

Dr. P. Hadjidoukas, Prof. C. Papadimitriou

ETH Zentrum, CLT E 13

CH-8092 Zürich

Project # 3

Issued: 15.05.2017

- In these scripts, we outline five computational engineering problems. Choose one and work either individually or in a group of two people. Communication is allowed to the extent that you do not copy the work of other groups.
- You are encouraged to contact one of the TAs in order to arrange a meeting to discuss your chosen project. These meetings are meant for clarifying and detailing the projects, give early feedback on your approach, *not* for correcting your code.
- In evaluating your work, we will consider your ability to analyze the problem and the hardware at your disposal, to appropriately apply the principles taught during the whole HPCSE class, and to report your reasoning and findings.
- The report (including text, code, figures) needs to be emailed before the day of the exam. If working in a group of two, each student has to write an *individual* report.

Molecular Dynamic Potentials

A potential in molecular dynamics (MD) is a function which computes the interaction energy of a set of particles. In this project we consider pairwise potentials, i.e. those which involve two particles. For more information you can check Wikipedia [1].

Lennard-Jones Potential

The Lennard-Jones (LJ) potential is a widely used tool to model particle interactions in MD simulations. It is a pairwise potential consisting of a repulsion and attraction terms:

$$V_{LJ}(r; \varepsilon, \sigma) = 4\varepsilon \left(\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right), \quad (1)$$

where ε, σ are material-specific constants and r is the distance between the interacting particles.

Buckingham Potential

The Buckingham potential is another popular model for computing interactions for a pair of particles. It has the following form:

$$V_{Bu}(r; A, B, C) = A \exp -Br - \frac{C}{r^6}, \quad (2)$$

where A, B, C are material-specific constants and r is the distance between the interacting particles.

Method description

We provide a brief description of the Bayesian inference theory. In the following small bold letters represent vectors while big bold letters represent matrices. Each random variable ξ is assumed to be continuous with a probability density function (PDF) denoted as $p(\xi)$.

We denote by $f(\mathbf{x}; \boldsymbol{\vartheta}) \in \mathbb{R}^M$ the output, or a quantity of interest (QoI), of a computational model with input $\mathbf{x} \in \mathbb{R}^{N_x}$ and parameters $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_{N_\vartheta}) \in \mathbb{R}^{N_\vartheta}$. We denote as $\mathbf{d} \in \mathbb{R}^{N_d}$ a vector of experimental data corresponding to the QoI f and the input parameters \mathbf{x} . The experimental data are linked with the computational model through the likelihood function, $p(\mathbf{d} | \boldsymbol{\vartheta}, \mathbf{x})$. A usual model assumption for the likelihood function involves a Gaussian,

$$p(\mathbf{d} | \boldsymbol{\vartheta}, \mathbf{x}) = \mathcal{N}(\mathbf{d} | f(\mathbf{x}; \boldsymbol{\vartheta}), \boldsymbol{\Sigma}), \quad (3)$$

where $\boldsymbol{\Sigma}$ is a covariance matrix that may be a function of $\boldsymbol{\vartheta}$. To simplify the notations, the conditioning on \mathbf{x} is omitted below. Prior information on the parameters $\boldsymbol{\vartheta}$ is encoded into the probability distribution with PDF $p(\boldsymbol{\vartheta} | \mathcal{M})$. We assume $\boldsymbol{\Sigma} = \sigma_n^2 \mathbf{I}$, where \mathbf{I} is the identity matrix in $\mathbb{R}^{N_\vartheta \times N_\vartheta}$ and $\sigma_n \in \mathbb{R}$ is *a priori* unknown. Bayes' theorem provides a tool for the inference of the parameters $\boldsymbol{\vartheta}$ conditioned on the observations \mathbf{d} ,

$$p(\boldsymbol{\vartheta} | \mathbf{d}, \mathcal{M}) = \frac{p(\mathbf{d} | \boldsymbol{\vartheta}, \mathcal{M}) p(\boldsymbol{\vartheta} | \mathcal{M})}{p(\mathbf{d} | \mathcal{M})}, \quad (4)$$

where $p(\mathbf{d} | \mathcal{M}) = \int p(\mathbf{d} | \boldsymbol{\vartheta}, \mathcal{M}) p(\boldsymbol{\vartheta} | \mathcal{M}) d\boldsymbol{\vartheta}$ is a normalization constant and \mathcal{M} stands for "model", which is a set of the assumptions regarding the likelihood and the prior. We remark that the denominator $p(\mathbf{d} | \mathcal{M})$, called model evidence, is used for model selection.

For the sampling of the distributions we will use the Transitional Markov Chain Monte Carlo (TMCMC) algorithm.

Problem setup

Your tasks in this project are the following:

- Extend the available OpenMP implementation of the TMCMC algorithm on distributed memory architectures, by applying the hybrid MPI + OpenMP programming model. You can use an appropriately large number of samples to evaluate the performance of your implementation.

- Extend the available OpenMP implementation on GPUs using the CUDA programming model. For this case, you can tune the parameters of the TMCMC algorithm so as to use the same length for all MCMC chains.
- Infer the parameters of both potentials using the quantum dimer calculations of argon. The data comes from Ref. [2] and is provided in the file `data.txt`. Report your setup: the likelihood, the number of samples per stage, the prior bounds of the parameters. Can you identify all the parameters? Which of the two potentials is preferred by the Bayesian model selection?

References

- [1] https://en.wikipedia.org/wiki/Molecular_dynamics#Potentials_in_MD_simulations.
- [2] A. M. Halpern and T. Haute. Structural and thermodynamic properties of the argon dimer a computational chemistry exercise in quantum and statistical mechanics. *J. Chem. Educ.*, 87(2), 2010.