The University of Texas at Austin

Freshman Research Initiative: Research Project

Gaussian Distributions and Global Optimization: An exploration into improved performance

Madhav Narayan

supervised by

Dr. Juliana Duncan and Mica Kohl

May 5, 2016

Abstract

There is great demand and use for efficient global optimization methods. While existing methods offer fast performance, they prove to be inefficient when applied to large systems. In this project, we sought to explore and identify optimal parameters for performing global optimization on molecular structures with Gaussian distribution. To see how different parameters affected performance, we used a basin hopping global optimization routine. Most parameters were kept constant, while the temperature and step size were varied to see if any particular values (or range of values) enhanced performance. Using these parameters, global optimization was performed on a system of 100 38-atom Lennard-Jones clusters. The optimizer's performance was quantified by observing the average number of force calls made. While the temperature had no apparent effect on the performance of the optimizer, we found that using a step size of 0.2 Å significantly increased performance. Moving forward, this can be used to make current optimization methods more efficient and can possibly be combined with other known optimal configurations to achieve even better performance.

The world's energy needs are constantly rising. In 1990, the total worldwide primary energy[1] consumption was 346.98 Quadrillion[2] British thermal units (Btu)[3]. In 22 years, global consumption rose by 51% to 524.08 Quadrillion Btu a year. The Energy Information Administration predicts[4] that world energy consumption will increase by 56% to 820 Quadrillion Btu in 2040. Moreover, it is estimated that 80% of this energy will be supplied by fossil fuels, with carbon-dioxide emissions projected to increase by 46% to 45 billion metric tons.

Natural (conventional) sources of energy can (and should) not be relied on to meet these increasing energy needs. The need of the hour is an assortment of efficient, easily available, and renewable sources of energy. Considering that automobiles are one of the major consumers of oil and gas, several research efforts have been directed towards the development of hybrid electric vehicles. IBM Research is working on a battery that can electrically power an automobile for 500 miles on a single charge[5]. Their efforts involve work towards achieving a battery capacity of 125 kWh. Similarly, Tesla Motors, which manufactures commerical electric vehicles, has been developing and reporting consistent improvements in the performance of their batteries[6].

In order to identify the most promising candidates and configurations for use in these batteries, it is

[1] "Energy in the form that it is first accounted for in a statistical energy balance, before any transformation to secondary or tertiary forms of energy." Glossary - U.S. Energy Information Administration (EIA). 5 May 2016. (http://www.eia.gov/tools/glossary/index.cfm?id=P)

[2] 1 Quadrillion = 10^{15} = 1,000,000,000,000,000

[3] International Energy Statistics - EIA. 5 May 2016. (http://www.eia.gov/cfapps/ipdbproject/)

[4] *International Energy Outlook 2014*, U.S. Energy Information Administration. (http://www.eia.gov/forecasts/ieo/)

[5] Girishkumar, G., B. McCloskey, A. C. Luntz, Sally Swanson, and W. Wilcke. "LithiumAir Battery: Promise and Challenges." LithiumAir Battery: Promise and Challenges. The Journal of Physical Chemistry Letters, 2 July 2010. Web. 5 May 2016.

[6] "Model S — Tesla Motors." Model S. Tesla, n.d. Web. 05 May 2016.

essential to have high-quality optimization methods that can model large systems and report on chemical systems and clusters that deliver the best performance.

While existing global optimization algorithms are fast, they are not always optimal: as the size of a system increases, the time taken to model and perform optimizations on this system increases significantly. A lot of important work has been done towards creating and employing efficient global optimization methods. Wales and Doye's seminal paper[7] describes the basin hopping technique, which uses hypersurface deformation to efficiently and accurately obtain the global minimum of an energy structure. Bernd Hartke reported several improvements in performance when performing optimization with phenotype variants of genetic algorithms[8] . To aid further improvements, The Wales Group at the University of Cambridge maintains The Cambridge Energy Landscape Database, a public collection of known lowest minima for a variety of systems.

In this project, we attempt to identify the optimal parameters for use in global optimization on a system of clusters using the basin hopping algorithm.

Methods

Global Optimization

Optimization is the process of making a system or function as effective and efficient as possible. This can involve producing better (more accurate) results, making a system faster, or having a system use less resources (such as space or memory). *Global* optimization means finding the maximum or minimum of a function across all input values: this is different from *local* optimization, which deals with optimizing a function over a particular range or interval. Usually, in the context of (computational) chemistry and in the context of this project, global optimization refers to obtaining the lowest energy configuration of a molecular system.

[7]Wales, David, and Jonathan Doye. "Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms." Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. The Journal of Physical Chemistry Letters, 10 July 1997. Web. 5 May 2016.

[8]Hartke, Bernd. "Global Cluster Geometry Optimization by a Phenotype Algorithm with Niches: Location of Elusive Minima, and Low-order Scaling with Cluster Size." Global Cluster Geometry Optimization by a Phenotype Algorithm with Niches: Location of Elusive Minima, and Low-order Scaling with Cluster Size. Journal of Computational Chemistry, 5 Nov. 1999. Web. 05 May 2016.

Basin Hopping

Basin Hopping[9] is an iterative global optimization algorithm used to obtain the global minima of a potential energy structure[10]. The algorithm is stochastic, and works by performing random transformations on a structure (set of co-ordinates), performing local minimization on these transformed co-ordinates, and accepting or rejecting its local optimizations based on the parameters passed to the optimizer.

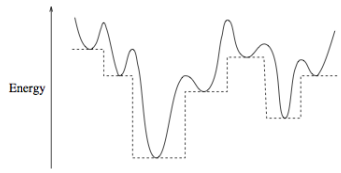

Figure 1: Schematic Diagram illustrating Basin Hopping[11]

Gaussian Distribution

Gaussian (Normal) Distribution is a type of distribution wherein most measurements tend to gather (collect) around the mean of the distribution[12]. These distributions yield good approximations even if errors in measurement are made. In the optimizer used for this experiment, the standard deviation for the Gaussian distribution was the step size (maximum displacement), and was also one of the variable parameters whose effect on performance we explored.

Setup

To run our optimizations, we used the Henkelman Research Group's[13] *Transition State Library for Atomic Simulation Environment*[14] (TSASE) software package. This library comes with a basin hopping global optimization routine that implements a steepest-descent limited memory version of the Broyden-Fletcher-Goldfarb-Shanno algorithm. When running our calculations, we kept certain parameters constant and altered the values of others. While default values[15] were used for most of the constant parameters, we used different (non-default) values for four fixed parameters. Table 1 details these changes.

Table 1: Fixed (Constant) Non-Default Parameters

[9] David J. Wales and Jonathan P. K. Doye, *Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms* The Journal of Physical Chemistry A 1997 101 (28), 5111-5116. DOI: 10.1021/jp970984n

[10] "Scipy.optimize.basinhopping." Basin Hopping Optimizer - SciPy. Web. 04 May 2016. (http://docs.scipy.org/doc/scipy-0.17.0/reference/generated/scipy.optimize.basinhopping.html)

[11] See footnote 9

[12] "Normal Distribution." Gale Encyclopedia of Psychology. 2001. Encyclopedia.com. 5 May. 2016 (http://www.encyclopedia.com).

[13] Henkleman Research Group, The University of Texas at Austin. (http://theory.cm.utexas.edu/henkelman/)

[14] TSASE: Transition State Library for ASE, http://theory.cm.utexas.edu/tsase/, 4 May 2016.

[15] Basin Hopping: Local Optimizers TSASE: Transition-State Library for ASE. Henkelman Research Group, The University of Texas at Austin. Web. 04 May 2016. (http://theory.cm.utexas.edu/tsase/optimizer.html#basin-hopping).

Parameter	Value
`fmax`	0.01
`mss`	0.05
`minenergy`	−173.9184 eV
`jumpmax`	5000 jumps

- `fmax`: The magnitude of the L2 (Euclidean) norm used as the convergence criteria for local optimization[16].
- `mss`: The largest step the basin hopping algorithm is permitted to take when performing local minimization on a set of co-ordinates.
- `minenergy`: The basin hopping algorithm stops execution when a lower potential energy than this minimum is encountered.
- `jumpmax`: The maximum permitted number of successive rejected basin hopping jumps. After 5000 consecutive rejected jumps, the algorithm accepts the next move, allowing for a more "global search of the potential energy surface"[17].

Using these fixed parameters, we altered the values of the two variables - temperature and step size - whose effect on performance we were investigating. As described below in Table 2, we selected five temperatures with increments of 2000 K and for each of these five temperatures, selected ten step sizes with increments of 0.1 Å. This made for fifty temperature-stepsize pairs, each of which was run as a seperate calculation on a system of 100 38-atom Lennard Jones clusters[18].

Table 2: Variable Parameters

Parameter	Function	Low	High	Increment	Total Values
`temperature`	Temperature	2000 K	10000 K	2000 K	5
`dr`	Step Size	0.1 Å	1.0 Å	0.1 Å	10

Results and Discussion

Results

Table 3 lists the average number of force calls made by the optimizer for each of the fifty temperature-step size combinations selected.

Table 3: Average Number of Force Calls

[16] See footnote 15
[17] See footnote 15
[18] The Cambridge Energy Landscape Database - Wales Group. 5 May 2016. (http://www-wales.ch.cam.ac.uk/CCD.html)

	2000 K	**4000 K**	**6000 K**	**8000 K**	**10,000K**
0.1 Å	225,837	183,430	222,604	206,247	216,865
0.2 Å	**88,195**	**78,114**	**99,681**	**91,832**	**81,110**
0.3 Å	147,207	193,826	189,681	155,084	161,022
0.4 Å	301,402	308,287	300,620	322,871	342,002
0.5 Å	444,348	458,016	496,749	477,744	445,932
0.6 Å	602,194	578,636	634,904	557,875	594,304
0.7 Å	675,132	681,77	719,074	701,488	672,093
0.8 Å	786,172	762,031	794,367	734,771	740,326
0.9 Å	867,163	849,732	845,143	870,483	801,095
1.0 Å	941,630	921,879	912,537	918,133	916,575

The data above indicates that the basin hopping optimizer made significantly fewer force calls when used with a step size of 0.2 Å. As can be seen, the average number of force calls is directly proportional to the step size: at and after a step size of 0.3 Å, the force calls only increase. Interestingly, this steady increase follows a sharp decrease at 0.2 Å. At this step size, on an average, 57% fewer force calls are made than at the preceding step size.

We also plotted this data (Figures 2 - 6) to visually observe the average number of force calls and see how the curve changed for different step sizes. Each of the five figures depicts the performance of the optimizer at the ten different step sizes for one temperature. As Figures 2 - 6 show, there is a minima at a step size of 0.2 Å for all five temperatures selected. Regardless of temperature, the optimizer is much more efficient when using a step size of 0.2 Å than it is for any of the other nine step sizes selected.

In addition to observing how different step sizes fared for each temperature, we also switched axes and observed how different temperatures performed for each step size. We were unable to identify any trends or similarities among the results and plots for the ten step sizes.

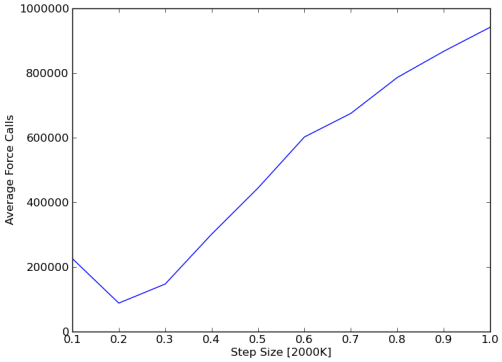

Figure 2: 10 Step Sizes (Å) at 2000 K

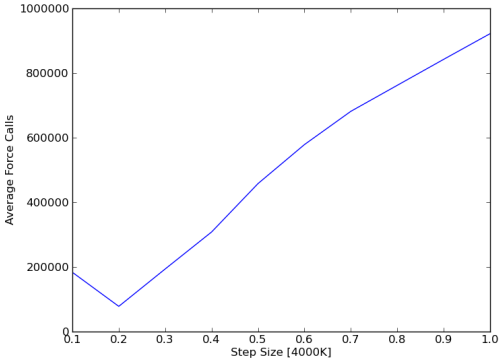

Figure 3: 10 Step Sizes (Å) at 4000 K

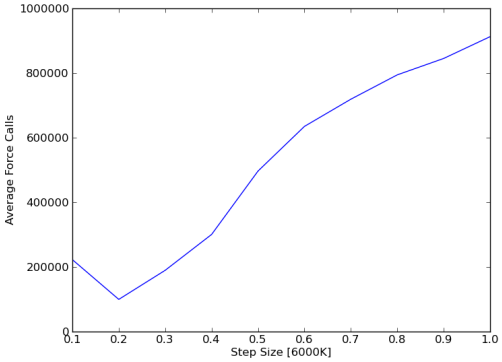

Figure 4: 10 Step Sizes (Å) at 6000 K

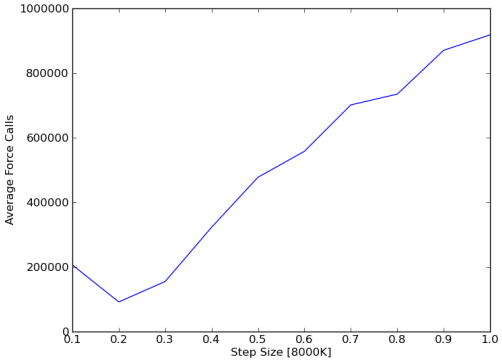

Figure 5: 10 Step Sizes (Å) at 8000 K

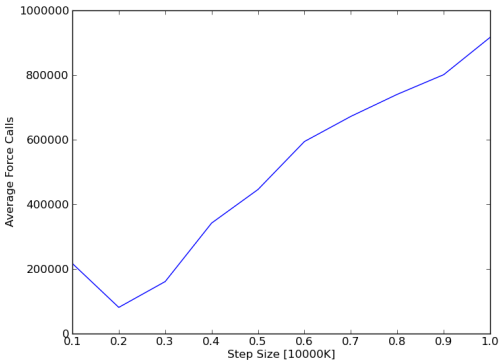

Figure 6: 10 Step Sizes (Å) at 10000 K

Discussion

It is clear that one value of the parameter we explored - 0.2 Å - yields significantly better performance than other values tested. This can be used to improve the performance of current optimization methods. Experiments and calculations can be designed keeping this optimized parameter in mind, and work towards optimizing other parameters with respect to this result can be conducted.

Going forward, we would like to continue exploring and idenfitying optimal configurations for global

optimization. A direction of particular interest is to use the same optimizer and scripts on larger, more complex systems[19] and with a wider range of fixed and variable parameters.

Conclusions

Ultimately, we sought to obtain configurations that could make existing optimization methods more efficient. To quantify whether one configuration was more "optimal" than another, we compared the two on the basis of the average number of force calls made. If one system made a fewer number of force calls than another, then we could conclude that the system more efficient than the other.

In conclusion, we reiterate our two main results:

1. **0.2 Å is the optimal step size** to use when performing global optimization (using the basin hopping algorithm) on systems with Gaussian distributions. The average number of force calls at 0.2 Å was orders of magnitude less than that at other step sizes.
2. **Temperature does not affect the performance of the optimizer.**

Acknowledgements

I would like to thank my instructor, Dr. Juliana Duncan, and my mentor, Ms. Mica Kohl, for their guidance and assistance with this project. I would also like to thank the other students in my group for their work, help, and enthusiasm.

References

[1] International Energy Statistics - EIA. 5 May 2016. (http://www.eia.gov/cfapps/ipdbproject/)

[2] International Energy Outlook 2014, U.S. Energy Information Administration. (http://www.eia.gov/forecasts/ieo/)

[3] Girishkumar, G., B. McCloskey, A. C. Luntz, Sally Swanson, and W. Wilcke. "LithiumAir Battery: Promise and Challenges." LithiumAir Battery: Promise and Challenges. The Journal of Physical Chemistry Letters, 2 July 2010. Web. 5 May 2016.

[4] "Model S — Tesla Motors." Model S. Tesla, n.d. Web. 05 May 2016.

[19] Systems with a larger number of clusters and clusters with a larger number of atoms.

[5] Wales, David, and Jonathan Doye. “Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms.” Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. The Journal of Physical Chemistry Letters, 10 July 1997. Web. 5 May 2016.

[6] Hartke, Bernd. “Global Cluster Geometry Optimization by a Phenotype Algorithm with Niches: Location of Elusive Minima, and Low-order Scaling with Cluster Size.” Global Cluster Geometry Optimization by a Phenotype Algorithm with Niches: Location of Elusive Minima, and Low-order Scaling with Cluster Size. Journal of Computational Chemistry, 5 Nov. 1999. Web. 05 May 2016.

[7] “Scipy.optimize.basinhopping.” Basin Hopping Optimizer - SciPy. Web. 04 May 2016. (http://docs.scipy.org/doc/scipy-0.17.0/reference/generated/scipy.optimize.basinhopping.html)

[8] Normal Distribution.” Gale Encyclopedia of Psychology. 2001. Encyclopedia.com. 5 May. 2016 (http://www.encyclopedia.com).

[9] Henkleman Research Group, The University of Texas at Austin. (http://theory.cm.utexas.edu/henkelman/)

[10] TSASE: Transition State Library for ASE, http://theory.cm.utexas.edu/tsase/, 4 May 2016.

[11] Basin Hopping: Local Optimizers TSASE: Transition-State Library for ASE. Henkelman Research Group, The University of Texas at Austin. Web. 04 May 2016. (http://theory.cm.utexas.edu/tsase/optimizer.html#basin-hopping).

[12] TSASE: Transition State Library for ASE, http://theory.cm.utexas.edu/tsase/, 4 May 2016.

[13] The Cambridge Energy Landscape Database - Wales Group. 5 May 2016. (http://www-wales.ch.cam.ac.uk/CCD.html)