Report on Project 2b: Bayesian Optimization

Alfred Juhlin Onbeck and Zhi Li

March 25, 2024

Task 1

Finding the global minimum of the potential energy surface (PES). The PES $E(x, y, z)$ is defined as the potential energy of the system with the ad-atom E_{ad} after subtracting the potential energy of just the surface system E_{surface}[1].

$$E = E_{\text{ad}} - E_{\text{surface}}. \tag{1}$$

The z-coordinate of the ad-atom is relaxed to a local minima,

$$E(x, y) = \min_{z} E(x, y, z), \tag{2}$$

so that only x, y will be freely varied parameters. Here in fig. 1 the PES has been evaluated in a grid with a spacing of $dx = dy = 0.1$ Å to get an estimate of where the global minimum is (approximately $(x, y) = (3.4, 0.7)$ Å).

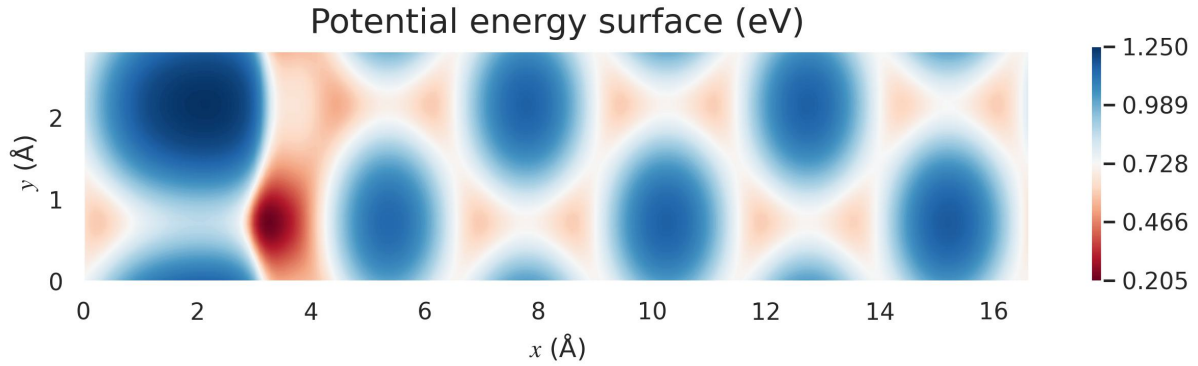

Figure 1: Visualisation of the potential energy surface $E(x, y)$ in the primitive cell.

Task 2

Estimating the global minimum through local searches using a gradient descent optimization method. Due to the bumpy surface as can be seen in fig. 1 with several local minima, a gradient descent method could struggle to find the global minimum. To show this a total of 400 local searches were run with a random starting position inside the primitive cell. In fig. 2 it shows that only ≈ 50 searches reached the global minimum ($(x, y) = (3.4, 0.7)$ Å) at $E \approx 0.20$ eV.

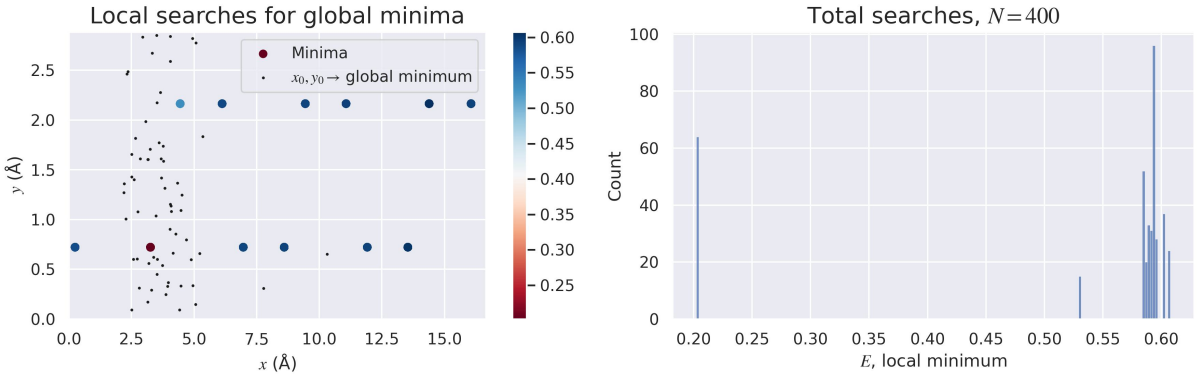

Figure 2: Local searches for the global minimum. Left figure, the initial positions of the searches that reach the global minimum are displayed as small black dots and the local minima are blue dots and the global minimum a red dot. Right figure, the frequency of the different minima reached are displayed as a histogram.

Task 3

Bayesian optimization through `GPy` to estimate the global minimum. The model will begin with 5 randomly selected samples in our primitive cell for optimization and then it will select a new sample based on the maximum of acquisition function $\rightarrow$ stack on top of previous samples $\rightarrow$ re-optimize the GPy model $\rightarrow$ refer to acquisition function for new sample $\rightarrow$ ..., until it has a sufficient amount of samples. Roughly 120 samples are sufficient for the model to get a good estimate of the global minimum of the PES. For our acquisition function we have used the lower confidence function[2],

$$A(x, y) = -\mu(x, y) + \beta\sigma(x, y), \tag{3}$$

where β will determine where new samples will be taken. A high β value will prioritise regions with high uncertainty (exploration) while low β will favour regions with low energy (exploitation). For our model we let $\beta = 3$ as some what of a middle ground between exploration and exploitation. The kernel of the `GPy` model is a combination of GPy.kern.RBF and GPy.kern.Bias, initially started with only the RBF kernel and later added the Bias kernel that appeared to improved the overall model. For the RBF kernel the length scale and variance has to be set and also the variance for the bias kernel. After various testing of different constraints and priors where we compared the model to the already evaluated PES, we concluded that a Gamma distributions did an adequate job for all three of the parameters. Specifically we let $\beta = 1$ for all three and for the RBF kernel $\alpha_{\text{length}} = 2$, $\alpha_{\text{var}} = 1.2$ and for the bias kernel $\alpha_{\text{var}} = 1.5$. In fig. 3 the GPy models prediction of the PES is shown along with the standard deviation $\sigma(x, y)$ and the acquisition function. The model infers that the global minimum is at $(x, y) = (3.4, 0.7)$ with $E \approx 0.204$ eV. Fig. 3 appears to be the most uncertain around the PES peaks as one would expect from a model that favour regions with low energy, it would probably more noticeable if more samples were used.

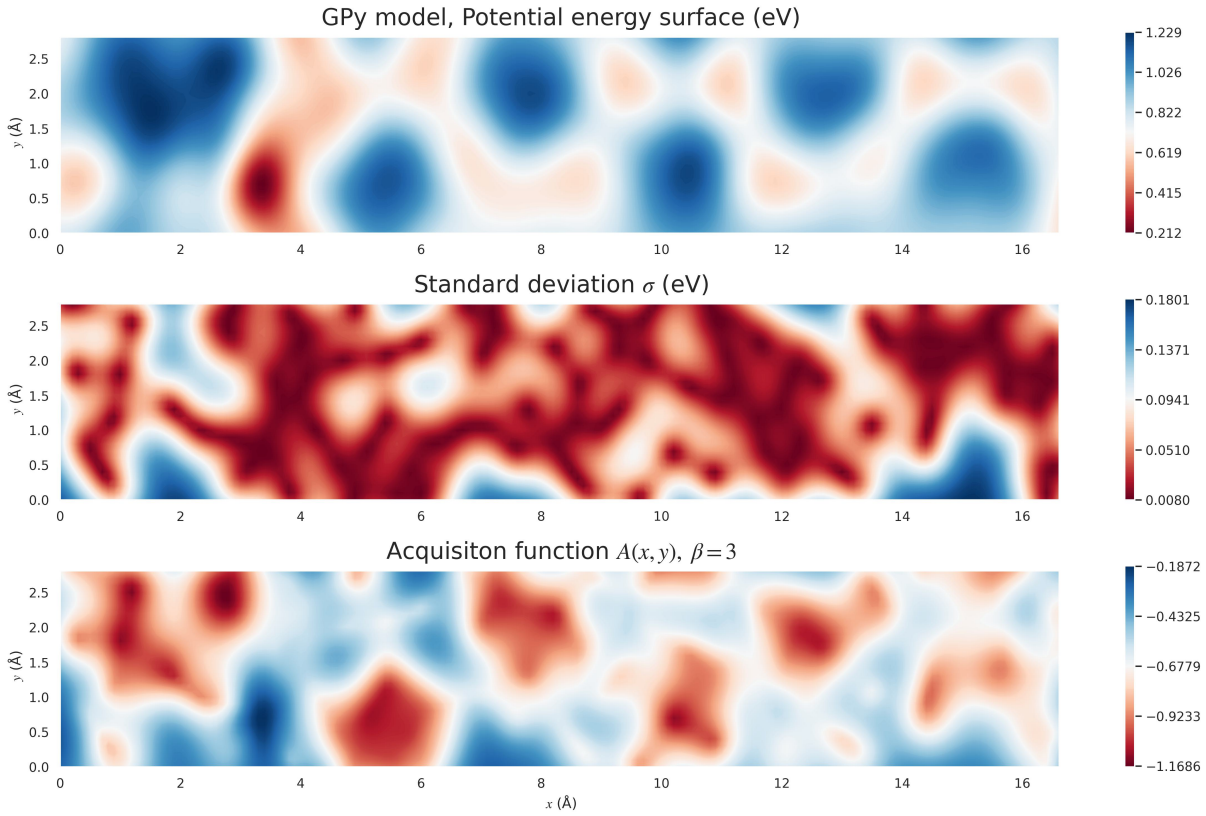

Figure 3: caption.

Task 4

General purpose model for sampling the entire PES. This model should prioritise regions with higher uncertainty and therefor we set $\beta = 100$. We will use the same priors for the RBF, bias kernel. As can be seen from fig. 4 this model should use more samples than the previous one, around 150 samples is suitable.

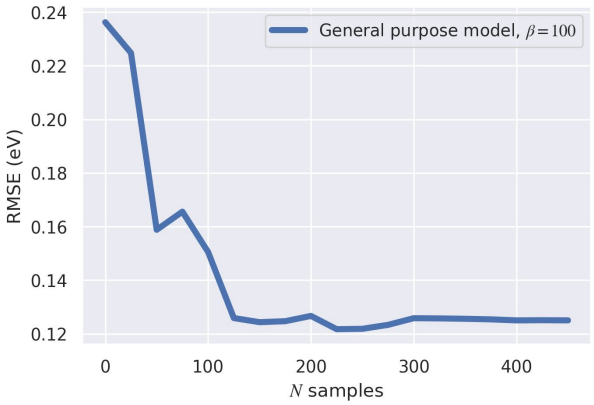

Figure 4: The relationship between the number of samples used and the RMSE.

In fig. 5 we can see the that model has a similar variance with compared to fig. 3, one small difference between them is that the general purpose model's σ is a bit smoother.

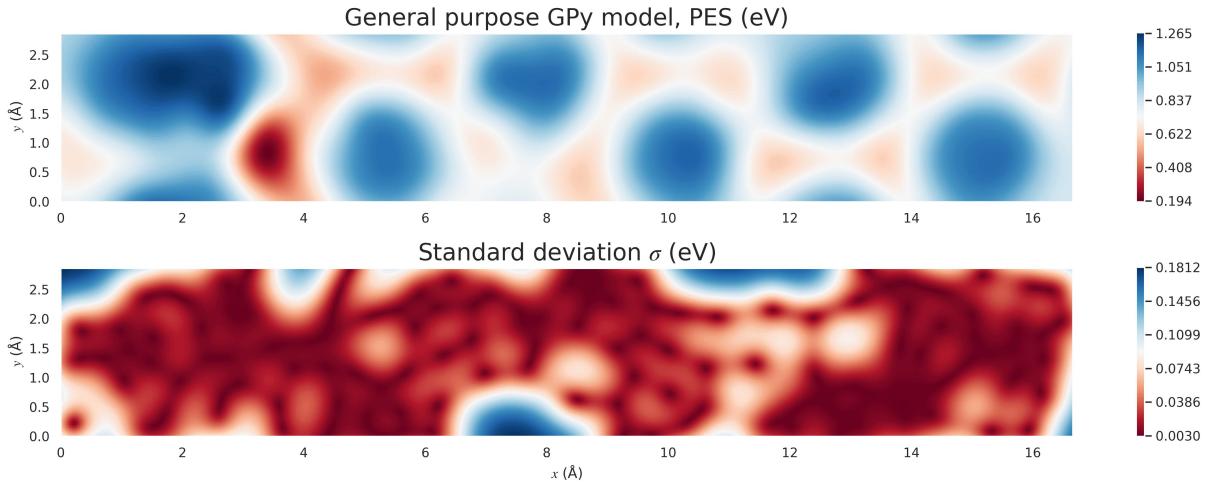

Figure 5: The predicted PES from the general purpose model and it's standard deviation.

In fig. 6 we examine the two models energy prediction along a path from the global minimum at $(x_{start}, y_{start}) = (3.4\,, 0.7\,)$ to a local minimum at $(x_{end}, y_{end}) = (11\,, 2.1\,)$,

$$(x, y) = (x_{start}, y_{start}) + \lambda(x_{end} - x_{start}, y_{end} - y_{start}), \tag{4}$$

where λ at is a parameter from 0 to 1. It is clear that the exploitation model (with a lower β) does a better job at predicting the energy in close proximity to the global minimum than the general purpose model. However, the general purpose model does have a lower variance on average over the path.

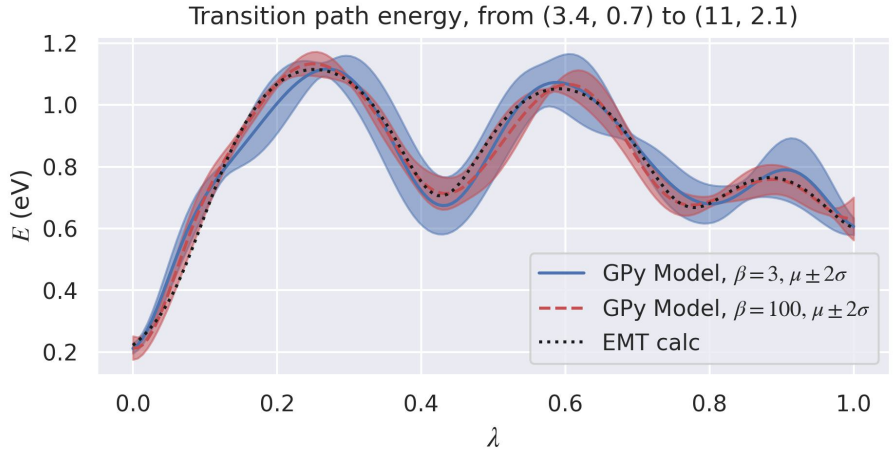

Figure 6: The predicted energy along a linear path from to global minimum to a local minimum for both models and the `EMT` calculator. The shaded blue, red regions correspond to two standard deviations away from their respective mean.

References

[1] L. Lista, *Statistical Methods for Data Analysis: With Applications in Particle Physics*, 3rd ed., ser. Lecture Notes in Physics. Springer Cham, 2023.

[2] S. Sisson, Y. Fan, and M. Beaumont, Eds., *Handbook of Approximate Bayesian Computation*, 1st ed. Chapman and Hall/CRC, 2018.