple) integration, or obtained through hyperrectangle approximation. In the latter situation, each CDF is approximated by a non-decreasing piecewise linear function.

In the second step, we can apply (7.3) to each point of a given n-run uniform design on C^s, so as to generate the corresponding points on $\mathscr{X}$. As we have reviewed, there are a rich collection of existing methods for constructing uniform designs on unit hypercubes. See the Appendix about the GLP method together with the elegant Fibonacci designs on unit squares.

Each uniform design on C^s leads to at maximum $s!$ different designs since there are $s!$ versions of IRT $T^{-1}_{(i_1,\ldots,i_s)}$ subject to different permutations. To determine the best design on $\mathscr{X}$, we employ the aforementioned CCD criterion as a measure of uniformity on arbitrary domains. According to Chuang and Hung [5], for any interior point $\mathbf{z}$ in $\mathscr{X}$, it can be treated as the Cartesian center to cut $\mathscr{X}$ into 2^s quadrants, then the ℓ_2 form of CCD is defined by

$$\mathrm{CCD}(X) = \left\{ \frac{1}{V(\mathscr{X})} \int_{\mathscr{X}} \frac{1}{2^s} \sum_{k=1}^{2^s} \left| \frac{N(\mathscr{X}_k(\mathbf{z}), X)}{n} - \frac{V(\mathscr{X}_k(\mathbf{z}))}{V(\mathscr{X})} \right|^2 d\mathbf{z} \right\}^{1/2}, \quad (7.4)$$

where $N(\mathscr{X}_k(\mathbf{z}), X)$ denotes the number of design points in $\mathscr{X}_k(\mathbf{z})$, and $V(\mathscr{X})$ and $V(\mathscr{X}_k(\mathbf{z}))$ denote the volumes of $\mathscr{X}$ and $\mathscr{X}_k(\mathbf{z})$, respectively. In practice, the integration over $\mathscr{X}$ can be approximated by Monte Carlo average over a large number of equal-spaced grid points (say, 100 grid points along each coordinate).

Note that in Algorithm 3, when the domain $\mathscr{X}$ is symmetric in two or more coordinates, some permutations can be relaxed and we only need consider all permutations for asymmetric coordinates. For the symmetric flexible regions in $\mathbb{R}^2$

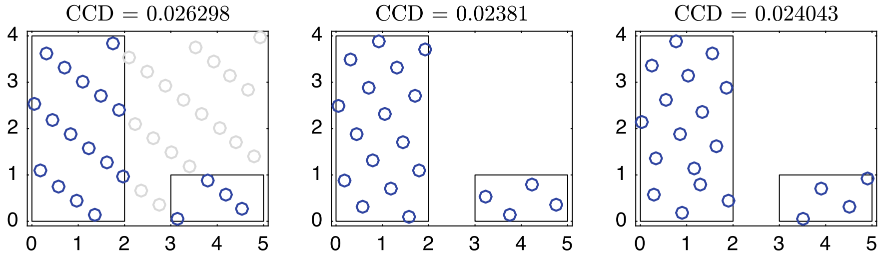

Fig. 7.3 Construction results of 20-run uniform designs on the two-rectangle domain in Fig. 7.1. Left panel: AR method; Center panel: IRT method with permutation (1, 2); Right panel: IRT method with permutation (2, 1)

to be discussed in next section, there is no need to consider permutations, so the evaluation of CCD criterion can be also skipped. Take the cylinder domain $\mathscr{X} = \{(x_1, x_2, x_3) : x_1^2 + x_2^2 \leq r^2, a \leq x_3 \leq b\} \subset \mathbb{R}^3$ as another example. We only need consider three permutations (1, 2, 3), (1, 3, 2) and (3, 1, 2).

Let us test the proposed IRT algorithm to construct the uniform design on the two-rectangle domain in Example 7.1. Using the LOO-Fibonacci design with $n = 20$ runs (see Fig. 7.8), we obtain the IRT construction results shown in Fig. 7.3. It turns out the permutation (x_1, x_2) leads to smaller CCD score than the permutation (x_2, x_1). In contrast, the AR method is also tested with 39-run uniform design on the outer rectangle with vertices ((0, 0), (5, 0), (5, 4), (0, 5)). The accepted 20-run sub-design within the domain of interest has a relatively worse CCD score.

7.3 Construction Results

In this section, we present the construction results by the IRT method for four kinds of experimental domains in $\mathbb{R}^2$. Numerical comparisons of effectiveness are conducted between the proposed method and the existing SR and AR methods.

7.3.1 Flexible Regions

The flexible regions on $\mathbb{R}^2$ controlled by a shape parameter $m > 0$ are defined by

$$\mathscr{X}_F^{(m)} = \left\{(x_1, x_2) \in [0, 1]^2 : \ |2x_1 - 1|^m + |2x_2 - 1|^m \leq 1\right\}. \tag{7.5}$$

Figure 7.4 shows the boundaries of such flexible regions with $m = \infty$, 2, 1 and 0.5, respectively. The circled points within each flexible region represent the constructed 88-run uniform designs by the proposed IRT method derived below.

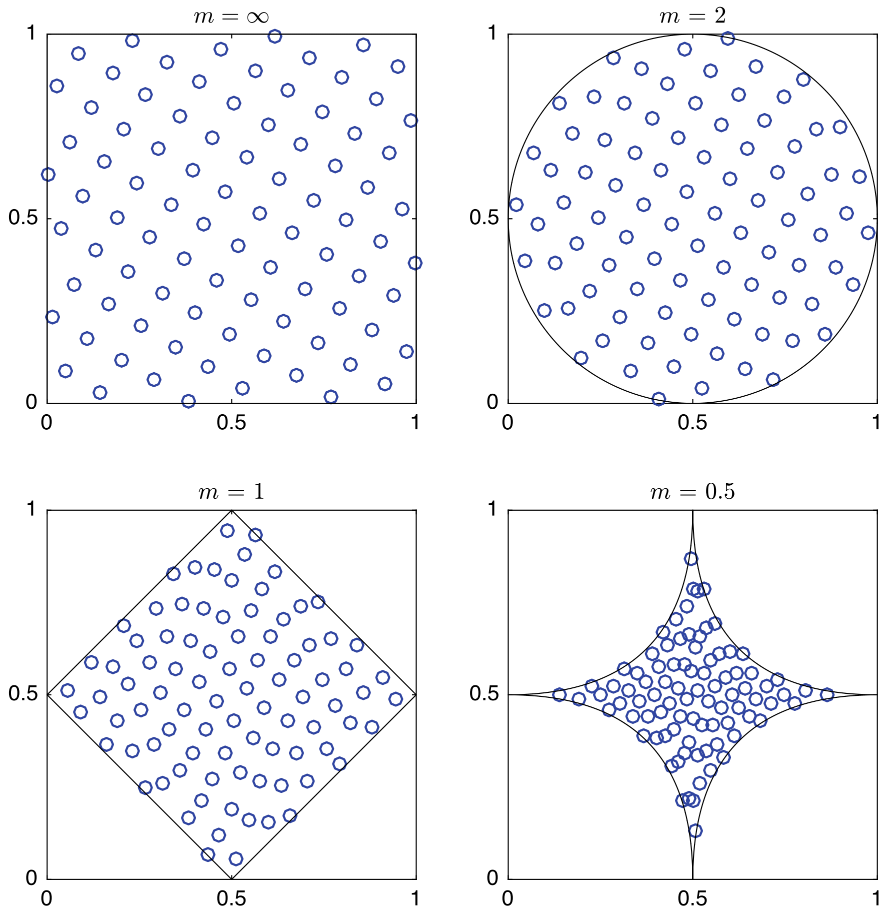

Fig. 7.4 Uniform designs with $n = 88$ runs on flexible regions with varying shape parameters. For $m = \infty$, the design is obtained by the GLP method (LOO-Fibonacci design) on the unit square (see Fig. 7.8). The design points in the $m = \infty$ case are then converted by the IRT method to the flexible regions for $m = 2$, 1 and 0.5

The flexible regions are symmetric in (x_1, x_2), so there is no need to consider permutations. For the random vector $X = (X_1, X_2) \sim \text{Uniform}(\mathscr{X}_F^{(m)})$, the marginal CDF $F_1(x)$ of the first component X_1 is

$$F_1(x) = \int_0^x \int_{0.5-(1-|2x_1-1|^m)^{1/m}/2}^{0.5+(1-|2x_1-1|^m)^{1/m}/2} \frac{1}{V(\mathscr{X}_F^{(m)})} dx_2 dx_1$$

where $V(\mathscr{X}_F^{(m)}) = \int_0^1 (1 - |2x_1 - 1|^m)^{1/m} dx_1$. Note that

$$\int_0^x (1-|2x_1-1|^m)^{1/m} dx_1$$
$$= \begin{cases} x, & m=\infty; \\ \frac{B\left(\frac{1}{m},\frac{1}{m}+1\right)}{2m}\left[1+\text{sign}(x-0.5)I_{|2x-1|^m}\left(\frac{1}{m},\frac{1}{m}+1\right)\right], & 0<m<\infty, \end{cases}$$

where $B(a,b)$ and $I_c(a,b)$ (with $a,b>0$, $c\in[0,1]$) are the values of Beta function and Incomplete Beta ratio with the following forms

$$B(a,b)=\int_0^1 t^{a-1}(1-t)^{b-1}dt, \quad I_c(a,b)=\frac{B_c(a,b)}{B(a,b)}=\frac{\int_0^c t^{a-1}(1-t)^{b-1}dt}{\int_0^1 t^{a-1}(1-t)^{b-1}dt}.$$

Therefore, the marginal CDF is given by

$$F_1(x)=\begin{cases} x, & m=\infty; \\ \frac{1}{2}+\frac{\text{sign}(x-0.5)}{2}\cdot I_{|2x-1|^m}\left(\frac{1}{m},\frac{1}{m}+1\right), & 0<m<\infty. \end{cases} \tag{7.6}$$

and its inverse is given by

$$F_1^{-1}(u)=\begin{cases} u, & m=\infty; \\ \frac{1}{2}+\frac{\text{sign}(u-0.5)}{2}\left(I^{-1}_{|2u-1|}\left(\frac{1}{m},\frac{1}{m}+1\right)\right)^{1/m}, & 0<m<\infty. \end{cases}$$

Moreover, the conditional CDF $F_{2|1}(x|x_1)$ of the second component given the value of the first component being x_1 is given by

$$F_{2|1}(x|x_1)=\begin{cases} x, & m=\infty; \\ \frac{x-\left(0.5-(1-|2x_1-1|^m)^{1/m}/2\right)}{(1-|2x_1-1|^m)^{1/m}}, & 0<m<\infty \end{cases} \tag{7.7}$$

and its inverse is given by

$$F_{2|1}^{-1}(u|x_1)=\begin{cases} u, & m=\infty; \\ 0.5+(u-0.5)(1-|2x_1-1|^m)^{1/m}, & 0<m<\infty. \end{cases}$$

Thus we obtain the analytical IRT $T^{-1}_{(1,2)}$ for $[0,1]^2 \to \mathscr{X}_F^{(m)}$ as follows:

$$T^{-1}_{(1,2)}((u_1,u_2))=\left(F_1^{-1}(u_1),\ F_{2|1}^{-1}\left(u_2|F_1^{-1}(u_1)\right)\right). \tag{7.8}$$

The effectiveness of the IRT method can be compared with traditional AR and SR methods. We use the CCD criterion to evaluate the uniform designs for $n=10,20,\ldots,100$ runs based on different construction methods. For the AR method, for each target n, we search the uniform designs on C^2 with sizes $n+1, n+2, \ldots$ in order to find such a design with exactly n runs falling into $\mathscr{X}_F^{(m)}$. Note that such AR method has the chance to find no appropriate design with the target number of runs. For the SR method, for $m = 2$, the method by Fang and Wang [9] is employed; for $m = 1$, the method by Tian et al. [16] is employed; for $m = 0.5$, there exists no SR method in the literature. All the needed uniform designs in C^2 are generated by the mixed GLP method (see Appendix). The numerical results for flexible regions with $m = 2, 1, 0.5$ are listed in Table 7.1. It can be found that IRT shows competitive performances in most cases.

Table 7.1 CCD scores for uniform designs constructed on the flexible regions with $m = 2, 1, 0.5$

m	n	AR	SR	IRT
2	10	–	0.054329	0.048877
	20	0.027465	0.030055	0.025930
	30	0.020286	0.024471	0.020323
	40	0.014839	0.021816	0.017132
	50	0.013721	0.018935	0.013708
	60	0.012722	0.015926	0.013522
	70	0.012529	0.013691	0.012670
	80	0.011813	0.013876	0.011595
	90	–	0.013461	0.011708
	100	0.012388	0.012752	0.011157
1	10	0.041046	0.047865	0.044335
	20	0.033795	0.045915	0.026233
	30	–	0.018906	0.021509
	40	0.018492	0.016612	0.018882
	50	0.012644	0.025909	0.015558
	60	0.015144	0.014044	0.015588
	70	0.015771	0.013879	0.014514
	80	0.015241	0.013489	0.014223
	90	0.017004	0.021852	0.013704
	100	0.012856	0.014312	0.013199
0.5	10	0.048221	–	0.048607
	20	0.035220	–	0.032163
	30	0.034672	–	0.027973
	40	0.028218	–	0.028656
	50	0.033651	–	0.023997
	60	0.025663	–	0.025844
	70	0.029249	–	0.025079
	80	0.024936	–	0.025046
	90	0.024674	–	0.024543
	100	0.020609	–	0.025098

7.3.2 Constrained Domain

Tian et al. [16] studied a tetragon shape of constrained domain for drug combination experiment, as defined by

$$\mathscr{X}_T = \left\{(x_1, x_2) \in \mathbb{R}_+^2 : 20 < 101.91 - 31.17x_1 - 9.56x_2 < 80\right\}. \tag{7.9}$$

They constructed a 19-run uniform design on this domain by the SR method, as shown in the left panel of Fig. 7.5. In this section, we apply the proposed IRT method for constructing a competitive uniform design on this specific constrained domain with the same number of runs.

First we can convert the domain $\mathscr{X}_T$ to the following symmetric domain

$$\mathscr{X}_{\tilde{T}} = \left\{(\tilde{x}_1, \tilde{x}_2) \in \mathbb{R}_+^2 : 21.91 < \tilde{x}_1 + \tilde{x}_2 < 81.91\right\},$$

where $\tilde{x}_1 = 31.17x_1$ and $\tilde{x}_2 = 9.56x_2$. Write $c_1 \equiv 21.91$ and $c_2 \equiv 81.91$, then it is easy to get the marginal CDF:

$$F_1(\tilde{x}_1) = \begin{cases} \frac{2\tilde{x}_1}{c_1+c_2}, & \text{if } 0 \le \tilde{x}_1 \le c_1; \\ 1 - \frac{(c_2-\tilde{x}_1)^2}{c_2^2-c_1^2}, & \text{if } c_1 < \tilde{x}_1 \le c_2. \end{cases} \tag{7.10}$$

For each fixed $\tilde{x}_1$, the conditional CDF $F_{2|1}(\tilde{x}_2|\tilde{x}_1)$ is given by the uniform distribution with the range $[c_1 - \tilde{x}_1, c_2 - \tilde{x}_1]$ if $\tilde{x}_1 \in [0, c_1]$ and the range $[0, c_2 - \tilde{x}_1]$ if $\tilde{x}_1 \in [c_1, c_2]$.

Using the IRT method based on a 19-run uniform design on C^2, we first obtain the transformed design on $\mathscr{X}_{\tilde{T}}$, then convert the design points back to $\mathscr{X}_T$ as plotted in the right panel of Fig. 7.5. This new design is more uniform than Tian et al. [16]'s result according to the CCD criterion.

7.3.3 Manifold Domain

Other than the flexible regions discussed in the previous section, we consider another special manifold domain defined by the ring constraint:

$$\mathscr{X}_R = \left\{(x_1, x_2) \in \mathbb{R}^2 : \frac{1}{4} \le x_1^2 + x_2^2 \le 1\right\}. \tag{7.11}$$

To get the marginal and conditional CDFs, instead of striving to derive the analytical forms, we adopt the following approximation method:

(1) Partition the x_1-coordinate from $[-1, 1]$ into $N = 1000$ equal-spaced intervals, each interval with mid-point $z_k = (2k-1)/N - 1$ for $k = 1, \ldots, N$.

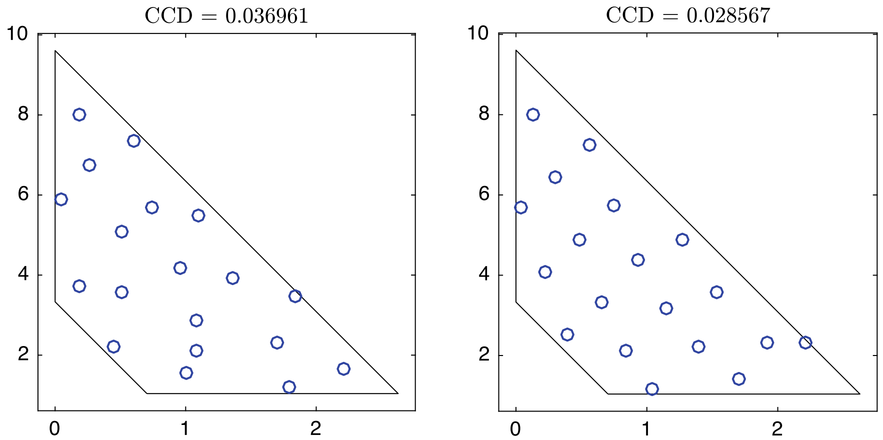

Fig. 7.5 Uniform designs with 19 runs on a constrained domain (with x_2 shifted 1.04 units upward). Left panel: SR method by Tian et al. [16]; Right panel: IRT method

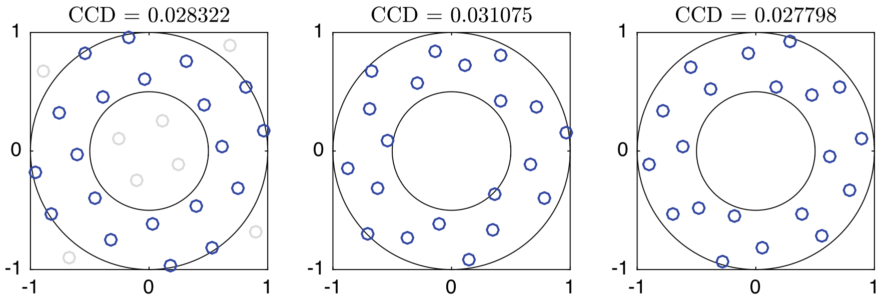

Fig. 7.6 Uniform designs with 20 runs on the ring domain. Left panel: AR method; Center panel: SR method by Zhang [19]; Right panel: IRT method

(2) Obtain the approximate marginal CDF for x_1 based on the midpoints:

$$\widehat{F}_1(z_k) = \frac{\sqrt{1-z_k^2} - \sqrt{1/4 - z_k^2} \cdot I\left(|z_k| \leq \frac{1}{2}\right)}{\sum_{k=1}^{N}\left(\sqrt{1-z_k^2} - \sqrt{1/4 - z_k^2} \cdot I\left(|z_k| \leq \frac{1}{2}\right)\right)}, \quad k = 1, \ldots, N. \tag{7.12}$$

(3) When x_1 takes discretized z_k-values, obtain the conditional CDF for $x_2|x_1$ by the uniform distribution with the range $\left[-\sqrt{1-z_k^2}, \sqrt{1-z_k^2}\right]$ if $|z_k| \geq 1/2$ or the range $\left[-\sqrt{1-z_k^2}, -\sqrt{1/4-z_k^2}\right] \cup \left[\sqrt{1/4-z_k^2}, \sqrt{1-z_k^2}\right]$ if $|z_k| \leq 1/2$.

Figure 7.6 (right panel) shows the IRT constructed 20-run uniform design on the ring domain. In contrast, on the left panel is the result by the AR method based on

Table 7.2 CCD scores for uniform designs constructed on the ring-shaped domain

n	AR	SR	IRT
10	0.051980	0.048791	0.047818
20	0.028322	0.031075	0.027798
30	0.016596	0.021944	0.017866
40	0.013043	0.017600	0.015126
50	0.013418	0.015682	0.014486
60	0.011472	0.012808	0.011033
70	0.009177	0.011868	0.010474
80	0.011507	0.011086	0.009580
90	0.009958	0.009751	0.009025
100	0.012203	0.009791	0.007863

28-run uniform design on the unit cube. On the center panel is the result by the SR method through $x_{i1} = \sqrt{u_{i1}}\sin(2\pi u_{i2})$, $x_{i2} = \sqrt{u_{i1}}\cos(2\pi u_{i2})$; see Zhang [19]. It is found that in this case the IRT outperforms AR and SR methods. Moreover, we run through $n = 10, 20, \ldots, 100$ to compare the three methods, with numerical results presented in Table 7.2. We can see that the IRT method always outperforms the SR method, and sometimes have better performance than the AR method.

7.3.4 Geographical Domain

Lastly, we consider the geographical domain that is usually rather irregular. In this section we consider the Land Map of China as the experimental domain $\mathscr{X}_{\text{map}}$. The entire domain consists of several closed subregions. It is difficult to determine the exact form of Rosenblatt transformation, so we use the approximated CDFs.

To approximate the marginal and conditional CDFs on the map domain, we find a rectangle to completely cover $\mathscr{X}_{\text{map}}$ and establish a cartesian coordinate system. The rectangle has the resolution of 1297×1083 pixels, and the contour of $\mathscr{X}_{\text{map}}$ contains $N = 675328$ pixels in total. The marginal CDF of x_1 is approximated by

$$\widehat{F}_1(x_1) = \frac{1}{N}\sum_{i=1}^{N} I(x_{i1} \le x_1), \quad x_1 = z_1, \ldots, z_{1297}. \tag{7.13}$$

For $x_1 = z_1, \ldots, z_{1297}$, the conditional CDF of $x_2|x_1$ is approximated by

$$\widehat{F}_{2|1}(x_2|x_1 = z_k) = \frac{1}{|\Omega_k|}\sum_{j\in\Omega_k} I(x_{j2} \le x_2), \tag{7.14}$$

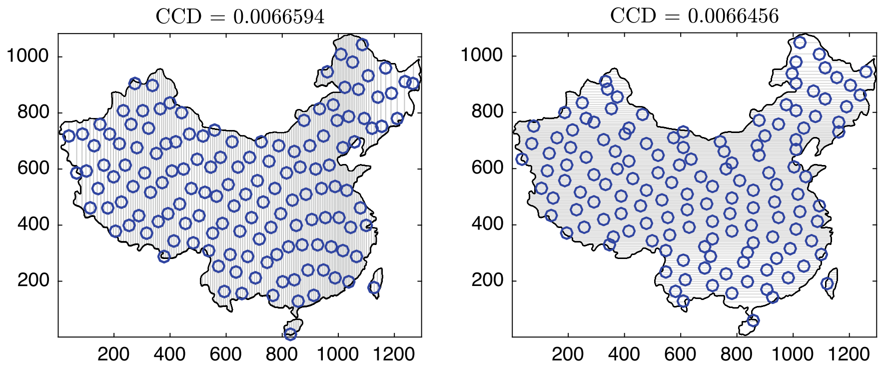

Fig. 7.7 Uniform designs with 143 runs on China map domain. Left panel: IRT method with permutation (1, 2); Right panel: IRT method with permutation (2, 1). The vertical and horizontal gray lines represent the partitions of the map for approximating the conditional CDFs

where Ω_k denotes the subset of pixels $\{j : x_{j1} = z_k\}$. Similarly for permutation (x_2, x_1), we can obtain $\widehat{F}_2(x_2)$ and $\widehat{F}_{1|2}(x_1|x_2)$.

Suppose we are given a 143-run LOO-Fibonacci design (see Fig. 7.8). We may use the IRT method with respect to permutations (x_1, x_2) and (x_2, x_1) to construct the corresponding uniform designs on $\mathscr{X}_{\text{map}}$. The construction results are visualized in Fig. 7.7, with the permutation (x_2, x_1) leading to a slightly lower CCD score. In each permutation, there are two points to represent the Hainan and Taiwan islands on the map, respectively.

7.4 Conclusion

The construction of uniform designs on irregular regions has been a relatively challenging task as compared with the case on regular regions. Inspired by the stochastic representation method in Fang and Wang [9], we propose to construct uniform designs on arbitrary domains by the inverse Rosenblatt transformation (IRT) based on marginal and conditional distributions. We have demonstrated how to use this method in multiple kinds of experimental domains in two-dimensional space, and the construction results are rather competitive and promising.

There are several interesting problems that are worth further study. First, the IRT method is proposed for not only two-dimensional space, but also higher dimensional space. In the latter case it is however computationally demanding. It is important to develop a highly efficient algorithm for approximating the marginal and conditional distribution functions. Second, there exist other manifold domains than the ring shape and flexible regions, e.g. the sphere and donut kinds of surfaces in high-dimensional space. It is interesting to extend the IRT method to these manifold cases. Third, in this study we find that the central composite discrepancy is an imperfect measure of uniformity on arbitrary domains. For example, it lacks the property of invariance under rotation. It is among our research plans to develop a better kind of discrepancy measure for space-filling designs on arbitrary domains with general distributional assumption.

Acknowledgements This work was supported by National Natural Science Foundation of China (11871288) and Natural Science Foundation of Tianjin (19JCZDJC31100).

Appendix: Good Lattice Point Method

The uniform designs constructed on the unit hypercubes by the GLP method are also known as the NT-nets [9], which uses the classical star-discrepancy for evaluating the uniformity of the candidate designs. In Algorithm 4 we write the GLP method using the centered-ℓ_2 discrepancy (CD2), a more popular criterion proposed by Hickernell [11]. Meanwhile, it is easy to check that the GLP designs (7.15) always include a point $\boldsymbol{x}_n = (1 - 1/2n, \ldots, 1 - 1/2n) \in C^s$. The leave-one-out (LOO) GLP method is to remove such a dummy point, then scale the remaining points by $n/(n-1)$ in all coordinates. Thus, in order to construct an n-run uniform design, we can use a mixed GLP method by selecting the lower-CD2 design between the GLP (with input n) and LOO-GLP (with input $n+1$) outputs.

It is well-known that for $s = 2$ and $n = F_k$ (Fibonacci numbers 5, 8, 13, 21, …), the lattice designs generated by $h_1 = 1$ and $h_2 = F_{k-1}$ enjoy the remarkable low star-discrepancy property [18]. It is of our interest to investigate whether such Fibonacci designs may also attain low discrepancy with respect to the CD2 criterion. As a key difference, the star-discrepancy is anchored at the origin of the unit hypercube, while the CD2 is anchored at the center. It turns out the Fibonacci designs are sub-optimal under CD2. Nevertheless, we find that the LOO-Fibonacci designs with $n = F_k - 1$ ($F_k \le 1597$) runs remarkably minimize the CD2 criterion among all the generating vectors for the mixed GLP method. See Fig. 7.8 about the LOO-Fibonacci designs with 20, 88 and 143 runs. See Table 7.3 for the numerical results based on

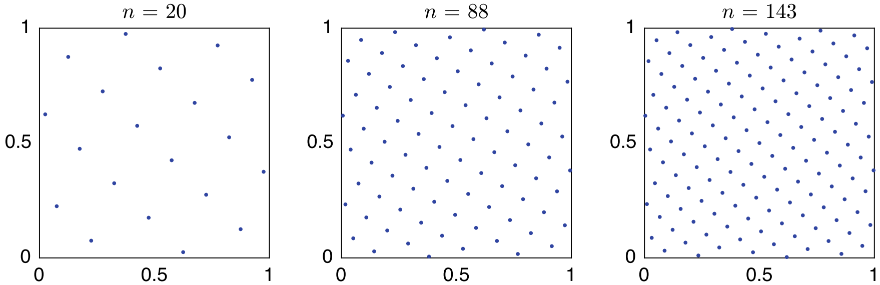

Fig. 7.8 Scatter plots of LOO-Fibonacci designs for $n = 20$, 88 and 143 runs

Algorithm 4 Good Lattice Point Method

Input: The number of factors s, and the number of runs n.

1: Form a generating vector $(h_1, h_2, \ldots, h_s)$ by choosing distinct positive integers that are less than and relatively prime to n.

2: Form the n-run lattice design $\boldsymbol{X} = [x_{ij}]_{n \times s}$ with entries

$$x_{ij} = \left\{ \frac{2ih_j - 1}{2n} \right\}, \ i = 1, \ldots, n, \ j = 1, \ldots, s \tag{7.15}$$

where $\{z\}$ is the factorial part of z.

3: Evaluate the criterion of the centered-ℓ_2 discrepancy

$$\mathrm{CD}_2(\boldsymbol{X}) = \left\{ \left(\frac{13}{12}\right)^s - \frac{2}{n} \sum_{i=1}^{n} \prod_{j=1}^{s} \left(1 + \frac{1}{2}\left|x_{ij} - \frac{1}{2}\right| - \frac{1}{2}\left|x_{ij} - \frac{1}{2}\right|^2\right) \right.$$
$$\left. + \frac{1}{n^2} \sum_{i,k=1}^{n} \prod_{j=1}^{s} \left(1 + \frac{1}{2}\left|x_{ij} - \frac{1}{2}\right| + \frac{1}{2}\left|x_{kj} - \frac{1}{2}\right| - \frac{1}{2}|x_{ij} - x_{kj}|\right) \right\}^{1/2} \tag{7.16}$$

4: Repeat Steps 1–3 for all distinct generating vectors. Output $\boldsymbol{X}^*$ with the lowest CD_2 value.

Table 7.3 LOO-Fibonacci designs with $h_1 = 1$ and $h_2 = F_{k-1}$ minimize the CD2 criterion for the mixed GLP method, where h_2^* represents the best found generating vectors per each method

$n = F_k - 1$	$h_2 = F_{k-1}$	h_2^* (LOO-GLP)	min-CD2	h_2^* (GLP)	min-CD2
4	3	2, 3	1.275E–01	3	1.350E–01
7	5	3, 5	7.631E–02	3, 5	8.122E–02
12	8	5, 8	4.557E–02	5	5.058E–02
20	13	8, 13	2.843E–02	9	3.133E–02
33	21	13, 21	1.764E–02	14, 26	1.947E–02
54	34	21, 34	1.117E–02	35	1.288E–02
88	55	34, 55	7.010E–03	37	7.661E–03
143	89	55, 89	4.456E–03	63	4.823E–03
232	144	89, 144	2.806E–03	147	3.115E–03
376	233	144, 233	1.784E–03	165	1.916E–03
609	377	233, 377	1.123E–03	256	1.224E–03
986	610	377, 610	7.128E–04	579	7.600E–04
1596	987	610, 987	4.484E–04	617	4.859E–04

exhaustive search up to $F_k = 1597$. From Table 7.3, it can be found that the LOO-Fibonacci designs with $n = F_k - 1$ also include $h_1 = 1$ and $h_2 = F_{k-2}$ as the optimal generating vector. This can be actually justified by the reflection-invariant property of the CD2 criterion.

References

1. Arnold, B.C., Castillo, E., Sarabia, J.M.: Families of multivariate distributions involving the Rosenblatt construction. J. Am. Stat. Assoc. **101**(476), 1652–1662 (2006)
2. Borkowski, J.J., Piepel, G.F.: Uniform designs for highly constrained mixture experiments. J. Qual. Technol. **41**, 35–47 (2009)
3. Chen, R.B., Shu, Y.H., Hung, Y., Wang, W.: Central composite discrepancy-based uniform designs for irregular experimental regions. Comput. Stat. Data Anal. **72**, 282–297 (2014)
4. Chen, R.B., Li, C.H., Hung, Y., Wang, W.: Optimal noncollapsing space-filling designs for irregular experimental regions. J. Comput. Graph. Stat. **28**(1), 74–91 (2019)
5. Chuang, S.C., Hung, Y.: Uniform design over general input domains with applications to target region estimation in computer experiments. Comput. Stat. Data Anal. **54**, 219–232 (2010)
6. Fang, K.T.: The uniform design: application of number theoretic methods in experimental design. Acta Math. Appl. Sin. **3**, 363–372 (1980)
7. Fang, K.T., Li, R.Z., Sudjianto, A.: Design and Modeling for Computer Experiments. Chapman and Hall/CRC, Boca Raton, FL (2006)
8. Fang, K.T., Liu, M.Q., Qin, H. and Zhou, Y.D.: Theory and Application of Uniform Experimental Designs. Springer, Singapore (2018)
9. Fang, K.T., Wang, Y.: Number-Theoretic Methods in Statistics. Chapman and Hall, London (1994)
10. Fang, K.T., Yang, Z.H.: On uniform design of experiments with restricted mixtures and generation of uniform distribution on some domains. Stat. Probab. Lett. **46**, 113–120 (2000)
11. Hickernell, F.: A generalized discrepancy and quadrature error bound. Math. Comput. **67**(221), 299–322 (1998)
12. Justel, A., Pena, D., Zamar, R.: A multivariate Kolmogorov-Smirnov test of goodness of fit. Stat. Probab. Lett. **35**(3), 251–259 (1997)
13. Lin, D.K.J., Sharpe, C., Winker, P.: Optimized U-type designs on flexible regions. Comput. Stat. Data Anal. **54**, 1505–1515 (2010)
14. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadelphia (1992)
15. Rosenblatt, M.: Remarks on a multivariate transformation. Ann. Math. Stat. **29**, 470–472 (1952)
16. Tian, G.L., Fang, H.B., Tan, M., Qin, H., Tang, M.L.: Uniform distributions in a class of convex polyhedrons with applications to drug combination studies. J. Multivar. Anal. **100**, 1854–1865 (2009)
17. Wang, Y., Fang, K.T.: A note on uniform distribution and experimental design. Kexue Tongbao **26**, 485–489 (1981)
18. Zaremba, S.K.: Good lattice points, discrepancy, and numerical integration. Annali di matematica pura ed applicata **73**(1), 293–317 (1966)
19. Zhang, R.C.: On a transformation method in constructing multivariate uniform designs. Stat. Sinica **6**, 455–469 (1996)