Randomized Maximum Likelihood via High-Dimensional Bayesian Optimization

V. Breaz  
The University of Nottingham  

Supervisor(s): R. Wilkinson, M. Iglesias  
The University of Nottingham  

PhD expected duration: Jul. 2018 - Oct. 2022  
Jul. 2018 - Oct. 2020 at The University of Sheffield  

Address: School of Mathematical Sciences, Mathematical Sciences Building, University Park, Nottingham, NG7 2RD, United Kingdom  

E-mail: valentin.breaz@nottingham.ac.uk

Abstract:  
Randomized Maximum Likelihood (RML) is an approximate posterior sampling methodology; it has been widely used in Bayesian inverse problems with complex forward models, particularly in petroleum engineering applications. The algorithm proceeds by first perturbing the data and the prior mean, and then optimizing the unnormalised log-posterior using these perturbed values (see Algorithm 1).

Algorithm 1 Randomized Maximum Likelihood (RML)  

$m_{\text{RML}}$ : number of samples required  

For $m \in [m_{\text{RML}}]$:  

1. Sample $D_m \sim \mathcal{N}(D, \Sigma_{\text{obs}})$ from the Gaussian likelihood, $D_m \in \mathbb{R}^p$.  
2. Sample $\mu_m \sim \mathcal{N}(\mu, \Sigma)$ from the Gaussian prior, $\mu_m \in \mathbb{R}^D$.  
3. Construct log-posterior w.r.t. the randomizations $(D_m, \mu_m)$  

$$O_m(x) := \log \mathcal{N}(f(x)|D_m, \Sigma_{\text{obs}}) + \log \mathcal{N}(x|\mu_m, \Sigma)$$  

4. Obtain $x_m^\star$ as the maximizer $x_m^\star = \arg \max_x O_m(x)$.  

We address solving the optimization problems (1) efficiently in the challenging case of a high-dimensional input space $\mathbb{R}^D$. We focus on the specific scenario where the log-likelihood has a low-dimensional active subspace, that is when $\log \mathcal{N}(f(x)|D_m, \Sigma_{\text{obs}}) \approx g_m(A_T x)$, where $g : \mathbb{R}^d \to \mathbb{R}$ with $d \ll D$, and $A \in \mathbb{R}^{D \times d}$ with $A_TA = I_d$. This scenario is widely encountered in various Bayesian inverse problems applications.

We summarize our methodological contributions below:

- we propose a new methodology for maximizing the RML objective functions (1) for high-dimensional input spaces $x \in \mathbb{R}^D$ via high-dimensional Bayesian Optimization (HD-BO) with random embeddings;  
- we propose a natural way to exploit the shared simulator $f(x)$ which is present in all of the objective functions (1), as well as an adjustment needed to incorporate a Gaussian prior distribution without a low-dimensional structure (see Figure 1 for the first two iterations in the simpler case of a Uniform prior, where the objective functions become $O_m(x) = \log \mathcal{N}(f(x)|D_m, \Sigma_{\text{obs}}) \approx g_m(A_T^m x)$);  
- in the limited budget setting, our methodology often outperforms alternative gradient-free optimization methods in a series of synthetic and real-world experiments (Figure 2);  
- we visualize the posterior distribution in the active subspace, together with the samples produced by our methodology; we show that the samples are indeed close to ‘true’ RML samples (collected via an infinite computational budget), while also covering well the high posterior density regions (Figure 3).
Short biography – Valentin Breza is a PhD student in Gaussian Processes and Bayesian Inverse Problems at The University of Nottingham. The PhD scholarship is covered by the university and TotalEnergies.