TIF345

ADVANCED SIMULATION AND MACHINE LEARNING

Project 2b:

Bayesian Optimization: Searching for the global minima

Author:
Simon Josefsson (simjos)
Tomas Lundberg (lutomas)

Examinator:
Paul Erhart

Tutor:
Eric Lindgren

December 8, 2021

Introduction

In this report, we will investigate the potential energy surface (PES) for an Au atom placed on top of an Au surface [1]. The atom that is placed on top of the surface is called an ad-atom. The resulting PES is both computationally costly to evaluate and contains many local minima. Thus, we will utilize Bayesian global optimization with the use of Gaussian processes (GPs). First, we will explore the PES by evaluating it over a dense grid with a solver based on effective medium theory (EMT). Second, to showcase the difficulty of optimizing the surface, we will perform local optimization on the PES and see how large fraction of the optimizations end up at the global minima. Third, we will deploy the Bayesian optimization with GPs with the goal of finding the global minima in as few EMT evaluations as possible. Fourth and finally, we will use a similar technique as previously but instead of finding the global minima as fast as possible, we will generate a general purpose approximation of the PES which can be used for further, and much faster, evaluation.

Task 1 - Analyzing the PES

Method: The energy of interest is given by

$$E(x, y, z) = E_{ad}(x, y, z) - E_{surface} \tag{1}$$

Where $E_{ad}(x, y, z)$ is the energy at the surface including the ad-atom at the given position, and $E_{surface}$, the constant energy of the surface without the ad-atom. Thus, for simplicity, we only calculate $E_{ad}(x, y, z)$, which will be proportional to $E(x, y, z)$. Moreover, we relax the z-coordinate to the lowest energy state such that we obtain $E_{ad}(x, y)$. We calculate the energy over a primitive cell grid, $x \in (0, 16.65653)$ [Å] and $y \in (0, 2.884996)$ [Å], with points spaced 0.1 Å, using an EMT potential solver of the `asap3` package. To account for atoms outside of the primitive cell that affects the energy, we perform the calculations using a supercell.

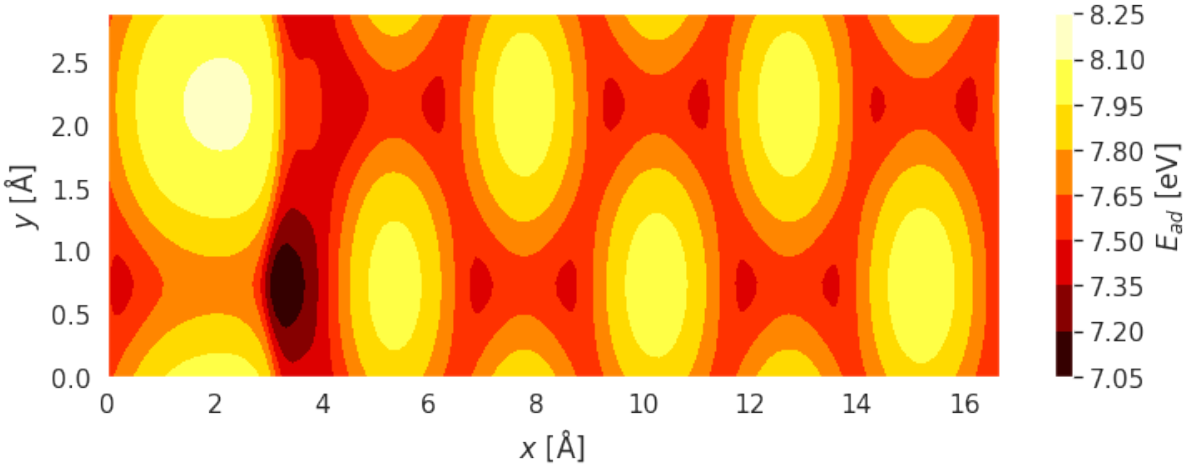


Figure 1: The PES of a primitive cell for x and y coordinates, evaluated on a grid with 0.1 Å between points using an EMT potential solver with relaxed z-coordinate. We note the global minima $E_{ad} \approx 7.077$ eV at $x \approx 3.2$ Å and $y \approx 0.8$ Å.

Results and Discussion: We show the results of the evaluated primitive cell PES in fig. 1. Remember that we have relaxed the z-coordinate to the lowest possible energy,

so we see the PES for x and y coordinates in Å as a heatmap. We note that the PES contains many local minima and maxima, with a distinct global minima $E_{ad} \approx 7.077$ eV at $x \approx 3.2$ Å and $y \approx 0.8$ Å. Furthermore, the standard deviation of the PES is around 0.2 eV which corresponds to a variance of about 0.04 eV^2. Also, when calculating the autocorrelation of the PES in x and y directions we obtain[1], that the autorrorelation passes $\exp(-1/2)$ around 0.5 Å. This corresponds roughly to the lengthscale we can expect. This will become relevant when we motivate our priors of the GPs in Task 3.

Task 2 - Local Search

Method: To show the difficulty of finding a global minima on a surface with many local minima, see fig. 1, we performed 500 local searches, using `scipy.optimize.minimize`, of randomly chosen starting points within the primitive cell. By visual inspection of fig. 1, we can expect the initialized points that fall roughly in the area $x \in [2, 5.5]$ [Å] and $y \in [0, 2]$ [Å], to end up at the global minima. This area corresponds to roughly 15% of the total area of the primitive cell, thus, we expect around 15% of the local searches to end up in the global minima.

Results and Discussion: The results of the 500 local searches are presented in fig. 2. We show the starting (*star*) and final position (*dot*) with a connecting line between them on top of the PES generated in the previous task. The fraction that ended up in the global minima was 14% which is close to what we expected.

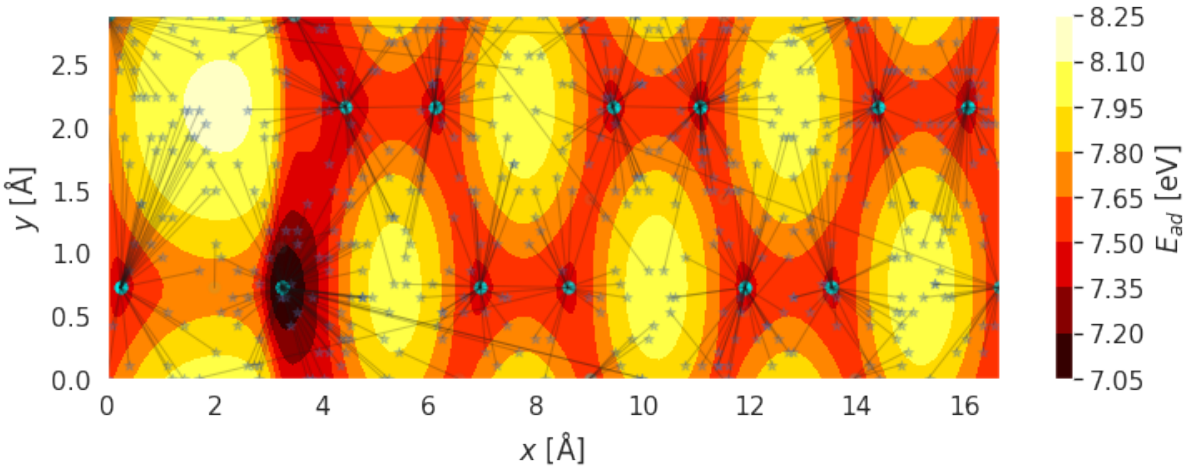


Figure 2: Starting (*blue stars*) and convergence points (*cyan dots*) connected by a *black* line on top of the PES calculated from Task 1 (see fig. 1) of 500 local optimizations with random starting points. We note that every local minima attracts the closely initialized starting points.

Task 3 - Bayesian Optimization

Method: To combat the issues with the local search approach in the previous task we now turn to global search and in particular to global Bayesian optimization with the help of Gaussian processes. We used a RBF and bias kernel to represent our PES. In line with what we found in Task 1, we wanted the priors of the RBF kernel to represent our observation that the lengthscale was around 0.5 Å and the variance around 0.04

[1]See figure in appended notebook. Not included in the report for lack of space.

eV^2. Thus we chose gamma priors with shape and rate parameters 2.5 and 3 for the lengthscale and 3 and 60 for the variance[2]. Moreover, we chose 5 points at random and calculated the energy using EMT for our initial data. To find the next point to sample, we fitted a GP on our data and maximized the lower confidence bound (LCB) acquisition function

$$A(x, y) = -\mu_{GP}(x, y) + \beta\sigma_{GP}(x, y), \tag{2}$$

where $\mu_{GP}(x, y)$ and $\sigma_{GP}(x, y)$ is the mean and std function of the GP, and β is a hyper parameter. The (x, y) corresponding to the maximum is the next point to sample. In particular, we performed 25 local searches over $A(x, y)$, and took the (x, y) that corresponded to the maximum value of the local searches. We terminated the search when the lowest energy found was lower than the one found in task 1, or if 150 iterations were reached.

To find the optimal β we did a line-scan for 33 values $\in [1, 5]$. For each β, ten models were fitted per β to get a better estimate of the number of evaluations before termination. Moreover, for a fair comparison, we set the random seed to be different for the ten evaluations per β, but the same across different values of β. This is because the initialization of points seemed to matter quite a lot for the time to convergence.

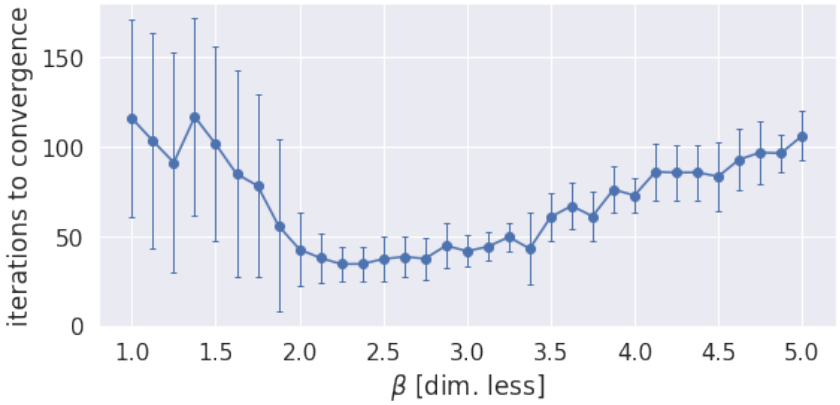


Figure 3: The mean, and standard deviation of the number of iterations to convergence for different values of β.

Results and Discussion: The result of the line-scan over β is found in fig. 3. We note a minimum, both in mean and standard deviation at $\beta = 2.25$. For larger values of β, i.e. more exploration, the average number of iterations to convergence goes up slowly, and so does the standard deviation. However, for lower values of β, that is , less exploration, the mean and in particular the standard deviation goes up drastically. This can be explained by our observation that the initial points of the model matters a lot. For example, if we have a low β, and no points are close to the global minima, then it is likely that the global minima will never be found. However, if one of the initial points are close to the global minima, then a low β will probably find the global minima even faster than for a higher β since it will exploit and sample the current minima more frequently. In fact, 7 out of the 10 fastest convergence times had $\beta < 2$.

To make use of the advantages of both exploration and exploitation one could employ a decreasing β such that there is more exploration early on and more exploitation

[2]Please see figure of the prior distributions in the appended notebook. Not inlcuded here for lack of space.

later. To avoid getting stuck at local minima, one could even combine this with some exploration technique inspired by ϵ-greedy [2]. That is, there is for every iteration some probability that a random point, independent of $A(x, y)$, is sampled. One could thus guarantee that a global optima will be found given enough time.

Nevertheless, for the model with optimal $\beta = 2.25$ found in fig. 3, we present the GP mean, GP sigma and acquisition function together with the samples in fig. 4. The model converged in 21 iterations and the RBF variance was 0.06 eV^2 and the lengthscale was 0.54 Å, close to what we expected. Remember that the acquisition function is the middle plot weighted by beta minus the top plot. As expected, the model samples the most close to the current minima. As a result, there are large regions that go completely unsampled where the uncertainty, thus, is large. The PES of this GP model is not very representative of the one shown in fig. 1 at all. In the next task, we will construct our acquisition function such that we obtain a more accurate representation of the PES.

Task 4 - Transition path barriers

Method: In the previous approach the goal was to find the global minimum in as few EMT calculations as possible. Here we will take a similar approach, but instead we want to calculate the PES more generally. With a more general potential over the entire PES, we can calculate numerous quantities of the the ad-atom such as diffusion barriers and thermodynamic properties [1]. To achieve a more general purpose model we set β to be large in eq. (2), more precisely we set it to be 1000. This implies a model with basically only exploration of the samplings.

As already mentioned, in this task we used a similar approach as in the previous task with the GPs scheme. We used the same priors for the kernel, but instead of terminating the search when the global minima was found, we simply looped over a specified number of iterations, while also tracking the root-mean-square error (RMSE) for each iteration.

Then we examined the energy landscape along a linear path from the global minimum to the local energy minimum close to $(11, 2.1)$. The points on the path were calculated by

$$(x, y) = (x_{start}, y_{start}) + \lambda(x_{end} - x_{start}, y_{end} - y_{start}) \tag{3}$$

for different $\lambda \in [-0.05, 1.05]$, where (x_{start}, y_{start}) are the global minimum point and (x_{end}, y_{end}) are the local minimum. To initialise the model we used the start and end points and three randomly chosen points in the cell.

Results and Discussion: In fig. 5a the RMSE per iteration is shown. We see that the rate of decline of the RMSE is rapidly declining with number of iterations. When training the model we chose to use 200 samples, where the decline in RMSE has decreased drastically and the balance between cost (number of iterations) and accuracy is sufficient for our case of examining global optimization. As predicted, the larger β-value leads to a sampling which is more evenly distributed over the primitive cell than in the previous task. In fig. 6 the sampling together with the GP mean, GP sigma and acquisition function are presented. We can see that the GP mean more closely resembles the PES in fig. 1 than the previous task and that the more even sampling leads to a smaller variance over the entire cell. The RBF variance was in this case 0.04 eV^2 and the lengthscale 0.62 Å, close to the mode of our priors.

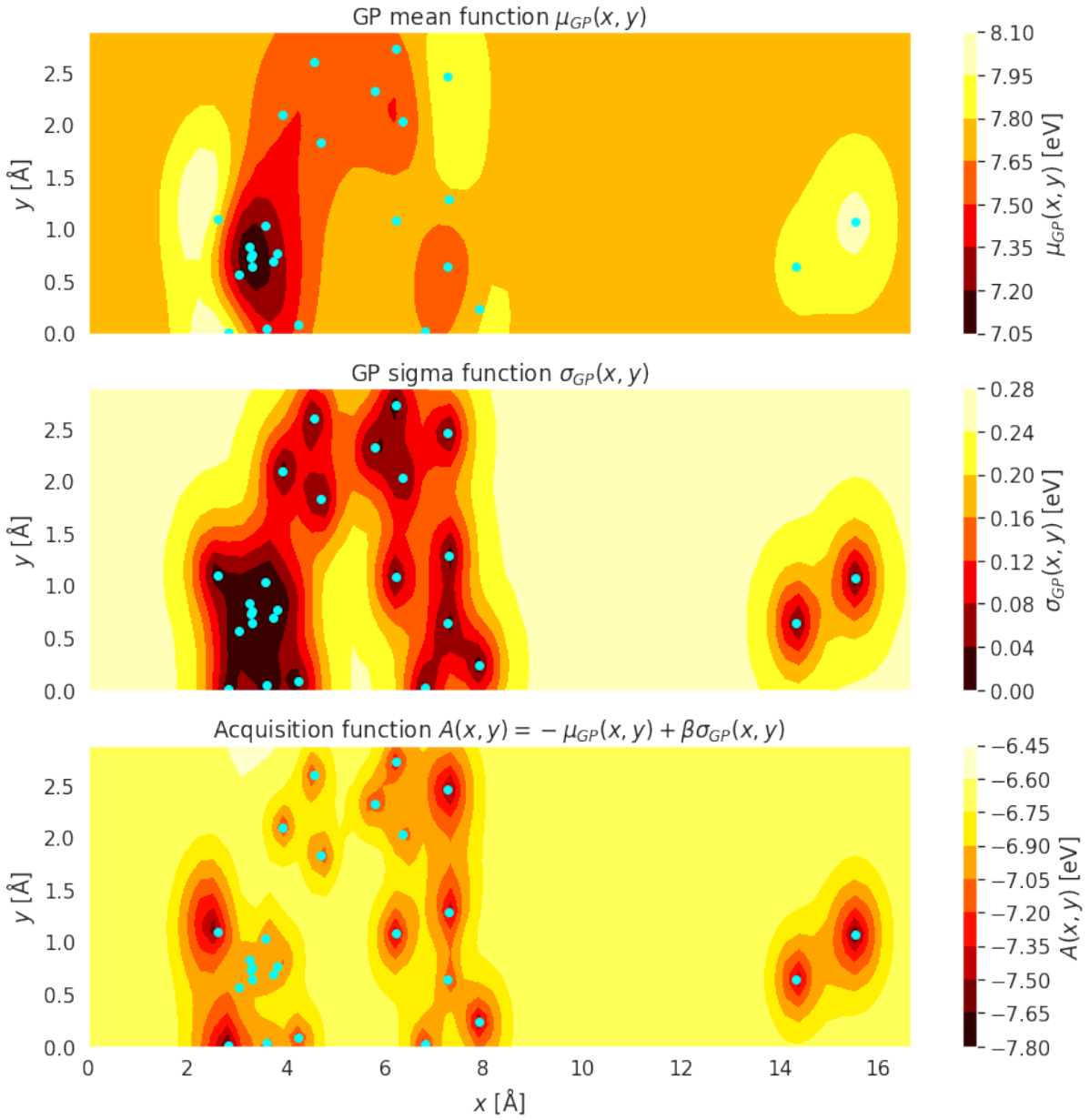
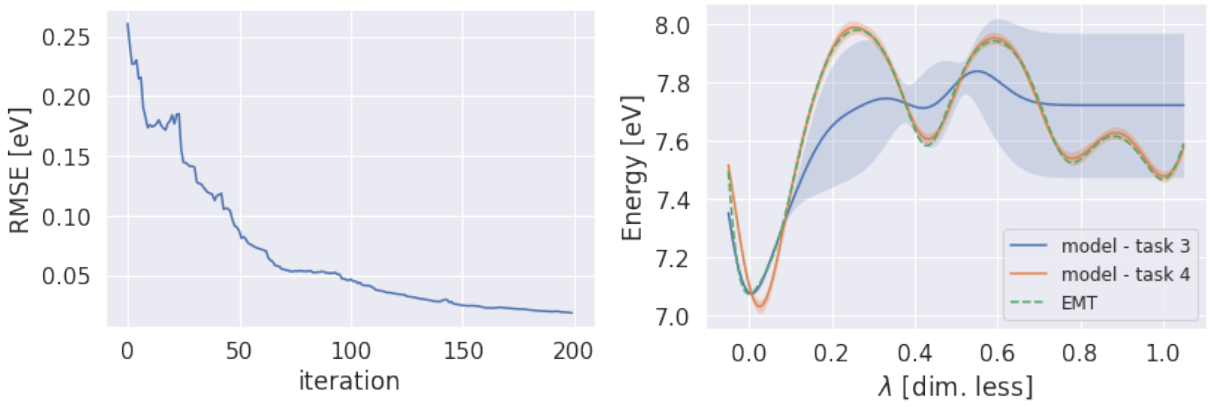


Figure 4: The resulting GP mean (*top*), GP sigma (*middle*) and acquisition (*bottom*) function of the GP-model with the β that converged the fastest in fig. 3, i.e. $\beta = X$. The sampled points are displayed on all heat maps. As expected, the area close to the global minima is sampled much more frequently than other areas.

It is interesting to note that there is more frequent sampling on the border of the primitive cell, see fig. 6. This might be because, even though we only display the primitive cell, of course, the GP sigma function could be calculated outside of the primitive cell, where we have not sampled. Thus, reducing the variance on the border is the best you can do to reduce the uncertainty outside of the primitive cell.

In fig. 5b the energy landscape for the general model and previous tasks models are presented. The energy landscape is taken along the linear path between the global minimum and a local minimum as indicated by the dashed line in fig. 6. We see that the general model gives an accurate estimate of the correct energy potential, that is calculated by EMT. The model from the previous task estimate the global minimum better than the more general model. This is not surprising as the two models focus on



(a) (b)

Figure 5: *(a)* The RMSE of the GP-model with $\beta = 1000$ over 200 iterations. *(b)* The energy along the transition path marked by the *dashed black line* in fig. 6, calculated with the best GP-model of task 3 (*blue*), the $\beta = 1000$ model of task 4 (*orange*) and through the EMT (*dashed in green)*. The shaded regions corresponds to $\pm 2\sigma$.

different things. The model of the previous task focuses only on effectively calculating the global minimum and does so with high confidence. However, this comes at the expense of the accuracy of the entire energy potential.

Conclusion

Naturally when choosing which model to use the application is at the center. As seen in fig. 5b the models are good at different things. If we would only be interested in the properties related to the global minimum we should choose the task 3 model with a β-value around 2.25 to get the properties as cheap and accurately as possible. If we are interested in properties connected to the entire potential we should instead use the more general model with a larger β-value.

This report shows the complexity of calculating global minimas and maximas, in the presence of many local extreme points. Moreover, we have showed the power and flexibility of optimizing with global Bayesian optimisation with the help of Gaussian processes with several applications.

References

[1] P. Erhart and E. Lindgren, "Tif345 project 2b: Bayesian optimization: Searching for the global minima," November 2021.

[2] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, second ed., 2018.

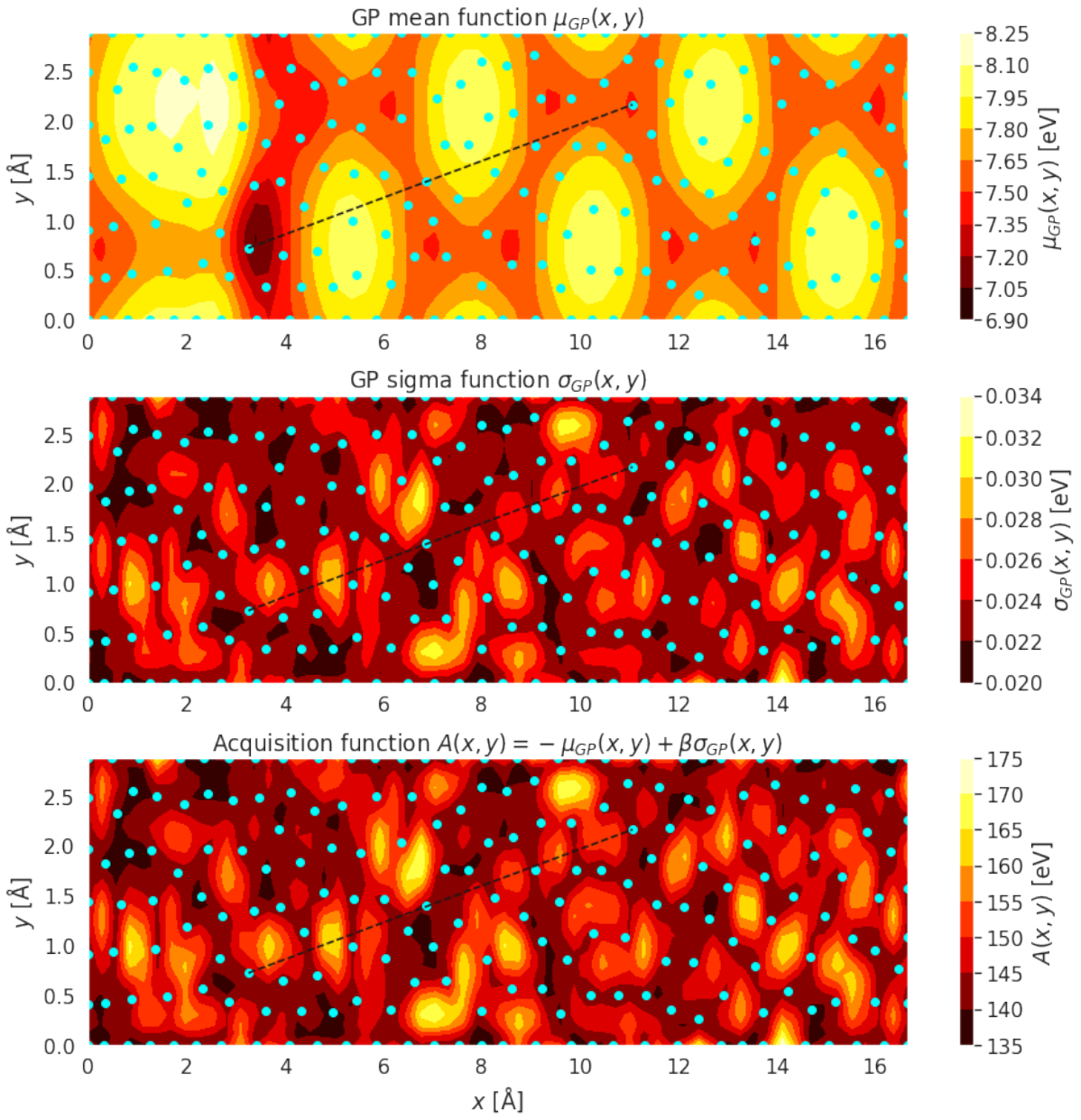


Figure 6: The resulting GP mean (*top*), GP sigma (*middle*) and acquisition (*bottom*) function of the GP-model constructed to sample the PES as evenly as possible with $\beta = 1000$. The 200 samples are shown together with the line represented by fig. 5b.