A new max-based compression algorithm for surrogate modelling. Application to the processing of neutronics data

Olivier TRUFFINET CEA Saclay - SERMA/LPEC laboratory
olivier.truffinet@cea.fr

Many research and industry codes tend to generate large amounts of intermediate data, for instance under the form of tabulated values of physical quantities. This data can be quite redundant (i.e of small effective dimension), justifying the use of lossy compression methods such as Singular Values Decomposition (SVD) for storing and accessing it more efficiently. But said redundancy could also be exploited in other ways, for instance by extrapolating it with some *surrogate model* to reduce unnecessary calculation on close configurations.

This is the lead we are following in this article, by introducing a new compression tool based on the Empirical Interpolation Method, an algorithm initially developed in the framework of partial differential equations [2]. Unlike SVD, this EIM-compression method is based on the infinite norm $\|\cdot\|_\infty$, and proceeds in a greedy way by iteratively trying to approximate the data and incorporating the chunks of information which cause the largest error. In the process, it provides a vector basis and a set of interpolation points, which can be used to approximate future data from very little information. The algorithm is competitive in terms of speed and accuracy, but also very suitable for parallelization and out-of-core computation (processing of data too large for the computer RAM).

We apply this algorithm to a neutronics problem: the computation of homogenized cross-sections. These quantities, which measure the interaction of neutrons with matter, are generated in large amounts (up to hundred of gigabytes) by nuclear reactor simulators, stored as tabulated multivariate functions, and have already been shown to be highly redundant [1]. The already mentionned EIM basis and interpolation points enable us to build an elementary surrogate model for these cross-sections: they make it possible to reconstruct a full grid of any of them using only its values at a small number of points. By training this model on a well-chosen subset of the data, we can greatly speed up the calculation by interpolating most of the sections instead of computing them explicitly. We assess the performance of this method on realistic nuclear data, and discuss the impact of several modeling choices - especially normalization, which is of great importance in the problem at hand.

Joint work with: Karim AMMAR, Bertrand BOURIQUET, Nicolas GERARD-CASTAING.

References

[1] M. Barrault & al. An empirical reduced-basis interpolation method: application to efficient discretization of partial differential equations. *C. R. Acad. Sci. Paris, Ser.I 339*, Numerical Analysis, 2004.

[2] D. Tomatis. A multivariate representation of compressed pin-by-pin cross sections. *EPJ Nuclear Sciences & Technologies*, 2021.