

Estimation of seismic fragility curves by sequential design of experiments

- **Author:** Clément Gauchy (CEA Saclay, École Polytechnique)
- **Supervisors:** Josselin Garnier (École Polytechnique), Cyril Feau (CEA Saclay)
- **PhD starting date:** November 2019

1 Abstract

Seismic probabilistic risk assessment studies consist in evaluating the probabilities of failure of mechanical structures when submitted to seismic ground motions. These studies are often concentrated on fragility curve estimation. The fragility curve is the probability of failure of the structure conditionally to a seismic intensity measure. However, its estimation requires computer experiments involving huge computation time. Such a computational burden makes crude Monte Carlo methods untractable, fragility curves estimation must then be economical in terms of sample size. We propose an algorithm of sequential planning of experiments by supposing a Gaussian process prior on the output of the mechanical computer model. The algorithm proposed is inspired from Sequential Uncertainty Reduction (SUR) methodology [1]. This type of Bayesian sequential design of experiments has been already developed for probability of failure estimation [1], quantile estimation [2], and excursion set estimation [3]. The sequential design of experiments algorithm will be applied on a numerical simulation of a piping system mock-up of a French Pressurized Water Reactor [4] implemented in CAST3M [5].

References

- [1] Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, and Emmanuel Vazquez. Sequential design of computer experiments for the estimation of a probability

of failure. *Statistics and Computing*, 22(3):773–793, April 2011. doi: 10.1007/s11222-011-9241-4.

- [2] Aurélie Arnaud, Julien Bect, Mathieu Couplet, Alberto Pasanisi, and Emmanuel Vazquez. Évaluation d’un risque d’inondation fluviale par planification séquentielle d’expériences. *42èmes Journées de Statistique*, 2010.
- [3] Dario Azzimonti, David Ginsbourger, Clément Chevalier, Julien Bect, and Yann Richet. Adaptive design of experiments for conservative estimation of excursion sets. *Technometrics*, 63(1):13–26, 2021. doi: 10.1080/00401706.2019.1693427.
- [4] F. Touboul, P. Sollogoub, and N. Blay. Seismic behaviour of piping systems with and without defects: experimental and numerical evaluations. *Nuclear Engineering and Design*, 192(2):243 – 260, 1999.
- [5] URL <http://www-cast3m.cea.fr/>.

Off-the-grid learning of sparse mixtures from a continuous dictionary

Clément Hardy

Ecole des Ponts, EDF R&D
clement.hardy@enpc.fr

supervised by Cristina Butucea (ENSAE), Jean-François Delmas (Ecole des ponts)
and Anne Dufloy (EDF R&D) since December 2019.

In many fields such as microscopy, astronomy, spectroscopy or imaging, signals that appear naturally have the structure of a sparse mixture of features belonging to a dictionary. We consider a general non-linear model where the signal is a finite mixture of an unknown, possibly increasing, number of features issued from a continuous dictionary parameterized by a real non-linear parameter. The signal is observed with Gaussian (possibly correlated) noise in either a continuous or a discrete setup. We propose an off-the-grid optimization method, that is, a method which does not use any discretization scheme on the parameter space, to estimate both the non-linear parameters of the features and the linear parameters of the mixture.

Using tail bounds for suprema of Gaussian processes, we bound the prediction error with high probability. Assuming a separation between the non-linear parameters, our prediction error bound is up to \log –factors similar to the rates attained by the Lasso predictor in the linear regression model. We also establish convergence rates that quantify with high probability the quality of estimation for both the non-linear parameters of the features and the linear parameters of the mixture.

Model Order Reduction and Bayesian Optimization for MDO problems

Inês Cardoso^{a,b} - 1st year PhD student (start date October 1st, 2021)

Supervisors: Sylvain Dubreuil^a, Michel Salaün^b, Nathalie Bartoli^a, Christian Gogu^b

a. DTIS, ONERA, Université de Toulouse, Toulouse, France

b. Université de Toulouse, CNRS, UPS, INSA, ISAE, Mines Albi, Institut Clément Ader (ICA),
Toulouse, France

Abstract:

Multidisciplinary analysis and optimization have become practical tools in aircraft design, since they allow to take into account the effect of several disciplines and their interaction in the overall performance of the vehicle [1]. This work proposes a framework for multidisciplinary optimization, by combining the approximation of the disciplinary solvers by Proper Orthogonal Decomposition and Interpolation (DPOD+I) models [2] with the Efficient Global Multidisciplinary Design Optimization (EGMDO) algorithm [3]. The DPOD+I approximation allows to address the problem of high-dimensional coupling variables, by using a surrogate training strategy based on model order reduction and Gaussian process interpolation. The obtained surrogates may then be enriched throughout the optimization, to assure a given precision. It is shown that this implementation allows for an important reduction of the number of expensive solver calls, compared to when no approximation is made. The combination of this method with the EGMDO algorithm further reduces the computational cost by using an Expected Improvement criterion to preferentially improve the surrogates in the areas of the design space found relevant for the search of the global optimum. In the context of an aeroelastic test case, which couples two linear disciplinary models, it is seen that, by combining the two methodologies, very few disciplinary solver enrichments are needed.

References

- [1] J. Jasa, J. Hwang, and J. Martins. “Open-source coupled aerostructural optimization using Python”. In: *Structural and Multidisciplinary Optimization* 57 (2018), pp. 1815–1827. DOI: 10.1007/s00158-018-1912-8.
- [2] G. Berthelin, S. Dubreuil, M. Salaün, N. Bartoli, and C. Gogu. “Disciplinary Proper Orthogonal Decomposition and Interpolation for the resolution of parametrized Multidisciplinary Analysis”. In: *International Journal for Numerical Methods in Engineering* (2022). Accepted Author Manuscript. DOI: 10.1002/nme.6981.
- [3] S. Dubreuil, N. Bartoli, C. Gogu, and T. Lefebvre. “Towards an efficient global multidisciplinary design optimization algorithm”. In: *Structural and Multidisciplinary Optimization* 62 (2020), pp. 1–27. DOI: 10.1007/s00158-020-02514-6.

ADAPTIVE IMPORTANCE SAMPLING FOR RELIABILITY ASSESSMENT OF A PIECEWISE DETERMINISTIC MARKOV PROCESS

Guillaume CHENNETIER^{1,2} & Hassane CHRAIBI¹ & Anne DUTFOY¹ & Josselin GARNIER²

¹ EDF Lab Paris-Saclay, Boulevard Gaspard Monge, 91120 Palaiseau, France.

guillaume.chennetier@edf.fr ; hassane.chraibi@edf.fr ; anne.dutfoy@edf.fr

² CMAP, École Polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France.

guillaume.chennetier@polytechnique.edu ; josselin.garnier@polytechnique.edu

Start date of the PhD: April 15, 2021.

Abstract. We are interested in assessing the reliability of industrial systems modeled by piecewise deterministic Markov processes (PDMPs). Mathematically, this translates into estimating the probability that a PDMP trajectory reaches a certain region of its state space before a fixed date. As the failures of the modeled systems are very rare, we cannot estimate the probability of these failures by standard Monte-Carlo. Instead, we propose an adaptive importance sampling method designed for PDMPs. The first axis of the method consists in the construction of a family of importance distributions based on the approximation of the committor function of the PDMP. The second one consists in jointly determining a good candidate within this family and estimating the probability of interest using a cross-entropy procedure with recycling of past samples.

Keywords. PDMP, importance sampling, rare event simulation, cross-entropy, fault tree analysis.

References

- [1] H. Chraibi, A. Dutfoy, T. Galtier et J. Garnier (2019), On the optimal importance process for piecewise deterministic Markov process. *ESAIM: PS.* 23:893–921.
- [2] H. Chraibi, J.-C. Houdebine et A. Sibler (2016), PyCATSHOO: Toward a new platform dedicated to dynamic reliability assessments of hybrid systems. *Proceedings PSAM13.*
- [3] M. H. A. Davis (1984), Piecewise-Deterministic Markov Processes: A general class of non-diffusion stochastic models. *J. R. Stat. Soc. B* 46:353–388.
- [4] B. Delyon et F. Portier (2018), Asymptotic optimality of adaptive importance sampling. *NIPS 2018.*
- [5] R. Rubinstein (1999), The Cross-Entropy method for combinatorial and continuous optimization. *Methodology and Computing in Applied Probability* 1:127–190.

DEFONTAINE, Théo, CECI, CERFACS/UMR5318 CNRS, PhD started in November, 2020

RICCI, Sophie, CECI, CERFACS/UMR5318 CNRS

LAPEYRE, Corentin, CECI, CERFACS/UMR5318 CNRS

Title:

Flood forecasting with Machine Learning in a scarce data layout

Abstract:

Floods are one of the major natural disasters occurring in the world, with climate change increasing their occurrence and severity. Reliable flood forecasting models are needed to enable better reactions of emergency relief services. This work reflects the capacities of Machine Learning models to improve discharge forecast results from empirical lag and route models based on hourly measured water level at gauge stations on the Garonne River. The scarce flood data (30000 points) gathered over the last 15 years is the stumbling block of this problem. Several learning algorithms are tested to predict floods in Toulouse at a 6-hour lead time from upstream stations providing hourly observations. A Linear Regression, a Gradient Boosting Regressor (Machine Learning) and a MultiLayer Perceptron (Neural Network, Deep Learning) are compared, using various strategies for learning, validating and predicting. Preliminary results show that these data-driven strategies score better than the reference empirical lag and route model. This paper highlights how AI algorithms allow to improve the accuracy of flood forecasts and how the layout of the limited volume of data influences their performance.

Shapley effect estimation in reliability-oriented sensitivity analysis with dependent inputs by importance sampling

Julien DEMANGE-CHRYST^{a,b}, François BACHOC^b, Jérôme MORIO^a

^a*ONERA/DTIS, Université de Toulouse, F-31055 Toulouse, France*

^b*Institut de Mathématiques de Toulouse, UMR5219 CNRS, 31062 Toulouse, France*

Keywords: Reliability-oriented sensitivity analysis, Target sensitivity analysis, Rare event estimation, Dependent inputs, Shapley effects, Importance sampling, Nearest-neighbour approximation.

Reliability-oriented sensitivity analysis aims at combining both reliability and sensitivity analyses by quantifying the influence of each input variable of a numerical model on a quantity of interest related to its failure. In particular, target sensitivity analysis focuses on the occurrence of the failure, and more precisely aims to determine which inputs are more likely to lead to the failure of the system. The Shapley effects are quantitative global sensitivity indices which are able to deal with dependent input variables. They have been recently adapted to the target sensitivity analysis framework. In this contribution, we investigate importance-sampling-based estimation schemes of these indices which are more efficient than the existing ones when the failure probability is small. Moreover, an extension to the case where only an i.i.d. input/output N -sample distributed according to the importance sampling auxiliary distribution is available is proposed. This extension allows to estimate the Shapley effects only with a data set distributed according to the importance sampling auxiliary distribution stemming from a reliability analysis without additional calls to the numerical model. In addition, we study theoretically the absence of bias of some estimators as well as the benefit of importance sampling, and finally, realistic test cases show the practical interest of the proposed methods.

PhD dates: 01/10/2021 - 30/09/2024

Sequential reliability analysis for offshore wind turbine fatigue assessment

E. FEKHARI

EDF R&D, Université Nice Côte d'Azur

Supervisor(s): B. Iooss (EDF R&D, Université Nice Côte d'Azur), V. Chabridon (EDF R&D), J. Muré (EDF R&D)

PhD duration: Jan. 2021 - Jan. 2024

Address: EDF Lab, 6 Quai Watier, 78401 Chatou, France

E-mail: elias.fekhari@edf.fr

Abstract:

Offshore wind turbines (OWT) are a growing part of the global electricity generation mix. These assets are subject to uncertain environmental conditions (wind and tidal), making long-term investment decisions riskier. This rapidly growing industry is regulated by international standards, imposing resistance bounds on the structures. The uncertain inputs related to the environment and the system itself can be propagated through a costly numerical model simulating the physics of the OWT. Then, one can retrieve an output random variable (e.g., the distribution of the material fatigue on the structure). To ensure that this variable of interest respects the standard resistance limit, one can estimate a threshold exceedance probability [5]. In this work, the goal is to combine targeted screening methods with sequential algorithms for reliability analysis to better estimate this risk measure using a small number of simulations. First, target screening using HSIC-based statistical tests [3] allows to narrow down the number of relevant input variables, reducing the domain explored. Then, an iterative procedure selects simulation points by optimizing over the input domain a specific acquisition function (dedicated to the reliability problem) exploiting information from a surrogate model [4] (often a Gaussian process regression). This work will explore and compare newly developed entropy-based acquisition functions[1] and recent methods for sampling in a constrained domain [2] to improve the reliability assessment.

References

- [1] D.A. Cole, R.B. Gramacy, J.E. Warner, G.F. Bomarito, P.E. Leser, and W.P. Leser. Entropy-based adaptive design for contour finding and estimating reliability. *Journal of Quality Technology*, 0:1–18, 2022.
- [2] C. Huang, V.R. Joseph, and D.M. Ray. Constrained minimum energy designs. *Statistics and Computing*, 31:80, 2021.
- [3] A. Marrel and V. Chabridon. Statistical developments for target and conditional sensitivity analysis: Application on safety studies for nuclear reactor. *Reliability Engineering & System Safety*, 214:107711, 2021.
- [4] M. Moustapha, S. Marelli, and B. Sudret. Active learning for structural reliability: Survey, general framework and benchmark. *Structural Safety*, 96:102174, 2022.
- [5] D. Wilkie and C. Galasso. Gaussian process regression for fatigue reliability analysis of offshore wind turbines. *Structural Safety*, 88:102020, 2021.

FELLMANN Noé

CENTRALE LYON-Institut Camille Jordan (UMR5208)-IFP Energies nouvelles

Encadrants :

HELBERT Céline (ECL-ICJ)

BLANCHET-SCALLIET Christopette (ECL-ICJ)

SPAGNOL Adrien (IFPEN)

SINOQUET Delphine (IFPEN)

Start of PhD : 01/12/2021

Sensitivity Analysis on excursion sets

In the black-box framework, Sensitivity Analysis (SA) aims at studying and quantifying the influence of the inputs on the variability of the output. For applications as robust optimization under constraints or robust inversion, Sensitivity Analysis can be Goal Oriented¹ (GOSA) in order to find the inputs that have the most impact on the constraints compliance. In such applications, the presence of uncertainties can be modelled by distinguish the inputs in two groups: the deterministic ones and the uncertain ones. Most SA methods use GOSA on the deterministic inputs. We aim at doing GOSA on the uncertain inputs. To do so we propose to carry out GOSA with set-valued outputs being excursion sets i.e. sets where the constraints are respected. However, most SA methods deal with scalar or vectorial outputs. A set-valued output space makes the adaptation of classical sensitivity analysis methods nontrivial. We propose three main ideas to carry out this adaptation. The first relies on random sets theory². The second adapts universal indices defined for general metric output space³ to set-valued outputs. The last one uses kernels-based sensitivity indices⁴ applied on sets.

¹ Adrien Spagnol, 'Kernel-Based Sensitivity Indices for High-Dimensional Optimization Problems' (phdthesis, Université de Lyon, 2020) <<https://tel.archives-ouvertes.fr/tel-03173192>>.

² 'Expectations of Random Sets', in *Theory of Random Sets*, ed. by Ilya Molchanov, Probability and Its Applications (London: Springer, 2005), pp. 145–94 <https://doi.org/10.1007/1-84628-150-4_2>.

³ Jean-Claude Fort, Thierry Klein, and Agnès Lagnoux, 'Global Sensitivity Analysis and Wasserstein Spaces', *SIAM/ASA Journal on Uncertainty Quantification*, 9.2 (2021), 880–921 <<https://doi.org/10.1137/20M1354957>>.

⁴ Sébastien da Veiga, 'Kernel-Based ANOVA Decomposition and Shapley Effects - Application to Global Sensitivity Analysis', 2021 <<https://hal.archives-ouvertes.fr/hal-03108628>> .

Robustness assessment using quantile-constrained Wasserstein projections

MAROUANE IL IDRISSE

EDF R&D, Institut de Mathématiques de Toulouse (IMT), SINCLAIR AI Lab

Supervisor(s): J.-M. Loubes (IMT), F. Gamboa (IMT), N. Bousquet (EDF R&D, SINCLAIR) and B. Iooss (EDF R&D, SINCLAIR)

PhD expected duration: Mar. 2021 - Mar. 2024

Address: EDF Lab Chatou, 6 Quai Watier, 78401 Chatou

E-mail: marouane.il-idrissi@edf.fr

Abstract:

Robust interpretation of model characteristics is widely recognized as a necessary task for both complex numerical models based on structural equations and models defined by complex patterns learned from data. Interpretive analyses both rely on assessing the influence of input information on the predicted output variable, and the desired robustness must be defined with respect to a possible misspecification of this information, as so-called covariate shift in machine learning. The study of numerical models is most often based on sensitivity analysis involving a probabilistic representation of input information, whereas the indices that help interpretability in machine learning are directly constructed from the data. In order to provide a generic and understandable framework for robustness studies, we define perturbations of input information relying on marginal quantile constraints and Wasserstein loss minimization. Associated to regularity constraints allowed by isotonic polynomial approximations, this approach allows to obtain discrete or smooth perturbations of the initial input information, while preserving the correlation structure of the latter. Numerical experiments conducted on real case-studies belonging to both fields, using a dedicated package, highlight the computational feasibility of such studies and obtaining robust scores that facilitate local and global interpretability.

Short biography – After graduating from ENSAI and Rennes 1 University in 2020, I started a CIFRE PhD track in 2021 at EDF R&D and Institut de Mathématiques de Toulouse, working on the development of interpretability methods for ML models. My research interests are at the crossroads between sensitivity analysis and explainable artificial intelligence methods, and more specifically applied cooperative game theory and probability measure perturbations.

Using Generative Adversarial Networks to constrain inverse problems resolution.

July 8, 2022

- **Name:** Matthéo SALDANHA
- **Affiliation:** CEA/ Orléans University
- **Supervisors:** Isabelle ABRAHAM & Cécile HABERSTICH (CEA) - Bruno GALERNE (Orléans University)
- **PhD:** started in October 2021.

In this work, we are interested in estimating physical characteristics of a deuterium sphere imploded by inertial confinement. Formally, this object is represented by a 2D or 3D mesh, denoted $X \in \mathbb{R}^n$ ($n \approx 10^6 - 10^9$, it is the number of voxels/pixels in the mesh). The physical characteristics are obtained from X-ray measurements. In the following, these measurements are represented by 2D images and denoted by $Y \in \mathbb{R}^m$ by ($m \approx 10^6$, it is the number of pixels). In this work, we are interested in solving the following inverse problem, it consists in finding \tilde{X} such that

$$\tilde{X} = \arg \min_{X \in \Omega} \|\mathcal{M}(X) - Y\| \quad (1)$$

where

- Ω is the space of acceptable physical objects.
- \mathcal{M} is an operator.

Solving inverse problems with very high-dimensional images Y and meshes X is difficult because the solution may be unstable. Moreover, often $n > m$ that leads to a non-unique solution. To overcome this issue, we model Ω using Generative Adversarial Networks (GANs), whose are neural networks generating images. The generated images are created by the GAN using existing examples $\{X_i\}_{i=1}^N$ of realistic objects and should represent the characteristics of the distribution of $\{X_i\}_{i=1}^N$. In this presentation, we will apply this methodology to solve the inverse problem on a toy case.

Gaussian processes indexed by clouds of points: a study

BABACAR SOW

Ecole Nationale Supérieure des Mines de Saint-Etienne

Supervisor(s): R. Le Riche (CNRS LIMOS), J. Pelamatti (EDF R&D), S. Zannane (EDF R&D) and M. Keller (EDF R&D)

PhD expected duration: Nov. 2021 - Oct. 2024

Address: Ecole Nationale Supérieure des Mines de Saint-Etienne, LIMOS, France

E-mail: babacar.sow@emse.fr

Abstract:

Optimizing the electrical production of a wind farm starts with finding the number and the optimal positions of the turbines that minimize the losses caused by wake interactions. The optimization variables are the number and the set of positions of the turbines and take the form of a cloud of points. In a classic way, this work deals with replacing the time-consuming simulations by a surrogate model, such as a Gaussian process, for later use in a Bayesian optimization framework.

This study addresses the problem of defining valid and efficient covariance kernels over variable-size clouds of points.

A common way is to compose conditionally negative definite metrics with specific functions such as Radial Basis Functions (RBF). This ensures the kernel validity, in terms of semi-positive definiteness.

Another way consists in associating measures (empirical probabilities) to the clouds of points through a deterministic mapping [1] and to define a kernel between laws.

We test different kernels built on the above principles.

We also propose an alternative kernel constructed by considering the vector of features relevant to our clouds such as the mean and eigen components of the covariance matrix.

The above mathematically valid kernels are compared in terms of prediction performances. For this, we create a benchmark of relevant test functions over clouds of points, including one that mimics wind farms power production, and a design of experiment problem.

Key words: Point clouds, Kernel, Hilbertian, Maximum Mean Discrepancy, Characteristic RKHS, Wasserstein distance, Gaussian Processes.

References

- [1] Tony Jebara and Risi Kondor. Bhattacharyya and expected likelihood kernels. In Learning theory and kernel machines, pages 57–71. Springer, 2003.

Short biography – Babacar Sow has received his Masters degree from Telecom Paris and Institut Polytechnique de Paris in Data Science And Applied Statistics. He currently pursues a PhD with Mines Saint-Etienne in collaboration with EDF. This work is funded by the ANR SAMOURAI project.

Sensitivity to statistical estimation uncertainties and probabilistic model identification

PhD expected duration : May 2021-May 2024

Charles Surget^{a,b}, Sylvain Dubreuil^a, Jérôme Morio^a, Cécile Mattrand^b, Jean-Marc Bourinet^b, Nicolas Gayton^b

^aONERA/DTIS, F-31055 Toulouse, France

^bSIGMA Clermont, F-63000 Clermont-Ferrand, France

Keywords: Small-data, Sensitivity, Trade-off.

Let us consider a black-box function $\phi : \mathcal{X} \rightarrow \mathbb{R}$. Its input \mathbf{X} is a continuous random vector drawn according to a probability density function (PDF) $f_{\mathbf{X}}$. One could be interested in estimating an expectation of a function τ of $\phi(\mathbf{X})$ by Monte Carlo Simulation (MCS):

$$\mathbb{E}_{f_{\mathbf{X}}} [\tau(\phi(\mathbf{X}))] = \int_{\mathbb{R}^d} \tau(\phi(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \approx \frac{1}{N_{\mathbf{X}}} \sum_{j=1}^{N_{\mathbf{X}}} \tau(\phi(\mathbf{X}^{(j)})). \quad (1)$$

The knowledge of the input probabilistic model is often restricted to a small $N_{\mathbf{D}}$ -sample $\tilde{\mathbf{D}} := \{\mathbf{D}^{(i)}, i = 1, \dots, N_{\mathbf{D}}\}$, with $\mathbf{D}^{(i)} \stackrel{i.i.d.}{\sim} f_{\mathbf{X}}$. Thus, the expectation is expressed as follows:

$$\mathbb{E}_{\hat{f}_{\mathbf{X}|\tilde{\mathbf{D}}}} [\tau(\phi(\mathbf{X}))] = \int_{\mathbb{R}^d} \tau(\phi(\mathbf{x})) \hat{f}_{\mathbf{X}|\tilde{\mathbf{D}}}(\mathbf{x}|\tilde{\mathbf{d}}) d\mathbf{x}, \quad (2)$$

where $\mathbf{X}^{(j)} \stackrel{i.i.d.}{\sim} \hat{f}_{\mathbf{X}|\tilde{\mathbf{D}}}$, the PDF estimate of $f_{\mathbf{X}}$. The estimator is then subject to a bi-level uncertainty source. The first one comes from the identification of $f_{\mathbf{X}}$ whereas the second one comes from the MCS sample.

The main objective here is to assess the variance of the estimator with respect to those two sources. This variance can be reduced by performing new mechanical tests and running new simulations, through a non-negligible cost. Hence, is it better to allocate the budget in the database or the MCS sample ?

The proposed work focuses on this trade-off. The issue is tackled by means of a sensitivity analysis to determine the contribution of each source. The relevance of the method is then illustrated on academic examples.