Bayesian optimization with sequential variable selection

M. Ben Salem
École des Mines de St-Etienne - ANSYS, Inc.

Supervisor(s): Olivier Roustant (École des Mines de St-Etienne), Fabrice Gamboa (Institut de Mathématiques de Toulouse) and Lionel Tomaso (ANSYS, Inc.)

Ph.D. expected duration: 2015-2018

Adress: 11 Avenue Albert Einstein, 69100 Villeurbanne, France

Email: malek.ben-salem@emse.fr

Abstract:

Bayesian optimization techniques have been successfully applied in various fields. In the same time, its use is generally limited to a moderate number of dimensions. Therefore, there is a great interests in Bayesian optimization algorithms suitable for high-dimensional problems. Here, we consider the case of functions when only few variables are influential. We propose an algorithm that consists in sequential Bayesian optimization combined with sequential variable's selection. The sampled points aims at accelerating the optimization and refining the variable's selection. To do so, new criteria are introduced within the Gaussian Process regression model and a specific set of stationary kernels. Without loss of generality, let us consider the double exponential kernel $k(\mathbf{x}, \mathbf{y}) = \exp(-\sum_{i=0}^{n} \theta_i (x_i - y_i)^2)$. For a given variable x_i, if the value of the kernel parameter θ_i is equal to zero then x_i does not impact on the variation of the surrogate model prediction and the variation of all its derived criteria in particular the expected improvement. The existence of non-influential variables $\mathbf{x}_i$ in an optimization algorithm make the search harder.

The basic idea consists in filtering some variables to make the optimization converges faster. The filtering is based on the value of their parameters $\theta_0 = (\theta_1, \ldots, \theta_d)$ estimated by Maximum Likelihood Estimation (MLE). For a given iteration, we consider I_M, respectively I_m, the set of indices of major, respectively minor variables. Eliminating the minor variables at once could lead to inaccurately filter some influential variables due a misleading estimation of the kernel parameters. Therefore, the proposed algorithm questions sequentially the initial variable splitting. In fact, we compute at each iteration the next points coordinates in I_M, $x_{|M}$ by the Efficient Global Optimization (EGO) [1] algorithm in the subspace of major variables.

For the subspace of minor parameters, the coordinates are computed using a novel criterion called the contrast. The maximum contrast strategy consists in choosing the values of minor parameters that maximizes the contrast between the actual kernel and a challenging kernel estimation. The challenging kernel is selected within a given value of likelihood ratio. We call the challenging kernel the maximum allowable doubt kernel defined by $\theta^\star$ where $\theta^\star$ is the solution of the following problem.

$$\begin{aligned} \underset{\theta}{\text{maximize}} \quad & \sum_{i \in I_m} \theta_i \\ \text{subject to} \quad & 2 \ln \left(\frac{L(\theta_0)}{L(\theta)} \right) < limit \end{aligned}$$

Essentially, if our variable splitting is relevant the optimization would continue to converge. If some influential variables are considered minor parameters, than we have to find a point that can

highlight that effect in the future iterations. The maximum contrast is the location where the kriging models defined by θ_0 and $\theta^\star$ "disagree" the most. We propose two criterion to compute the contrast of a point x:

- The expected likelihood ratio after adding the point x.
- The differences between the two model predictions at x.

The performances of our algorithm, called MADMEC, are compared to a similar algorithm where the minor variables of the sampled points are chosen randomly and the classical EGO. The results show the usefulness of our approach. We display an example in Figure.1 where we show the evolution of the minimum value for the Hartmann6 function in 25 dimension (19 useless dimension).

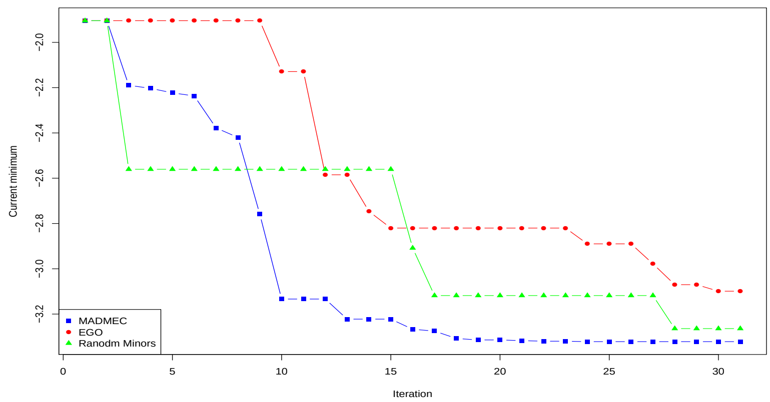

Figure 1: Evolution of the best value of Hartmann6 function embedded in 25 dimension optimization using several algorithms.

References

[1] Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

Short biography – After an engineering diploma from the french engineering school ISIMA (Clermont-Ferrand) and a Research Masters Degree from the Blaise Pascal university (Clermont-Ferrand), Malek Ben Salem started a PhD thesis at École des Mines Saint-Étienne. He is funded by a CIFRE grant from the ANSYS company, subsidized by the French National Association for Research and Technology (ANRT, CIFRE grant number 2014/1349) and he works on surrogate models aggregations, prediction uncertainty quantification and sequential design.