RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

FRANCE 2030
PROGRAMME DE RECHERCHE
NUMÉRIQUE POUR L'EXASCALE

Exascale Optimization

WP5

Presented by

Prof. El-Ghazali Talbi

Objectives and tasks of WP5

Objectives

- Solving large-scale optimization problems (decision variables, many-objectives, expensive objectives, big data) using Exascale optimization algorithms
- Continuous, and discrete optimization problems

Tasks

1. **Exascale discrete and continuous optimization**
   - Exact optimization (Branch and bound, tree search)
   - Heuristic optimization (Computational intelligence)
2. **Exascale surrogate-based and Bayesian optimization**
   - Parallel coupling of surrogates, optimization and sampling
3. **Exascale shape optimization**
   - Involving multi-physics models
4. **Exascale optimization for AutoML** (Optimization of deep/large ML models)

Main Partners

- Inria Bonus
- Unistra

- 2 Phds
- Still to recrut
  - 1 Engineer
  - 1 Postdoc

Sommaire

1. Intra-WP meeting

Future work

Scientific challenges

- **Exascale Parallel Bayesian Optimization based on Fractal optimization**

- **Resilience against faults**
- **Fault-tolerance strategies**
  - Same strategies for both classical and Bayesian optimization
  - Process failure → Recovery mechanisms → Reconstruction of fractals (low memory and time complexity)
  - Global failure → Checkpointing → Reduced I/O, distributed algorithm
  - Recovery → Independant subproblems, local surrogates

Future work

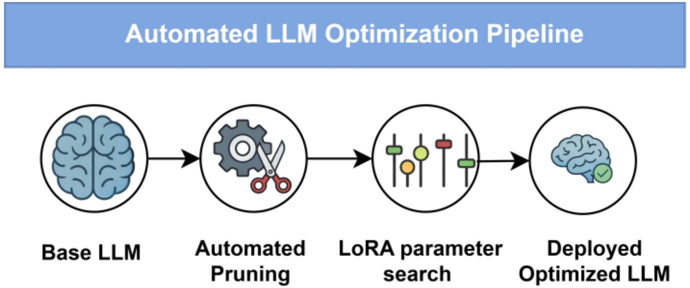
Scientific challenges

- **Generic framework for PDE-constrained optimization**
  - Shape optimization under parametric PDE constraints
  - Shape optimization with geometric constraints
  - High-dimensional shape optimization
  - Potential applications:
    - Mechanics,
    - Heat transfer

Future work

Scientific challenges

- **Application to AutoML - LLMs**
  - Fine tuning
  - Pruning



- **Application to AutoML - Spiking Neural Networks**
  - Hyper-parameter optimization
  - Architecture Search

Problem consists in finding the optimal ML model a^* maximizing f (accuracy & other objectives):

$$a^* = arg\ max_{a \in \mathrm{A}} f(\Theta(a, d_{train}), d_{valid}) = arg\ max_{a \in \mathrm{A}} f(a)$$

- Regression: RMSE (Root Mean Squared Error)
- Multi-class classification: Cross-entropy

Future work

Recrutment strategy WP5

- **Already recruted: 2 PhD students (INRIA Bonus, Uni Strasbourg)**
- **To recrut 1 Engineer**
  - Exascale [Bayesian] optimization algorithms
  - Software Zellij
- **To recrut 1 Postdoc** Exascale AutoML
  - Heterogeneous computing
  - Quantum-HPC → Neuromorphic-HPC → Quantum-Neuromorphic-HPC

2. Scimba and WP5

Scimba and AutoML

- **Scimba** : Python library that implements varying Scientific Machine Learning (SciML) methods for PDE problems, as well as tools for hybrid numerical methods.
- Solves parametric PDEs using various nonlinear approximation spaces such as neural networks, low-rank approximations, and nonlinear kernel methods.
- Compatible with both space–time algorithms (PINN, Deep Ritz) and time-sequential ones (discrete PINNs, neural Galerkin and neural semi-Lagrangian schemes).

Scimba and AutoML

- Current PINNs are not very sensitive to hyperparameters,
  - involve homogeneous & small architectures
  - Efficiency of Natural gradient

- Nowadays, Neural Architecture Search (NAS) for PINNs not essential. Good optimization methods (e.g. Natural gradient, Anagram, SSBroyden) are more crucial.

- More complex future architectures including Transformers, …
  - Hyper-parameter optimization
  - Neural Architecture Search (NAS) PINNs

- Use of Auto-ML strategy: Automatic selection of deep neural networks for neural operator-type methods

Scimba and Multi-objective optimization

- Multi-objective optimisation in PINNs (one residual per equation + boundary conditions)
  - Navier-Stokes : 3 equations ou plus
  - Many EDPs → Many objectives (Loss functions) → Pareto optimal solutions
  - Domain decomposition → Multi-PINNs (FBPINNs, Finite Basis PINNs)

- ## Multi-objective optimization
  - **Pareto-based optimization to generate the Pareto front**
  - **Extension of Exascale Fractal optimization to multi-objective optimization**

3. Uranie and WP5

Uranie is a software framework dedicated to uncertainty quantification sensitivity analysis, calibration and or generation of surrogate-models, optimisations, etc.

- Design-of-experiments generation

- Surrogate-models generation and training

- Sensitivity analysis methods

- Uncertainty dedicated visualisations tools

- Parametric optimisation methods (mono and multi-criteria, with or without constraint)

- Calibration methods, with or without uncertainties, with *a priori* knowledge for Bayesian estimations for instance.

Uncertainty quantification for optimization and AutoML

- Global sensitivity analysis for optimization with variable selection

- **Hilbert-Schmidt Independence Criterion (HSIC)**: HSIC-based sensitivity analysis to interpret and optimize ML models hyperparameters

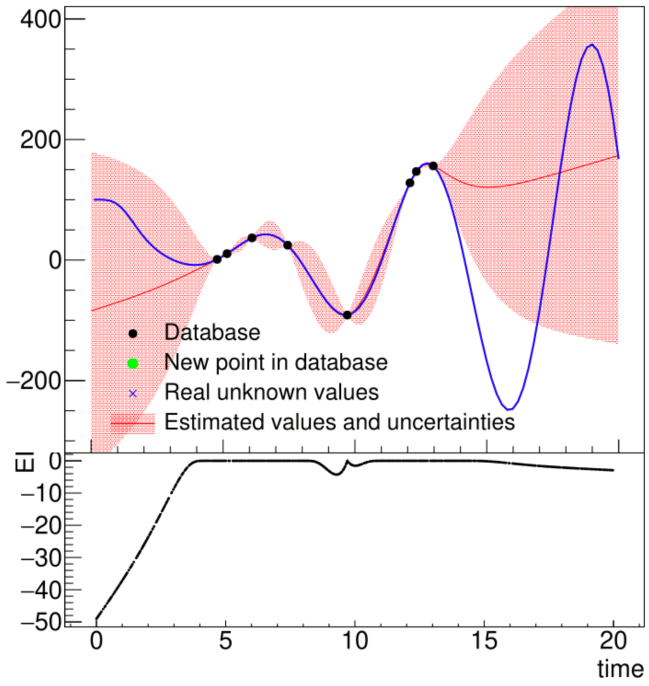
- Hyperparameter analysis to design online hierarchical multi-step approach involving different variables

- Design of more algorithms for hyperparameter optimization for **LLMs (Large Langage Models)** and **SNNs (Spiking Neural Networks)**

- Achieving competitive and cost-effective performance across machine learning and scientific tasks.
- Make hyperparameter optimization more interpretable.

Bayesian optimizer ?

- Uranie: Use of the EGO method (Efficient Global Optimization) → Intrinsically Sequential Algorithm
- Towards Exascale Bayesian Optimization Algorithms



Design

- Search space decomposition based on Fractals
- Low complexity (exploration, exploitation, tree search)
- Massively parallel → Exascale
- Handle very expensive objective functions

Implementation

- Will handle fault-tolerance
- Efficient heterogeneous implementation (load balancing, …)
- Portability on various heterogeneous architectures (CPUs, GPU)