Optimisation Bayésienne en grande dimension: application en physique des réacteurs nucléaires.

Clément Gauchy [1]

[1] *Université Paris-Saclay, CEA, Service de Génie Logiciel pour la Simulation, 91191 Gif-sur-Yvette, France*
clement.gauchy@cea.fr

Résumé. La quantification des incertitudes (UQ) est une étape cruciale lors de la conception de systèmes industriels complexes, en particulier dans l'industrie nucléaire. Il peut être nécessaire au cours d'une étude d'UQ d'effectuer l'optimisation d'un critère qui est obtenu par une simulation multiphysiques, cette simulation est souvent coûteuse en terme de temps de calcul et de ressources informatiques. Dans cet article, nous étudions un problème jouet consistant à trouver une nappe de puissance optimale d'un réacteur à eau pressurisée (REP) fictif vis à vis d'un critère scalaire, une nappe de puissance étant une distribution spatiale de la puissance à l'intérieur du cœur du réacteur nucléaire. Le critère scalaire pour une nappe de puissance donnée est calculé en utilisant une fonction non linéaire d'une distance entre cette nappe de puissance et une nappe de puissance de référence. Tout d'abord, une réduction de dimension sera appliquée à l'ensemble de données des nappes de puissance. Ensuite, nous comparerons plusieurs algorithmes d'optimisation Bayésienne en termes de performance de regret.

Mots-clés. Statistique appliquée, Optimisation Bayésienne, Réduction de dimension.

Abstract. Uncertainty Quantification (UQ) is a crucial step when designing complex industrial systems, especially when it comes to nuclear industry. It could be necessary during an UQ study to perform optimization of a criterion which is obtain by a multiphysics simulation, this simulation is often costly in terms of computation time and high-performance computing resources. In this article, we study a toy problem of finding a optimal power map of a fictitious Pressurized Water Reactor (PWR) with respect to a scalar criterion, a power map being the spatial distribution of power inside the reactor core. The optimality score for a given power map is computed using a nonlinear function of a distance measure between the considered power map and a reference power map. First, a dimension reduction will be applied on the power maps dataset. Then, we will benchmark several Bayesian Optimization algorithms in terms of their regret performance.

Keywords. Applied statistics, Bayesian Optimization, Dimensionality reduction.

1 Introduction

In a Nuclear Power Plant (NPP), power is delivered by the fuel inside the reactor core: the fission reaction inside the fuel realeases heat power that is absorbed by the water from the

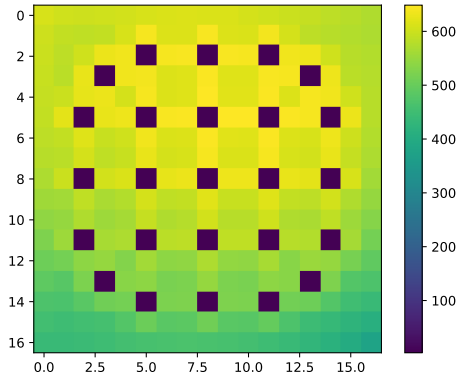
primary circuit. A reactor core in a PWR is composed of fuel assemblies, each assemblies contains 264 fuel rods and 24 tubes that contains the control rods and guide tubes. The fuel assemblies is then arranged in a 17×17 square lattice. Figure 2.4 of Jacquemain (2015) show a diagram of a fuel assembly. The spatial distribution of heat power inside the reactor core is coined power map and is critical for the exploitation of the reactor and its safety. The increase of computational power since the mid 20th century allows engineers to use numerical simulators to study the fission reaction inside the reactor core and hence computing power maps. During studies for designing the reactor core and assessing the safety of the NPP, several power maps are generated to explore different scenarios during the lifetime of the NPP. For each power map, a multiphysics computation may be done to validate the design and the safety of the NPP. In general, the multiphysics simulation is way costlier in computation time and resources than the simulation of a power map. It is thus preferable to minimize the number of multiphysics simulation to be done. In this paper, we will benchmark several Bayesian optimization (BO) algorithms Wang et al. (2023) with respect to an optimality score, this optimality score will emulate the costly multiphysics computation. The benchmark will be done on a dataset of radial power maps of a mock-up PWR reactor core simulated using the neutronics simulation code APOLLO3 developed in CEA Schneider et al. (2016).

2 Description of the data

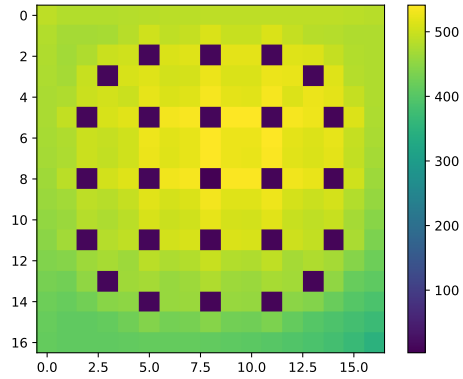
The data used in this paper are $N = 1080$ power maps $(\mathbf{P}_k)_{1\leq k\leq N}$ of a specific fuel assembly of a mock-up PWR reactor core. A power map $\mathbf{P} = (P_{ij})_{1\leq i,j\leq 17}$ is a matrix of dimension 2 composed of $d = 17\times 17 = 289$ values P_{ij} with $P_{ij} \in \mathbb{R}^+$. Hence, we can see that the data are high-dimensional. Figure 1 shows a power map observation as well as descriptive statistics. Note that some locations have always zero values: this corresponds to the positions of the guide tubes, where the neutron absorber rods are introduced to control the nuclear fission reaction. Hence, the number of variables that are informative for our problem for each power map is $17 \times 17 - 25 = 264$. The optimality score for each map is designated by the function g such that for a power map $\mathbf{P}$ we have:

$$g(\mathbf{P}) = \frac{\|\mathbf{P} - \mathbf{P}_{\text{ref}}\|_\infty - m}{\sigma} + \sin\left(\pi \frac{\|\mathbf{P} - \mathbf{P}_{\text{ref}}\|_\infty - m}{\sigma}\right)^2 , \tag{1}$$

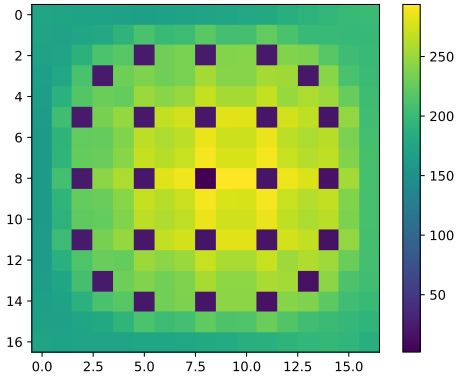
where $m = \frac{1}{N}\sum_{k=1}^{N} \|\mathbf{P}_k - \mathbf{P}_{\text{ref}}\|_\infty$ and $\sigma^2 = \frac{1}{N}\sum_{k=1}^{N} (\|\mathbf{P}_k - \mathbf{P}_{\text{ref}}\|_\infty - m)^2$ and $\mathbf{P}_{\text{ref}}$ is a reference power map consisting of a uniform spatial distribution of the mean total power computed over the dataset, with the total power of a power map defined as the sum of the power in each cell of a power map. The optimality score was designed to emulate a multiphysics simulation of a PWR reactor core with a power map as an input. It is designed to be highly nonlinear in order to reproduce the behavior of chained or coupled numerical simulation between neutronics, thermohydraulics and mechanics.



(a) Example of a power map from the dataset of the mock-up PWR reactor core



(b) Empirical mean power map over the 1080 observations



(c) Empirical standard deviation power map over the 1080 observations

Figure 1: Descriptive statistics of the power maps dataset.

The problem to be solved in this article is to determine $\mathbf{P}_{\max}$ such that:

$$\mathbf{P}_{\max} = \underset{1 \leq k \leq N}{\operatorname{argmax}}\, g(\mathbf{P}_k) \ . \tag{2}$$

The main challenge of (2) is to find $\mathbf{P}_{\max}$ with only the knowledge of $(y_k = g(\mathbf{P}_k))_{k \in \mathcal{I}}$ where $\mathcal{I} \subset \{1, \ldots, N\}$ with $|\mathcal{I}|$ the smaller possible. Indeed, we want to find the solution of our optimization problem with the less number of multiphysics computation possible, here emulated by the optimality criterion g. The mathematical tools used to solved this problem will be at first a dimensionality reduction step in order to compress the information given by the 264 power values of each power map and then the use of Bayesian optimization techniques. Bayesian optimization algorithms are designed to optimize a costly objective function with the less number of calls to the function, these methods are adapt to our problem. These two steps are described in the next section.

3 Dimensionality reduction

The first step in our study is to reduce the dimensionality of our data, the most popular method for achieving this task is the Principal Component Analysis (PCA) Jackson (1991). Indeed, its popularity is due to its simplicity of computation (it is a convex optimization problem) and its easy intrepretation of the results. Define $\bar{\mathbf{P}} = 1/N \sum_{k=1}^{N} \mathbf{P}_k$ the empirical mean power map. Using linear algebra we can express a power map $\mathbf{P}$ as follow:

$$\mathbf{P} = \bar{\mathbf{P}} + \sum_{k=1}^{N} \alpha^{(k)} \mathbf{P}^{(k)} \ , \tag{3}$$

where $\mathbf{P}^{(k)}$ are the eigenvectors of the covariance matrix of the observed power maps $(\mathbf{P}_k)_{1 \leq k \leq N}$, ranked by decreasing eigenvalues order and

$$\alpha^{(k)} = \langle \mathbf{P}^{(k)}, \mathbf{P} - \bar{\mathbf{P}} \rangle \ ,$$

The main idea of PCA is to truncate the sum in Equation (3) at $M < N$:

$$\mathbf{P} \approx \bar{\mathbf{P}} + \sum_{k=1}^{M} \alpha^{(k)} \mathbf{P}^{(k)} \ , \tag{4}$$

The choice of M is governed by the part of variance explained by the truncated sum:

$$v(M) = 1 - \sum_{k=1}^{M} (\alpha^{(k)})^2 \ , \tag{5}$$

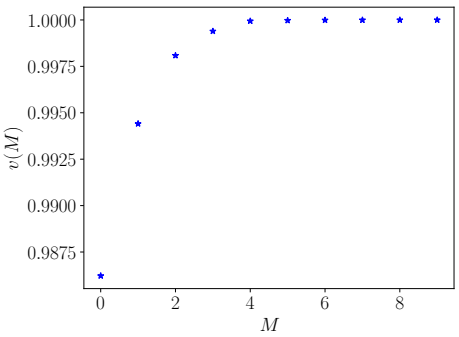


Figure 2: Evolution of the explained variance $v(M)$ with respect to M

Figure 2 show the evolution of the explained variance $v(M)$ with respect to M. We can observe that the 5 first principal components explain more than 99% of the variance. Hence,

the dimension of data can be drastically reduced from 264 to 5. It is thus possible to express the power maps by a low-dimensional representation thanks to the following application:

$$\varphi : \mathbf{P} \mapsto (\alpha^{(i)})_{1 \leq i \leq 5} \tag{6}$$

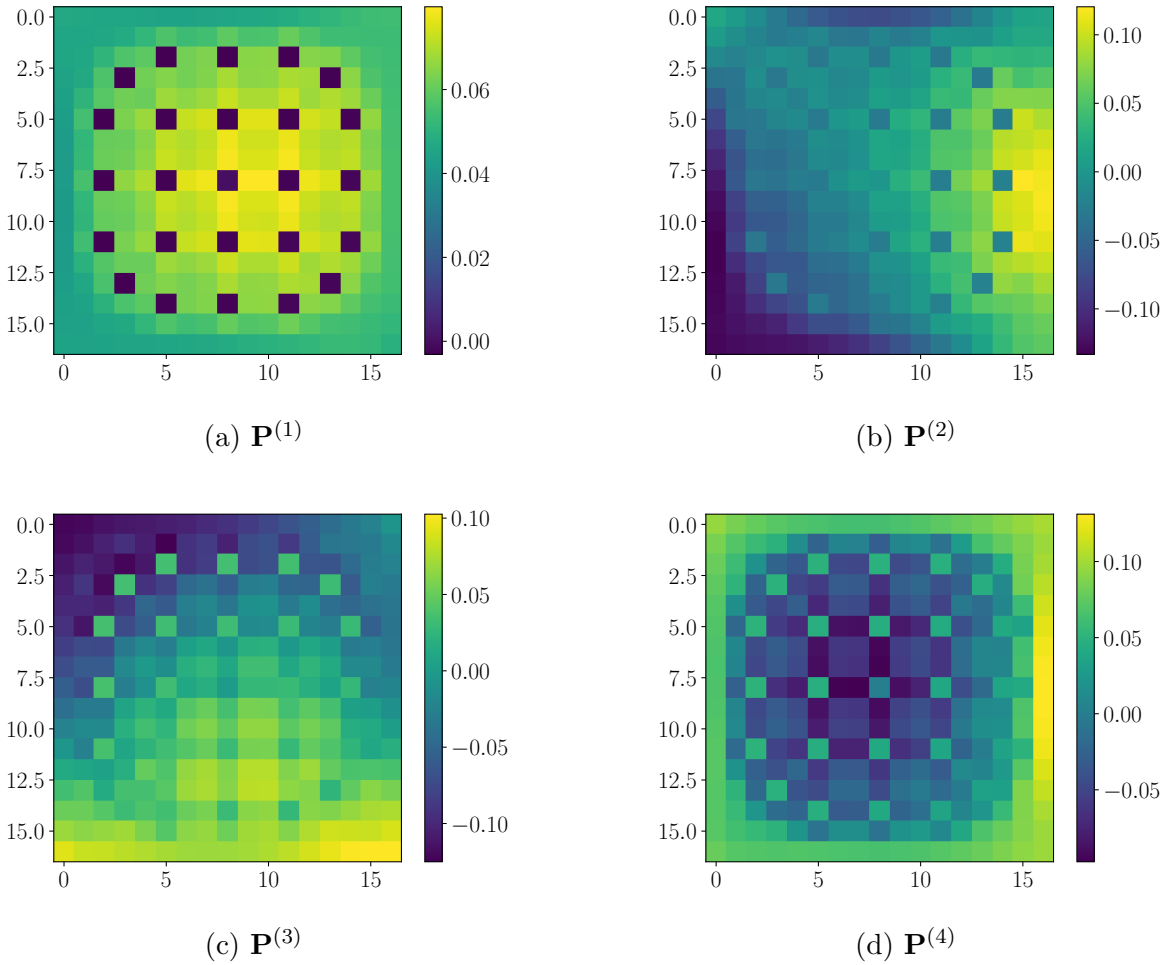
The optimization problem defined in Equation (2) is then reformulate as follow:

$$\mathbf{P}_{\max} = \underset{1 \leq k \leq N}{\operatorname{argmax}} \, \tilde{g}(\boldsymbol{\alpha}_k) \ , \tag{7}$$

where $\boldsymbol{\alpha}_k = \varphi(\mathbf{P}_k) = (\alpha_k^{(1)}, \ldots, \alpha_k^{(M)})$ is the low-dimensional representation of the power map $\mathbf{P}_k$ and $\tilde{g}$ is defined by

$$\tilde{g}(\boldsymbol{\alpha}_k) = g(\bar{\mathbf{P}} + \sum_{k=1}^{M} \alpha_k^{(k)} \mathbf{P}^{(k)}) \ .$$

The main advantage of PCA is its capacity to give visual interpretation of high-dimensional



(a) $\mathbf{P}^{(1)}$ (b) $\mathbf{P}^{(2)}$

(c) $\mathbf{P}^{(3)}$ (d) $\mathbf{P}^{(4)}$

Figure 3: Graphical representation of the four first eigenvectors of the PCA decomposition defined in Equation (4).

data. Figure 3 shows the four first eigenvector of the PCA decomposition. The first eigenvector is associated to high power values in the center of the power map. The second and third eigenvectors are associated to power values on the right and on the bottom of the power map, while the fourth eigenvector concerns the opposite variations of the power between the center and the edges of the power map.

The next section will be dedicated to the benchmark of several Bayesian optimization techniques for solving the optimization problem defined in Equation (7).

4 Bayesian optimization

Bayesian optimization aims at optimize objective functions that are costly to evaluate. The main idea is to incorporate prior belief into the objective function in the form of a Gaussian process prior in order to guide sampling of new computations of the costly function, and thus achieving a good compromise between exploration & exploitation of the computational budget. First, we will consider a Gaussian process (GP) $\tilde{G}$ for emulating the optimization criterion $\tilde{g}$. The GP is characterized by a prior mean $\mu(\cdot)$ and a covariance function $c(\cdot,\cdot)$. For $1 \leq n \leq N$, we define $\mathcal{I}_n \subset \{1,\dots,N\}$ a subset of indices of cardinal n and $\mathcal{D}_n = (\boldsymbol{\alpha}_k, \tilde{g}(\boldsymbol{\alpha}_k))_{k\in\mathcal{I}_n}$ the dataset of observations of the optimization criterion. The choice of the next computation is done through the help of the optimization of an acquisition function: these are auxillary functions evaluated using the posterior distribution of the objective function. Several acquistion functions exist in the literature. We will benchmark in this article one of the most popular one, Expected Improvement (EI) Močkus (1975):

$$\begin{aligned} \mathrm{EI}(\boldsymbol{\alpha};\mathcal{D}_n) &= \mathbb{E}_{\tilde{G}}\left[\left(\tilde{G}(\boldsymbol{\alpha}) - \tilde{G}^*\right)_+\right] \\ &= (\mu_n(\boldsymbol{\alpha}) - \tilde{G}^*)\Phi\left(\tfrac{\mu_n(\boldsymbol{\alpha})-\tilde{G}^*}{\sigma_n(\boldsymbol{\alpha})}\right) + \sigma_n(\boldsymbol{\alpha})\Phi\left(\tfrac{\mu_n(\boldsymbol{\alpha})-\tilde{G}^*}{\sigma_n(\boldsymbol{\alpha})}\right) \end{aligned} \tag{8}$$

where $\tilde{G}^* = \mathrm{argmax}_{k\in\mathcal{I}_n}\, \tilde{g}(\boldsymbol{\alpha}_k)$, $\mu_n(\cdot)$ and $\sigma_n(\cdot)$ are respectivly the mean and standard deviation of $\tilde{G}$ conditionally to $\mathcal{D}_n$. Note that the EI acquisition function can be computed analytically without any need of Monte-Carlo sampling. The EI is optimized to determine the next datapoint denoted by $\boldsymbol{\alpha}_{(n+1)}$ such that:

$$\boldsymbol{\alpha}_{(n+1)} = \underset{k\in\{1,\dots,N\}/\mathcal{I}_n}{\mathrm{argmax}}\ \mathrm{EI}(\boldsymbol{\alpha}_k;\mathcal{D}_n)$$

Another classical Bayesian optimization algorithm is denoted by Thompson sampling. The core principles of this method is to sample $\tilde{G}_n \sim (\tilde{G}|\mathcal{D}_n)$ and to solve the optimization problem:

$$\boldsymbol{\alpha}_{(n+1)} = \underset{k\in\{1,\dots,N\}/\mathcal{I}_n}{\mathrm{argmax}}\ \tilde{G}_n(\boldsymbol{\alpha}_k)$$

The next section is dedicated to a benchmark of these two Bayesian optimization method on the power maps dataset.

5 Numerical experiments and results

The two Bayesian optimization algorithms presented in Section 4 will be benchmarked on a dataset of 1080 simulated power maps. Also, a random sampling strategy will also be performed in order to compare the Bayesian optimization strategies with respect to a naive strategy of sampling at random the power maps in the dataset. We choose $M = 4$ for the low-dimensional representation of the power maps. Before starting the Bayesian optimization algorithm, 5 power maps are sampled at random in the dataset to train and learn the hyper-paremeters of the Gaussian process, which are then fixed during the rest of the algorithm. 100 replications of each strategies are thus performed to take into account the randomness of the sampling of the 5 first power maps and the one from the sampling of the Gaussian process posterior in the Thompson sampling. The different strategies are quantitatively benchmarked by their regret at step n:

$$R_n = g(\mathbf{P}_{\max}) - \max_{k \in \mathcal{I}_n} g(\mathbf{P}_k) \ ,$$

The numerical results of the benchmark is presented in Figure 4. The best strategy given by

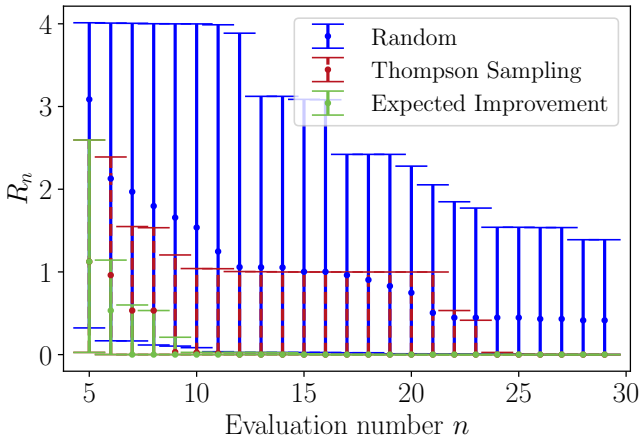


Figure 4: Results from the benchmark of 100 replications of each strategy. The vertical bars represent the interquantiles ranges between the 10% and 90% quantile of the regret R_n at each step n.

our results is the EI strategy. Indeed the regret reach the value 0 for all the 100 replications after only 16 computations of the optimization criterion g, while Thompson sampling need 25 computations. However, it is important to notice that the two Bayesian optimization algorithms outperformed very well a naive random strategy.

6 Conclusion & perspectives

In this article, a toy UQ problem of designing a optimal power map is studied. A optimality criterion is associated to each power map to emulate a costly multiphysics simulation taking a power map as an input. Given the high dimensionality of the power map data, a dimension

reduction procedure by PCA is performed to represent each power map by a low-dimensional vector. After this step, two Bayesian optimization algorithms are benchmarked with respect to their regret values. The Expected Improvement algorithm shows the best performances. A perspective of this work could be to improve the dimensionality reduction by enforcing physics constraints. For instance, the power maps have always positive values, they thus lies in a specific quadrant of $\mathbb{R}^{264}$. More complex physics constraints based on neutronics could be implemented. Another perspective is to implement parallel EI algorithms in the same fashion as in Ginsbourger et al. (2008), Wang et al. (2020). Indeed, parallel computing is now common in industrial practices and thus new Bayesian optimization algorithms needs to be developed for the case when the costly computer simulation is ran in parallel.

7 Acknowledgments

The author adress its warmful thanks to Mr. Blaise MATHON for the generation of the power maps dataset in APOLLO3. As well as Dr. Nathan GRENIER, Mr. Thibaut LOPEZ and Mr. Damien RAGUENES for the fruitful discussions, without which this article would not have been possible.

References

Ginsbourger, D., Le Riche, R. & Carraro, L. (2008), A Multi-points Criterion for Deterministic Parallel Global Optimization based on Gaussian Processes, Technical report.

Jackson, J. (1991), *A user's guide to principal components*, Wiley series in probability and mathematical statistics, Wiley.

Jacquemain, D. (2015), *Nuclear Power Reactor Core Melt Accidents. Current State of Knowledge*, EDP Sciences.

Močkus, J. (1975), On bayesian methods for seeking the extremum, *in* G. I. Marchuk, ed., 'Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 400–404.

Schneider, D., Dolci, F., Gabriel, F., Palau, J.-M., Guillo, M. & Pothet, B. (2016), APOLLO3® CEA/DEN deterministic multi-purpose code for reactor physics analysis, *in* 'PHYSOR 2016 – Unifying Theory and Experiments in the 21st Century', Sun Valley, United States. APOLLO3® is a registered trademark of CEA.

Wang, J., Clark, S. C., Liu, E. & Frazier, P. I. (2020), 'Parallel bayesian global optimization of expensive functions', *Operations Research* **68**(6), 1850–1865.

Wang, X., Jin, Y., Schmitt, S. & Olhofer, M. (2023), 'Recent advances in bayesian optimization', *ACM Comput. Surv.* **55**(13s).