2025 RT-UQ PhD day

SAMO satellite event

April 22, 2025 in Grenoble, France

Contents

1 RT-UQ PhD student day, April 22, 2025

1.1 Plenary lecture

Chairman: Anthony Nouy

Jessica Hoffmann, Google DeepMind

Title: Epidemics on Graphs under Uncertainty

Abstract: Epidemic processes can model anything that spreads. As such, they are a useful tool for studying not only human diseases, but also network attacks, spikes in the brain, the propagation of real or fake news, the spread of viral tweets, and other processes. This talk focuses on epidemics spreading on an underlying graph. Currently, most state-of-the-art research in this field assumes some form of perfect observation of the epidemic process. This is an unrealistic assumption for many real-life applications, as the recent COVID-19 pandemic tragically demonstrated: data is scarce, delayed, and/or imprecise for human epidemics, and symptoms may appear in a non-deterministic fashion - if they appear at all. We show in this work not only that the algorithms developed previously are not robust to adding noise into the observation, but that some theoretical results cannot be adapted to this setting. In other words, uncertainty fundamentally changes how we must approach epidemics on graphs.

1.2 PhD student talks

Chairwoman: Céline Helbert

A distributional perturbation method based on the Fisher-Rao distance for robustness analysis in uncertainty quantification, Ketema Baalu Belay [et al.]

Addressing the Rashomon Effect through ranking aggregation, Sessa Claudia [et al.]

Chairman: Sébastien Da Veiga

Surrogate to Poincaré inequalities on manifolds for dimension reduction in nonlinear feature spaces, Pasco Alexandre [et al.]

Mollifiers to enhance gradient-based dimension reduction, Verdière Romain [et al.]

Variance-Informed Subspace: a Gradient-free Dimension Reduction for Adaptive Bayesian Inference, Polette Nadege [et al.]

Chairman: Olivier Roustant

Sequential transport for density estimation and its applications, Zanger Benjamin [et al.]

Learning signals defined on graphs with optimal transport and Gaussian process regression, Carpintero Perez Raphaël [et al.]

Importance Sampling in high dimension, Beh Jason [et al.]

1.3 PhD student posters

Chairman of poster blitz: Guillaume Perrin

Louis Allain - "Kernel-based uncertainty quantification of machine learning models: assessment and first advances"

Nils Baillie - "Variational inference for approximate reference priors using neural networks"

Clément Cardoen - "Moment approach for model reduction of parameter-dependent conservation laws in Wasserstein spaces"

Leonardo Chiani - "gsaot: an R package for Optimal Transport-based sensitivity analysis"

Mathis Deronzier - "Block-Additive Gaussian Processes under Monotonicity Constraints"

Baptiste Ferrere - "Generalized Hoeffding Decomposition for Models with Bernoulli Inputs"

Lisa Garcia - "From the analysis of experimental shock dynamic films to Bayesian calibration of physical models"

Lisanne Gossel - "Scale-bridging in a complex model hierarchy for development of an iron-fueled energy cycle"

Edgar Jaber - "Bayesian calibration for hybrid prognostics of steam generators clogging"

Fatima-Zahrae El-Boukkouri - "General reproducing properties in RKHS with application to derivative and integral operators"

Guerlain Lambert - "Surrogate-based active learning for Sobol' indices estimation with dependent inputs"

Exauce Luweh Adjim Ngarti - "Robust parameter estimation using variational inference and generative neural networks"

Ioannis Mavrogiannis - "Uncertainty Quantification and Sensitivity Analysis of Energetic Particles from a Neutral Beam Injection in a Nuclear Fusion Plasma"

Charles Miranda - "Optimal sampling Tensor-Train approximation of backward stochastic differential equations"

Mahamat Hamdan Nassouradine - "Prediction of physical fields under linear constraint"

Adrian Padilla Segarra - "Reconstruction of fluid flow fields from data using Gaussian process regression with physics–informed kernels"

Anthony Quintin - "Optimal experimental designs under uncertainties in fracture toughness test campaigns"

Nathan Ricard - "Analysis and improvement of the convergence of physics-informed neural networks"

Angélique Saillet - "Sensitivity analysis & implementation of a multi-fidelity approach for the biogeochemical model Eco3M-MED-CN in an 1DV configuration"

Marie Temple-Boyer - "Generic framework for decision-making models in risk analysis"

A distributional perturbation method based on the Fisher-Rao distance for robustness analysis in uncertainty quantification

Baalu Belay Ketema[†,1,2], Nicolas Bousquet[§,2], Francesco Costantino[§,1], Fabrice Gamboa[§,1], Bertrand Iooss [§,1,2], Roman Sueur [§,2]

[†] PhD student (presenting author). [§] PhD supervisors

PhD expected duration: Feb. 2023 – Feb. 2026

[1] Institut de mathématiques de Toulouse, Paul Sabatier University,
{baalu-belay.ketema,francesco.costantino,fabrice.gamboa}@univ-toulouse.fr

[2] EDF R&D,
{baalu-belay.ketema,roman.sueur,nicolas.bousquet,bertrand.iooss}@edf.fr

Abstract

Robustness analysis is a subdomain of uncertainty quantification that deals with the uncertainty propagation through a computer code G (assumed costly) of its input probability distributions. More specifically, the inputs $(X_1, ..., X_d)$ of G take uncertain values which usually correspond to physical measurements. Therefore, the uncertainty on these inputs X_i is modeled by a probability distribution f_i that is determined through physical experiments, hence f_i is itself uncertain. The goal of a robustness analysis method is then to: (a) take into account the uncertainty on f_i, this is done by defining a distributional perturbation method; (b) assess the impact on a quantity of interest (QoI) of the output $Y = G(\mathbf{X})$ of a perturbation on the inputs through robustness indices.

In this talk, we will present these two aspects of a specific robustness analysis method initially proposed in [1].

In the first part of this presentation, we will focus on the distributional perturbation method that is based on the Fisher-Rao distance on parametric families of probability distributions $\mathcal{P} = \{g_\theta\}_{\theta\in\Theta}$. This particular distance, on $\mathcal{P}$, derived from the Fisher information metric, has a geometric origin: it is the length of the shortest path connecting two points in $\mathcal{P}$. The Fisher-Rao distance presents many interesting properties for the purpose of robustness analysis. One of these properties is the universality of a (small) distance value $\delta > 0$ in two different families: if $\mathcal{P} = \{g_\theta\}_{\theta\in\Theta}$ and $\mathcal{Q} = \{h_\xi\}_{\xi\in\Xi}$ are two parametric families of probability distributions with their respective Fisher-Rao distance $d_\mathcal{P}$ and $d_\mathcal{Q}$, then if $d_\mathcal{P}(g_\theta, g_{\theta'}) = d_\mathcal{Q}(h_\xi, h_{\xi'}) = \delta > 0$ we can conclude that g_θ and $g_{\theta'}$ are "as far apart as" h_ξ and $h_{\xi'}$. This is a consequence of the Cramér-Rao lower bound and holds even though these distributions belong to two different families each possessing its own Fisher-Rao distance.

In the second part of this presentation, we will explain how the impact of a distributional perturbation on the inputs is assessed through robustness indices. These indices are denoted $S_{i\delta}$ for each input X_i and a perturbation level δ. They quantify the relative variation of the QoI of the output before and after perturbation of the input distributions

$$S_{i\delta} = \frac{Q(Y^{i\delta}) - Q(Y)}{Q(Y)},$$

where $Q(Y)$ is a QoI of Y, usually a quantile, and Y and $Y^{i\delta}$ are respectively the initial and perturbed output of G. Since these statistical quantities are not explicitly known, a statistical estimation method is necessary which takes into account the computation cost of G. The estimation method that we use is importance sampling. A central limit theorem is available for the estimators of both $Q(Y)$ and $Q(Y^{i\delta})$ but the asymptotic variance depends respectively on the density function of Y and $Y^{i\delta}$ which is hard to estimate [1, 2]. For this reason, we decided to construct non-asymptotic confidence intervals for the output QoI based on well known concentration inequalities [3]. This will be illustrated on different analytical and industrial cases.

Short biography (PhD student)

I did my bachelor's and master's degree in mathematics at Paul Sabatier University. I then pursued a PhD career in mathematics at EDF R&D in collaboration with the Mathematics Institute of Toulouse (IMT). The main goal in my thesis is to perform robustness analysis of computer codes for uncertainty quantification using information geometry. I am currently a second year PhD student and I was funded by the Labex CIMI (Toulouse) for the first few months of my PhD and by EDF for the remaining duration.

References

[1] C. Gauchy, J. Stenger, R. Sueur, and B. Iooss. An information geometry approach to robustness analysis for the uncertainty quantification of computer codes. *Technometrics*, 64(1):80–91, 2022.

[2] P. W. Glynn. Importance sampling for Monte Carlo estimation of quantiles. In *Mathematical Methods in Stochastic Simulation and Experimental Design: Proceedings of the 2nd St. Petersburg Workshop on Simulation*, pages 180–185. Citeseer, 1996.

[3] B. B. Ketema, N. Bousquet, F. Costantino, F. Gamboa, B. Iooss, and R. Sueur. Non-asymptotic confidence intervals for importance sampling estimators of quantiles. In *55ème Journées de statistique de la SFdS*, Bordeaux, France, May 2024.

Addressing the Rashomon Effect through ranking aggregation

C. Sessa[†,1], E. Borgonovo[2], A. Cillo[§,1], G.P. Crespi[§,1], X. Lu[3]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Dec. 2022 – Dec. 2025

[1] LIUC Business University, Castellanza, Italy
{csessa, acillo, pcrespi}@liuc.it

[2] Department of Decision Sciences, Bocconi University, Milan, Italy
emanuele.borgonovo@unibocconi.it

[3] SKEMA Business School, Université Côte d'Azur, Paris, France
xuefei.lu@skema.edu

Abstract

When dealing with prediction problems, analysts rely on variable importance measures and global sensitivity measures to understand the predictive power of variables and uncover the relationships in the data [10]. When the data generating process (DGP) is unknown, analysts typically train machine learning models to use as surrogates, and derive explanations for the patterns in the data computing the variable importance of the best performing model. The validity of this approach is threatened by the Rashomon Effect [2], whereby multiple models achieve similar predictive accuracy but offer different and sometimes conflicting explanations for the underlying patterns. Indeed, the Rashomon Set [5] – the collection of all almost-optimal prediction models – can be seen both as a challenge and an opportunity for analysts: while this adds uncertainty to inference, it also allows for broader exploration of potential explanations.

A number of studies have succeeded in framing a procedure to compute or approximate the Rashomon Set for some specific model classes [11, 12, 4, 9]. Few attempts, however, have been made to explain the relationships in the data by exploiting the whole Rashomon Set [5, 4]. In this work, we propose a novel methodological framework that leverages all the models in the Rashomon Set to produce more reliable and consistent insights into variable importance. Our idea is to view the Rashomon Set for a dataset as a collection of agents, each expressing its own possibly different preference for the features, much like how different experts may offer varying interpretations of the same data. The strength of this preference corresponds to the importance of each variable for the prediction, quantified through an importance measure. By transforming the importance vectors for all the models into rankings and then aggregating them, our method allows analysts to generate a consensus ranking which reflects the preferences of the entire Rashomon Set, offering a comprehensive view on the mechanisms in the data. We draw upon the established literature on ranking aggregation techniques [3, 6, 8] to combine the individual importance rankings into a unified ranking that is robust to model variability.

The proposed framework complements existing variable importance measures and provides analysts with a powerful tool to handle model multiplicity in practical applications. We validate our methodology using both simulated data from known DGPs and real-world datasets, to demonstrate how the framework reconciles conflicting signals from multiple models and produces an

importance ranking of variables that is more aligned with the true DGP. We test different aggregation techniques to show how the choice of the technique impacts the consensus ranking. Furthermore, we provide theoretical results on the structure of the Rashomon Set for the specific class of linear regression models. In particular, we clarify the connection between the coefficients of linear models in the Rashomon Set and the permutation importance measure [1], a widely used measure in machine learning, exploring its relation to total indices [7].

Short biography (PhD student)

Claudia Sessa is a PhD Candidate at LIUC Business University in Castellanza, Italy. Previously, she obtained a MSc in Data Science and Business Analytics (cum laude) and a BSc in Economics, Management and Computer Science, both from Bocconi University in Milan. Her research lies at the intersection of operations research and machine learning.

References

[1] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[2] Leo Breiman. Statistical modeling: The two cultures. *Statistical Science*, 16(3), August 2001.

[3] Wade D. Cook. Distance-based and ad hoc consensus models in ordinal preference ranking. *European Journal of Operational Research*, 172(2):369–385, July 2006.

[4] Jiayun Dong and Cynthia Rudin. Exploring the cloud of variable importance for the set of all good models. *Nature Machine Intelligence*, 2(12):810–824, 2020.

[5] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177):1–81, 2019.

[6] Dorit S. Hochbaum and Asaf Levin. Methodologies and algorithms for group-rankings decision. *Management Science*, 52(9):1394–1408, September 2006.

[7] Toshimitsu Homma and Andrea Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, April 1996.

[8] Xue Li, Xinlei Wang, and Guanghua Xiao. A comparative study of rank aggregation methods for partial and top ranked lists in genomic applications. *Briefings in Bioinformatics*, 20(1):178–189, August 2017.

[9] Kota Mata, Kentaro Kanamori, and Hiroki Arimura. Computing the collection of good models for rule lists. In *Proc. the 18th International Conference on Machine Learning and Data Mining (MLDM 2022)*, 2022.

[10] Brian D. Williamson, Peter B. Gilbert, Noah R. Simon, and Marco Carone. A general framework for inference on algorithm-agnostic variable importance. *Journal of the American Statistical Association*, 118(543):1645–1658, January 2022.

[11] Rui Xin, Chudi Zhong, Zhi Chen, Takuya Takagi, Margo Seltzer, and Cynthia Rudin. Exploring the whole rashomon set of sparse decision trees. *Advances in Neural Information Processing Systems*, 35:14071–14084, 2022.

[12] Chudi Zhong, Zhi Chen, Jiachang Liu, Margo Seltzer, and Cynthia Rudin. Exploring and interacting with the set of good sparse generalized additive models. *Advances in Neural Information Processing Systems*, 36, 2024.

Surrogate to Poincaré inequalities on manifolds for dimension reduction in nonlinear feature spaces

A. PASCO[†,1], A. NOUY[§,1]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Dec. 2022 – Nov. 2025

[1] Laboratoire de Mathématiques Jean Leray, Ecole Centrale de Nantes, Nantes Université
{alexandre.pasco,anthony.nouy}@ec-nantes.fr

Abstract

This work focuses on approximating a differentiable function $u : \mathbb{R}^d \to \mathbb{R}$ with $d \gg 1$ by a composition of functions $f \circ g$ where $g : \mathbb{R}^d \to \mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}$. The approximation error is assessed in the L^2_μ-norm where μ is some probability measure on $\mathbb{R}^d$. The approach considered is two-staged. Firstly the feature map g is selected among some prescribed functional class by minimizing some function $\mathcal{J}$ involved in the upper bound of the approximation error

$$\min_{f:\mathbb{R}^m\to\mathbb{R}} \mathbb{E}_\mu(|u - f \circ g|^2) \leq C_\mu \mathcal{J}(g), \tag{1}$$

which is based on Poincaré inequalities and requires evaluations of ∇u.

Secondly the function f is built using classical regression methods. Until recently, bounds of the form (1) were only available for linear feature maps g. This framework has been extensively studied under the name Active Subspace, see for example [2, 4], and the solution is given by the eigenvectors of the matrix $\mathbb{E}(\nabla u \nabla u^T) \in \mathbb{R}^d$. This approach is easy to implement, computationally efficient, has robust theoretical guarantees for some classical probability laws μ, and showed good performances in various numerical applications. However, there are many functions u for which such an approximation with $m < d$ is known to be not efficient.

Therefore, recent works consider non-linear feature maps in order to produce better dimension reduction. More especially, we will focus on the work from [1, 3] in which authors leverage Poincaré inequalities on smooth manifolds to obtain a bound of the form (1) for non-linear g. Although there are less theoretical guarantees, their numerical experiments showed improved performances compared with linear featuring. However, minimizing $\mathcal{J}$ is now much harder than finding eigenvectors of some matrix, and can only be done using iterative descent methods.

In this work we consider feature maps as in [1], of the form $g(x) = G^T\Phi(x)$ with $G \in \mathbb{R}^{K\times m}$ and where $\Phi : \mathbb{R}^d \to \mathbb{R}^K$, $K \geq d$, is fixed. We study a new quantity, denoted $\mathcal{L}(g)$, which can be expressed as the minimal singular value of some positive semi-definite matrix. We show that for a compact set of polynomial feature maps with $m = 1$, for some class of probability distributions, any minimizer g^* of $\mathcal{L}$ satisfies the sub-optimality result

$$\mathcal{J}(g^*) \lesssim \min_g \mathcal{J}(g)^\beta,$$

where $0 < \beta \leq 1$ is some constant which depends on the degree. We also extend this approach to the case $m > 1$ as well as for simultaneously learning a parametrized family of functions $u_y \in L^2_\mu$ by $y \in \mathcal{Y}$, although the theoretical results are weaker. Finally, we provide numerical examples to illustrate the performances of g^*, both as the feature map used in the regression step, or as the initializer for some iterative descent algorithm for minimizing $\mathcal{J}$.

Short biography (PhD student)

I graduated from Ecole Centrale de Nantes in 2022 as generalist engineer. It included 2 years focused on applied mathematics, from machine learning to numerical analysis. I continued at ECN by starting my PhD thesis on December 2022, funded by ANR COFNET, focusing on compositional function networks for nonlinear model reduction, for forward and inverse problems.

References

[1] Daniele Bigoni, Youssef Marzouk, Clémentine Prieur, and Olivier Zahm. Nonlinear dimension reduction for surrogate modeling using gradient information. *Information and Inference: A Journal of the IMA*, 11(4):1597–1639, December 2022.

[2] Paul G. Constantine, Eric Dow, and Qiqi Wang. Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces. *SIAM J. Sci. Comput.*, 36(4):A1500–A1524, January 2014.

[3] Romain Verdière, Clémentine Prieur, and Olivier Zahm. Diffeomorphism-based feature learning using Poincaré inequalities on augmented input space, December 2023.

[4] Olivier Zahm, Paul G. Constantine, Clémentine Prieur, and Youssef M. Marzouk. Gradient-Based Dimension Reduction of Multivariate Vector-Valued Functions. *SIAM J. Sci. Comput.*, 42(1):A534–A558, January 2020.

Mollifiers to enhance gradient-based dimension reduction

R. Verdière[†,1], C. Prieur[§,1] O. Zahm[§,1],

[†] PhD student (presenting author). [§] PhD supervisor
PhD expected duration: Sept. 2022 – Aug. 2025

[1] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK
{romain.verdiere,olivier.zahm}@inria.fr ; clementine.prieur@univ-grenoble-alpes.fr

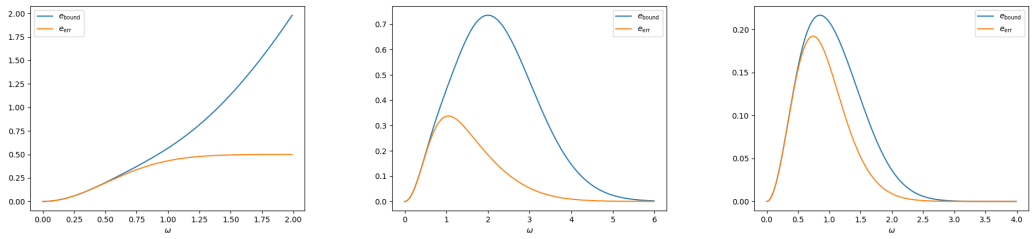
Abstract

Modern computational models for scientific and engineering applications typically involve a large number of input parameters and are expensive-to-evaluate both in time and resources. Replacing the model with an accurate and fast-to-evaluate surrogate (or approximation) offers a viable workaround in many applications. Approximating such high-dimensional functions with classical approximation tools such as polynomials, wavelets or neural networks is, however, a difficult task. This is even aggravated in the small sample regime where one only has access to a little number of model evaluations. One way to address this challenge is to reduce the input dimension beforehand. This consists in approximating the *model* $x \to u(x)$ as the composition of two functions: a *feature map* $x \to z = g(x)$ which extracts the relevant features of the input variables, and a *profile function* $z \to f(z)$ which regresses the model output on the features. The feature map can be built by minimizing an upper bound of the reconstruction error $\min_f \mathbb{E}[(u(\boldsymbol{X}) - f \circ g(\boldsymbol{X}))^2]$ obtained with Poincaré-type inequalities. When the feature map is linear this strategy reduces to Active Subspace [4, 2]. The case of nonlinear feature maps has been studied in [1] for polynomial feature maps and in [3, 5] for diffeomorphism-based feature maps. The bound derived from Poincaré inequality is proportional to the L^2-norm of model gradients, therefore, this strategy works well for slowly varing functions for which the bound is tight. For oscillatory model with large gradient norms, however, the bound reveals too loose to build a meaningfull feature map and the method fails.

In this talk we demonstrate that working with a mollified version of the model $(u \star \rho_\sigma)$ is a good strategy to circumvent this issue as it allows to obtain sharper Poincaré error bounds and to reduce the dimension efficiently using gradient-based techniques. Here ρ_σ is the gaussian kernel with 0 mean and $\sigma^2 I_d$ covariance, $\star$ is the convolution operator and we call σ the mollifying parameter. We demonstrate that the reconstruction error when using a mollified version of the model can be bounded by the sum of two terms: one that vanishes when the mollifying parameter goes to zero and one that is proportional to the Poincaré error bound of the mollified model. This bound shows the trade-off between mollification and dimension reduction: for strongly mollified models the first term is large and the second one quite small and the other way around when the model is less mollified. Based on this result, we propose an iterative algorithm for dimension reduction. More precisely, we introduce a sequence $\sigma_1 > \sigma_2 > \ldots > \sigma_p \geq 0$ of decreasing mollifying parameters. Then at the first iteration we approximate a strongly mollified version of the model $u_1^* = u \star \rho_{\sigma_1}$ with a feature map g_1 and a profile function f_1. At the next interation the algorithm approximates a slightly less mollified version of the residual model $u_2^* = (u - f_1 \circ g_1) \star \rho_{\sigma_2}$ with a feature map g_2 and a profile function f_2. This proccess iterates p times and at the end the original model u is approximated by $\sum_{1 \leq i \leq p} f_i \circ g_i$.

Let us illustrate on some example the impact of the mollification step on the accuracy of Poincaré error bound. We consider the analytical toy model $u(x) = \sum_{i=1}^{d} a_i \sin(\omega_i x_i)$, where a_i, ω_i, x_i are

respectivelly the ith components of vectors $a \in \mathbb{R}^d, \omega \in \mathbb{R}^d, x \in \mathbb{R}^d$. We aim at approximating u by $f \circ g$ with g a projector onto $\{e_1, ..., e_d\}$ the canonical basis of $\mathbb{R}^d$. Here g is a linear feature map and $g = U^\top \sum_{i \in \tau} e_i e_i^\top$ where $\tau \subset \{1, \dots, d\}$, $\#\tau = m$ and $U = [e_i]_{i \in \tau} \in \mathbb{R}^{d \times m}$. In this framework, and for $\boldsymbol{X} \sim \mathcal{N}(0, I_d)$, we compare the minimal reconstruction error for $u_\sigma = u * \rho_\sigma$ with the one obtained by minimizing Poincaré error bound. We perform the comparison for $a_i = 1$, $i = 1, ..., d$ and for different values of σ. In this situation the reconstruction error is equal to $\frac{1}{2} \sum_{i \in -\tau} e^{-\omega_i^2 \sigma^2}(1 - e^{-2\omega_i^2})$ and the Poincaré error bound is equal to $\frac{1}{2} \sum_{i \in -\tau} \omega_i^2 e^{-\omega_i^2 \sigma^2}(1 + e^{-2\omega_i^2})$, where $-\tau$ is the complementary set of τ in $\{1, \dots, d\}$. We can compare the 2 functions $e_{\text{err}}(\omega) = \frac{1}{2} e^{-\omega^2 \sigma^2}(1 - e^{-2\omega^2})$ and $e_{\text{bound}}(\omega) = \frac{1}{2} \omega^2 e^{-\omega^2 \sigma^2}(1 + e^{-2\omega^2})$ to understand how the error and the bound behave for different values of the mollifying parameter σ. Figure 1 clearly shows that the reconstruction error and the error bound become closer together as the value of σ grows.



(a) $\sigma = 0$ (b) $\sigma = 0.5$ (c) $\sigma = 1$

Figure 1: Plots of e_{err} and e_{bound} according to ω for different values of σ

Short biography

I am a former student of ENSTA Paris where I studied mathematics and computer science. At the end of my engineering studies, to deepen my knowledge in general mathematics, I prepared for and passed mathematical aggregation. Finally, after a short experience as a teacher, I moved to Grenoble to work as PhD student with Clémentine Prieur and Olivier Zahm on nonlinear dimension reduction for function approximation. This thesis is funded by the French Research Agency (ANR).

References

[1] Daniele Bigoni, Youssef Marzouk, Clémentine Prieur, and Olivier Zahm. Nonlinear dimension reduction for surrogate modeling using gradient information. *Information and Inference: A Journal of the IMA*, 11(4):1597–1639, 2022.

[2] Paul G. Constantine, Eric Dow, and Qiqi Wang. Active subspace methods in theory and practice: Applications to kriging surfaces. *SIAM Journal on Scientific Computing*, 36(4):A1500–A1524, January 2014.

[3] Romain Verdière, Clémentine Prieur, and Olivier Zahm. Diffeomorphism-based feature learning using Poincaré inequalities on augmented input space. working paper or preprint, December 2023.

[4] Olivier Zahm, Paul Constantine, Clémentine Prieur, and Youssef Marzouk. Gradient-based dimension reduction of multivariate vector-valued functions. 2019.

[5] Guannan Zhang, Jiaxin Zhang, and Jacob Hinkle. Learning nonlinear level sets for dimensionality reduction in function approximation. *Advances in Neural Information Processing Systems*, 32, 2019.

Variance-Informed Subspace: a Gradient-free Dimension Reduction for Adaptive Bayesian Inference

N. Polette[†,1,2], P. Sochala[§,1], O. Le Maître[§,3], A. Gesret[§,2]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Nov. 2022 – Oct. 2025

[1] CEA, DAM, DIF, F-91297 Arpajon, France
{pierre.sochala}@cea.fr

[2] Mines Paris PSL, Geosciences center, Fontainebleau, France
{nadege.polette,alexandrine.gesret}@minesparis.psl.eu

[3] CMAP, CNRS, Inria, École Polytechnique, IPP, Palaiseau, France
{olivier.le-maitre}@polytechnique.edu

Abstract

Inverse problems are encountered in many applications whenever one search for information about a physical system based on measurements [7]. In this work, we are interested in estimating a physical field thanks to a set of indirect observations $\boldsymbol{d}$. The Bayesian inference is an attractive approach for adressing such problems, as it provides a full estimation of the unknown parameters distributions. In that framework, the aim is to estimate the posterior probability of the field parameters $\boldsymbol{x}$ based on the observations

$$\pi_{\text{post}}(\boldsymbol{x}|\boldsymbol{d}) \propto \mathcal{L}(\boldsymbol{d}|\boldsymbol{x})\pi_{\text{prior}}(\boldsymbol{x}), \tag{1}$$

where $\mathcal{L}$ is the likelihood of the observations given a field and π_{prior} the prior probability of the field. The posterior distribution is then sampled with Markov Chain Monte–Carlo (MCMC) [4]. In order to accelerate the MCMC sampling, the forward model predictions are replaced with surrogate models based on polynomial chaos (PC) expansions [8, 5]. In order to reduce the input dimension of the surrogate model, a parsimonious representation of the field is introduced by means of the Karhunen-Loève (KL) decomposition, on the assumption that the field of interest is a particular realization of a Gaussian random field. Despite this parametrization, several hundred inputs could be required to represent accurately a two-dimensional field. This is expensive with regard to both the forward model surrogate training and the MCMC convergence.

Linear dimension reduction techniques have been developed to decrease the number of parameters to infer. These techniques assume that most of the information provided by the likelihood can be captured by a low-dimensional linear subspace. The input parameter space is decomposed into two subsets

$$\boldsymbol{x} = Ax_a + A_\perp x_i, \tag{2}$$

where x_a is informed by the likelihood, while x_i is constrained by the prior. The posterior distribution (1) rewrites

$$\pi_{\text{post}}(\boldsymbol{x}|\boldsymbol{d}) \propto \mathcal{L}(\boldsymbol{d}|x_a)\pi_{\text{prior}}(x_a)\pi_{\text{prior}}(x_i|x_a), \tag{3}$$

such that only x_a is sampled during the MCMC procedure. Several methods to define the linear transformation operator A have been developed. Cui et al. [3] build a Likelihood-Informed Subspace (LIS) which relies on the Hessian of the log-likelihood. The optimality of such construction

has been proven in [6] for the linear case. Constantine et al. [2] adapt the Active Subspace (AS) approach [1] to the Bayesian framework by using the misfit gradient. In both methods, the curvature of the log-posterior density is more constrained by the log-likelihood than by the prior along the subspace directions.

This study presents a new construction for the linear transformation operator A. The general idea is inspired from the work of [6] which states that, in the linear case, approximating the posterior covariance is equivalent to approximating its inverse. Instead of relying on the Hessian of the log-likelihood, the approximation of the inverse posterior covariance involves the ratio of the posterior and the prior variances.

For nonlinear inverse problems, we propose to generalize this variance ratio. The low-dimensional subspace is defined as the directions in which the posterior variance is drastically reduced in comparison to the prior variance. This method is gradient-free. We show on state-of-the-art examples that it is sufficient for unimodal posteriors, while some adjustments are required in the case of multimodal results. An application on a two-dimensional field inference case illustrates the interest of the method for high-dimensional problems.

Short biography (PhD student)

Nadège Polette studied applied mathematics at ENPC as well as at Sorbonne Université. Her PhD is funded by CEA. The supervision is ensured by Dr. A. Gesret (Geosciences center, École des Mines de Paris), Dr. P. Sochala (CEA), and Dr. O. Le Maître (CNRS, CMAP). Her PhD falls within the aim of the CTBTO to detect and analyse seismic events. The goal is to develop numerical methods for solving inverse problems applied to geophysical events analysis.

References

[1] P. G. Constantine. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*. SIAM Spotlights. SIAM, 2015.

[2] P. G. Constantine, C. Kent, and T. Bui-Thanh. Accelerating MCMC with Active Subspaces. *SIAM Journal on Scientific Computing*, 38(5):A2779–A2805, January 2016.

[3] T. Cui, J. Martin, Y. M. Marzouk, A. Solonen, and A. Spantini. Likelihood-informed Dimension Reduction for Nonlinear Inverse Problems. *Inverse Problems*, 30(11):114015, November 2014.

[4] A. Doucet, A. Smith, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Information Science and Statistics. Springer New York, 2013.

[5] R. G. Ghanem and P. D. Spanos. Stochastic Finite Element Method: Response Statistics. In *Stochastic Finite Elements: A Spectral Approach*, pages 101–119. Springer New York, New York, NY, 1991.

[6] A. Spantini, A. Solonen, T. Cui, J. Martin, L. Tenorio, and Y. M. Marzouk. Optimal Low-Rank Approximations of Bayesian Linear Inverse Problems, July 2015.

[7] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*, volume xii. SIAM, January 2005.

[8] N. Wiener. The Homogeneous Chaos. *American Journal of Mathematics*, 60(4):897–936, 1938.

Sequential transport for density estimation and its applications

B. Zanger[†,1], O. Zahm[1], T. Cui[2], M. Schreiber[§,1]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Oct. 2022 – Sep. 2025

[1] Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France
{benjamin.zanger, olivier.zahm}@inria.fr
{martin.schreiber}@univ-grenoble-alpes.fr

[2] School of Mathematics and Statistics, University of Sydney, Australia
tiangang.cui@sydney.edu.au

Abstract

Transport-based methods are receiving growing interest because of their ability to sample easily from the approximated density. These methods aim at building a deterministic diffeomorphism $\mathcal{T}$, also called a transport map, which pushes forward an arbitrary reference probability density ρ_{ref} to a given target probability density π to be approximated. Typically, the transport map $\mathcal{T}$ is parameterized *e.g.* by invertible neural networks and fitted using variational methods of the form

$$\min_{\mathcal{T}\in\mathcal{M}} \mathrm{D}(\pi||\mathcal{T}_\sharp\,\rho_{\text{ref}}) \tag{1}$$

with a statistical divergence $\mathrm{D}(\cdot\,||\cdot)$, typically the (reversed) KL-divergence. An emerging strategy for this problem is to first estimate π by π and then to compute a map $\mathcal{T}$ which exactly pushes forward ρ_{ref} to π, see [2, 1]. Among the infinitely many maps $\mathcal{T}$ which satisfy $\mathcal{T}_\sharp\,\rho_{\text{ref}} = \pi$, the Knothe–Rosenblatt (KR) map is rather simple to evaluate since it requires only computing the cumulative distribution functions (CDFs) of the conditional marginals of π. In general, problem (1) is difficult to solve when π is multimodal or when it concentrates on a low-dimensional manifold. The solution proposed in [1] consists in introducing an arbitrary sequence of bridging densities

$$\pi^{(1)}, \pi^{(2)}, \ldots, \pi^{(L)} = \pi, \tag{2}$$

with increasing complexity. The sequential strategy consists in building L transport maps $\mathcal{Q}_1, \ldots, \mathcal{Q}_L$ one after the other by solving

$$\min_{\mathcal{Q}_\ell\in\mathcal{M}} \mathrm{D}(\pi^{(\ell)}||(\mathcal{T}_{\ell-1}\circ\mathcal{Q}_\ell)_\sharp\,\rho_{\text{ref}}), \quad \text{where} \quad \mathcal{T}_{\ell-1} = \mathcal{Q}_1\circ\ldots\circ\mathcal{Q}_{\ell-1}. \tag{3}$$

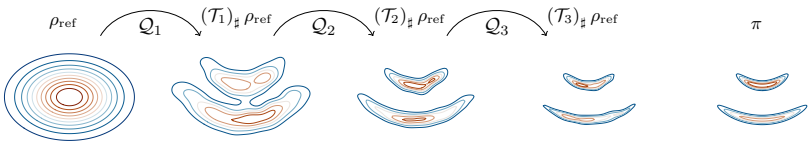


Figure 1: Visualization of the approximation of a bimodal density π (right) using $L = 3$ intermediate tempered densities estimated using SoS (4) and a Gaussian reference density ρ_{ref}.

For suitable statistical distances, so that $\mathrm{D}(\pi||\mathcal{T}_\sharp\rho) = \mathrm{D}(\mathcal{T}^\sharp\pi||\rho)$, these problems are equivalent to estimating the pullback density $(\mathcal{T}_{\ell-1})^\sharp\pi^{(\ell)}$ with an intermediate approximation $\rho^{(\ell)} = (\mathcal{Q}_\ell)_\sharp\,\rho_{\mathrm{ref}}$.

In our work, we contribute to this methodology as follows.

First, we employ Sum-of-Squares (SoS) densities to approximate the intermediate densities $\rho^{(\ell)}$ using α-divergences $\mathrm{D}_\alpha(\cdot||\cdot)$. We sequentially solve the variational density approximation problem as in Equation (3) with D_α as the statistical divergence and where

$$\rho^{(\ell)}(\boldsymbol{x}) = \left(\Phi(\boldsymbol{x})^\top A_\ell \Phi(\boldsymbol{x})\right)\rho_{\mathrm{ref}}(\boldsymbol{x}), \tag{4}$$

for some arbitrary orthonormal basis function Φ in $L^2(\rho_{\mathrm{ref}})$. Here, the positivity of the matrix $A_\ell \succeq 0$ ensures the density $\rho^{(\ell)}$ to be positive. Since the α-divergence is defined for general *unnormalized* densities, it is not necessary to know the normalizing constant of π beforehand. α-divergences $\mathrm{D}_\alpha(\cdot||\cdot)$ with parameter $\alpha \in \mathbb{R}$ include the Hellinger distance and KL-divergence, which have been used in previous works. The proposed SoS densities permit to efficiently normalize the estimated unnormalized density and to compute the KR map $\mathcal{Q}_\ell$ such that $(\mathcal{Q}_\ell)_\sharp\,\rho_{\mathrm{ref}} = \rho^{(\ell)}$. This combined use of α-divergence for performing SoS density estimation results in a *convex* optimization problems which can be efficiently solved using off-the-shelf toolboxes.

Second, we extend the methodology to the scenario where *only samples* $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}$ from π are available, as opposed to *point-evaluations* of the target density π. In this scenario, we propose to use diffusion-based bridging densities $\pi^{(\ell)}(\boldsymbol{x})$ where the distribution follows a time–inversed diffusion process such as the Ornstein-Uhlenbeck process with time parameters $t_{\ell-1} \leq t_\ell$ and $t_L = 0$. This idea is at the root of diffusion models.

Third, we present a novel convergence analysis using the geometric properties of α-divergences. This analysis unifies and extends previous analyses proposed in [3, 1] and, more interestingly, it guides the choice of bridging densities. In particular, we show that a smart choice of β_ℓ for tempered densities or of t_ℓ for diffusion-based densities yield a convergence rate of $\mathcal{O}(1/L^2)$ with respect to the number of layer L.

Last, we give an outlook for using sequential measure transport to solve optimal transport problems, where we mitigate the difficulty of estimating the optimal coupling by a sequence of entropic regularized problems.

We demonstrate the capability of our proposed method in unsupervised learning, Bayesian inverse, and optimal transport problems in moderate dimensions.

Short biography (PhD student)

I am an electrical engineer, now pursuing a PhD in applied mathematics at the Université Grenoble Alpes while working at INRIA in the AIRSEA team. In my thesis, I work on estimation of densities using compositional methods and its applications.

References

[1] Tiangang Cui and Sergey Dolgov. Deep composition of tensor-trains using squared inverse Rosenblatt transports. *Foundations of Computational Mathematics*, September 2021. arXiv:2007.06968 [cs, math, stat].

[2] Sergey Dolgov, Karim Anaya-Izquierdo, Colin Fox, and Robert Scheichl. Approximation and sampling of multivariate probability distributions in the tensor train decomposition. *Statistics and Computing*, 30(3):603–625, May 2020.

[3] Josephine Westermann and Jakob Zech. Measure transport via polynomial density surrogates, November 2023. arXiv:2311.04172 [cs, math, stat].

Learning signals defined on graphs with optimal transport and Gaussian process regression

Raphaël Carpintero Perez[†,1,2], Sébastien Da Veiga[§,3], Josselin Garnier[§,2], Brian Staber[§,1]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Nov. 2022 – Nov. 2025

[1] Safran Tech, Digital Sciences & Technologies, 78114 Magny-Les-Hameaux, France
{raphael.carpintero-perez, brian.staber}@safrangroup.com

[2] Centre de Mathématiques Appliquées, Ecole polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France
{raphael.carpintero-perez, josselin.garnier}@polytechnique.edu

[3] Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France
sebastien.da-veiga@ensai.fr

Abstract

In computational physics, machine learning has now emerged as a powerful complementary tool to explore efficiently candidate designs in engineering studies. In this context, we would like to be able to easily predict fields defined on meshes corresponding to new geometries without the need for costly simulations. While some methods like Graph Neural Networks [4] are intrinsically designed to predict signals defined on graphs or point clouds, a natural question is the extension of general scalar output regression models to such complex outputs. Changes between input geometries in terms of both size and adjacency structure in particular make this transition non-trivial. Another key challenge is the obtention of predictive uncertainties, which is crucial to certify the quality of results, to assist sequential design of experiments or to plug the models into Bayesian optimization workflows.

In this work, we propose an innovative strategy for Gaussian process regression where inputs are large and sparse graphs with continuous node attributes and outputs are signals defined on the nodes of the associated inputs. The methodology relies on the combination of regularized optimal transport [3], dimension reduction techniques [2], and the use of Gaussian processes [5] indexed by graphs. It extends previous work on Gaussian processes with Sliced Wasserstein Weisfeiler Lehman graph kernels [1] previously limited to scalar outputs. In addition to enabling signal prediction, the main point of our proposal is to come with confidence intervals on node values.

We illustrate the efficiency of the method with regression tasks involving large graphs from mesh-based simulations in computational fluid dynamics and mechanics [1]. Train datasets are made up of a few hundred graphs with their respective 2D/3D coordinates, where adjacency matrices vary between several inputs, and output fields represent physical quantities of interest on the nodes such as the pressure or the temperature. In Figure 1, we show predictions and associated uncertainties for two test samples of a problem in computational mechanics.

[1] Datasets: https://plaid-lib.readthedocs.io/en/latest/source/data_challenges.html

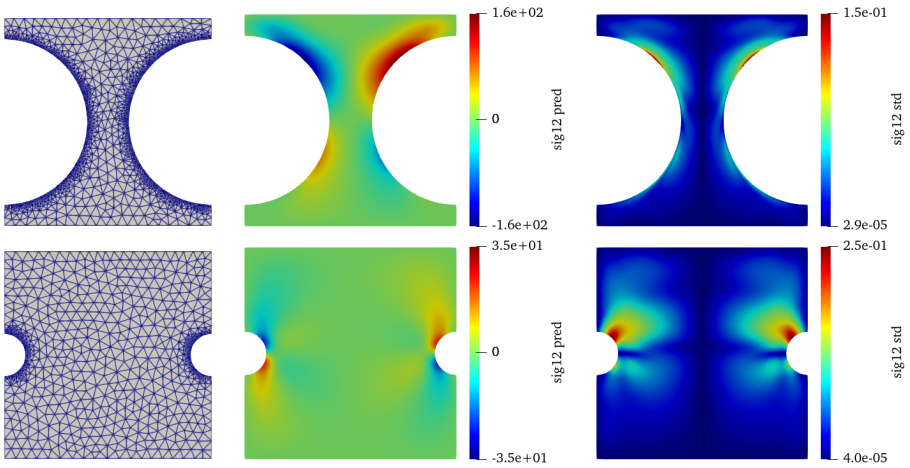


Figure 1: Predictions for two test meshes from the Tensile2d dataset (top and bottom lines). From left to right: the input mesh, the predicted field and the posterior standard deviation of the Gaussian process regression.

Short biography (PhD student)

I have a double Master's degree in Mathematics and Computer Science for Data Science (MIDS) from the University Paris Cité (formerly Diderot).
This thesis is a joint work between the Ecole polytechnique and Safran Tech.
It is part of the ANR SAMOURAI (ANR-20-CE46-0013).

References

[1] Raphaël Carpintero Perez, Sébastien Da Veiga, Josselin Garnier, and Brian Staber. Gaussian process regression with sliced Wasserstein Weisfeiler-Lehman graph kernels. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, pages 1297–1305, 2024.

[2] Katiana Kontolati, Dimitrios Loukrezis, Dimitrios G Giovanis, Lohit Vandanapu, and Michael D Shields. A survey of unsupervised learning methods for high-dimensional uncertainty quantification in black-box-type problems. *Journal of Computational Physics*, 464:111313, 2022.

[3] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[4] Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter Battaglia. Learning mesh-based simulation with graph networks. In *International Conference on Learning Representations*, 2020.

[5] Christopher Ki Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*. MIT press, 2006.

Importance Sampling in high dimension

J. Beh[†,1,2], F. Simatos[§,2], J. Morio[§,1,2]

[†] PhD student (presenting author). [§] PhD supervisor

PhD expected duration: Jan. 2023 – Dec. 2025

[1] ONERA/DTIS, Université de Toulouse, F-31055 Toulouse
{jason.beh,jerome.morio}@onera.fr

[2] Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, 31000 Toulouse
florian.simatos@isae-supaero.fr

Abstract

Rare event analysis often involves the estimation of the rare event probability $p = \mathbb{P}_f(X \in A)$, where $f = N(0, I)$ is the d-dimensional standard Gaussian distribution, which is a fairly general setting owing to isoprobabilistic transformations [8]. Whereas Adaptive Splitting [3] concerns the modification of the trajectories of the samples towards the region of interest A, Importance Sampling (IS) considers an auxiliary distribution g which allocates more probability mass in A than f. Given n_g samples $(Y_i)_{i=1\dots n}$ generated according to g to whom $\mathbb{1}(\cdot \in A)\, f$ is absolutely continuous, the IS estimator is written as

$$\hat{p}_g = \frac{1}{n_g} \sum_{i=1}^{n_g} \frac{f(Y_i)}{g(Y_i)} \mathbb{1}(Y_i \in A)$$

In low dimension, IS estimators are often employed due to the desired variance reduction property compared to Monte Carlo estimator. However, in high dimension, IS estimators suffer from convergence issues and become extremely sensitive to the choice of auxiliary distribution. This motivates a theoretical study on the convergence of IS estimators in the high-dimensional setting, $d \to +\infty$.

As $d \to +\infty$, two settings can arise: either the probability to be estimated is bounded away from zero: $\inf_d p > 0$, or the probability tends to zero with the dimension: $p \to 0$. The first setting $\inf_d p > 0$, considered by [1, 4], occurs when p involves a stochastic process which is approximated by a finite sum of random variables by principal component analysis. Then, the probability to estimate becomes p_d, which tends to $p > 0$ when $d \to \infty$. In this setting, We will discuss our work on the convergence of the Cross-Entropy scheme [2] as well as its projection and improved variants [5, 10, 9].

The second setting, $p \to 0$ as $d \to \infty$, considered by [7, 6], occurs in specific settings such as in Highly Reliable Markovian Systems or in static network reliability estimation. This setting is more complex since the properties of IS estimators are reliant on the rate of convergence of p to 0. To tackle this setting, we first establish necessary and sufficient conditions for general IS estimators to be consistent, and conditions to verify a Central Limit Theorem towards a normal distribution. We then translate these conditions into the necessary rate of growth of the sample size n_g for various auxiliary distributions in a classical large deviation setting, $A = \{x \in \mathbb{R}^d : \sum_{j=1}^{d} x(j) \geq d^\gamma\}$ with $\gamma > 1/2$. It will be observed that the 'optimal' Gaussian density for IS largely depends on the error metric considered.

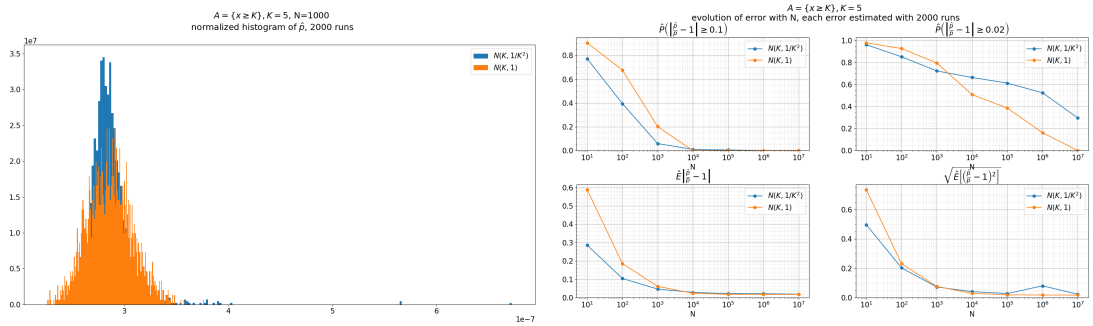


Figure 1: The histogram ($n_g = 1000$) and the evolution of usual error metrics with n_g of $\hat{p}_g$ for two choices of auxiliary density: which is better?

Short biography (PhD student)

I hold an M. Sc. in engineering from ISAE-SUPAERO in Toulouse. I completed my final-year internship at ISAE-SUPAERO which led to my current PhD thesis co-funded by EUR-MINT and ONERA, under the supervision of F. Simatos and J. Morio. The main goal of the thesis is to study the cause of the well-known curse-of-dimensionality for adaptive importance sampling, and to provide means to circumvent it based on this knowledge.

References

[1] S. Au and J. Beck. Important sampling in high dimensions. *Structural Safety* 25.2 (2003), pp. 139–163.

[2] J. Beh, Y. Shadmi, and F. Simatos. Insight from the Kullback–Leibler divergence into adaptive importance sampling schemes for rare event analysis in high dimension. *arXiv preprint* (2023). arXiv: `2309.16828 [math.ST]`.

[3] F. Cérou, A. Guyader, and M. Rousset. Adaptive multilevel splitting: Historical perspective and recent results. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.4 (2019), p. 043108.

[4] S. Chatterjee and P. Diaconis. The sample size required in importance sampling. *The Annals of Applied Probability* 28.2 (2018).

[5] M. El Masri, J. Morio, and F. Simatos. Improvement of the cross-entropy method in high dimension for failure probability estimation through a one-dimensional projection without gradient estimation. *Reliability Engineering & System Safety* 216 (2021), p. 107991.

[6] A. Guyader and H. Touchette. Efficient Large Deviation Estimation Based on Importance Sampling. *Journal of Statistical Physics* 181.2 (2020), pp. 551–586.

[7] P. L'Ecuyer et al. Asymptotic robustness of estimators in rare-event simulation. *ACM Trans. Model. Comput. Simul.* 20.1 (2010).

[8] R. Lebrun and A. Dutfoy. Do Rosenblatt and Nataf isoprobabilistic transformations really differ? *Probabilistic Engineering Mechanics* 24.4 (2009), pp. 577–584.

[9] I. Papaioannou, S. Geyer, and D. Straub. Improved cross entropy-based importance sampling with a flexible mixture model. *Reliability Engineering & System Safety* 191 (2019), p. 106564.

[10] F. Uribe et al. Cross-Entropy-Based Importance Sampling with Failure-Informed Dimension Reduction for Rare Event Simulation. *SIAM/ASA Journal on Uncertainty Quantification* 9.2 (2021), pp. 818–847.