AIAA-2003-2008

Initial Application and Evaluation of a Promising New Sampling Method for Response Surface Generation: Centroidal Voronoi Tessellation*

V. Romero†, J. Burkardt◊, M. Gunzburger◊, J. Peterson◊, T. Krishnamurthy‡

ABSTRACT

A recently developed Centroidal Voronoi Tessellation (CVT) sampling method is investigated here to assess its suitability for use in response surface generation. CVT is an unstructured sampling method that can generate nearly uniform point spacing over arbitrarily shaped M-dimensional parameter spaces. For rectangular parameter spaces (hypercubes), CVT appears to extend to higher dimensions more effectively and inexpensively than "Distributed" and "Improved Distributed" Latin Hypercube Monte Carlo methods, and CVT does not appear to suffer from spurious correlation effects in higher dimensions and at high sampling densities as quasi-Monte-Carlo methods such as Halton and Sobol sequences typically do. CVT is described briefly in this paper and its impact on response surface accuracy in a 2-D test problem is compared to the accuracy yielded by Latin Hypercube Sampling (LHS) and a deterministic structured-uniform sampling method. To accommodate the different point patterns over the parameter space given by the different sampling methods, Moving Least Squares (MLS) for interpolation of arbitrarily located data points is used. It is found that CVT performs better than LHS in 11 of 12 test cases investigated here, and as often as not performs better than the structured sampling method with its deterministically uniform point placement over the 2-D parameter space.

†Sandia National Laboratories, Albuquerque, NM. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the U.S. Department of Energy's National Nuclear Security Administratn. under Contract DE-AC04- 94AL85000.

◊School of Computational Science and Information Technology, Florida State University, Tallahassee, FL. M. Gunzburger also supported in part by the Computer Science Research Institute, Sandia National Laboratories, under contract 18407.

‡NASA Langley Research Center, Hampton, VA.

1. Introduction

It is often beneficial in statistical sampling and response surface generation to sample "uniformly" over a 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. As elaborated in [7], quantitative aspects of uniformity involve: 1) the equality with which points are spaced relative to one another (are they all nominally the same distance from their nearest neighbors?); 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 3) isotropy or lack thereof in the coverage pattern. Each of these aspects of uniformity can be quantified by one or more mathematical measures considered in [7]. Here we will not directly invoke these measures, but we mention them to indicate a quantitative underpinning to the notion of uniformity. Rather, for the 2-D examples in this paper we invoke our experience that 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 mentioned above, and a good 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 ([4]). Recent efforts to modify LHS to get an even more uniform distribution of points over the paramter space have included Distributed Hypercube Sampling (DHS, [17]) and Improved Distributed Hypercube Sampling (IHS, [2]).

DHS improves on the uniformity of an initial LHS distribution by enforcing sampling uniformity over two dimensions at a time instead of just one as in LHS. Parameter pairings are shuffled, in turn for each parameter pair in the hypercube, to increase the overall uniformity over the hypercube as indicated by the particular figure of merit employed in [17]. Though the quality metric employed can be a misleading indicator of volumetric sampling uniformity, the method did substantially improve uniformity of 100 samples on a 16-D hypercube. Because the number of pairs to consider quickly increases with the dimensionality of the hypercube, the swapping and quality-evaluation operations can become computationally expensive in high dimensions. However, it was also found that the potential for improvement is constrained more and more as each next parameter pair is considered, so most of the improvement occurs with manipulation of the early pairs anyway.

An improvement to DHS was made in [2]. IHS initially conceived of the hypercube volume as a whole in trying to manipulate point locations to achieve a uniform point distribution over the parameter space. For a given number N of sample points and dimensionality M of the hypercube, each point's "equal share" of the hypervolume could be approximated as a hypersphere, and from this an approximate optimal point-to-point distance (hypersphere diameter) could be calculated. A global repositioning scheme would then perturb initial point locations to simultaneously drive all point-to-point distances up or down toward the target optimal value. However, it was found that the operation count for the procedure grew very quickly with N and M, and that "the search space becomes unsearchably large even for small bin numbers [N] and modest dimensionality [M]." A fallback to application of the scheme to hypercube subspaces of affordably reduced dimension was implemented with good success in achieving better uniformity of 100 sample points in a 7-D hypercube than either SRS or LHS (though again, the measure of uniformity used was potentially misleading). Extrapolative comparisons against DHS results suggested that IHS was indeed an improvement on DHS.

A number of potential whole-hypercube-volume approaches are reviewed and some new ones are presented in [8], along with some quantitative metrics related to visual/sensory point uniformity in 2-D. Many of these 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 are seemingly more promising for high dimensions, but have not yet been investigated enough.

Recently, an efficient whole-domain approach, "Centroidal Voronoi Tessellation" (CVT), has been developed ([12]) for implementing the principles of Centroidal Voronoi diagrams ([5],[18]) which subdivide arbitrarily shaped domains in arbitary-dimensional space into arbitrary numbers of nearly uniform subvolumes. The CVT methodology and several applications are discussed in [5]. A recent application at Sandia National Laboratories involved uniform unstructured node placement throughout irregular 2-D and 3-D geometries ([3]) for "meshless" or "particle" type computational mechanics methods (e.g., [1]). However, the CVT methodology is easily and inexpensively extensible to higher dimensions, especially for hypercube volumes. Since this is the case, it is natural to ask whether CVT can be applied for: A) statistical sampling over arbitrary-dimensional spaces of random variables; and B) whether it can serve as a method for generating favorable point distributions for more accurate response surfaces than other unstructured sampling schemes like LHS can produce.

Improved response surface generation can significantly impact many areas of high-dimensional modeling approximation found in optimization and uncertainty quantification. In this paper we take a first step toward assessing the potential of CVT for improved response-surface generation (item B above), and also mention some preliminary findings regarding statistical sampling performance (item A above) which are documented elsewhere ([23]).

Finally in this introduction we mention the so-called "Quasi- Monte Carlo" (QMC) quasi- or sub- random low-discrepency sequence methods (see *e.g.* [19]) that 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 faster times faster than LHS over a large range of test problems. For resolving response probabilities, Halton and

Hammersley were found in [20] 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 Halton sequences in 16-D (ref. [17]) and 40-D (ref. [20]). Sometimes a modification can be found to suppress or delay the onset of spurious correlation, as a fix from the literature implemented in [20] shows for Halton sequences.

In comparison, initial investigation in [7] indicates that for 2-D, 7-D, and 20-D test cases examined so far, CVT provides 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 is detected at higher dimensions (*i.e.*, for the 20-D case).

Lastly, it is believed that correlation structure for correlated variables can be introduced into CVT sampling by the rank correlation procedure [10] employed in [9] for SRS and LHS and in [11] for Hammersley QMC.

2. A CVT Sampling Method for Generating Nearly Uniform Point Placement in a Hypercube

2.1 Introduction to CVT and an Implementation using Probabilistic Construction Methods

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.

For each Voronoi region V_i, we can define its center of mass or mass centroid $\tilde{z}_i$ by

$$\tilde{z}_i = \frac{\int_{V_i} x\,dx}{\int_{V_i} dx} \qquad \text{for } i=1,...,N.$$

A *centroidal Voronoi tessellation* (CVT) is a special Voronoi tessellation with the property that

$$z_i = \tilde{z}_i \qquad \text{for } i = 1,...,N, \qquad (1)$$

i.e., each generating point z_i is itself the mass centroid of the corresponding Voronoi region V_i. General Voronoi tessellations do not satisfy the CVT property (1).

The special nature of *centroidal* Voronoi tesselations (CVTs) require their construction, *i.e.*, given a hypercube in M dimensions and a positive integer N, determine N points z_i (i=1,...,N), such that (1) is satisfied. In the past, CVT's were constructed either by deterministic methods typified by Lloyd's iteration [15], or probabilistic methods typified by MacQueen's random algorithm [16]. It is important to note with regard to the latter that although CVTs are deterministic they can be converged to through probabilistic sampling methods like MacQueen's method.

Although attractive due to its simplicity, the convergence of MacQueen's method is extremely slow. Therefore, the study of various accelerations of MacQueen's method, as well as its generalization and parallelization, are active areas of research. In [12], new superior probabilistic CVT construction algorithms were introduced, implemented, and tested. These algorithms can be viewed as generalizations of both the simple MacQueen random and Lloyd deterministic algorithms. The key to the efficiency of the new methods is that many points are sampled before each averaging step is performed. It is very important to point out that the probabilistic algorithms do not require, at any stage, the explicit construction of Voronoi diagrams nor the determination of the centers of mass of the Voronoi cells. Moreover, the new algorithms are highly amenable to fully scalable parallelization, as was demonstrated in [12]. The methods presented in [12] constitute the most efficient methodology for constructing CVT's and, as a result, in many applications they render the CVT concept superior or at least competitive with other approaches.

The CVT concept and the algorithms just mentioned for their construction can be generalized in many ways (see [5] for details). For example, instead of the hypercube, general regions in M-dimensional space can be treated, and points can be distributed non-uniformly according to a prescribed density function as will be shown later.

2.2 Examples and Discussion of SRS, LHS, CVT, Halton, & Hammersley Point Placements in 2-D

Figure 1 compares three SRS and three corresponding CVT pointsets for 100 samples in 2D unit hypercube. The three SRS pointsets were generated with the Monte Carlo sampling software [9] for different initial seeds (Seed1 = 123456789, Seed2 = 192837465, Seed3 = 987654321) and uncorrelated Uniform marginal density functions for the two input dimensions P1 and P2. The three corresponding CVT pointsets were generated by using the three SRS sets as initial conditions (point locations) to begin the CVT iterations. The three CVT sets are all relatively similar –visually and quantitatively (see

[7] for quantitative comparisons)– even though starting from three very different initial conditions given by the noticably different SRS pointsets. In all cases the CVT set is much more uniform (visually and quantitatively) than its associated SRS set.

Figure 2 shows an analogous comparison between three LHS sets and three corresponding CVT sets. The LRS pointsets were generated with the software [9] for the same three initial seeds (Seed1, Seed2, Seed3). The corresponding CVT pointsets were generated from LHS initial conditions. The three CVT sets are much more uniform (visually and quantitatively) than the associated LHS sets. The LHS sets do not appear to be significantly more uniform that the SRS sets in Figure 1, but quantitatively they are somewhat more uniform. The CVT sets from the different LHS and the SRS initial conditions are of relatively similar uniformity.

Figures 3 and 4 respectively show CVT sets corresponding to Halton and Hammersley initial pointsets. The Halton pointset is noticably and quantitatively more uniform than any of the LHS sets, the Hammersley set is noticably and quantitatively more uniform than the Halton set, and the CVT sets are noticably and quantitatively more uniform than the Hammersley set.

Hence, the figures show that 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. The performance of these various samplesets for integration and statistical sampling of function response is discussed in [23]. Preliminary indications are that CVT performs best overall.

Figure 5 compares three SRS and three corresponding CVT sets for 100 samples Normally distributed over a 2D unit hypercube. The SRS pointsets were generated with the software [9] from the aforementioned initial seeds and a prescribed 2-D joint-Normal probability density function of uncorrelated Normal marginal distributions in P1 and P2 (means=0.5, standard deviations =1/6). Figure 6 shows analogous LHS pointsets with their coresponding CVT sets. It is immediately clear in both figures that CVT gives the pointsets with the least clustering and most regular *weighted* coverage of the space according to the prescribed JPDF, independent of the starting initial conditions (pointset). LHS appears to give the next best regularity in weighted coverage, with SRS looking the least regular. The performance of the samplesets in these figures for resolving several different statistical metrics of function response (mean, variance, response probability) is discussed in [23]. Preliminary indications are that CVT does not perform as well as SRS or LHS.

3. Examination of CVT as a Point-Placement Technique for Response Surface Generation

The properties of CVT would seem to make it a good candidate for uniformly sampling a function over some (hypercubic or irregular) parameter space for the purpose of building a response surface approximation to that function over the space. Here we put this hypothesis to an initial test.

3.1 Importance of Seeking a Uniform Sampling Distribution over a Response-Surface Parameter Space, and the Suitability of CVT for this

In general, for reasonably arbitrary functions for which no information is known *a priori* regarding the functional variation over the parameter space (therefore, adaptive point-spacing over the domain cannot be intelligently applied), it stands to reason that uniformity of the (at least initial set of) sampling points over the space is desirable. Thus, some regions of the space will not go unsampled or under-sampled relative to other more densely sampled regions where the value of each sample would typically be marginalized somewhat. Especially for smooth continuous functions over the space, such marginalization increases with redundancy in the sampling, which accompanies any point "clustering" or "clumping" introduced by the sampling method.

Examples of point clustering and clumping, with corresponding relative under-sampling in other regions, is shown in Figures 1-4 to be most pronounced for SRS, then 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.

3.2 Interpolation for Unstructured Sampling of a Parameter Space, and the Moving Least Squares Method

Given a set of sampling points over a parameter space, the quality of the response surface approximation (RSA) also depends on the particular method used to fit and interpolate the data.

Fitting and interpolating data from unstructured sampling in even a simple hypercubic arbitrary-dimensional parameter space is a complex topic ([6]). Three popular scalable fitting/interpolation methods that the authors have some initial experience with are Kriging

([22] and [14]); Moving Least Squares ([14]); and Radial Basis Function methods ([13]). Though our experience is very limited, of these, MLS appears to present a good balance of response surface accuracy, smoothness, robustness, and ease of use. Therefore, we use MLS in this paper to generate response surfaces from our 2-D data sets. The particular implementation of MLS we use is described in [14]. Here we use quadratic polynomial interpolant basis functions, which requires at least (M+1)(M+2)/2 (*i.e*. 6 for 2-D) sample points within a given evaluation point's local radius of influence.

3.3 Response-Surface Sampling Efficiency in 2-D

2-D Test Function and Metric of Fitting Quality

Figure 7 shows an analytic multimodal surface defined by the equation

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

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

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

Note that this function is not separable in the two input parameters $p1$ and $p2$ and therefore does not give unfair advantage to rectangular structured sampling schemes which might align samples along the parameter directions.

This is the "target" or "exact" function to be approximated by MLS response surfaces based on various pointsets as described in the following. A simple measure of quality of approximation over the parameter space can be calculated as:

$$\text{approx. spatially avgd. }|\text{error}| = \frac{\sum_{i=1}^{441} |\text{exact}_i - \text{predicted}_i|}{441}$$

where exact and predicted values in the summation come from respective evaluation of the exact function and the particular response surface approximation at 441 equally spaced points on a 21x21 square grid overlaid on the domain. This measure is an expedient approximation to the global average integrated absolute error over the domain, which would require a much more involved calculation. The value of the current metric can depend strongly on the grid used if it underresolves the functions. However, as Figure 7 shows, a 21x21 plotting grid is sufficiently dense to achieve adequate representation of the target function.

Sampling Placement with CVT, LHS, and PLS

Figure 8 shows pointsets of 9, 13, 25, and 41 samples deterministically spaced in a 2-D hypercube according to a particular notion of maximum or ideal uniformity. These will serve as a benchmark against which the sampling performance of CVT and LHS point sets will be compared for accurate response surface approximation (RSA) generation.

The point placements in Figure 8 come from a structured-sampling "incremental experimental design" scheme called Progressive Lattice Sampling (PLS, [21]). PLS prescribes fixed numbers and locations of sample points for various levels of coverage of the parameter space. The number of points at each PLS level in the design is fixed by the particular level L and the dimension M of the space. The required number of sample points quickly grows with increasing PLS Level and/or space dimension. Thus, it is desirable to seek an unstructured sampling method such as CVT that has full freedom in the number of sampling points that can be applied, but which will still effectively maximize uniformity of coverage over the parameter space. In 2-D we can benchmark the effective sampling uniformity of CVT against the "asymptotic" structured uniformity of the 2-D PLS sets, to see how well CVT does against a standard of "ideal" uniformity. Comparisons will also be made to LHS, which is perhaps the most popular general unstructured sampling method for response surface generation in optimization and uncertainty analysis.

Figures 9 - 12 show three LHS and three corresponding CVT pointsets of 9, 13, 25, and 41 samples. It is immediately apparent that the point arrangements of the CVT sets are in many cases quite similar to the structured uniform arrangements of the PLS sets for the same number of samples. A systematic difference in point placement does exist, however, due to the fact that the PLS samples extend to and lie on the boundaries of the 2-D domain, whereas the CVT samples exist within a concentric square subdomain of the full 2-D hypercube.

Regarding this difference, it was considered in [21] that it may indeed be preferable in the general case to place the PLS samples in a smaller concentric subdomain of the full hypercube[1] in order to associate the sample points with more centrally surrounding hypercube subvolumes of more nearly equal size. This principle arises from non-adaptive numerical quadrature where, in

[1] This was not done, however, because the optimal size of a concentric sub-hypercube was not obvious, and would presumably change for each new PLS level as the total number of samples changes.

the general case when integrating an unknown function, the expectation is that optimal quadrature accuracy will coincide with a sampling of the parameter space that allows the domain to be subdivided so that the quadrature points lie at the centers of equal-volume compactly surrounding integration regions. Thus, the principle that governs optimal point coverage in quadrature is a direct complement to the Centroidal Voronoi diagraming principle that governs subdivision of the space into equal Centroidal Voronoi regions or subvolumes and locates CVT points at the centers of those regions. Therefore, there is precedent to expect that the CVT sampling in Figures 9 - 12 may outperform the corresponding PLS samplings in Figure 8. In fact, it is conjectured (though this has not yet been checked) that, by most meaningful quantitative measures of uniformity over the hypercube, the CVT sets will actually turn out to be more uniform than the initial ideals of uniformity in Figure 8.

MLS Response Surface Errors using PLS, CVT, LHS

Tables 1 - 4 list the RSA fitting-error indicators for MLS interpolation of the various pointsets. (The MLS radius of influence was the same for all 9-sample sets, all 13-sample sets, etc., but did change as the sample size changed. The radii of influence were optimized with a 1-D parameter study to get the lowest RSA errors with the PLS pointsets.)

Table 1. Response Surface Spatially Averaged Absolute Error (fit to 9 data points)

		LHS	CVT	PLS
		avg. |error|	avg. |error|	avg. |error|
REALIZATION	1	0.1010	0.1152	0.1544
	2	0.2019	0.1155	“
	3	0.1204	0.1158	“
	mean	0.14110	0.1150	0.1544
	std. dev.	$3x10^{-3}$	$5x10^{-7}$	0

Table 2. Response Surface Spatially Averaged Absolute Error (fit to 13 data points)

		LHS	CVT	PLS
		avg. |error|	avg. |error|	avg. |error|
REALIZATION	1	0.1133	0.0646	0.0587
	2	0.0856	0.0636	“
	3	0.0928	0.0629	“
	mean	0.09722	0.06371	0.05867
	std. dev.	$2x10^{-4}$	$7x10^{-7}$	0

Table 3. Response Surface Spatially Averaged Absolute Error (fit to 25 data points)

		LHS	CVT	PLS
		avg. |error|	avg. |error|	avg. |error|
REALIZATION	1	0.0439	0.0246	0.0219
	2	0.0404	0.0264	“
	3	0.0344	0.0258	“
	mean	0.03957	0.02560	0.02192
	std. dev.	$2x10^{-5}$	$8x10^{-7}$	0

Table 4. Response Surface Spatially Averaged Absolute Error (fit to 41 data points)

		LHS	CVT	PLS
		avg. |error|	avg. |error|	avg. |error|
REALIZATION	1	0.0192	0.0134	0.0191
	2	0.0215	0.0130	“
	3	0.0254	0.0133	“
	mean	0.02203	0.01324	0.01910
	std. dev.	$1x10^{-5}$	$4x10^{-8}$	0

3.4 Discussion of Response Surface Results

At every population level (9, 13, 25, 41 samples), CVT is found to perform significantly better than LHS on average over the three realizations (seeds), and in fact performs significantly better than LHS in 11 of the 12 individual cases. Furthermore, the variability of CVT results (standard deviation of the estimates) is considerably smaller than for LHS. Thus, here CVT shows both more *precision* **and** *accuracy* than LHS.

Compared to PLS, at 9 and 41 samples CVT yields significantly better results on all three trials than the deterministically placed uniform PLS set does. This may be because, though more uniform than CVT, PLS goes to the boundaries of the parameter space (see footnote 1). The strong showing of CVT versus PLS is especially encouraging given that the MLS influence radii used were those that minimize PLS fitting error.

4. Concluding Remarks

CVT appears to be a promising unstructured sampling method for response surface generation and perhaps other sampling purposes as well. Preliminary investigation with the 2-D test function here revealed that the uniform hypercube sampling properties of CVT yielded better MLS response surface results than the popular LHS method, and results about as good as the idealized deterministically uniform PLS structured sampling method.

However, much more work needs to be done before CVT can be concluded to be typically better than LHS for general RSA applications in high-dimensional modeling approximation settings (such as in uncertainty analysis and optimization).

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, Denver, CO.

[3] Burkardt, J.S., M.D. Gunzburger, J.S. Peterson, R. Brannon, "User Manual and Supporting Information for Library of Codes for Centroidal Voronoi Point Placement and Associated Zeroth, First, and Second Moment Determination," Sandia National Laboratories report SAND2002-0099, printed Feb. 2002.

[4] 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 Novermber 2002.

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

[6] Franke, R., and G.M. Nielson, "Scattered Data Interpolation and Applications: A Tutorial and Survey," in *Geometric Modeling: Methods and Applications*, 1990, Springer, Berlin

[7] Burkardt, J.V., M.D. Gunzburger, J.S. Peterson, V.J. Romero, "Comparison of Sampling Methods for Achieving a Uniform Point Distribution in a Hypercube," to be submitted for the 4th International Symposium on Uncertainty Modeling and Simulation (ISUMA'03), Sept. 21-24 2003, College Park, MD.

[8] 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.

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

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

[13] Krishnamurthy, T., "Response Surface Approximation with Augmented and Compact Support Radial Basis Functions," paper AIAA-2003-7561, 44th

Structures, Structural Dynamics, and Materials Conference, April 7-10, Norfolk, VA.

[14] Krishnamurthy, T., and V.J. Romero, “Construction of Response Surfaces with Higher Order Continuity and its Application to Reliability Engineering,” paper AIAA-2002-1466, 43rd Structures, Structural Dynamics, and Materials Conference, April 22-25, Denver, CO.

[15] Lloyd, S., “Least squares quantization in PCM,” *IEEE Transactions in Information Theory*, Vol. 28, 1982, pp. 129-137.

[16] MacQueen, J., “Some methods for classification and analysis of multivariate observations,” Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. I, Ed. by L. Le Cam and J. Neyman, University of California, Berkeley, 1967, pp. 281-297.

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

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

[19] 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.

[20] 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.

[21] 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.

[22] Romero, V.J., L.P. Swiler, A.A. Giunta, “Application of Finite Element, Global Polynomial, and Kriging Response Surfaces in Progressive Lattice Sampling Designs,” paper PMC2000-175, 8th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability, July 24-26, 2000, University of Notre Dame, IN.

[23] Romero, V.J., J.S. Burkardt, M.D. Gunzburger, J.S. Peterson, “Initial Evaluation of Centroidal Voronoi Tessellation Sampling Method for Integration and Statistical Sampling” to be submitted for the 4th International Symposium on Uncertainty Modeling and Simulation (ISUMA’03), Sept. 21-24 2003, College Park, MD.

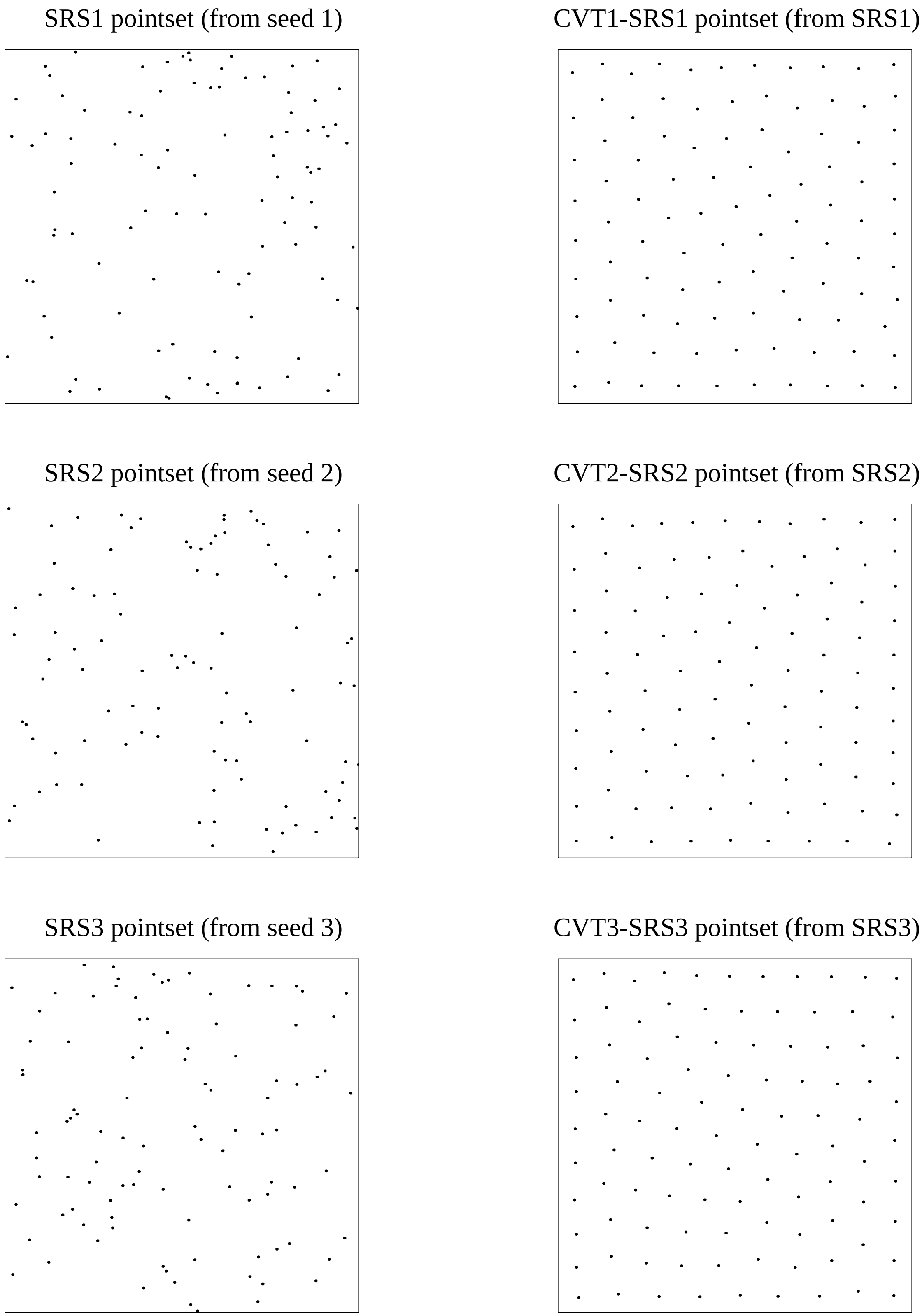

Figure 1: 100-point sample sets on 2-D unit hypercube for: A) Left Column– uniform JPDF SRS Monte Carlo with three different initial seeds; and B) Right Column– corresponding uniform JPDF CVT sets starting from SRS sets as initial conditions.

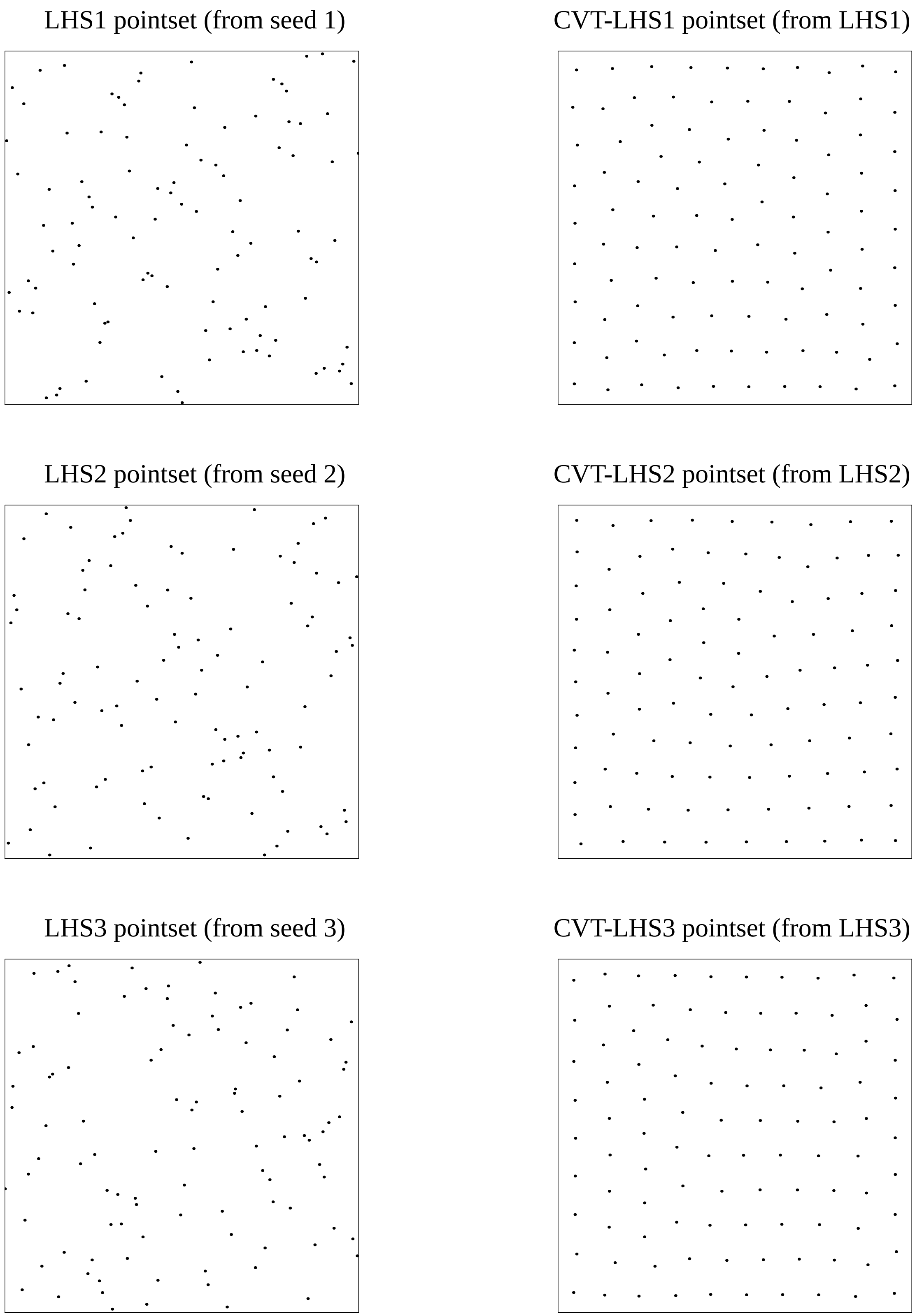

Figure 2: 100-point sample sets on 2-D unit hypercube for: A) Left Column– uniform JPDF LHS Monte Carlo with three different initial seeds; and B) Right Column– corresponding uniform JPDF CVT sets starting from LHS sets as initial conditions.

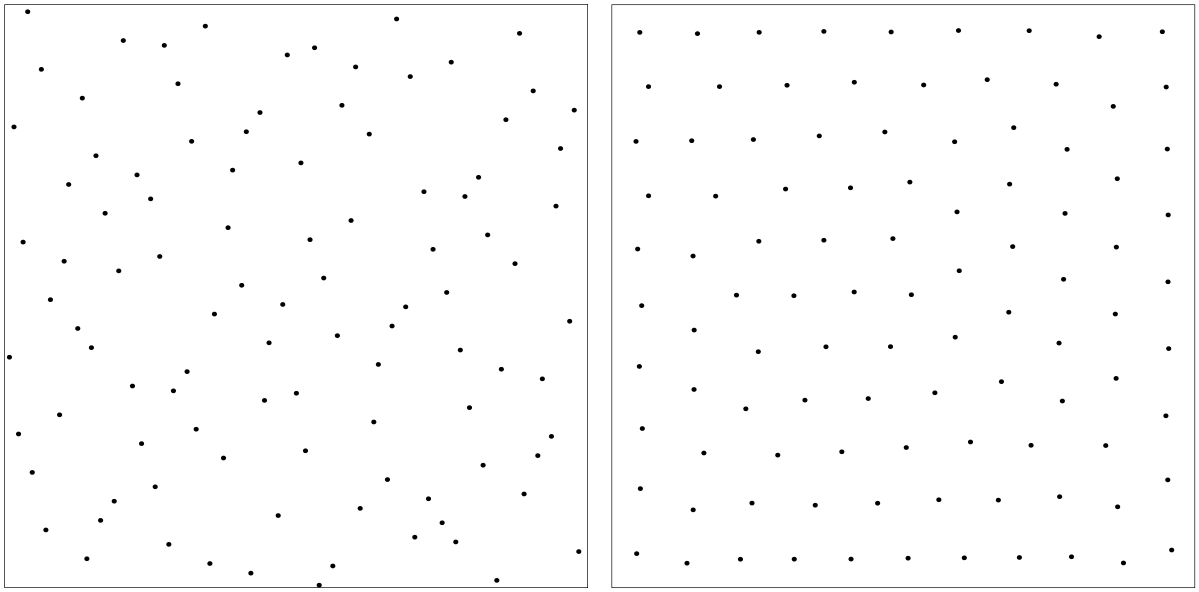

Figure 3: 100-point sample sets on 2-D unit hypercube for:
A) Left plot– Halton QMC sequence;
B) Right plot– corresponding CVT set starting from the Halton set as initial conditions.

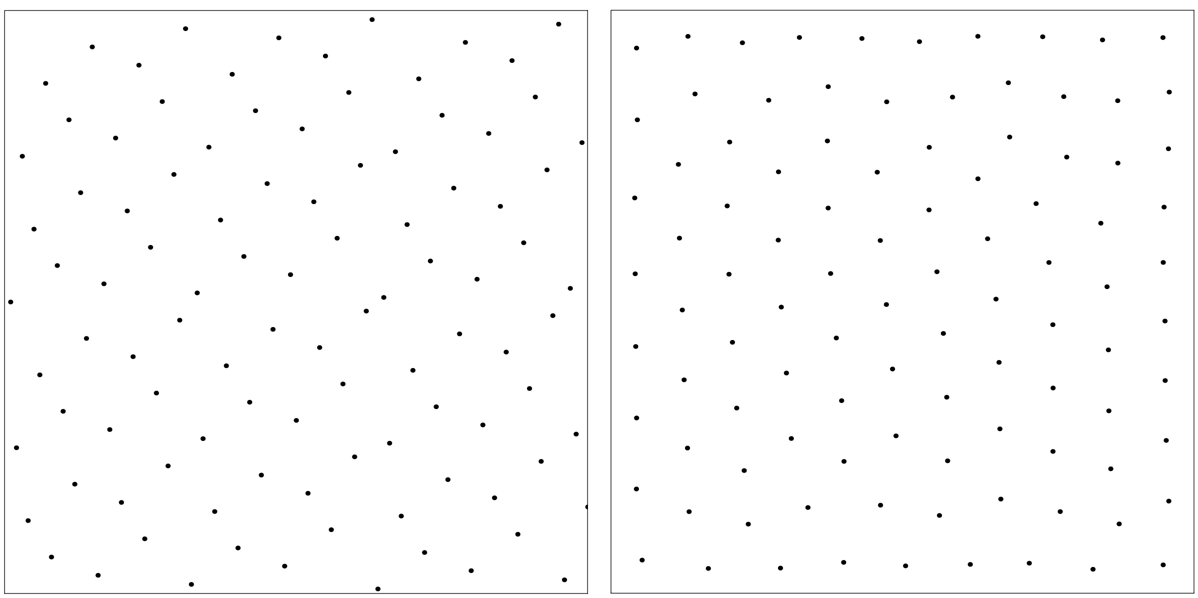

Figure 4: 100-point sample sets on 2-D unit hypercube for:
A) Left plot– Hammersley QMC sequence;
B) Right plot– corresponding CVT set starting from the Hammersley set as initial conditions.

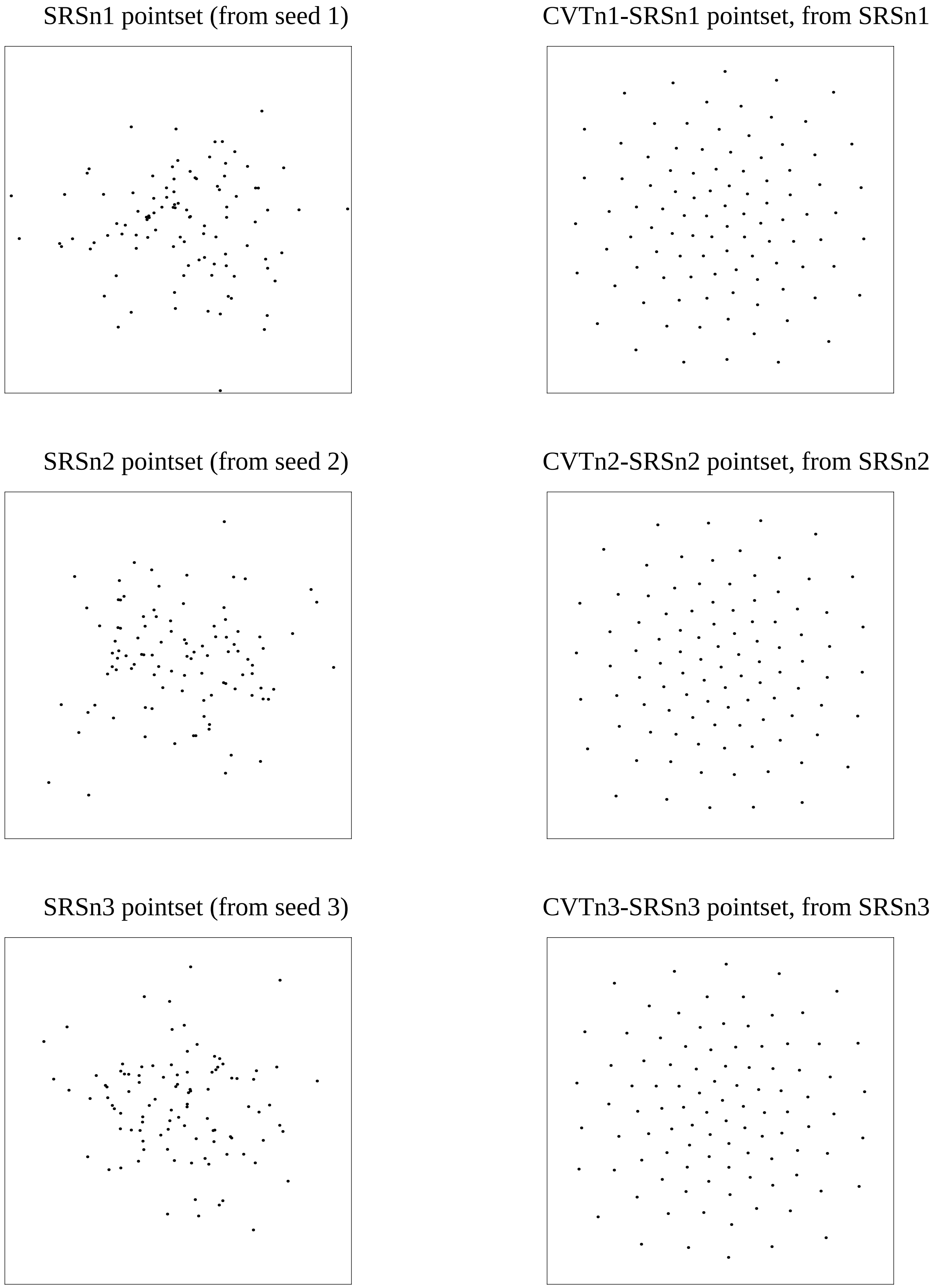

Figure 5: 100-point sample sets on 2-D unit ypercube for: A) Left Column– Normal JPDF SRS Monte Carlo with three different initial seeds; and B) Right Column– corresponding Normal JPDF CVT sets starting from SRS sets as initial conditions.

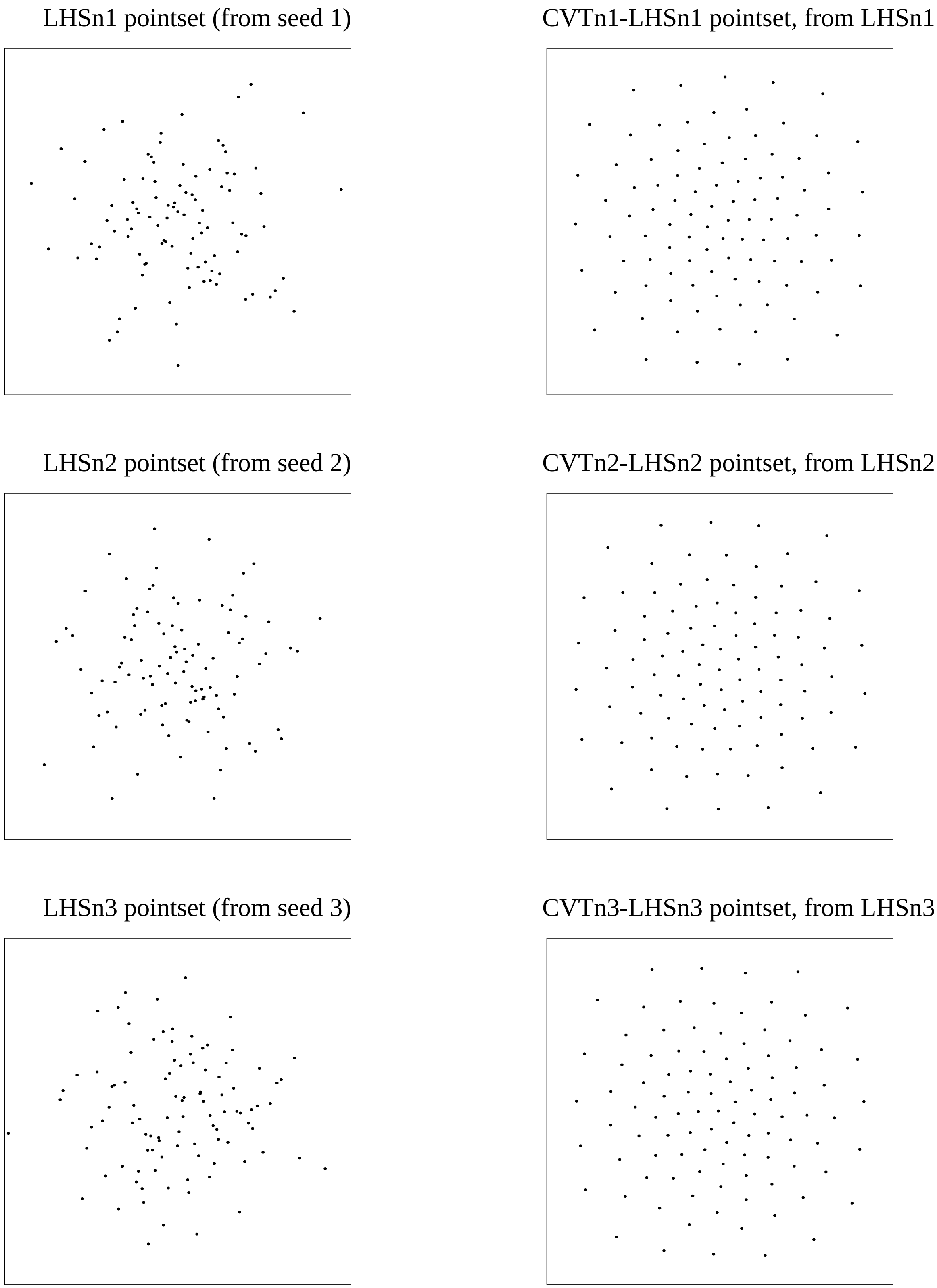

Figure 6: 100-point sample sets on 2-D unit ypercube for: A) Left Column– Normal JPDF LHS Monte Carlo with three different initial seeds; and B) Right Column– corresponding Normal JPDF CVT sets starting from LHS sets as initial conditions.

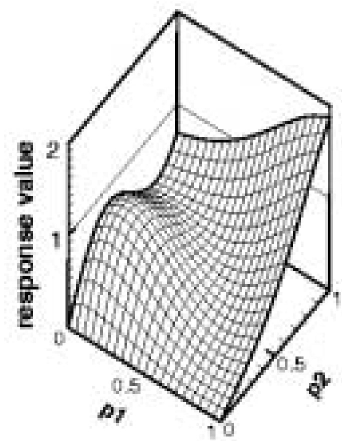

Figure 7: 2-D model function ("exact function") for response surface fitting tests.

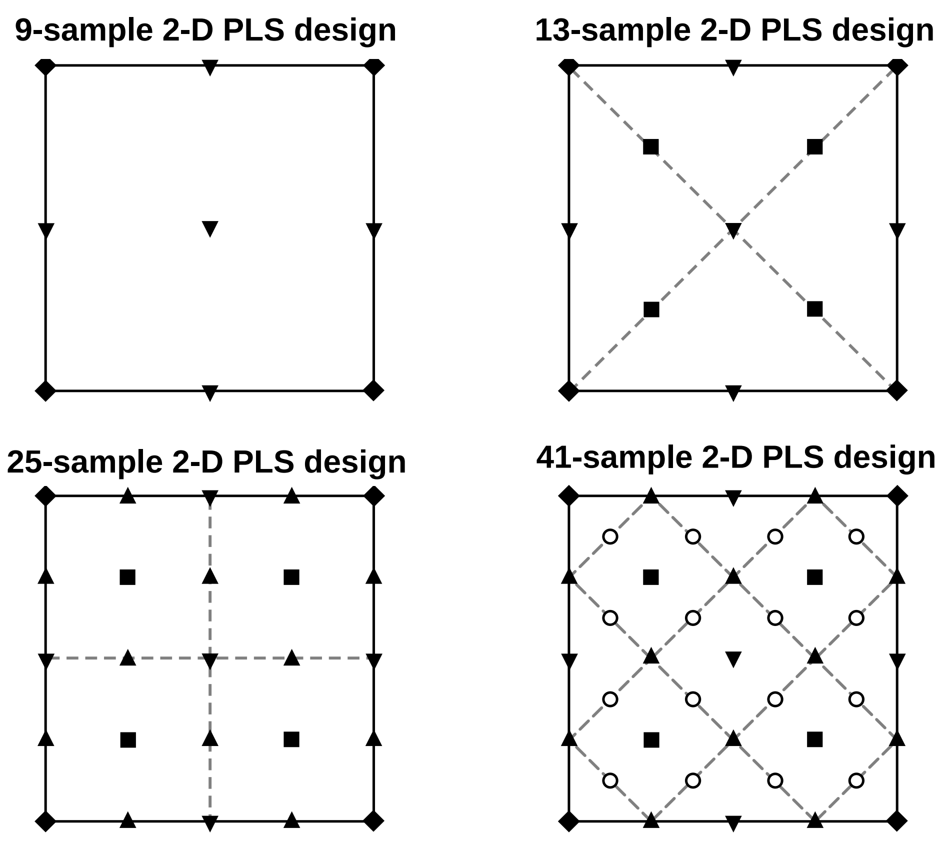

Figure 8: 2-D Progressive Lattice Sampling (PLS) Incremental Experimental Design.

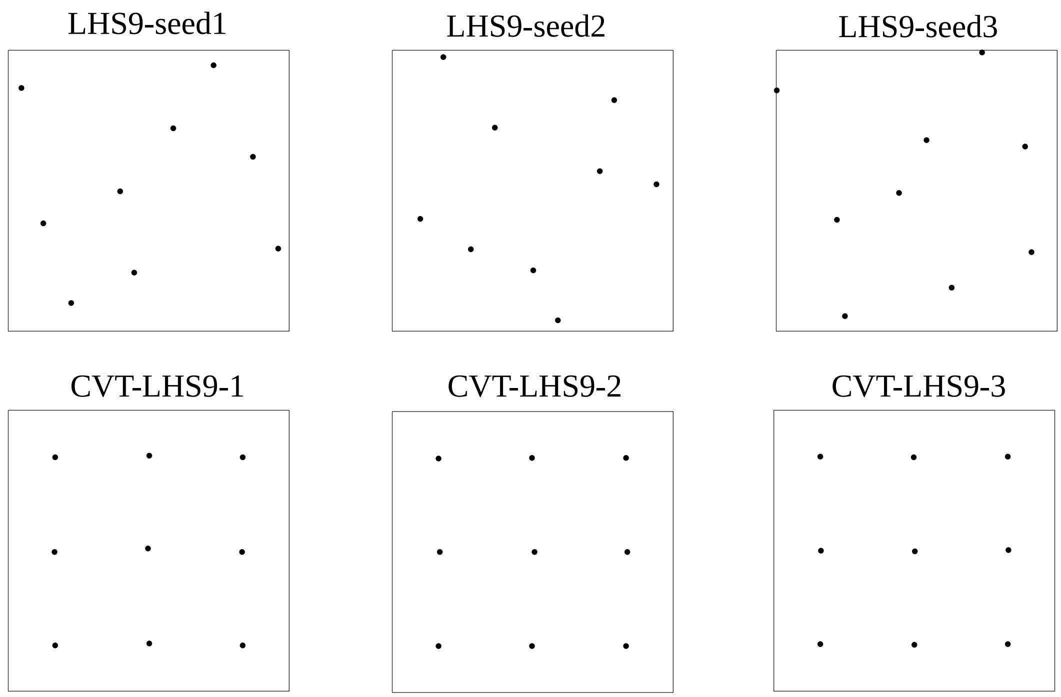

Figure 9: 9-point sample sets on 2-D unit Hcube for: A) Top Row– LHS Monte Carlo with three different initial seeds; and B) Bottom Row– corresponding CVT sets starting from LHS sets.

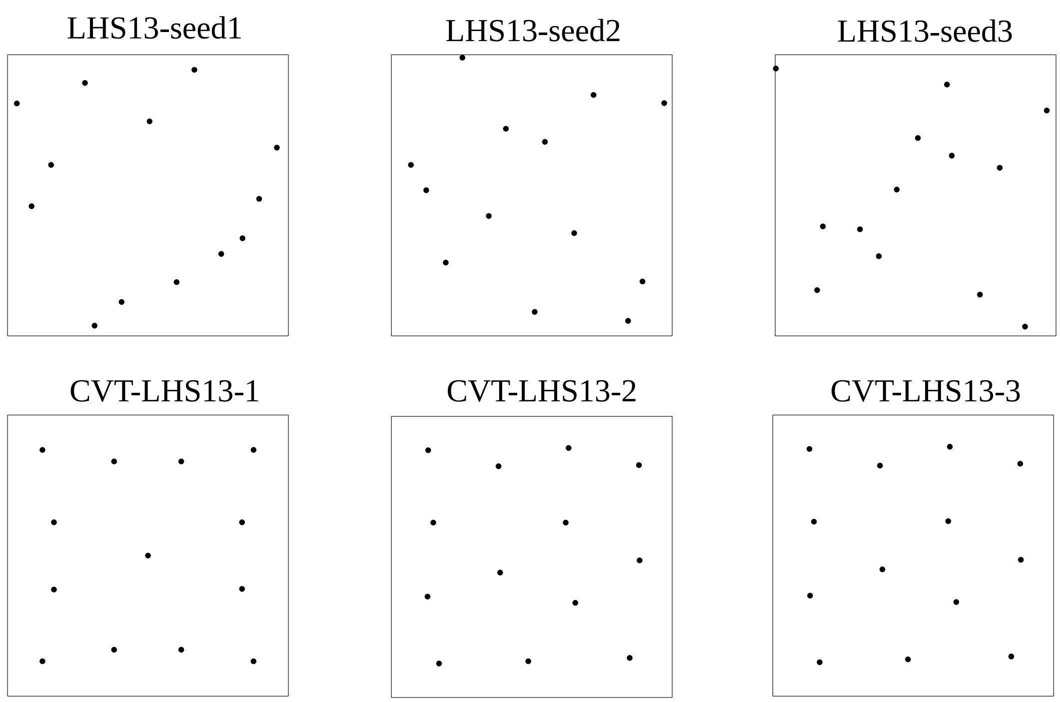

Figure 10: 13-point sample sets on 2-D unit Hcube for: A) Top Row– LHS Monte Carlo with three different initial seeds; and B) Bottom Row– corresponding CVT sets starting from LHS sets.

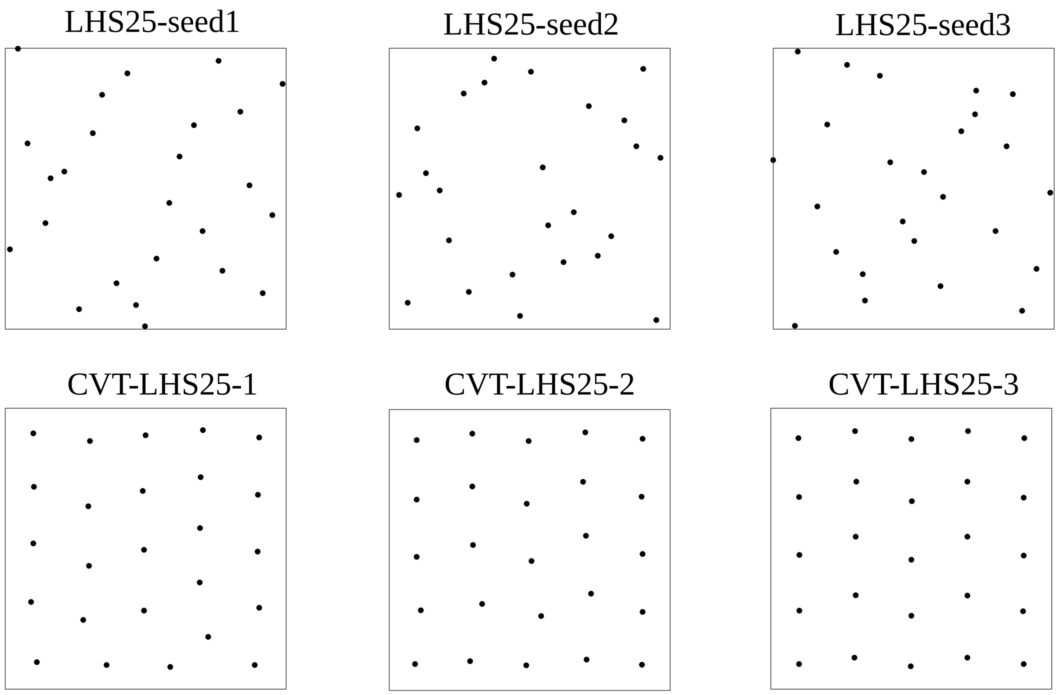

Figure 11: 25-point sample sets on 2-D unit Hcube for: A) Top Row– LHS Monte Carlo with three different initial seeds; and B) Bottom Row– corresponding CVT sets starting from LHS sets.

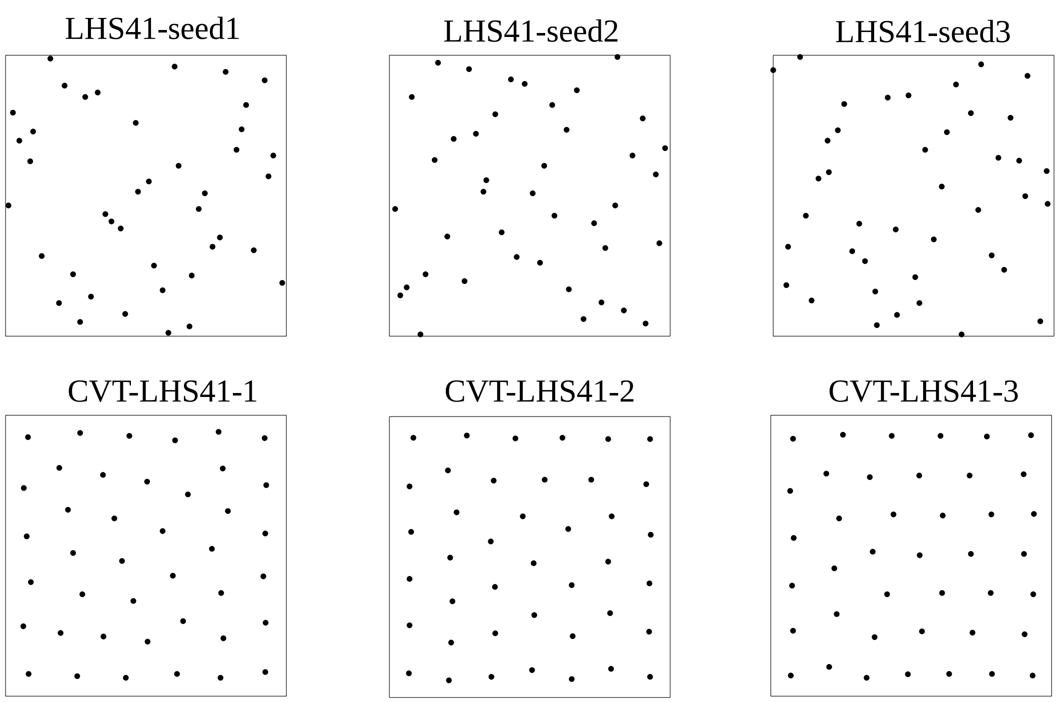

Figure 12: 41-point sample sets on 2-D unit Hcube for: A) Top Row– LHS Monte Carlo with three different initial seeds; and B) Bottom Row– corresponding CVT sets starting from LHS sets.