Output-Weighted Importance Sampling for Uncertainty Quantification

Antoine Blanchard* and Themistoklis Sapsis*

*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

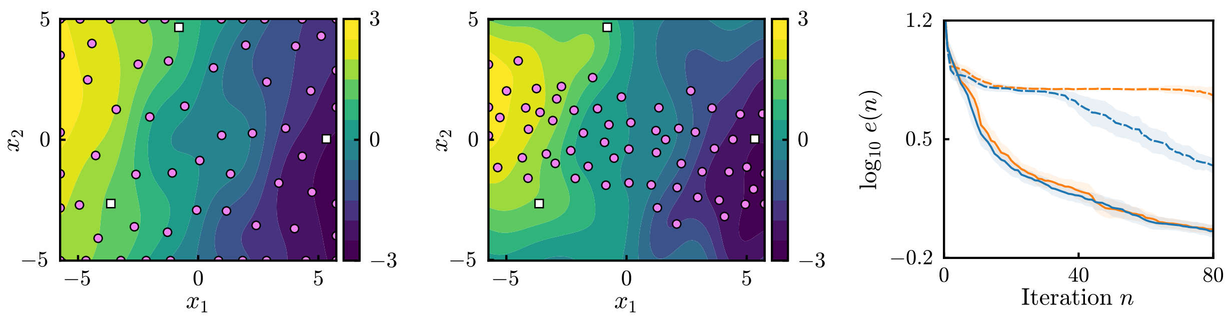
Abstract. We introduce a class of acquisition functions for sample selection that leads to faster convergence in applications related to Bayesian experimental design and uncertainty quantification of high-dimensional, strongly nonlinear systems. The approach follows the paradigm of active learning, whereby existing samples of a black-box function are utilized to optimize the next most informative sample. The proposed acquisition functions leverage the properties of the likelihood ratio, a quantity that acts as a probabilistic sampling weight and guides the active-learning algorithm towards regions of the input space that are deemed most relevant.

Introduction

Modern societies have reached such high levels of sophistication that real-world systems have become far too intricate to design, optimize and analyze using traditional techniques. Conceptually, these systems can be viewed as "black boxes" which can be learned using standard machine-learning techniques by making a series of queries and fitting a Gaussian process (GP) to the resulting input–output pairs. In practice, each query may take days or even weeks to produce a result, which means that each sample point must be selected gingerly. This difficulty is exacerbated when the black box has a large number of input parameters, possesses strongly nonlinear features, or has the ability to generate extreme events [1]. Active-learning methods aim to optimize the selection of each individual sample instead of drawing them in bulk from a carefully crafted distribution as with randomized sampling [2]. A critical issue in active learning is the choice of acquisition function, i.e., the criterion used to select which sample to query next in an optimal manner. We introduce a class of acquisition functions which guide the search algorithm toward regions of the input space that are associated with unusual output values associated with rare events. The key issue is that the proposed criteria contain a mechanism that accounts for the importance of the output relative to the input, and therefore are not limited to extreme-event quantification but can be applied to any problem related to uncertainty quantification.

Results and discussion

Our goal is to design acquisition functions that put a premium on the output values of previously visited data points while being computationally tractable. To quantify the importance of the output relative to the input, we utilize the likelihood ratio $w(\mathbf{x}) = p_{\mathbf{x}}(\mathbf{x})/p_\mu(\mu(\mathbf{x}))$, where $p_{\mathbf{x}}$ reflects uncertainty in the input and p_μ denotes the pdf of the GP posterior mean conditioned on the input. The likelihood ratio acts as a sampling weight, assigning to each point $\mathbf{x}$ a measure of "relevance" defined in probabilistic terms. We mathematical derive two novel likelihood-weighted acquisition functions, namely, $a_{\textit{IVR-LW}}(\mathbf{x}) = \sigma^{-2}(\mathbf{x}) \int \operatorname{cov}^2(\mathbf{x}, \mathbf{x}') w(\mathbf{x}')\,\mathrm{d}\mathbf{x}'$ and $a_{\textit{US-LW}}(\mathbf{x}) = \sigma^2(\mathbf{x}) w(\mathbf{x})$, where $\sigma^2(\mathbf{x})$ and $\operatorname{cov}(\mathbf{x}, \mathbf{x}')$ are the posterior variance and posterior covariance of the GP model, respectively. To demonstrate the benefits of the likelihood ratio in uncertainty quantification, we consider the stochastic oscillator of Mohamad and Sapsis [3] and monitor the error $e(n)$ in the estimated pdf of the mean displacement as a function of the number of samples n collected by the algorithm. Figure 1 shows that the likelihood ratio helps identify critical regions of the input space more efficiently than otherwise. As a result, the surrogate model constructed using the likelihood-weighted acquisition functions is able to predict the statistics of the output much more accurately than with traditional selection criteria.



(a) Learning with $a(\mathbf{x}) = \sigma^2(\mathbf{x})$ (b) Learning with $a(\mathbf{x}) = \sigma^2(\mathbf{x})w(\mathbf{x})$ (c) Error in estimated pdf

Figure 1: Progression of the active-learning algorithm (a) without and (b) with the likelihood ratio, as more samples (circles) are added to the initial dataset (squares); and (c) error in estimated pdf for the proposed criteria (solid) and their unweighted counterparts (dashed).

References

[1] Albeverio S., Jentsch V., Kantz H. (1999) Extreme events in nature and society. Springer Verlag, NY.

[2] Sacks J., Welch W. J., Mitchell T. J., Wynn H. P. (1989) Design and analysis of computer experiments. *Stat. Sci.* **4**:409–423.

[3] Mohamad M. A., Sapsis T. P. (2018) Sequential sampling strategy for extreme event statistics in nonlinear dynamical systems. *Proc. Natl. Acad. Sci.* **115**:11138–11143.