Initial Evaluation of Centroidal Voronoi Tessellation Method for Statistical Sampling and Function Integration[*]

Vicente J. Romero
Sandia National Laboratories[†]
Albuquerque, NM

John V. Burkardt, Max D. Gunzburger[◊], Janet S. Peterson
School of Computational Science and Information Technology
Florida State University, Tallahassee, FL

Abstract

A recently developed Centroidal Voronoi Tessellation (CVT) unstructured sampling method is investigated here to assess its suitability for use in statistical sampling and function integration. CVT efficiently generates a highly uniform distribution of sample points over arbitrarily shaped M-Dimensional parameter spaces. It has recently been shown on several 2-D test problems to provide superior point distributions for generating locally conforming response surfaces. In this paper, its performance as a statistical sampling and function integration method is compared to that of Latin-Hypercube Sampling (LHS) and Simple Random Sampling (SRS) Monte Carlo methods, and Halton and Hammersley quasi-Monte-Carlo sequence methods. Specifically, sampling efficiencies are compared for function integration and for resolving various statistics of response in a 2-D test problem. It is found that on balance CVT performs best of all these sampling methods on our test problems.

1. Background

For reasons that will become clear later, it is often beneficial in statistical sampling and function integration to sample "uniformly" over the applicable parameter space.

Such uniformity, while conceptually simple and intuitive on a qualitative level, is on a quantitative level somewhat complicated to describe and quantify mathematically. Quantitative aspects of uniformity involve: 1) the equality with which points are spaced relative to one another in the parameter space (are they all nominally the same distance from one another?); 2) uniformity of point density over the entire domain of the parameter space (i.e., uniform "coverage" of the whole domain by the set of points, and not just good uniformity within certain regions of the space); and 3) isotropy in the point placement pattern. Each of these aspects of uniformity can be quantified by several mathematical measures. We will not discuss these measures further here, but we mention them to say that quantitative measures do exist for the intuitive notion of uniformity. We find that in 2-D the visual-intuitive sense of uniformity obtained by viewing a distribution of samples in a square (2-D hypercube) correlates very strongly with the quantitative quality measures mentioned above. Thus, in 2-D the eye is an excellent integrator of the different aspects of uniformity listed above, and a very accurate discriminator of uniformity or lack thereof –or at least in judging whether one particular layout of sample points is more uniform than another.

Much effort has been applied in the literature to the problem of achieving uniform placement of N samples over M-dimensional hypercubes, where M and N are both arbitrary. It is well recognized that Simple-Random sampling (SRS) Monte Carlo does not do a particularly good job of uniformly spreading out the sample points. The popular Latin Hypercube Sampling (LHS) method generally does a much better job of uniformly spreading out the points. This is due to the greater sampling regularity over each individual parameter dimension before the individually generated parameter values are randomly combined into parameter sets which define the coordinates of the sampling points ([5]).

Recent efforts to modify LHS to get an even more uniform distribution of points over the paramter space have included Distributed Hypercube Sampling (DHS, [12]) and Improved Distributed Hypercube Sampling (IHS, [2]). The fundamentals and history of these are reviewed in [18]. Though the quantitative measure of uniformity used for comparisons in [2] and [12] was somewhat flawed, it does appear that DHS gives better sampling uniformity

than LHS, and that IHS gives better sampling uniformity than DHS but is increasingly more computationally expensive as the dimensionality of the paramter space increases. We have recently become aware of another LHS variant, "Optimal Symmetric LHS" (OSLHS, [20]) which also seems to improve the spatial uniformity of LHS samples. Its computational cost and performance relative to DHS and IHS are not yet known, however.

A number of other potential approaches for achieving uniform point placement that are not evolved from an LHS basis are reviewed (and some new ones are presented) in [7]. There, some quantitative metrics related to visual/sensory perception of point uniformity in 2-D are reviewed and some new ones presented. Many of these non-LHS-based approaches appear to work very well in 2-D, but it is said that some of the methods may not be applicable or may not perform well in more than two dimensions, and some clearly will not scale up to high dimensions affordably. Others seem more promising for high dimensions, but have not yet been investigated enough.

The so-called "Quasi- Monte Carlo" (QMC) quasi- or sub- random low-discrepency sequence methods (see *e.g.* [14]) can often achieve reasonably uniform sample placement in hypercubes. The strength of these sequence methods (Halton, Hammersley, Sobol, etc.), is that they can produce fairly uniform point distributions even though samples are added one at a time to the parameter space. The one-at-a-time incremental sampling of QMC (and SRS) enables these methods to have better efficiency prospects than CVT and LHS-type methods in the area of error estimation and control. Not only this, the results achieved are often quite good. For resolving the mean and standard deviation of response measures, Hammersley sequences were found in [11] to converge to within 1% of exact results 3 to 100 times faster than LHS over a large range of test problems. For resolving response probabilities, Hammersley and modified-Halton were found in [15] to perform roughly the same as LHS on balance over several test problems.

However, when the hyperspace dimension becomes moderate to large and/or the sampling density becomes high, some (perhaps all?) sequences suffer from spurious correlation of the samples. This is shown for standard Halton sequences in 16-D (ref. [12]) and 40-D (ref. [15]). Sometimes a modification can be found to suppress or delay the onset of spurious correlation, as a fix from the literature implemented in [15] shows for Halton sequences.

Recently, a long-recognized approach for achieving uniformity of point placement in M-dimensional volumes, called "Centroidal Voronoi Tessellation" (CVT), has been made computationally efficient ([10]) for implementing the principles of Centroidal Voronoi diagrams ([6],[13]).

These diagrams subdivide arbitrarily shaped domains in arbitary-dimensional space into arbitrary numbers of nearly uniform subvolumes, or Voronoi cells/regions. Given a set of N points $\{z_i\}$ (i=1,...,N) in an M-dimensional hypercube, the Voronoi region or Voronoi cell V_j (j=1,...,N) corresponding to z_j is defined to be all points in the hypercube that are closer to z_j than to any of the other z_i's. The set $\{V_i\}$ (i=1,...,N) is called a Voronoi tessellation or Voronoi diagram of the hypercube, the set $\{z_i\}$ (i=1,...,N) being the generating points or generators. A *centroidal* Voronoi tessellation (CVT) is a special Voronoi tessellation with the property that each generating point z_i is itself the mass centroid of the corresponding Voronoi region V_i.

Although CVTs are deterministic, they can be converged to with probabilistic sampling methods. In [10], new probabilistic CVT construction algorithms were introduced, implemented, and tested. These methods are generally much more efficient than previous deterministic and probabilistic methods for constructing CVTs.

The CVT concept and the algorithms in [10] for their construction can be generalized in many ways (see [6] for details). For example, instead of a hypercube, general regions in M-dimensional space can be treated. This feature has been exploited with great success (see [6]) for discretizing arbitrary 2-D and 3-D domain volumes for computational mechanics analysis with meshless analogues of finite element methods (e.g., [1]). Furthermore, points can be distributed non-uniformly according to a prescribed density function over the space. For instance, reference [18] shows several CVT point sets spaced according to a bi-Normal joint probability density function. Thus, CVT can be used for Monte-Carlo-like sampling in problems containing multiple random variables. In this regard, we surmise that correlation structure for correlated random variables can be introduced into CVT sampling with the rank correlation procedure [8] employed in [9] for SRS and LHS, and in [11] for Hammersley QMC.

Figure 1 compares three LHS and three corresponding CVT pointsets for 100 samples in a 2D unit hypercube. The three LHS pointsets were generated with [9] for different initial seeds (Seed1 = 123456789, Seed2 = 192837465, Seed3 = 987654321) and a Uniform joint probability density function over a unit-hypercube parameter space. The three corresponding CVT pointsets were generated ([3]) by using the LHS sets as initial conditions (point locations) to begin the CVT iterations. In all cases the CVT set is much more uniform (visually and quantitatively) than its associated LHS set. All three CVT sets are relatively similar visually and quantitatively, even though starting from three very different initial conditions given by the LHS sets.

The LHS sets exhibit significant clustering and non-uniformity of the points. The LHS sets do not appear to be significantly more uniform than three analogous SRS sets shown in [18], and which will be used here in later comparisons, but quantitatively they are significantly more uniform ([4]). CVT sets from the three different SRS initial sets are shown in [18]. The different LHS and SRS initial conditions do not have much of an impact on final CVT uniformity, so CVT appears to be robust in this regard.

Figures 2 and 3 show Halton and Hammersley pointsets and the corresponding CVT sets started from them. The Halton pointset is noticably and quantitatively more uniform than any of the LHS sets; the Hammersley set is even more uniform than the Halton set; and the CVT sets are even more uniform than the Hammersley set.

Hence, CVT places samples much more uniformly in the 2D hypercube than SRS and LHS, and even more uniformly than the low-discrepancy Halton and Hammersley QMC sequences. This is true regardless of the initial conditions (sample sets) that CVT starts from ([4]). In initial investigations [4] for 2-D, 7-D, and 20-D test cases, CVT has provided greater sampling uniformity than Halton, Hammersley, Sobol, SRS, LHS, DHS, and IHS according to a meaningful subset of nonflawed quantitative quality measures. Additionally, no degradation of sampling uniformity has been detected in higher dimensions (*i.e.*, for the 20-D case).

It is therefore natural to ask whether CVT can be applied for: A) statistical sampling over arbitrary-dimensional spaces of input random variables to calculate various statistics of output response behavior; B) function integration over arbitrarily shaped domains; and C) whether it can serve as a method for generating favorable point distributions for improved response-surface accuracy.

A preliminary positive indication regarding item C) for response surface generation is presented in [18]. There, CVT was shown on several 2-D test problems to provide superior point distributions for generating locally-conforming Moving Least Squares response surfaces. Point distributions by CVT, SRS, LHS, and a structured sampling method with deterministically uniform point placement ([17]) were tried in the study.

In this paper we take a first step toward examining the potential of CVT for improved statistical sampling and function integration (sections 2 and 3 respectively). We compare against results from SRS, LHS, Halton, and Hammersley sampling. Our discussion clarifies the connection between traditional statistical sampling and function integration. We use this connection to contemplate the prospects of CVT vs. other sampling methods.

2. Evaluation of CVT as a Statistical Sampling Method

2.1. 2-D Model Problem and Statistical Measures of Response for Performance Evaluation of Sampling Methods

Figure 4 shows an analytic multimodal function describing system response r as a function of two system inputs p1 and p2:

$$r(p1,p2) = \left[0.8\kappa + 0.35\sin\left(2.4\pi\frac{\kappa}{\sqrt{2}}\right)\right][1.5\sin(1.3\theta)] \quad \text{EQ 1}$$

on the domain $0 \le p1 \le 1$ and $0 \le p2 \le 1$,

where $\kappa = \sqrt{(p1)^2 + (p2)^2}$, $\theta = \text{atan}\left(\frac{p2}{p1}\right)$.

A statistical problem arises if p1 and p2 are random variables. In that case, any particular realization $p1_i$ and $p2_i$ of the stochastic variables yields a deterministic response r_i as given by the above functional relationship. An ensemble of responses accompanies the different realizations of p1 and p2 as they vary stochastically or randomly according to their individual propensities, or joint propensities if the two variables are correlated. In probability theory a joint probability density function defined over the input parameter space, JPDF(p1,p2), is used to model the relative likelihood of achieving an input combination p1,p2 corresponding to the point (p1,p2) in the p1-p2 coordinate plane. The JPDF function is defined for every point in the p1-p2 parameter plane and integrates over the plane into a value of unity.

The JPDF likelihood function for attaining various input combinations maps through the response function $r(p1,p2)$ into a corresponding likelihood function for response values. Operationally, the resulting response probability density function, PDF(r), can be approached closer and closer via Monte Carlo sampling as more and more parameter sets or realizations $(p1,p2)_i$ are randomly generated from the governing input JPDF and are propagated through the response function $r(p1,p2)$ into response realizations r_i. The response realizations are distributed in the response space (*i.e.*, along the response coordinate axis r) with a density that, as more and more samples are added, trends toward the exact PDF of response.

Very often, only certain statistical measures of the PDF of response are desired or can be reasonably estimated. Response mean, μ_r, and standard deviation, σ_r, can be estimated directly from the mean $\hat{\mu}_r$ and standard deviation $\hat{\sigma}_r$ of the population or set $\{r_i\}$ of realizations. We have the following definitions:

$$\hat{\mu}_r = \frac{1}{N} \sum_{i=1}^{N} r_i \qquad \text{EQ 2}$$

$$\hat{\sigma}_r = \left[\frac{1}{N-1} \sum_{i=1}^{N} (r_i - \hat{\mu}_r)^2 \right]^{\frac{1}{2}} \qquad \text{EQ 3}$$

where N is the number of realizations or "samples" of response.

Also often of interest is the probability of response exceeding (or not exceeding) some particular threshold value r_T. Exceedence probability is very simply estimated as the ratio of the number of calculated response values at or above the given threshold value, to the total number of samples, N. As the number of response realizations increases, the estimate (quotient) trends toward greater accuracy, *i.e.*, toward the actual exceedence probability. This is of course also true for the estimates $\hat{\mu}_r$ and $\hat{\sigma}_r$ of response mean and standard deviation.

2.2. Comparison of Response Statistics from Various Sampling Methods

Here we compare estimates of response mean, standard deviation, and exceedence probabilities as obtained from various sampling methods we have previously introduced: CVT, SRS, LHS, and Halton and Hammersley sequences.

We start with *e.g.* the 100-sample pointsets in Figure 1, which correspond to a uniform JPDF over the input parameter space of our model response function (Figure 4). We map these sets of samples through our response function EQ 1 to obtain corresponding response sets, and then calculate the aforementioned statistics of these populations.

We then compare the calculated statistics of each set to each other and to "reference values" obtained from using three million SRS samples at parameter sets generated by the sampling code [9]. The reference values are actually averages of three results, each obtained from one million samples generated from random initial seeds "X", "Y", and "Z" (different from seeds 1, 2, and 3 used to generate the 100-sample LHS sets in Figure 1).

Three "replicate" sets of one million samples each were used in preference to one set of three million samples so that empirical confidence intervals (CI) on the calculated averages could be compared against their classical CI to reaffirm or caveat them. (Recent research ([16], [19]) has shown that for SRS, empirical CI appear to be somewhat more accurate than classical CI.) Empirical CI are formed by assuming the calculated statistic (response mean, standard deviation, or exceedence probability) is a random realization from a Normal or nearly Normal distribution about the exact result. Hence a T-distribution with 3 - 1 =

2 degrees of freedom can be used to get confidence intervals about the small-sample average of the three replicates. Thus, for 95% empirical CI the following formula is used:

$$95\% \text{ confidence half-interval} = 4.303 \frac{\hat{\sigma}_{est}}{\sqrt{3}} \qquad \text{EQ 4}$$

where $\hat{\sigma}_{est}$ is the sample standard deviation (*cf.* EQ 3) of the three estimates.

2.2.1. Mean of Response. Tables 1 and 2 show the calculated means, along with nominal errors from the reference mean, $\hat{\mu}_{ref}$. The reference mean $\hat{\mu}_{ref}$ =0.581608 is the average of the three means from three SRS sets of 10^6 samples each. The standard deviation of the estimated means is $\hat{\sigma}_{est}$ =0.0002278. Thus, empirical 95% half-CI by EQ 4 are 0.000566. When the reference mean is calculated based on the entire population of N=3x10^6 samples, the value doesn't change from the averaged value based on three separate 10^6-sample sets, but classical CI can be computed. The classical 95% half-CI from standard statistical formulas is somewhat smaller, at 0.000388. Using the larger CI (empirical) we say that with at least 95% certainty the true response mean μ lies within the range $\hat{\mu}_{ref}$ ±0.000566 = (0.582174, 0.581042).

From Table 2, then, even using the larger CI to account for uncertainty in the reference value we can definitively say that all three SRS results have errors that are an order of magnitude greater than for the LHS and CVT sampling. This is true of individual errors and of the average error. In writing up these results we notice that the individual SRS errors are all of the same algebraic sign and of fairly large magnitude relative to the CVT and LHS results. A measure of sampling bias is given by the average of the errors, which is two orders of magnitude larger (=0.0407) for the SRS results than the average bias in the LHS (-0.0003) and CVT (=0.0006) results. A slight bias in the code used to generate the 100-sample SRS sets ([3]) may be suggested. (Note that the 10^6 SRS sets that the reference statistics were derived from were generated from a different code, [9].) Alternatively, we could simply be seeing a chance anomaly in the generated SRS data. We cannot tell which is the case from our limited number of trials.

In any case, the standard deviation of the estimates is an order of magnitude higher for SRS than for LHS or CVT. This is somewhat expected, as LHS and CVT (as well as DHS, IHS, and QMC sequence methods) are known to perform better than SRS in terms of variance reduction of the estimates.

Taking the uncertainty in the reference value into account, we can conclude to well over 95% certainty that CVT is more accurate than LHS in realization 3 and less accurate in realization 2. To almost 95% certainty we can

conclude that CVT does better than LHS in realization 1. These conclusions are reflected in a slightly lower average error magnitude for CVT than for LHS. Their average error (bias) and variance of the estimates is comparable. We note that the CVT results are obtained from the LHS results as initial conditions for the CVT iterations. In cases 1 and 3 the magnitude of the error declined after the CVT iterations, and in case 2 the magnitude increased. (These results hold above and beyond considerations of uncertainty in the reference result).

Table 2 shows that the Hammersley result is, to well over 95% certainty, an order of magnitude better than the Halton result. The Halton result is considerably improved by CVT iterations. Within the uncertainty in the reference result it cannot be determined whether CVT actually did improve the Hammersley result, but a significant improvement according to the nominal values in the table is indicated.

2.2.2. Standard Deviation of Response. Tables 3 and 4 show the estimates of the standard deviation of our response. Nominal errors from the reference value $\hat{\sigma}_{ref}$ =0.343208 are also shown. This value is the average of the three standard deviations calculated from the three 10^6 SRS sets. The standard deviation of these three estimates is $\hat{\sigma}_{est}$ =0.000327. Empirical 95% half-CI by EQ 4 are 0.000813. We can then say that to at least 95% certainty that the true response standard deviation σ lies within the range $\hat{\sigma}_{ref}$ ±0.000813 = (0.344021, 0.342395)

From Table 3, even accounting for uncertainty in the reference value, we can conclude to well over 95% certainty that the SRS and LHS errors are for all three trials an order of magnitude greater than the CVT errors. This is true of individual errors and for average error magnitude as well. The algebraic average of the signed errors shows that the average bias in the LHS and CVT results is similar, and smaller than the average bias in the SRS results by an order of magnitude. The standard deviation of the estimates is similar for LHS and SRS, and these are an order of magnitude larger than with CVT. Thus, in these three trial calculations of response standard deviation, CVT shows an order of magnitude improvement over SRS and LHS in both the error magnitude and standard deviation of the estimates.

Table 4 shows that, even given the uncertainty in the reference value, we can definitively say that the Halton and Hammersley errors are reduced by an order of magnitude with the CVT iterations. The initial and improved Hammersley results are an order of magnitude better than the initial and improved Halton results.

2.2.3. Response Exceedence Probability for r_T=0.2. The reference value for exceedence probability (EP)

corresponding to a response threshold level of r_T=0.2 is $\hat{P}_{ref}$ =0.870984. This value is the average of the three EPs calculated from the three 10^6 SRS sets. The standard deviation of these three estimates is $\hat{\sigma}_{est}$ =0.000257. Empirical 95% half-CI by EQ 4 are 0.000639. When the reference EP is calculated based on the entire population of N=3×10^6 samples, the value doesn't change from the averaged value based on three separate 10^6-sample sets, but classical CI can be computed. The classical 95% half-CI from standard statistical formulas is somewhat smaller, at 0.000379. Using the larger (empirical) 95% half-CI we can then say that at least to 95% certainty the true probability $P_{0.2}$ of response exceeding the threshold value r_T=0.2 lies within the range $\hat{P}_{ref}$ ±0.000639 = (0.871623, 0.870345).

Since the test sets were limited to 100 sample points, derived probabities can only be resolved in increments of 0.01. Thus, for the r_T=0.2 case, a derived result cannot be more accurate than 0.87 or 0.88 –both of which are equally valid estimates of the true probability, 0.870984±0.000639, which lies between the attainable values 0.87 and 0.88. In other words, any error in an estimate of 0.87 or 0.88 is due to resolution error from the limited number of samples, and not to a fault or inferiority of the sampling method's point placement scheme or resulting pattern. Hence, in judging the performance of our sampling methods, in Tables 5 and 6 we take results of 0.87 and 0.88 as exact results, and quantify errors therefrom. Accordingly, a sample-set result of *e.g.* 0.89 would entail an error of +0.01 here, and a result of *e.g.* 0.85 would entail an error of -0.02.

From Table 5 we can conclude to well over 95% certainty that the SRS and LHS errors are for all three trials significantly greater than the CVT errors. This is true of individual errors, and for average error magnitude as well (which for CVT is an order of magnitude less than that of SRS and LHS –LHS actually being significantly worse than SRS in this set of trials). The algebraic average of the signed errors shows that the average bias in the LHS and SRS results is similar, and about 50% smaller for CVT. The standard deviation of the estimates is an order of magnitude less for the CVT results than for the LHS and SRS, with SRS significantly better than LHS according to this metric. Thus, in these calculations of exceedence probability, CVT shows an order of magnitude improvement over SRS and LHS in both the error magnitude and standard deviation of the estimates.

Table 6 shows that the Halton error is significantly larger than the Hammersley error. CVT reduces the -0.02 Halton error to zero within our ability to distinguish error here, but does not improve the -0.01 error of the Hammersley result.

2.2.3. Response Exceedence Probability for r_T=0.5. The reference value for exceedence probability corresponding to a response threshold level of r_T=0.5 is $\hat{P}_{ref}$ =0.555050. This value is the average of the three EPs calculated from the three 10^6 SRS sets. The standard deviation of these three estimates is $\hat{\sigma}_{est}$ =0.000209. Empirical 95% half-CI by EQ 4 are 0.000519. The reference EP when calculated based on the entire population of N=3×10^6 samples yields classical 95% half-CI of 0.000562, very close to the empirical value. Using the larger (classical) 95% half-CI we can then say to at least 95% certainty that the true probability $P_{0.5}$ of response exceeding the threshold value r_T=0.5 lies within the range $\hat{P}_{ref}$ ±0.000562 = (0.555612, 0.554488).

Since the test sets were limited to 100 sample points, for the r_T=0.5 case a derived result cannot be more accurate than 0.55 or 0.56 –both of which are equally valid estimates of the true probability, 0.555050±0.000562, which lies between the attainable values 0.55 and 0.56. Hence, in judging the performance of our sampling methods, in Tables 7 and 8 we take results of 0.55 and 0.56 as exact results, and quantify errors therefrom. Accordingly, a sample-set result of *e.g.* 0.57 would entail an error of +0.01 here, and a result of *e.g.* 0.53 would entail an error of -0.02.

From Table 7 we see that the average bias of the SRS samples is an order of magnitude larger for SRS than for LHS and CVT. In the case, LHS bias is smaller than CVT bias by about 50%, but both are small. The standard deviation of the estimates is the same order of magnitude for SRS, LHS, and CVT, with CVT having the smallest standard deviation, then LHS, then SRS. Average error magnitude is also least for CVT, then for LHS, then for SRS.

Table 9 shows that for this problem the Halton error of 0.02 is actually better than the Hammersley error of -0.03. CVT reduces both these errors to zero within our ability to distinguish error here.

3. Relationship of Statistical Sampling to Function Integration

The volume integral of a continuous differentiable function f over some parameter volume V can be written

$$\int_V f\, dV = \int_{p\,\in\,V} f(\underset{\sim}{p})dV(\underset{\sim}{p}) \approx \sum_{i\,=\,1,\,N} f(\underset{\sim}{p}_i)(\Delta V)_i \qquad \text{EQ 5}$$

where $\underset{\sim}{p}$ are the coordinates of a differential volume element of integration within V, and the finite summation is the discrete analogue approximation of the integral.

Following the most common precepts of discrete numerical integration, the N subvolumes $(\Delta V)_i$ are non-overlapping regions which taken together occupy the entire in-

tegration domain, and a function evaluation point $\underset{\sim}{p}_i$ associated with each subvolume $(\Delta V)_i$ is located within the subvolume. Ideally the location is such that the value $f_i = f(\underset{\sim}{p}_i)$ is best representative over the subvolume so that

$$\int_{(\Delta V)_{ii}} f\, dV = \bar{f}_i(\Delta V)_i \approx f_i(\Delta V)_i, \qquad \text{EQ 6}$$

where $\bar{f}_i$ is the mean value of the function f over subvolume $(\Delta V)_i$. If equalities EQ 6 hold for all N subvolumes, then EQ 5 becomes an equality and the finite discrete summation equals the exact integral.

In practice, when knowing nothing about the function f before hand, we can generally best hope to approach equalities EQ 6 as follows. We attempt to subdivide the domain into "compact" subdomains of nearly or exactly equal volumes and place the function-evaluation or sampling points $\underset{\sim}{p}_i$ at the effective centers (usually center of "mass") of their compactly surrounding regions (subvolumes). In this way we improve the prospects that f_i is representative of the mean of f over the subvolume so that equality in EQ 6 is approached.

function value over the subvolume such that the equality EQ 6 is most closely approached. This is directly in line with the Centroidal Voronoi principle that governs subdivision of the space into equal Centroidal Voronoi regions and locates CVT points at the centers of those regions. Therefore, as explained next there is reason to expect that CVT sampling will perform relatively well for function integration.

If we subdivide the domain into N compactly surrounding Voronoi cells, in the ideal limit of equal cell volumes we have $(\Delta V)_i = V/N$ and the discrete summation in EQ 5 becomes

$$\sum_{i\,=\,1,\,N} f(\underset{\sim}{p}_i)(\Delta V)_i = \sum_{i\,=\,1,\,N} f(\underset{\sim}{p}_i)\cdot\frac{V}{N} = V\sum_{i\,=\,1,\,N}\frac{f_i}{N}. \quad \text{EQ 7}$$

Comparing EQ 5, EQ 7, and EQ 2 we have, in the terminology of EQ 2,

$$\int_V r\, dV = V\sum_{i\,=\,1,\,N}\frac{r_i}{N} = V\left(\frac{1}{N}\sum_{i\,=\,1}^N r_i\right) = V\cdot\hat{\mu}_r. \qquad \text{EQ 8}$$

We see that the volume integral of a function $r(\underset{\sim}{p})$ over some domain is equal to the volume V of the domain multiplied by the mean of the PDF(r) which results from mapping a uniform PDF over $V(\underset{\sim}{p})$ through the function $r(\underset{\sim}{p})$. That is, we can estimate the integral of a function over some domain by uniformly sampling the function over the domain and multiplying the mean $\hat{\mu}$ of the sample results by the volume V of the integration region. This is

the connection between statistical sampling and function integration.

For $(\Delta V)_i$ not equal to V/N, EQ 7 is not an equality and therefore EQ 8 is not an equality and hence the correspondence between statistical sampling and function integration suffers. However, the correspondence improves as the number of samples increases –provided the sample points remain well spread out over the domain so the associated subvolumes becomes more nearly equal in size approaching V/N. If compactness of the subvolumes is retained about centrally located sample points, then the decreasing subvolume size (as N increases) also means that the point value f_i becomes more and more representative of the mean of f over the subvolume so that equality in EQ 6 is approached.

Because of the properties of CVT, the rate of expected error decrease as sample size N increases would appear to be superior to other unstructured sampling methods (at least the ones examined here: SRS, LHS, Halton, and Hammersley). We plan to explore this in future papers, especially the connection between expected rate of error decrease and sampling uniformity (which governs the associated subvolumes' size equivalence, compactness, and sample-point centrality).

4. Concluding Remarks

Uniform CVT sampling would seem to be a natural best choice among other unstructured sampling methods for function integration –certainly better than the commonly used SRS ([14]). CVT's better performance than SRS, LHS, Halton, and nominally Hammersley on estimates of mean response μ tend to support this conjecture.

Certainly, in non-adaptive function integration (and by association, in uniform sampling to find the mean of a function over some parameter space), it stands to reason that uniformity of the sample points over the space is desirable. Thus, some regions of the space would not be over-sampled -where a high density of samples would tend to marginalize the information value of each individual sample- at the expense of undersampling or not sampling other regions of the space. Information marginalization increases with redundancy in the sampling, which accompanies point "clustering" or "clumping".

Examples of point clustering and clumping, with corresponding relative under-sampling in other regions, are shown in Figures 1-3 (and many more figures in [18]) to be most pronounced for SRS, next most for LHS, then Halton, then Hammersley, and finally CVT. Certainly, CVT yields the most uniform placement of samples. In fact, in all CVT pointsets there are no instances of discernable variation in sampling density over the parameter space -no clumping or clustering.

In our results, greater sampling uniformity generally correlated with better accuracy in the calculated response statistics. SRS usually but not always performed worst of all methods. LHS and Halton were generally better than SRS and not quite as good as Hammersley and CVT. Hammersley was often, but not always, as good as CVT.

The variability of CVT results (standard deviation of the estimates) was considerably smaller than for SRS and usually LHS as well, generally showing both more *precision* **and** *accuracy* than the more popular SRS and LHS methods.

However, much more work needs to be done before CVT can be concluded to be typically best for general applications of statistical sampling and function integration. In particular, we have not yet assessed CVT statistical performance under the much more common situation where non-uniform input random variables yield a non-uniform JPDF over the parameter space. This will be the next important test for CVT concerning its prospects as a general statistical sampling method.

Certainly, for function integration and point placement for response-surface generation (see [18]), CVT already appears very promising relative to other structured and unstructured sampling methods. Especially for irregular (non-hypercube) interpolation and integration domains, the regularity of CVT sampling over the domain is a large part of the reason why CVT is already recognized to hold great promise for the application of 2-D and 3-D meshless finite-element methods.

REFERENCES

[1] Atluri, S.N., and T. Zhu, "A New Meshless Local Petrov-Galerkin Approach in Computational Mechanics," *Computational Mechanics*, Vol. 22, 1998.

[2] Beachkofski, B.K., and R.V. Grandhi, "Improved Distributed Hypercube Sampling," paper AIAA-2002-1274, 43rd Structures, Structural Dynamics, and Materials Conference, April 22-25, 2002, Denver, CO.

[3] Burkardt, J.V., M.D. Gunzburger, J.S. Peterson, research sampling software

[4] Burkardt, J.V., M.D. Gunzburger, J.S. Peterson, 2003, preliminary research findings not yet published.

[5] Conover, W.M., "On a Better Method for Selecting Input Variables," 1975 unpublished Los Alamos National Laboratories manuscript, reproduced as Appendix A of "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems" by J.C. Helton and F.J. Davis, Sandia National Laboratories report SAND2001-0417, printed November 2002.

[6] Du, Q., V. Faber, M. Gunzburger, "Centroidal Voronoi tessellations: applications and algorithms," *SIAM Review.* Vol. 41, 1999, pp. 637-676.

[7] Hanson, K.M., "Quasi-Monte Carlo: halftoning in high dimensions?," to appear in Proceedings of SPIE 5016 (2003), *Computational Imaging*, C.A. Bouman and R.L. Stevenson, eds.

[8] Iman, R.L., and Conover, W.J., "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables," *Communications in Statistics*, B11(3) 1982a, pp. 311-334.

[9] Iman, R.L., and Shortencarier, M.J., "A FORTRAN77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples to Use with Computer Models," Sandia National Laboratories report SAND83-2365 (RG), printed March 1984.

[10] Ju, L., Q. Du., M. Gunzburger, "Probabilistic algorithms for centroidal Voronoi tessellations and their parallel implementation," *Parallel Computing*, Vol. 28, 2002, pp. 1477-1500.

[11] Kalagnanam, J.R., and Diwekar, U.M., "An Efficient Sampling Technique for Off-line Quality Control," *Technometrics*, August 1997, Vol.39, No.3.

[12] Manteufel, R.D., "Distributed Hypercube Sampling Algorithm," paper AIAA-2001-1673, 42nd Structures, Structural Dynamics, and Materials Conference, April 16-19, 2001, Seattle, WA.

[13] Okabe, A., B. Boots, K. Sugihara, S. Chui, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, 2nd Edition, Wiley, Chichester, 2000.

[14] Press, W.H., S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd Edition, Cambridge University Press, 1992.

[15] Robinson, D., and Atcitty, D., "Comparison of Quasi- and Pseudo- Monte Carlo Sampling for Reliability and Uncertainty Analysis," paper AIAA-1999-1589, 40th Structures, Structural Dynamics, and Materials Conference, April 12-15, St. Louis, MO.

[16] Romero, V.J., "Effect of Initial Seed and Number of Samples on Simple-Random and Latin-Hypercube Monte Carlo Probabilities –Confidence Interval Considerations," paper PMC2000-176, 8th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability, Univ. of Notre Dame, IN, July 24-26, 2000.

[17] Romero, V.J., and S.D. Bankston, "Progressive-Lattice-Sampling Methodology and Application to Benchmark Probability Quantification Problems," Sandia National Laboratories report SAND98-0567, printed March 1998.

[18] Romero, V.J., J.S. Burkardt, M.D. Gunzburger, J.S. Peterson, T. Krishnamurthy, "Initial Application and Evaluation of a Promising New Sampling Method for Response Surface Generation: Centroidal Voronoi Tessellation," paper AIAA-2003-2008, 44th Structures, Structural Dynamics, and Materials Conference (5th AIAA Non-Deterministic Approaches Forum), Apr. 7-10, 2003, Norfolk, VA.

[19] Romero, V.J., and A.L. Mellen, 2002, preliminary research findings not yet published.

[20] Ye, K.Q., W. Li, A. Sudjianto, "Algorithmic Construction of Optimal Symmetric Latin Hypercube Designs," *Journal of Statistical Planning and Inference*, Vol. 90 (2000), pp. 145-159.
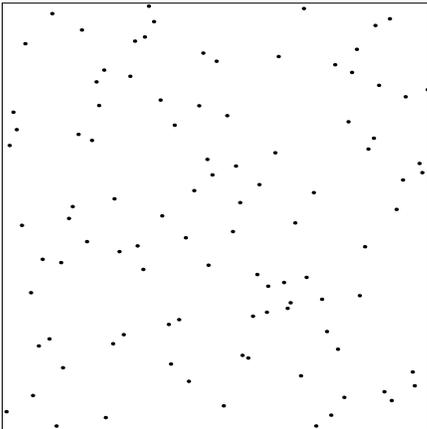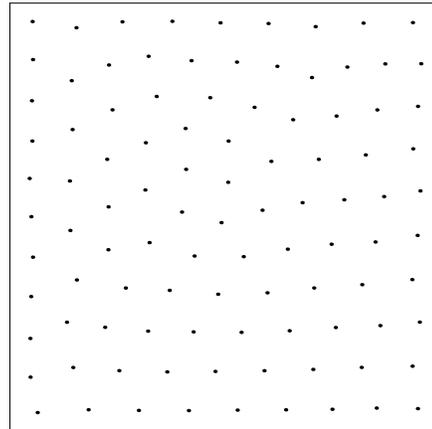
LHS1 pointset (from seed 1)
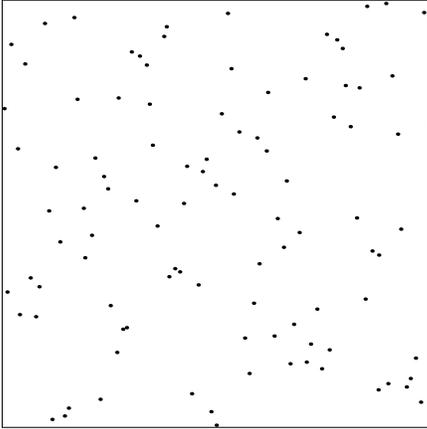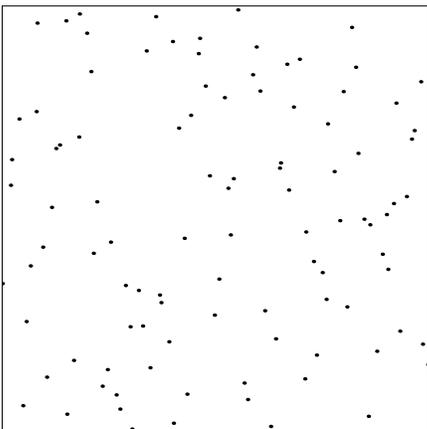
CVT-LHS1 pointset (from LHS1)
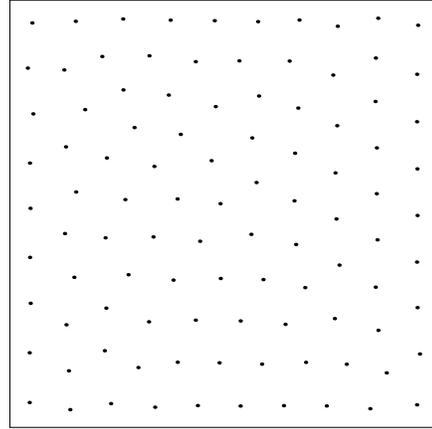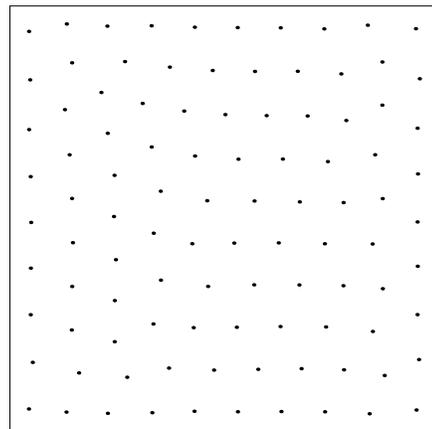
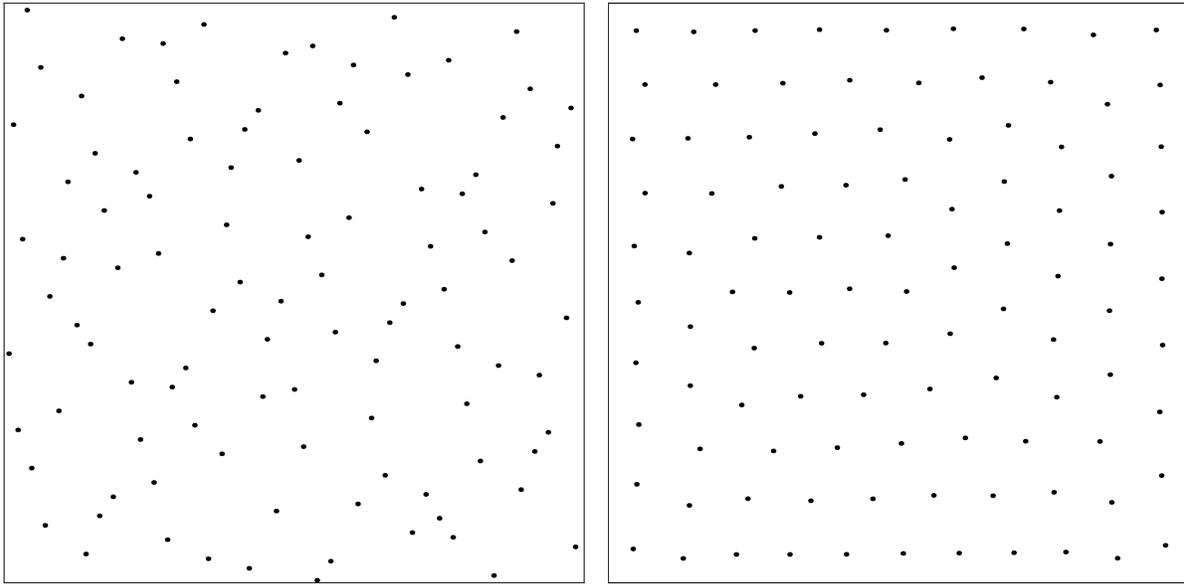LHS2 pointset (from seed 2)

CVT-LHS2 pointset (from LHS2)

LHS3 pointset (from seed 3)

CVT-LHS3 pointset (from LHS3)

Figure 1. 100-point sample sets on a 2-D unit hypercube for: A) Left Column– uniform JPDF LHS Monte Carlo with three different initial seeds; and B) Right Column– corresponding uniform JPDF CVT sets starting from LHS sets as initial conditions.

Figure 2. 100-point sample sets on 2-D unit hypercube for:
A) Left plot– Halton QMC sequence;
B) Right plot– corresponding CVT set starting from the Halton set as initial conditions.
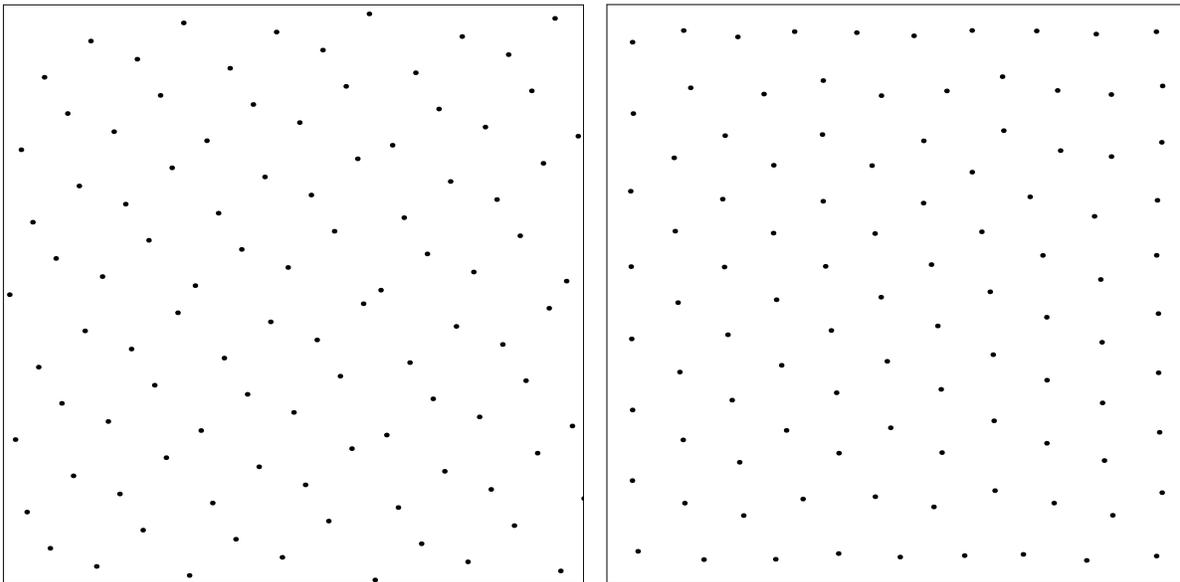
Figure 3. 100-point sample sets on 2-D unit hypercube for:
A) Left plot– Hammersley QMC sequence;
B) Right plot– corresponding CVT set starting from the Hammersley set as initial conditions.
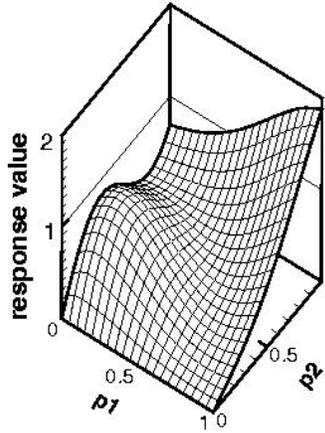
Figure 4. 2-D model function for system response as a function of input parameters p1 and p2.

Table 1. Calculated response means (100 samples, Uniform 2D JPDF)

		SRS		LHS		CVT (LHS)	
		$\hat{\mu}_r$	$\hat{\mu}_r$ error	$\hat{\mu}_r$	$\hat{\mu}_r$ error	$\hat{\mu}_r$	$\hat{\mu}_r$ error
REALIZATION	1	0.63153	+0.0499	0.58472	+0.0031	0.57948	-0.0021
	2	0.62511	+0.0435	0.58035	-0.0013	0.58675	0.0051
	3	0.61019	+0.0286	0.57891	-0.0027	0.58035	-0.0013
	average	0.622277	+0.0407	0.581327	-0.0003	0.582193	+0.00059
	std. dev.	0.010949	0.0109	0.003026	0.00302	0.003970	0.00400
	avg. error magnitude		0.0407		0.0024		0.0020

Table 2. Calculated response means (100 samples, Uniform 2D JPDF)

Halton		Hammersley		CVT	
$\hat{\mu}_r$	$\hat{\mu}_r$ error	$\hat{\mu}_r$	$\hat{\mu}_r$ error	$\hat{\mu}_r$	$\hat{\mu}_r$ error
0.56891	-0.0127			0.57455	-0.0071
		0.57533	-0.0063	0.58719	+0.0056

Table 3. Calculated response standard deviations (100 samples, Uniform 2D JPDF)

		SRS		LHS		CVT (LHS)	
		$\hat{\sigma}_r$	$\hat{\sigma}_r$ error	$\hat{\sigma}_r$	$\hat{\sigma}_r$ error	$\hat{\sigma}_r$	$\hat{\sigma}_r$ error
REALIZATION	1	0.39227	+0.04906	0.37505	+0.03184	0.34135	-0.00186
	2	0.37978	+0.03657	0.35397	+0.01076	0.33821	-0.00500
	3	0.32844	-0.01477	0.30729	-0.03592	0.33800	-0.00521
	average	0.366830	+0.023622	0.345437	+0.002228	0.339187	-0.003428
	std. dev.	0.033828	0.033828	0.034677	0.034677	0.001876	0.002220
	avg. error magnitude		0.033467		0.026174		0.003428

Table 4. Calculated response standard deviations (100 samples, Uniform 2D JPDF)

Halton		Hammersley		CVT	
$\hat{\sigma}_r$	$\hat{\sigma}_r$ error	$\hat{\sigma}_r$	$\hat{\sigma}_r$ error	$\hat{\sigma}_r$	$\hat{\sigma}_r$ error
0.32942	-0.01379			0.34565	-0.00600
		0.34565	+0.00244	0.34392	+0.00071

Table 5. Calculated response exceedence probabilities, threshold=0.2 (100 samples, Uniform 2D JPDF)

		SRS		LHS		CVT	
		$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error	$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error	$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error
REALIZATION	1	0.85	-0.02	0.84	-0.03	0.86	-0.01
	2	0.86	-0.01	0.86	-0.01	0.87	0.0
	3	0.89	+0.01	0.90	+0.02	0.87	0.0
	average	0.867	-0.0067	0.867	-0.0067	0.867	-0.0033
	std. dev.	0.021	0.0153	0.031	0.0252	0.006	0.0058
	avg. error magnitude		0.0133		0.02		0.0033

Table 6. Calculated response exceedence probabilities, threshold=0.2 (100 samples, Uniform 2D JPDF)

Halton		Hammersley		CVT	
$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error	$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error	$\hat{P}_{0.2}$	$\hat{P}_{0.2}$ error
0.85	-0.02			0.87	0.0
		0.86	-0.01	0.86	-0.01

Table 7. Calculated response exceedence probabilities, threshold=0.5 (100 samples, Uniform 2D JPDF)

		SRS		LHS		CVT	
		$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error	$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error	$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error
REALIZATION	1	0.56	0.	0.53	-0.02	0.55	0.
	2	0.57	+0.01	0.55	0.	0.58	0.02
	3	0.62	+0.06	0.57	+0.01	0.56	0.
	average	0.5833	+0.023	0.550	-0.003	0.563	+0.007
	std. dev.	0.0321	0.032	0.020	0.015	0.015	0.012
	avg. error magnitude		0.023		0.01		0.007

Table 8. Calculated response exceedence probabilities, threshold=0.5 (100 samples, Uniform 2D JPDF)

Halton		Hammersley		CVT	
$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error	$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error	$\hat{P}_{0.5}$	$\hat{P}_{0.5}$ error
0.58	+0.02			0.56	0.
		0.52	-0.03	0.56	0.