2SC7691 – Optimization of a seismic exploration campaign for infrastructure protection

Instructors: Stephane Vialle, Filippo GATTI
Department: DOMINANTE - MATHÉMATIQUES, DATA SCIENCES, DOMINANTE - INFORMATIQUE ET NUMÉRIQUE
Language of instruction: ANGLAIS
Campus: CAMPUS DE PARIS - SACLAY
Workload (HEE): 80
On-site hours (HPE): 48,00

Description

Project topic in partnership with CEA-DAM.

After the accident in Fukushima (Japan), the use of high-performance computing has become increasingly common in estimating the seismic risk associated with nuclear power plants. These tools are of strategic importance not only in the context of the design of new installations, but also in order to study the performance of existing power plants in the face of extreme vents, not anticipated during their design. In this context, the SEISM Institute (of which CentraleSupélec and CEA are founders) was founded in 2012. It is a French scientific grouping, comprising academic and industrial partners (including CEA and EDF), with the 'objective of bringing together the various know-how in seismology and earthquake engineering to improve the prediction of the seismic response of critical sites and structures in France, as well as the assessment of the associated risk.
In this context, this project concerns the optimization of a geophysical exploration campaign on an experimental site, using its digital twin, built using a wave propagation code (SEM3D) in development between CentraleSupélec, CEA and the Institute of Globe Physics. SEM3D simulates the propagation of seismic waves over large 3D domains, with domain decomposition on a Cartesian (or spherical) mesh. It also integrates the site topography and complex geological structures. The project therefore consists of solving an inverse problem in order to optimize - using SEM3D - the geological configuration of the site of interest. This optimization is based on the *Reverse Time Migration* method (i.e. resolution by adjoint problem). The optimization strategy provides for many realistic simulations, from source to sensors (forward) and back-propagation of the misfit (backward) in order to be able to update iteratively the mechanical properties of the subsoil. Indeed, given the size of the site of interest (~10 km wide) and the spatial resolution sought (~100m), although SEM3D is parallelized and distributed on supercomputers, each wave propagation simulation can last

several hours on many shared computing cores. For this, at each iteration, the *Forward* and *Backward* steps must be properly chained with an appropriate job scheduling strategy (launch of batch calculations). Finally, the number of sensors for in situ recordings must be reduced, given the associated costs, in terms of sensors, acquisition campaigns and storage of the data obtained.

The objective of this study is therefore threefold:

1. propose a geology model minimizing the difference between simulation and records,

2. minimize the number of sensors required to arrive at a model at a reasonable financial cost (considering their spatial layout),

3. manage to design this solution over the duration of the project with high-performance computers and with a limited quota of computing hours.

For this purpose, an optimization loop will be developed using the wave propagation simulation code as efficiently as possible: by sparingly exploring the space of possible configurations, to economically find a good solution.

Technical details of the system:
The studied system consists of a sedimentary basin surrounded by outcropping bedrock, possible candidate for the construction of a new nuclear power plant. To evaluate the seismic response of the site and to propose possible earthquake scenarios, one needs to know:

- the 3D geometry of the geological layers,
- the mechanical properties of these layers.

This information is fundamental for the definition of site effects on seismic energy radiated by an active fault.

Quarter number
ST7

Prerequisites (in terms of CS courses)
First year courses:

- SG1 common course "Systèmes d'Information et Programmation" (1CC1000)
- ST2 common course "Algorithmique et complexité" (1CC2000)

Courses of the ST:

- ST7 common course "Optimisation" (2CC3000)
- ST7 specific course "Méthodes et algorithmes parallèles pour l'optimisation" (2SC7610)

Others prerequisites:

- Parts of common course "CIP - Convergence, Intégration et Probabilités" (1SL1000)
- Parts of common course "EDP - Equations aux dérivées partielles" (1SL1500)
- Knowledge of linear algebra will also be needed

Syllabus

Main tasks of the study:

- Presentation of the subject at the beginning of the first part of the project (intervention of CEA-DAM):
    - presentation of theoretical fundamentals in wave propagation in complex configurations,
    - presentation of scientific issues related to design earthquake-resistant nuclear power plants, to the evaluation of the seismic response of the site and the estimation of the associated risk.
- Formalization of the problem (and of its characteristic sizes) represented by the geophysical investigation to characterize the seismic response of a nuclear site. Choice of an optimization method adapted to problem. Development of the algorithm on a verification case, supported by the analytical solution.
- Identification of SEM3D code parameters and their respective impacts on its parallel execution time, handling of computing resources of the Moulon Mesocentre.
- Conception and implementation in Python of an optimization code calling SEM3D. The optimization code will itself be parallelized if its algorithm allows it (we would then have a parallel master code calling on request a parallel code).
- Test and debugging of the complete parallel optimization code on the Moulon Mésocentre supercomputers, on small and medium problems (low frequency and / or small size of the domain).
- Application to a real case (in terms of the size of the area studied, and maximum frequency):
    - test larger problems on a larger number of nodes and CPU cores (approach of scaling),
    - analysis of the quality of the solution found and of the performance of calculations,

- code optimization to improve the solution found AND, if required, the performance of calculations,
- Estimate the minimum number of sensors needed to find a workable solution based on available resources and maximum duration of the study.
- The study will conclude with the submission of a report and a presentation to be evaluated: the quality of the solution found, the effectiveness and the possible extensions of the crafted optimization code, and the management of the computational resources and quota during the project.

Rmk : Different student groups will experiment different optimization methods, but they will test all several examples of *Reverse Time Migration* (terrestrial and marine geophysical prospecting).

Class components (lecture, labs, etc.)

Part 1 (40HEE):

- Steps 1 and 2: course complements, handling of computing resources, formalization of the problem, choice of an optimisation algorithm.
- Step 3 and 4: Python numerical implementation of the optimisation algorithm used for the Reverse Time Migration method. Result analysis of the geophysical propecting campaigns, in order to choose the parameters to optimize and to identify the sources of uncertainty and experimental noise.
- Step 5: first implementation of the optimisation algorithm on parallel machine, evaluation of the result quality and computation performances on small and medium size problems, and comparison to a reference analytical solution.

- Midterm report (slides and progress summary) and talk about current results and future work (part 2).

Part 2 - *final sprint* (40HEE):

- Step 6: experiments on real problems with different configurations, identification of the sources of performance locks and losses, and (if required) code improvement to push the limits of supportable problem size.
- Step 7: estimate of the number of sensors/records required for a stable optimisation, function of the available computing resources and of the maximum duration of the project.

- Final report (slides and extended abstract) and global talk about the project approach and results.

Grading
This project will be evaluated by a midterm talk at the end of part 1 (40HEE), and by a final talk at the end of part 2 (*final sprint* 40HEE). Talks will be done by the entire team, but will lead to individual marks in case of strongly heterogeneous teams.Each talk evaluation will consider the overall quality of the talk, of the slides and of the progress summary. Each talk mark will be 50% of the total mark.

Resources
Teaching staff:

- **F. Gatti** (CentraleSupélec & MSSMat)
- **M. Bertin** (CEA-DAM)

Workplace and computing resources:

- Students will work at CentraleSupelec, in a classroom with electrical outlets and reliable wifi Internet access (except lockdown).
- Students will use their laptops to connect to remote **PC clusters** at **Moulon Mesocenter.**
- Final oral exam will take place at CentraleSupelec the last afternoon of the project.

Learning outcomes covered on the course

When finishing the course, the students will be able to:

- **Learning Outcome 1 (AA1):** identify and parametrize an optimisation method adapted when each evaluation/iteration requires many computing resources and computation time,
- **Learning Outcome 2 (AA2):** implement and debug a sequential or parallel Python code on supercomputer (developped from scratch or using libraries), calling distributed C/C++ computing kernels,
- **Learning Outcome 3 (AA3):** deploy intensive computing applications on remote resources,
- **Learning Outcome 4 (AA4):** identify the limitations of the study, function of the available computing resources,
- **Learning Outcome 5 (AA5):** manage a computing resource quota, during an intensive computing campaign.

Description of the skills acquired at the end of the course

- C4: Have a sense of value creation for his company and his customers
- C7: Know how to convince
- C8: Lead a project, a team