S1 Oracle Multiple Steps look-ahead Expected Loss

Denote by \( \eta_n = \min\{y_0, y_1, y_2, \ldots, y_{n-1}\} \) the value of the best visited location when looking at \( n \) evaluations in the future. Note that \( \eta_n \) reduces to the current best location \( \eta \) in the one step-ahead case. It is straightforward to see that

\[
\min(y_n, \eta_n) = \min(y, \eta).
\]

It holds that

\[
\Lambda_n(x_s, I_0, F_n(x_s)) = \int \min(y, \eta) \prod_{j=1}^{n} p(y_j | I_{j-1}, F_n(x_s)) dy_1 \ldots dy_n
\]

where the integrals with respect to \( x_2 \ldots dx_n \) are \( p(x_j | I_{j-1}, F_n(x_s)) = 1, \quad j = 2, \ldots, n \) since we don’t need to optimise for any location and \( p(y_j | x_j, I_{j-1}, F_n(x_s)) = p(y_j | I_{j-1}, F_n(x_s)) \). Notice that

\[
\prod_{j=1}^{n} p(y_j | I_{j-1}, F_n(x_s)) = p(y_n | I_{n-1}, F_n(x_s)) \prod_{j=1}^{n-1} p(y_j | I_{j-1} F_n(x_s))
\]

\[
= p(y_n, y_{n-1} | I_{n-2}, F_n(x_s)) \prod_{j=1}^{n-2} p(y_j | I_{j-1} F_n(x_s))
\]

\[
\ldots
\]

\[
= p(y_n, y_{n-1}, \ldots, y_2 | I_1, F_n(x_s)) \prod_{j=1}^{2} p(y_j | I_{j-1} F_n(x_s))
\]

\[
= p(y | I_0, F_n(x_s))
\]

and therefore

\[
\Lambda_n(x_s, I_0, F_n(x_s)) = \mathbb{E}[\min(y, \eta)] = \int \min(y, \eta) p(y | I_0, F_n(x_s)) dy
\]

S2 Formulation of the Oracle Multiple Steps look-ahead Expected Loss to be computed using Expectation Propagation

Assume that \( y \sim \mathcal{N}(y; \mu, \Sigma) \). Then we have that

\[
\mathbb{E}[\min(y, \eta)] = \int_{\mathbb{R}^n} \min(y, \eta) \mathcal{N}(y; \mu, \Sigma) dy
\]

\[
= \int_{\mathbb{R}^n-(\eta, \infty)^n} \min(y) \mathcal{N}(y; \mu, \Sigma) dy + \int_{(\eta, \infty)^n} \eta \mathcal{N}(y; \mu, \Sigma) dy.
\]

The first term can be written as follows:

\[
\int_{\mathbb{R}^n-(\eta, \infty)^n} \min(y) \mathcal{N}(y; \mu, \Sigma) dy = \sum_{j=1}^{n} \int_{P_j} y_j \mathcal{N}(y; \mu, \Sigma) dy
\]

where \( P_j := \{ y \in \mathbb{R}^n - (\eta, \infty)^n : y_j \leq y_i, \quad \forall i \neq j \} \). We can do this because the regions \( P_j \) are disjoint and it holds that \( \cup_{j=1}^{n} P_j = \mathbb{R}^n - (\eta, \infty)^n \). Also, note that the \( \min(y) \) can be replaced within the integrals since within each \( P_j \) it holds that \( \min(y) = y_j \). Rewriting the integral in terms of indicator functions we have that

\[
\sum_{j=1}^{n} \int_{P_j} y_j \mathcal{N}(y; \mu, \Sigma) dy = \sum_{j=1}^{n} \int_{\mathbb{R}^n} y_j \prod_{i=1}^{n} t_{j,i}(y) \mathcal{N}(y; \mu, \Sigma) dy
\]

(S.1)

where \( t_{j,i}(y) = \mathbb{I}\{y_i \leq y_j\} \) if \( j = i \) and \( t_{j,i}(y) = \mathbb{I}\{y_j \leq y_i\} \) otherwise.
The second term can be written as

$$
\int_{(\eta, \infty)^n} \eta \mathcal{N}(y; \mu, \Sigma) dy = \eta \int_{\mathbb{R}^n} \prod_{i=1}^n h_i(y) \mathcal{N}(y; \mu, \Sigma) dy
$$

(S.1)

where $h_i(y) = \mathbb{1}\{y_i > \eta\}$. Merge (S.1) and (S.2) to conclude the proof.

S3 Synthetic functions

In this section we include the formulation of the objective functions used in the experiments that are not available in the references provided.

<table>
<thead>
<tr>
<th>Name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>SinCos</td>
<td>$f(x) = x \sin(x) + x \cos(2x)$</td>
</tr>
<tr>
<td>Alpine2-$q$</td>
<td>$f(x) = \prod_{i=1}^q \sqrt{x_i} \sin(x_i)$</td>
</tr>
<tr>
<td>Cosines</td>
<td>$f(x) = 1 - \sum_{i=1}^2 (g(x_i) - r(x_i))$ with $g(x_i) = (1.6x_i - 0.5)^2$ and $r(x_i) = 0.3 \cos(3\pi(1.6x_i - 0.5))$.</td>
</tr>
</tbody>
</table>

Table 1: Functions used in the experimental section.

S4 Evaluating the effect of the loss function

To isolate the impact of the acquisition function on the performance of the optimisation we run an experiment in which the function to optimise is sampled from the GP used to perform the optimisation. In particular, we use the square exponential kernel with variance and length-scale fixed to 1 and we solve problems of dimensions 1 and 2 in $[0, 1]$ and $[0, 1] \times [0, 1]$ respectively. The average minimum value obtained by the GP-LCB, MPI, EL and GLASSES is shown in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>1D</th>
<th>2D</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP-LCB</td>
<td>-1.90</td>
<td>-1.28</td>
</tr>
<tr>
<td>MPI</td>
<td>-2.09</td>
<td>-1.15</td>
</tr>
<tr>
<td>EL</td>
<td>-2.35</td>
<td>-1.34</td>
</tr>
<tr>
<td>GLASSES</td>
<td>-2.38</td>
<td>-1.37</td>
</tr>
</tbody>
</table>

Table 2: Average min. results for 1D and 2D problems in which random samples from the model used to perform the optimisation are taken as objectives. GLASSES achieves the best results of the used acquisitions.
### S5 Standard deviation of the ‘gap’ measures

<table>
<thead>
<tr>
<th>Function</th>
<th>MPI</th>
<th>LCB</th>
<th>EL</th>
<th>GL-2</th>
<th>GL-3</th>
<th>GL-5</th>
<th>GL-10</th>
<th>GL-H</th>
</tr>
</thead>
<tbody>
<tr>
<td>SinCos</td>
<td>0.1502</td>
<td>0.1442</td>
<td>0.1221</td>
<td>0.0707</td>
<td>0.1429</td>
<td>0.1749</td>
<td>0.0862</td>
<td>0.0499</td>
</tr>
<tr>
<td>Cosines</td>
<td>0.0377</td>
<td>0.0368</td>
<td>0.0548</td>
<td>0.0394</td>
<td>0.0389</td>
<td>0.0417</td>
<td>0.0633</td>
<td>0.0135</td>
</tr>
<tr>
<td>Branin</td>
<td>0.0060</td>
<td>0.0121</td>
<td>0.0004</td>
<td>0.0020</td>
<td>0.0146</td>
<td>0.0036</td>
<td>0.0005</td>
<td>0.0030</td>
</tr>
<tr>
<td>Six-hump Camel</td>
<td>0.0065</td>
<td>0.0199</td>
<td>0.0063</td>
<td>0.0080</td>
<td>0.0104</td>
<td>0.0080</td>
<td>0.0096</td>
<td>0.0092</td>
</tr>
<tr>
<td>McCormick</td>
<td>0.0093</td>
<td>0.0091</td>
<td>0.0242</td>
<td>0.0152</td>
<td>0.0135</td>
<td>0.0128</td>
<td>0.0116</td>
<td>0.0129</td>
</tr>
<tr>
<td>Dropwave</td>
<td>0.0473</td>
<td>0.0595</td>
<td>0.0558</td>
<td>0.0293</td>
<td>0.0320</td>
<td>0.0238</td>
<td>0.0229</td>
<td>0.0407</td>
</tr>
<tr>
<td>Powers</td>
<td>0.0073</td>
<td>0.0073</td>
<td>0.0071</td>
<td>0.0186</td>
<td>0.0063</td>
<td>0.0147</td>
<td>0.0059</td>
<td>0.1415</td>
</tr>
<tr>
<td>Ackley-2</td>
<td>0.0491</td>
<td>0.0103</td>
<td>0.1197</td>
<td>0.1061</td>
<td>0.1349</td>
<td>0.1005</td>
<td>0.1171</td>
<td>0.1637</td>
</tr>
<tr>
<td>Ackley-5</td>
<td>0.0196</td>
<td>0.0181</td>
<td>0.1146</td>
<td>0.1809</td>
<td>0.1433</td>
<td>0.1401</td>
<td>0.1779</td>
<td>0.1361</td>
</tr>
<tr>
<td>Ackley-10</td>
<td>0.0015</td>
<td>0.0016</td>
<td>0.1519</td>
<td>0.0011</td>
<td>0.0020</td>
<td>0.0019</td>
<td>0.1386</td>
<td>0.1209</td>
</tr>
<tr>
<td>Alpine2-2</td>
<td>0.0957</td>
<td>0.0903</td>
<td>0.1132</td>
<td>0.0848</td>
<td>0.0534</td>
<td>0.0822</td>
<td>0.0878</td>
<td>0.0439</td>
</tr>
<tr>
<td>Alpine2-5</td>
<td>0.0679</td>
<td>0.0577</td>
<td>0.0579</td>
<td>0.0835</td>
<td>0.0878</td>
<td>0.0808</td>
<td>0.0777</td>
<td>0.0814</td>
</tr>
</tbody>
</table>

Table 3: Standard deviation of the average ‘gap’ measure (5 replicates) across different functions. EL-k is the expect loss function computed with k steps ahead at each iteration. GLASSES is the GLASSES algorithm, MPI is the maximum probability of improvement and GP-LCB is the lower confidence bound criterion.