Finding minimum energy paths using Gaussian process regression*

Olli-Pekka Koistinen
Helsinki Institute for Information Technology HIIT
Department of Computer Science, Aalto University
Science Institute and Faculty of Physical Sciences, University of Iceland
Department of Applied Physics, Aalto University
olli-pekka.koistinen@aalto.fi

Freyja B. Dagbjartsdóttir
Science Institute and Faculty of Physical Sciences, University of Iceland

Vilhjálmur Ásgeirsson
Science Institute and Faculty of Physical Sciences, University of Iceland

Aki Vehtari
Helsinki Institute for Information Technology HIIT
Department of Computer Science, Aalto University
aki.vehtari@aalto.fi

Hannes Jónsson
Science Institute and Faculty of Physical Sciences, University of Iceland
Department of Applied Physics, Aalto University
hj@hi.is

Abstract

In computational chemistry, systems of atoms are often characterized by a potential energy surface defined as a function of the atom coordinates. When determining the mechanism and rate of transitions between different configurations, it is useful to find a minimum energy path connecting the energy minimum points corresponding to the initial and final state. Such paths are frequently calculated using the nudged elastic band method, where an initial path is iteratively shifted to the nearest minimum energy path. The computational effort can be large, since the method typically requires hundreds of expensive evaluations of energy and its gradient. Here, we show how the number of such evaluations can be reduced by an order of magnitude using a Gaussian process regression approach where an approximate energy surface is generated and refined in each iteration.

1 Introduction

In computational chemistry, systems of atoms are often characterized by a potential energy surface (PES), a function of the atom coordinates, where the local minimum points correspond to the stable

*This is a summary of an article published under the name "Nudged elastic band calculations accelerated with Gaussian process regression" in *the Journal of Chemical Physics* **147**, 152720 (2017).

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

states of the system. When determining the mechanism and rate of transitions between two states, it is useful to find a minimum energy path (MEP) connecting the corresponding energy minimum points. The most interesting point on the MEP is its highest energy point, which is a first-order saddle point on the PES. Mathematically, a MEP can be defined with a requirement that any point on the path is at an energy minimum in all directions perpendicular to the path. With this definition, there can potentially be multiple MEPs connecting the two states.

The nudged elastic band (NEB) method is a commonly used iterative approach to find MEPs [1]. In the NEB method, the path between two local minimum points on the energy surface is represented by a discrete set of points, referred to as 'images'. Starting from an initial guess, the idea is to move the images downwards on the energy surface but keep them evenly spaced. The images are moved along an NEB force vector, which is a resultant of two components: negative energy gradient projected on the direction perpendicular to the path and a spring force (parallel to the path), which tends to keep the images evenly spaced. The path has converged to an MEP when the magnitude of the NEB force vector is zero (to some tolerance) for all images.

Each NEB calculation typically involves on the order of a hundred evaluations of the energy and its gradient for each one of the images, and the path is typically represented by five to ten images. The evaluations were initially performed mostly using analytical potential energy functions, but nowadays electronic structure calculations are also used extensively in NEB applications. Since a typical electronic structure calculation takes on the order of tens of CPU minutes or more, the NEB calculations can become computationally demanding. In addition, the calculation may need to be repeated if there are several possible final states for the transition. Thus, it would be valuable to find ways to accelerate NEB calculations. To get the most out of the computationally intensive electronic structure calculations, the information obtained from them should be exploited better to decrease the number of NEB iterations instead of forgetting it after one iteration.

2 Gaussian process regression approach

The use of machine learning to accelerate MEP and saddle point calculations has been introduced by Peterson [2], who applied neural networks to construct an approximate energy surface for which NEB calculations were carried out. After relaxation of the path on the approximate energy surface, the accurate energy and gradient were evaluated at the images of the relaxed path to see whether or not the path had converged on an MEP on the true energy surface. If true convergence had not been reached, the new energy and gradient calculated at the location of the images were added to the training data set and the model was updated. This procedure was repeated iteratively until the approximate energy surface was accurate enough for converging on the true MEP.

Proof-of-principle results have also been presented where Gaussian process regression (GPR) [3-6] is applied to accelerate NEB calculations [7]. Since the calculations are largely based on the gradient vector of the energy surface, straightforward inclusion of derivative observations [8-12] and prediction of derivatives can be seen as advantages of GPR for this application. It is also easy to encode prior assumptions about the smoothness properties of the energy surface into the covariance function of the Gaussian process (GP) model or learn about these properties from the data. Analytical expressions for the posterior predictions conditional on the hyperparameters of the GP model allow both fast predictions and reliable estimation of uncertainties. The predictive performance of GPR has been shown to be competitive with other machine learning methods especially when the number of observations is small [13].

The GPR approach to MEP calculations is extended here by introducing two algorithms to accelerate climbing image nudged elastic band (CI-NEB) calculations, where one of the images is made to converge to a small tolerance on the highest energy maximum along the MEP [14]. The basic GPR approach is referred to as the all-images-evaluated (AIE) algorithm, where the accurate energy and its gradient are evaluated at all intermediate images of the CI-NEB path before the approximation to the energy surface is updated (see figure 1).

In a more advanced algorithm, the accurate energy and its gradient are evaluated at only one image before a new approximate energy surface is constructed. We refer to the latter as the one-image-evaluated (OIE) algorithm (see figure 2). As a probabilistic model, a GP expresses the energy predictions as probability distributions, which means that the uncertainty of the prediction can also be estimated, e.g., as the variance of the posterior distribution. This uncertainty estimate is used by the

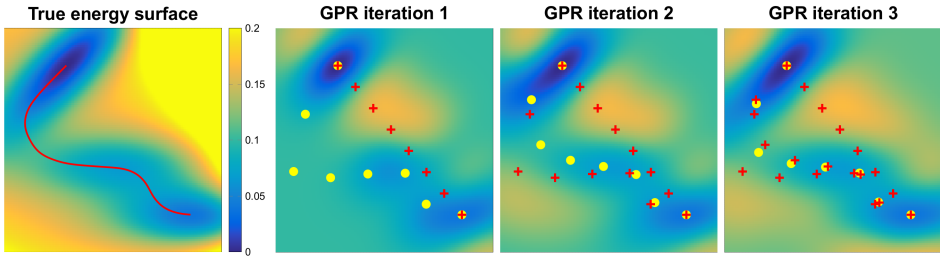

Figure 1: Illustration of the progression of the AIE algorithm. Far left: Two-dimensional Müller-Brown energy surface [15] and the true minimum energy path (red curve). Three panels to the right: The approximate energy surface predicted by GPR after one, two and three iterations. The red + signs show the points at which the accurate energy and its gradient have been evaluated. The yellow disks show the CI-NEB relaxed on the approximate energy surface of each GPR iteration. After each GPR iteration, final convergence of the path is checked by accurate evaluations, which are then added to the training data for the following GPR iteration. Final convergence is confirmed after 24 evaluations.

OIE algorithm to select the image to be evaluated in such a way as to give maximal improvement of the model. This algorithm has similarities with Bayesian optimization in the sense that uncertainties of a GP model are used to select the locations of new evaluations. However, our approach to find a local minimum energy path, where the optimal location of an image depends also on the locations of the neighbouring images, is not fully analogous to the standard framework of Bayesian optimization, where the GP is used to model a single objective function and the global optimum is searched by exploring the whole domain.

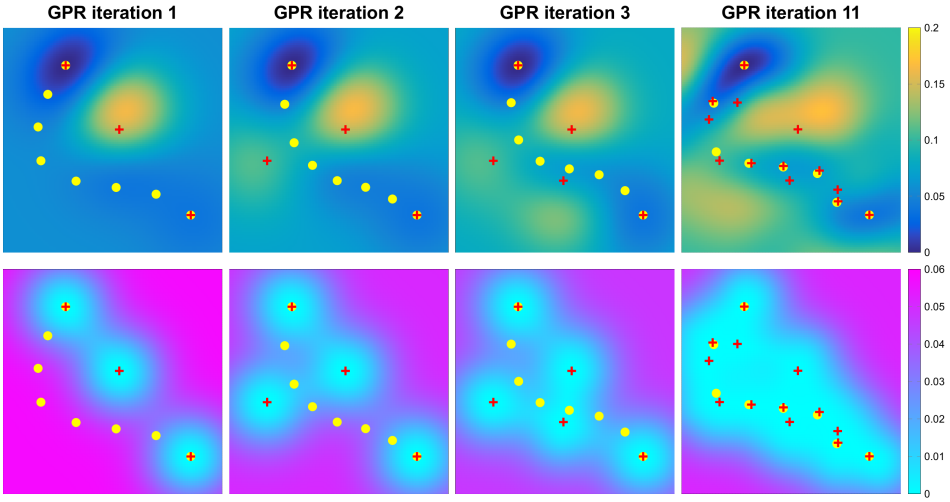

Figure 2: Illustration of the progression of the OIE algorithm. Upper panel: The approximate energy surface predicted by GPR after one, two, three and eleven iterations. After iterations 1, 2 and 3, the accurate energy and its gradient are evaluated at the image where the estimated uncertainty is largest and the observed data are then added to the training data set for the following GPR iteration. Lower panel: The estimated uncertainty (standard deviation) of the energy approximation. After eleven iterations, the path is not displaced further but the final convergence is checked by evaluating the energy and gradient at each intermediate image one by one, Final convergence is confirmed after 17 evaluations.

Another extension of the GPR approach applies when the overall goal is to estimate the forward and backward transition rates, which requires evaluating the Hessian matrices including second derivative information of the PES at the two minimum points. The evaluation of the Hessians can be done already before finding the MEP to improve the GPR approximations (helps especially in the beginning when there is little information about the PES).

3 Experiments

A test problem that has been used in several studies of algorithms for finding MEPs and saddle points involves an island of seven atoms on the (111) surface of a face-centered cubic (FCC) crystal [16,17]. The initial, saddle point and final configurations of the atoms for the 13 lowest activation energy transitions, labeled from A to M, are shown in figure 3. The interaction between the atoms is described with a simple Morse potential to make the implementation of the benchmark easy.

The average number of energy and gradient evaluations for transitions C – M as a function of the number of degrees of freedom is shown in figure 3.[1] For the smallest number of degrees of freedom, 21, only the seven island atoms are allowed to move while all the substrate atoms are frozen. For the larger numbers of degrees of freedom, some of the surface atoms are also allowed to move. Starting with the AIE algorithm, the use of the Hessian input reduces the number of evaluations by about 20%, but the transition to the OIE algorithm has an even larger effect, a reduction to a half. The OIE results represent savings of an order of magnitude with respect to the regular CI-NEB calculation.

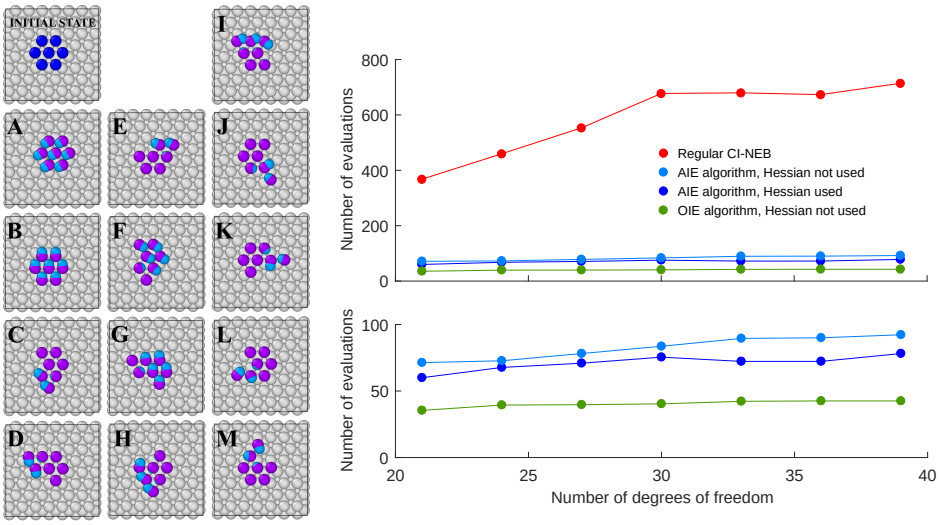

Figure 3: Left: The initial (dark blue disks in uppermost left column), saddle point (light blue disks) and final configuration (purple disks) for the 13 heptamer island transitions. Right: Average number of energy evaluations for the conventional NEB and variations of the GPR approach.

4 Conclusions

The results presented here show that the GPR approach can reduce the number of energy and gradient evaluations needed in CI-NEB calculations of MEPs by an order of magnitude. This is important since a large amount of computer time is used in such calculations. In the future, it will be important to test the GPR approach also on more complex systems to be able to fully assess its utility and to develop the methodology further.

Acknowledgements

We thank Professor Andrew Peterson for helpful discussions. This work was supported by the Icelandic Research Fund and the Academy of Finland (Grant No. 278260). We acknowledge the computational resources provided by the Aalto Science-IT project and the Computer Services of the University of Iceland.

[1] Transitions A and B are not included in the averages shown in figure 3 because the regular CI-NEB required an anomalously large number of iterations for some of the intermediate numbers of degrees of freedom. The results of the GPR algorithms were, however, similar for all numbers of degrees of freedom tested here.

References

[1] H. Jónsson, G. Mills, and K. W. Jacobsen. Nudged elastic band method for finding minimum energy paths of transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, 1998), pp. 385–404.

[2] A. A. Peterson. Acceleration of saddle-point searches with machine learning. *J. Chem. Phys.* **145**, 074106 (2016).

[3] A. O'Hagan and J. F. C. Kingman. Curve fitting and optimal design for prediction. *J. R. Stat. Soc. B* **40**, 1 (1978).

[4] D. J. C. MacKay. Introduction to Gaussian processes. In *Neural Networks and Machine Learning*, edited by C. M. Bishop (Springer Verlag, 1998), pp. 133–166.

[5] R. M. Neal. Regression and classification using Gaussian process priors (with discussion). In *Bayesian Statistics*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Oxford University Press, 1999), vol. 6, pp. 475–501.

[6] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning* (MIT Press, 2006).

[7] O.-P. Koistinen, E. Maras, A. Vehtari, and H. Jónsson, Minimum energy path calculations with Gaussian process regression. *Nanosyst.: Phys. Chem. Math.* **7**, 925 (2016). A slightly corrected version available as e-print arXiv:1703.10423.

[8] A. O'Hagan. Some Bayesian numerical analysis. In *Bayesian Statistics*, edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Oxford University Press, 1992), vol. 4, pp. 345–363.

[9] C. E. Rasmussen. Gaussian processes to speed up hybrid Monte Carlo for expensive Bayesian integrals. In *Bayesian Statistics*, edited by J. M. Bernardo, A. P. Dawid, J. O. Berger, M. West, D. Heckerman, M. J. Bayarri, and A. F. M. Smith (Oxford University Press, 2003), vol. 7, pp. 651–659.

[10] E. Solak, R. Murray-Smith, W. E. Leithead, D. J. Leith, and C. E. Rasmussen. Derivative observations in Gaussian process models of dynamic systems. In *Advances in Neural Information Processing Systems*, edited by S. Becker, S. Thrun, and K. Obermayer (MIT Press, 2003), vol. 15, pp. 1057–1064.

[11] J. Riihimäki and A. Vehtari. Gaussian processes with monotonicity information. In *Proceedings of Machine Learning Research*, edited by Y. W. Teh and M. Titterington (PMLR, 2010), vol. 9, pp. 645–652.

[12] A. P. Bartók and G. Csányi. Gaussian approximation potentials: a brief tutorial introduction. *Int. J. Quantum Chem.* **115**, 1051 (2015).

[13] J. Lampinen and A. Vehtari. Bayesian approach for neural networks – review and case studies. *Neural Netw.* **14**, 257 (2001).

[14] G. Henkelman, B. P. Uberuaga, and H. Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **113**, 9901 (2000).

[15] K. Müller and L. D. Brown. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor. Chim. Acta* **53**, 75 (1979).

[16] G. Henkelman, G. H. Jóhannesson, and H. Jónsson. Methods for finding saddle points and minimum energy paths. In *Progress in Theoretical Chemistry and Physics*, edited by S. D. Schwartz (Kluwer Academic, 2000), vol. 5, pp. 269–300.

[17] S. T. Chill, J. Stevenson, V. Ruhle, C. Shang, P. Xiao, J. D. Farrell, D. J. Wales, and G. Henkelman. Benchmarks for characterization of minima, transition states and pathways in atomic, molecular, and condensed matter systems. *J. Chem. Theory Comput.* **10**, 5476 (2014).