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

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