Subject: Advanced Sampling for Medical Simulations Using Machine Learning Approaches

General Context

Computer simulation of physical processes is generally based on solving various multi-dimensional integral equations. In general, integrands in such simulations cannot be represented in closed form, and the integral equations rely on stochastic and quasi-stochastic sampling techniques known as Monte Carlo sampling/integration. The primary goal of Monte Carlo sampling/integration is to provide the lowest possible error at a fixed computational cost, while avoiding undesirable side effect such as bias or aliasing. The present research project targets the conception of innovative approaches for improving the performances of the most advanced existing schemes used in Monte Carlo sampling/integration, by combining our expertise with the power of machine learning approaches which made tremendous progress in the last years.

The student will be assisted by world-class experts in the theory of sampling and in medical simulation, who will help to propose innovative approaches combining the most advanced ideas and techniques known to date.

The goal

The goal of the present research project is many-fold.

- First, the student will develop a tangible approach to estimate the degree of uniformity of N-dimensional sequences of sampling distributions. This estimation will be used to tailor a reliable measure used in the optimization process in order to obtain sequences of sampling distributions, well-suited for a specific domain of application. Typical applications that we aim to improve are physical processes which happen in medical simulations, namely in particle-tissue or in wave-tissue interactions in the beam therapy (e.g., X-ray or hadron therapy). We expect that machine-learning techniques will greatly improve the performance of the simulation. The goal here is to build a machine-learning based estimation of the discrepancy. The bottleneck lies in the fact that we cannot compute exact values for the discrepancy but rather upper and lower bounds or numerical surrogates (optimal transport-based for example). The goal here is to aggregate all these partial and approximate information on the discrepancy to learn the discrepancy metric. As a starting point towards this challenging goal, a siamese network or, alternatively, a triplet network will provide network baselines to get a first metric estimation.

- Second, it is well known that importance sampling strongly impacts the error in stochastic sampling. Importance sampling is efficient when the probability density function (PDF) of the integrand is known or at least precisely estimated. In this part of the project, the student will develop machine learning-based approaches to estimate PDF of the integrand from the sparse available data. This is a challenging task, which may become successful for a variety of physical simulations of the particle-tissue interactions. To do so, the student will build a training set of integrands and corresponding numerical integral values and learn from it, via for example embedding learning with an encoder-decoder or more simply a network. This integral approximation will then be used as a control variate to accelerate the Monte Carlo convergence. While this can look as an easy task, designing the training dataset so as to reliably estimate integrals afterwards is a challenge, in particular the way integrands will be fed to the neural network must be carefully chosen.

Tasks

- Explore experimental data from X-ray or hadron therapy in order to provide a clear model for simulation.
- Study the simulation context based on experimental data.
- Study errors in physically based simulations in the context of particle-tissue interactions (X-ray or hadron therapy): their convergency rate, bias, aliasing.
- Propose a machine-learning approach for optimization of sampling distributions in order to improve the convergency rate. Evaluate the performance of the proposed approach and compare it to the state-of-the art ones.
- Propose a machine-learning approach for estimation of PDF from sparse data. Evaluate the benefits of the proposed approach in the context of importance sampling.

Details

Supervisor: Victor Ostromoukhov https://perso.liris.cnrs.fr/victor.ostromoukhov/ (https://perso.liris.cnrs.fr/victor.ostromoukhov/)
Where: University Claude Bernard Lyon1/LIRIS-CNRS lab
When: academic year 2020-2021 and/or summer 2021
Financing: ANR MOCAMED grant
Possible Extension: for a successful Master student, there is a possibility to continue with a PhD financed by the ANR MOCAMED grant

References

Introductory-level material:

- Alexander Keller, Per Christensen, Matt Pharr, Abdalla Ahmed, Victor Ostromoukhov, Iliyan Georgiev, My favorite Samples, SIGGRAPH 2019 Course, Los Angeles, USA. https://sites.google.com/view/myfavoritesamples (https://sites.google.com/view/myfavoritesamples)
  In-depth material:
- Art B. Owen, Monte Carlo theory, methods and examples, 2013. https://statweb.stanford.edu/~owen/mc/ (https://statweb.stanford.edu/~owen/mc/)
- Christiane Lemieux, Monte Carlo and Quasi-Monte Carlo Sampling. Springer Series in Statistics, 2009. https://www.springer.com/gp/book/9780387781648 (https://www.springer.com/gp/book/9780387781648)
  Sampling and integration in physically based rendering:
- Matt Pharr, Wenzel Jacob, Greg Humphreys, Physically Based Rendering: From Theory to Implementation (Third Edition), 2018. https://pbrt.org/ (https://pbrt.org/)