

---

# Active Learning for Approximation of Expensive Functions with Normal Distributed Output Uncertainty

---

Joachim van der Herten  
Ivo Couckuyt  
Dirk Deschrijver  
Tom Dhaene

Ghent University, Technologiepark 15, B-9052 Ghent, BELGIUM

JOACHIM.VANDERHERTEN@INTEC.UGENT.BE  
IVO.COUCKUYT@INTEC.UGENT.BE  
DIRK.DESCHRIJVER@INTEC.UGENT.BE  
TOM.DHAENE@INTEC.UGENT.BE

## Abstract

When approximating a black-box function, sampling with active learning focussing on regions with non-linear responses tends to improve accuracy. We present the FLOLA-Voronoi method introduced previously for deterministic responses, and theoretically derive the impact of output uncertainty. The algorithm automatically puts more emphasis on exploration to provide more information to the models.

## 1. Introduction

Consider an unknown multivariate black-box function  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ . Our aim is obtaining an approximation of the response  $\tilde{f}$  using machine learning methodology (regression) by evaluating  $f$  on distinct points  $\mathbf{p}_i \in \mathbb{R}^d$ . The corresponding responses are subject to normal distributed uncertainty  $\lambda$ :  $y_i \sim \mathcal{N}(f(\mathbf{p}_i), \lambda)$ . However, assuming evaluation of  $f$  involves a significant cost, only a limited amount of  $N$  evaluations can be performed. Intelligent choice of these evaluations (*experimental design*) can result in significantly better approximations.

Traditional Design of Experiments for computer experiments select all  $N$  data points prior to evaluation. These space-filling methodologies do not allow exploiting information obtained by each evaluation. In contrast, we present an active learning approach developed in the context of surrogate modeling known as FLOLA-Voronoi (van der Herten et al., 2015). Starting from a small set of initial evaluations, iteratively new evaluation(s) are chosen based on an analysis of the previous responses. As a result, the algorithm increases the information density in non-linear areas that are more difficult to approximate. Originally

the algorithm was formulated for deterministic responses, we analyze the impact of uncertain outputs theoretically in Section 2, and illustrate the effect in Section 3.

## 2. Method

We present the FLOLA-Voronoi method, an improved version of the LOLA-Voronoi method (Crombecq et al., 2010) which has been applied successfully previously (Deschrijver et al., 2011; Rosenbaum & Schulz, 2012; Deschrijver et al., 2012). The method can operate without intermediate model training, balances exploration and exploitation and remains efficient for high-dimensional functions. Two scores (computed by two algorithms) are aggregated for each previously evaluated point  $\mathbf{p}$ . The aggregated scores are ranked, and the highest ranked points are chosen to select a new point for evaluation local to the chosen point, using for instance the maximin criterion.

### 2.1. Exploration

The Voronoi exploration algorithm (Crombecq et al., 2009) estimates the size of the voronoi cells whose centroids corresponds to earlier evaluated data points: each point is assigned a score proportional to the relative size of the cell with respect to the other cells. As only a ranking of cell sizes is required, a Monte-Carlo estimate is sufficiently accurate and no full Voronoi tessellation needs to be computed. This algorithm only considers the d-dimensional input space and is insensitive to  $\lambda$ .

### 2.2. Exploitation

For a previously evaluated point  $\mathbf{p}_r$ , neighbouring points are selected to compute a gradient estimate  $\mathbf{g}$  by solving a linear system of equations. Denote the set  $N(\mathbf{p}_r)$  as the set containing the indices of the neighbouring points of  $\mathbf{p}_r$ . The original crosspolytope selection criterion presented in (Crombecq et al., 2010) does not scale well to

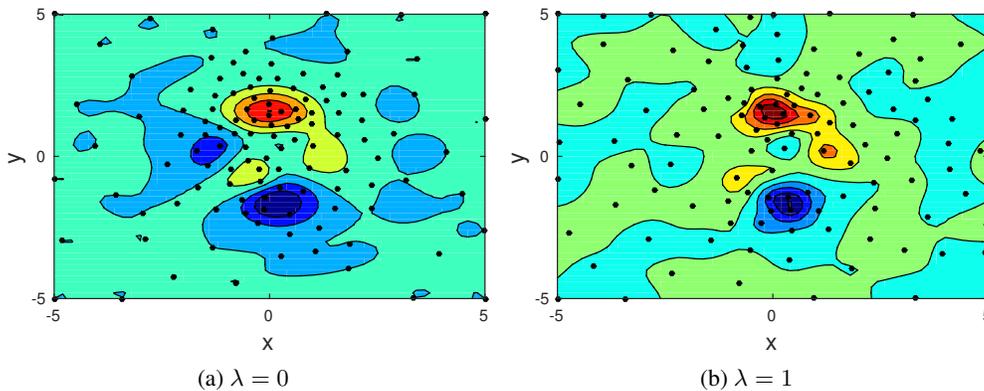


Figure 1. FLOLA-Voronoi applied to the *Peaks* problem, with and without noise. Clearly, the noiseless case is more focussed on the non-linear area whereas the noisy case results in more exploration.

high-dimensional input spaces, however this can be resolved by using a faster selection scheme (van der Hertén et al., 2015). The non-linearity error  $E$  sums the differences between the linear prediction and the response obtained from the simulator. This score depends on the output and therefore is sensitive to  $\lambda$ . We now show how much the uncertainty impacts  $E$ , under the assumption of normal distributed noise:  $y_i = f(\mathbf{p}_i) + \epsilon_i$  with  $\epsilon_i \sim \mathcal{N}(0, \lambda)$ . By expanding the terms and applying the triangular inequality:

$$E(\mathbf{p}_r) \leq \sum_{i \in N(\mathbf{p}_r)} |f(\mathbf{p}_i) - (f(\mathbf{p}_r) + \mathbf{g} \cdot (\mathbf{p}_i - \mathbf{p}_r))| + \sum_{i \in N(\mathbf{p}_r)} \zeta_i. \quad (1)$$

The upper bound of the score corresponds to the score for deterministic responses, complemented with a sum of noise terms  $\zeta_i = |\epsilon_i - \epsilon_r|$ . We denote the sum of the noise terms as random variable  $X$ . All  $\epsilon$  are normally distributed with zero mean, so the difference of both variance terms is also normally distributed with zero mean and variance  $2\lambda$ . This means each  $\zeta_i$  is distributed according to a folded normal distribution:

$$\zeta_i \sim \mathcal{FN} \left( 2\sqrt{\frac{\lambda}{\pi}}, 2\lambda \left( 1 - \frac{2}{\pi} \right) \right).$$

The distribution of  $X$  depends on  $T = |N(\mathbf{p}_r)|$ , and the shape of its probability density function resembles the log-normal distribution, its explicit formulation can be calculated but is a lengthy expression. Defining  $u(t) = (t - 1)^2 + 1$  and  $v(t) = t^2 - 2t + 2$ , the expectancy and variance are given by following expressions:

$$\begin{aligned} \mathbb{E}[X] &= \frac{2}{\pi} \sqrt{\lambda(\pi - 2)} (T + (T - 1) \sqrt{u(T)}), \\ \text{Var}[X] &= \frac{2\lambda(\pi - 2) (4(T - 1) + 5\pi v(T))}{\pi^2} - \mathbb{E}[X]^2. \end{aligned}$$

The variance of  $X$  perturbs the non-linearity score most. From the equation it can be derived that higher  $\lambda$  and a higher  $T$  cause more variance on the distribution of  $X$ . Using the neighbourhood selection mechanism of FLOLA-Voronoi the latter can be countered by including only the most significant neighbours in Equation 1 (instead of all).

### 3. Illustration

The perturbations caused by  $\text{Var}[X]$  impact the ranking of FLOLA-Voronoi causing more exploration, as the exploitation score becomes less decisive. This is illustrated by selecting 120 data points from the same function (*Peaks*) twice, once with output noise and once without for a constant  $T$ . The sample distributions together with a contour plot of a fitted Kriging model are shown in Figure 1. Clearly, the perturbed non-linearity scores cause more exploration which is a desirable property.

### 4. Conclusion

We presented the FLOLA-Voronoi algorithm for efficient model-independent sequential sampling of expensive black-box functions, and analysed the impact of uncertain response values on the algorithm. We have shown how the exploitation component of the algorithm is affected, and illustrated the resulting enhanced exploration. This is a desirable property in the presence of output uncertainty, as the model typically requires more information for linear areas.

### Acknowledgements

This research has (partially) been funded by the Inter university Attraction Poles Programme BESTCOM initiated by the Belgian Science Policy Office. Ivo Couckuyt is a post-doctoral research fellow of the Research Foundation Flanders (FWO).

## References

- Crombecq, K., Couckuyt, I., Gorissen, D., and Dhaene, T. Space-filling Sequential Design Strategies for Adaptive Surrogate Modelling. In *Soft Computing Technology in Civil, Structural and Environmental Engineering (CSC 2009)*, volume 92, pp. Paper-50, 2009.
- Crombecq, K., Gorissen, D., Deschrijver, D., and Dhaene, T. A Novel Hybrid Sequential Design Strategy for Global Surrogate Modelling of Computer Experiments. *SIAM Journal of Scientific Computing*, 33(4):1948–1974, 2010.
- Deschrijver, Dirk, Crombecq, Karel, Nguyen, Huu Minh, and Dhaene, Tom. Adaptive Sampling Algorithm for Macromodeling of Parameterized-Parameter Responses. *Microwave Theory and Techniques, IEEE Transactions on*, 59(1):39–45, 2011.
- Deschrijver, Dirk, Vanhee, Filip, Pissoort, Davy, and Dhaene, Tom. Automated near-field scanning algorithm for the EMC analysis of electronic devices. *IEEE Transactions on Electromagnetic Compatibility*, 54(3):502–510, 2012.
- Rosenbaum, Benjamin and Schulz, Volker. Comparing sampling strategies for aerodynamic Kriging surrogate models. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, 92(11-12):852–868, 2012.
- van der Herten, J, Couckuyt, I, Deschrijver, D, and Dhaene, T. A Fuzzy Hybrid Sequential Design Strategy for Global Surrogate Modeling of High-Dimensional Computer Experiments. *SIAM Journal on Scientific Computing*, 37(2):A1020–A1039, 2015.