Design of Experiments in mixed continuous and discrete space

Thi Thoi TRAN
IFP Energies nouvelles

Supervisor(s): Dr. Sébastien Da Veiga (Safran Tech), Dr. Delphine Sinoquet (IFP Energies Nouvelles) and Prof. Marcel Mongeau (ENAC, Université Paul Sabatier)

Ph.D. expected duration: Sept. 2018 - Sept. 2021

Adress: IFP Energies Nouvelles, 1-4 avenue Bois-Préau, 92582 Rueil-Malmaison Cedex, France

Email: thi-thoi.tran@ifpen.fr

Abstract: Design of experiments (DoE) are used in various contexts such as optimization or uncertainty quantification based on a time-consuming numerical simulator. It aims to select a limited number of values to assign to the simulator input variables that give a maximal knowledge on the simulator outputs of interest. One motivating application is the optimal design of turbine blades in an helicopter engine [1], which takes as inputs mixed continuous and binary variables.

In this study, we propose two new methods for space-filling designs in mixed continuous and discrete space defined as

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

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). Although the literature on space-filling DoEs is vast in the continuous case, we focus here on a particular setting relying on kernel-embedding of probability distributions, since it can be more easily generalized to the mixed variables case, as we will see below. In this framework, several previous works reformulated the space-filling DoE problem as the minimization of the *maximum mean discrepancy* (MMD) between an empirical measure (corresponding to the DoE) and a target distribution. For two probability distributions P and Q and a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ with positive definite kernel k, the MMD distance between P and Q is defined as

$$\mathrm{MMD}(\mathrm{P}, \mathrm{Q}) = \|\mu_\mathrm{P} - \mu_\mathrm{Q}\|_\mathcal{H} \tag{2}$$

where $\mu_\mathrm{P} = \int k(x, .)d\mathrm{P}(x)$ and $\mu_\mathrm{Q} = \int k(x, .)d\mathrm{Q}(x)$ are the *kernel embeddings* of P and Q, respectively, and serve as representations of the probability distributions. The RKHS framework makes it possible to write the distance with only expectations of kernel functions: $\mathrm{MMD}^2(\mathrm{P}, \mathrm{Q}) = \mathbb{E}_{\xi,\xi'\sim\mathrm{P}}k(\xi, \xi') + \mathbb{E}_{\zeta,\zeta'\sim\mathrm{Q}}k(\zeta, \zeta') - 2\mathbb{E}_{\xi\sim\mathrm{P},\zeta\sim\mathrm{Q}}k(\xi, \zeta)$.

Coming back to our DoE setting, when P is an empirical measure $\frac{1}{n}\sum_{i=1}^n \delta_{s_i}$ of n points $s_i \in [a, b]^d$ in a hypercube and the target measure Q is the uniform distribution on $[a, b]^d$, the squared MMD actually writes as well-known discrepancy measures, depending on the choice of the kernel k [2]. But the choice of the target Q is not restricted, and the case where it is also given as an empirical measure $\frac{1}{N}\sum_{i=1}^N \delta_{t_i}$ with N typically much larger than n is of particular interest. Indeed this case corresponds to applications where one is given a large sample from a probability distribution and the goal is to find a small subset of samples which best represents the underlying unknown distribution [3]. The problem then writes as the optimization problem

$$\min_{s_1,s_2,\ldots,s_n} \mathrm{MMD}^2\Big(\frac{1}{n}\sum_{i=1}^n \delta_{s_i}, \frac{1}{N}\sum_{i=1}^N \delta_{t_i}\Big). \tag{3}$$

When the points s_i are chosen among the points in the large sample t_i, [4] proposed *kernel herding*, a greedy sequential algorithm for solving this minimization problem. On the contrary when the

points s_i are not restricted, it is possible to design efficient optimization strategies as in [3]. In the following we assume that we are in the first situation.

Unfortunately all the MMD-based approaches proposed so far only consider kernels for continuous variables only. Our goal is thus to generalize this point of view to mixed continuous / discrete variables. The starting point is a user-defined distance in the discrete space that characterizes any prior information available from the type of problem we address: for instance in the turbomachine blade application mentioned before, the *necklace* distance introduced by [5] accounts for the cyclic symmetry of the design problem. Once such a distance is specified, two roads can be taken:

1. The *greedy-MDS* approach, where we build a continuous encoding of the discrete variables. This is achieved by first computing all the pairwise distances in the discrete space for the large sample and then applying Multi-Dimensional Scaling [6]. The original kernel-herding algorithm with a standard kernel on continuous variables is then used in the space consisting of the continuous input variables and the encoded discrete variables obtained with MDS. The final DoE is ultimately retrieved by examining the selected subsamples and find the MDS association between the discrete variables and their continuous encoding;

2. The *adapted-greedy* method, which directly integrates a kernel on mixed variables inside kernel herding. Building a kernel from a distance is not straightforward since one has to ensure that the kernel must be positive definite. To perform this step we rely on the soft string kernel recently introduced by [7]. For our particular necklace distance, the mixed kernel writes
$$K_{mixed}(z_i, z_j) = e^{\lambda \|x_i - x_j\|^2} \sum_{\omega \in \Omega} e^{-\gamma\{d_{neck}(y_i,\omega) + d_{neck}(y_j,\omega)\}},$$
where $\lambda, \gamma > 0$ are kernel hyperparameters and Ω is the set that contains all the feasible discrete elements.

We apply the two proposed methods to three different types of DoE problems (mixed integers, mixed binaries with cyclic symmetry and time series). The obtained results illustrate the good performances of the methods and the wide range of applications they can address.

References

[1] Moustapha Mbaye. *Conception robuste en vibration et aéroélasticité des roues aubagées de turbomachines*. PhD thesis, Université Paris-Est Marne la vallée, 2009.

[2] Fred Hickernell. A generalized discrepancy and quadrature error bound. *Mathematics of computation*, 67(221):299–322, 1998.

[3] Simon Mak, V Roshan Joseph, et al. Support points. *The Annals of Statistics*, 46(6A):2562–2592, 2018.

[4] Yutian Chen, Max Welling, and Alex Smola. Super-samples from kernel herding. pages 109–116, 01 2010.

[5] Thi-Thoi Tran, Delphine Sinoquet, Sébastien Da Veiga, and Marcel Mongeau. An Adapted Derivative-Free Optimization Method for an Optimal Design Application with Mixed Binary and Continuous Variables. In *6th International Conference on Computer Science, Applied Mathematics and Applications, ICCSAMA 2019*, Hanoï, Vietnam, December 2019.

[6] Joseph Bernard Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29:1–27, 1964.

[7] Lingfei Wu, Ian En-Hsu Yen, Siyu Huo, Liang Zhao, Kun Xu, Liang Ma, Shouling Ji, and Charu Aggarwal. Efficient global string kernel with random features: Beyond counting substructures. New York, USA, 2019. Association for Computing Machinery.

Short biography – Thi-Thoi Tran got her bachelor's degree in mathematics from Hanoi university of sciences and then graduated with a Master's degree in optimization at Limoges University. She is currently a third year PhD student at IFPEN. The thesis, funded by Safran Tech, focuses on nonlinear optimization problem with mixed continuous and discrete variables for black-box simulators.