Internship position:

Solving sparse linear systems with modular and adaptive mixed precision Krylov methods

Internship level: Master 2

Duration: 6 months approximately

Location: Laboratoire LIP6, Sorbonne Université, 4 place Jussieu, Paris 5e

Contact: Théo Mary (theo.mary@lip6.fr), Emmanuel Agullo (emmanuel.agullo@inria.fr), Luc Giraud (luc.giraud@inria.fr), Fabienne Jézéquel (fabienne.jezequel@lip6.fr), Pierre Jolivet (pierre.jolivet@lip6.fr)

Salary: on the basis of CNRS internship allowances (27.30 euros/working day and reimbursement of part of transport costs)

Scientific context:
Solving sparse linear systems is one of the fundamental problems of scientific computing. Iterative methods based on Krylov subspaces (conjugate gradient, GMRES, etc.) are capable of scaling up to very large problems, but however require costly linear algebra operations to ensure their rapid convergence: preconditioning, orthogonalization, etc. Solving large sparse linear systems in the way that is both frugal and reliable therefore constitutes one of the major current challenges in the field.

This internship seeks to tackle this challenge by developing mixed precision methods [1], which exploit several different levels of precision in order to best optimize their resource consumption. Indeed, modern computing architectures have several precisions implemented in hardware, notably double (64 bits), single (32 bits), and half (16 bits) precisions. Computations carried out in low precision (32 or even 16 bits) are much faster and more energy efficient. However, most scientific computing applications require a final solution equivalent to 64 bits.

Main objectives:
The main objective of the internship is therefore to develop Krylov methods in mixed precision that only reduce the precision at certain well-chosen places. This objective will be addressed through two axes: modularity and adaptivity.

Modularity consists in considering each of the core kernels used by a method (in the case of Krylov: matrix-vector product, preconditioning, orthogonalization, etc.) as being independent. For example, some studies have shown the interest of performing the preconditioned matrix-vector product with a precision different from the rest of the kernels [2]. The internship will seek to develop a completely modular Krylov method, in particular by dissociating the preconditioned matrix-vector product into two kernels: the product with the matrix and the product with the preconditioner [3, chapter 7].

Adaptivity consists in dynamically adapting the precisions of each operation of a method during its execution, according to the data provided as input [1, section 14]. This class of approaches has, for example, proven efficient for the matrix-vector product [4]. The internship will seek to develop adaptive versions of other more complex kernels, such as orthogonalization or preconditioning, and will study the synergy of this type of approach with inexact Krylov methods [5] which reduce the overall precision during iterations.

Each of the two axes involves both a mathematical component, the aim of which is to carry out theoretical error analyses to rigorously determine the places where precision can be reduced without impacting the final quality of the solution, and a computer science component, whose goal is to develop efficient implementations of these new methods to exploit modern parallel computing architectures.

Work environnement:
The internship will take place in the PEQUAN team of the LIP6 laboratory of Sorbonne University (Paris), and will be carried out in collaboration with the CONCACE team of the Inria Bordeaux center. At least one short-term stay in Bordeaux is planned during the internship.

Continuation with a PhD thesis:
The internship could lead to a PhD position, in the context of the national project NUMPEX lead by CNRS, CEA, and Inria, whose goal is to design and develop the software stack that will be used by the future exascale computers with applications to climate, the energy transition, health, AI and the industry.

References to go deeper:
[1] N. Higham, T. Mary. Mixed precision algorithms in numerical linear algebra, https://hal.archives-ouvertes.fr/hal-03537373
[2] P. Amestoy, A. Buttari, N. Higham, J.-Y. L'Excellent, T. Mary, B. Vieublé. Five-precision GMRES-based Iterative Refinement, https://hal.archives-ouvertes.fr/hal-03190686
[3] B. Vieublé. Mixed precision iterative refinement for the solution of large sparse linear systems, https://filedn.eu/lyDamYXd3jSjd2K2sKty8BJ/thesis.pdf
[4] S. Graillat, F. Jézéquel, T. Mary, R. Molina. Adaptive precision sparse matrix-vector product and its application to Krylov solvers, https://hal.science/hal-03561193
[5] L. Giraud, S. Gratton and J. Langou. Convergence in backward error of relaxed GMRES, https://doi.org/10.1137/040608416