BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF CIVIL ENGINEERING

FAKULTA STAVEBNÍ

INSTITUTE OF STRUCTURAL MECHANICS

ÚSTAV STAVEBNÍ MECHANIKY

A DYNAMICAL PARTICLE SYSTEM AS A DRIVER FOR OPTIMAL STATISTICAL SAMPLING

DYNAMICKÝ ČÁSTICOVÝ SYSTÉM JAKO ÚČINNÝ NÁSTROJ PRO STATISTICKÉ VZORKOVÁNÍ

SHORTENED VERSION OF DOCTORAL THESIS

ZKRÁCENÁ VERZE DIZERTAČNÍ PRÁCE

AUTHOR   Ing. Jan Mašek
AUTOR PRÁCE

SUPERVISOR   prof. Ing. Miroslav Vořechovský, Ph.D.
VEDOUCÍ PRÁCE

BRNO 2018

Table of Contents

II

1   Introductory remarks

Due to the formerly unseen power of the contemporary *throughput-oriented* hardware, numerical simulations of vast, detailed models are conducted to complement or even replace costly physical experiments. Quite commonly, the output of these models is in fact a result of a complex deterministic function $g(\boldsymbol{X})$ of a vector of input variables, $\boldsymbol{X}$. The actual simulation of such a model, i.e. the evaluation of $g(\boldsymbol{X})$ for a given $\boldsymbol{X}$ might be and typically is a computationally intensive task. Still, running a computer simulation of a complex model for hours or days turns out cheaper than conducting an expensive physical experiment. Furthermore, in case of numerical models that require heavy computational effort, one might attempt to approximate the response of these complex models in a similar sense as these approximate the costly physical experiments. Such subsequent simplification efforts of numerical models that are expensive to solve are commonly recognized as *surrogate models*, *metamodels* or *response surface models*, see e.g. [1]. Based on known outputs from a limited number of *cleverly designed* runs of the more detailed model, a response surface model aims to construct a substitute approximative function of the input vector $\boldsymbol{X}$. Such a metamodel is constructed to offer an orders of magnitude faster evaluation in comparison to the approximated model while retaining a reasonably accurate output.

It comes out natural that when conducting physical experiments, simulating a numerical model or constructing a metamodel, to capture the behavior of any modelled system, realizations of a physical or numerical experiment shall (i) be as many as possible in compliance with computational or other costs and (ii) shall be designed with as diverse initial configurations as possible to provide a maximal amount of new information about the behavior of the system.

Efforts to optimize configurations of a finite number of realizations of an experiment in order to obtain unbiased and accurate results while reducing computational costs are often referred to as the Design of Experiments (DoE) [2, 3]. To design a configuration of a planned experiment, one has to set the value of each random variable that is considered by the numerical model or is under control in case of a physical experiment. An underlying assumption for computer experimentation is that these simulations are fully deterministic.

1.1 Monte Carlo estimation of properties of functions of random vectors

Approximation of properties of functions of random vectors can be computed in a similar fashion as the Monte Carlo method is able to approximate the solution of a deterministic process. The approximated value is typically considered to be a scalar value or a result of a scalar function of an array of outputs of the model.

Let us consider a random vector $\boldsymbol{X}$ that, instead of its joint probability distribution function, $f_X(x)$, is described by its independent continuous random variables X_1, X_2, $\ldots, X_{N_{\mathrm{var}}}$ along each marginal.

Commonly evaluated properties of random variables are their respective statistical moments [4]. In a generalized sense, a statistical moment of a random variable of an nth order is defined as follows:

$$\mu_n = \int\limits_{-\infty}^{\infty} (x - c)^n \ f(x) \, \mathrm{d}x, \tag{1}$$

where c is defines an offset towards which the moment is calculated. Quite commonly, *central moments* are calculated by substituting c for the mean value of random variable, μ_i:

$$\mu = \int\limits_{-\infty}^{\infty} x \ f(x) \, \mathrm{d}x. \tag{2}$$

Furthermore, let us introduce a function $g(\boldsymbol{X}) = g(X_1, X_2, \ldots, X_{N_{\mathrm{var}}})$ that transforms the random vector $\boldsymbol{X}$ into a random variable $Z = g(\boldsymbol{X})$. The function $g(\boldsymbol{X})$ might be a kind of analytic function or can be considered to be an unknown operator representing a numerical model, input of which is the random vector $\boldsymbol{X}$. The mean value of μ_z can be solved as follows:

$$\mu_z = \mu(g(\boldsymbol{X})) = \int\limits_{-\infty}^{\infty} z \ f_Z(z) \, \mathrm{d}z. \tag{3}$$

Since the density $f_Z(z)$ is generally not known, the integration is performed over the domain of input vector $\boldsymbol{X}$ and the transformation is weighed by the

joint density function f_X:

$$\mu_z = \int_{x_1} \int_{x_2} \cdots \int_{x_{N_{\mathrm{var}}}} g(\boldsymbol{x})\, f_X(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{4}$$

Due to the independence among the input variables, one can substitute the product of marginal density functions for the joint density function $f_X(\boldsymbol{x})$. Moreover, by substituting for $f_i(x_i)\mathrm{d}x_i = F_i(x_i)$ in Equation (4) results in:

$$\mu_z = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, , \ldots, x_{N_{\mathrm{var}}}) \prod_{i=1}^{N_{\mathrm{var}}} \mathrm{d}F_i(x_i). \tag{5}$$

At this point, it is possible to estimate the above mean value by using the Monte Carlo method. However, this would require sampling of points across the design domain of each marginal x_i, i.e. from negative to positive infinity. Another obstacle is that these sampling points must represent equal weights[1]. In this case, each sampling point must represent equal probability $\frac{1}{N_{\mathrm{sim}}} = f_i(x_i)\mathrm{d}x_i = \mathrm{d}F_i(x_i)$. In other words, the distribution of sampling points along x_i must be uniform with respect to probabilities, not x_i itself. Therefore, yet another transformation is proposed by substituting $\mathrm{d}F_i(x_i) = \mathrm{d}u_i$, where u_i is the probability $P(X_i < x)$, explore Figure 1:

$$u_i(x) = \int_{-\infty}^{x} f_i(t)\,\mathrm{d}t = F_i(x). \tag{6}$$

Then, performing a spatially uniform sampling of points in the *domain of probabilities*, $\boldsymbol{U} = (U_0, U_1, \ldots, U_{N_{\mathrm{var}}-1})$, results also in an uniform sampling with respect to probabilities in the *real space* and that is what we have been after. The problem is hence transformed into the domain of *uniformly distributed probabilities* U_i, as oppose to the original space of $X_1, X_2, \ldots, X_{N_{\mathrm{sim}}}$:

$$\mu_z = \int_{0}^{1} \cdots \int_{0}^{1} g\left(F_1^{-1}(u_1), \ldots, F_{N_{\mathrm{var}}}^{-1}(u_{N_{\mathrm{var}}})\right) \underbrace{\mathrm{d}u_1 \ldots \mathrm{d}u_{N_{\mathrm{var}}}}_{\mathrm{d}\boldsymbol{u}^{N_{\mathrm{var}}}}. \tag{7}$$

---

[1]In the example of a definite integral, these "weights" are the partial volumes represented by each point, $\frac{1}{N_{\mathrm{sim}}}$. That is because the volume is uniformly distributed across the domain, unlike the probability densities $f_i(x_i)$.
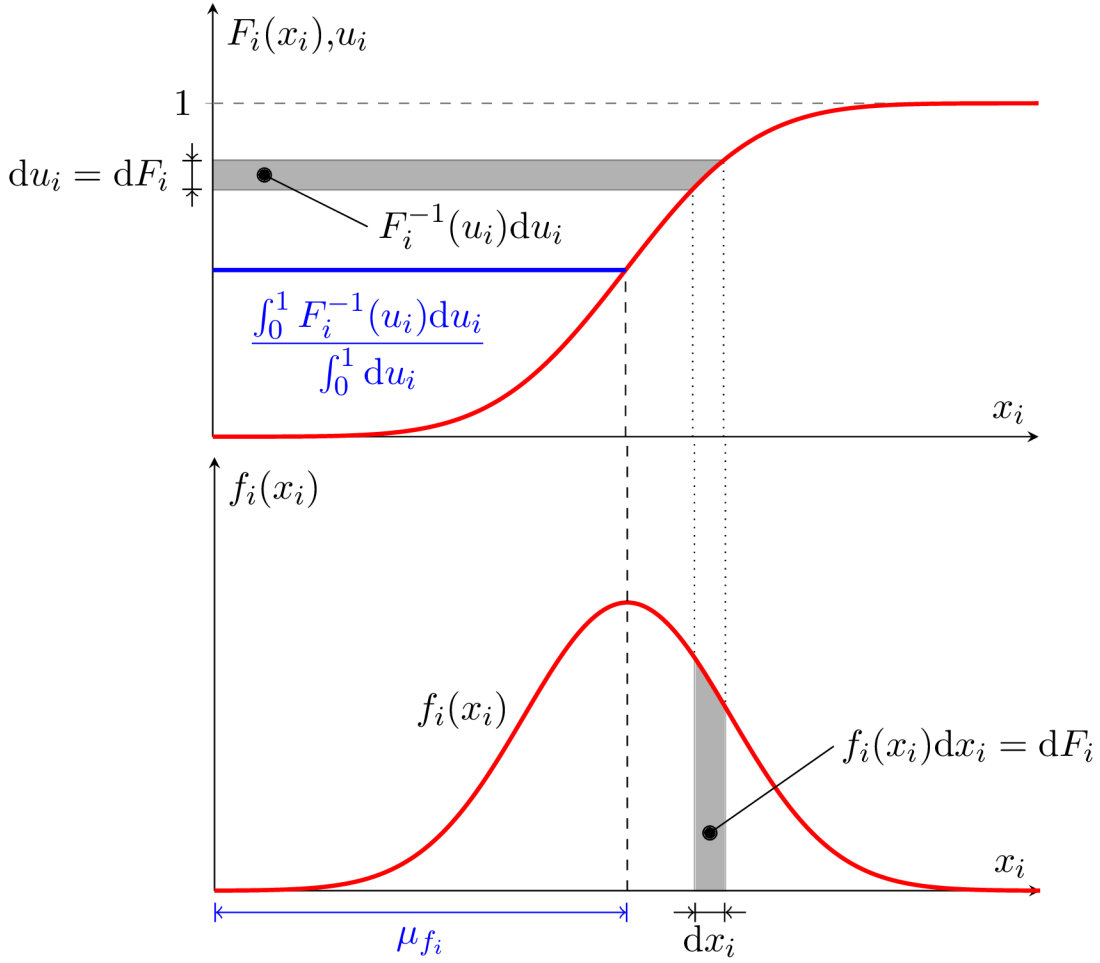
Figure 1: Computation of the mean value μ_{f_i} of random variable X_i.

The result of the Monte Carlo approximation using N_{sim} points, the following statistics (average) is obtained to estimate the *expected value* $E[g(X)]$:

$$\mu_z = \mu(g(\boldsymbol{X})) = E[Z] \approx \frac{1}{N_{\text{sim}}} \sum_{i=1}^{N_{\text{sim}}} g\left(F_1^{-1}(u_1), \ldots, F_{N_{\text{var}}}^{-1}(u_{N_{\text{var}}})\right). \quad (8)$$

It should be noted that apart from linear functions, one cannot substitute $E[g(X)]$ for $g(E[X])$.

1.1.1   Approximation of failure probability

Approximation of the failure probability or failure rate using Monte Carlo is similar to the discussed definite integral. The probability of failure, p_{f}, of the studied system $g(\boldsymbol{X})$ is essentially equal to the volume of the probability domain

$\boldsymbol{U} = (U_1, \ldots, U_{N_{\mathrm{var}}})$ where values of the function $g\left(F_1^{-1}(u_1), \ldots, F_{N_{\mathrm{var}}}^{-1}(u_{N_{\mathrm{var}}})\right)$ signalize a *failure event*. Similarly to the approximation of a definite integral, a kind of Indicator function:

$$I\left[g\left(F_{\boldsymbol{X}}^{-1}(\boldsymbol{u})\right)\right] = \begin{cases} 1 & \text{if } g\left(F_{\boldsymbol{X}}^{-1}(\boldsymbol{u})\right) \Rightarrow \text{failure event} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

can be integrated over U for an exact solution of p_{f}:

$$p_{\mathrm{f}} = \int\limits_0^1 \cdots \int\limits_0^1 I\left[g\left(F_{\boldsymbol{X}}^{-1}(\boldsymbol{u})\right)\right] \, \mathrm{d}u_1 \ldots \mathrm{d}u_{N_{\mathrm{var}}} = \iiint\limits_{D_{\mathrm{f}}} \mathrm{d}u_1 \ldots \mathrm{d}u_{N_{\mathrm{var}}}, \tag{10}$$

where D_{f} is the failure domain in the unit hypercube. Finally, the approximation in the sense of Monte Carlo can be computed as follows:

$$p_{\mathrm{f}} \approx \frac{1}{N_{\mathrm{sim}}} \sum_{\mathrm{sim}=1}^{N_{\mathrm{sim}}} I\left[g\left(F_{\boldsymbol{X}}^{-1}(\boldsymbol{u}_{\mathrm{sim}})\right)\right]. \tag{11}$$

2   General requirements on point samples

Even from such a brief introduction to the Monte Carlo method, one can perceive that the distribution of sampling points across the design domain is a crucial factor that influences both convergence and error of Monte Carlo-type approximation. A recognized term that emphasizes the need for *well distributed* sampling points in the design domain is the Koksma-Hlawka inequality [5, 6]:

$$\left| \frac{1}{N_{\mathrm{sim}}} \sum_{i=0}^{N_{\mathrm{sim}}-1} g(u_i) - \int_U g(u) \, \mathrm{d}u \right| \leq V(g) \, D_{N_{\mathrm{sim}}}^*(\mathcal{D}). \tag{12}$$

The Koksma-Hlawka inequality sets an upper bound on the error of solution approximation (left hand side) as a product of two independent terms: (i) $V(g)$, which depends on the variation (fluctuation) of the function $g(U)$, see [6], and (ii) $D_{N_{\mathrm{sim}}}^*(\mathcal{D})$ which is the measure of *discrepancy* (i.e. non-uniformity) of the set of sampling points, $\mathcal{D}$.

As is assumed, the function $g(U)$ that represents e.g. a numerical model is considered to be unknown or, at least, beyond control. Therefore, according to Equation (12), the only remaining way how to decrease the approximation error (*reduce the variance*) is to ensure that the layout of sampling points is as *uniform* as possible.

2.1   Distance-based criteria of uniformity

An inherent advantage of the distance-based criteria against discrepancy is that these may provide more information about the topology of the sample. For example, one can study the distribution of pair-wise distances or seek for points that violate a particular criterion the most. If identified, such points or pairs of points then might be a subject of a conscious optimization effort.

One of natural, visual properties of a "uniformly" distributed set of points is that no pair of points should be situated overly close to each other. Therefore positions of the points that are responsible for the minimal mutual distance $L_{\min}$ should be set farther apart to *maximize* this *minimum distance*. This very idea has been proposed by the Maximin criterion [7].

An inverse notion to the Maximin criterion is the Minimax, also see [7]. It employs the requirement that no pair of points should be overly distant from each other. That way, two points that are responsible for the maximal distance $L_{\max}$ should be brought closer together to *minimize* this *maximum distance.*

Instead of considering only the extremal mutual distances $L_{\min}$ and $L_{\max}$, certain distance-based criteria propose to take into account all pair-wise distances. An example of such a family of criteria is the ϕ_p criterion, see [8]. The ϕ_p criterion proposes a scalar-valued criterion function to rank designs of sampling points:

$$\phi_p = \left(\sum_{i=1}^{N_{\text{sim}}-1} \sum_{j=i+1}^{N_{\text{sim}}} \frac{1}{d^p\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)} \right)^{\frac{1}{p}}, \tag{13}$$

where the exponent p upon the metric $d\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)$ is considered as an arbitrary positive integer. The authors propose to use the Manhattan or Euclidean metrics. The ϕ_p criterion proposes to minimize the term (13) to reach an optimal design. The advantage of the criteria of ϕ_p family is that the scalar value of the criterion depends on all mutual distances of points while, in a way, retaining the property of the Maximin criterion:

$$\lim_{d^p\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) \to 0} \left(\phi_p\right) = \infty. \tag{14}$$

Further, with rising the exponent p, the contribution of pairs of closest points will become increasingly dominant. In the limit case of $p = \infty$, the designs that minimize the ϕ_∞ criterion are the designs obtained by the Maximin criterion.

Already in 1977, a predecessor criterion to the ϕ_p was proposed by Audze and Eglajs [9]. The Audze-Eglajs criterion considers the Euclidean distance metric and a particular value of the exponent $p = 2$. Not only that the Audze-Eglajs criterion is a part of the ϕ_p family, it also proposes a remarkable notion of a physical analogy between a set of sampling points and a set of charged, *mutually repelling particles*, see Figure 2 for illustration.
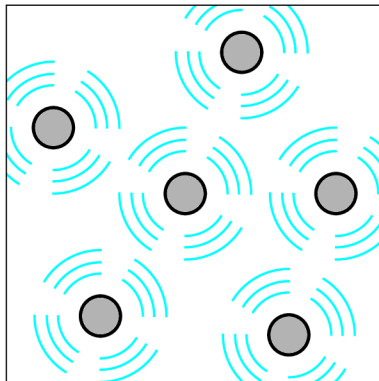

Figure 2: A system of mutually repelling particles within a hypercube $[0, 1]^2$.

2.1.1   Optimization criteria combined with LHS

The LHS method [10, 11] is often being paired with various optimization criteria in pursuit of a subsequent reduction of its variance and/or improvement of convergence of the estimator, if biased. The simplest approach possible is to use a selected optimization criterion only as an objective function when evaluating randomly generated LHS samples. A typical example of such an "optimization" is the `lhsdesign()` function used in the MATLAB software [12] that generates a set of LHS samples and, according to a user-specified parameter, returns the sample with either (i) the lowest correlation or (ii) the best value of the Maximin criterion. In the specific, although much used, case of Matlab, one should proceed with caution. Matlab merely selects among N_t (user-defined) randomly permuted designs. This approach renders the actual optimization very inefficient, if not entirely unused, see also in [13].

For purposes of an actual optimization, LHS can be suitably coupled with heuristic optimization methods. The changes of positions of points within the sample are typically done by mutual swapping of coordinates between a pair

of points that have been randomly selected or identified by the criterion. For instance, minimization of sample correlation has been proposed [13] by swapping of coordinates of randomly selected points, controlled by the algorithm of simulated annealing [14]. Similar strategy of swapping of LHS coordinates governed by simulated annealing has been suggested in [15] by utilizing the Audze-Eglajs criterion. There, the advantage that the Audze-Eglajs criterion (and the entire family of ϕ_p-criteria, respectively) is able to identify the pair of points that violate the criterion the most has been utilized for choosing the coordinates to be swapped. Also, the notion of using periodically repeated design domain first appeared in [16, 15]. An optimization of LHS samples using simulated annealing combined with Maximin or ϕ_p-criterion in [8]. However, it was proposed only to swap randomly selected pairs of coordinates.

3 Refinement of the ϕ-criterion

It has been mentioned above that the family of ϕ_p optimization criteria does offer the advantageous property of evaluating a sample by investigating mutual distances between all pairs of points. That way, it is able to provide the coupled optimization algorithm not only with the scalar value that evaluates the uniformity of the sample as a whole but also with contribution of each pair of points. The Audze-Eglajs criterion proceeds even further, as it proposes to understand its value as the amount of potential energy stored within a system of charged, mutually repelling particles:

$$E^{AE} = \sum_{i=1}^{N_{\text{sim}}-1} \sum_{j=i+1}^{N_{\text{sim}}} \frac{1}{L_{ij}^2}. \tag{15}$$

To represent potential energy, the Audze-Eglajs criterion omits the exponent of $1/p$ above the whole sum, compare Equations (13) and (15).

During the recent years, it has been shown that the Audze-Eglajs criterion suffers from existence of boundaries of the design space [15, 16]. A remedy of this behavior was proposed [15, 16], assuming periodically extended design hypercube and thus achieving a design domain without boundaries, see Figure 4b. Since then, it has been proved that optimization of point layouts by the introduced Periodic Audze-Eglajs (PAE) criterion combined

with LHS switching yields statistically uniform designs (from design to design) and to well distributed set of points in each single point layout, especially in two-dimensional design domain. For higher dimensions, corrupted designs have been observed. Their relatively good uniformity has actually been due to LHS that simply does not allow the malfunctioning criterion to emerge in full effect.

3.1 Periodic extension of design domain

The ϕ_p criterion proposes to use the Euclidean distance metric. This metric is arguably much desired for its property of directional independence, or isotropy. Let us assume that the points i and j with their mutual intersite distance L_{ij} are repelled by the force F_{ij} induced by the potential energy E_{ij}:

$$E_{ij}(L_{ij}) = \frac{1}{L_{ij}^p} = \int_{\infty}^{L_{ij}} F_{ij}(x)\mathrm{d}x. \tag{16}$$

By differentiating the energy potential with respect to the distance, L_{ij}, the *repulsive* force is obtained, also see in Figure 3. Neglecting the constant coefficient, the repulsive force is proportional to:

$$F_{ij}(L_{ij}) \propto \frac{1}{L_{ij}^{p+1}}. \tag{17}$$

Equation (17) can be understood as the constitutive law governing the interaction of particles.

As the particle system contains N_{sim} interacting particles, the total potential energy of the system is a sum of contributions from all $\binom{N_{\mathrm{sim}}}{2}$ individual pairs:

$$E^{AE} = \sum_{i=1}^{N_{\mathrm{sim}}-1} \sum_{j=i+1}^{N_{\mathrm{sim}}} E_{ij} = \sum_{i=1}^{N_{\mathrm{sim}}-1} \sum_{j=i+1}^{N_{\mathrm{sim}}} \frac{1}{L_{ij}^p}. \tag{18}$$

The total potential energy in Equation (18) represents the value of the Audze-Eglajs criterion to be minimized.

The Euclidean distance between points i and j in N_{var}-dimensional space, L_{ij}, can be expressed as a function of their coordinates:

$$L_{ij} = \sqrt{\sum_{v=1}^{N_{\mathrm{var}}} (x_{i,v} - x_{j,v})^2} = \sqrt{\sum_{v=1}^{N_{\mathrm{var}}} (\Delta_{ij,v})^2}, \tag{19}$$
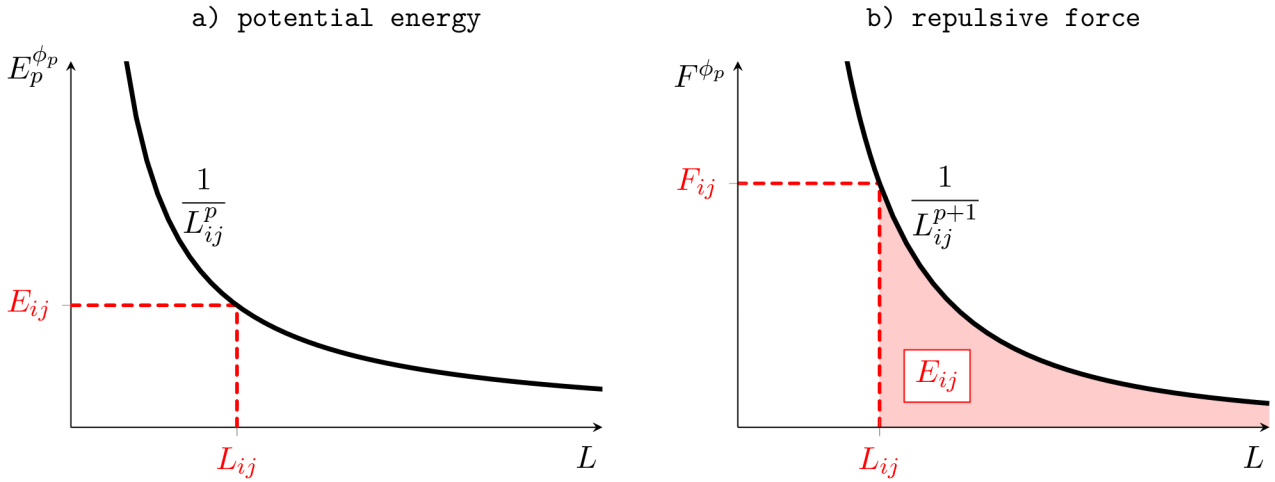
Figure 3: a) Potential energy of repulsive interaction between a pair of particles. b) Repulsive force acting on particles.

where $\Delta_{ij,v} = |x_{i,v} - x_{j,v}|$ is the difference in their positions projected onto the axis v. A simple and efficient improvement that considers a periodic extension of the design space has been proposed in [15]. After some simplification, one can derive equations for *periodic* Audze-Eglajs criterion (PAE) by replacing $\Delta_{ij,v}$ in Equation (19) with its periodic variant:

$$\overline{\Delta}_{ij,v} = \min(\Delta_{ij,v},\, 1 - \Delta_{ij,v}). \tag{20}$$

With such a redefined projection, a new metric is obtained and the distance between points i and j, called the periodic length $\overline{L}_{ij}$, becomes the actual shortest linear path between point i and the nearest image of point j [15], also see Figure 4:

$$\overline{L}_{ij} = \sqrt{\sum_{v=1}^{N_{\mathrm{var}}} \left(\overline{\Delta}_{ij,v}\right)^2}. \tag{21}$$

We note that using the nearest image of point j with respect to point i does not cover a true periodic repetition of the design domain. In a complete periodic repetition, an infinite number of images of point j would interact with point i. The presented approach is a simplification that has been shown in [15] to yield identical results to the fully repeated system in case of sufficient point count, N_{sim}. If the number of points in the original domain is too low for
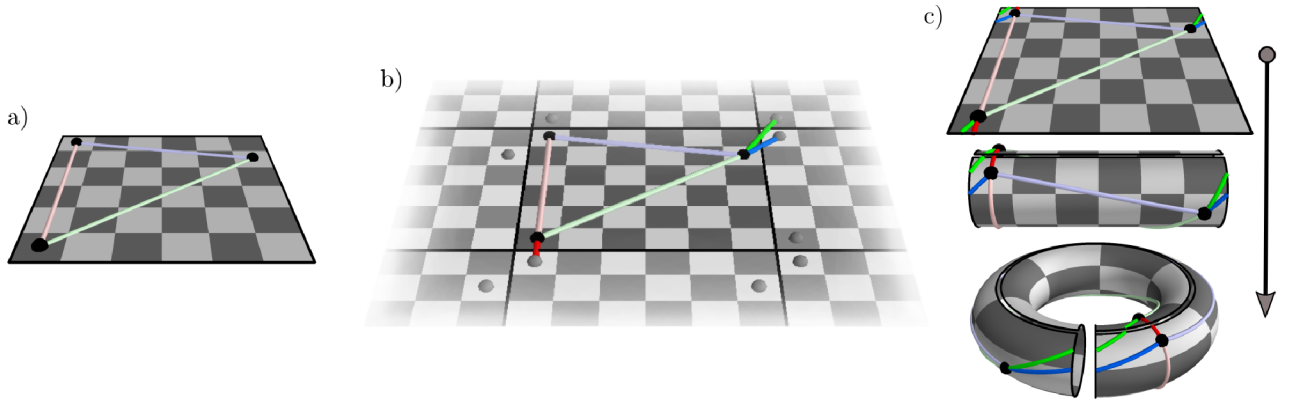
Figure 4: Illustration of periodically repeated planar domain. a) the original two-dimensional design domain with pale colored intersite distances. b) periodically repeated design domain with eight additional images of each particle. Periodic distances are rich colored. c) folding the design domain into a torus is another possible illustration of a periodical domain. Note that the computed distances are not defined on the toroidal surface.

assembly of the desired self-similar pattern[2], considering additional periodical images of particles is advised.

In this section we consider a generalized model in which a certain number of periodic repetitions of the original design domain are considered. Using the nearest image of point j with respect to point i, as considered in Equation (20), does not cover a true periodic repetition of the design domain. We argue that the above presented approach is a simplification that can be shown to yield identical results to the fully repeated system in case of sufficient point count N_{sim}. If the number of points in the original domain is too small to carry enough information about the pattern of a periodically repeated system, making a periodic extension to a sufficient level is desirable. In a true periodic domain, an infinite number of images of point j would interact with point i. When a finite number of copies of the design domain is considered, not only the real particle j, but also all periodically repeated images of the particle j

---

[2]Simplest self-similar space-filling patterns can be assembled from simplest objects which contain volume in the particular dimension N_{var}: line in 1D (2 points), triangle in 2D (3 points), tetrahedron in 3D (4 points), etc.

will contribute to the potential:

$$\phi_{\left(\overline{L},p,k_{\max}\right)} = \sum_{i=1}^{N_{\mathrm{sim}}-1} \sum_{j=i+1}^{N_{\mathrm{sim}}} \left(\frac{1}{\overline{L}_{ij}^{p}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)} + \sum_{k=1}^{k_{\max}} \sum_{c=1}^{c_{\max}} \frac{1}{L_{ij}^{p}\left(\boldsymbol{x}_i, \boldsymbol{x}_j + \mathbf{s}_c\right)} \right), \quad (22)$$

where $k_{\max}$, introduced as an additional parameter, is the number of added periodical extensions (envelopes) of the design space. In the fully repeated system $k_{\max} = \infty$ and analogically, for a non-extended system $k_{\max} = 0$. Therefore $\phi_{\left(\overline{L},p\right)} = \phi_{\left(\overline{L},p,0\right)}$.

A single "level" of periodic extension adds another envelope of periodically repeated images of all other particles around each point, see Figure 5 and through Figure 8. Such an extension does provide additional information about the point layout within the domain.
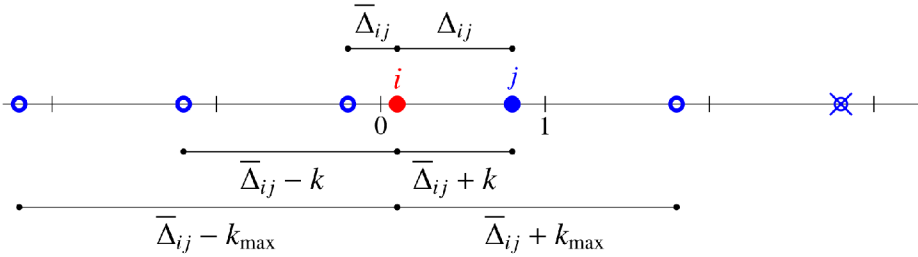

Figure 5: 1D example of the periodic extension of level $k_{\max} = 2$.

3.2   Selection of value of the distance exponent p

In the work, the power is derived to be at least $p \geq N_{\mathrm{var}} + 1$. With this power, the interaction is *dominated by short-range interactions*. With such a sufficient exponent p, the convergence of the potential energy $\phi_{(L,p)}$ or better $\phi_{\left(\overline{L},p\right)}$ towards infinity for a uniform distribution of points is a *power law*. Such a convergence signalizes self-similarity of the problem or absence of a length scale.

This can be shown by studying the behavior of the radial part of the integral of the potential over the volume V of N_{var}-dimensional domain. The potential energy for a uniform design reads:

$$I = \int_{N_{\mathrm{var}}} \frac{1}{L^p} \, \mathrm{d}^{N_{\mathrm{var}}} V, \quad (23)$$

where L is used to denote one-dimensional distance between points (the symbol d is not used to avoid confusion with the symbol d for the differential). Transforming this into polar coordinated gives:

$$I = \int_{N_{\mathrm{var}}} \varphi \; \mathrm{d}^{N_{\mathrm{var}}} V \; |J| \; \frac{1}{L^p} \; \mathrm{d}L, \tag{24}$$

where $|J|$ is the Jacobian. Therefore, the integral is performed over the product $L^{N_{\mathrm{var}}-1-i}$. Performing just the radial integration leads to:

$$I_r = \int \frac{L^{N_{\mathrm{var}}-1}}{L^p} \mathrm{d}L = \int L^{N_{\mathrm{var}}-1-p} \mathrm{d}L. \tag{25}$$

For $p = 2$ as used in the AE criterion, we get the behavior described above (i.e. $I \propto 1/L^2$). Using $p = N_{\mathrm{var}}$ leads to:

$$I_r = \int L^{-1} \mathrm{d}L = \ln(L), \tag{26}$$

which diverges logarithmically and the interaction is still long-ranged. Using $p = N_{\mathrm{var}} + 1$ yields

$$I_r = \int L^{-2} \; \mathrm{d}L = \frac{1}{L}, \tag{27}$$

which has the desired asymptotic behavior dominated by short-ranged interactions. Using powers $p > N_{\mathrm{var}} + 1$ only increases the (asymptotically constant) ratio between short-range and long-range interactions.

Figure 6 shows the convergence of the normalized potential energy $\phi_{(\overline{L},p)}$ with rise of the number of particles, N_{sim}. Instead of presenting the results for the point count, N_{sim}, we introduce a variable ℓ_{char}, the *characteristic length* that involves also the dimension of the space and therefore expresses the *saturation of the design domain* with integration points (particles). The characteristic length is defined as:

$$\ell_{\mathrm{char}} = \frac{1}{\sqrt[N_{\mathrm{var}}]{N_{\mathrm{sim}}}}. \tag{28}$$

The geometrical meaning of the characteristic length can be illustrated using a regular orthogonal grid of points in N_{var}-dimensional hypercube $[0,1]^{N_{\mathrm{var}}}$. The characteristic length represents the smallest distance between points within
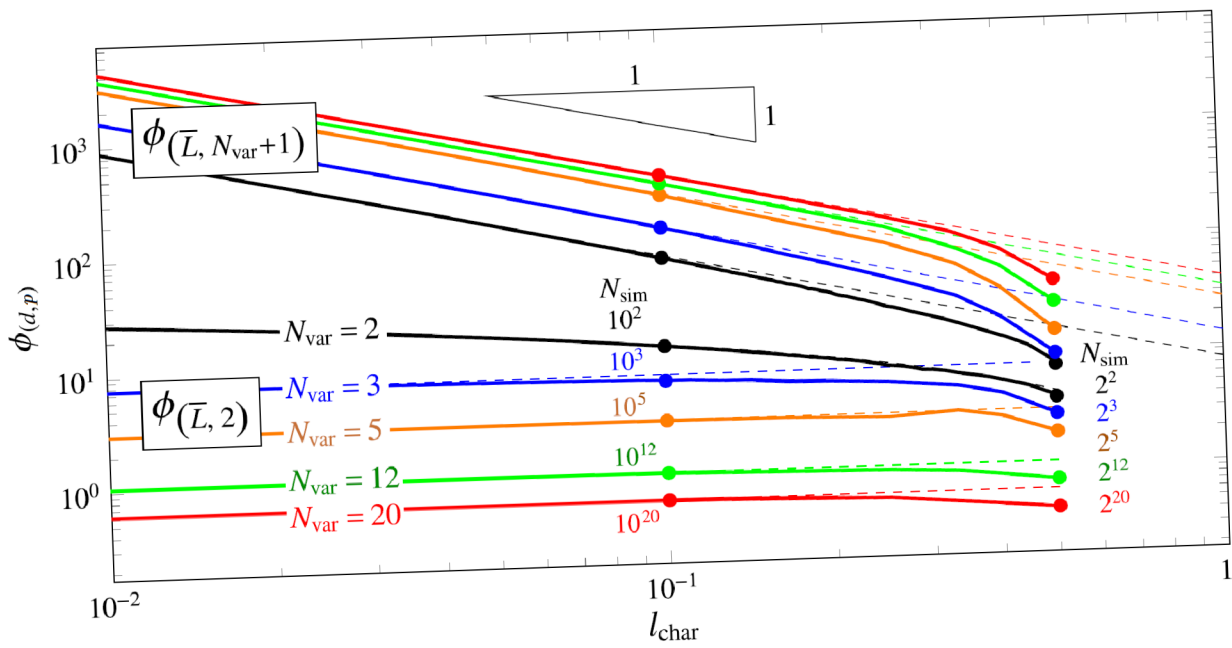
Figure 6: Convergence of the normalized potential energy $\phi_{(\overline{L},p)}$ depending on the exponent p.
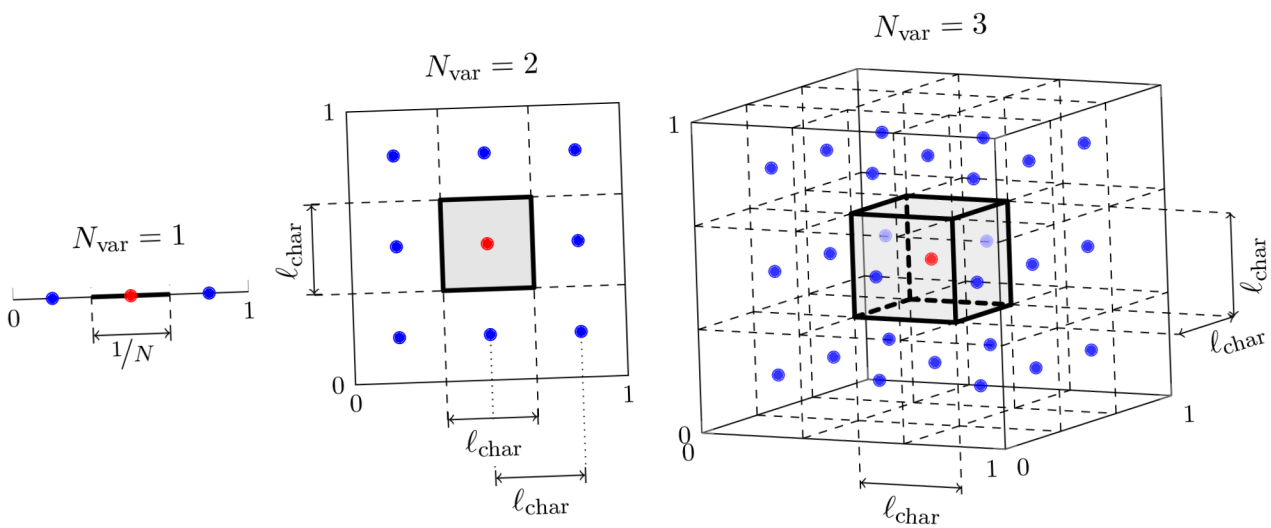

Figure 7: Geometrical illustration of the characteristic length, ℓ_{char}.

the grid. If N is the number of points over each dimension, the characteristic length can be derived as follows:

$$\ell_{\text{char}} = \frac{1}{N} = \left| N_{\text{sim}} = N^{N_{\text{var}}} \right| = \frac{1}{\sqrt[N_{\text{var}}]{N_{\text{sim}}}}. \tag{29}$$

At the same time, ℓ_{char} is the characteristic side length of the hypercubical volume belonging to each point in the space of sampling probabilities ($\ell_{\text{char}}^{N_{\text{var}}} = dU = 1/N_{\text{sim}}$), see Figure 7.

It can be seen that with the original exponent value $p = 2$ in the dimension $N_{\text{var}} = 2$ ($p = N_{\text{var}}$), the potential energy of the system does not converge to a power law but diverges logarithmically, roughly:

$$\phi_{\left(\overline{L},p\right)} \approx \pi \ln(N_{\text{sim}}) + \underbrace{\frac{1}{\sqrt{N_{\text{sim}}}} - \frac{1}{N_{\text{sim}}}}_{\to 0}. \tag{30}$$

In higher dimensions $N_{\text{var}} \geq 3$, the exponent $p = 2$ further leads to convergence of the potential energy to a constant, see [17].

Using the above proposed exponent $p = N_{\text{var}} + 1$, the potential energy value tends to a power law as $N_{\text{sim}} \to \infty$. Note the universality here: the quality of the criterion does not depend on sample size, N_{sim}, nor on the dimension, N_{var}:

$$\phi_{\left(\overline{L},N_{\text{var}}+1\right)} \propto \frac{1}{\ell_{\text{char}}}. \tag{31}$$

Such a behavior is desired as the designs for a given dimension, N_{var}, tend to have a universal self-similar pattern and the dependence on sample size disappears (no length scale is present). Thus the character of the criterion is kept independent of N_{sim} and the proportion between short-range interactions and long-range interactions is constant. This stabilization is obtained for a sufficient number of points within the design domain (a kind of N_{var}-dimensional tile). The self-similarity manifested is by the power law dependence (a straight line in Figure 6). When the exponent is taken even higher ($p > N_{\text{var}} + 1$), the self-similar regime is achieved for even smaller number of points (greater ℓ_{char}).

Graphs in Figure 6 suggest that there must be link between (a) the exponent (responsible for the proportion between long- and short-range interactions) and, (b) the number of "dummy" copies of the design domain that also modify the proportions.
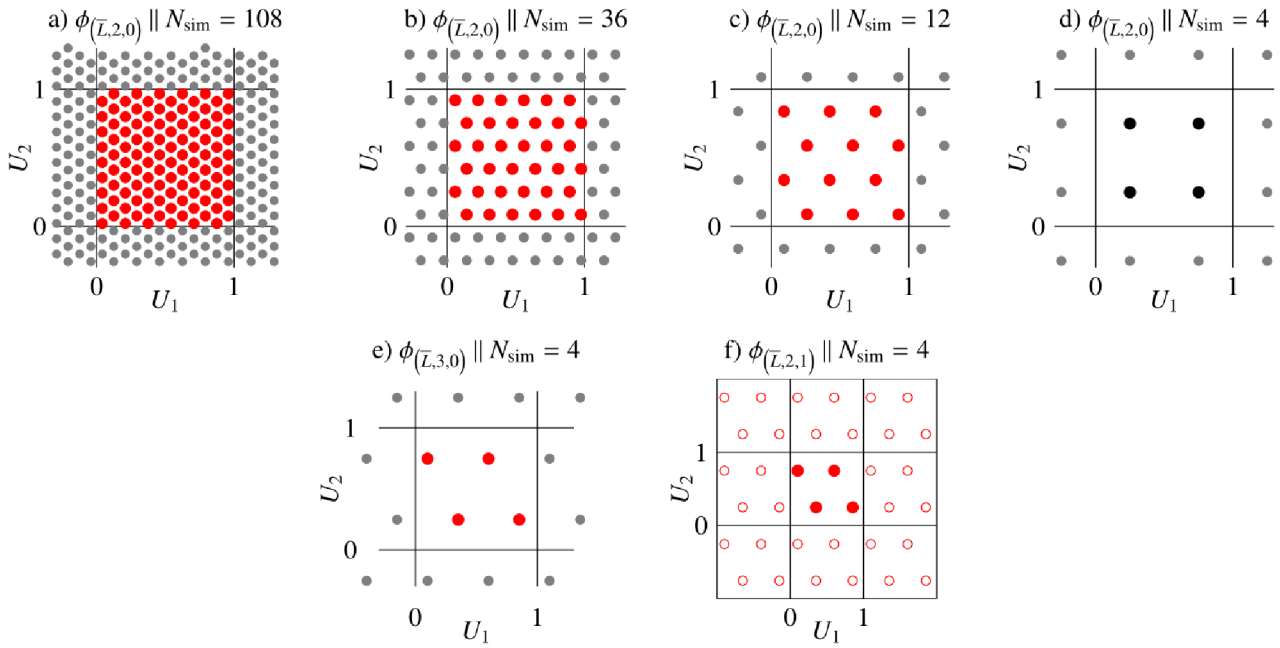
Figure 8: Disappearing of the quality of a self-similar design and the effect of remedies proposed.

4   Physical analogy: a particle system

The Audze-Eglajs criterion, and its generalization, the ϕ_p criterion, propose to understand the layout of design points as a system of interacting (mutually repelling) particles and evaluates the amount of potential energy stored within all interactions, see Figure 4 for illustration.

Instead of utilizing the ϕ_p criterion as a norm minimized using combinatorial or heuristic optimization for a fixed set of LHS coordinates [13], this work proposes to solve the physically analogical problem by simulating a discrete dynamical system of mutually repelling particles, recall Figure 2. The coordinates of particles of the dynamical system, after reaching the static equilibrium, may be directly understood as coordinates of the design points, see Figure 9.

Essential remarks about derivation using Lagrangian mechanics are provided in what follows. The general value of exponent p will be utilized for derivation of *equations of motion* of the dynamical particle system. Through the following content, mind also Figure 10 that illustrates the interaction of two particles i and j.

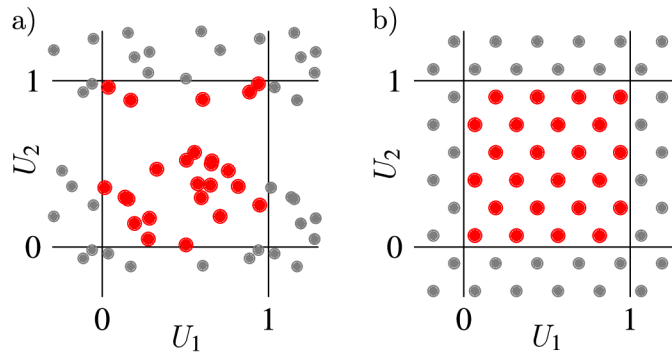To begin, let us state that the dynamical behavior of a mechanical system

Figure 9: Illustration of the optimization process of $N_{\text{sim}}=24$, $N_{\text{var}}=2$. a) initial randomized sample, b) optimized sample.

with a finite number of degrees of freedom can be described by the Lagrange's function, or shortly Lagrangian, $\mathcal{L}$. Sometimes also called a *kinetic potential*, the Lagrangian $\mathcal{L}$ is a functional; a sum of formulations of kinetic and potential energy. In case of the ϕ_p-conditioned dynamical system, the Lagrangian can be described as follows:

$$\mathcal{L} = E_k + E_p, \tag{32}$$

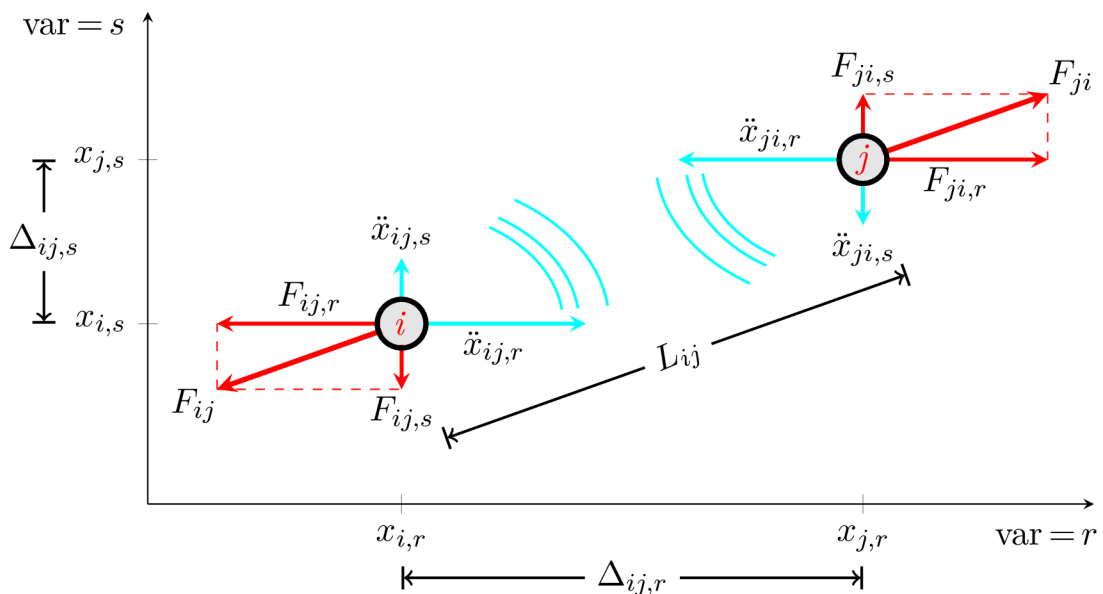with the kinetic energy of the particle system E_k being a simple sum of kinetic

Figure 10: Interaction of two charged particles.

energies of all particles of equal mass m:

$$E_k = \frac{1}{2} m \sum_{i=1}^{N_{\mathrm{sim}}} \sum_{v=1}^{N_{\mathrm{var}}} \dot{x}_{i,v}^2, \tag{33}$$

where $\dot{x}_{i,v} = \frac{\mathrm{d}}{\mathrm{d}t} x_{i,v}$ is the velocity of ith particle in dimension v.

We now consider a generalized formulation of the potential energy, E_p, employing an arbitrary value of the exponent, p, upon the mutual distances, $\overline{L}_{ij}$. The potential energy of the model can be written as a sum of energies stored within all mutual inter-particle interactions:

$$E_p = \sum_{i=1}^{N_{\mathrm{sim}}-1} \sum_{j=i+1}^{N_{\mathrm{sim}}} \frac{1}{\overline{L}_{ij}^p}, \tag{34}$$

where the exponent is now considered as a general integer, p, (similarly to the ϕ-criterion [8]) and the metric considered is the periodic length, $\overline{L}_{ij}$, see Equation (21).

Further, it is needed to calculate the derivatives of Lagrangian $\mathcal{L}$ with respect to all state variables. In the case at hand, the state variables are the coordinates $x_{i,v}$ and velocities $\dot{x}_{i,v}$ of all particles in each dimension. Obeying the Lagrange's equations of the second kind:

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial \mathcal{L}}{\partial \dot{x}_{i,v}} \right) = \frac{\partial \mathcal{L}}{\partial x_{i,v}}, \tag{35}$$

one can start off with the assumption that, apart from the derivatives with respect to the time t, the kinetic energy E_k is further differentiable only with respect to velocities $\dot{x}_{i,v}$ and the potential energy E_p (34) is differentiable only with respect to coordinates $x_{i,v}$. Therefore, the left-hand side of Equation (35) is rather easily obtainable:

$$\frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial \mathcal{L}}{\partial \dot{x}_{i,v}} \right) = \frac{\mathrm{d}}{\mathrm{d}t} \left(\frac{\partial E_k}{\partial \dot{x}_{i,v}} \right) = m\, \ddot{x}_{i,v}, \tag{36}$$

with $\ddot{x}_{i,v} = \frac{\mathrm{d}}{\mathrm{d}t} \dot{x}_{i,v}$ being the acceleration of the ith particle in the dimension v. The right-hand side of Equation (35) becomes:

$$\frac{\partial \mathcal{L}}{\partial x_{i,v}} = \frac{\partial E_p}{\partial x_{i,v}} = \sum_{j \neq i}^{N_{\mathrm{sim}}} \left(\frac{1}{\overline{L}_{ij}^{p+1}} \frac{\overline{\Delta}_{ij,v}}{\overline{L}_{ij}} \right). \tag{37}$$

The resulting equation of motion of ith particle in vth dimension as assembled from Equations (35), (36) and (37) finally reads:

$$\ddot{x}_{i,v} = \frac{1}{m} \sum_{\substack{j=1 \\ j \neq i}}^{N_{\text{sim}}} \frac{\overline{\Delta}_{ij,v}}{\overline{L}_{ij}^{p+2}}. \tag{38}$$

Note that these are equations of motion of a conservative dynamical system as defined by the energy potential (32) which does not cover any form of energy dissipation.

The motion of the dynamical system is therefore described by a system of independent equations. This awareness is of high importance while considering the possibilities for solution method and its computer implementation. It means that accelerations $\ddot{x}_{i,v}$ of each particle in each dimension can be solved separately, without solving a system of equations.

5   Performance of the parallelized solution

Due to the heavy computational demands of such a hyper-dimensional particle system, it was necessary to develop an efficient, massively parallel solution using a GPGPU platform. Specifically, the Nvidia CUDA archiecture has been used. Due to the essentially arbitrary extent (number of particles N_{sim} and dimensions N_{var}) of the particle system at hand, two main implementation branches have been developed: (i) a more restricted solution utilizing the fast but limited on-chip memory and (ii) a more general (safer) solution using strictly the global memory of the GPU. A great portion of the doctoral thesis is devoted to the actual implementation as well as to the CUDA platform itself. The extent of the GPGPU part is beyond the scope of this Ph.D. thesis, considering the implementation part as a tool for obtaining the optimized samples of interest.

The main message that needs to be conveyed here is that the developed massively parallel solution essentially allowed feasible simulation of the particle model. Figure 11 shows the computational time required for solving 10^5 steps of the dynamical simulation in 2 dimensions ($N_{\text{var}} = 2$). The two **CUDA implementation**s are compared to the initial, single-thread **CPU implementation in C**. Clearly, a major solution speedup has been reached.
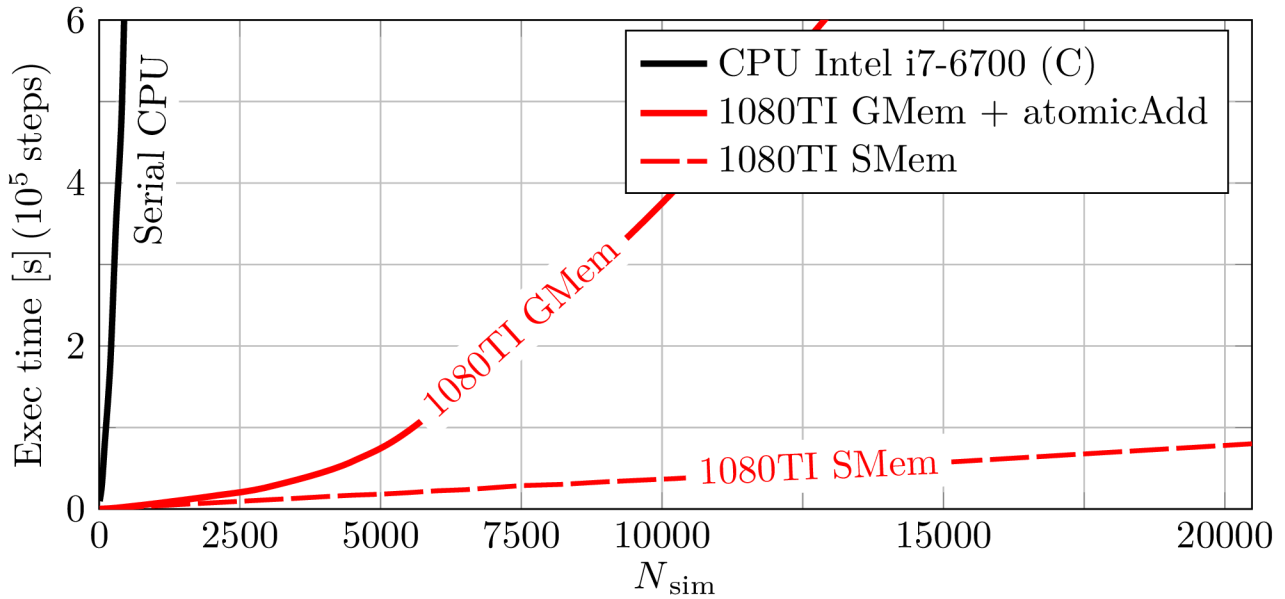
Figure 11: Performance comparison of the all-pairs $\mathcal{O}\left(N^2\right)$ solution while scaling N_{sim} in constant dimension of $N_{\text{var}} = 2$.

Next, we investigate the execution time when scaling both parameters of the problem; the number of particles, N_{sim}, as well as the number of dimensions, N_{var}. Figure 12a compares the execution time of solution of 10^3 steps of the $\mathcal{O}\left(N^2\right)$ interaction.

Figure 12a shows that the parallel execution starts with linear dependency on N_{sim}. As the device starts to reach its peak bandwidth, the execution time scaling tends to $\mathcal{O}(N^2)$. As shown in Figure 12b, in the expected range of computation, the speedup in solution time by CUDA grows linearly with the number of particles, reaching up to $250\times$.
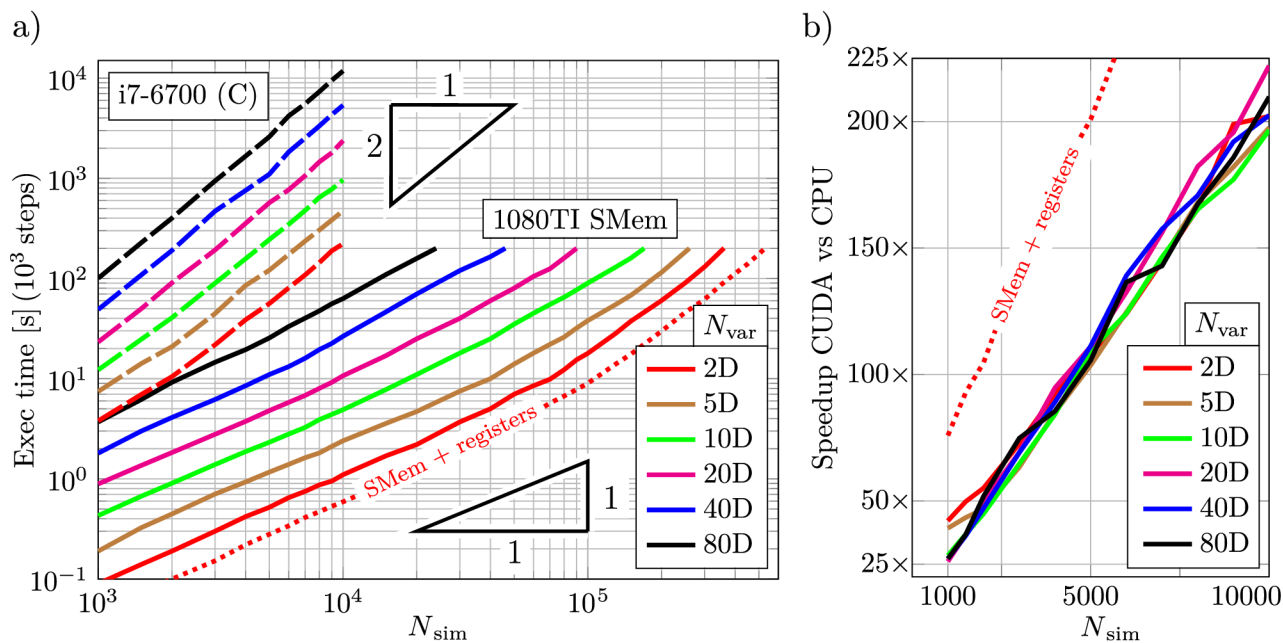
Figure 12: a) Performance comparison of the $\mathcal{O}\left(N^2\right)$ solution as executed by CPU (dashed) and GPU (solid and dotted lines) while scaling both N_{sim} and N_{var}. b) The achieved speedup of solution.

6   Numerical integration performance

In the work, the performance of the dynamically optimized samples was tested in various use-cases of Monte Carlo-type approximation. To demonstrate the characteristic properties of dynamically optimized samples and the refined ϕ_p criterion, respectively, these samples will be compared to a plain Monte Carlo integration, LHS-random samples and LHS samples optimized by the Periodic Audze-Eglajs criterion [15]. The nomenclature of the sampling methods used is summarized in Table 1.

Marking	Optimization method used
MC RAND	Random Monte Carlo samples
LHS RAND	LHS samples with a random order of points evaluated at strata median, see e.g. [13].
DYN ϕ_p Intersite	Optimized by the dynamical system using the refined ϕ_p potential, intersite Euclidean metric, bordered design domain
DYN ϕ_p Periodic	Optimized by the dynamical system using the refined ϕ_p potential, shortest-distance Euclidean metric, periodic design domain
DYN ϕ_p Periodic +LHS	DYN ϕ_p Periodic samples latinized after pre-rotation
LHS PAE	Samples optimized by mutual swapping of LHS coordinates governed by Periodic Audze-Eglajs criterion as proposed in [15].

Table 1: Marking of compared optimization methods.

It has shown as convenient to represent the results using a normalized scale of ℓ_{char} instead the actual number of integration points N_{sim} to express each domain is filled with integration points. For convenience, additional horizontal axis with the corresponding values of N_{sim} will be also provided. In the thesis, only very brief preview of the obtained results is shown.

6.1   Approximation of definite integral

An approximation of a definite integral is one benchmark problem in the work. Multiple variants of such a problem have been tested, primarily for demonstration of the effect of introducing periodic boundary conditions into the ϕ_p criterion in order to achieve statistically uniform samples (unbiased estimation).

One of such results is shown in Figure 13 that compares performance of selected sampling methods in estimation of the highlighted area. It is very clear that without using a periodically repeated design domain, an unbiased estimation cannot be obtained because the lack of statistical uniformity of samples. Furthermore, the dynamically optimized samples **DYN ϕ_p Periodic** bring a major reduction in variance of estimation. Due to the periodically repeated design domain, such samples exhibit statistical uniformity, i.e. their estimates are unbiased. The **DYN ϕ_p Intersite** samples optimized in a bounded domain tend to oversaturate boundaries of the domain and therefore are inherently biased.
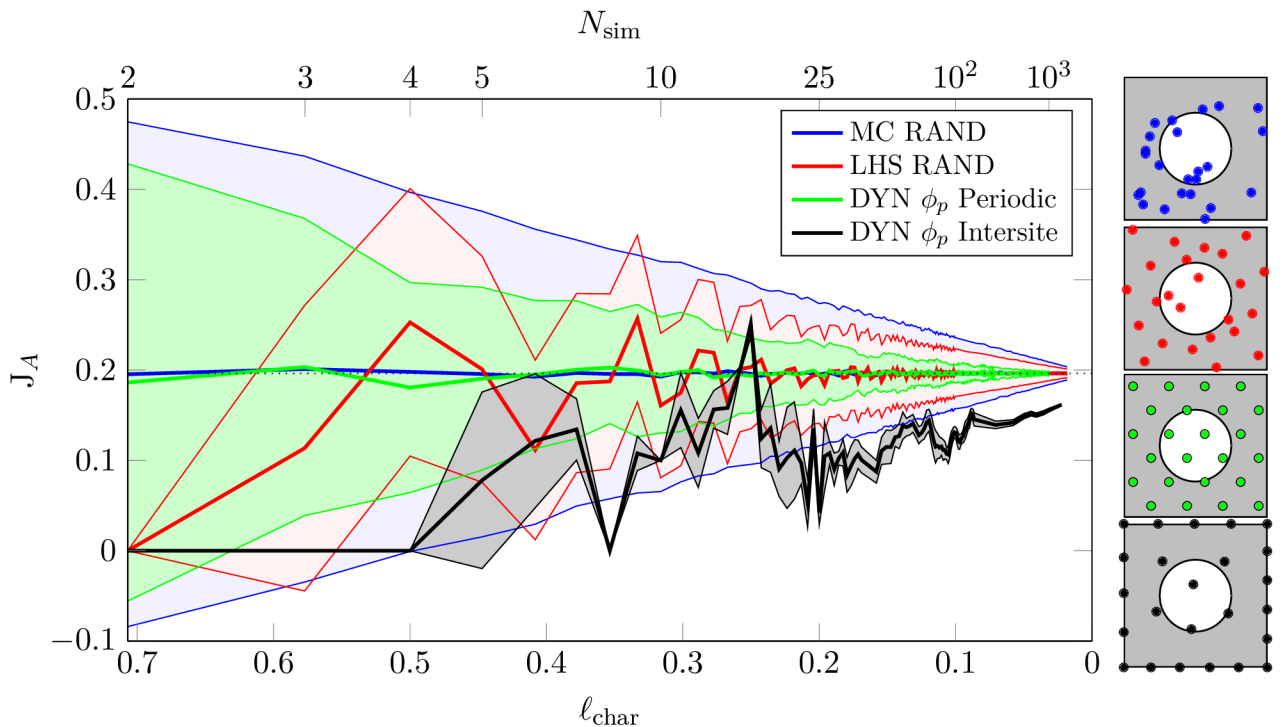

Figure 13: Estimated mean value of the circular area, J_A, within a unit square (ave, ave $\pm$ ssd).

6.2   Sum of exponentials of normal variables

One of the numerical examples in this work aims to represent an approximation of characteristics of a function of random variables. The selected problem has been featured in [18] and also in [15] for evaluating of performance of LHS samples optimized by switching governed by the Audze-Eglajs criterion that considers periodically repeated design domain (periodic Audze-Eglajs – PAE).

The studied example considers a function $g_{\exp}(\boldsymbol{X})$ of input random vector $\boldsymbol{X}$:

$$g_{\exp}(\boldsymbol{X}) = \sum_{v=1}^{N_{\mathrm{var}}} \exp\left(-X_v^2\right) = \sum_{v=1}^{N_{\mathrm{var}}} \exp\left(F_v^{-1}(u_v)^2\right). \tag{39}$$

The input random variables $X_1, \ldots, X_{N_{\mathrm{var}}}$ are considered as independent and identically distributed (IID) random variables with standard normal distribution $\mathcal{N}(0,1)$.

The numerical results of this example are presented for the case of $N_{\mathrm{var}} = 2$. Results of estimation of the mean value $\mu_{\exp}$ (see Figure 14), and the standard deviation $\sigma_{\exp}$ (see Figure 15), are presented.

Because the function $g_{\exp}(\boldsymbol{X})$ represents a sum of independent marginals, any deterministic LHS sample is going to estimate the mean value $\mu_{\exp}$ without any variance. In the two-dimensional case, the poor performance of **DYN** ϕ_p **Intersite samples** is also shown to remind the improvement of using periodically repeated design domain.

The DYN ϕ_p Periodic samples bring reduction in variance of the estimated values. In fact, using the **DYN** ϕ_p **Periodic samples** leads to enhanced rate of convergence. If using DYN ϕ_p Periodic samples the attained asymptotic rate of convergence is $\mathcal{O}(1/\sqrt{N_{\mathrm{sim}}^3})$, whereas **Monte Carlo samples** exhibit the expected $\mathcal{O}(1/\sqrt{N_{\mathrm{sim}}})$ rate of convergence.
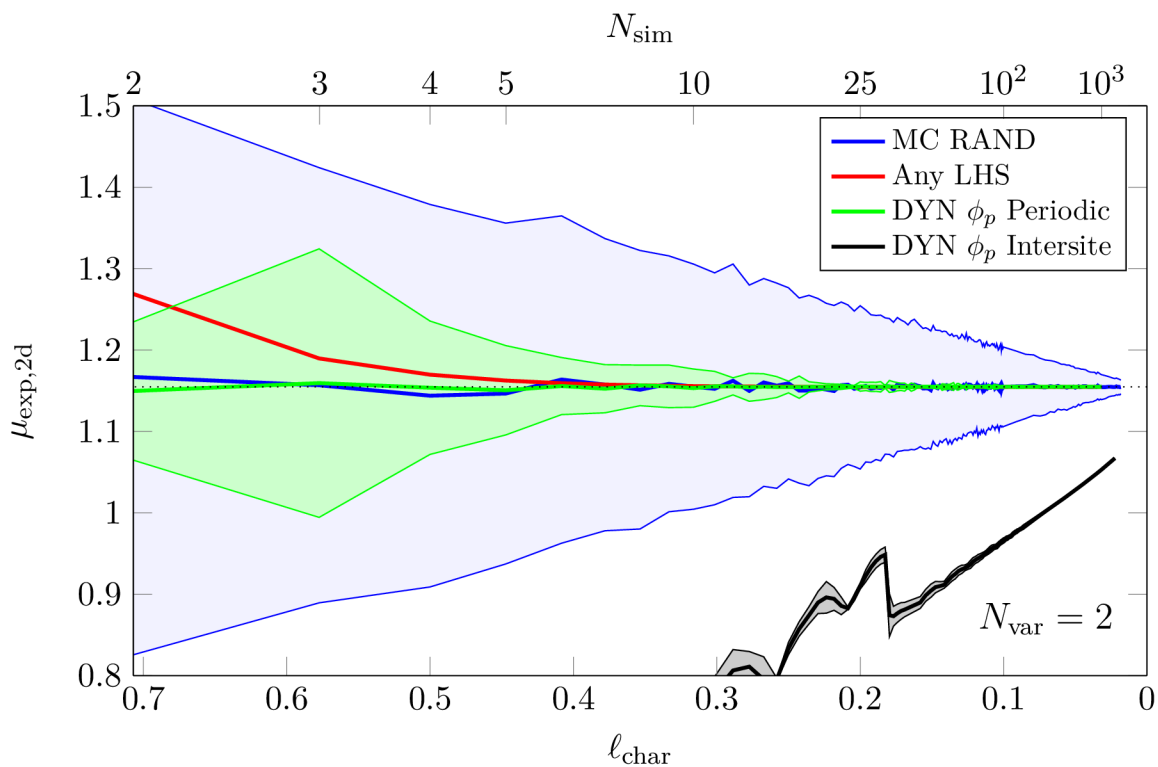
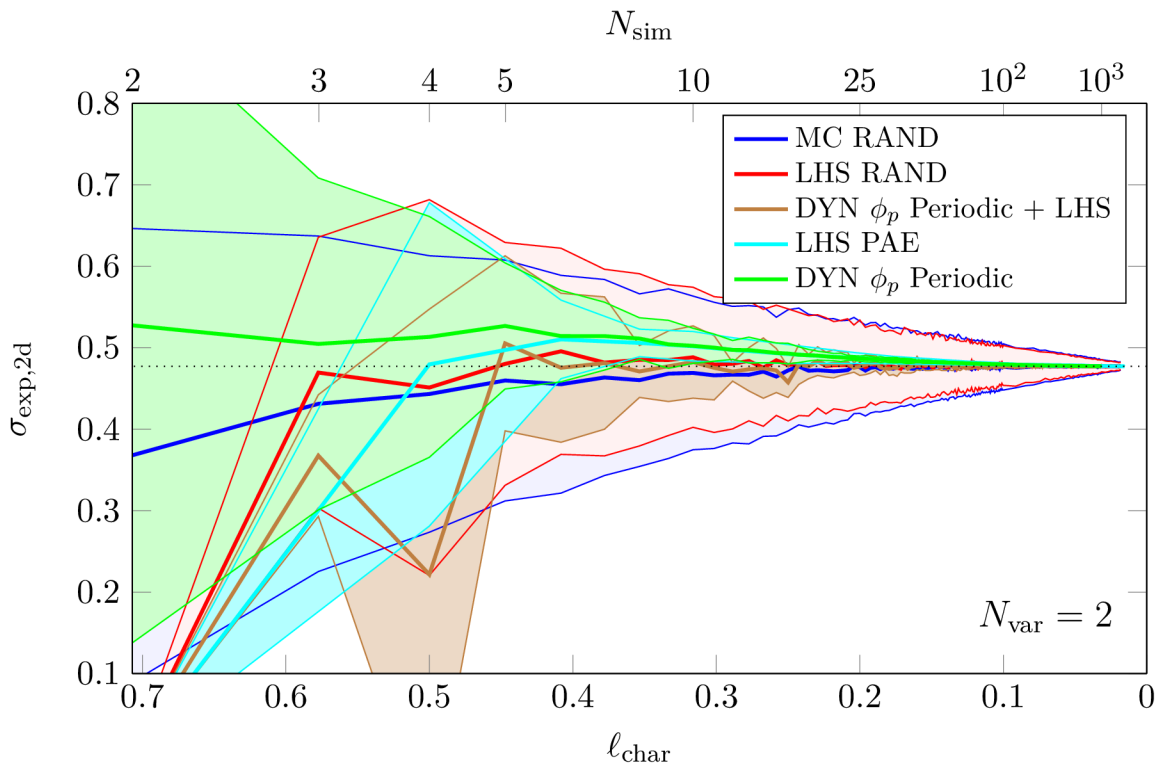Figure 14: Estimated mean value of $g_{\exp,\text{2d}}$ (ave, ave $\pm$ ssd).

Figure 15: Estimated standard deviation of $g_{\exp,\text{2d}}$ (ave, ave $\pm$ ssd).

6.3   Engineering example – Failure of a structure

Assume now an engineering problem of estimation of deflection and failure probability of a truss structure. The model is similar to the truss model used in [19] where it served as one of benchmark problems for comparison of methods for estimating structural reliability. The setup of the numerical experiment is illustrated in Figure 16.
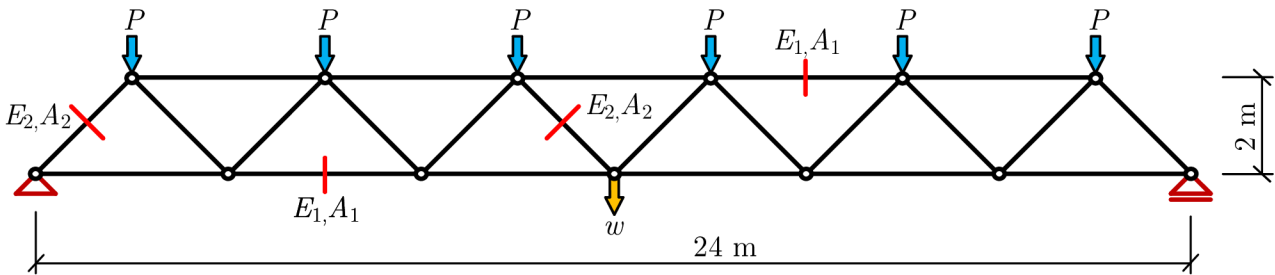

Figure 16: An illustration of the studied model of truss structure.

The input random vector of the model consists of five random variables: the material properties of horizontal members (Young's modulus E_1 and cross-section area A_1), the material properties of diagonal members (Young's modulus E_2 and cross-section area A_2) and the loading forces P in top joints. The properties of input random variables are summarized in Table 2. The task is to estimate the mean value and standard deviation of the midspan deflection w and the probability of failure, respectively. The failure is defined to occur when the deflection w exceeds the given value of $0.11\,\mathrm{m}$.

Random variable			Distribution	Mean value	c.o.v.
Young's modulus	E_1	[GPa]	Log-normal	210	0.1
Young's modulus	E_2	[GPa]	Log-normal	210	0.1
Cross-section area	A_1	[m^2]	Log-normal	2.0×10^{-3}	0.1
Cross-section area	A_2	[m^2]	Log-normal	1.0×10^{-3}	0.1
Load	P	[kN]	Gumbel	50	0.15

Table 2: Random variables of truss example and their properties.

This engineering example has been solved using multiple approaches to

estimation of failure probability. Along with the original 5d model, its reduced 3d equivalent model has been derived. Both variants have been also evaluated using the Importance sampling method. For a detailed analysis of the results, see the dissertation thesis. Only essential remarks will be provided here.

6.4   Comparison of approaches to estimation of p_f

For each individual sampling method, the convergence of estimation of failure probability is plotted as attained in the 5d space of the original problem and the benefit of using IS around the design point in 5d and the reduced 3d space.

Through the Figures 18, 19, 20, and 21, one can study the effect of reducing the dimension and IS for MC RAND, LHS RAND, and both versions of DYN ϕ_p samples. The following plots use a universal measure that is the number of integration points (particles), N_sim, which in practical sense represents the number of numerical experiments required.

It can be concluded that both variants of the dynamically optimized samples perform well also in this engineering example, yielding a rather significant decrease of required computing time. Compare especially the highlighted numbers of model evaluations required to attain 10% CoV in estimation of p_f.
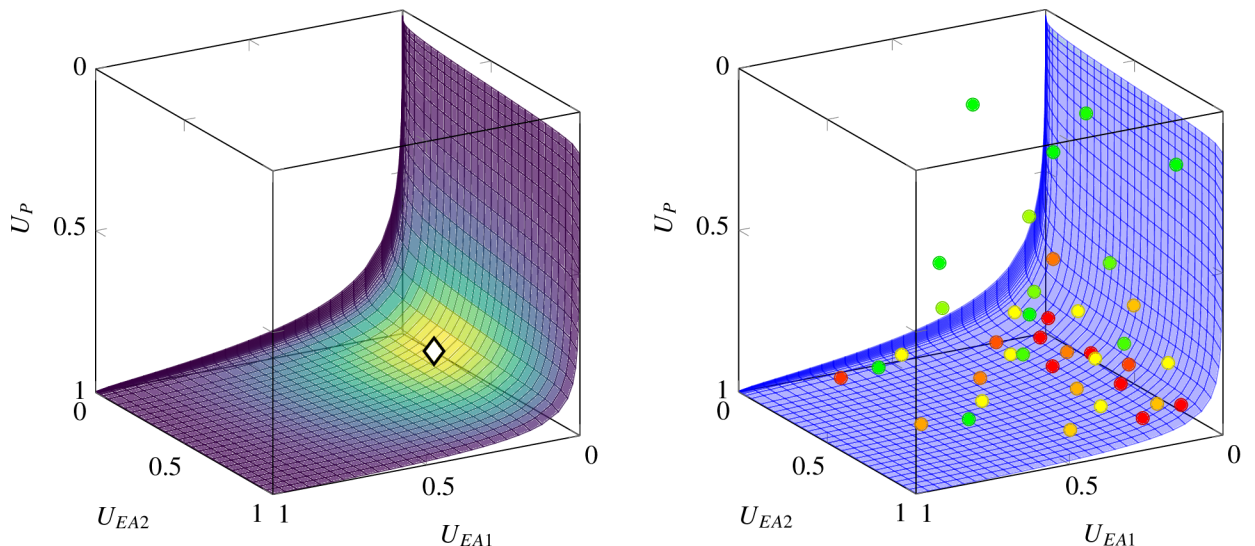

Figure 17: Left: The probability on the failure surface of the reduced 3d model and the position of the design point in the probability space. Right: Results of simulations in sampling points.
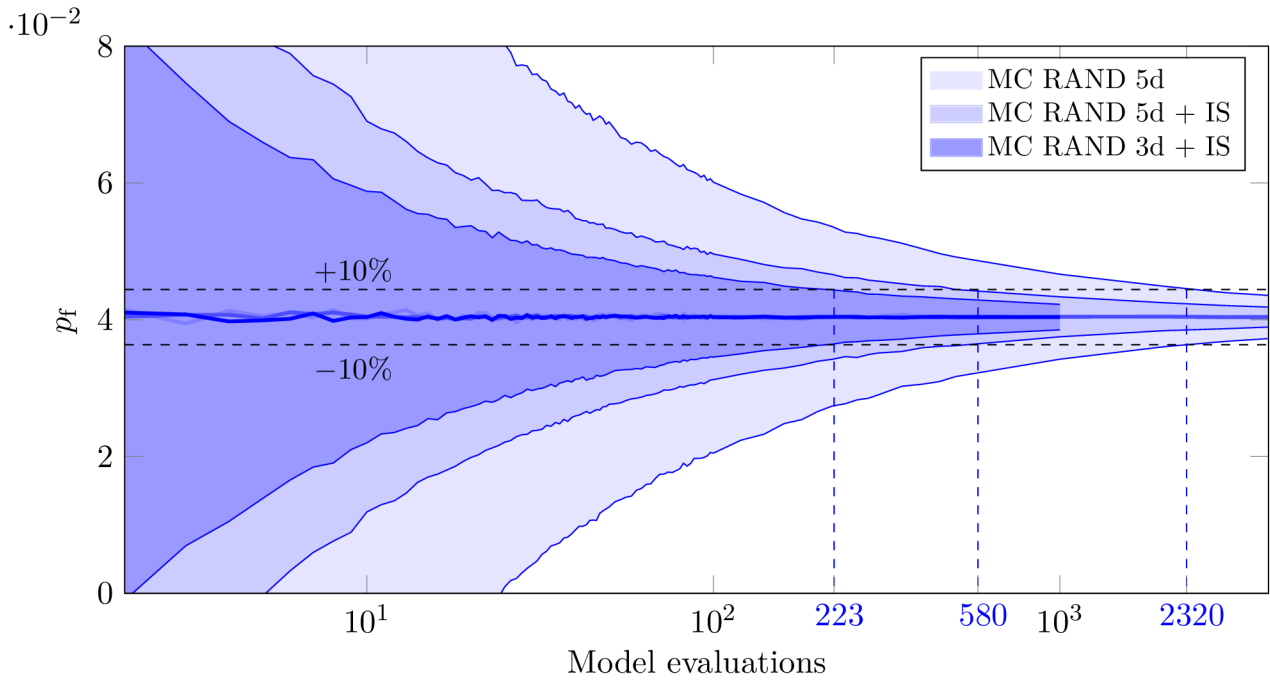
Figure 18: Monte Carlo approximation: comparison of convergence of estimated failure probability, p_{f}, in the 5d space of the original problem and the benefit of using IS in 5d and the reduced 3d space (ave, ave $\pm$ ssd).

Figure 19: Random LHS approximation: comparison of convergence of estimated failure probability, p_{f}, in the 5d space of the original problem and the benefit of using IS in 5d and the reduced 3d space (ave, ave $\pm$ ssd).

Figure 20: DYN ϕ_p Periodic samples: comparison of convergence of estimated failure probability, p_f, in the 5d space of the original problem and the benefit of using IS in 5d and the reduced 3d space (ave, ave $\pm$ ssd).
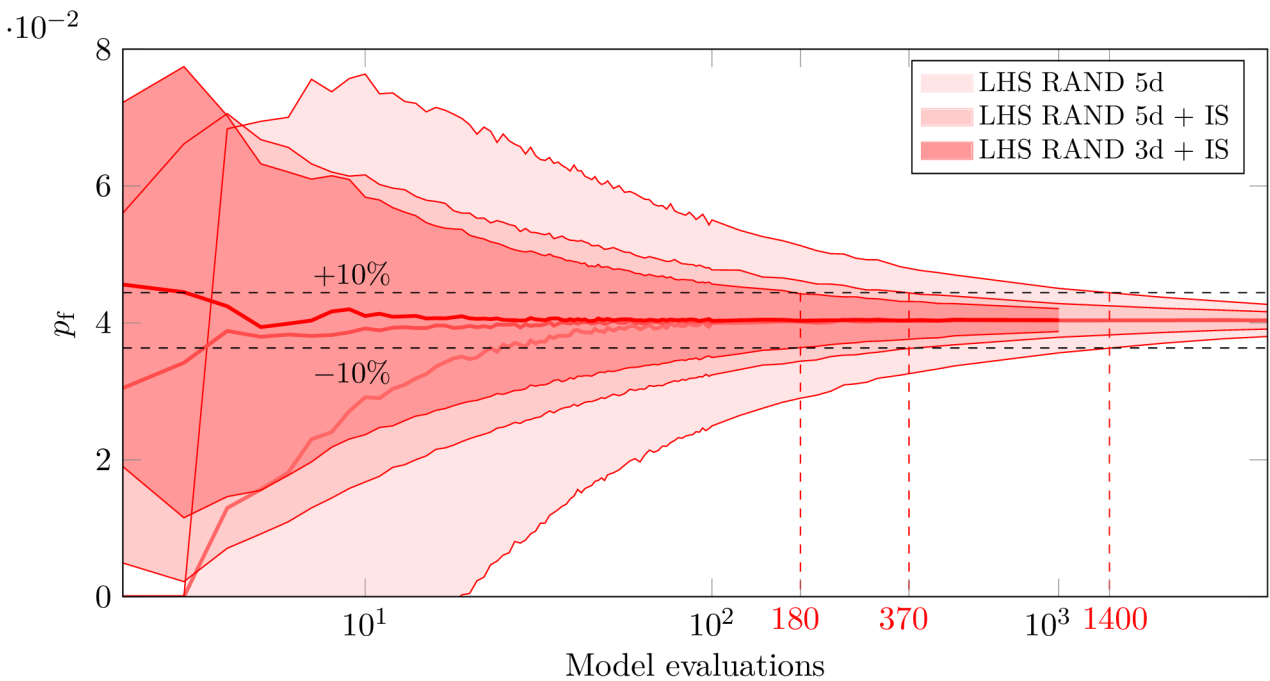

Figure 21: Latinized DYN ϕ_p Periodic samples: comparison of convergence of estimated failure probability, p_f, in the 5d space of the original problem and the benefit of using IS in 5d and the reduced 3d space (ave, ave $\pm$ ssd).
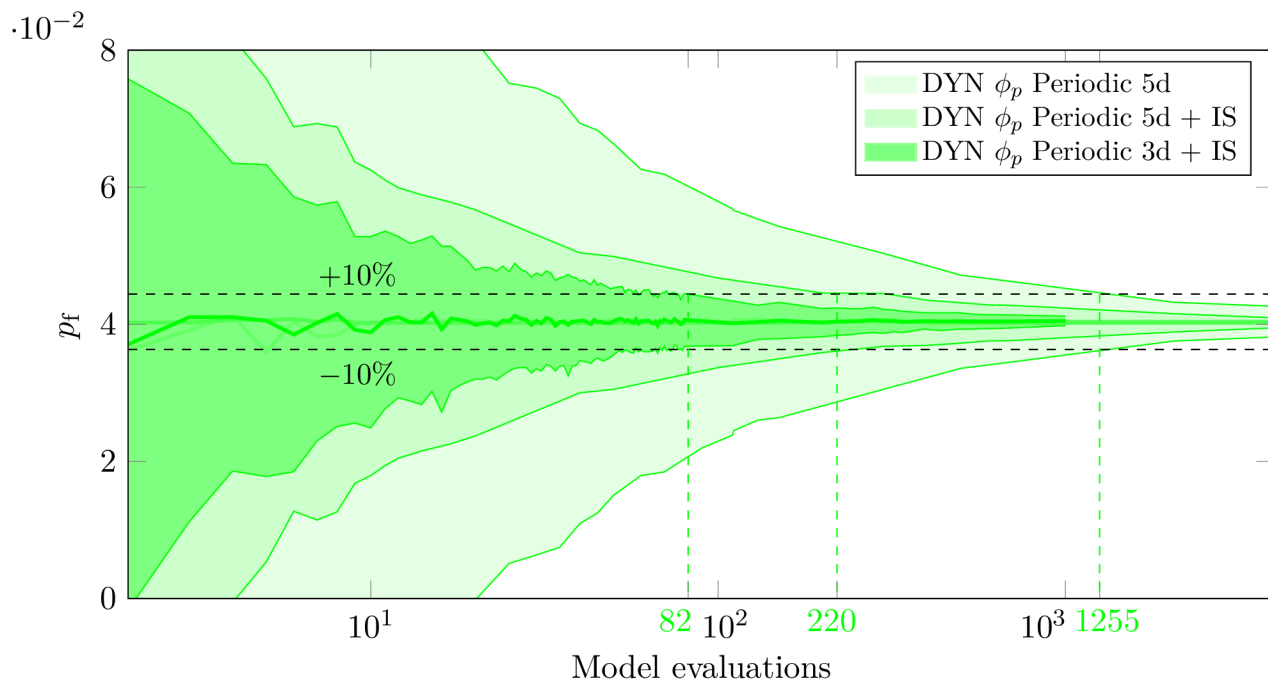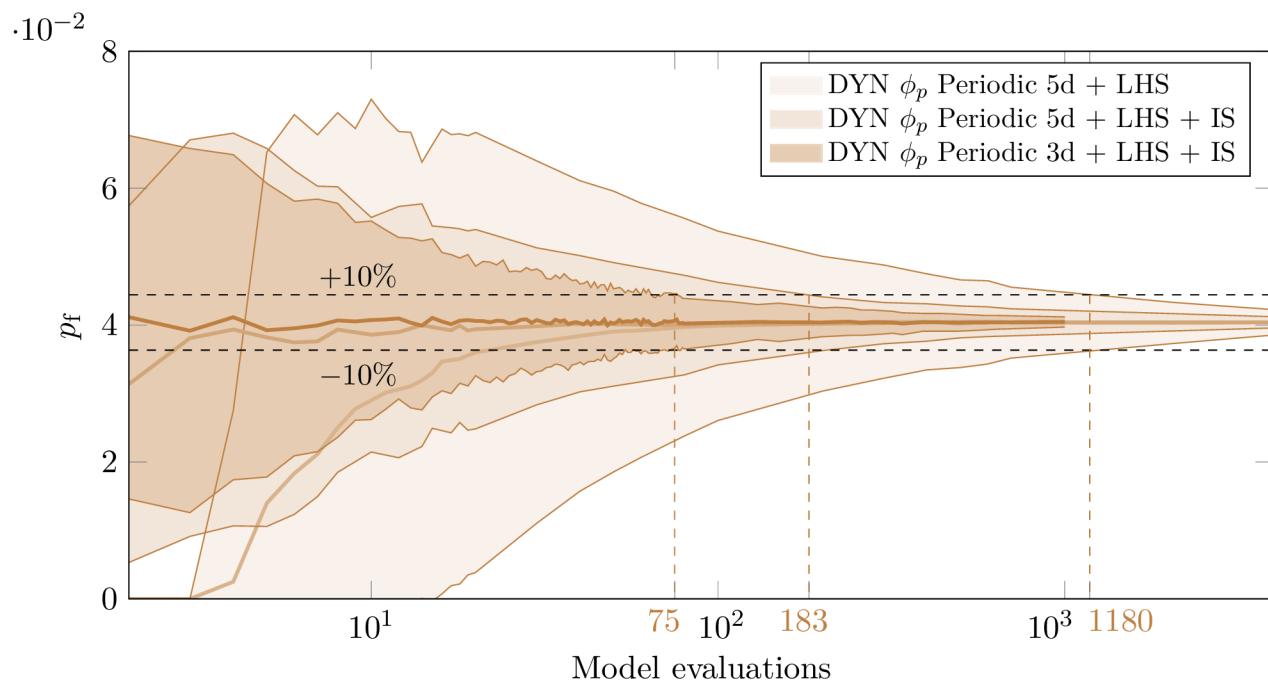
7   Concluding remarks

The presented doctoral thesis was devoted to development of a new efficient tool for optimization of point samples that are spatially and statistically uniform. The primary use-case of these point sets is intended to be the usage as optimized sets of integration points in statistical analyses of computer models using Monte Carlo type integration.

The first task of the work has been a survey of literature concerning properties of existing statistical sampling methods a criteria used for evaluation and/or optimization of uniformity of point sets. An increased attention has been devoted to the distance-based ϕ_p criterion [8]. A substantial part of this work is dedicated to the refinement of the entirely general definition of the ϕ_p criterion aiming at:

- reaching *statistical uniformity* of optimized samples. This means that the probability of appearance of a sampling point is equal all across the design domain. That way, using such samples shall yield statistically unbiased estimation,

- attaining *well distributed* points within each individual point sample, preferably in *noncollapsible* patterns, leading to a decrease in variance of Monte Carlo estimation,

- attaining *self-similar* patterns of design points, that is to obtain ideally optimized patterns for any sample size, N_{sim}, and dimension N_{var},

- developing a criterion that also *takes into account the dimension* of the problem, N_{var}.

It can be concluded that these goals have been completed by (i) incorporating the periodic extension of the design domain as first suggested in [15], (ii) deriving the minimal value of the exponent p within the ϕ_p criterion that results in self-similarity of the criterion value itself and also in self-similar patterns in arbitrary dimension, N_{var}, and finally, by (iii) introducing latinization of samples as a possible post-processing step.

As the next cornerstone of the work can be considered the task of actually materializing the notion of the related Audze-Eglajs criterion [9] that suggests to understand the evaluated set of points as a system of mutually repelling particles.

The actual numerical simulation of the derived dynamical particle system is a task with heavy computational demands, far beyond capabilities of a single CPU thread. A substantial effort has been therefore allocated to author's self-study of the massively parallel computing platform Nvidia CUDA.

It can be stated that the dynamically optimized samples may not necessarily surpass the LHS samples optimized by combinatorial optimization (switching) as in [13]. However, their performance in Monte Carlo estimation is not worse and the efficiency of the dynamical optimization is far superior to the combinatorial approach. The conclusions above indicate that initial objectives of the work were accomplished.

7.1   Future work

One of future tasks directly connected to the present work is to create an open-access database to offer the optimized point samples for other research fellows.

The next paper in preparation concerns replacing the Qhull algorithm by the fast parallelized approximation of Voronoï diagram in relevant sample optimization techniques (such as the Minimax criterion [7]). Also, demonstrating the performance of samples gradually extended by the developed sample extension algorithm is considered as important, especially for engineering applications.

Due to the encountered numerical issues, the next goal is to develop a numerically stable, pseudo-dynamical particle system that is governed by the Maximin criterion, which is the limit case of the ϕ_p criterion ($p = \infty$).

Another future topic is the development of an efficient tool for sampling optimization that yields *directionally invariant*, hyper-dimensional samples. Such samples, invariant with respect to rotation, are considered desirable when dealing with functions that contain strong interactions of input random variables.

One of goals for more distant future is to expand the gained knowledge into another field of interest. That is to begin the development of an efficient solver for lattice particle models of materials that will benefit from the modern massively parallel platforms (such as CUDA [20] or OpenCL [21]) in the largest possible extent.

References

[1] A. Forrester, A. Keane, et al. *Engineering design via surrogate modelling: a practical guide.* John Wiley & Sons, 2008.

[2] R. A. Fisher. *The design of experiments.* Oliver And Boyd; Edinburgh; London, 1937.

[3] V. L. Anderson and R. A. McLean. *Design of experiments: a realistic approach.* Vol. 5. CRC Press, 1974.

[4] Wikipedia contributors. *Moment (mathematics) — Wikipedia, The Free Encyclopedia.* `https://en.wikipedia.org/w/index.php?title=Moment_(mathematics)&oldid=838597128`. [Online; accessed 12-June-2018]. 2018.

[5] J. F. Koksma. "Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1". In: *Mathematica B* 11 (1942/1943), pp. 7–11.

[6] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods.* Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics, 1992. ISBN: 0-89871-295-5.

[7] M. Johnson, L. Moore, and D. Ylvisaker. "Minimax and maximin distance designs". In: *Journal of Statistical Planning and Inference* 2.26 (1990), pp. 131–148. ISSN: 0378-3758. DOI: `10.1016/0378-3758(90)90122-B`.

[8] M. D. Morris and T. J. Mitchell. "Exploratory designs for computational experiments". In: *Journal of Statistical Planning and Inference* 43.3 (1995), pp. 381–402. ISSN: 0378-3758. DOI: `10.1016/0378-3758(94)00035-T`.

[9] P. Audze and V. Eglājs. "New approach for planning out of experiments". In: *Problems of Dynamics and Strengths* 35 (1977). (in Russian), pp. 104–107.

[10] W. J. Conover. "On a Better Method for Selecting Input Variables". 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. 1975.

[11] M. D. McKay, W. J. Conover, and R. J. Beckman. "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code". In: *Technometrics* 21 (1979), pp. 239–245. DOI: 10.1080/00401706.1979.10489755.

[12] M. Grant, S. Boyd, and Y. Ye. *CVX: Matlab software for disciplined convex programming.* 2008.

[13] M. Vořechovský and D. Novák. "Correlation control in small sample Monte Carlo type simulations I: A Simulated Annealing approach". In: *Probabilistic Engineering Mechanics* 24.3 (2009), pp. 452–462. ISSN: 0266-8920. DOI: 10.1016/j.probengmech.2009.01.004.

[14] D. Novák, M. Vořechovský, and B. Teplý. "FReET: Software for the statistical and reliability analysis of engineering problems and FReET-D: Degradation module". In: *Advances in Engineering Software* 72 (2014), pp. 179–192. DOI: 10.1016/j.advengsoft.2013.06.011.

[15] J. Eliáš and M. Vořechovský. "Modification of the Audze–Eglājs criterion to achieve a uniform distribution of sampling points". In: *Advances in Engineering Software* 100 (2016), pp. 82–96.

[16] M. Vořechovský and J. Eliáš. "Improved formulation of Audze-Eglājs criterion for space-filling designs". In: *ICASP12, the 12th International Conference on Applications of Statistics and Probability in Civil Engineering held in Vancouver, Canada on July 12-15, 2015.* Ed. by T. Haukaas. The University of British Columbia. 2015, pp. 1–8. DOI: 10.14288/1.0076173. URL: http://hdl.handle.net/2429/53478.

[17] V. Sadílek and M. Vořechovský. "Evaluation of pairwise distances among points forming a regular orthogonal grid in a hypercube". In: *Journal of Civil Engineering and Management* (2018), in press.

[18] M. Vořechovský. "Hierarchical refinement of Latin hypercube samples". In: *Computer-Aided Civil and Infrastructure Engineering* 30.5 (2015), pp. 394–411.

[19] S. H. Lee and B. M. Kwak. "Response surface augmented moment method for efficient reliability analysis". In: *Structural safety* 28.3 (2006), pp. 261–272.

[20] Nvidia. *Nvidia CUDA C/C++.* `https://developer.nvidia.com/cuda-toolkit`. Since 2007.

[21] K. Group. *The open standard for parallel programming of heterogeneous systems.* `https://www.khronos.org/opencl/`. Since 2009.

About author

Personal data

Jan Mašek

Institute of Structural Mechanics

Faculty of Civil Engineering

Brno University of Technology

Veveří 95, 635 00 Brno

Email: masek.j@fce.vutbr.cz

Website: www.fce.vutbr.cz/STM/masek.j

Specialization, research interests

Mathematical modeling based on physical discretization of continuum.
Parallelization of solution using the nVidia CUDA architecture.
Large deflections, chaotic behavior of deterministic dynamical systems.

Programming skills

JAVA, C, CUDA C/C++, LaTeX, HTML, CSS, Python

Teaching Experience

BD001	Fundamentals of Structural Mechanics, Czech
BD002	Elasticity and Plasticity, Czech

Courses and other activities

06/2018	ECCOMAS Advanced Course on Computational Structural Dynamics, Institute of Thermomechanics, Czech Academy of Sciences, Prague, CZ
01/2018	Computational challenges in the reliability assessment of engineering structures. Netherlands Organization for Applied Scientific Research, Delft, NL

10/2017	ANSYS Mechanical APDL 1, SVS FEM Brno, CZ
10/2017	ANSYS Mechanical APDL 2, SVS FEM Brno, CZ
11/2017	ANSYS Mechanical APDL Dynamics, SVS FEM Brno, CZ

Participation in research projects

from 01/2018 to 12/2018	STM FAST	Research of topological features of discrete processes and their use for optimization of statistical sampling. (FAST-J-18-5254)
from 01/2018 to 12/2018	UTKO FEKT, STM FAST	Low-energy device for secured transmission from physical measurements according to Industry 4.0. (FAST/FEKT-J-18-5365)
from 01/2017 to 12/2017	STM FAST	Development of theory and metodics for optimized design of experiment using dynamical particle system. (FAST-J-17-4564)
from 01/2016 to 12/2018	STM FAST	Development of advanced sampling methods for statistical analysis of structures. (GA16-22230S)
from 01/2016 to 12/2016	STM FAST	Optimization of design of experiment using a discrete dynamical particle system. (FAST-J-16-3486)
from 01/2016 to 12/2016	UTKO FEKT, STM FAST	Aggregation gateway for secured data transmissions from instantaneous measurements of physical quantities. (FAST/FEKT-J-16-3344)
from 06/2015 to 12/2015	UTKO FEKT	Smart Multi-Purpose Home Gateway 2.0. Proof of concept demonstrator. Research and development of OSGi-based smart hub platform. (2015_HS18557077)

Software development

| Since 2016 | CUDA Hypercube | CUDA C / JAVA application serving for optimization of point samples for design of experiment. |
| Since 2014 | GTDiPS GUI | Free JAVA application serving for advanced transformations of large multidimensional point sequences. Utilized by the staff and students of the Department of Structural mechanics. Developed in cooperation with doc. Frantík. |

Intellectual properties

| 2017 | Utility model | Aggregation gateway for secured data transmissions from instantaneous measurements of physical quantities. |

Abstract

The presented doctoral thesis aims at development a new efficient tool for optimization of uniformity of point samples. One of use-cases of these point sets is the usage as optimized sets of integration points in statistical analyses of computer models using Monte Carlo type integration. The tasks of the work concern a survey of currently used criteria for evaluation and/or optimization of uniformity of point sets. A critical evaluation of their properties is presented, leading to suggestions towards improvements in spatial and statistical uniformity of resulting samples. A refined variant of the general formulation of the ϕ_p optimization criterion has been derived by incorporating the periodically repeated design domain along with a scale-independent behavior of the criterion.

Based on a notion of a physical analogy between a set of sampling points and a dynamical system of mutually repelling particles, a hyper-dimensional N-body system has been selected to be the driver of the developed optimization tool. Because the simulation of such a dynamical system is known to be a computationally intensive task, an efficient solution using the massively parallel GPGPU platform Nvidia CUDA has been developed. An intensive study of properties of this complex architecture turned out as necessary to fully exploit the possible solution speedup.

Apart from statistical uniformity, the samples optimized by efficient dynamical simulations also tend to consist of uniform, self-similar point patterns. Furthermore, the desired self-similarity of samples also results in well optimized point layouts in all subspaces of the design domain. Due to the advantageneous formulation of the particle model, the possibility of additional sample size extension one-by-one is inherently possible. The performance of the dynamically optimized samples if used Monte Carlo estimation has proven to be superior to the commonly used sampling methods and comparable to samples optimized by brute-force combinatorial optimization. Due to the efficiency of the parallelized solution, obtaining the dynamically optimized samples requires orders of magnitude lower computational time.

Keywords