Nested Metasampling for Material Science

State of the art and objectives

We propose an interdisciplinary project involving different fields such as Bayesian data analysis, statistical physics and applied mathematics for the efficient exploration and minimization of complex scalar functions. The project is centered on the *Nested Sampling* optimization algorithm [1, 2] based on Bayesian statistics to explore the relevant parameter space recursively effectively, obtain the topology thereof, and assess the quality and reliability of the obtained optimum within a pre-determined statistical error.

Nested sampling works with a finite set of *live points* in the multi-dimensional space in which one repeatedly substitutes the *worst* live point with a *better* one.[1] This evolves towards a narrower and narrower "nested" volumes of parameter space, in the end to obtain an optimum. Within that process the parameter space is efficiently sampled in order not only to converge to an optimum but also to present a view on the overall landscape, allowing a better assessment of the result and to access the topology of the function to be optimized. The method, which was proposed 20 years ago, has been mainly applied to data analysis [2–4]; recently, we contributed to assess its application to quantum statistical physics by simplifying the calculation of the quantum partition function Q [5]. The evaluation of Q, which involves summing the quantity $\exp(-E(\vec{x};\{a_i\})/k_B T)$ over all accessible states [6], is crucial, as all thermodynamic properties can be derived from it.[2] For a single system configuration $\vec{x}$, distinct $E(\vec{x};\{a_i\})$ models are available, each corresponding to a given accuracy and computational complexity. The bottleneck is the speedy evaluation of $E(\vec{x};\{a_i\})$ for as large as possible samples of complex systems.

To date, the most used theory in material science is the *Density Functional Theory* (DFT), which allows at most the simulation of a few hundreds of atoms for a few picoseconds, at a viable computational cost. Moreover, the DFT energy function has no analytic expression. An alternative are *machine-learning* generated energy $E(\vec{x};\{a_i\})$ models. These can provide excellent results at a much lower cost (up to 10^3 smaller than the DFT) but are strongly dependent on the training process as, while they interpolate well, extrapolations will generate physically unacceptable errors. Therefore, scrutinizing the quality of machine-learning models, once trained via a set of atomic configurations and the corresponding DFT energy and gradients, can be unworkable, especially for high-energy configurations that can often be outside the initial training set. Furthermore, the neural network weights form a very large parameter space and have no physical meaning so far.

Finally, analytical interaction models, such as the Lennard-Jones or Morse potentials, variants and combinations thereof, based on physical principles and fitted to theoretical and experimental data, can be much more efficient, but are usually limited to a narrow range of conditions. Finding the correct analytical expression and optimal set of parameters is a rather lengthy process requiring both expertise and intuition. This results in a maze of different expressions adapted to specific situations. Indeed, the extension of analytical interatomic energies $E(\vec{x};\{a_i\})$ to a wide ensemble of conditions and the systematic assessment of their reliability is still an open problem, although some models date back to a century ago. In this respect, nested sampling can tell us to what extent a given model is reliable and whether adding another parameter to it makes it better or not. This will provide a highly valuable mathematical guideline in this quest that will be complemented by additional mathematical methods for the reduction of pertinent parameters [7, 8]. Moreover, because the parameters $\{a_i\}$ result from highly non-linear fitting, nested sampling (or, better, *metasampling* as we work in the parameter space) can describe the topology of the whole parameter space, including secondary minima.

Working plan

1. For a given material, compute a database of statistically uncorrelated atomic configurations $\{\vec{x}\}$ including the energy $E(\vec{x})$ and gradients $\nabla_{x_j} E(\vec{x})$ for each configuration as obtained via the DFT or similar methods;

---

[1] The quality of live points is measured in terms of a positive-definite object function to be minimized.

[2] $E(\vec{x};\{a_i\})$ depends on a vector of dimension N_{dof} that defines the microscopic state and some parameters $\{a_1, \ldots, a_P\}$. The partition function sums up over the N_{dof} variables.

2. Choose a starting model for the potential among those available in the literature, preferably with a small number of parameters. Using the results from the DFT as a reference, evaluate the likelihood of that model;

3. Increase the complexity, flexibility by adding new parameters and simplifying the problem with mathematical tools, while assessing the improvement or lack of it using nested sampling and the computational speed-up;

4. Compare the performance of the new model in reproducing DFT data;

5. Either use the improved model in standard simulations, or use nested sampling again to sample more precisely the configurational phase space for a partition function evaluation and statistical (quantum) properties of the material. A similar strategy can be applied to find effective Hamiltonians working on a relevant subset of the microscopic space ($N_{eff} \ll N_{dof}$).

6. When studying the quantum effects of light nuclei in condensed matter (such as H diffusion in real materials, with applications for energy production and storage or giant isotope effects – see below), apply our *extended partition function* method [5] that we successfully tested in small rare-gas clusters.

As first applications, we choose some selected problems in materials science as inspired by recent works:

- High-pressure ice and water/ammonia alloys as observed inside planets in which the combination of hydrogen bonds and nuclear quantum effects [9] is at work.

- NaOD versus NaOH. The former does exhibit a structural quantum phase transition while the latter isotope does not [10], which indicates giant nuclear quantum effects. Both the partition function issue and a satisfactory effective Hamiltonian remain to be set up.

Perspectives and candidate's skills

The proposed work will deal both with practical studies of real materials and formal developments of the method. Mathematical input was vital in previous work and will remain so. The rigorous assessment of model potentials $E(\vec{x}; \{a_i\})$ on the basis of Bayes' statistics could have deep consequences on their improvement as well on the simplification of machine-learning potentials. The new mathematical tools that will be developed all along the project could spread in other fields, such as statistics and data analysis.

The ideal candidate has a background in quantum and statistical physics, applied mathematics and good programming skills. He/She will benefit from a very open scientific and interdisciplinary environment: the project brings together physicists with distinct expertise (Fabio Finocchi for the theory of solid state and Martino Trassinelli for Bayesian methods for atomic spectroscopy at INSP), in close collaboration with Julien Salomon at INRIA for his mathematical skills in model selection algorithms. We also expect to benefit from the SCAI environment, exchanging with researchers in applied mathematics and statistical learning.

References

[1] J. Skilling, AIP Conf. Proc. **735** (2004), 395–405.
[2] G. Ashton, N. Bernstein, J. Buchner, et al., Nat. Rev. Methods Primers **2** (2022), 39.
[3] **M. Trassinelli**, Nucl. Instrum. Methods B **408** (2017), 301–312.
[4] L. Maillard, **F. Finocchi, and M. Trassinelli**, Entropy **25** (2023), 347.
[5] L. Maillard, P. Depondt, **F. Finocchi**, S. Huppert, T. Plé, **J. Salomon, and M. Trassinelli**, in preparation for J. Chem. Phys.
[6] B. Szekeres, L. B. Pártay, and E. Mátyus, J. Chem. Theory Comput. **14** (2018), 4353–4359.
[7] B. Haasdonk, **J. Salomon**, and B. Wohlmuth, SIAM J. Numer. Anal.
[8] O. Burkovska, B. Haasdonk, **J. Salomon**, and B. Wohlmuth, SIAM J. Financ. Math. **6** (2015), 685–712.
[9] N. Avallone, S. Huppert, P. Depondt, L. Andriambariarijaona, F. Datchi, S. Ninet, T. Plé, R. Spezia, and **F. Finocchi**, Phys. Rev. Lett. **133** (2024), 106102.
[10] S. Schaack, E. Mangaud, E. Fallacara, S. Huppert, P. Depondt, and **F. Finocchi**, Phys. Rev. Lett. **131** (2023), 126101.