Approximation of differential entropy in Bayesian optimal experimental design

Yuya Suzuki
Aalto University
yuya.suzuki@aalto.fi

Coauthor(s): Chuntao Chen, Tapio Helin, Nuutti Hyvönen

We present a method to approximate differential entropy, which is then used to maximize the expected information gain, commonly used in Bayesian Optimal Experimental Design (BOED). We consider the standard additive noise model

$$y = \mathcal{G}(x;\xi) + \varepsilon,$$

where $y \in \mathbb{R}^d$ is the observed data, $\mathcal{G}$ denotes the forward map, $x \in \mathcal{X}$ is the quantity of interest, ξ is the design variable, and ε is the additive noise. We assume that ε is Gaussian and does not depend on the design ξ. We want to maximize expected information gain,

$$\begin{aligned} U(\xi) &:= \iint_{\mathbb{R}^d \times \mathcal{X}} \left[\log \frac{\pi(y|x;\xi)}{\pi(y;\xi)}\right] \pi(x,y;\xi)\mathrm{d}x\mathrm{d}y \\ &= -\int_{\mathbb{R}^d} \log\left(\pi\left(y;\xi\right)\right) \pi(y;\xi)\mathrm{d}y + \iint_{\mathbb{R}^d \times \mathcal{X}} \log\left(\pi\left(y|x;\xi\right)\right) \pi(y|x;\xi)\mathrm{d}y\mu(x)\mathrm{d}x \end{aligned}$$

where $\pi(x,y;\xi)$ stands for the joint distribution, $\pi\left(y|x;\xi\right)$ is the likelihood, $\pi\left(y;\xi\right)$ is the evidence, and $\mu(x)$ is the prior. In this setting, the second term above is independent of ξ. Thus we only need to maximize the first term, the differential entropy of the evidence.

We assume that evaluation of the forward map $\mathcal{G}$ is the dominant cost of the problem, and we want to reduce the number of forward map evaluations, say M, as much as possible. To this end, we construct a surrogate model of the evidence $\pi(y;\xi)$. We show that our proposed method can attain the convergence rate $\mathcal{O}(M^{-1/2})$, and this can be accelerated by quasi-Monte Carlo methods. We also numerically demonstrate our method for a PDE problem where the dimensionality of $\mathcal{X}$ is high.