INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 4, Number 3-4, Pages 729–743

© 2007 Institute for Scientific
Computing and Information

LATINIZED, IMPROVED LHS, AND CVT POINT SETS IN HYPERCUBES

YUKI SAKA, MAX GUNZBURGER, AND JOHN BURKARDT

(Communicated by Steve Hou)

Abstract. It is shown how an arbitrary set of points in the hypercube can be Latinized, i.e., can be transformed into a point set that has the Latin hypercube property. The effect of Latinization on the star discrepancy and other uniformity measures of a point set is analyzed. For a few selected but representative point sampling methods, evidence is provided to show that Latinization lowers the star discrepancy measure. A novel point sampling method is presented based on centroidal Voronoi tessellations of the hypercube. These point sets have excellent volumetric distributions, but have poor star discrepancies. Evidence is given that the Latinization of CVT points sets greatly lowers their star discrepancy measure but still preserves superior volumetric uniformity. As a result, means for determining improved Latin hypercube point samples are given.

Key Words. Latin hypercube sampling, quasi-Monte Carlo sampling, centroidal Voronoi tessellations, uniform sampling

1. Introduction

Point sampling in regions in $\mathbb{R}^d$ is useful in many areas of scientific computing and, depending on the specific application, it comes in different forms in terms of the dimensionality of the region and the cardinality and distribution of the samples. One example is the numerical integration of high-dimensional functions in hypercubes. In such applications, quadrature points must be chosen from a possibly very high-dimensional space in such a way that the quadrature error asymptotically converges to zero at a rate which is independent or weakly dependent on the dimension. Since usually no prior assumption is made about the smoothness or variation of the integrands, the points are sampled uniformly from the hypercube. Another example is mesh generation for which the points are typically chosen to belong to a low-dimensional complicated domain. The points are used to define a discrete approximant of some function which is hoped to converge to the true function as the number of points goes to infinity. One also hopes that the points are distributed in such a way so that optimal convergence rates are obtained, the discrete problem is well conditioned, and as few points as possible are used to achieve a desired accuracy. This often results in the need for nonuniform point distributions in general regions. A third example is in the design of experiments of both the laboratory and computational type. Here, parameters are chosen to define the experimental setup. Since experiments of either type may be expensive, and since

Received by the editors March 20, 2006.
2000 *Mathematics Subject Classification.* 62K10, 51E05, 62K05, 62L05, 62H10.

often many parameters serve to define an experiment, one would like to sample as few points as possible from a possibly high-dimensional parameter space. In addition, with the absence of any prior information about the system, one may need to sample uniformly in the parameter volume. Hence, this is a case of sparse, uniform sampling in high dimensions. A particular design of experiment problem arises in the model reduction of complex systems, where parameters are chosen to generate high-fidelity simulations called snapshots from which the reduced-order model is derived. The choice of snapshots is crucial to the accuracy of reduced-order models since those models can only capture the information contained in the snapshots. Since the high-fidelity simulation is expensive, one would like to sample parameter space as sparsely as possible.

In this paper, we focus on uniform sampling in the hypercube, possibly in high dimensions. Because the sense of "uniformity" depends on the application, we consider two strategies for defining uniformly distributed point sets. One strategy is aimed at producing points sets whose projections onto lower-dimensional faces of the hypercube are themselves well distributed. Discrepancy measures are usually used to evaluate the quality of such point sets. There have been many ways proposed for defining low-discrepancy points sets in hypercubes [12]. Here, in an attempt to improve the discrepancy measure of arbitrary point sets, we introduce a simple procedure that *Latinizes* any point set, i.e., that converts a point set into another set of "nearby" points that has the Latin hypercube sampling property.

The second strategy for defining uniformly distributed point sets is aimed at producing point sets that are well-distributed volumetrically in the hypercube. We introduce a specific strategy for accomplishing this goal that is based on minimizing the variance or second moment of tessellations associated with the point set. We refer to the resulting point sets as CVT (for centroidal Voronoi tessellation) point sets. There are several measures that can be used to evaluate the quality of volumetrically distributed point sets; here, we use several measures based on the Voronoi diagrams associated with point sets. CVT point sets, although superior with respect to volumetric measures of quality, have poor discrepancies. The Latinization of CVT point sets significantly improves their associated discrepancy measures.

Through some computational examples, we test the quality of representative methods for defining point sets and compare them to Latinized and CVT point sets. The comparisons are made with respect to several quality measures.

2. Quality measures for point sets

We use two types of measures to determine the quality of point sets in a hypercube. The first examines the uniformity of the set projected onto lower-dimensional faces while the second only looks at the volumetric uniformity of the points. Throughout, $H = [0,1]^d$ will denote the d-dimensional hypercube.

2.1. The star discrepancy. The *star discrepancy* [12, 13] of a point-set $Z = \{z_i\}_{i=1}^N \subset H$ is given by

$$D^*(Z) = \sup_{B_0 \subset H} \left| \frac{\#(Z \cap B_0)}{N} - \mu(B_0) \right|,$$

where $B_0 = [0, v_1] \times \cdots \times [0, v_d]$ for some $v_1, \ldots, v_d \in [0,1]$.

The star-discrepancy measures how well the point set can approximate the volume of axis-parallel boxes. This measure turns out to play an important role in high-dimensional integration. The Koksma-Hlawka inequality [12, 13] states that

if $f : H \mapsto \mathbb{R}$ has a finite total variation $\mathcal{V}(f)$, then the following error bound holds for the simple sampling-and-averaging quadrature rule based on a point set $Z = \{z_i\}_{i=1}^N$:

$$\left| \int_{[0,1]^d} f(x)\,dx - \frac{1}{N}\sum_{i=1}^{N} f(z_i) \right| \leq \mathcal{V}(f) D^*(Z). \tag{1}$$

2.2. Uniformity measures based on Voronoi tessellations. The Voronoi diagram corresponding to a given set of points can be used to define several measures of the uniformity of the point set. Here, we recall four of these measures. We note that these are volumetric measures of uniformity in the sense that they view point sets as objects in $\mathbb{R}^d$ and do not especially measure the uniformity of point sets viewed in lower dimensions. Detailed descriptions of these and other measures can be found in [2, 3, 5, 6]. Here, we merely note that these measures attempt to assess one or more of the hallmarks of volumetrically uniform point sets, namely, equal spacing between points, good coverage of the hypercube (e.g., no large regions having relatively few points), and some degree of isotropy.

Let $\|\cdot\|$ denote the Euclidean distance in $\mathbb{R}^d$ and μ the Lebesgue measure on H. A tessellation of H is a set $\{V_i\}_{i=1}^N$ having the following properties: for $i = 1, \ldots, N$, $V_i \subset H$ and is closed, $\cup_{i=1}^N V_i = H$, and $\mu(V_i \cap V_j) = 0$ whenever $i \neq j$. A *Voronoi tessellation* of H corresponding to a given set $\{z_i\}_{i=1}^N$ of N points belonging to H is given by, for $i = 1, \ldots, N$,

$$V_i = \{\, x \in H \;:\; \|x - z_i\| \leq \|x - z_j\| \quad \text{for } j = 1, \ldots, N,\, j \neq i \,\}.$$

Such a set $\{V_i\}_{i=1}^N$ is also referred to as a *Voronoi diagram* of H, the members of the set of points $\{z_i\}_{i=1}^N$ are referred to as the *generators* of the Voronoi tessellation, and each V_i is referred to as the *Voronoi region* or *Voronoi cell* corresponding to z_i.

For each Voronoi region V_i, let $\mu(V_i)$ denote its volume, let h_i denote the maximum distance between its generator z_i and the points in V_i, and let χ_i denote the minimum distance from its generator z_i to any of the other generators, i.e., for $i = 1, \ldots, N$,

$$\mu(V_i) = \int_{V_i} dx, \qquad h_i = \max_{x \in V_i} \|x - z_i\|, \qquad \text{and} \qquad \chi_i = \min_{j=1,\ldots,N,\, j \neq i} \|z_i - z_j\|.$$

Also, for each Voronoi region V_i, let $\overline{x}_i$ denote its center of mass and let M_i denote its second-moment tensor with respect to its center of mass, i.e.,

$$\overline{x}_i = \frac{1}{\mu(V_i)} \int_{V_i} x\,dx \qquad \text{and} \qquad M_i = \frac{1}{\mu(V_i)} \int_{V_i} (x - z_i)(x - z_i)^T\,dx.$$

Let $T_i = \operatorname{trace}(M_i)$ and $\overline{T} = \frac{1}{N}\sum_{i=1}^N T_i$. Also, let D_i denote the determinant of the deviatoric tensor associated with V_i, i.e., $D_i = \det(M_i - \frac{1}{N} T_i)$. Then, the four volumetric uniformity measures we will use are defined by

$$\begin{cases} h = \max\limits_{i=1,\ldots,N} h_i \\ \chi = \max\limits_{i=1,\ldots,N} \left(2\dfrac{h_i}{\chi_i}\right) \\ \tau = \max\limits_{i=1,\ldots,N} |T_i - \overline{T}| \\ \gamma = \max\limits_{i=1,\ldots,N} |D_i|. \end{cases} \tag{2}$$

In each case, the smaller the value of the uniformity measure, the more uniform is the point set; again, see [2, 3, 5, 6] for details.

3. CVT point sets

In Section 2.2, we defined Voronoi tessellations $\{V_i\}_{i=1}^N$ of a given point set $\{z_i\}_{i=1}^N$ in the hypercube $H = [0,1]^d$. The *centers of mass* or *centroids* of given a tessellation $\{W_i\}_{i=1}^N$ of H are the points $\{z_i^*\}_{i=1}^N$ such that, for $i = 1, \ldots, N$,

$$z_i^* = \frac{1}{|W_i|} \int_{W_i} x \, dx \,,$$

where $|W_i|$ denotes the measure of W_i. Note that if $\{V_i\}_{i=1}^N$ is a Voronoi tessellation corresponding to a given set of points $\{z_i\}_{i=1}^N$ and if, for $i = 1, \ldots, N$, z_i^* is the centroid of the Voronoi region V_i, then, in general, $z_i^* \neq z_i$ for $i = 1, \ldots, N$, i.e., the generating points of a Voronoi tessellation are, in general, *not* the centers of mass of the corresponding Voronoi regions; see the left image in Figure 1.

Given any set $\{w_i\}_{i=1}^N$ of N points in H and given any tessellation $\{W_i\}_{i=1}^N$ of H, we define, for $i = 1, \ldots, N$, the second moment of each pair $\{w_i, W_i\}$ by

$$\mathcal{E}_i(w_i, W_i) = \int_{W_i} \|x - w_i\|^2 \, dx \,. \tag{3}$$

Note that, at this point, we do not assume any relation between a point w_i and its associated volume W_i; in particular, we do not presume that $w_i \in W_i$ or that $\{w_i, W_i\}_{i=1}^N$ defines a Voronoi tessellation of H. The average of the second moments is then given by

$$\mathcal{E}\left(\{w_i\}_{i=1}^N, \{W_i\}_{i=1}^N\right) = \frac{1}{N} \sum_{i=1}^N \int_{W_i} \|x - w_i\|^2 \, dx \,. \tag{4}$$

We then pose the following problem: *among all possible sets of N points belonging to H and all possible tessellations of H into N subregions, find a set of points and a tessellation that minimizes $\mathcal{E}(\cdot, \cdot)$.* It is known that solutions of this problem, i.e., minimizers of $\mathcal{E}(\cdot, \cdot)$, are special Voronoi tessellations for which *the generating points are also the centroids* of the corresponding Voronoi regions [4]. We call such a tessellation a *centroidal Voronoi tessellation*, or, for short, a CVT. The functional $\mathcal{E}(\cdot, \cdot)$ is referred to as the "CVT-energy." See the right image of Figure 1 for an example of a CVT.

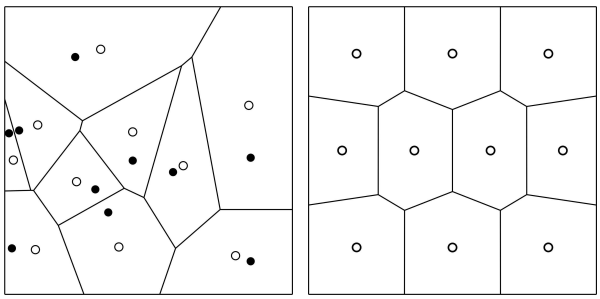

FIGURE 1. *Left: ten randomly chosen points in a square (the dots), the correspoding Voronoi tessellation of the square (the polygons), and the centers of mass of the Voronoi regions (the circles). Right: a ten point CVT of the square; the circles are simultaneously the generators of the Voronoi tessellation and the centers of mass of the Voronoi regions.*

The special nature of CVT point sets require that they be constructed. In low dimensions, i.e., two or at most three, they can be constructed by Lloyd's method [9] which is the obvious iteration consisting of first guessing the locations of the generators, then computing the Voronoi diagram for those points, then computing the centers of mass of the resulting Voronoi cells, and them using those centers of mass as the new guesses for locations of the generators. See [4, 9] for details. In higher dimensions, MacQueen's method [10] provides a sampling and averaging technique for constructing CVT's that does not require the construction of Voronoi diagrams or the determination of centers of mass of polyhedral domains. See [4, 10] for details. MacQueen's method is very slow to converge; more efficient probabilistic algorithms, i.e., algorithms based on sampling, have been developed, including some that are eminently parallelizable; see [8].

What does a CVT look like? Without any other constraints, it is easy to see that the optimal shape that minimizes the second moment (3) for a fixed volume is the ball. This is easy to see as follows. Let B denote a unit ball, W any measurable set of equal volume, and w an arbitrary point belonging to W. We can take $w = 0$ by translation. We have that $\mu(W - B \cap W) = \mu(B - B \cap W)$ so that

$$\begin{aligned}\int_{W-B\cap W} \|x\|^2\,dx &\geq \int_{W-B\cap W} dx = \mu(W - B \cap W)\\ &= \mu(B - B \cap W) = \int_{B-B\cap W} dx \geq \int_{B-B\cap W} \|x\|^2\,dx\,.\end{aligned}$$

Therefore,

$$\begin{aligned}\int_{B} \|x\|^2\,dx &= \int_{B\cap W} \|x\|^2 + \int_{B-B\cap W} \|x\|^2\,dx\\ &\leq \int_{B\cap W} \|x\|^2 + \int_{W-B\cap W} \|x\|^2\,dx = \int_{W} \|x\|^2\,dx\,.\end{aligned}$$

Obviously, balls cannot tessellate H so that a CVT cannot consist of balls. However, intuitively speaking, a CVT would have the Voronoi regions be as close to balls as possible while tessellating H by hyperplanes. In the plane, it has in fact been proved that the minimizer of the CVT-energy is a hexagonal lattice, hexagons being the closest shape to a disc which can tile the plane. In fact, Gersho's conjecture of vector-quantization [7] states that the minimizer of the CVT-energy results in Voronoi regions having equal volumes and shapes that are rotated or translated copies of the same shape; this conjecture has been proven valid in $\mathbb{R}^2$ [11]. Boundaries place additional constraints on the minimizers of the CVT-energy so that a regular lattice cannot be achieved. However, in the sense that the CVT-energy is minimized, CVT tessellations come as close as possible to a regular lattice. See Figure 2 for a further example of a CVT in $\mathbb{R}^2$.

Following [7], under the assumption that Gersho's conjecture holds, we may estimate the CVT-energy corresponding to the optimal CVT points as a function of dimension. For each Voronoi region V_j, we define the normalized moment as

$$M_j = \frac{1}{|V_j|^{\frac{2+d}{d}}} \int_{V_j} \|x - z_j\|^2\,dx\,.$$

We can see that M_j is a translation, rotation, and scale invariant quantity when z_j is also appropriately transformed. Under Gersho's conjecture, all the M_j's and $|V_j|$'s are essentially the same, i.e., for $i = 1, \ldots, N$, $M_j \approx M$ for some M and

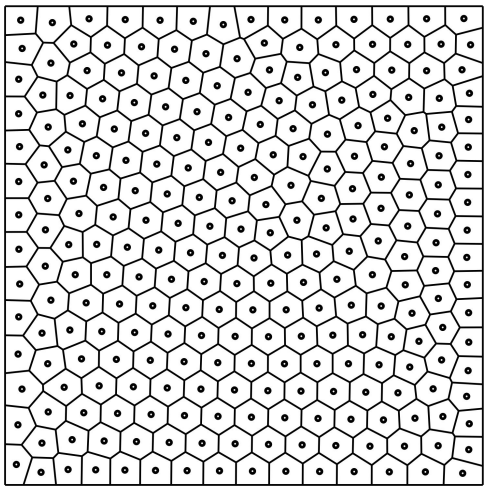

FIGURE 2. *A 256-point CVT of the square. Away from the boundary, the Voronoi regions approximate a regular hexagonal tiling.*

$|V_j| \approx 1/N$. Then,

$$\begin{aligned}\mathcal{E}\left(\{z_i\}_{i=1}^N, \{V_i\}_{i=1}^N\right) &= \sum_{j=1}^N \int_{V_j} \|x - z_j\|^2 \, dx = \sum_{j=1}^N |V_j|^{\frac{2+d}{d}} \int_{V_j} \frac{\|x - z_j\|^2}{|V_j|^{\frac{2+d}{d}}} \, dx \\ &= \sum_{j=1}^N |V_j|^{\frac{2+d}{d}} M_j \approx \frac{1}{N^{\frac{2}{d}}} M \, .\end{aligned} \tag{5}$$

It is interesting to note that the CVT-energy is strongly dependent on the dimension d. In fact, the bound (5) is relevant because it may be applied to obtain an error bound for a quadrature rule using CVT point sets as quadrature points. The following result is shown in [14]. We consider the following quadrature rule for approximating the integral of a function $f \in C^2(H)$:

$$\int_H f(x) \, dx \approx \sum_{i=1}^N f(z_i) \mu(V_i) \, . \tag{6}$$

By Taylor's theorem, we obtain

$$\begin{aligned}\int_H f(x) \, dx &= \sum_{i=1}^N \int_{V_i} \left[f(z_i) + \nabla f(z_i) \cdot (x - z_i) + (x - z_i) \cdot \left(\nabla^2 f(\eta_i)(x - z_i) \right) \right] dx \\ &= \sum_{i=1}^N \Big[f(z_i)|V_i| + \nabla f(z_i) \cdot \int_{V_i} (x - z_i) \, dx \\ &\qquad + \int_{V_i} (x - z_i) \cdot \left(\nabla^2 f(\eta_i)(x - z_i) \right) dx \Big] \, .\end{aligned}$$

Since z_i is the centroid of the corresponding Voronoi region V_i, the first-order term vanishes. Therefore,

$$\begin{aligned}\left| \int_H f(x) \, dx - \sum_{i=1}^N |V_i| f(z_i) \right| &\leq \|f\|_{C^2(H)} \sum_{i=1}^N \int_{V_j} \|x - z_i\|^2 \, dx \\ &= \|f\|_{C^2(H)} \mathcal{E}\left(\{z_i\}_{i=1}^N, \{V_i\}_{i=1}^N\right) .\end{aligned} \tag{7}$$

Thus, if Gersho's conjecture holds, we see that the quadrature error is bounded from above by a term of $O(\frac{1}{N^{2/d}})$. Note that if h_i denotes the diameter of the Voronoi region V_i, and if $h = \max_{i=1,\ldots,N} h_i$, then (7) implies that the error for the quadrature rule (6) is of $O(h^2)$. One should note the difference between the sampling and averaging quadrature rule of (1) and the volume-weighted quadrature rule of (6).

4. Latinization of point sets

A *Latin-hypercube sample* (LHS) can be defined as follows. Let $H = [0,1]^d$ denote the d-dimensional hypercube. For $i = 1,\ldots,N$ and $j = 1\ldots,d$, let $z_i(j)$ denote the j-th coordinate of the i-th point z_i. Let P_j, $j = 1,\ldots,d$, denote d random permutations of the set $\{1,\ldots,N\}$ and, for $i = 1,\ldots,N$, let $P_j(i)$ denote the i-th element of P_j. Then, we choose

$$z_i(j) = \frac{P_j(i) - U_{ji}}{N} \quad \text{for } i = 1,\ldots,N \text{ and } j = 1\ldots,d, \tag{8}$$

where U_{ij} denotes a uniform random variable taking values on the unit interval. As a result, each slab of width $\frac{1}{N}$ along all coordinate directions contains exactly one point; we refer to this property as the *LHS property*. For example, in two-dimensions, we subdivide the unit square into an N by N array of bins, i.e, small squares of size $\frac{1}{N}$ by $\frac{1}{N}$. Then, we sample N points in such a way that each row and column of the array contains exactly one point. If, instead of (8), we use

$$z_i(j) = \frac{P_j(i) - .5}{N} \quad \text{for } i = 1,\ldots,N \text{ and } j = 1\ldots,d,$$

we obtain a *centered* LHS for which all the points are located at the centers of their bins.

Any pointset can be transformed into an LHS by a simple procedure we call *Latinization*. Suppose we are given a point set $Z = \{z_i\}_{i=1}^N$. For any integer k such that $1 \le k \le d$, we define:

- the *k-th reordering* of Z to be the point set $\{R_k z_i\}_{i=1}^N$ obtained by reordering Z according to the value of the k-th coordinates of the z_i's; ties can be arbitrarily broken;
- the *k-th shift* of Z to be the point set $\{S_k z_i\}_{i=1}^N$ such that

$$S_k z_i(j) = \begin{cases} z_i(j) & \text{if } j \neq k \\ \dfrac{i - U_{ji}}{N} & \text{if } j = k, \end{cases} \tag{9}$$

  where again U_{ij} denotes a uniform random variable taking values on the unit interval.

Then, starting with any point set $Z = \{z_i\}_{i=1}^N$, the corresponding Latinized point set $\{Lz_i\}_{i=1}^N$ is given by

$$Lz_i = \left(\Pi_{k=1}^d (S_k R_k)\right) z_i \quad \text{for } i = 1,\ldots,N.$$

By construction, the Latinized point set is an LHS. The k-th shift moves the reordered points parallel to the k-th axis while preserving the k-th coordinate ordering of the points. Latinization is the result of applying the shift to all coordinates.

Clearly, we could obtain a centered LHS from any point set by replacing U_{ji} in (9) by .5. For such a centered Latinized strategy, we define the *k-th shift energy* of

a point set $Z = \{z_i\}_{i=1}^N$ with respect to k-th shift to be

$$ES_k(Z) = \sup_{i=1,\ldots,N} \left\| \big(R_k z_i\big)(k) - \big(S_k R_k z_i\big)(k) \right\| = \sup_{i=1,\ldots,N} \left\| \big(R_k z_i\big)(k) - \frac{1}{N}(i - .5) \right\|.$$

The k-th shift energy is a measure of how close the reordering of the point set Z is to a point set that is completely uniform with respect to the k-th axis.

4.1. Discrepancies of Latinized point sets. A question that naturally arises is: how much does Latinization distort the star discrepancy of the original point set? We prove that under certain hypotheses, the star-discrepancy of the Latinized point set has the same asymptotic bound as that of the original point set. To do so, we need the following preliminary result.

Lemma 4.1. *Let $Z = \{z_i\}_{i=1}^N$ denote a point set in H. If $D^*(Z) \leq f(N)$, then $ES_1(Z) \leq f(N)$.*

Proof. Let $B = [0, s_i] \times [0, 1] \times \ldots \times [0, 1]$; then,

$$|s_i - s_i'| = |s_i - \frac{i}{N}| = \left| \mu(B) - \frac{\#(B \cap Z)}{N} \right| \leq f(N).$$

This is exactly the discrepancy of the box B which we know is at most $f(N)$ from the hypothesis. □

Theorem 4.2. *If $D^*(Z) \leq f(N)$ and $f(N) \geq \frac{C}{N}$ for some $C > 0$, then $D(S_1(Z)) \leq (2 + \frac{1}{C}) f(N)$.*

Proof. Let $B_0 = [0, c_1] \times \ldots \times [0, c_d]$. Let $A = c_2 \cdots c_d$. Let k be the largest index such that $s_k' \leq c_0$. Let $B_1 = [0, s_k] \times [0, c_2] \times \ldots \times [0, c_d]$. Note that $B_1 \cap Z$ and $B_0 \cap S_1(Z)$ contain the same number of points, since if v_j is in B_1, $v_j(1) = s_j \leq s_k$, so that $j \leq k$. This implies that $S_1 v_j(1) = \frac{j}{N} \leq \frac{k}{N} = s_k'$. Also, $v_j(k) = S_1 v_j(k) \in [0, c_k]$ for all $k \neq 1$. Thus, $S_1 v_j \in B_0$. On the other hand, if $S_1 v_j \in B_0$, by the maximality of k, $S_1 v_j(1) = s_j' \leq s_k'$. So $j \leq k$ and this implies $s_j \leq s_k$, so that $v_j \in B_1$. Thus, $v_j \in B_1$ if and only if $S_1 v_j \in B_0$. Also, because of how k was chosen, $|c_1 - s_k'| \leq \frac{1}{N}$. Also, $|s_k - s_k'| \leq f(N)$ by Lemma 4.1. Therefore,

$$|c_1 - s_k| \leq |c_1 - s_k' + s_k' - s_k| \leq |c_1 - s_k'| + |s_k' - s_k| \leq f(N) + \frac{1}{N}.$$

Therefore,

$$\begin{aligned} \left| \frac{\#(B_0 \cap S_1(Z))}{N} - \mu(B_0) \right| &= \left| \frac{\#(B_1 \cap Z)}{N} - A c_1 \right| \\ &= \left| \frac{\#(B_1 \cap Z)}{N} - A s_k + A s_k - A c_1 \right| \leq \left| \frac{\#(B_1 \cap Z)}{N} - A s_k \right| + A \left| s_k - c_1 \right| \\ &\leq f(N) + A(f(N) + \frac{1}{N}) \leq \Big(2 + \frac{1}{C}\Big) f(N). \end{aligned}$$

This is true for any box B_0; therefore, $D^*(S_1(Z)) \leq (2 + \frac{1}{C}) f(N)$. □

Thus, we see that the discrepancy of the kth shift is of the same order as the original point set, given that the original point set has discrepancy greater than $\frac{C}{N}$ for some C. A simple corollary of the above is that a point set after Latinization has the discrepancy bounded by $O(f(N))$, where the constant in the big-O notation depends on d.

4.2. CVT energy of Latinized point sets. We next ask what happens to the CVT energy of a point set $Z = \{z_i\}_{i=1}^N$ after it is Latinized? At this point, we need not assume that the point set is a CVT point set.

Suppose we start with a set of points $Z = \{z_i\}_{i=1}^N$ in H and a tessellation $\{V_1, ..., V_N\}$ of H such that $z_i \in V_i$ and the following properties hold:

(1) $\mu(V_i) \sim \frac{1}{N}$ in the sense that there exists $0 \leq \gamma < 1$ such that for all $i = 1, \ldots, N$, $\mu(V_i) \in [((1+\gamma)N)^{-1}, ((1-\gamma)N)^{-1}]$;
(2) for given $R > 0$, $|x(j) - z_i(j)| \leq R$ for all $x \in V_i$ and $j \leq d$.

The first property defines the notion of a quasi-uniform tessellation of H.

Then, we can obtain the following bound on the shift energy of the pointset. In the following lemma, we treat the first axis. The other axes follows in the same way. Hence, we let $s_i = z_i(1)$ and $s'_i = (S_1 R_1 z)_i(1) = \frac{i}{N}$.

Proposition 4.3. *Let the pointset Z generate a Voronoi diagram which is a quasi-uniform partition in the sense that the above two requirements are met. Then,*

$$ES_1(P) \leq R + \gamma.$$

Proof. Let $H_r = \{v \in H \ : \ (v - re_1) \cdot e_1 \leq 0\}$; clearly, H_r is the subset of H consisting of those points with a first coordinate less than or equal to r. Fix a $k \in \{1 \ldots N\}$ and let $A = \cup_{i=1}^k V_i$, $r_h = \sup_{x \in A} x(1)$, and $r_l = \inf_{x \in A^c} x(1)$. Then, $A \subset H_{r_h}$ and $\text{int}(H_{r_l}) \subset A$. Since $\mu(A) \sim \frac{k}{N}$, we have

$$(1-\gamma)\mu(H_{r_l}) \leq \frac{k}{N} \leq (1+\gamma)\mu(H_{r_h}).$$

But $\mu(H_r) = r$ so we have

$$(1-\gamma)r_l \leq \frac{k}{N} \leq (1+\gamma)r_h.$$

Suppose first that $\frac{k}{N} \geq s_k$. Since A is closed, there exists $x \in A$ such that $x(1) = r_h$. Then, $x \in V_p$ for some $p \leq k$ so $s_p \leq s_k$. Therefore,

$$0 \leq \frac{k}{N} - s_k \leq (1+\gamma)r_h - s_k \leq (1+\gamma)x(1) - s_p \leq R + \gamma.$$

Hence, $|s'_k - s_k| \leq R + \gamma$. On the other hand, if $\frac{k}{N} < s_k$, let $x \in \cup_{i=k+1}^N V_i$ such that $x(1) = r_l$. Then, $x \in V_p$ such that $p > k$ so $s_p > s_k$. Therefore,

$$0 \leq s_k - \frac{k}{N} \leq s_k - (1-\gamma)r_l \leq s_p - (1-\gamma)x(1) \leq R + \gamma$$

so that in this case we also have that $|s'_k - s_k| \leq R + \gamma$. □

We can now roughly bound the change in CVT energy due to a shift if the neighbors of each Voronoi region do not change after the shift. Let J_k be set of indices associated with the neighboring Voronoi regions of V_k. Let δ_k be such that $z_k + \delta_k = S_1 R_1 z_k$. Let Z' and $\{V'_k\}_{k=1}^N$ respectively denote the point set and Voronoi regions after the shift.

Note that due to the fact that $V_k = (V_k \cap V'_k) \cup \left(\cup_{j \in J_k} (V_k \cap V'_j) \right)$ and $V'_k = (V_k \cap V'_k) \cup \left(\cup_{j \in J_k} (V'_k \cap V_j) \right)$,

$$\begin{aligned} |\mathcal{E}(Z') - \mathcal{E}(Z)| &= \frac{1}{N} \sum_{k=1}^N \Bigg[\int_{V_k \cap V'_k} \|z_k + \delta_k - x\|^2 - \|z_k - x\|^2 \, dx \\ &\quad + \sum_{j \in J_k} \left(\int_{V'_k \cap V_j} \|z_k + \delta_k - x\|^2 \, dx + \int_{V_k \cap V'_j} \|z_k - x\|^2 \, dx \right) \Bigg]. \end{aligned}$$

Each of the three terms can be bounded as follows:

$$\begin{aligned}\int_{V_k\cap V_k'} \|z_k+\delta_k-x\|^2-\|z_k-x\|^2\,dx &= \int_{V_k\cap V_k'} 2\langle z_k-x+\frac{1}{2}\delta_k,\delta k\rangle \\ &\le \mu(V_k\cap V_k')2(R+\gamma)^2 \le C_1\frac{(R+\gamma)^2}{N},\end{aligned}$$

$$\sum_{j\in J_k}\int_{V_k'\cap V_j}\|z_k+\delta_k-x\|^2\,dx \le \sum_{j\in J_k}\mu(V_k'\cap V_j)4(R+\gamma)^2 \le C_2|J_k|\frac{(R+\gamma)^2}{N},$$

and

$$\sum_{j\in J_k}\int_{V_k\cap V_j'}\|z_k-x\|^2\,dx \le \sum_{j\in J_k}\mu(V_j'\cap V_k)4(R+\gamma)^2 \le C_3|J_k|\frac{(R+\gamma)^2}{N}.$$

Therefore, $|\mathcal{E}(Z)-\mathcal{E}(Z')| \le C\frac{(R+\gamma)^2}{N}$.

Note that the CVT energy of a uniform partition of H is bounded by $\frac{R^2}{N}$; therefore, the above bound indicates that a shift changes the CVT energy by the same order as the CVT energy of the original set.

5. Numerical experiments and conclusions

Five methods are used to sample 100 and 1000 points in the two, three, and seven-dimensional hypercube; the sampling methods used are centroidal Voronoi tessellation (CVT), Halton (HAL), Hammersley (HAM), Latin hypercube (LHS), and (IHS) which is the improvement on Latin hypercube described in [1]. Latinized versions (LCVT, LHAL, and LHAM) of the first three types of point sets are also determined.

The star discrepancy of each point set was approximately determined by the method of [15]; the results in seven dimensions are less accurate than the corresponding results in two and three dimensions. For each point set, approximations to the four Voronoi diagram-based measures described in Section 2.2 and the CVT energy (see (4)) were also determined using intense sampling methods to evaluate the integrals. Again, the results in seven dimensions are less accurate than the corresponding results in two and three dimensions.

We note that some of the results, especially in seven dimensions, are very sensitive to initialization. For example, the construction of CVT points sets require choosing initial positions for the points; an iterative process is then used to move the points to the CVT locations. The results of that process are sometimes sensitive to the initial position of the points.

The results obtained are given in Tables 1–6 (ignoring, for the moment, the last two columns). Of course, we present a very limited number of test runs so that drawing definitive conclusions with regard to the relative merits of the different point sets is not possible. However, the inferences listed below which can be drawn from the tables are also consistent with the results of other tests we have performed.

The star discrepancy

1. *Latinization improves the star discrepancy* of every CVT, Halton, and Hammersley point set.
2. Latinized Hammersley point sets consistently have the smallest star discrepancy.
3. CVT point sets have the largest star discrepancy, but the Latinization of CVT point sets greatly reduces that discrepancy to a value comparable to that for some of the other point sets.

The four Voronoi diagram-based measures

1. CVT point sets are consistently the best according to these measures.
2. LHS point sets are consistently the worst; IHS point sets are much better.
3. Latinization raises the measures for the CVT point sets, but LCVT point sets are still consistently better than any of the other points sets.
4. Latinization mostly but not always raises these measures for the Halton and Hammersley point sets.
5. Other than the CVT and LCVT point sets that have lower values and the LHS point sets that have higher values, the other points sets have similar values.

The CVT energy

1. Naturally, CVT point sets have the lowest value of the CVT energy since they are, by design, minimizers of that energy.
2. A Latinized CVT point set has higher CVT energy than that of its parent CVT point set, but generally lower energy than that of the other point sets.
3. Latinization seems to have little effect on the CVT energy of Halton and Hammersley point sets.

Summary

1. If the star discrepancy measure is of most importance, then the Latinized Hammersley point sets seem to be preferred.
2. If the four Voronoi diagram-based measures are of most importance, then the CVT point sets seem to be preferred.
3. If both measures are of interest, Latinized CVT point sets seem to provide the best compromise.
4. LHS and IHS point sets are not competitive in either category.

These observations are reinforced by a visual inspection of the point sets in two dimensions. The eye is a very good discerner of volumetric uniformity and nonuniformity, i.e., it can easily detect disparities in the spacing between points, the lack or presence of large areas that contain no points, and the isotropy or lack thereof of the point distribution. For example, consider the top row of Figure 3 which, visually, is arranged, left to right, in increasing disorder but which also is arranged according to increasing volumetric measures; see Table 1. From the the second row, one also sees that Latinization has little effect on the Halton and Hammersley points sets, but has a definite disordering effect on the CVT point set. Finally, from the third row, one sees that the IHS point set determined by the algorithm of [1] is a definite improvement over the Latin hypercube sampling point set, but that both are considerably less volumetrically uniform than the CVT point set.

The eye is not so good at picking out which point sets have smaller or larger star discrepancy. One exception, perhaps, is the CVT point set which is depicted at the top left of Figure 3. Not only does one see a lattice-like structure in the interior of the square, but even more influential with respect to the star discrepancy, the points are very well aligned along the sides of the square. This latter feature can be useful in some setting, e.g., grid generation, but is disastrous with respect to the star discrepancy since it results in very poor point distributions after the point set is projected onto the sides of the square.

The fact that the CVT point set "knows" so well where the boundary is located can be counteracted by having it "ignore" the boundary, i.e., by constructing a periodic (with respect to all coordinates) CVT. Now, the Voronoi cell corresponding to a point near the boundary of the hypercube A^s is not cut off by that boundary, but instead wraps around to the opposite $(s-1)$-dimensional face. This is an approach we are currently studying; some preliminary results seem promising; see the last two columns of Tables 1–6 and the last row of Figure 3 which provide the same information for periodic CVT points (CVTP) and their Latinizations (LCVTP) as was given for the other point sets. In particular, we see that the value of the star discrepancy of the CVTP points sets is much lower than that for the CVT point sets. Also, the LCVTP point sets have lower value of the star discrepancy than do the LHS and IHS points sets, and in this respect, they begin to rival the Halton and Hammersley point sets. On the other hand, with respect to most of the volumetric measures, the CVTP and LCVTP points sets seem to be superior to the Halton, Hammersley, Latin hypercube sampling, and improved LHS point sets and are only outdone by the CVT point sets.

Clearly more tests and theoretical studies need to be done, especially in higher dimensions and/or for larger point sets. However, CVT and periodic CVT point sampling seems to be useful in some settings and the Latinization of point sets such as Hammersley seems to produce very much improved Latin hypercube samples. Future work will also focus on the the effects of Latinization and on the relative merits of the different sampling strategies when point sets are used for the different purposes mentioned in the introduction.

References

[1] B. Beachkofski and R. Grandhi, Improved distributed hypercube sampling; AIAA Paper 2002-1274, AIAA, Washington, 2002.

[2] J. Burkardt, M. Gunzburger, J. Peterson, and R. Brannon , User manual and supporting information for library of codes for centroidal Voronoi point placement and associated zeroth, first, and second moment determination; SAND Report SAND2002-0099, Sandia National Laboratories, Albuquerque, 2002.

[3] J. Burkardt, M. Gunzburger, J. Peterson, and Y. Saka, Uniformity measures for point distributions, in preparation.

[4] Q. Du, V. Faber, and M. Gunzburger, Centroidal Voronoi tessellations: applications and algorithms; SIAM Review 41, 1999, 637-676.

[5] Q. Du, M. Gunzburger, and L. Ju, Constrained centroidal Voronoi tessellations for surfaces; SIAM J. Sci. Comput. 24 2003, 1488-1506.

[6] Q. Du, M. Gunzburger, and L. Ju, Voronoi-based finite volume methods, optimal Voronoi meshes, and PDEs on the sphere; Comp. Meth. Appl. Mech. Engrg. 192 2003, 3933-3957.

[7] R. Gray and D. Neuhoff, Quantization; IEEE Trans. Inform. Theory 44 1998, 2325-2383.

[8] L. Ju, Q. Du, and M. Gunzburger, Probabilistic algorithms for centroidal Voronoi tessellations and their parallel implementation; Parallel Comput. 28 2002, 1477-150.

[9] S. Lloyd, Least squares quantization in PCM; IEEE Trans. Inform. Theory 28 1982, 129-137.

[10] J. MacQueen, Some methods for classification and analysis of multivariate observations; in Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol I., L. Le Cam and J. Neyman, eds., University of California, Berkeley, 1967, 281-297.

[11] D. Newman, The hexagon theorem; IEEE Trans. Inform. Theory 28 1982, 2.

[12] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, SIAM, Philadelphia, 1992.

[13] A. Owen, Multidimensional variation for quasi-Monte Carlo; Technical Report, Stanford University, 2004.

[14] G. Pages, A space quantization method for numerical integration; J. Comp. Appl. Math. 89 1997, 1.

[15] E. Thiemard, An algorithm to compute bounds for star discrepancy; J. Complexity 17 2001, 850-88.

Measure	CVT	LCVT	HAL	LHAL	HAM	LHAM	LHS	IHS	CVTP	LCVTP
$D^* \times 10^2$	8.89	2.74	5.05	2.91	3.80	2.78	6.06	3.71	5.86	3.65
h	.072	.104	.125	.126	.140	.134	.165	.118	.109	.108
χ	1.56	7.19	5.18	5.87	3.91	4.19	15.7	4.51	2.17	2.46
$\tau \times 10^3$	.31	.59	1.64	1.18	1.20	1.13	2.83	1.14	.81	.78
$\gamma \times 10^7$	.26	2.50	6.63	6.22	4.79	4.75	22.3	4.45	1.10	1.02
$\mathcal{E} \times 10^3$	1.64	1.84	2.22	2.25	1.96	1.96	3.03	1.97	1.73	1.81

TABLE 1. 100 points in two dimensions.

Measure	CVT	LCVT	HAL	LHAL	HAM	LHAM	LHS	IHS	CVTP	LCVTP
$D^* \times 10^3$	31.1	1.5	7.29	5.64	4.81	3.71	19.2	8.87	16.3	8.42
$h \times 10^2$	2.29	3.10	3.60	3.69	4.08	4.07	5.93	3.72	3.34	3.35
χ	1.52	3.59	7.31	7.82	11.4	13.0	45.1	4.82	2.42	2.35
$\tau \times 10^4$	.63	.67	1.21	1.18	1.34	1.35	4.06	1.28	.97	1.12
$\gamma \times 10^9$	1.14	2.18	9.56	11.1	3.98	3.71	29.9	5.06	3.25	3.68
$\mathcal{E} \times 10^3$	.163	.171	.221	.221	.183	.184	.315	.182	.168	1.69

TABLE 2. 1000 points in two dimensions.

Measure	CVT	LCVT	HAL	LHAL	HAM	LHAM	LHS	IHS	CVTP	LCVTP
$D^* \times 10^2$	24.3	6.95	6.76	5.89	6.43	4.98	9.18	6.83	1.5	6.02
h	.200	.265	.312	.321	.269	.268	.350	.281	.281	.276
χ	1.94	15.2	9.64	9.69	8.00	7.33	13.7	5.29	2.53	2.85
$\tau \times 10^3$	1.40	4.58	5.57	5.37	6.11	5.24	1.6	5.57	4.67	4.16
$\gamma \times 10^8$	.19	1.24	3.88	3.98	1.19	1.49	11.7	4.31	1.70	3.13
$\mathcal{E} \times 10^2$	1.14	1.37	1.55	1.55	1.41	1.41	1.83	1.41	1.31	1.31

TABLE 3. 100 points in three dimensions.

Measure	CVT	LCVT	HAL	LHAL	HAM	LHAM	LHS	IHS	CVTP	LCVTP
$D^* \times 10^2$	11.8	1.97	1.64	1.48	1.33	1.17	3.28	2.00	2.53	1.83
h	.082	.114	.136	.136	.130	.129	.167	.129	.116	.123
χ	1.83	4.74	1.4	1.3	6.43	6.52	31.8	6.19	2.32	2.36
$\tau \times 10^3$	.63	1.43	1.49	1.63	1.28	1.24	3.00	1.63	1.83	2.02
$\gamma \times 10^{10}$	.38	6.43	5.22	6.93	2.25	2.99	17.0	3.59	2.99	2.60
$\mathcal{E} \times 10^3$	2.42	2.63	3.13	3.14	2.84	2.84	3.70	2.73	2.60	2.60

TABLE 4. 1000 points in three dimensions.

Measure	CVT	LCVT	HAL	LHAL	HAM	LHAM	LHS	IHS	CVTP	LCVTP
D^*	.870	.250	.248	.203	.237	.200	.228	.234	.222	.210
h	.718	.817	.956	.919	.863	.863	.908	.904	.964	.948
χ	2.63	5.18	4.05	4.03	4.69	4.73	6.56	4.05	3.26	3.23
$\tau \times 10^2$	.91	2.94	3.53	3.43	2.94	3.14	5.24	5.46	5.25	4.65
$\gamma \times 10^{14}$	.000108	1.79	6.15	6.36	1.8	11.3	14.3	17.8	2.26	1.89
$\mathcal{E}$	.156	.211	.212	.212	.206	.206	.215	.215	.203	.203

TABLE 5. 100 points in seven dimensions.

Measure	CVT	LCVT	HAL	LHAL	HAM	LHAM	LHS	IHS	CVTP	LCVTP
D^*	.741	.173	.132	.128	.125	.120	.139	.139	.137	.133
h	.552	.582	.600	.600	.587	.585	.622	.632	.569	.573
χ	2.80	4.41	5.85	5.86	4.36	4.34	7.55	3.78	2.77	2.74
$\tau \times 10^2$	1.23	2.64	2.84	2.70	4.69	4.69	3.16	3.39	3.03	2.68
$\gamma \times 10^{15}$	.012	.96	4.80	4.94	3.00	2.99	2.3	6.53	3.19	3.44
$\mathcal{E} \times 10^2$	7.86	9.50	9.80	9.81	9.83	9.83	1.1	9.87	9.38	9.36

TABLE 6. 1000 points in seven dimensions.

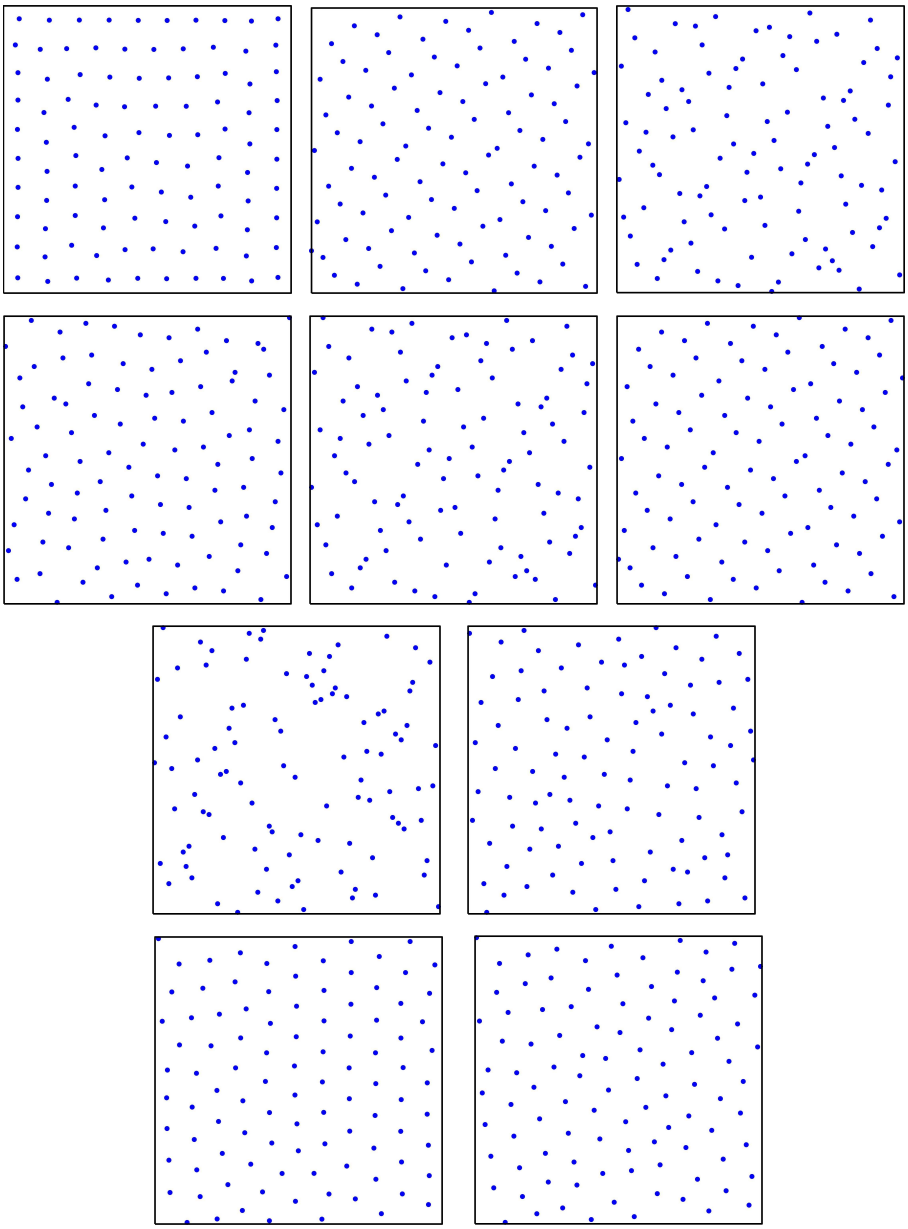

FIGURE 3. *100 points in the square; top row: CVT, HAL, and HAM; second row: LCVT, LHAL, and LHAM; third row: LHS and IHS; bottom row: CVTP and LCVTP.*

School of Computational Science Florida State University, Tallahassee FL 32306-4120
E-mail: saka@scs.fsu.edu, gunzburg@scs.fsu.edu, burkardt@scs.fsu.edu