Exploring Methods for Finding Maxima Using Bayesian Optimisation

Rebekah Fearnhead[1], Daniel Dodd[2]

[1]University of Durham, [2]STOR-i, Lancaster University

1. Introduction

- The global optimisation of black-box functions which are expensive and potentially gradient-free is an important problem in industry, for example in training self driving cars.
- Bayesian optimisation is an approach which has been shown to obtain better results, with fewer evaluations, compared to other methods such as random-search based methods.
- The general idea is to construct a probabilistic model of the object function which can then be used to sequentially decide where to evaluate it next.
- This model can then be improved using a surrogate function which adds extra noise to the sampled data and these two methods will be compared.

2. Bayesian Optimisation

- A prior distribution is placed over the function, and with each new observation, this model is refined using Bayesian posterior updating.
- Gaussian processes are known to make well-calibrated predictions and are therefore a common choice for the prior.
- Using the posterior, acquisition functions are induced to evaluate the utility of the candidate points for the next evaluation of the objective function.

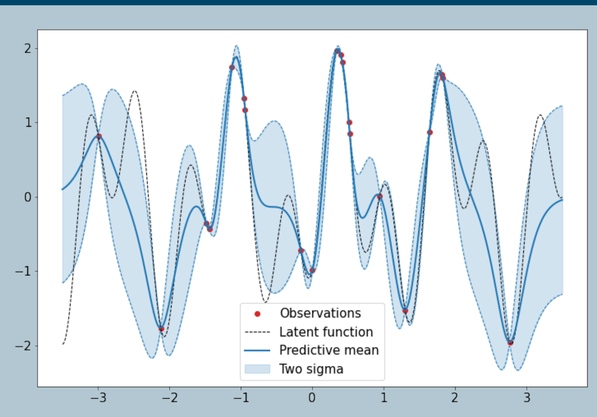

Figure: A graph of the predicted distribution of a function (blue) and two standard deviations of the prediction, given the observed data points (red), compared to the true function (black).

3. Acquisition Functions

Acquisition functions are used to decide the most beneficial places to sample to find the maximum (or minimum) of a function. There are different types of functions that place different weights on exploration (points with high uncertainty), and exploitation (points where the model prediction is high).

- **Improvement Based Policies**
  These focus on favouring points that are likely to improve on the current maximum value, τ.
  - Probability of Improvement

$$\alpha_{PI}(\mathbf{x}; \mathcal{D}_n) := \mathbb{P}[v > \tau] = \Phi\left(\frac{\mu_n(\mathbf{x}) - \tau}{\sigma_n(\mathbf{x})}\right).$$

  - Expected Improvement

$$\alpha_{EI}(\mathbf{x}; \mathcal{D}_n) := (\mu_n(\mathbf{x}) - \tau)\Phi\left(\frac{\mu_n(\mathbf{x}) - \tau}{\sigma_n(\mathbf{x})}\right) + \sigma_n(\mathbf{x})\phi\left(\frac{\mu_n(\mathbf{x}) - \tau}{\sigma_n(\mathbf{x})}\right).$$

- **Optimistic Policies**
  These aim to be optimistic by looking at the upper confidence bounds for the predicted values of the data points that can be sampled. The hyper-parameter, β_n can be adjusted, with higher values placing more weight on exploration.
  - Upper Confidence Bound

$$\alpha_{UCB}(\mathbf{x}; \mathcal{D}_n) := \mu_n(\mathbf{x}) + \beta_n\sigma_n(\mathbf{x}).$$

4. Problems with Acquisition Functions

- Traditional acquisition functions are often trapped sampling a small area after locating a local optima.
- This is caused by the greater importance placed on exploitation rather than exploration.
- Surrogate functions aim to resolve this by increasing the posterior variance which encourages exploration.

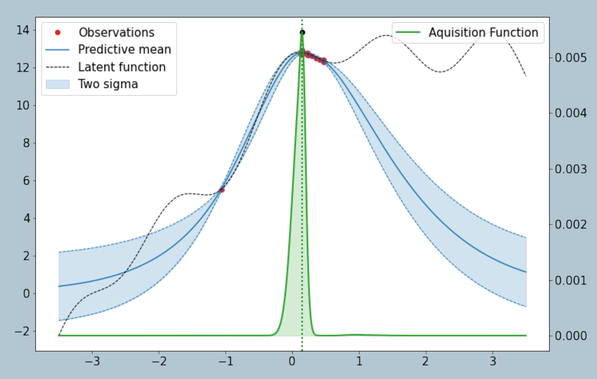

Figure: A graph of the expected improvement acquisition function getting trapped at a local maximum of the function being optimised.

5. Surrogate Functions

- Surrogate functions improve the performance of acquisition functions by adding noise to the model which increases the posterior variance to encourage exploration.
- This can be shown as

$$f(\mathbf{x}) = g(\mathbf{x}, \mathbf{h}), \qquad g \sim \mathcal{GP}, \qquad \mathbf{h} \sim \mathcal{N}(0, \sigma_h),$$

  where g is a well behaved function following a GP prior distribution.
- This therefore adds non-linear interactions between a random variable, $\mathbf{h}$ and the sampled values of $\mathbf{x}$ to the predicted function distribution.

6. Method Comparison

- The differences between using surrogate functions and traditional methods on three different functions are illustrated below.
- This simulation is done in Python using the GPJax module.
- For each simulation, the starting position is the same for expected improvement both with and without using a surrogate function, with the same data point values given.
- A graph showing regret is then plotted to show how quickly each of the methods find the global maximum by plotting the difference between the current largest value found and the true global maximum.

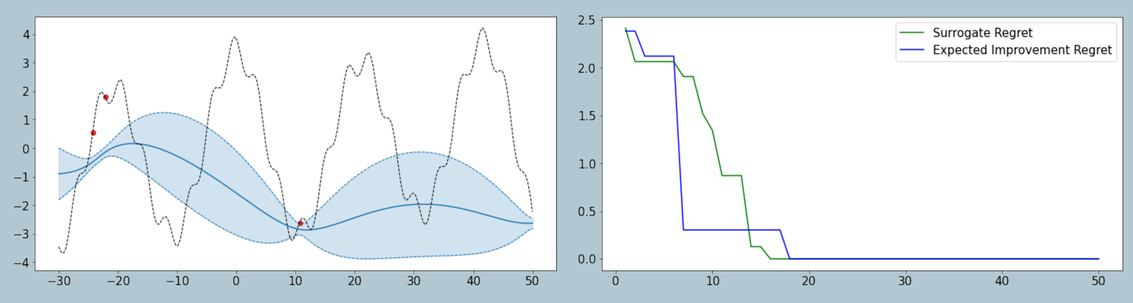

Figure: A graph showing the starting data points and smooth function being sampled to find the maximum and the regret of each method after more data points are sampled.

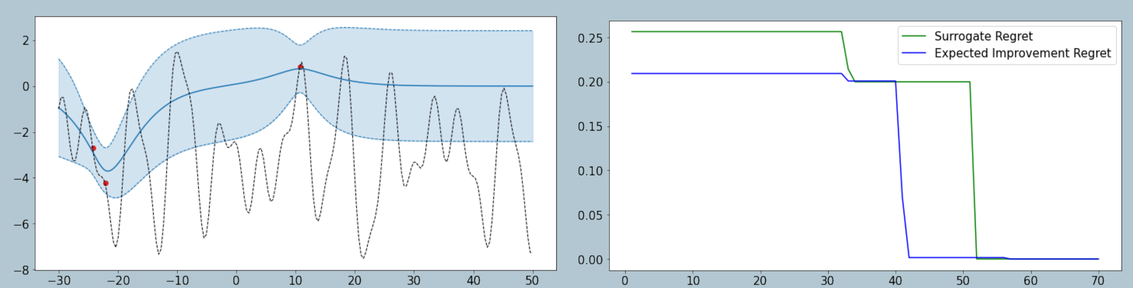

Figure: Similar graphs for a function with more local maxima.

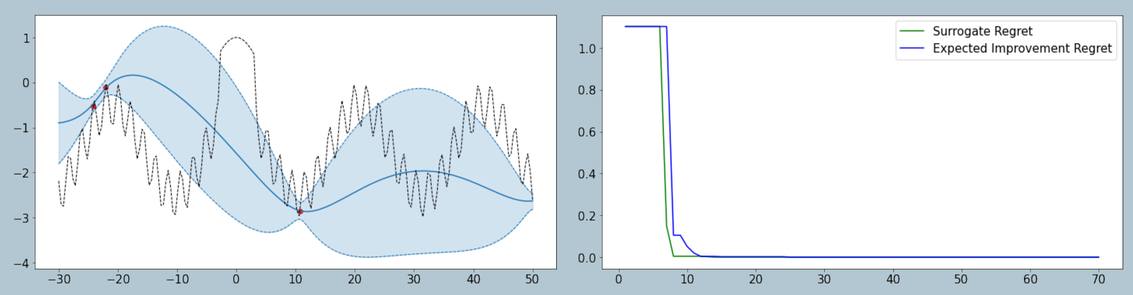

Figure: Similar graphs for a function partially formed by a step function containing the global maximum.

7. Observations and Conclusions

- From the tested functions above, it can be seen that for the smooth function and the step function, using a surrogate function before applying Expected Improvement outperformed just using acquisition function.
- However, for the function with lots of local maxima, the acquisition function performs better. Even though the surrogate model gets a regret value of 0 first, using expected improvement gets close to the maximum quicker. This could be because the function already is quite noisy and unpredictable so the posterior variance is already large.
- To further investigate the improvements on Bayesian Optimisation that using a surrogate function method causes, the standard deviation of $\boldsymbol{h}$ and the functions the method is tested on could be changed.

8. References

Bodin, E., Kaiser, M., Kazlauskaite, I., Dai, Z., Campbell, N., & Ek, C. H. (2020, November). Modulating surrogates for Bayesian Optimization. In *International Conference on Machine Learning* (pp. 970-979). PMLR.

Rasmussen, & Williams, Christopher K. I. (2006). *Gaussian processes for machine learning*. MIT Press.

Shahriari, Swersky, K., Wang, Z., Adams, R. P., & de Freitas, N. (2016). Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE, 104*(1), 148–175.

rebekah.m.fearnhead@durham.ac.uk