RESEARCH ARTICLE

Isovolumetric adaptations to space-filling design of experiments

Koushyar Komeilizadeh[1] · Arne Kaps[1] · Fabian Duddeck[1]

Abstract

A brief review of methods in design of experiments and criteria to determine space-filling properties of a set of samples is given. Subsequently, the so-called curse of dimensionality in sampling is reviewed and used as motivation for the proposal of an adaptation to the strata creation process in Latin hypercube sampling based on the idea of nested same-sized hypervolumes. The proposed approach places samples closer to design space boundaries, where in higher dimensions the majority of the design space volume is located. The same idea is introduced for Monte Carlo considering an affordable number of samples as an a-posteriori transformation. Both ideas are studied on different algorithms and compared using different distance-based space-filling criteria. The proposed new sampling approach then enables more efficient sampling for optimization especially for high-dimensional problems, i.e. for problems with a high number of design variables.

1 Introduction

Design of experiments (DoE) deals with the problem of distributing samples in a design space. As such, it is highly relevant in various fields of applications ranging from physical experiments over optimization problems to all kinds of modeling problems. Every population-based optimization as well as every surrogate model requires some form of DoE.

✉ Arne Kaps
arne.kaps@tum.de

1 TUM School of Engineering and Design, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany

DoE in its original form predates the computer era and was initially aimed at physical experiments. With the advent of computers and consequently numerical simulations, the need for different and more flexible ways of sampling has emerged and as a result the DoE literature has vastly expanded. In the original branch, called classical DoE, sampling is deterministic, i.e., the location of samples is predefined. In many classical designs, like full-factorial, samples are located in regular patterns with many samples close to design space boundaries. Then, the variation of the response can be estimated over the entire range of the inputs. For full factorial designs, many sample points are at the boundaries, especially the lower-levels ones. The number of samples in factorial designs grows very fast with dimensions, and only certain number of samples are possible. More information on classical DoE can for example be found in Montgomery (2009).

In modern DoE, which is also called *design of computer experiments* (DoCE), more focus is put on uniformity and space-filling properties of samples across the design space. To that end, multiple space-filling criteria have been suggested [for example in Johnson et al. (1990), Morris and Mitchell (1995)]. Overall, they can be divided into two groups – uniformity-based measures like discrepancy or entropy and distance-based measures (Garud et al. 2017). Discrepancy in this context can be thought of as the deviation of samples from a uniform distribution.

In modern DoE, the sampling itself is usually (quasi) stochastic. One popular class of approaches are (Quasi) Monte Carlo ((Q)MC) methods, which are based on samples being picked (quasi) randomly from the design space. The second popular class of approaches, which also includes Latin hypercube sampling (LHS), is based on subdivision of the design space (also called stratification) and subsequent random selection of the created strata. In LHS, each dimension is usually divided into uniform strata and subsequent sampling performed in a way that each stratum in each dimension only contains a single sample. DoE can be one-shot like LHS and (Q)MC or in addition designed to be sequential. In the latter case, the new samples are defined based on keeping a desired property like space-filling (Wu et al. 2017; Li et al. 2017) and in addition being non-collapsing (Crombecq et al. 2011). Also, the output (response) may be taken into account to balance exploration and exploitation. In other words, the sequential design may be adaptive based on the output values. A more comprehensive review of DoE can be found in Garud et al. (2017). DoE is used in many research areas such as numerical integration (Harase 2019), uncertainty quantification (Abdar et al. 2021) and construction of surrogate models (Chen et al. 2006; Alizadeh et al. 2020). Since surrogate models represent a rather efficient approximation to the underlying functions, they are extensively used in various types and classes of optimization, such as, multifidelity and multidisciplinary (Zadeh et al. 2016), Bayesian (Hebbal et al. 2021), robust design (Wurm and Bestle 2016) and so on. DoE has also been used as a baseline for validation of an optimization algorithm (Frank et al. 2018). More information on surrogate modeling using DoE can be found in Fang et al. (2017) and Yondo et al. (2018). In this work, the focus lies on enhancing Latin hypercube and Monte Carlo sampling. The most relevant uniformity criteria as well as sampling algorithms will be introduced in more details later.

One big challenge in DoCE is commonly named the *curse of dimensionality*. It describes the problems that arise as the number of design variables, and hence

the dimension of the design space, becomes large. Roughly speaking, the number of required samples to fill out the design space grows exponentially. Also intuitions that are obtained from two- or three-dimensional space considerations may become invalid.

Given the current computational power and lower dimensions, one can often obtain the desirable results even when evaluations are expensive. As the number of considered inputs increases, the required number of samples can get out of reach fast due to the curse of dimensionality. In recent years, some suggestions have been made on how to alleviate the effects of curse of dimensionality. Vořechovský and Eliáš (2020) and Vořechovský and Mašek (2020) suggested an adaptation of uniformity criteria for a general-purpose, including (optimization of) LHS. One consequence of the curse of dimensionality is the fact, that one needs to be more economical with the sampling scheme. Adaptivity, i.e., placing samples according to the demand, would be a favorite method. However, in practice, the underlying function is not known in advance and the first stage of an adaptive scheme would more likely be a uniform design. MC, QMC and (optimal) LHS provide various levels of uniformity. (Sample) uniformity means samples are spread over the space evenly, and one tries to improve this uniformity via methods such as optimization of a distance based criterion as will be discussed later. This means that in the case of numerical integration not only sample uniformity but also statistical uniformity are required, which makes QMC methods and especially Sobol suitable candidates and neither MC nor LHS. See Vořechovský and Eliáš (2020) for more detailed discussion of uniformity and requirements for numerical integration. In order to see the difficulty of uniformity in higher dimensions, consider the following example. A d-dimensional unit hypercube $[0, 1]^d$ can be divided into 2^d hyper-octants that all share the point $[0.5, 0.5, …]$. So, in dimension fifteen, there will be $2^{15} = 32768$ hyper-octants. The number of affordable evaluations is often less or in the same order (e.g., for simplified models). This means in case of expensive-enough simulations, many of the hyper-octants may remain without a single sample while others would have one or even multiple samples. Sobol sampling avoids this by filling hyper-octants one by one (Kucherenko et al. 2015), while LHS and MC do not.

Here, we do not try to improve the sample uniformity of these methods, but to have more samples at regions with higher shares of total design space volume. Hence, our method is applicable to optimization, prediction and response surfaces but not numerical integration as stated above. We propose a variation of the stratification process in LHS. Instead of using equally sized strata, our approach varies the size of strata depending on the dimensionality of the design space and number of samples. Just as in LHS, all dimensions are treated the same way. We compare our approach to the standard MC and LHS using common space-filling criteria that will be introduced later.

This paper is structured as follows. Initially, we review some space-filling criteria and sampling methods in section 2 and 3 respectively. Then, we briefly discuss the so-called curse of dimensionality and its implications to Latin hypercube sampling in section 4. Based on that, we introduce our isovolumetric weighting approach and outline its expected benefits and disadvantages in section 5. Initial results are discussed

in section 6. We conclude with a summary of our findings and an outlook to possible applications.

2 Space-filling criteria

As mentioned before, in modern DoE without any prior knowledge about the design space, the target is usually to fill as much of the design space as uniformly as possible. In this regard, two major criteria classes help to measure the space-filling properties of a sampling method. One is based on uniformity and the other one on distance. Other criteria such as entropy will not be discussed here.

The following summary is based on Garud et al. (2017). First, we define discrepancy. The aim is to fill a hyper-rectangular design space called $S \in \mathbb{R}^d$. Consider H as a subspace of S with the volume $V(H) = \Delta x_1 \Delta x_2 ... \Delta x_d$. If $\sharp(\bullet)$ represents the number of samples in a space, then the L_∞-discrepancy is defined as

$$D = \sup_H \left| \frac{\sharp H}{\sharp S} - V(H) \right|. \tag{1}$$

The lower the discrepancy, the more uniformly the samples fill the design space. The interpretation is that the proportion of samples in a subspace should be proportional to the subspace volume. This formula is not straightforward for numerical implementation and has in fact been shown to be an NP-hard problem (Gnewuch et al. 2009). Hence, different versions based on L_2 norms have been proposed [for example by Hickernell (1998)]. They are of little practical relevance in medium- to high-dimensional setups due to their complexity or the high number of required points. Especially in DoE methods based on optimizing a space-filling criterion, the second class of criteria is far more popular [compare for example Table 1 in Garud et al. (2017)]. Hence, the focus in the following will lie on distance-based criteria. However, it is expected that the adaptations presented below would increase the discrepancy of sample sets, i.e. worsen results in these criteria. Distance-based criteria are usually more popular when optimizing different DoEs for their space-filling properties. An overview of the most popular criteria is given in Table 1. Of the shown criteria, Φ_p, potential energy and Maximin are most commonly used for optimizing sampling, e.g. in OLHS, due to their low computation times. Among the listed criteria, all but Minimax try to put samples further from each other.

Distance-based criteria can be combined to create new ones, e.g., Bhattacharyya summed up the weighted intersite distance and projective distance. Intersite distance is a form of Maximin criterion and projected distance helps to generate samples with unique coordinates; so by removing dimensions, samples would not coincide in any projection (Bhattacharyya 2018).

Table 1 Some distance-based criteria for uniformly distributing samples in the design space.

Method Name	Equation	Explanation
Morris and Mitchell Φ_p	$\Phi_p = \left[\sum_{k=1}^{N-1}\sum_{j=k+1}^{N} d(x^{(j)},x^{(k)})^{-p}\right]^{1/p}$ (lower is better)	Suggested in Morris and Mitchell (1995). The lower Φ_p, the more uniform the design.
Audze and Eglais potential energy	$\Phi_{PE} = \sum_{k=1}^{N-1}\sum_{j=k+1}^{N} d(x^{(j)},x^{(k)})^{-2}$ (lower is better)	Special case of Morris and Mitchell Φ_p criterion with Euclidean distance suggested originally by Audze and Eglais (1977). The square root of the result is not considered due to its monotonicity.
Maximin	$\Phi_{Mm} = \min_{j\neq k}\left[d(x^{(j)},x^{(k)})p\right]$ (higher is better)	Suggested in Johnson et al. (1990). $d(x^{(j)},x^{(k)})$ is defined as the distance function in the original publication. It is redefined here for consistency.
Minimax	$\Phi_{mM} = \max\left[\min_{j=1,2,...,n}\left[d(x,x^{(j)})p\right]\right]$ (lower is better)	Suggested in Johnson et al. (1990). Determining this criterion is more difficult than the Maximin criterion as it includes all points in the design space (Pronzato 2017).
Minimum spanning tree (MST)		Samples are considered as nodes in a graph and they are connected while the sum of distances (i.e. edge lengths) is minimized, involving all samples and without making closed loops. This is called a tree and it is made by starting from one sample and adding others step-wise (Dussert et al. 1986). At the end, if the mean of all edge lengths is large and the corresponding standard deviation is low, the current set of samples has a good space-filling character.

$x^{(j)}$ is the j-th sample. d represents the number of dimensions and N the number of samples. Here, $d(x^{(j)},x^{(k)}) = \sum_{i=1}^{d}\left|x_i^{(j)} - x_i^{(k)}\right|$ and $d(x^{(j)},x^{(k)})p$ is a distance function with $p > 0$ defining the distance norm

3 Sampling methods

Two of the most extensively used sampling methods – (Quasi) Monte Carlo Sampling (MCS) and LHS – are described in the following.

MCS is very popular in a number of applications due to the simplicity of the approach. Samples are drawn pseudo-randomly from the design space. However, when the intention is to have samples uniformly distributed in a hypercube, more efficient methods under the name of Quasi Monte Carlo (QMC) were developed which are based on low-discrepancy sequences of prime numbers. Popular examples are Hammersley, Halton, Faure and Sobol sequences (see for example Niederreiter (1992) or Antinori (2017) for more details). Low-discrepancy is considered here with respect to uniform space-filling property. Although QMC methods are designed to have better space-filling properties than MC, they are still prone to some problems. Halton samples may line up in 2D projections (Glasserman 2013) and points of Sobol sequences may pair with each other and create a pattern of clusters and gaps in 2D projections (Joe and Kuo 2008). It should be noted that improvements have been suggested trying to keep the merits of the approaches. For example, Bratley and Fox provided an implementation of Sobol for up to 40 dimensions (Bratley and Fox 1988), which was increased later to 1111 dimensions (Joe and Kuo 2003), followed by further improvements in (Joe and Kuo 2008) where some of the poor 2D projections were fixed while keeping Sobol a fast method. Halton's performance has also been enhanced to have a lower discrepancy by tuning the possible permutations of its sequence using an evolutionary algorithm (De Rainville et al. 2012). Re-randomization of QMC has also been suggested, which preserves their low-discrepancy properties while providing an unbiased estimator of errors in numerical integration (L'Ecuyer and Lemieux 2002).

Another common space filling method is Latin hypercube (LH). The construction of a Latin hypercube design (LHD) for N samples in d dimensions works as follows. In each dimension the space is divided into N strata of equal probability. That means that the design space is divided into N^d cells. N of these cells are selected at random with the condition that in each dimension each stratum may only contain one sample (McKay et al. 1979). Each sample can be placed in the center of its cell or randomly located within (Rajabi et al. 2015). In this work, we consider the centered case.

One advantage of LH is that, if some dimensions are removed, the remaining design is still an LHD, even though space-filling properties and correlation may deteriorate (Viana 2013). More details on practical use and remaining challenges of LHS can also be found in Viana (2016).

Like QMC methods, samples may not fill the space uniformly enough and considerable correlations exist between dimensions. One example is the case, where all samples are lined up on the main diagonal of the design space, thus exhibiting perfect linear correlation and low space filling while still being a valid LHD. Therefore, modifications have been proposed to improve LHD. Ye introduced an orthogonal LHD in which the correlation is zero between all dimensions (Ye

1998). However, as he also mentioned, this does not necessarily translate into good space-filling properties. In addition, an orthogonal LHD for a given number of samples may not exist for some dimensions. These problems were improved considerably in Cioppa and Lucas (2007), by allowing small correlations (e.g., within [-0.3,0.3]) and a condition number of slightly greater than one. Authors also optimized these nearly orthogonal designs to obtain better uniformity as well. Ye proposed the symmetric LHD (Ye et al. 2000), in which for any row "i" of an LHS there exists another row in the design which is the "ith" row's reflection through the center. This initially symmetric design shows better optimality compared to standard LHS and therefore optimization is less costly to get a good end result. If LHD have uniformity in each edge, in the distributed hypercube sampling (DHS) the two-dimensional projection of samples are tried to be well distributed by having a low coefficient of variance of the minimum distance between the projected samples (Manteufel 2001). In an improvement to DHS, an optimal distance constraint is constructed based on the ratio of the space volume and the number of samples. Then, several samples are generated and their minimum distances to each sample in a subspace are calculated. The sample that has the minimum distance closest to the mentioned optimal distance constraint is selected next (Beachkofski and Grandhi 2002).

Optimal LH (OLH) design is among the most powerful improvements of LHD that can provide good space-filling properties. However, OLH can take significant time for higher dimensions and number of samples. For this combinatorial problem, the number of outcomes grows extremely fast as $(N!)^d$ (Viana 2013). Therefore, creating a high number of LH designs and comparing them based on an optimality criterion is not an efficient way of optimizing and hence a couple of methods are suggested to use an optimization procedure and not just selecting the best out of many designs. In these OLH, the objective function is one of the uniformity criteria discussed earlier (see Table 1). A common procedure used within optimization to change the design is the columnwise-pairwise procedure, in which a column (i.e., a dimension) in the design is selected and two elements in the column will exchange places (i.e., their corresponding level value). Morris and Mitchell used Simulated Annealing as the optimizer with both column and element selection realized randomly. If the new design is better based on the potential energy criterion (Table 1, second row), it will replace the old one, but if it is not there is still a chance that it will be accepted (Morris and Mitchell 1995). OLH design (OLHD) as proposed in Ye et al. (2000) also employs the columnwise-pairwise idea, however, this time, new designs are obtained not randomly but deterministically and inferior designs are never accepted. More elaborate optimization schemes such as Enhanced Stochastic Evolutionary algorithm (ESE) (Jin et al. 2005) have been proposed but are not further investigated here. OLHD with periodic structure which employs ESE has been proposed in Husslage et al. (2011) for up to 10 dimensions. It should be noted that optimization has also been employed to create designs with the desired correlation among inputs (Vořechovský and Novák 2009).

After the brief introduction above, we illustrate the performance of MC, LHD and OLHD in two dimensions. These are the methods that will be the focus of the present work. In the case of the OLH, optimization is done using an in-house code
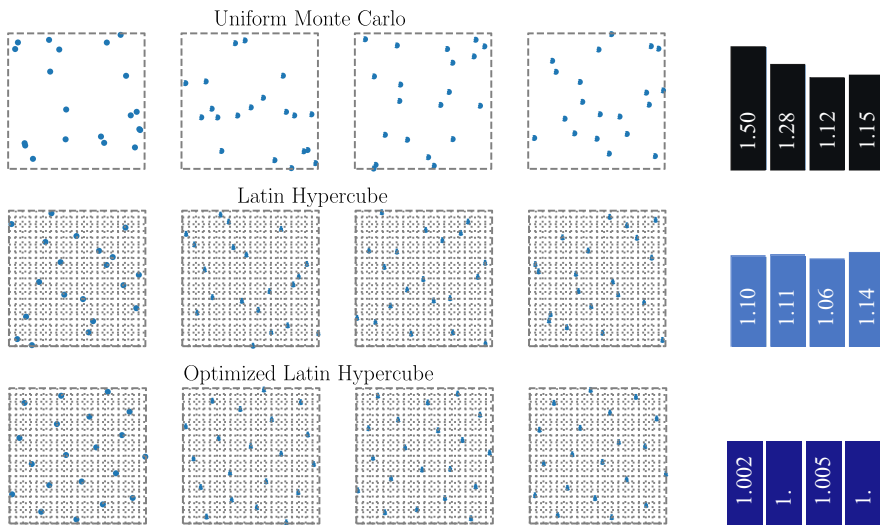
Fig. 1 Four cases of two-dimensional samples generated by uniform MC on top, LH in the middle and OLH at the bottom are shown on the left side. On the right, the relative values of the potential energy criterion for each case are given within the corresponding bar plots. The values are normalized with respect to the lowest one

of the method suggested by Morris and Mitchell (1995). Figure 1 compares uniformity of MC, LH and OLH based on the potential energy (PE) criterion in two dimensions. It can be seen that MC has the highest PE, therefore it is the worst in terms of uniformity and OLH is the best. The variation among the four exemplary cases is the lowest in OLHD and highest in MC, indicating that OLHD is also the most robust. The results are according to our expectations concerning the relative behavior of these methods. Note that, Fig. 1 is just illustrative to show uniformity. However, in terms of time consumption and based on our experience in low dimensions, the employed OLHD is also more efficient than selecting the best results out of many generated samples. The question remains whether the relative performance remains the same in higher dimensions. Before discussing this, we review concisely the consequences of increasing the number of dimensions in the next section.

4 Curse of dimensionality

When increasing the input dimensionality, an exponential growth in the required number of samples to represent the output well-enough is expected. Still, in higher dimensions intuitions we have from two- or three-dimensional space can easily become invalid. For example, the volume of a unit hypersphere increases with dimension, but only up to dimension five and then strictly decreases (Verleysen 2003).

If placing samples randomly, the higher the number of dimensions, the more samples are found close to boundaries of the design space. As the number of dimensions

increases, a higher share of the unit hypercube volume is found close to the boundaries. A similar discussion can be found for example in Lange et al. (2018).

The considered distance norm in this work is Euclidean, i.e., the "L_k, $k = 2$" norm. As dimensionality increases, this norm differentiates far away points less and less. Hence, depending on the application and dimension, one may use more suitable norms (Aggarwal et al. 2001; Vořechovský and Mašek 2020).

Obviously, these issues of high dimensionality do not happen by going from one dimension to the next, but more gradually. Typically the range of uniformity of a series of LH designs is within that of a series of MC designs in higher dimensions. The best MC design out of several DoEs can easily be more uniform than the best (optimized) LH design. In lower dimensions the range of LH designs is almost completely more uniform than MC designs. This effect can be observed in the figures in sect. 6. We observed above that as dimension increases, most of the design space volume ends up close to the boundaries rather than the center. When applying LHS in applications with higher dimensionality it seems less optimal. The bin creation in standard LHS divides each dimension uniformly. Hence, many samples may be generated in the center of the design space even though it only contains a small share of the total hypercube volume. A similar argument applies to MC sampling where obtaining desired properties such as uniformity or a special point of focus may be hard if the number of samples is not large. This observation leads us to our proposed stratum design for LH design that is introduced in the following.

5 Isovolumetric weighting approach

As discussed before we want to obtain an approach that naturally places more samples closer to design space boundaries. This focus on the boundaries should increase with an increasing number of dimensions. In common LHD we divide each dimension independently into N uniform strata where N is the number of samples we want to generate. Now let us assume that N is even. This should not be a limitation in high dimensional situations where N is commonly in the hundreds or thousands. What happens if we regard the strata as $N_v = \frac{N}{2}$ nested hypervolume shells? The outermost hypervolume would obviously be the largest. Also each hypervolume except the innermost could contain at maximum $2d$ samples with d the dimensionality of the design space. The innermost hypervolume may contain a maximum of 2 samples.

We propose to enforce all of the nested hypervolumes to be of the same size (compare differently colored regions in Fig. 2). Assuming the design space to be a unit hypercube, the stratum boundaries p_i and sizes a_j in each dimension for the standard LHS are given by $p_i = \frac{i}{N}$ and $a_j = \frac{1}{N}$. When enforcing our condition of identical volume for different hypervolume shells we can replace these two equations with

$$p_i = \begin{cases} 0.5\left(1 - \left(\frac{N_v+1-i}{N_v}\right)^{1/d}\right), & i \in \{1, 2, ..., N_v\} \\ 0.5\left(1 + \left(\frac{i-(N_v+1)}{N_v}\right)^{1/d}\right), & i \in \{N_v + 1, N_v + 2, ..., N + 1\} \end{cases} \tag{2}$$
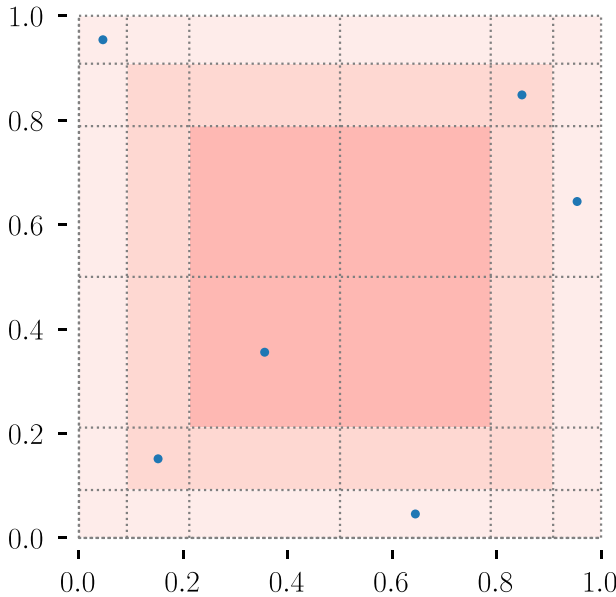
Fig. 2 Isovolumetric weighted sampling for six samples. This is basically a LH, while strata of same probability now have different width. Regions of same color have the same area. The beginning and end of strata, i.e., where dashed lines intersect an axis are symmetric with respect to the center $x_i = 0.5$ and the values are [0.091752, 0.211325, 0.500000, 0.788675, 0.908248] (see Eq. (2))

$$a_j = p_{j+1} - p_j \quad ,j \in \{1, 2, ..., N\}. \tag{3}$$

The equations can easily be adapted to any hyper-rectangular design space by transformation to a unit hypercube. A two-dimensional unit hypercube example is visualized in Fig. 2. It can be seen that the outer cells have smaller one-dimensional projections in each direction. All three regions of same color have the same area. Since all individual cells have an equal probability of being sampled, this means more focus on the design space boundaries. Requiring all regions to be of the same size naturally leads to the equations presented above.

Additionally, we apply our idea to (Q)MC designs as an a-posteriori transformation. The respective equation for each coordinate of a QMC sample point $\tilde{x}_k^{(j)}$ reads

$$x_k^{(j)} = \begin{cases} 0.5\left(1 - \left(1 - 2\tilde{x}_k^{(j)}\right)^{1/d}\right), & \tilde{x}_k^{(j)} \in [0, 0.5) \\ 0.5\left(1 + \left(2\tilde{x}_k^{(j)} - 1\right)^{1/d}\right), & \tilde{x}_k^{(j)} \in [0.5, 1] \end{cases}. \tag{4}$$

Here, $x_k^{(j)}$ denotes the k-th coordinate of the transformed sample point $\boldsymbol{x}^{(j)}$. This equation is derived from Eq. (2). We consider the limit case for a high number of samples N. When replacing N_v by N and taking into account that i/N is the coordinate value in the limit we get Eq. (4).

One advantage of our adaptations is that they are free of empirical parameters that would have to be chosen or tuned. Additionally, both approaches only need to be evaluated once per DoE creation so they should not add computational cost to the respective sampling approaches. This claim will be confirmed below. We expect our approach to increase space-filling properties if the number of design space dimensions is higher than five.

A slight disadvantage of the approach is the required even number of samples. As mentioned before, we believe this should not be a big constraint, especially in medium to high dimensional applications.

6 Discussion

Both adaptations are implemented into an in-house DoE code that does not use other sampling packages. Pseudo-code for both implementations can be found in the "Appendix". Results have also been reproduced with the SMT Python package (Bouhlel et al. 2019) by applying Eq. (4) to the output of the blackbox DoE code.

We investigate how both our sampling methods, isovolumetric LH (IVLH) and the isovolumetric transformed MC (IVMC), compare to their standard counterparts in the following. To show the potential of the methods, we additionally include OLH and its isovolumetric variation (OIVLH) into the comparisons. As optimization algorithm we use simulated annealing (SA) [as suggested by Morris and Mitchell (1995)] with an initial temperature $T = 0.01\Phi_{PE}$, cooling parameter $\alpha = 0.999$ and a maximum number of 50, 000 iterations. The optimization is performed with respect to the potential energy criterion Φ_{PE} (see Table 1). We use SA here due to the simplicity of the approach. We also experimented with the enhanced stochastic evolutionary (ESE) algorithm (Jin et al. 2005). It may show slight improvement in space-filling and can reduce computational costs compared to SA. However, all statements regarding comparison between our isovolumetric approach and standard methods are totally unaffected by the choice of optimization algorithm. We compare 100 designs for each method in each setup. Overall three distance-based space-filling criteria are displayed, with the focus on the potential energy criterion due to its simplicity. Φ_{mM} (see Table 1) is calculated as it rewards samples closer together, which contradicts our approach. However, it is of limited practical usability due to very long computation times in medium to high dimensional problems. For the calculation of Φ_{mM} we use the Markov-Chain-MC estimation suggested in Pronzato (2017). We choose the parameters of the estimation according to the ones used in the original publication except for $\epsilon = 0.1$ for the sake of performance.

As an initial setup we use a five-dimensional design space where 100 samples are to be generated. Results for different space-filling criteria are shown in Fig. 3. For the Maximin criterion we see no change for MC between our approach and the standard. Both, modified LHS and OLH perform slightly worse in that criterion compared to standard versions albeit showing lower variance. As expected, all our isovolumetric approaches also perform worse in the Minimax criterion with no significant difference between the different methods. For the potential energy criterion, we see that OIVLH, IVMC and IVLH perform similarly to the three standard
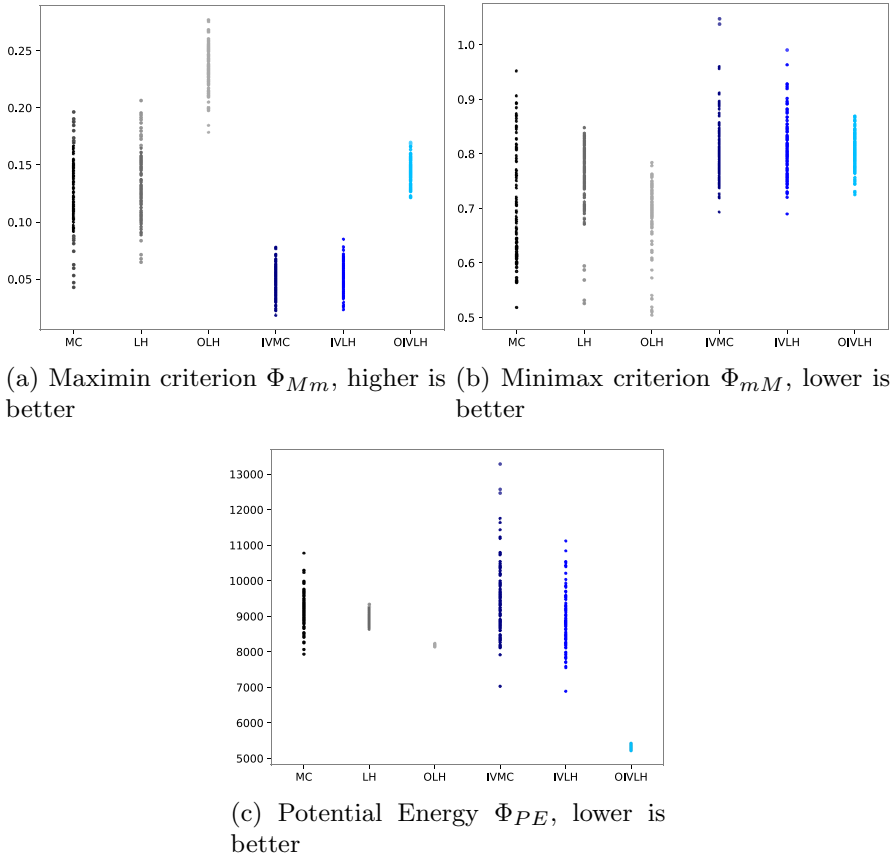
(a) Maximin criterion Φ_{Mm}, higher is better

(b) Minimax criterion Φ_{mM}, lower is better

(c) Potential Energy Φ_{PE}, lower is better

Fig. 3 Criteria for the setup of drawing 100 samples from five-dimensional unit hypercube. 100 cases are simulated for each method. All criteria are calculated for the same samples

approaches with slightly higher variances; this is especially true compared to (O) LH. However, OIVLH reaches significantly better results. This makes sense, as we use this criterion as objective in the optimization algorithm of OLH.

This example was designed to showcase how our approach performs in a lower dimensional setup. Considering that it was designed for high dimensionality, the results are as expected.

In Table 2 the average creation time for each of the methods along with standard errors are listed. The table serves to compare the computational cost between the standard variants of the methods and our isovolumetric adaptations. As expected, our method does not change the computational cost for LH and OLH. For MC sampling our adaptation seems to add a little time as it is applied as an a-posteriori transformation.

The second example we use is a 20-dimensional design space with 400 samples to be generated. Different space-filling criteria calculated for the same set of respective

Table 2 100 samples from a five-dimensional unit hypercube: Comparison of creation times for different DoE methods between standard variant and isovolumetric adaptation.

Variant	Method		
	MC [ms]	LH [ms]	OLH [s]
Standard	0.009 ± 0.000	0.191 ± 0.004	1.185 ± 0.001
Isovolumetric	0.034 ± 0.001	0.193 ± 0.003	1.147 ± 0.002

All times are mean and standard error of 100 repetitions

samples are depicted in Fig. 4. For this case, the results in potential energy criterion are very clear. All our proposed methods achieve significantly better results than all classic methods. This is largely expected as our method is designed to put samples further apart and closer to design boundaries. Interestingly, this shows only partially for the Maximin criterion. While the majority of cases still shows an improvement over the respective classic methods, there is a non-negligible number of outliers,
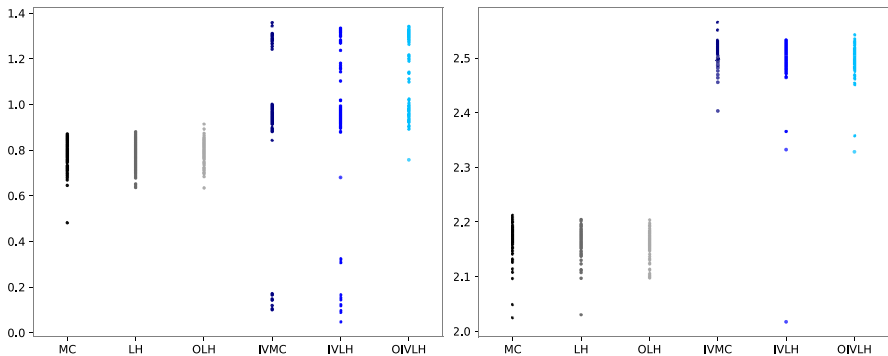
(a) Maximin criterion Φ_{Mm}, higher is better

(b) Minimax criterion Φ_{mM}, lower is better
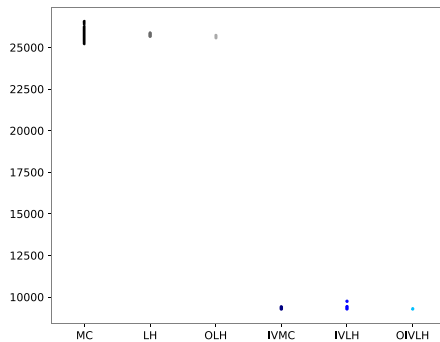
(c) Potential Energy Φ_{PE}, lower is better

Fig. 4 Criteria for the setup of drawing 400 samples from 20-dimensional unit hypercube. 100 cases are simulated for each method. All criteria are calculated for the same samples

especially for IVMC and IVLH methods. The reason is that the Maximin criterion is highly sensitive to pairs of points being in close proximity of each other. Our transformation adds another step in the DoE creation process that may end up creating this kind of point pair. In the Minimax criterion the ranges of values are similar for all methods (except for one outlier in IVLH). Our modified methods perform slightly worse here than in the 5D example.

Obviously, different results in different criteria are due to the definitions of the criteria. Minimax rewards pushing samples together while still covering the whole design space, while the other criteria improve when samples are as far apart as possible. This highlights that care should be taken when choosing a criterion to optimize the sampling. As our method pushes samples towards boundaries it naturally favors Φ_p - type criteria (i.e., Maximin, potential energy, etc.).

Another interesting observation is how the isovolumetric transformation is far more effective than the optimization for potential energy and Maximin criterion while the computational cost is at the same time much less (e.g., compare IVLH to standard OLH in Table 3). Here, in contrast to the two-dimensional example the optimization mostly reduces the variance of the criteria (over several runs) without largely improving the mean score (compare Fig. 4c). We suspect that the reason the isovolumetric transformation outperforms the optimization lies in the structural change it is based on. In the common LH the same width assumption forces the optimizer to put samples in the middle of space while in higher dimensions a large part of the volume is around the boundaries. This does not mean that the isovolumetric approach can overcome the required number of samples in higher dimensions, it just tries to be resourceful and place the samples considering their number and density of the space. To confirm these conclusions, further work is necessary.

In Table 3, the average creation time for each of the methods along with standard errors are listed. The table serves to compare the computational cost between the standard variants of the methods and our isovolumetric adaptations. The observations here are the same as before for the five-dimensional example and completely match with expectations.

As previously mentioned, our adaptation was designed with high dimensions in mind. Hence, we performed another study looking into the number of dimensions necessary for our approach to gain an advantage. The results are depicted in Fig. 5. We compare all six different approaches w.r.t. the Φ_{PE} criterion. The number of design space dimensions d is shown on the abscissa. The number of samples for each calculation is defined as $50 * d$. The colored areas represent the range of results for 100 repetitions of the sampling in each number of dimensions.

Table 3 400 samples from a 20-dimensional unit hypercube: Comparison of creation times for different DoE methods between standard variant and isovolumetric adaptation.

Variant	Method		
	MC [ms]	LH [ms]	OLH [s]
Standard	0.060 ± 0.001	1.273 ± 0.005	15.63 ± 0.018
Isovolumetric	0.221 ± 0.002	1.317 ± 0.004	15.70 ± 0.020

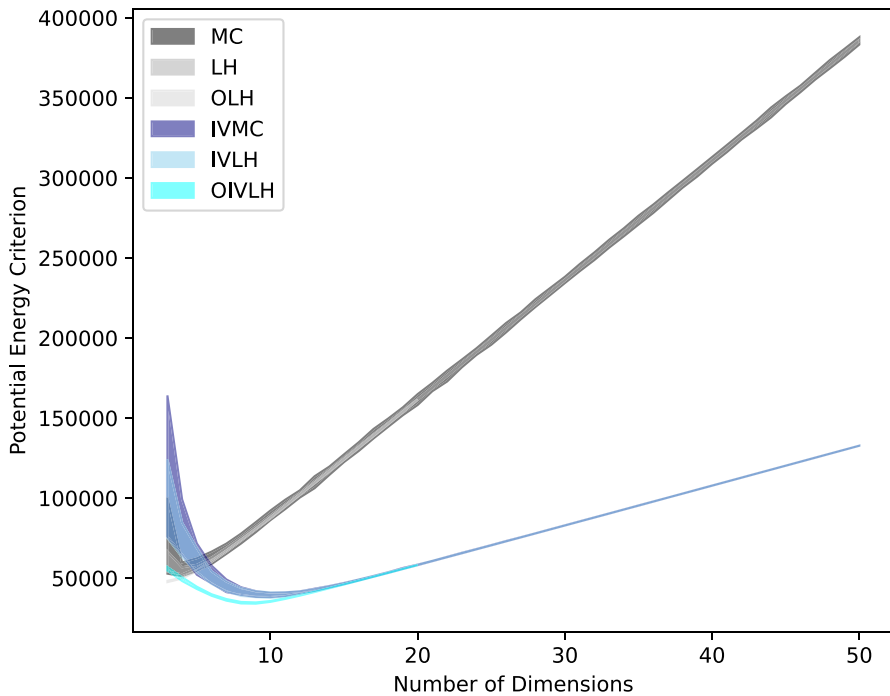All times are mean and standard error of 100 repetitions

Fig. 5 Potential energy criterion for different sampling methods is plotted over the number of dimensions. The respective number of samples is chosen to be fifty times the number of dimensions. The shaded areas represent ranges of values for 100 repetitions. Grey curves from dark to bright depict MC, LHS and OLHD. Blue curves represent the respective isovolumetric version

The overall value of Φ_{PE} grows as the dimension increases. For a dimensionality below five, the variability in all methods increases significantly (see Fig. 5). As the available design space shrinks, the probability for samples being very close together increases. In this range, our method shows no real advantage compared to the respective standard approaches. Interestingly also around five dimensions the volume of a unit hypersphere becomes maximal (Verleysen 2003). As the number of dimensions increases beyond five it can be seen that for our adaptations the slope of the curve is lower than for the standard methods. It can also be seen that in our case beyond around five (or ten if isovolumetric transformation is applied) dimensions O(IV)LH offers no real improvement compared to (IV)LH and (IV)MC, neither in the mean value nor in the range of results. Hence, in higher dimensional applications, it may be feasible to simply create a number of LH or MC designs and choose the best instead of employing an optimization algorithm. Here, a significantly better optimizer in higher dimensions, the number of the created samples, and a proper criterion are decisive.

The results of a similar study investigating the influence of the number of samples are depicted in Fig. 6. Here, the number of samples is shown on the abscissa and the number of dimensions is kept constant at 20. The potential
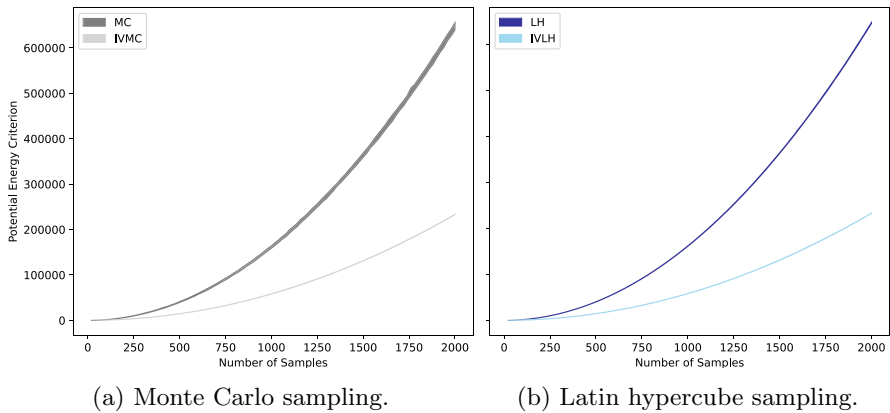
(a) Monte Carlo sampling.   (b) Latin hypercube sampling.

Fig. 6 Potential energy criterion for different sampling methods is plotted over the number of samples. A 20-dimensional design space is considered. The shaded areas represent ranges of values for 100 repetitions. Dark grey and dark blue curves depict MC and LH respectively. Light grey and blue curves represent the respective isovolumetric version

energy criterion is used for the comparison. Again, 100 repetitions are performed for each method with the depicted colored areas representing the ranges of the results. The potential energy criterion generally increases as the number of samples increases. This is not surprising as more samples actually have to be closer together. For our isovolumetric methods the order of the increase appears to be lower, i.e. the more samples we have the better they perform compared to their standard counterpart. The improvement seems legitimate as our isovolumetric method allows samples to better utilize areas close to design space boundaries which may increase distances between samples.

Projection to lower dimensions, especially to two dimensions, is one of the aspects that is considered in uniform space filling literature [see for example Damblin et al. (2013)]. Having space-filling two-dimensional projections is considered desirable; we have to show that, by eliminating other dimensions, the design is still space-filling. The question is however, if having uniform two-dimensional projections implies that the actual d-dimensional design is also always the better space-filling design. We question here the assumption that uniformity of projections in lower dimensions always translates into better higher dimensional space-filling properties, especially for significantly higher dimensions. One reason to raise such doubt is based on observing how LHS behaves. LHS is indeed uniform in one dimension but, this can get easily less and less in higher dimensions, while at least it also does not follow the space density.

Having a space-filling design is normally not the final goal itself, but it addresses for example, the efficient usage of an optimization algorithm. It remains to be investigated if improving space-filling criteria translates to increased performances in follow-up applications. In addition, not all dimensions may have the same effect on the output and the amount of effective dimensions and the order of interaction terms between inputs can vary. In some studies, these changes are taken into account

when comparing LHS and (Q)MC, see Kucherenko et al. (2015) for example. We discussed the results for a couple of designs. However, a big enough number of samples can have a larger impact than different designs (Liu 2005).

7 Conclusions

Design of experiments is an important step in many areas of applications from physical experiment over optimization to uncertainty quantification. Especially in computational applications, a common objective is to distribute samples as evenly as possible across the design space. In this work we move away from this uniformity objective by pushing samples to design space boundaries thus creating a kind of importance sampling. We propose adaptations to both (quasi) Monte Carlo and Latin hypercube sampling which increase the space-filling properties according to the popular potential energy criterion of the samples in higher dimensional applications. Our adaptation does not increase the computational requirements of LHS and does not require any empirical parameters. It can also be applied to other sampling methods as an a-posteriori transformation. We compare our adapted approach to the respective standard sampling methods in different distance-based space-filling criteria. We investigate the number of design space dimensions from which our approach gains an advantage over standard LHS or (Q)MC sampling. Additionally, we study how potential energy develops for a different number of samples at fixed number of dimensions. In both studies, the proposed isovolumetric methods get much better potential energy scores than their standard counterparts for dimensions larger than five and an affordable amount of samples. Our proposed method provides a complementary result to both standard LHS and optimal LHS by preferring boundaries of space as opposed to more uniform designs. As mentioned earlier, the term *space-filling* in higher dimensions is not clear and in practice we do not need a unique definition either. Space-filling design is rather part of a procedure with the aim of approximating an underlying function, either for fitting a response surface or for an optimization. We are interested to realize these procedures in the most efficient way. Based on the promising results presented above we list here some questions that require further investigation and may be the subject of future studies in the context of DoE or optimization:

- When there are more than five inputs, does the isovolumetric transformation perform better than the available LHS, (Q)MC in surrogate model construction? Will the rate of success be higher for surrogate models that change faster around the boundaries?
- Since the proposed method puts samples differently than the original LHS, how about combining both? Can the concept of isovolumetric sampling be used to create bins with samples partly in the center and partly close to the boundaries? Does this lead to a decent method on average?
- What if sample uniformity is also introduced within the proposed isovolumetric hypershells? Does this turn out to be universally decent from low to high dimensions? How would it compete with a QMC method like Sobol?

- Does applying the isovolumetric transformation at higher dimensions (above five) help with more efficient sensitivity analysis [see e.g., Kucherenko et al. (2015)]? How does it improve efficiency in uncertainty analysis, in comparison to QMC methods [see e.g., Hou et al. (2019)]?

Appendix

Algorithm 1 Variation on LHS bin creation

Require:
    N: Number of samples to be created (required to be even)
    d: Number of design space dimensions
Ensure:
    x: Array of LHS samples (size N x d)
1:   $N_v \leftarrow N/2$
2:   $sb \leftarrow array(N+1)$
3:   $i \leftarrow 1$
4:   **while** $i <= N+1$ **do** $\triangleright$ Create strata boundaries using Equation (2)
5:      **if** $i <= N_v$ **then**
6:         $sb[i] \leftarrow 0.5 \left(1 - ((N_v + 1 - i)/N_v)^{1/d} \right)$
7:      **else if** $i > N_v$ **then**
8:         $sb[i] \leftarrow 0.5 \left(1 + ((i - N_v - 1)/N_v)^{1/d} \right)$
9:      **end if**
10:      $i \leftarrow i + 1$
11:   **end while**
12:   $ss \leftarrow diff(sb)$ $\triangleright$ Strata sizes as defined in Equation (3)
13:   $refpts \leftarrow sb[1:N] + ss$
14:   $j \leftarrow 1$
15:   **while** $j <= d$ **do**
16:      $order \leftarrow random.permutation(range(N))$
17:      $x[:,j] \leftarrow refpts[order, j]$
18:   **end while**
19:   **return** x

Algorithm 2 A-posteriori transformation of blackbox DoE samples

---

Require:

  $\tilde{x}$: Samples created by a blackbox DoE code

  d: Number of design space dimensions

1: $x \leftarrow array(size(\tilde{x}))$

2: **for** $\tilde{x}_k^{(j)}$, $x_k^{(j)}$ in $\tilde{x}$, x **do**   ▷ Transformation as defined in Equation (4)

3:      **if** $\tilde{x}_k^{(j)} < 0.5$ **then**

4:          $x_k^j \leftarrow 0.5 \left(1 - \left(1 - 2\tilde{x}_k^{(j)} \right)^{(1/d)} \right)$

5:      **else if** $\tilde{x}_k^{(j)} >= 0.5$ **then**

6:          $x_k^j \leftarrow 0.5 \left(1 + \left(2\tilde{x}_k^{(j)} - 1 \right)^{(1/d)} \right)$

7:      **end if**

8: **end for**

9: **return** x

Data availability The detailed information of the proposed methods and the corresponding software employed in this paper can be found in the corresponding sections. This includes the algorithms and their hyper parameters.

Code availability Concerning the new sampling approach, a pseudo-code is given in the "Appendix". In addition, the used in-house Python code can be obtained from the corresponding author on reasonable request.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Abdar M, Pourpanah F, Hussain S et al (2021) A review of uncertainty quantification in deep learning: techniques, applications and challenges. Inf Fus. https://doi.org/10.1016/j.inffus.2021.05.008

Aggarwal CC, Hinneburg A, Keim DA (2001) On the surprising behavior of distance metrics in high dimensional space. In: International conference on database theory, Springer, Uppsala, Sweden, pp 420–434

Alizadeh R, Allen J, Mistree F (2020) Managing computational complexity using surrogate models: a critical review. Res Eng Design 31:275–298. https://doi.org/10.1007/s00163-020-00336-7

Antinori G (2017) Uncertainty analysis and robust optimization for low pressure turbine rotors. PhD thesis, Technische Universität München, Munich, Germany, http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20160804-1279009-0-6

Audze P, Eglais V (1977) New approach to the design of experiments. Probl Dyn Strength 35:104–107

Beachkofski B, Grandhi R (2002) Improved distributed hypercube sampling. In: 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and materials conference, p 1274

Bhattacharyya B (2018) A critical appraisal of design of experiments for uncertainty quantification. Arch Computat Methods Eng 25(5):727–751. https://doi.org/10.1007/s11831-017-9211-x

Bouhlel MA, Hwang JT, Bartoli N et al (2019) A python surrogate modeling framework with derivatives. Adv Eng Soft. https://doi.org/10.1016/j.advengsoft.2019.03.005

Bratley P, Fox B (1988) Algorithm 659: implementing Sobol's quasi-random sequence generator. ACM Trans Math Softw 14(1):88–100. https://doi.org/10.1145/42288.214372

Chen VC, Tsui KL, Barton RR et al (2006) A review on design, modeling and applications of computer experiments. IIE Trans 38(4):273–291. https://doi.org/10.1080/07408170500232495

Cioppa TM, Lucas TW (2007) Efficient nearly orthogonal and space-filling Latin hypercubes. Technometrics 49(1):45–55. https://doi.org/10.1198/004017006000000453

Crombecq K, Laermans E, Dhaene T (2011) Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling. Eur J Oper Res 214(3):683–696. https://doi.org/10.1016/j.ejor.2011.05.032

Damblin G, Couplet M, Iooss B (2013) Numerical studies of space-filling designs: optimization of Latin hypercube samples and subprojection properties. J Simul 7(4):276–289. https://doi.org/10.1057/jos.2013.16

De Rainville FM, Gagné C, Teytaud O et al (2012) Evolutionary optimization of low-discrepancy sequences. ACM Trans Model Comput Simul. doi 10(1145/2133390):2133393

Dussert C, Rasigni G, Rasigni M et al (1986) Minimal spanning tree: a new approach for studying order and disorder. Phys Rev B 34:3528–3531. https://doi.org/10.1103/PhysRevB.34.3528

Fang J, Sun G, Qiu N et al (2017) On design optimization for structural crashworthiness and its state of the art. Struct Multidisc Optim 55:1091–1119

Frank CP, Marlier RA, Pinon-Fischer OJ et al (2018) Evolutionary multi-objective multi-architecture design space exploration methodology. Optim Eng 19(2):359–381. https://doi.org/10.1007/s11081-018-9373-x

Garud SS, Karimi IA, Kraft M (2017) Design of computer experiments: a review. Comput Chem Eng 106:71–95. https://doi.org/10.1016/j.compchemeng.2017.05.010

Glasserman P (2013) Monte Carlo methods in financial engineering, vol 53. Springer Science & Business Media, chap Quasi-Monte Carlo. https://doi.org/10.1007/978-0-387-21617-1

Gnewuch M, Srivastav A, Winzen C (2009) Finding optimal volume subintervals with k points and calculating the star discrepancy are np-hard problems. J Complex 25(2):115–127. https://doi.org/10.1016/j.jco.2008.10.001

Harase S (2019) Comparison of Sobol' sequences in financial applications. Monte Carlo Methods Appl 25(1):61–74. https://doi.org/10.1515/mcma-2019-2029

Hebbal A, Brevault L, Balesdent M et al (2021) Bayesian optimization using deep Gaussian processes with applications to aerospace system design. Optim Eng 22(1):321–361. https://doi.org/10.1007/s11081-020-09517-8

Hickernell F (1998) A generalized discrepancy and quadrature error bound. Math comput 67(221):299–322. https://doi.org/10.1090/S0025-5718-98-00894-1

Hou T, Nuyens D, Roels S et al (2019) Quasi-Monte Carlo based uncertainty analysis: sampling efficiency and error estimation in engineering applications. Reliab Eng Syst Safety 191(106):549. https://doi.org/10.1016/j.ress.2019.106549

Husslage BG, Rennen G, Van Dam ER et al (2011) Space-filling Latin hypercube designs for computer experiments. Optim Eng 12(4):611–630. https://doi.org/10.1007/s11081-010-9129-8

Jin R, Chen W, Sudjianto A (2005) An efficient algorithm for constructing optimal design of computer experiments. J Stat Plan Inference 134(1):268–287. https://doi.org/10.1016/j.jspi.2004.02.014

Joe S, Kuo FY (2003) Remark on algorithm 659: implementing Sobol's quasirandom sequence generator. ACM Trans Math Softw 29(1):49–57. https://doi.org/10.1145/641876.641879

Joe S, Kuo FY (2008) Constructing Sobol sequences with better two-dimensional projections. SIAM J Sci Comput 30(5):2635–2654. https://doi.org/10.1137/070709359

Johnson ME, Moore LM, Ylvisaker D (1990) Minimax and maximin distance designs. J Sat Plan Inference 26(2):131–148. https://doi.org/10.1016/0378-3758(90)90122-B

Kucherenko S, Albrecht D, Saltelli A (2015) Exploring multi-dimensional spaces: a comparison of Latin hypercube and quasi Monte Carlo sampling techniques. preprint, arXiv:1505.02350

Lange VA, Fender J, Duddeck F (2018) Relaxing high-dimensional constraints in the direct solution space method for early phase development. Optim Eng 19(4):887–915. https://doi.org/10.1007/s11081-018-9381-x

L'Ecuyer P, Lemieux C (2002) Recent advances in randomized Quasi-Monte Carlo methods. Springer US, New York, pp 419–474. https://doi.org/10.1007/0-306-48102-2_20

Li W, Lu L, Xie X et al (2017) A novel extension algorithm for optimized Latin hypercube sampling. J Stat Comput Simul 87(13):2549–2559. https://doi.org/10.1080/00949655.2017.1340475

Liu L (2005) Could enough samples be more important than better designs for computer experiments? In: 38th Annual simulation symposium, IEEE, pp 107–115

Manteufel R (2001) Distributed hypercube sampling algorithm. In: 19th AIAA applied aerodynamics conference, p 1673

McKay M, Beckman R, Conover W (1979) Acomparisonof three methodsforselecting valuesofinputvariablesinthe analysisofoutputfrom acomputercode. Technometrics 21(2):239–245

Montgomery DC (2009) Design and analysis of experiments, 7th edn. Wiley, Hoboken

Morris MD, Mitchell TJ (1995) Exploratory designs for computational experiments. J Stat Plan Inference 43(3):381–402. https://doi.org/10.1016/0378-3758(94)00035-T

Niederreiter H (1992) Random number generation and quasi-Monte Carlo methods. SIAM, New Delhi. https://doi.org/10.1137/1.9781611970081

Pronzato L (2017) Minimax and maximin space-filling designs: some properties and methods for construction. J de la Société Française de Stat 158(1):7–36

Rajabi MM, Ataie-Ashtiani B, Janssen H (2015) Efficiency enhancement of optimized Latin hypercube sampling strategies: application to Monte Carlo uncertainty analysis and meta-modeling. Adv Water Resour 76:127–139. https://doi.org/10.1016/j.advwatres.2014.12.008

Verleysen M (2003) Learning high-dimensional data. Nato Sci Series Sub Series III Comput Syst Sci 186:141–162

Viana FAC (2013) Things you wanted to know about the Latin hypercube design and were afraid to ask. In: 10th World congress on structural and multidisciplinary optimization, Orlando, FL, USA

Viana FAC (2016) A tutorial on Latin hypercube design of experiments. Quality and Reliab Eng Int 32(5):1975–1985. https://doi.org/10.1002/qre.1924

Vořechovský M, Eliáš J (2020) Modification of the maximin and ϕ_p criteria to achieve statistically uniform distribution of sampling points. Technometrics 62(3):371–386. https://doi.org/10.1080/00401706.2019.1639550

Vořechovský M, Mašek J (2020) Distance-based optimal sampling in a hypercube: energy potentials for high-dimensional and low-saturation designs. Adv Eng Soft 149(102):880. https://doi.org/10.1016/j.advengsoft.2020.102880

Vořechovský M, Novák D (2009) Correlation control in small-sample Monte Carlo type simulations i: a simulated annealing approach. Probab Eng Mech 24(3):452–462. https://doi.org/10.1016/j.probengmech.2009.01.004

Wu Z, Wang D, Okolo PN et al (2017) Efficient space-filling and near-orthogonality sequential Latin hypercube for computer experiments. Comput Methods Appl Mech Eng 324:348–365. https://doi.org/10.1016/j.cma.2017.05.020

Wurm A, Bestle D (2016) Robust design optimization for improving automotive shift quality. Optim Eng 17(2):421–436. https://doi.org/10.1007/s11081-015-9290-1

Ye KQ (1998) Orthogonal column Latin hypercubes and their application in computer experiments. J Am Stat Assoc 93(444):1430–1439. https://doi.org/10.1080/01621459.1998.10473803

Ye KQ, Li W, Sudjianto A (2000) Algorithmic construction of optimal symmetric Latin hypercube designs. J stat plan Inference 90(1):145–159. https://doi.org/10.1016/S0378-3758(00)00105-1

Yondo R, Andrés E, Valero E (2018) A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. Prog Arospace Sci 96:23–61

Zadeh PM, Mehmani A, Messac A (2016) High fidelity multidisciplinary design optimization of a wing using the interaction of low and high fidelity models. Optim Eng 17(3):503–532. https://doi.org/10.1007/s11081-015-9284-z

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.