Statistica Sinica **6**(1996), 455-469

ON A TRANSFORMATION METHOD IN CONSTRUCTING MULTIVARIATE UNIFORM DESIGNS

Runchu Zhang

Nankai University

Abstract. Based on the work of Wang and Fang (1990a,b), this paper extends the definition of discrepancy of a set of points with respect to a distribution. To construct multivariate uniform designs, two transformations, discrepancy-preserving transformation and density-preserving transformation, and a method of separating variables are introduced. Also, for some special domains useful in statistics, some related transformations are given. In particular, as an application of the method of separating variables, we give an approach of constructing a uniform design in the Stiefel manifold and outline its applications in projection pursuit methods.

Key words and phrases: Computer experiment, Latin hypercube, number-theoretic method, projection pursuit, Stiefel manifold, uniform design.

1. Introduction

In recent years, a number of statisticians have paid attention to uniform designs and contributed a number of papers to explore the related theory and methods. Wang and Fang (1981) proposed a number-theoretic method of constructing uniform designs. Afterward, they gave a more systematic elaboration of their method and spelled out several application examples (Wang and Fang (1990a,b)). McKay, Conover, and Beckman (1979) proposed Latin hypercube sampling as a general sampling technique, which was also used in a computer experiment. Stein (1987) and Welch et al. (1992) developed this technique further and gave a series of results. Owen (1992) and Tang (1993) improved Latin hypercube sampling by using orthogonal arrays.

Uniform designs have been used in more and more fields such as numerical integration, computer experiments, statistical simulations and other statistical areas. For example, in finding the integral of a function or estimating statistical parameters, one uses uniform designs to improve precision and speed up convergence. In computer experiments, one uses uniform designs efficiently to estimate parameters and screen out important components of a model, etc. The above two methods of constructing uniform designs use different ways to select points in a given hypercube. One is to use the number-theoretic method based on Hua and Wang (1981). A set of points obtained by this method possesses a good representation in terms of lower discrepancy in the sense of Weyl (1916), even though the sample size is small. Another uses Latin hypercube sampling or lattice sampling. This method is convenient for studying statistical properties. In many practical problems, although the methods are different, they have often led to similar practical conclusions.

In some practical applications, one needs to consider uniform design problems not only in a unit hypercube but also in other domains. For this reason, it is necessary to discuss how to extend the notion of a uniform design in a hypercube to a general domain. Wang and Fang (1990a,b) expounded some methods of generating a set of points with good representation in some special domains. The present paper intends to explore this further. First, in Section 2 we extend the definition of discrepancy in Wang and Fang (1981) and introduce two types of transformations, discrepancy-preserving transformations and density-preserving transformations. In Section 3, we give some examples on constructing uniform designs in some special areas useful in statistics. Although some of these examples have been considered in Wang and Fang (1990a,b), there are some new implications here. We introduce a method of separating variables for constructing uniform designs in Section 4 and use this method to propose an approach of constructing uniform designs in the Stiefel manifold, which is useful, for example, in projection pursuit methods.

2. Discrepancy and Transformations

First of all, we simply recall the definition of discrepancy introduced in Wang and Fang (1981).

Let X be a continuous random vector in R^p with cumulative distribution function $F(\boldsymbol{x}) = F(x_1, \ldots, x_p)$. For any two points $\boldsymbol{x}_1 = (x_{11}, \ldots, x_{1p})$ and $\boldsymbol{x}_2 = (x_{21}, \ldots, x_{2p})$ in R^p, if $x_{1i} \leq x_{2i}, i = 1, \ldots, p$, we write $\boldsymbol{x}_1 \leq \boldsymbol{x}_2$. As usual, $F(\boldsymbol{x}) = P(X \leq \boldsymbol{x})$. Suppose that $\mathcal{F}_n = \{\boldsymbol{x}_i, i = 1, \ldots, n\}$ is a set of points in R^p and for any $\boldsymbol{x} \in R^p$ we denote by $N(\mathcal{F}_n, \boldsymbol{x})$ the number of points satisfying $\boldsymbol{x}_i \leq \boldsymbol{x}$. Then we call

$$D_F(\mathcal{F}_n) = \sup_{\boldsymbol{x} \in R^p} \Big| \frac{1}{n} N(\mathcal{F}_n, \boldsymbol{x}) - F(\boldsymbol{x}) \Big| \tag{1}$$

the discrepancy of $\mathcal{F}_n$ with respect to F. If there exists a set of points $\mathcal{F}_n^* = \{\boldsymbol{x}_i^*, i = 1, \ldots, n\} \subset R^p$ such that

$$D_F(\mathcal{F}_n^*) = \inf_{\mathcal{F}_n} D_F(\mathcal{F}_n), \tag{2}$$

then $\mathcal{F}_n^*$ is said to be a set of best representation points (brp) of F. In particular, if X has a uniform distribution in domain D, then call $\mathcal{F}_n^*$ satisfying the above conditions a set of best uniform representation points (*burp*) in D. In this case, we have $F(\boldsymbol{x}) = |D_{\boldsymbol{x}}|/|D|$, where $D_{\boldsymbol{x}}$ stands for the subdomain $(\boldsymbol{y} \leq \boldsymbol{x}) \cap D$ of D and $|D_{\boldsymbol{x}}|$ the volume of $D_{\boldsymbol{x}}$. From this, we see that the intuitive meaning of a *burp* $\mathcal{F}_n^*$ set is that the ratio of the number of points of $\mathcal{F}_n^*$ in a subdomain $D_{\boldsymbol{x}}$ to n, the total number of points in D, is uniformly close to the ratio of the volume of subdomain $D_{\boldsymbol{x}}$ to that of D. This just shows some uniformity of the set of points $\mathcal{F}_n^*$ located in D. For a general random vector X, a *brp* $\mathcal{F}_n^*$ set means that the ratio of the number of points of $\mathcal{F}_n^*$ in a domain $(\boldsymbol{y} \leq \boldsymbol{x})$ is uniformly close to the probability $P(X \leq \boldsymbol{x})$, which indicates the closeness between the sampling distribution $\mathcal{F}_n^*$ and the distribution F of X. Usually, for any distribution F and given n, it is difficult to find a *brp* set of F. Because of this, we also need the following definition: $\mathcal{F}_n$ is called a set of good representation points (*grp*) of F if $D_F(\mathcal{F}_n) = O(n^{-1+\epsilon})$ as $n \to \infty$, where $0 < \epsilon < \frac{1}{2}$. Hua and Wang (1981) gave a method for finding the *grp* sets in a hypercube and listed some related parameter tables.

To find a *grp* (*brp*) set for a given distribution, we introduce the following definition.

Definition 1. Let $Y = T(X)$ be a one-to-one differentiable transformation from a random vector $X = (X_1, \ldots, X_p)$ to a random vector $Y = (Y_1, \ldots, Y_p)$, X and Y respectively having distributions $F(\boldsymbol{x})$ and $G(\boldsymbol{y})$. Assume that $\mathcal{F}_n = \{\boldsymbol{x}_i; i = 1, \ldots, n\}$ is a set of points in R^p and $\mathcal{F}'_n = \{\boldsymbol{y}_i; \boldsymbol{y}_i = T(\boldsymbol{x}_i), i = 1, \ldots, n\}$ is the image of $\mathcal{F}_n$ through T. If for any $\mathcal{F}_n$, the discrepancy of $\mathcal{F}_n$ with respect to F equals the discrepancy of $\mathcal{F}'_n$ with respect to G, i.e.

$$D_F(\mathcal{F}_n) = D_G(\mathcal{F}'_n), \tag{3}$$

then we call T a discrepancy-preserving transformation (DIPT).

Obviously, a DIPT $Y = T(X)$ can ensure that a *grp* (*brp*) set of F is transformed to a *grp* (*grp*) set of G, and the inverse of a DIPT is also a DIPT provided its Jacobian is not zero. So we can find a *grp* (*brp*) set of G from that of F by using a DIPT.

Example 1. If the c.d.f. $F(\boldsymbol{x})$ of a random vector $X = (X_1, \ldots, X_p)$ has the form $\prod_{i=1}^p F_i(x_i)$, where $F_i(x_i)$ is the c.d.f. of component X_i, then the transformation $Y_i = F_i(X_i),\ \ i = 1, \ldots, p$, is a DIPT.

Example 2. If two random vectors $X = (X_1, \ldots, X_p)$ and $Y = (Y_1, \ldots, Y_p)$ have the relationship $Y = AX + B$, where A is a matrix with nonnegative elements and B a constant vector, then the transformation is a DIPT.

Proof. Since for any $\boldsymbol{x}_1 \leq \boldsymbol{x}_2$ we have $\boldsymbol{y}_1 = A\boldsymbol{x}_1 + B \leq A\boldsymbol{x}_2 + B = \boldsymbol{y}_2$, then for any set of points $\mathcal{F}_n$ we have $N(\mathcal{F}_n, \boldsymbol{x}) = N(\mathcal{F}'_n, \boldsymbol{y})$, where $\boldsymbol{y} = A\boldsymbol{x} + B$. Also, for any $\boldsymbol{x} \in R^p$, it follows that

$$F(\boldsymbol{x}) = P(X \leq \boldsymbol{x}) = P(AX + B \leq A\boldsymbol{x} + B) = P(Y \leq \boldsymbol{y}) = G(\boldsymbol{y}).$$

Therefore $D_F(\mathcal{F}_n)$=$D_G(\mathcal{F}'_n)$ is valid for all $\mathcal{F}_n$, proving the discrepancy-preserving property.

We say that a transformation $Y = T(X)$ is order-preserving if $\boldsymbol{x}_1 \leq \boldsymbol{x}_2$ implies that $\boldsymbol{y}_1 = T(\boldsymbol{x}_1) \leq T(\boldsymbol{x}_2) = \boldsymbol{y}_2$. From the proof of Example 2 we get the following corollary immediately.

Corollary 1. *An order-preserving transformation $Y = T(X)$ is also DIPT.*

The above method is based on an easy to find *grp* set and a DIPT. Usually, it is also difficult to find a DIPT and an F whose *grp* set can be determined. In this case, we need the following more general discrepancy definition.

Let a random vector X be defined in a domain D and have a c.d.f. $F(\boldsymbol{x})$. Suppose that we have a one-to-one transformation $X = T(V)$ so that the random vector V obtained through T is defined in an image domain D' and has a c.d.f. $G(\boldsymbol{v})$. Assume further that $\mathcal{F}_n = \{\boldsymbol{x}_i;\ \boldsymbol{x}_i \in D, i = 1, \ldots, n\}$ is a set of points in D and $\mathcal{F}'_n = \{\boldsymbol{v}_i;\ \boldsymbol{x}_i = T(\boldsymbol{v}_i), i = 1, \ldots, n\}$ is the set of image points in D' of $\mathcal{F}_n$. If we denote $D'_{\boldsymbol{v}} = (V \leq \boldsymbol{v}) \cap D'$, $D_{\boldsymbol{x}} = T(D'_{\boldsymbol{v}})$ and also use $N(\mathcal{F}_n, D_{\boldsymbol{x}})$ to indicate the number of points of $\mathcal{F}_n$ in $D_{\boldsymbol{x}}$, then we have $N(\mathcal{F}_n, D_{\boldsymbol{x}}) = N(\mathcal{F}'_n, D'_{\boldsymbol{v}}) = N(\mathcal{F}'_n, \boldsymbol{v})$, $F(D_{\boldsymbol{x}}) = G(D'_{\boldsymbol{v}}) = G(\boldsymbol{v})$ (here $F(D_{\boldsymbol{x}}) = \int_{D_{\boldsymbol{x}}} F(d\boldsymbol{x})$ and $G(D'_{\boldsymbol{v}}) = \int_{D'_{\boldsymbol{v}}} G(d\boldsymbol{v})$), and hence

$$\Big|\frac{1}{n}N(\mathcal{F}_n, D_{\boldsymbol{x}}) - F(D_{\boldsymbol{x}})\Big| = \Big|\frac{1}{n}N(\mathcal{F}'_n, \boldsymbol{v}) - G(\boldsymbol{v})\Big|. \tag{4}$$

We call

$$D_F^T(\mathcal{F}_n) = \sup_{\boldsymbol{x} \in D}\Big|\frac{1}{n}N(\mathcal{F}_n, D_{\boldsymbol{x}}) - F(D_{\boldsymbol{x}})\Big| \tag{5}$$

a transformation discrepancy (t-discrepancy) of $\mathcal{F}_n$ with respect to F under T. If there exists an $\mathcal{F}_n^*$ such that $D_F^T(\mathcal{F}_n^*) = \inf_{\mathcal{F}_n} D_F^T(\mathcal{F}_n)$, then we call the $\mathcal{F}_n^*$ a *brp* (*tbrp*) set of transformation of F under T. Similarly, we also have the definitions of a *tgrp* set and *tgurp* set of a distribution.

Clearly, by (4) we have $D_F^T(\mathcal{F}_n) = D_G(\mathcal{F}'_n)$ and $D_F^T(\mathcal{F}_n^*) = D_G(\mathcal{F}_n^{*'})$. Evidently, the definition of t-discrepancy depends on the transformation T used. But it has the following general implication. A set *tbrp* $\mathcal{F}_n^*$ of F is such that, for any $D_{\boldsymbol{x}}$, the ratio of the number of points of $\mathcal{F}_n^*$ in $D_{\boldsymbol{x}}$ to n is uniformly close to the probability $P(X \in D_{\boldsymbol{x}}) = P(D_{\boldsymbol{x}})$; it just shows uniformity of $\mathcal{F}_n^*$ with respect to distribution $F(\boldsymbol{x})$. In this sense, we can say that this definition is essentially similar to that of discrepancy given above, especially when the transformation

T is convex. Therefore in applications we do not distinguish between them and use the same notations *brp*, *grp*, and *gurp* for both.

From the definition above, note that, for any continuous random vectors X and V, if there is a one-to-one differentiable transformation $X = T(V)$ relating them, then we can always find a *brp* set from one random vector to the other. In particular, to do this, we recommend an important transformation below.

Let X be a continuous random vector with p.d.f. $f(\boldsymbol{x})$ and c.d.f. $F(\boldsymbol{x})$. Let $F(x_i|x_1,\ldots,x_{i-1})$ denote the conditional c.d.f. of X_i given $X_1 = x_1,\ldots,X_{i-1} = x_{i-1}, i = 2,\ldots,p$. Let

$$\begin{aligned} V_1 &= F(X_1), \\ V_i &= F(X_i|X_1,\ldots,X_{i-1}), \quad i = 2,\ldots,p. \end{aligned} \tag{6}$$

This is the well-known Rosenblatt transformation (Rosenblatt (1952)) and denote it by R^*.

Note that, for every i, V_i has a uniform distribution on $[0,1]$ which does not depend on $V_1,\ldots,V_{i-1}$, and hence $V_1,\ldots,V_p$ are i.i.d. $U[0,1]$ random variables. It is easy to see that the Jacobian of R^* is $f(\boldsymbol{x})$. So R^* is a one-to-one differentiable transformation and it can always change a continuous random vector in any domain to the random vector with uniform distribution in a unit hypercube. Therefore, the significance of the R^* transformation lies in the fact that it may allow us to find a *tgrp* set for many random vectors by finding a *grp* set in a unit hypercube, provided that the p.d.f. $f(\boldsymbol{x})$ of the random vector X has suitable properties so that we can get $X = (R^*)^{-1}(V)$.

In particular, we are interested in finding a *gurp* set in any domain. For this, in the following we consider another transformation.

Definition 2. Let D be a domain in R^p, and $\boldsymbol{x} = (x_1,\ldots,x_p)$ stand for a point in D and $\boldsymbol{v} = (v_1,\ldots,v_n)$ stand for a point in D'. If a one-to-one differentiable transformation $X = T(V)$:

$$x_i = g_i(v_1,\ldots,v_p), \quad i = 1,\ldots,p, \tag{7}$$

transforms the domain D to a domain D' in R^p, and its Jacobian

$$|J| \equiv c,$$

where $J = (\frac{\partial g_i}{\partial v_j})$ and c is a constant, then we say the transformation (7) is density-preserving (DEPT).

By density-reserving, we mean that, through a DEPT $X = T(V)$ the density of V is the same as that of X up to a constant factor. In particular, if X

has a uniform distribution in D, then V also has a uniform distribution in D'. Obviously, the inverse of a DEPT is also a DEPT.

Under a DEPT, the volume elements in the two domains have the relationship $\bigwedge_{i=1}^{p} dx_i = c \bigwedge_{i=1}^{p} dv_i$, in the exterior product symbol, where $\bigwedge_{i=1}^{p} dx_i = dx_1 \cdots dx_p$, or more simply $(d\boldsymbol{x}) = c(d\boldsymbol{v})$.

Sometimes a domain D under study in R^p is a hypersurface of q-dimension $(q < p)$ instead of p-dimension. In this case, we also consider a corresponding DEPT.

Definition 3. Let D be a q-dimensional domain in R^p $(q < p)$. If a random vector X defined in D has a one-to-one differentiable parameter representation $X = T(V)$:

$$x_i = g_i(v_1, \ldots, v_q), \quad i = 1, \ldots, p, \tag{8}$$

in an image domain D' in R^q, and its transform determinant satisfies

$$|T'T| \equiv c, \quad \text{where} \quad T(p \times q) = (\frac{\partial g_i}{\partial v_j}), \tag{9}$$

then we call (8) a DEPT from D in R^p to D' in R^q.

Similarly, if X has a uniform distribution in D, then the V obtained by a DEPT still has a uniform distribution in D' and *vice versa.* Also, we use the exterior product symbol to denote the relationship of differential elements between these two domains: $(d\boldsymbol{x}) = c(d\boldsymbol{v})$. But $(d\boldsymbol{x})$ here is the exterior differential form of $\boldsymbol{x}$ in D that stands for an area element of this hypersurface and $(d\boldsymbol{v})$ is still a volume element in D'.

It is easy to see that a DEPT T has the following properties:

1. If D_1 is a measurable subdomain of D and D_1' is the image of D_1 through T, denoted by $D_1 = T(D_1')$, then $|D_1| = |T(D_1')| = c|D_1'|$, where $|D|$ stands for the volume (or area) of a domain D.
2. In the definition of general discrepancy above, a DEPT T has a uniformity-preserving property, i.e., it transforms a *gurp* set in D into a *gurp* set in D', where $D = T(D')$.

3. Examples

In this section, we use the two transformations above to find uniform designs in several special domains which are commonly used. These examples demonstrate how to transform *gurp* sets in a unit hypercube to those in a general domain by using a DIPT or a DEPT.

First we consider an example of finding *gurp* in a p-dimensional unit ball or ball ring, which is fundamental for finding *gurp* for many other domains.

Example 3. Let $D = \{\boldsymbol{x} = (x_1, \ldots, x_p);\ 0 \le \delta \le \sum_{i=1}^p x_i^2 \le 1\}$. The following two transformations are its DEPT and DIPT respectively:

$$\begin{aligned}
x_1 &= (pv_1)^{1/p} \sin(F_2^{-1}(v_2)) \cdots \sin(F_{p-1}^{-1}(v_{p-1})) \sin v_p, \\
x_2 &= (pv_1)^{1/p} \sin(F_2^{-1}(v_2)) \cdots \sin(F_{p-1}^{-1}(v_{p-1})) \cos v_p, \\
x_j &= (pv_1)^{1/p} \prod_{i=2}^{p-j+1} \sin(F_i^{-1}(v_i)) \cos(F_{p-j+2}^{-1}(v_{p-j+2})),\ j = 3, \ldots, p-1, \\
x_p &= (pv_1)^{1/p} \cos(F_2^{-1}(v_2)),
\end{aligned} \tag{10}$$

where $\frac{\delta^p}{p} \le v_1 \le \frac{1}{p}$, $0 \le v_j \le B(\frac{1}{2}, \frac{1}{2}(p-j+1))$, $j = 2, \ldots, p-1$, $0 \le v_p \le 2\pi$, $F_j(x) = \int_0^x \sin^{p-j} x dx$, F_j^{-1} stands for the inverse function of F_j, $B(x, y)$ denotes the Beta function,

$$\begin{aligned}
v_1 &= \frac{1}{p} t_1, \\
v_j &= B\Big(\frac{1}{2}, \frac{1}{2}(p-j+1)\Big) t_j, \quad j = 2, \ldots, p-1, \\
v_p &= 2\pi t_p,
\end{aligned} \tag{11}$$

and $\delta^p \le t_1 \le 1$, $0 \le t_j \le 1$, $j = 2, \ldots, p$.

The density-preserving property of (10) can be deduced as follows.

First, note that the transformation

$$\begin{aligned}
x_1 &= r \sin\theta_1 \cdots \sin\theta_{p-2} \sin\theta_{p-1}, \\
x_j &= r \sin\theta_1 \cdots \sin\theta_{p-j} \cos\theta_{p-j+1}, \quad j = 2, \ldots, p,
\end{aligned} \tag{12}$$

where $\delta \le r \le 1$, $0 \le \theta_j \le \pi$, $j = 1, \ldots, p-2$, $0 \le \theta_{p-1} \le 2\pi$, and with a convention $\sin\theta_0 = 1$, has the differential relationship

$$\bigwedge_{i=1}^{p} dx_i = r^{p-1} \sin^{p-2}\theta_1 \sin^{p-3}\theta_2 \cdots \sin\theta_{p-2} \Big(\bigwedge_{i=1}^{p-1} d\theta_i\Big) \bigwedge dr. \tag{13}$$

Again, let

$$\begin{aligned}
v_1 &= F_1(r) = \int_0^r r^{p-1} dr = \frac{1}{p} r^p, \\
v_j &= F_j(\theta_{j-1}) = \int_0^{\theta_{j-1}} \sin^{p-j} x dx, \quad j = 2, \ldots, p-1, \\
v_p &= \theta_{p-1},
\end{aligned} \tag{14}$$

where the domain of $(v_1,\ldots,v_p)$ is

$$\frac{\delta^p}{p}\le v_1\le\frac{1}{p},\ 0\le v_j\le B\Big(\frac{1}{2},\frac{1}{2}(p-j+1)\Big),\ j=2,\ldots,p-1,\ 0\le v_p\le 2\pi,$$

since $F_j(\pi)=B(\frac{1}{2},\frac{1}{2}(p-j+1))$.

Substituting (14) into (12) we obtain (10), and from (13) and (14) we see that $\bigwedge_{i=1}^p dx_i=\bigwedge_{i=1}^p dv_i$, showing that (10) is a DEPT. By Example 2, (11) is obviously a DIPT.

Example 4. Consider the p-dimensional unit sphere $S_p=\{\boldsymbol{x}=(x_1,\ldots,x_p);\ \sum_{i=1}^p x_i^2=1\}$. Its DEPT and DIPT can be respectively taken as

$$\begin{aligned}
x_1&=\sin(F_1^{-1}(v_1))\cdots\sin(F_{p-2}^{-1}(v_{p-2}))\sin v_{p-1},\\
x_2&=\sin(F_1^{-1}(v_1))\cdots\sin(F_{p-2}^{-1}(v_{p-2}))\cos v_{p-1},\\
x_j&=\prod_{i=1}^{p-j}\sin(F_i^{-1}(v_i))\cos(F_{p-j+1}^{-1}(v_{p-j+1})),\quad j=3,\ldots,p-1,\\
x_p&=\cos(F_1^{-1}(v_1)),
\end{aligned}\tag{15}$$

where $0\le v_j\le B(\frac{1}{2},\frac{1}{2}(p-j))$, $j=1,\ldots,p-2$, $0\le v_{p-1}\le 2\pi$, $F_j(x)=\int_0^x\sin^{p-j-1}x\,dx$, F_j^{-1} is the inverse function of F_j, and

$$\begin{aligned}
v_j&=B\Big(\frac{1}{2},\frac{1}{2}(p-j)\Big)t_j,\quad j=1,\ldots,p-2,\\
v_{p-1}&=2\pi t_{p-1},
\end{aligned}\tag{16}$$

where $0\le t_j\le 1$, $j=1,\ldots,p-1$.

Similar to Example 3, we can show the density-preserving property of (15). The only difference is that the transform determinant here is $|T'T|=\prod_{i=1}^{p-2}\sin^{p-i-1}\theta_i$ and its differential relationship is $(d\boldsymbol{x})=\prod_{i=1}^{p-2}\sin^{p-i-1}\theta_i(\bigwedge_{i=1}^{p-1}d\theta_i)$.

Example 5. Consider the p-dimensional simplex $D_s=\{(x_1,\ldots,x_p);\ \sum_{i=1}^p x_i=1$ with $x_i>0\}$. We have its DEPT

$$\begin{aligned}
x_1&=((p-1)v_1)^{1/(p-1)}((p-2)v_2)^{1/(p-2)}\cdots(2v_{p-2})^{1/2}\frac{v_{p-1}}{\sqrt{p}},\\
x_2&=((p-1)v_1)^{1/(p-1)}((p-2)v_2)^{1/(p-2)}\cdots(2v_{p-2})^{1/2}(1-\frac{v_{p-1}}{\sqrt{p}}),\\
x_j&=\prod_{i=1}^{p-j}((p-i)v_i)^{1/(p-i)}(1-((j-1)v_{p-j+1})^{1/(j-1)}),\quad j=3,\ldots,p-1,\\
x_p&=1-((p-1)v_1)^{1/(p-1)},
\end{aligned}\tag{17}$$

where $0 \le v_j \le \frac{1}{p-j}$, $j = 1, \ldots, p-2$, $0 \le v_{p-1} \le \sqrt{p}$, and DIPT

$$v_j = \frac{1}{p-j} t_j, \quad j = 1, \ldots, p-2, \quad v_{p-1} = \sqrt{p} t_{p-1}, \tag{18}$$

where $0 \le t_j \le 1,\ j = 1, \ldots, p-1$.

To deduce the density-preserving property of (17), we need to use an intermediate transformation

$$\begin{aligned} x_1 &= \sin^2\theta_1 \cdots \sin^2\theta_{p-2}\sin^2\theta_{p-1}, \\ x_j &= \sin^2\theta_1 \cdots \sin^2\theta_{p-j}\cos^2\theta_{p-j+1}, \quad j = 2, \ldots, p, \end{aligned} \tag{19}$$

$(0 \le \theta_j \le \pi,\ j = 1, \ldots, p-1,\ 0 \le \theta_p \le 2\pi)$ which has the transform determinant

$$|T'T| = \sqrt{p} \sum_{i=1}^{p-1} \sin^{2(p-i)-1}\theta_i \cos\theta_i. \tag{20}$$

The rest of the deduction is similar to Example 3.

Usually, for simplicity we combine the two transformations into one.

Example 6. Let $D = \{\boldsymbol{x} = (x_1, \ldots, x_p);\ \sum_{i=1}^p x_i \le 1 \text{ with } x_i > 0\}$. We have a DEPT

$$\begin{aligned} x_1 &= (pv_1)^{1/p}((p-1)v_2)^{1/(p-1)} \cdots (2v_{p-1})^{1/2} v_p, \\ x_j &= (pv_1)^{1/p} \prod_{i=1}^{p-j} ((p-i)v_{i+1})^{1/(p-i)} (1 - ((j-1)v_{p-j+2})^{1/(j-1)}),\ j = 2, \ldots, p-1, \\ x_p &= (pv_1)^{1/p}(1 - ((p-1)v_2)^{1/(p-1)}), \end{aligned} \tag{21}$$

where $v_j = \frac{1}{p-j+1} t_j$, $0 \le t_j \le 1$, $j = 1, \ldots, p$.

Example 7. Let $D = \{(x_1, \ldots, x_p);\ 0 \le x_1 \le \cdots \le x_p \le 1\}$. We can take a DEPT

$$x_j = \prod_{i=1}^{p-j+1} t_i^{1/(p-i+1)}, \quad j = 1, \ldots, p,$$

where $0 \le t_i \le 1$, $i = 1, \ldots, p$.

Example 8. Let $D = \{(x_1, \ldots, x_p);\ \sum_{i=1}^p x_i = 0 \text{ and } \sum_{i=1}^p x_i^2 = 1\}$. We take a transformation $X = A_1 V$, where $(\boldsymbol{a}, A_1)$ is a $p \times p$ orthogonal matrix with $\boldsymbol{a} = \frac{1}{\sqrt{p}}(1, \ldots, 1)'$ as its first column, which changes domain $D_1 = \{(v_1, \ldots, v_{p-1});\ \sum_{i=1}^{p-1} v_i^2 = 1\}$ to domain D. Clearly, the transformation determinant is $|A_1' A_1| \equiv 1$. This means that the transformation is a DEPT. Therefore , we can find *gurp* in D by finding *gurp* in D_1, which has been done in Example 4.

The domain in this example is useful in statistics especially in experimental design, for we can consider D as the set of contrast directions for comparison.

4. The Method of Separating Variables and Uniform Design in the Stiefel Manifold

First we give the following lemma and corollary.

Lemma 1. *Let D be some domain in space R^p and let a random vector $X = (X_1, \ldots, X_p)$ have uniform distribution on D. Assume that*

$$x_j = x_j(u_1, \ldots, u_s, v_1, \ldots, v_t), \quad j = 1, \ldots, p, \quad s + t = p, \tag{22}$$

is a one-to-one differentiable transformation which maps the domain D onto a product space $D_1 \times D_2$, where $\boldsymbol{u} = (u_1, \ldots, u_s) \in D_1$ and $\boldsymbol{v} = (v_1, \ldots, v_t) \in D_2$, and assume it satisfies the differential form relationship

$$\bigwedge_{i=1}^{p} dx_i = \Big(\bigwedge_{i=1}^{s} du_i\Big)\Big(\bigwedge_{j=1}^{t} dv_j\Big). \tag{23}$$

Then, (22) is a DEPT, and U and V have uniform distributions in D_1 and D_2 respectively. Conversely, if U and V have uniform distributions on D_1 and D_2 respectively and (22) satisfies (23), then X has a uniform distribution in D.

Proof. By Definition 2 the density-preserving property of (22) is a direct result. The second part of the conclusion is obvious. Also the converse is clear.

By Lemma 1 the following corollary follows immediately.

Corollary 2. *Under the assumptions of Lemma 1, if $\mathcal{F}_l = \{\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip}), i = 1, \ldots, l\}$ is a gurp set in D, then $\mathcal{L}_l = \{\boldsymbol{u}_i = (u_{i1}, \ldots, u_{is}), i = 1, \ldots, l\}$ and $\mathcal{M}_l = \{\boldsymbol{v}_i = (v_{i1}, \ldots, v_{it}), i = 1, \ldots, l\}$ are respectively a gurp set in D_1 and D_2, where $x_{ij} = x_j(u_{i1}, \ldots, u_{is}, v_{i1}, \ldots, v_{it})$, $j = 1, \ldots, p$, $i = 1, \ldots, l$.*

We give the method above a name, the method of separating variables, for the method can transform a *gurp* set in a domain into two *gurp* sets in two domains separately.

Using this method, we consider uniform designs in the Stiefel manifold. First, we describe an application of such a design in the projection pursuit method.

In projection pursuit (PP) methods, one usually observes the projections of a set of high-dimensional data on some low-dimensional space (projection space) to explore the structure and character of this set of data (Huber (1985)). What we can do is select a finite number of projection spaces to observe. So, how to choose the projection spaces becomes an important problem in PP methods.

As is well known, when we have no prior knowledge about the structure and character of these data, to get most information about it, the selected projection spaces should be scattered uniformly in all the possible projection spaces. The existing method of finding projection spaces is to first take a projection space as an original position and then rotate the space according to some rule to get all spaces we want and to observe the projection on these spaces (Friedman and Tukey (1974)). But we are not sure if the selected spaces have uniformity in all the spaces. Furthermore, if the dimension of the considered space is very high, say 20, it is not easy for the rotation method to give consideration to every possible position of projection spaces, because the angles needed to rotate are too numerous and the position of the projection space in a high-dimensional space is too complicated. In practice, in order to quickly find a best projection space, we can consider the following strategy. First choose a few projection spaces to observe so that we get a rough understanding of the set of data to determine an initial position around which we might find a better structure of the data, and then take a careful observation around the initial position to get the best projection space. However, how to choose the few projection spaces, remains perhaps a very difficult problem for the existing method. Under these cases, a new method of choosing projection spaces with uniformity is needed.

Since every m-dimensional projection space is determined by its orthogonal coordinate system, all the possible m-dimensional projection spaces can be regarded as the Stiefel manifold $V_{m,n} = \{H_1(n \times m);\ H_1'H_1 = I_m\}$ with a set of m standard orthogonal vectors as a point. Our purpose is to choose a set of points $\mathcal{F}_l = \{H_{1i}; H_{1i} \in V_{m,n}, i = 1, \ldots, l\}$ for a given l such that $\mathcal{F}_l$ has uniformity in $V_{m,n}$.

According to Muirhead (1982), we can define a measure in $V_{m,n}$ by the differential form:

$$(H_1'dH_1) = \bigwedge_{i=1}^{m} \bigwedge_{j=i+1}^{n} \boldsymbol{h}_j' d\boldsymbol{h}_i, \tag{24}$$

where $H_1 = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_m]$ and $H_2 = [\boldsymbol{h}_{m+1}, \ldots, \boldsymbol{h}_n]$ such that $[H_1 \vdots H_2]$ is an orthogonal matrix. $H_1'H_2 = 0$. This measure is invariant under orthogonal transformations, and so is called a Haar invariant measure. In fact, the above differential form stands for the volume element of $V_{m,n}$ and the volume of $V_{m,n}$ is $2^m \pi^{mn/2}/\Gamma_m(\frac{1}{2}n)$, where $\Gamma_m(a) = \pi^{m(m-1)/4} \prod_{i=1}^{m} \Gamma(a - \frac{1}{2}(i-1))$.

Thus, actually we need to find a *gurp* set in $V_{m,n}$ with respect to the measure $(H_1'dH_1)$.

Lemma 2. *Let Z be an $n \times m$ $(m \leq n)$ matrix of rank m and write $Z = H_1T$, where $H_1 \in V_{m,n}$ and T is an upper-triangular matrix with positive diagonal elements. Let H_2 (as a function of H_1) be an $n \times (n-m)$ matrix such that*

$H = [H_1 \vdots H_2]$ is an orthogonal matrix. Denote $H_1 = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_m]$ and $H_2 = [\boldsymbol{h}_{m+1}, \ldots, \boldsymbol{h}_n]$. Then

$$(dZ) = \prod_{i=1}^{m} t_{ii}^{n-i}(dT)(H_1' dH_1), \tag{25}$$

where $(H_1' dH_1)$ is as given in (24) (Muirhead (1982, p63)).

Remark. We use the following facts. Let $Z = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m]$ be an $n \times m$ matrix of rank m and $M_{m,n}$ denote all of $n \times m$ matrices of rank m. Use Z^v to denote the stretched vector of Z, i.e. $Z^v = (\boldsymbol{z}_1', \ldots, \boldsymbol{z}_m')$, $\|\boldsymbol{v}\|$ the norm of vector $\boldsymbol{v}$, and D the unit sphere in R^{mn} with centre at the origin. Let $D_M = \{Z;\ Z \in M_{m,n}, \|Z^v\|^2 \leq 1\}$. Because $\mathcal{J} = \{Z; Z \in M_{m,n}, |Z'Z| = 0\}$ is a low-dimensional hypersurface, the volume of $\mathcal{J}$ in R^{mn} is equal to zero. Therefore, the volume of D_M in D is equal to that of D (D_M being Lebesgue-measurable). We can treat D and D_M as being the same in some sense. For example, let a random vector X have a uniform distribution in D; then we can say that X also has the uniform distribution in D_M because $P(X \in D - D_M) = 0$.

Thus, we introduce an important theorem which can be used to construct a uniform design in the Stiefel manifold.

Theorem 1. *Let $\mathcal{F}_l = \{Z_i; Z_i \in D_M, i = 1, \ldots, l\}$ be a gurp set in D_M. Then $\mathcal{L}_l = \{H_{1i}, i = 1, \ldots, l\}$ is a gurp set in $V_{m,n}$ with respect to the measure $(H_1' dH_1)$, where for every i, H_{1i} satisfies $Z_i = H_{1i}T_i, H_{1i} \in V_{m,n}$ and T_i is an upper-triangular matrix with positive diagonal elements.*

Proof. Consider the transformation

$$Z = H_1 T, \tag{26}$$

where H_1 and T are as in Lemma 2. From Lemma 2 it follows that (26) is a one-to-one differentiable transformation and satisfies the equality (25). Furthermore, let

$$u_{ij} = \begin{cases} t_{ij}, & i < j, \\ \frac{1}{n-i+1} t_{ii}^{n-i+1}, & i = j, \\ 0, & i > j. \end{cases} \tag{27}$$

Writing (27) in matrix form $U = K(T)$, where both U and T are upper-triangular matrixes with positive diagonal elements, we have

$$(dU) = \prod_{i=1}^{m} t_{ii}^{n-i}(dT). \tag{28}$$

Using $T = K^{-1}(U)$ to denote the inverse function of (27) and substituting it into (26), we get the transformation

$$Z = H_1 K^{-1}(U) \tag{29}$$

which has the differential form

$$(dZ) = (dU)(H_1' dH_1). \tag{30}$$

This means that (29) is a DEPT.

Now we discuss what the image domains of D_M are under transformations (26) and (29). First, note that, for any $H_1 \in V_{m,n}$, because $H_1' H_1 = I_m$, we have

$$|Z'Z| = |T'T| = |(K^{-1}(U))'(K^{-1}(U))| \tag{31}$$

and

$$\|Z^v\|^2 = \|T^v\|^2 = \|(K^{-1}(U))^v\|^2, \tag{32}$$

where T and $K^{-1}(U)$ satisfy (26) and (29) respectively.

Write the domains

$$D_M = \Big\{ Z(n \times m) = (z_{ij});\ \ |Z'Z| > 0, \|Z^v\|^2 = \sum_{i,j} z_{ij}^2 \le 1 \Big\},$$

$$D_1' = \Big\{ T(m \times m) = (t_{ij});\ t_{ij} = 0 \text{ for all } i > j,\ t_{ii} > 0 \text{ for all } i, \text{ and } \|T^v\|^2 = \sum_{i \le j} t_{ij}^2 \le 1 \Big\},$$

and

$$D_1'' = \Big\{ U(m \times m) = (u_{ij});\ u_{ij} = 0 \text{ for all } i > j, u_{ii} > 0 \text{ for all } i, \text{ and } \|(K^{-1}(U))^v\|^2 = \sum_{i=1}^m ((n-i+1)u_{ii})^{2/(n-i+1)} + \sum_{i<j} u_{ij}^2 \le 1 \Big\}.$$

Therefore, by (31) and (32), under (26) the product domain $D_1' \times V_{m,n}$ is the image domain of D_M and under (29) the product domain $D_1'' \times V_{m,n}$ is that of D_M. Furthermore, since (29) is a DEPT and we have (30). By Lemma 1, then, when Z has a uniform distribution in D_M, U has a uniform distribution in D_1'' and H_1 has a uniform distribution in $V_{m,n}$ with respect to the measure $(H_1' dH_1)$. Finally, according to Corollary 2, when $\mathcal{F}_l$ is a *gurp* set in D_M, through the transformation (29) we get a *gurp* set $\mathcal{L}_l$ in $V_{m,n}$. The proof is completed.

Finally, by Theorem 1 we give the following algorithm for finding *gurp* in $V_{m,n}$:

1. By using the number-theoretic method (Wang and Fang (1981)) or other methods (such as the OA-based Latin hypercube (Owen (1992) and Tang (1993)) and uniform design sampling (Zhang and Wang (1994)), we generate a *gurp* set in an mn-dimensional unit cube of R^{mn}, say $\mathcal{F}_l = \{\boldsymbol{t}_i = (t_{i1}, \ldots, t_{i,mn}), i = 1, \ldots, l\}$;

2. Substitute all the points of $\mathcal{F}_l$ into formula (11) with $p = mn$ to get a *gurp* set, say $\mathcal{F}_l' = \{\boldsymbol{v}_i = (v_{i1}, \ldots, v_{i,mn}), i = 1, \ldots, l\}$, in the following hyper-rectangle D' of R^{mn}:

$$\begin{aligned} &0 \leq v_1 \leq \frac{1}{mn}, \\ &0 \leq v_i \leq B(\frac{1}{2}, \frac{1}{2}(mn - i + 1)), \quad i = 2, \ldots, mn - 1, \\ &0 \leq v_{mn} \leq 2\pi; \end{aligned} \tag{33}$$

3. Substitute all the points of $\mathcal{F}_l'$ into formula (10) with $p = mn$ to generate a *gurp* set in the mn-dimensional unit ball in R^{mn}, say $\mathcal{F}_l^{''} = \{\boldsymbol{x}_i = (x_{i1}, \ldots, x_{i,mn}), i = 1, \ldots, l\}$;
4. Using the matrix transformation $Z = H_1 T$ (formula (26)), for every i, take Z_i such that $Z_i^v = (x_{i1}, \ldots, x_{i,mn}) \in \mathcal{F}_l^{''}$ and find $H_{1i} \in V_{m,n}$. After doing so, we get the required *gurp* set $\mathcal{F}_l^{'''} = \{H_{1i}; H_{1i} \in V_{m,n}, i = 1, \ldots, l\}$ in $V_{m,n}$.

Especially, when $m = 1$, $V_{1,n}$ is a unit sphere in R^n. In this case, by Theorem 1, to find a *gurp* set in $V_{1,n}$ is to find a *gurp* set on the unit sphere in R^n. To do this, we only need to normalize all the vectors in $\mathcal{F}_l^{''}$ which is a *gurp* set in the unit ball in R^n obtained by step 3. This just verifies the correctness of (15) which is used to find *gurp* on a unit sphere.

We note the number of independent variables in H_1 is $(2mn - m(m+1))/2$. But here we have used mn variables to generate the $m \times n$ column-orthogonal matrix H_1, so $m(m+1)/2$ freedom degrees of variable are wasted. Using a similar transformation, we have another method of generating uniform designs in the Stiefel manifold, which does not waste freedom degrees but needs more mathematical proof (for details see Zhang and Fang (1993)).

Acknowledgements

The author wishes to thank Professors C. F. Jeff Wu, W. J. Welch, and Dr. Youyi Chen for helpful discussions and comments. The research was supported by the Natural Sciences and Engineering Research Council of Canada and the Manufacturing Research Corporation of Ontario, under the sponsorship of Professor C. F. Jeff Wu, while the author was visiting the University of Waterloo, and partly was supported by NNSF of China. Also, the author thanks the referees for their comments and suggestions.

References

Friedman, J. H. and Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. IEEE Trans. Comput. **23**, 881-889.

Hua, L. K. and Wang, Y. (1981). Applications of Number Theory to Numerical Analysis. Springer-Verlag, Berlin and Beijing.

Huber, P. J. (1985). Projection pursuit (with discussion). Ann. Statist. **13**, 435-525.

McKay, M. D., Conover, W. J. and Beckman, R. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics **21**, 239-245.

Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory. John Wiley, New York.

Owen, A. B. (1992). Orthogonal arrays for computer experiments, integration and visualization. Statist. Sinica **2**, 439-452.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. Ann. Math. Statist. **23**, 470-472.

Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. Technometrics **29**, 143-151.

Tang, B. (1993). Orthogonal array-based Latin hypercubes. JASA **88**, 1392-1397.

Wang, Y. and Fang, K. T. (1981). A note on uniform distribution and experimental design. Kexue Tongbao **26**, 485-489.

Wang, Y. and Fang, K. T. (1990a). Number theoretic methods in applied statistics. Chinese Ann. Math. Ser.B **1**, 51-65.

Wang, Y. and Fang, K. T. (1990b). Number theoretic methods in applied statistics (II). Chinese Ann. Math. Ser.B **3**, 384-394.

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J. and Morris, M. D. (1992). Screening, Predicting, and Computer Experiments. Technometrics **34**, 15-25.

Weyl, H. (1916). Uber die Gleichversteiluing der Zahlen mod. Eins. Math. Ann. **77**, 313-352.

Zhang, R. C. and Fang, K. T. (1993). On design of column-orthogonal matrix with uniform distribution. Technical Report, Dept. of Math., HKBC, Hong Kong.

Zhang, R. C. and Wang, Z. J. (1994). Uniform design sampling and its fine properties. The Fifth Japan-China Symposium on Statistics, Okayama, 343-346.

Department of Mathematics, Nankai University, Tianjin 300071.

(Received November 1991; accepted July 1995)