BASC: Applying Bayesian Optimization to the Search for Global Minima on Potential Energy Surfaces

Shane Carr SHANE.CARR@WUSTL.EDU
Roman Garnett GARNETT@WUSTL.EDU
Cynthia Lo CLO@WUSTL.EDU
Washington University, 1 Brookings Dr., St. Louis, MO 63130 USA

Abstract

We present a novel application of Bayesian optimization to the field of surface science: rapidly and accurately searching for the global minimum on potential energy surfaces. Controlling molecule–surface interactions is key for applications ranging from environmental catalysis to gas sensing. We present pragmatic techniques, including exploration/exploitation scheduling and a custom covariance kernel that encodes the properties of our objective function. Our method, the Bayesian Active Site Calculator (BASC), outperforms differential evolution and constrained minima hopping — two state-of-the-art approaches — in trial examples of carbon monoxide adsorption on a hematite substrate, both with and without a defect.

1. Introduction

The study of chemical processes on solid surfaces, which was recently recognized with the 2007 Nobel Prize in Chemistry, is important for cleaner energy and environmental applications ranging from catalysis to gas sensing.

In particular, heterogeneous catalysis is the study of how solid materials can influence the rates of chemical reactions. These materials provide binding sites for the reacting gas molecules and orient them to facilitate chemical bond breaking and forming. A few examples include the work of Sivula et al. (2011), who used hematite (α-Fe_2O_3) for water splitting, Song (2006), who compared several CO_2-selective catalysts for treating flue gases of power plants via carbon capture, and Xu et al. (2012), who reviewed the use of iron oxide for wastewater treatment.

Proceedings of the 33^{rd} International Conference on Machine Learning, New York, NY, USA, 2016. JMLR: W&CP volume 48.

The process of a molecule binding to a surface (called *adsorption*) involves searching for energetically favorable binding sites. Although any site on the surface is theoretically possible, the most likely sites are the most thermodynamically stable (i.e., those that have the lowest potential energy). This gives rise to an important question: given a solid surface and a gas molecule, which adsorption configuration corresponds to the globally minimized potential energy?

From a mathematical point of view, this boils down to an optimization problem — efficiently performing global energy minimization in a multidimensional space, while simultaneously ensuring that the identities of the molecule and surface remain unchanged.

Researchers have attempted several different approaches to solving this problem. One of the more popular solutions to date, *constrained minima hopping* (Peterson, 2014), enforces molecular identity by placing constraints on the distance between atoms, and uses a stochastic global optimization algorithm based on basin hopping (Wales & Doye, 1997) to explore different molecule–surface configurations. However, constrained minima hopping is computationally expensive, often requiring several hundred density functional theory (DFT) calculations in an average run.[1]

In this paper, we present the Bayesian Active Site Calculator (BASC), a novel method for predicting adsorption configurations. We enforce the identity of the molecule and surface by formulating the problem as the minimization of a low-dimensional objective function (Section 2). We model this objective using a Gaussian process with a custom covariance kernel function that enforces the periodicity of and relationship between the dimensions of the objective function (Section 3). We then use Bayesian optimization (Section 4) to converge to the global solution with relatively few function evaluations (Sections 5 and 6).

[1] DFT is a computational method for calculating the total electronic energy given a particular atomic configuration by self-consistently solving a nonlinear system of equations related to the Schrödinger equation; see Section 4.3.

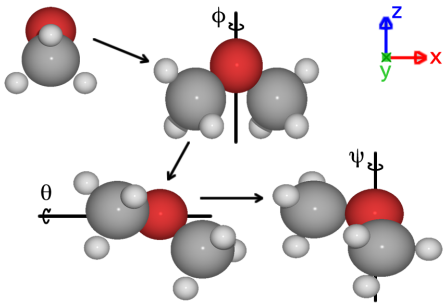

Figure 1. The three Euler angles fully specify the orientation of a rigid body in space. This image shows the process of rotating a dimethyl ether molecule by $\pi/3$ radians about each axis.

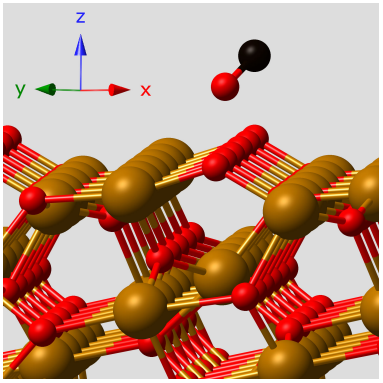

Figure 2. An example configuration of CO on hematite. The red atoms are oxygen; brown is iron; and black is carbon. The C atom in CO is positioned at $\langle x, y \rangle = \langle 0.5, 0.5 \rangle$, and it is $z = 2.2\,\text{Å}$ above the surface. The molecule is rotated with Euler coordinates $\langle \theta, \psi \rangle = \langle \frac{\pi}{3}, \frac{\pi}{3} \rangle$. The symmetry of the CO molecule nullifies the Euler coordinate ϕ. The surface is periodic across the boundaries of the unit cell, shown in black.

Our study here will be focused on CO (carbon monoxide) — one of the two components of syngas, adsorbing onto hematite — a widely used metal-oxide catalyst.

2. Parameterizing the Objective Function

Traditional methods for energy minimization involve optimization over multiple dimensions — namely, the three spatial coordinates of each of its N atoms. As an adsorbate molecule grows in size, so do the number of dimensions in the optimization problem, which scales as $3N$. So, for even a relatively small molecule such as dimethyl ether (CH_3OCH_3), its 9 atoms correspond to 27 degrees of freedom, which must then be systematically explored to find the most stable structural configuration.

Because we want to maintain the identity of the molecule through the optimization procedure, however, our problem is intrinsically lower-dimensional. We consider six dimensions that are capable of approximating most configurations of small molecules.

2.1. Six Basic Parameters

First, we need to specify the location of the center of the molecule along the surface plane. We can do this using two parameters, x and y. These two parameters are periodic across the boundaries of the unit cell.[2] We always take x and y to be fractional coordinates in the given direction, such that $x, y \in [0, 1)$.

Second, we need to specify how "close" the molecule is to the surface. We can do this with another parameter, which we will call z. We take z to be given in Å.[3]

[2]The term *unit cell* refers to a section of the surface, usually shaped like a parallelogram, that infinitely repeats itself to form the entire surface.

[3]An Angstrom (Å) is a unit of measurement equal to 10^{-10} meters, very near the length of an atomic bond.

Finally, we need to specify the orientation of the molecule. The orientation of a rigid body in space is fully specified by three parameters, ϕ, θ, and ψ, commonly known as the *Euler angles*. We always take ϕ, θ, and ψ to be measured in radians. See Figure 1 for an illustration of the Euler angles.

Molecules that are symmetric about an axis, like CO and CO_2, need only two Euler coordinates rather than three. Single-atom adsorbates do not need any Euler coordinates.

We can therefore approximate the configuration of a molecule on a catalyst surface with just six parameters, and sometimes fewer. An example configuration is shown in Figure 2.

2.2. Additional Parameters

The six parameters of Section 2.1 neglect structural changes of the molecule and of the surface. We consider two types of structural changes: *free structural parameters* and *deformation by adsorption*.

2.2.1. Free Structural Parameters

Consider a molecule like ethane, C_2H_6. In addition to the three Euler angles, ethane needs one more parameter — the internal angle of rotation about the C–C bond — to fully define its orientation. This parameter is periodic, and it could be added to the objective function.

2.2.2. Deformation by Adsorption

It is usually the case that the interaction of the molecule with the surface causes one or the other to deform in a way that would otherwise not occur in free space. For example, the

length of an internal bond may increase or decrease, and the binding of a molecule to the surface may cause the atoms on the surface to move.

These changes are difficult to parameterize. Depending on the molecule, some number of parameters could potentially be added to correspond to common ways that the molecule could deform. The number of parameters would increase dramatically, though, for complex surfaces and molecules.

Fortunately, our results for a CO molecule on a hematite surface indicate that deformation by adsorption can be neglected. The user can run an additional relaxation step to converge BASC's result to the true global minimum.[4]

3. Designing an Appropriate Kernel

Bayesian optimization (see Section 4.1) operates by making observations of the objective function and adding those observations to a Gaussian process model. A *Gaussian process*, or GP (Rasmussen & Williams, 2006), is a Bayesian method popular as a technique for performing nonlinear regression.

A *kernel function*, or *covariance function*, is a measure of the "similarity" between two points based on their locations in the parameter space of an objective function (Duvenaud, 2014, chap. 2). A GP uses a kernel function coupled with a prior belief $\mu(x)$ (the *mean function*) to make predictions of the objective function by weighing the influence of known observations over an unknown point in space.

The classical example of a kernel function is the *squared exponential*, or SE:

$$k_{SE}(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2\ell^2}(x_2 - x_1)^2\right) \qquad (1)$$

where ℓ is the *length scale* and σ^2 is the *output variance*. The length scale defines the "region of influence" of a point within the parameter space; the influence of an observation decreases as one considers points farther and farther away from the observation. The variance defines the "expected deviation" of the function away from its average value.

We use the SE kernel for modeling the z dimension. For the other two axes, we use variations on it, which are explained below. To combine everything into one all-encompassing multi-dimensional kernel, we multiply the kernels together using the method presented in Duvenaud (2014, chap. 2).

[4] A "relaxation step" means converging the full-dimensional system (three coordinates for each atom) to a local potential energy minimum, typically using a classical optimization algorithm like L-BFGS (Liu & Nocedal, 1989) and typically until the forces on every atom are less than 0.05 eV/Å.

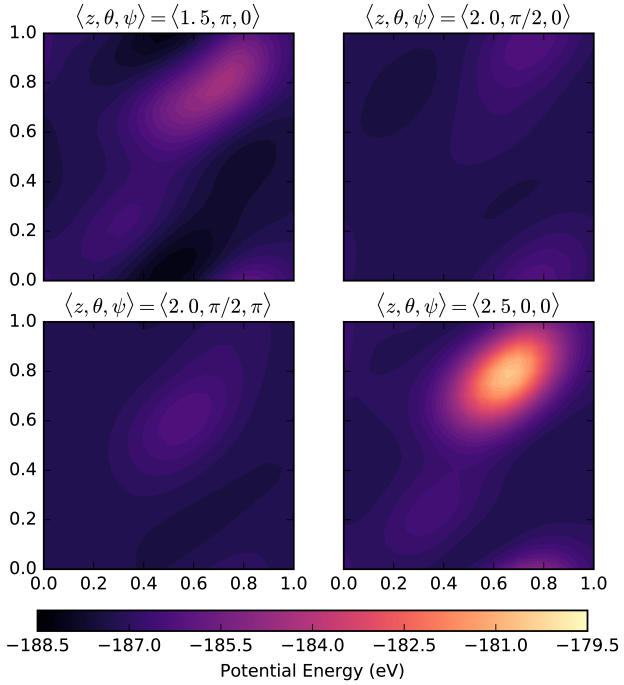

Figure 3. Four slices of the x–y plane from the objective function for CO on α-Fe_2O_3 (hematite). The global minimum is near the center of the dark area in the first slice.

3.1. Standard Periodic Kernel (x and y)

Figure 3 shows several slices along the x–y plane of the objective function. The figures indicate that the objective function is rather smooth and periodic in these dimensions.

To model x and y, we employ a periodic kernel proposed in MacKay (1998) and documented in Rasmussen & Williams (2006, eqn. 4.31), which has become known as the *standard periodic* (SP) kernel. We use the following one-dimensional form, implemented by The GPy authors (2012–2015):

$$k_P(x_1, x_2) = \sigma^2 \exp\left(-\frac{1}{2\ell^2}\sin^2\left((x_2 - x_1)\frac{\pi}{p}\right)\right) \qquad (2)$$

where p is the *period*.

3.2. Spherical Kernel (Euler angles)

The three Euler angles (see Figure 1) have trigonometric relationships that benefit from a specialized kernel.

To illustrate the relationships, consider a linear molecule with its axis of symmetry aligned to the z axis. First we perform the θ rotation about the x axis, followed by the ψ rotation about the z axis. This is analogous to walking along the surface of a sphere: you start at the north pole, you walk south along a line of longitude, and then you walk west along a line of latitude.

Figure 4 shows how a slice of the objective function along

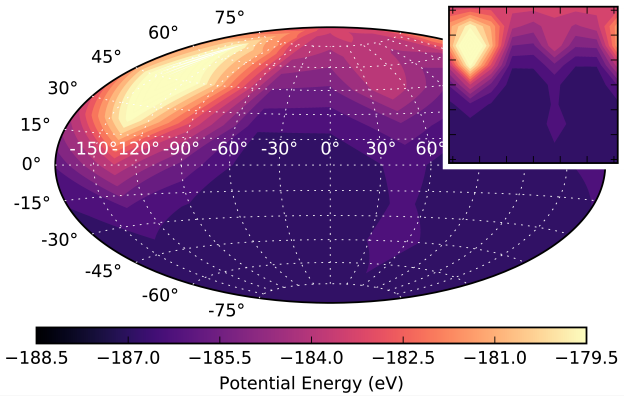

Figure 4. Two projections of a slice of the θ–ψ plane from the objective function for CO on hematite at $x = y = 0$ and $z = 1.5$ Å. The color scale is the same as in Figure 3.

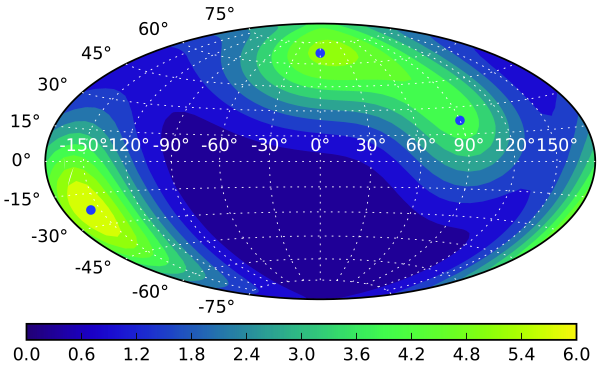

Figure 5. A GP using the spherical kernel function (Equation 4) with $\ell = \pi/6$ and $\sigma^2 = 5$ fit to three observations, shown in blue, with values from top to bottom of 5, 4, and 6. The GP has a constant mean of zero and noise variance of 0.1^2.

the θ–ψ plane for CO on hematite conforms well to a spherical projection. How can we represent this spherical relationship in a kernel function?

A handful of methods for representing kernel functions in spherical space have been proposed (Berman, 1980; Brauchart et al., 2014). For example, Solin & Särkkä (2014) present an approach making use of eigenfunction expansions of the Laplace operator and show how it can be used to model temperature patterns on the surface of the earth. Paciorek's approach (2003) projects the sphere into Euclidean space and uses a convolution in the spherical domain to combine the kernels for each point.

We propose an approach based on the observation that the SE and SP kernels are of the same form: an exponential function of a distance metric (Euclidean distance in the SE kernel, and periodic distance in the SP kernel). Our spherical kernel takes the same form, but uses the *great circle distance* as its metric:[5]

$$d_{GC}(\hat{\theta}_1, \hat{\psi}_1, \hat{\theta}_2, \hat{\psi}_2) = \arcsin(\sin\hat{\theta}_1 \sin\hat{\theta}_2 + \cos\hat{\theta}_1 \cos\hat{\theta}_2 \cos(\hat{\psi}_2 - \hat{\psi}_1)) \quad (3)$$

where $\hat{\theta}_i$ is the inclination (or "latitude") and ψ_i is the azimuth (or "longitude"). The conversion from our Euler angles is:

$$\hat{\theta}_i = \frac{\pi}{2} - \theta_i \qquad \hat{\psi}_i = \psi_i$$

The kernel function can then be written as:

$$k_{SPH}(\theta_1, \theta_2, \psi_1, \psi_2) = \sigma^2 \exp\left(-\frac{1}{2\ell^2} d_{GC}^2\right) \quad (4)$$

To illustrate this kernel, we show in Figure 5 an example GP using the kernel to fit three observations. Observe that the kernel makes the function periodic around the sphere.

For sufficiently small ℓ, Equation 4 is positive-definite to numerical precision and can be safely used as a kernel function. Empirically, we drew the first 150 points of the Sobol sequence (Burkardt & Chisari, 2011) for a slice of our objective function and fitted a GP using the spherical kernel to that data. We then incremented ℓ in steps of 0.001 until we found a value of ℓ that produced a covariance matrix that was not positive-definite to machine precision. The largest value of ℓ that our GP could handle was 0.977, or 0.311π. In the remainder of our work, we have bounded the value of of ℓ at 0.3π.

4. Optimizing the Objective Function

In Section 2, we showed how we can represent the problem at hand as a low-dimensional objective function, and in Section 3, we documented a kernel to model it. In this section, we explain Bayesian optimization and how we have applied it in BASC.

4.1. Bayesian Optimization

Bayesian optimization (BO) is a method for finding the global minimum of an expensive, black-box oracle function over some bounded set of parameters. It was first documented by Mockus et al. (1978), but it remained largely confined to theoretical literature until being revived by Jones et al. (1998) and finding several practical applications, most popularly the optimization of machine learning algorithms (Snoek et al., 2012). Compared with other global optimization algorithms, BO has an advantage when the objective function is expensive (more expensive than the optimization routine) and lacks a well-defined mathematical represen-

[5] The form in Equation 3 is not numerically stable for small distances, so in practice, we use the haversine form.

tation (Brochu et al., 2010). In the last two years, BO has begun seeing practical applications outside of machine learning, in areas such as robotics (Heping et al., 2015), bioengineering (Luna & Martinez, 2014), and mechanical engineering (Sterling et al., 2015). However, to the best of our knowledge, this is the first documented application of BO to the field of surface science.

BO starts by considering a set of known observations of the oracle function. It fits a GP to those observations, and then it picks a new point to evaluate by maximizing the *expected improvement* (EI) over the parameter space. EI is a non-convex function that has higher values near promising observations, lower values near bad observations, and medium values in unexplored areas.[6] The point that maximizes EI will be evaluated by the oracle function and added to the GP. This process repeats itself until the user is satisfied with the result.

4.2. Exploration and Exploitation Scheduling

BO, and GPs in general, assume that the objective function is drawn from the distribution encoded by the kernel. With our real-life potential energy surface, this is unlikely to be the case. Instead, we can manipulate the hyperparameters to control how BO chooses points to evaluate.

The output variance σ^2 can be used to control the trade-off between so-called *exploration* (evaluating a point in an unknown area of the function space) and *exploitation* (evaluating a point near a previous, promising observation). We use a decaying function to schedule a region of exploration followed by a region of exploitation:

$$\sigma^2 = B \exp\left(1 - (nf)^2\right) \tag{5}$$

where n is the current iteration number, B is a fixed constant, and f is what we call the *influence fraction* of an observation: the fraction of the total parameter space that is influenced by the observation. We typically assume that "influence" travels for 2 length scales in each dimension. This function is chosen such that the first f^{-1} observations (in which $\sigma^2 > B$) are capable of exploring the entire parameter space, with all subsequent observations (in which $\sigma^2 < B$) transitioning into exploitation mode. Equation 5 is shown in Figure 6.

4.3. Computing the Potential Energy

Our parameterization takes a vector of numbers and outputs a set of atomic coordinates. To convert those atomic coordinates into a scalar potential energy, we turn to the field of computational chemistry. There are two primary approaches: one involving classical mechanics and empirical force fields, such as that of Lennard–Jones (1924), and the other involving quantum mechanics, such as that of Kohn and Sham (1965).

[6] Maximizing EI is actually just one example of what is more generally known as an *acquisition function*.

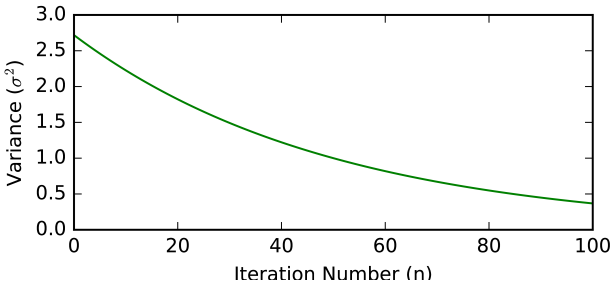

Figure 6. Equation 5 with B=1 and f=0.02, designed to schedule exploration in the first 50 iterations and exploitation in the second 50 iterations.

Kohn and Sham's method, density functional theory (KS–DFT), is a reformulation of the Schrödinger equation to express the electron density as a function of the electron coordinates; since the electron density is the squared modulus of the wavefunction, the approaches are, in principle, exact. An initial electron density is correlated one-to-one to the total electronic energy, which is a sum over kinetic and potential (including electron-electron and electron-nuclear) terms. The method then involves an iterative procedure to converge the electronic energy as a functional of the electronic coordinates.

With the electronic structure being calculated on the fly using quantum mechanics, DFT is naturally more rigorous and accurate than classical mechanics, so it has been successfully used to model structure and properties in extended systems, such as semiconductors (Jensen, 2007). We therefore use DFT for all of our calculations.

4.4. Details of Our Implementation

In BASC, we make use of the following domain-specific implementation details.

Initialization We initialize the GP by evaluating the objective function at a point from the Sobol sequence (Burkardt & Chisari, 2011). Each time we ran BASC, we used a different initial observation.

Noise We set the likelihood variance (the "noise") to be an arbitrary small, but nonzero, number. We found empirically that settings larger than around 10^{-4} caused the algorithm to get stuck evaluating points in the close vicinity of a promising observation due to the observation having a low confidence interval.

Period We fix the period for x and y at 1, since we have defined x and y to be in fractional coordinates.

Length Scales We can use our knowledge about the topology of the objective function to set reasonable lengths scales. In the x and y directions, we set the length scales to 1 Å, a conservative value on the order of the length of an atomic bond, expressed in fractional coordinates depending on the size of the unit cell. In the other dimensions, we use 0.25 Å in z and $\frac{\pi}{6}$ in the spherical kernel.

Mean Function To ensure that the mean function, $\mu(x)$, is within the true range of the objective function, after each iteration, we set $\mu(x)$ to be a constant equal to the mean of all previous observations.

Scheduling We set the constant B for the scheduling function (Equation 5) to be fixed at 10 eV, a value slightly larger than the overall variance of the objective function. Based on our choices for length scales, the influence fraction turns out to be $f = 0.0177$.

Maximizing Expected Improvement Our implementation uses differential evolution internally to maximize the expected improvement. This is one source of stochasticism in our implementation of BO.

DFT Configuration We use the LDA exchange-correlation functional (Perdew & Wang, 1992) and a pseudo partial wave electron density basis set using GPAW (Mortensen et al., 2005; Enkovaara et al., 2010).

Hematite Surface We obtained a crystal structure for α-Fe_2O_3 (hematite) from Materials Project. We used the $R\bar{3}C$ space group, corresponding to material ID "mp-24972" (Jain et al., 2013; Ong et al., 2015). We cut the 001 surface from the crystal structure, with two unit cell layers and 15 Å of vacuum. We then relaxed the surface using DFT and L-BFGS. The bottom unit cell layer was frozen, while the top layer was allowed to move. The relaxed hematite surface was used as a basis for all further calculations.

CO Molecule Our CO (carbon monoxide) molecule has the oxygen atom 1.128 Å below the carbon atom along the z axis. The center of the carbon atom is always the reference point for the x, y, and z parameters.

5. State of the Art

We will compare BASC to two other methods: differential evolution (an alternative global optimization routine) and constrained minima hopping (a domain solution).

Table 1. Hyperparameter settings for Differential Evolution.

CASE	POP.	TOL.	MUTATION	RECOMB.
DEFAULT	15	0.01	0.5 TO 1	0.7
AGGRESSIVE	5	0.01	0.5 TO 1	0.9

5.1. Differential Evolution

Differential evolution (DE), first documented by Storn and Price (1997), is a general routine for the global optimization of a bounded objective function.

DE, which is a type of genetic algorithm, starts with a fixed number of candidates (the *population*), which should be reasonably well-distributed over the parameter space. It calls the objective function for each of those candidates, and at each step, it performs a "mutation" process in which the next generation of candidates are more likely to inherit properties from the "best" candidates in the previous generation.

Like BO, DE has hyperparameters needing to be set. We ran two different cases of DE: one with default, conservative settings, and one with aggressive settings.[7] Aggressive parameters cause the algorithm to terminate more quickly, but they risk converging to a local minimum. These settings are listed in Table 1.

We gave DE the same objective function as BO. We used the "*best1bin*" strategy from the implementation in SciPy.

5.2. Constrained Minima Hopping

In contrast, *constrained minima hopping* (CMH) is a routine for optimizing adsorbate-surface structures, documented by Peterson (2014). It builds on Goedecker's earlier routine called *minima hopping* (2004).

CMH starts with an initial molecule–surface configuration, and it lets the molecule relax to a local potential energy minimum (via L-BFGS). Each local minimization requires around 25 function evaluations for CO on hematite. Through this process, the atoms of the molecule are allowed to move independently of one another, but a Hookean constraint is applied to ensure that the molecular identity is maintained. Once a local minimum is found, it is recorded. The system is then randomly permuted by a fixed amount of energy and allowed to relax again. If the new local minimum is the same as the previous local minimum, the next random permutation will add extra energy. This process is repeated several times until the user is satisfied with the result.

In our runs, CMH evaluated the atomic forces and potential energy using DFT, like the parameterized objective function.

[7] The "default" settings are those in the SciPy implementation of DE as of SciPy version 0.15.

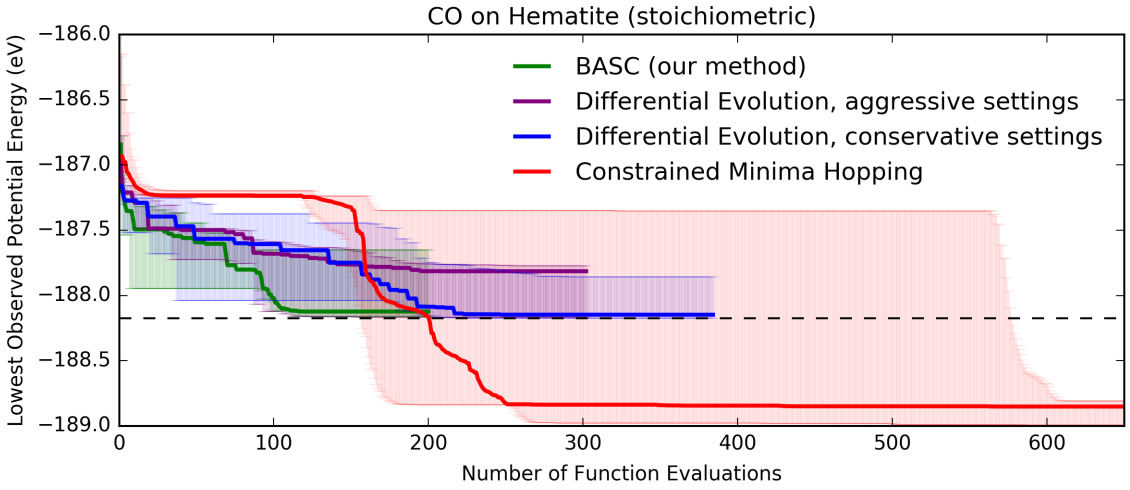

Figure 7. Ranges of best observation after each call to DFT for CO on hematite. The error bars show max and min, and the solid lines show the median. At least four runs of each routine are included. The dashed black line shows the minimum for the parameterized objective function; since CMH allows for deformation of the molecule and surface atoms, it is capable of reaching a lower energy.

6. Results

To evaluate its performance, we compared BASC (our Bayesian optimization-based method) to two state-of-the-art routines — differential evolution (DE) and contrained minima hopping (CMH) — to find the most energetically stable binding site of CO on a hematite surface.

We considered two different structures of hematite: one stoichiometric, and the other with an oxygen vacancy (defect) on the surface. Oxygen-deficient, or "reduced," catalysts have a strong propensity to redistribute electrons, and they are increasingly popular for applications including photocatalytic reactors and battery technology (Wang et al., 2012).

6.1. Stoichiometric Hematite

In the stoichiometric (defect-free) system, we ran four routines: BASC, DE with conservative settings, DE with aggressive settings, and CMH. All of the routines found the same solution — the CO molecule oriented C-down directly over an Fe atom on the surface. This configuration is shown in Figure 8. The corresponding coordinates in our objective function's parameter space are shown in Table 2.

Table 2. Solution of the objective function for CO on hematite.

X	Y	Z	θ	ψ
0.493	0.023	1.5 Å	3.077	0.009

Figure 7 plots the minimum energy observed from DFT as a function of the number of calls to DFT. BASC converges the most quickly, converging to 0.1 eV above the minimum

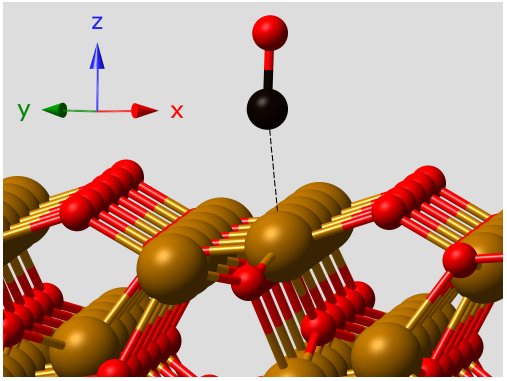

Figure 8. All of the routines identify the above configuration for CO on hematite.

of the parameterized objective function in 100 iterations in the median case and 94 iterations in the best case. DE with the default settings reaches that threshold in 192 iterations in the median case and 181 iterations in the best case. DE with the aggressive settings reaches it in 93 iterations in the best case, but it does not reach it in the median case.

For comparison, CMH is also shown on the plot. CMH does not use the same parameterized objective function as BASC and DE; in particular, it is allowed to perform the "deformation by adsorption" that we discussed in Section 2.2.2. This allows it to achieve a potential energy that is smaller than the minimum possible from the parameterized objective function. However, the solution it finds is virtually identical to the one from the objective function, with only minor shifts in atom positions; this validation confirms the accuracy of our approach. The most notable difference is

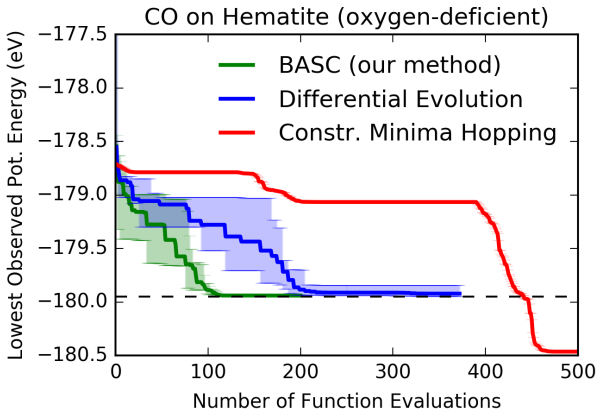

Figure 9. Range of minimum potential energies after each call to DFT for CO on oxygen-deficient hematite. Four runs each of BASC and DE are included; only one run of CMH is included.

that CMH's C–O bond length is around 1.141 Å (up from 1.128 Å), suggesting that bond length may be a beneficial free structural parameter to add to the objective function. Performing an unconstrained L-BFGS relaxation step on BASC's solution, with a cost of 26 more DFT function calls, yields a configuration closer to that of CMH, at -188.81 eV.

6.2. Oxygen-Deficient Hematite

To prepare the oxygen-deficient surface, we first removed the oxygen atom that was nearest the top of the surface, located near $x = y = 0.75$. We then re-relaxed the empty surface (using DFT and L-BFGS).

We ran BASC, differential evolution, and constrained minima hopping on this surface. We ran DE with only the conservative settings, since we found from Figure 7 that the aggressive settings lead to instability.

As before, all three routines agree on the solution. In this case, it is the CO molecule oriented C-down directly over an Fe atom, but with the CO tilted over the vacancy site, as shown in Figure 10. This result makes sense physically, because the oxygen atom in CO should be attracted to the oxygen vacancy site in the surface. The solution after relaxing the BASC result with L-BFGS has the CO molecule more severely tilted toward the vacancy site, corresponding to an energy of -180.74 eV, which exceeds CMH's best result of -180.46 eV.[8] The coordinates are shown in Table 3.

In addition, it is reassuring that BASC performs well in the oxygen-deficient system. Figure 9 plots the minimum observed energy versus the number of calls to DFT. On average, BASC finds a configuration within 0.1 eV of the minimum by iteration 96, while DE finds it by iteration 198.

[8]CMH's solution has the CO molecule positioned over one of the other Fe atoms, but still tilted toward the vacancy site.

Table 3. Solution of the objective function for CO on oxygen-deficient hematite.

X	Y	Z	θ	ψ
0.401	0.955	1.5	2.909	4.64

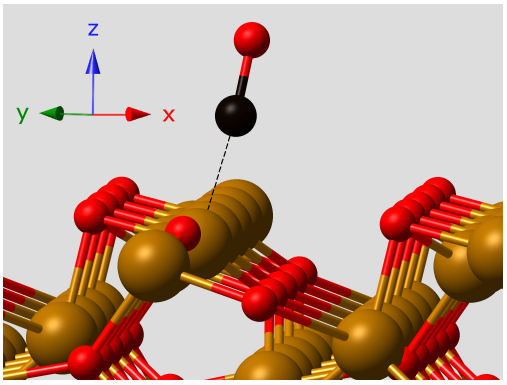

Figure 10. All routines identify the above configuration for CO on hematite with a defect. The atoms are in slightly different positions than Figure 8 because the atoms in the surface were allowed to relax after the oxygen was removed. The CO is positioned above the same Fe atom as in Figure 8, but its oxygen atom on top is "leaning" toward the vacancy site.

7. Conclusion

This paper documents a novel application of Bayesian optimization to a field of study outside the bounds of traditional machine learning.

We started by considering the problem in the field of surface science: determining the optimal molecule–surface configuration. We modeled the problem as an optimization problem with a well-defined objective function. We derived a custom kernel for the objective function. We presented a pragmatic scheduling function for running Bayesian optimization on a non-ideal objective function. We then compared our method with two others—differential evolution and constrained minima hopping—and found that our approach was accurate and had the best performance in terms of number of calls to DFT. Our Bayesian optimization approach to efficiently sample the search space has broad impact for surface science at environmental interfaces.

Our code is available at:
`https://gitlab.com/caml/basc`

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. IIA–1355406.

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575 (John, 2014).

This work used several Python libraries, including GPy for Gaussian processes (The GPy authors, 2012–2015), GPAW for DFT calculations (Mortensen et al., 2005; Enkovaara et al., 2010; Larsen et al., 2009), ASE for atomic manipulations and LJ calculations (Bahn & Jacobsen, 2002), and SciPy for numerical routines (Jones et al., 2001–2016). Much of our work used the IPython graphical user interface to Python (Prez & Granger, 2007). Most graphs were generated using the Python interface to Matplotlib (Hunter, 2007).

References

Bahn, S. R. and Jacobsen, K. W. An object-oriented scripting interface to a legacy electronic structure code. *Computing in Science & Engineering*, 4(3):56–66, 2002.

Berman, S. M. Isotropic Gaussian processes on the Hilbert sphere. *The Annals of Probability*, pp. 1093–1106, 1980.

Brauchart, J., Saff, E., Sloan, I., and Womersley, R. QMC designs: Optimal order Quasi Monte Carlo integration schemes on the sphere. *Mathematics of Computation*, 83 (290):2821–2851, 2014.

Brochu, E., Cora, V. M., and De Freitas, N. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*, 2010.

Burkardt, J. and Chisari, C. SOBOL: The Sobol Quasirandom Sequence, 2011. URL http://people.sc.fsu.edu/~jburkardt/py_src/sobol/sobol.html.

Duvenaud, D. *Automatic model construction with Gaussian processes*. PhD thesis, University of Cambridge, 2014.

Enkovaara, J., Rostgaard, C., Mortensen, J. J., Chen, J., Dułak, M., Ferrighi, L., Gavnholt, J., Glinsvad, C., Haikola, V., and Hansen, H. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *Journal of Physics: Condensed Matter*, 22(25):253202, 2010.

Goedecker, S. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *The Journal of Chemical Physics*, 120(21):9911–9917, 2004.

Heping, C., Binbin, L., Gravel, D., Zhang, G., and Biao, Z. Robot learning for complex manufacturing process. pp. 3207–11, 2015.

Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, 9(3):90–95, 2007. URL http://matplotlib.org/.

Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., and Ceder, G. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1 (1):011002, 2013.

Jensen, F. *Introduction to Computational Chemistry*. Wiley, 2007.

John, T. XSEDE: Accelerating Scientific Discovery. 16(5): 62–74, 2014. Computing in Science and Engineering.

Jones, D. R., Schonlau, M., and Welch, W. J. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

Jones, E., Oliphant, T., Peterson, P., et al. SciPy: Open source scientific tools for Python, 2001–2016. URL http://www.scipy.org/.

Jones, S. J. E. On the determination of molecular fields: From the equation of state of a gas. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 106(738):463–477, 1924.

Kohn, W. and Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140:A1133–A1138, 1965.

Larsen, A. H., Vanin, M., Mortensen, J. J., Thygesen, K. S., and Jacobsen, K. W. Localized atomic basis set in the projector augmented wave method. *Physical Review B*, 80(19):195112, 2009.

Liu, D. and Nocedal, J. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, 1989.

Luna, M. and Martinez, E. A Bayesian Approach to Run-to-Run Optimization of Animal Cell Bioreactors Using Probabilistic Tendency Models. *Industrial & Engineering Chemistry Research*, 53(44):17252–17266, 2014.

MacKay, D. J. Introduction to Gaussian processes. *NATO ASI Series F Computer and Systems Sciences*, 168:133–166, 1998.

Mockus, J., Tiesis, V., and Zilinskas, A. The application of Bayesian methods for seeking the extremum. *Towards Global Optimization*, 2(117–129):2, 1978.

Mortensen, J. J., Hansen, L. B., and Jacobsen, K. W. Real-space grid implementation of the projector augmented wave method. *Physical Review B*, 71(3):035109, 2005.

Ong, S. P., Cholia, S., Jain, A., Brafman, M., Gunter, D., Ceder, G., and Persson, K. A. The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. *Computational Materials Science*, 97:209–215, 2015.

Paciorek, C. J. *Nonstationary Gaussian processes for regression and spatial modelling*. PhD thesis, Carnegie Mellon University, 2003.

Perdew, J. P. and Wang, Y. Accurate and simple analytic representation of the electron–gas correlation energy. *Physical Review B*, 45(23):13244, 1992.

Peterson, A. A. Global optimization of adsorbate-surface structures while preserving molecular identity. *Topics in Catalysis*, 57(1-4):40–53, 2014.

Prez, F. and Granger, B. E. IPython: a system for interactive scientific computing. *Computing in Science & Engineering*, 9(3):21–29, 2007. URL http://ipython.org/.

Rasmussen, C. E. and Williams, C. K. I. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

Sivula, K., LeFormal, F., and Grtzel, M. Solar Water Splitting: Progress Using Hematite (α-Fe2O3) Photoelectrodes. *ChemSusChem*, 4(4):432–449, 2011.

Snoek, J., Larochelle, H., and Adams, R. P. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems 25*, pp. 2951–2959, 2012.

Solin, A. and Särkkä, S. Hilbert space methods for reduced-rank Gaussian process regression. *arXiv preprint arXiv:1401.5508*, 2014.

Song, C. Global challenges and strategies for control, conversion and utilization of CO2 for sustainable development involving energy, catalysis, adsorption and chemical processing. *Catalysis Today*, 115(14):2–32, 2006.

Sterling, D., Sterling, T., YuMing, Z., and Heping, C. Welding parameter optimization based on Gaussian process regression Bayesian optimization algorithm. In *Proceedings of the 2015 IEEE International Conference on Automation Science and Engineering (CASE)*, pp. 1490–1496, 2015.

Storn, R. and Price, K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

The GPy authors. GPy: A gaussian process framework in python. http://github.com/SheffieldML/GPy, 2012–2015.

Wales, D. J. and Doye, J. P. 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*, 101(28):5111–5116, 1997.

Wang, G., Ling, Y., and Li, Y. Oxygen-deficient metal oxide nanostructures for photoelectrochemical water oxidation and other applications. *Nanoscale*, 4:6682–6691, 2012.

Xu, P., Zeng, G. M., Huang, D. L., Feng, C. L., Hu, S., Zhao, M. H., Lai, C., Wei, Z., Huang, C., Xie, G. X., and Liu, Z. F. Use of iron oxide nanomaterials in wastewater treatment: A review. *Science of The Total Environment*, 424:1–10, 2012.