MASCOT-NUM 2022
Clermont Ferrand, France
7-9 June 2022

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_{RML} : number of samples required
For $m \in [m_{RML}]$:

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

$$O_m(x) := \log \mathcal{N}_p(f(x)|\mathcal{D}_m, \Sigma_{\text{obs}}) + \log \mathcal{N}_D(x|\mu_m, \Sigma) \quad (1)$$

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 $\log \mathcal{N}_p(\mathcal{D}|f(x), \Sigma_{\text{obs}}) \approx g(A^T x)$, where $g : \mathbb{R}^d \to \mathbb{R}$ with $d \ll D$, and $A \in \mathbb{R}^{D \times d}$ with $A^T A = I_d$. This scenario is widely encountered in various Bayesian inverse problems applications.

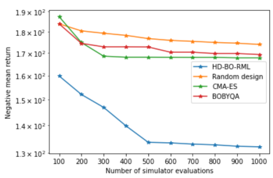
We summarize our methodological contributions below:

- 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 $O_m(x) = \log \mathcal{N}_p(f(x)|\mathcal{D}_m, \Sigma_{\text{obs}}) \approx g_m(A_m^T x)$);
- 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).

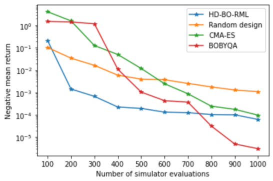
$$
\begin{array}{c} \\ O_1(y) \\ O_2(y) \\ \vdots \\ O_M(y) \end{array}
\begin{array}{c} \begin{array}{cccc} 1 & \dots & n & n+1 \end{array} \\
\begin{pmatrix}
y_1, O_1(y_1) & \vdots & y_n, O_1(y_n) & \color{teal}{y_{n+1}} \\
y_1, O_2(y_1) & \vdots & y_n, O_2(y_n) & \\
\vdots & \vdots & \vdots & \\
y_1, O_M(y_1) & \vdots & y_n, O_M(y_n) &
\end{pmatrix} \end{array}
\rightarrow
\begin{array}{c} \begin{array}{cccc} 1 & \dots & n & n+1 \end{array} \\
\begin{pmatrix}
y_1, O_1(y_1) & \vdots & y_n, O_1(y_n) & \color{orange}{y_{n+1}, O_1(y_{n+1})} \\
y_1, O_2(y_1) & \vdots & y_n, O_2(y_n) & \color{orange}{y_{n+1}, O_2(y_{n+1})} \\
\vdots & \vdots & \vdots & \vdots \\
y_1, O_M(y_1) & \vdots & y_n, O_M(y_n) & \color{orange}{y_{n+1}, O_M(y_{n+1})}
\end{pmatrix} \end{array}
$$

$$
\rightarrow
\begin{array}{c} \begin{array}{cccc} 1 & \dots & n+1 & n+2 \end{array} \\
\begin{pmatrix}
y_1, O_1(y_1) & \vdots & \color{orange}{y_{n+1}, O_1(y_{n+1})} & \\
y_1, O_2(y_1) & \vdots & \color{orange}{y_{n+1}, O_2(y_{n+1})} & \color{teal}{y_{n+2}} \\
\vdots & \vdots & \vdots & \\
y_1, O_M(y_1) & \vdots & \color{orange}{y_{n+1}, O_M(y_{n+1})} &
\end{pmatrix} \end{array}
\rightarrow
\begin{array}{c} \begin{array}{cccc} 1 & \dots & n+1 & n+2 \end{array} \\
\begin{pmatrix}
y_1, O_1(y_1) & \vdots & \color{orange}{y_{n+1}, O_1(y_{n+1})} & \color{orange}{y_{n+2}, O_1(y_{n+2})} \\
y_1, O_2(y_1) & \vdots & \color{orange}{y_{n+1}, O_2(y_{n+1})} & \color{orange}{y_{n+2}, O_2(y_{n+2})} \\
\vdots & \vdots & \vdots & \vdots \\
y_1, O_M(y_1) & \vdots & \color{orange}{y_{n+1}, O_M(y_{n+1})} & \color{orange}{y_{n+2}, O_M(y_{n+2})}
\end{pmatrix} \end{array}
$$

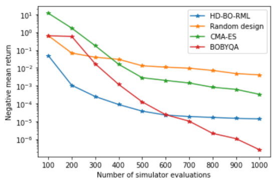
Figure 1: To condense notation, we write $n := n_0$ for the initial points and $M := m_{RML}$ for the RML objectives. Also, we suppress the superscript k corresponding to the random embedding R_k, i.e., $y_q := y_q^k$ and $O_m(y) := O_m(R_k y)$ for $q \in [n]$ and $m \in [M]$. From top to bottom, according to the arrows: first HD-BO iteration to collect $y_{n+1} := y_{n+1}^k = \arg\max_y a_{n+1}^k(y)$ (acquisition function) from $O_1(y) \sim \text{GP}$, then perform $f(R_k y_{n+1}^k)$ which generates data points $y_{n+1}, O_m(y_{n+1}) := y_{n+1}^k, O_m(R_k y_{n+1}^k)$ for all $m \in [M]$; second HD-BO iteration to collect $y_{n+2} := y_{n+2}^k = \arg\max_y a_{n+2}^k(y)$ from $O_2(y) \sim \text{GP}$, then perform $f(R_k y_{n+2}^k)$ which generates data points $y_{n+2}, O_m(y_{n+2}) := y_{n+2}^k, O_m(R_k y_{n+2}^k)$ for all $m \in [M]$.



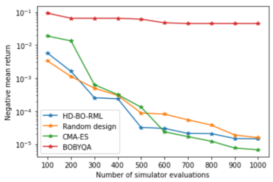
(a) Elliptic PDE



(b) Ebola



(c) MHD

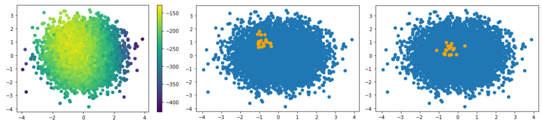


(d) HIV

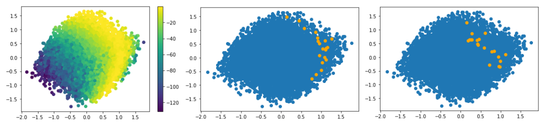
Figure 2: Given the samples $x_m^\star$ selected by our procedure (HD-BO-RML) and alternative competing gradient-free optimization methods, we plot negative mean returns

$$\frac{1}{m_{RML}} \sum_{m=1}^{m_{RML}} -O_m(x_m^\star)$$

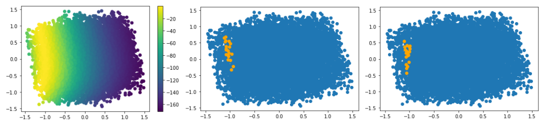
versus the computational budget (averaged over 5 optimization trials, lower values are better)



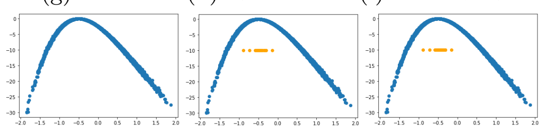
(a) PDE (b) RML oracle (c) HD-BO-RML



(d) Ebola (e) RML oracle (f) HD-BO-RML



(g) MHD (h) RML oracle (i) HD-BO-RML



(j) HIV posterior (k) RML oracle (l) HD-BO-RML

Figure 3: Posterior landscape in the active subspace (left), oracle RML samples (middle) and RML samples obtained by our procedure (right). The RML samples are displayed in orange.

Short biography – Valentin Breaz 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.