Design of experiments in mixed continuous and discrete space

Thi Thoi TRAN [1,3], Sébastien Da Veiga [2], Delphine SINOQUET [1], Marcel MONGEAU [3]

[1] IFP Energies nouvelles [2] Safran Tech [3] ENAC, Université de Toulouse

Context

Design of Experiments (DoEs) :

- Used in exploration, optimization and uncertainty quantification, when coupled with time-consuming numerical simulators
- Aim : select a limited number of values to assign to the simulator input variables that give a maximal knowledge on the simulator outputs of interest

Context

- Vast literature on space-filling DoEs for continuous variables
- Less extensive for mixed continuous and discrete variables

Objective 1 : Propose space-filling DoE in mixed continuous and discrete space :

$$\mathscr{D} = \{z = (x, y) \in \mathbb{R}^m \times \mathbb{I}^n\}. \quad (1)$$

- DoEs are applied to integer encoding of discrete variables

Objective 2 : Relevant DoEs for a large range of applications

Literature review

- Extended work from LHS for continuous to mixed variables : independent Latin Hypercubes (**Santner2003**), Sliced Latin Hypercube (**Quian2012**)

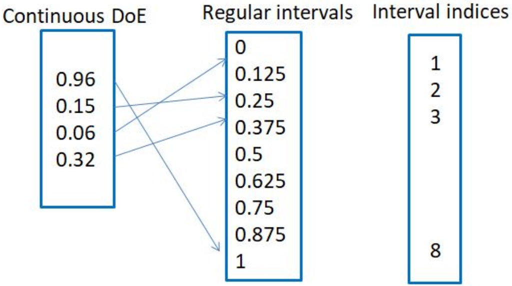
  Require large numbers of points to obtain a space-filling DoE
  $\Longrightarrow$ Not reasonable for expensive simulators !

- In derivative-free optimization context : rounding LHS (R-LHS) (**Costa2018**), i.e. a classical LHS for relaxed integer variables, then values are rounded to obtain discrete values

  Duplicated points

Literature review

- Projected DoEs : based on the projection of continuous values to integers representing the indices of discrete variables
- Example : generate DoE with 4 integers $\in \{1, 2, \dots, 8\}$



- Coupled with various DoEs : LHS, Sobol sequences (A-LHS, A-Sobol)

No prior information is used, e.g., cyclic symmetry, periodicity, correlations,...

Objective 3 : DoEs taking into account some given prior information

Outline

Table of Contents

Problem statement

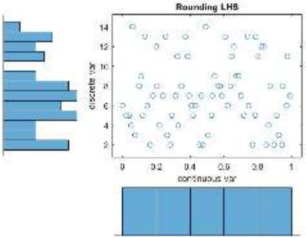
Select a limited number of values to assign to the simulator input variables that give a maximal knowledge of the mixed space under study :

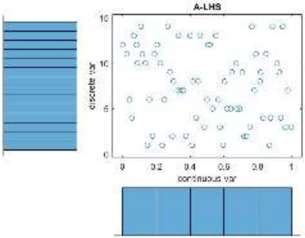
$$\mathscr{D} = \{z = (x, y) \in \mathbb{R}^m \times \mathbb{I}^n\},$$

where $x \in \mathbb{R}^m$, $y \in \mathbb{I}^n$ are the continuous and discrete variables, respectively, and $\mathbb{I}$ denotes the discrete space (*e.g.* integer, binary or categorical variables)

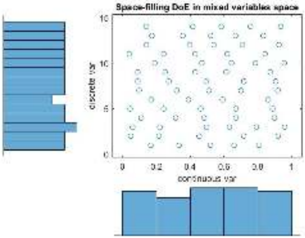
Space-filling DoEs for mixed continuous and discrete variables

- We do not want





- We want well distributed points in both continuous and discrete domains



Space-filling DoEs for continuous variables

Based on different criteria :

- Geometrical criteria :
  1. Minimax DoE : minimize the maximal distances between any points and DoE points
  2. Maximin DoE : maximize the minimal distance
- Low discrepancy criterion of the DoE points
- Kernel-embedding of distributions

Objective : extend DoEs based on kernel-embedding distributions to mixed discrete variables

DoEs based on kernel-embedding distribution (I)

Main idea :

- Based on kernel-embedding of probability distributions, (**Hickernell1998**)
- Minimization of the maximum mean discrepancy (MMD) between an empirical measure corresponding to the DoE and a target distribution

Principle :

- A set of candidate points from the target distribution $X = \{x_1, x_2, \ldots, x_N\}$, N is large
- Aim : Approximate X by a subset of points $\bar{X} = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{nDoE}\} \subset X, nDoE << N$

Kernel herding : illustration

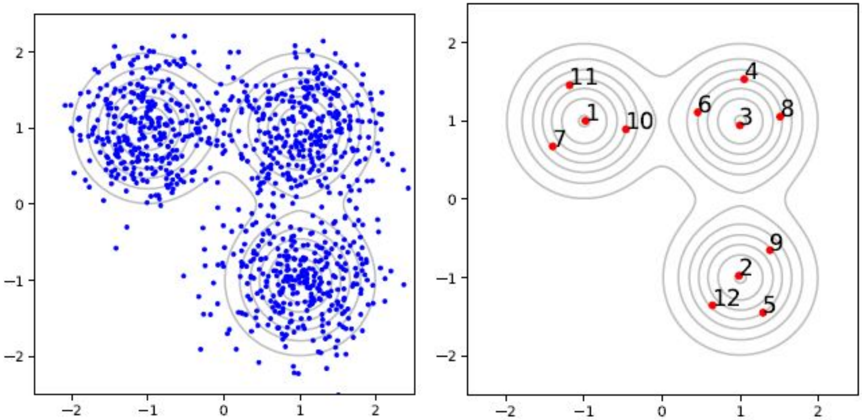


Figure – Illustration of kernel herding applied to a mixture of Gaussian for continuous variables (**Chen2010**)

DoEs based on kernel-embedding distribution (II)

The problem is solved by minimizing $\text{MMD}^2(\mathbb{P}_X, \mathbb{P}_{\bar{X}})$:

$$\min_{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{nDoE}} \text{MMD}^2\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{x_i}, \frac{1}{nDoE}\sum_{j=1}^{nDoE}\delta_{\bar{x}_j},\right),$$

where the MMD is defined as follows :

$$\text{MMD}^2(\mathbb{P}_X, \mathbb{P}_{\bar{X}}) = \|\mu_{\mathbb{P}_X} - \mu_{\mathbb{P}_{\bar{X}}}\|^2_{\mathcal{H}_k},$$

with $\mathcal{H}_k$ a Reproducing Kernel Hilbert Space (RKHS), k a positive definite (p.d) kernel and $\mu_{\mathbb{P}} = \int k(x,.)d\mathbb{P}(x)$ the kernel embedding of the probability distribution $\mathbb{P}$.

DoEs based on kernel-embedding distribution (III)

- Interestingly, the RKHS framework allows to simplify MMD into :

$$\mathrm{MMD}^2(\mathbb{P}_X, \mathbb{P}_{\bar{X}}) = \mathbb{E}_{\xi,\xi' \sim \mathbb{P}_X} k(\xi, \xi') + \mathbb{E}_{\zeta,\zeta' \sim \mathbb{P}_{\bar{X}}} k(\zeta, \zeta') - 2\mathbb{E}_{\xi \sim \mathbb{P}_X, \zeta \sim \mathbb{P}_{\bar{X}}} k(\xi, \zeta).$$

- The Kernel herding approach is a greedy sequential algorithm to solve the minimization problem (**Chen2010**) :

$$\bar{x}_1^* = \underset{\bar{x} \in X}{\operatorname{argmax}} \frac{1}{N} \sum_{j=1}^{N} k(\bar{x}, x_j).$$

For $t = 1, \ldots, nDoE - 1$:

$$\bar{x}_{t+1}^* = \underset{\bar{x} \in X}{\operatorname{argmax}} \frac{1}{N} \sum_{j=1}^{N} k(\bar{x}, x_j) - \frac{1}{t+1} \sum_{i=1}^{t} k(\bar{x}, \bar{x}_t^*).$$

⚠

- The current kernel herding algorithm only considers kernels for continuous variables
- The requirement of p.d kernel can be difficult to extend to mixed discrete variables
- **Our proposal : We focus on cases where a distance for discrete variables is already defined** (which will also account for prior information)

Extending kernel herding to mixed variables

- Two proposals to use this distance inside kernel herding

  1. We build a continuous encoding of the discrete variables. The original kernel herding algorithm with a standard p.d continuous kernel is then used in this transformed space
  2. We directly define a p.d kernel for mixed discrete variables which is integrated inside kernel herding

First proposal : *Greedy-MDS* approach

- Main idea : we build a continuous encoding of discrete variables

  1. We compute all the pairwise distances in the discrete space for the large sample from the target
  2. We apply Multidimensional Scaling (MDS), (**Kruskal1964**), to this distance matrix $\Longrightarrow$ a continuous representation of discrete variables
  3. We build the target distribution in this transformed space by completing the encoded distribution of the original discrete variables with a sliced space filling design for continuous variables
  4. We apply kernel herding with a classical kernel for continuous variables, e.g. Gaussian kernel
  5. We retrieve the final DoE by "de-encoding" the selected samples into the original discrete values (thanks to the MDS correspondence)

Second proposal : *Adapted-Greedy* approach

- Main idea : we directly integrate an appropriate kernel for mixed discrete variables inside kernel herding
- This requires a p.d kernel for mixed continuous and discrete variables
- Adapt kernel to application which accounts for the prior information on mixed variables, e.g., symmetry, correlations,...

Relevant approach for a large range of applications

- Two examples :

  1. The mixed continuous and **binary** kernel (**Hutter2014**), no prior information
  2. The soft string kernel (**Wu2019**) which is defined from **any distance**, thus being able to include a prior

Specific kernels

- Mixed continuous and **binary** kernel (**Hutter2014**)

$$K^H_{mixed}(z_i, z_j) = \exp\left(-\lambda_c \cdot \|x_i - x_j\|^2 - \lambda_b d_H(y_i, y_j)\right),$$

d_H the Hamming distance, hyper-parameters $\lambda_b = n$ and λ_c computed thanks to the rule of thumb :

$$\lambda_c = \lambda_b \frac{\text{median}(d_H(y_i, y_j))}{\text{median}(\|x_i - x_j\|_2^2)}. \quad (2)$$

- The soft string kernel (**Wu2019**)

$$k^{soft}(y_i, y_j) = \sum_{\omega \in \Omega} e^{-\gamma\{d(y_i,\omega) + d(y_j,\omega)\}},$$

d is a user-defined distance. Hence we propose a mixed soft string kernel :

$$K^{soft}_{mixed}(z_i, z_j) = \exp\left(-\lambda_c \cdot \|x_i - x_j\|^2\right) \; k^{soft}(y_i, y_j),$$

hyper-parameters λ_c, γ are chosen as above.

Illustrations of our adapted-greedy approach

- Problem with 1 continuous variable, 1 categorical variable with 14 levels

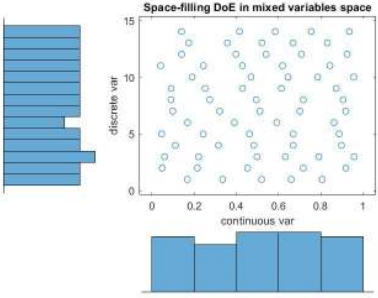


Figure – DoE obtained by adapted-greedy approach

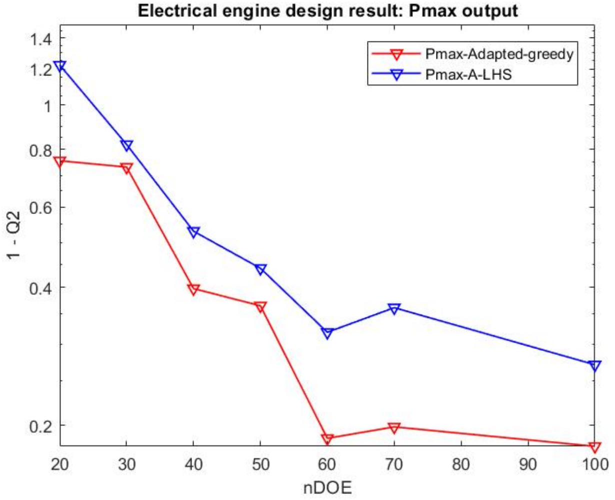
Table of Contents

DoEs for mixed continuous and discrete variables

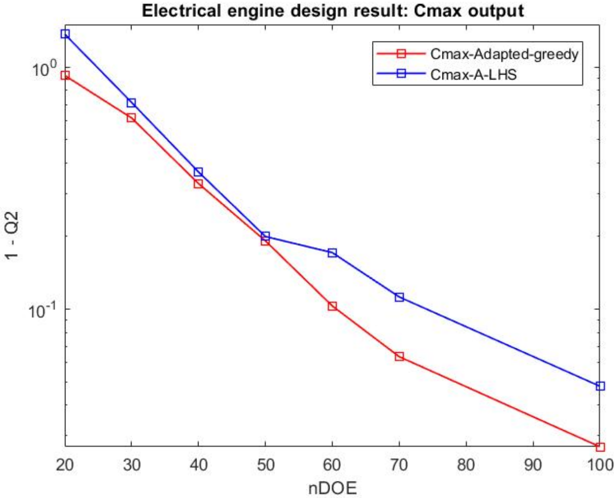
Example 1 : we use a DoE to build a surrogate model

- Test case : the maximal torque (Cmax) and the maximal power (Pmax) of an electrical engine
- Inputs : 1 categorical (rotor type), 2 integer (number of wires and coils) and 1 continuous variables (length of the rotor). We encode all discrete variables as binary ones
- We compare two DoE methods : kernel herding approach (Adapted Greedy), projected LHS (A-LHS)
- We use the p.d. kernel proposed by (**Hutter2014**)
- The surrogate model is a kriging model with adapted mixed discrete kernel, (**Qian2008**)

Numerical results for DoEs with mixed discrete variables



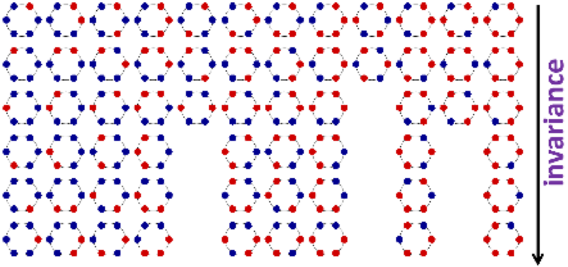
Numerical results for DoEs with mixed discrete variables



DoEs for mixed continuous and binary variables with cyclic symmetry property (I)

Example 2 : we build a DOE to approximate the mean of a function with symmetrical inputs

- We assume that the input variables are binary arrangements invariant by rotation : concept of **necklace**, e.g. 1001 ~ 1100 ~ 0110...



- In (**Tran2021**), we proposed the *necklace distance* :

$$d_{neck}(y, y') = \min_{i=1,2,\dots,n} d_H(y, Rot^i(y'))$$

DoEs for mixed continuous and binary variables with cyclic symmetry property (II)

- Tested functions : $\sin(\|\cdot\|)$, Branin, Hartman3, Perm6 transformed into cyclic symmetry problems

Functions	$m \times n$	# necklaces
$\sin(\|\cdot\|)$	2×7	20
Branin	1×7	20
Hartman3	3×6	14
Perm6	5×5	10

- RMSE of the mean of the function for each necklace $F_i(.)$:

$$\text{RMSE}(\delta_{nDoE}) = \sqrt{\frac{1}{n_{neck}} \sum_{i=1}^{n_{neck}} \left(F_i(X^i_{ref}) - F_i(\delta^i_{nDoE}) \right)^2},$$

n_{neck} is the number of distinct necklaces, $\delta^i_{nDoE}, i = 1, 2, \ldots, n_{neck}$ are the DoE points for a given necklace i

Mixed continuous and binary with cyclic symmetry property : results

$$nDoE = kn_{neck}$$

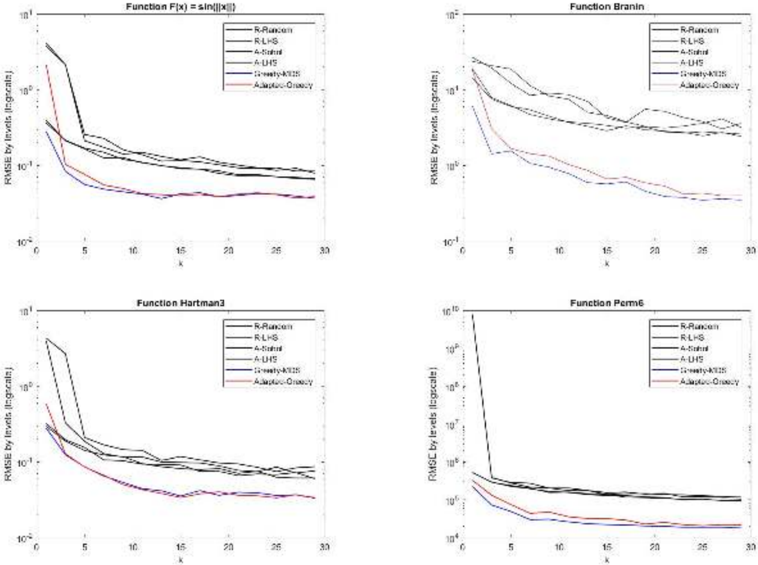


Table of Contents

Conclusions and perspectives

- Main results
  1. Extended DoEs to mixed discrete variables thanks to kernel herding approach
  2. Different p.d kernels adapted to the type of problems
  3. Promising results for various types of mixed discrete problems
- Perspectives
  1. Enlarge the application domain of our DoE methods, e.g., other objects as time series, images...with adapted kernels

Bibliography I

[Costa2018] A. Costa, G. Nannicini, *RBFOpt : An open-source library for black-box optimization with costly function evaluations. Mathematical Programming Computation (2018), 597-629*

[Bremner] D. Bremner et al. Necklaces, convolutions, and $X + Y$

[Cuturi2007] M. Cuturi, J. Vert, O. Birkenes and T. Matsui, A Kernel for Time Series Based on Global Alignments, 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP, 2007.

[Elamri2020] El Amri Mohamed Reda, Helbert Celine, Lepreux Olivier, Zuniga Miguel Munoz, Prieur Clémentine and Sinoquet Delphine, Data-driven stochastic inversion via functional quantization,Statistics and Computing, 2020,https://hal-ifp.archives-ouvertes.fr/hal-02291766

[Quian2012] Peter Z. G. Qian, Sliced Latin Hypercube Designs, Journal of the American Statistical Association, 2012. http://www.jstor.org/stable/23239678

Bibliography II

[Santner2003] Thomas J. Santner, Brian J. Williams and William I. Notz, The design and analysis of computer experiments, Springer, New York, NY, 2003. https://doi.org/10.1007/978-1-4757-3799-8

[Kruskal1964] Joseph Bernard Kruskal, Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, Psychometrika, 1964. https://doi.org/10.1007/BF02289565

[Hickernell1998] Fred Hickernell, A generalized discrepancy and quadrature error bound, 1998, Mathematics of computation

[Chen2010] Yutian Chen, Max Welling and Alex Smola, Super-Samples from Kernel Herding, 2010, proceedings

[Wu2019] Lingfei Wu, Ian En-Hsu Yen, Siyu Huo, Liang Zhao, Kun Xu, Liang Ma, Shouling Ji and Charu Aggarwalm Efficient Global String Kernel with Random Features : Beyond Counting Substructures, 2019, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining

Bibliography III

[Hutter2014] *Frank Hutter, Lin Xu, Holger H. Hoos and Kevin Leyton-Brown, Algorithm runtime prediction : Methods and evaluation, 2014, Artificial Intelligence*

[Tran2021] *Thi Thoi Tran, Delphine Sinoquet, Sébastien da Veiga, Marcel Mongeau, Derivative-free mixed binary necklace optimization for cyclic-symmetry optimal design problems, 2021, submitted to Optimization and Engineering, available in ⟨hal-03170761⟩*

[Dixon1975] *Dixon, L.-C.-W and Szegö, G.-P, The global optimization problem : An introduction, 1975, In : Dixon, L.C.W, Szegö, G.P (eds.) Towards Global Optimization, North Holland*

[Neumaier2014] *Neumaier, A., Neumaier's collection of test problems for global optimization, 2014, Retrieved in May 2014,* `http://www.mat.univie.ac.at/~neum/glopt/my_problems.html`

Perspectives : time series applications

Objective : build DOEs for a functional variable to compute the mean of a function

- Applied to Max-stable curves (200)
- DoEs methods : Reduced LHS (LHS + PCA), L_2 greedy quantization method (L2 quantization + PCA) (**ElAmri2019**), adapted-greedy
- Kernel : normalized global alignment kernel, (**Cuturi2007**)
- Measure the expectation estimation errors of function

$$f : V \to \max_t V_t |0.1 \cos(a \max_t V_t) \sin(b)(a + b \min_t V_t)^2| \int_0^T (30 + V_t)^{\frac{ab}{20}} dt, a = 2.95, b = 3.97$$

DoEs for time series : kernel

Kernel used in adapted-greedy approach : the normalized global alignment kernel for same length time series, **Cuturi2007** :

$$K^{GAK}(V_1, V_2) = K(V_1, V_2) - \frac{1}{2}(K(V_1, V_1) + K(V_2, V_2)) \quad (3)$$

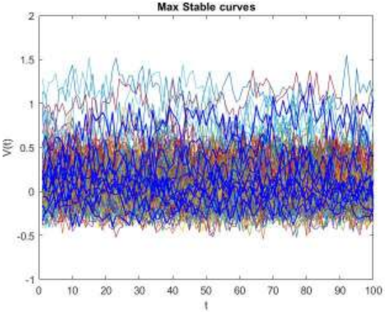
where K is the global alignment kernel

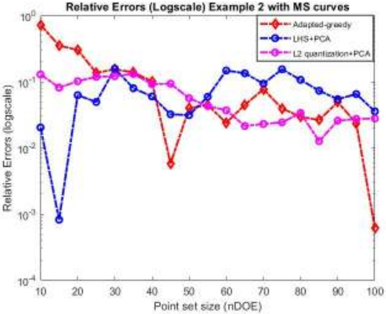
$$K(V_1, V_2) = \prod_{i=1}^{|V_2|} e^{-\phi_\sigma(V_1^i, V_2^i)} \quad (4)$$

with

$$\phi_\sigma(V_1^i, V_2^i) = \frac{1}{2\sigma^2}\|V_1^i - V_2^i\|^2 + \log(2 - e^{-\frac{\|V_1^i - V_2^i\|^2}{2\sigma^2}})$$

Preliminary results on time series





- Adapted-greedy is competitive with other methods
- L2 quantization method requires a preliminary step of dimension reduction
- Adapted-greedy gains in terms of CPU time

DoEs for mixed continuous and binary variables with cyclic symmetry property

Objective : build DOE to compute the mean of a function

- Tested functions : $\sin(\|x\|)$, Branin, Hartman3, Perm6
- Build cyclic symmetry property function : restrict the last continuous variable to take only a finite number of values in the discretized-interval set :

$$X^{end} := \left\{ \underline{x}^{end} + (w-1)\frac{\bar{x}^{end} - \underline{x}^{end}}{l-1} : w = 1,2,\ldots,l \right\}$$

- Each level corresponds to a distinct binary arrangement of size n, for example

$$\min_{x \in [\underline{x},\bar{x}],\, y \in \{0,1\}^2} f(x,y) := \begin{cases} F(x,1), & \text{if } y = (0,0) \\ F(x,2), & \text{if } y = (0,1),(1,0) \\ F(x,3), & \text{if } y = (1,1) \end{cases} \quad (5)$$