Suppl. materials: 'GLASSES: Relieving The Myopia Of Bayesian Optimisation'

Javier González
University of Sheffield
Dept. of Computer Science &
Chem. and Biological Engineering
j.h.gonzalez@sheffield.ac.uk

Michael Osborne
University of Oxford
Dept. of Engineering Science
mosb@robots.ox.ac.uk

Neil D. Lawrence
University of Sheffield
Dept. of Computer Science
n.lawrence@sheffield.ac.uk

S1 Oracle Multiple Steps look-ahead Expected Loss

Denote by $\eta_n = \min\{\mathbf{y}_0, y_*, y_2 \dots, y_{n-1}\}$ the value of the best visited location when looking at n evaluations in the future. Note that η_n reduces to the current best lost η in the one step-ahead case. It is straightforward to see that

$$\min(y_n, \eta_n) = \min(\mathbf{y}, \eta).$$

It holds hat

$$\Lambda_n(\mathbf{x}_*|\mathcal{I}_0, \mathcal{F}_n(\mathbf{x}_*)) \quad = \quad \int \min(\mathbf{y}, \eta) \prod_{j=1}^{n} p(y_j|\mathcal{I}_{j-1}, \mathcal{F}_n(\mathbf{x}_*)) \mathrm{d}y_* \dots \mathrm{d}y_n$$

where the integrals with respect to $\mathbf{x}_2 \dots \mathrm{d}\mathbf{x}_n$ are $p(\mathbf{x}_j|\mathcal{I}_{j-1}, \mathcal{F}_n(\mathbf{x}_*)) = 1$, $j = 2, \dots, n$ since we don't need to optimise for any location and $p(y_j|\mathbf{x}_j, \mathcal{I}_{j-1}, \mathcal{F}_n(\mathbf{x}_*)) = p(y_j|\mathcal{I}_{j-1}, \mathcal{F}_n(\mathbf{x}_*))$. Notice that

$$\begin{aligned}
\prod_{j=1}^{n} p(y_j|\mathcal{I}_{j-1}, \mathcal{F}_n(\mathbf{x}_*)) &= p(y_n|\mathcal{I}_{n-1}, \mathcal{F}_n(\mathbf{x}_*)) \prod_{j=1}^{n-1} p(y_j|\mathcal{I}_{j-1}\mathcal{F}_n(\mathbf{x}_*)) \\
&= p(y_n, y_{n-1}|\mathcal{I}_{n-2}, \mathcal{F}_n(\mathbf{x}_*)) \prod_{j=1}^{n-2} p(y_j|\mathcal{I}_{j-1}\mathcal{F}_n(\mathbf{x}_*)) \\
&\dots \\
&= p(y_n, y_{n-1}, \dots, y_2|\mathcal{I}_1, \mathcal{F}_n(\mathbf{x}_*)) \prod_{j=1}^{2} p(y_j|\mathcal{I}_{j-1}\mathcal{F}_n(\mathbf{x}_*)) \\
&= p(\mathbf{y}|\mathcal{I}_0, \mathcal{F}_n(\mathbf{x}_*))
\end{aligned}$$

and therefore

$$\Lambda_n(\mathbf{x}_*|\mathcal{I}_0, \mathcal{F}_n(\mathbf{x}_*)) = \mathbb{E}[\min(\mathbf{y}, \eta)] = \int \min(\mathbf{y}, \eta) p(\mathbf{y}|\mathcal{I}_0, \mathcal{F}_n(\mathbf{x}_*)) d\mathbf{y}$$

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

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

$$\begin{aligned}
\mathbb{E}[\min(\mathbf{y}, \eta)] &= \int_{\mathbb{R}^n} \min(\mathbf{y}, \eta) \mathcal{N}(\mathbf{y}; \mu, \Sigma) d\mathbf{y} \\
&= \int_{\mathbb{R}^n - (\eta, \infty)^n} \min(\mathbf{y}) \mathcal{N}(\mathbf{y}; \mu, \Sigma) d\mathbf{y} + \int_{(\eta, \infty)^n} \eta \mathcal{N}(\mathbf{y}; \mu, \Sigma) d\mathbf{y}.
\end{aligned}$$

The first term can be written as follows:

$$\int_{\mathbb{R}^n - (\eta, \infty)^n} \min(\mathbf{y}) \mathcal{N}(\mathbf{y}; \mu, \Sigma) d\mathbf{y} = \sum_{j=1}^{n} \int_{P_j} y_j \mathcal{N}(\mathbf{y}; \mu, \Sigma) \mathrm{d}\mathbf{y}$$

where $P_j := \{\mathbf{y} \in \mathbb{R}^n - (\eta, \infty)^n : y_j \leq y_i, \ \forall i \neq j\}$. 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 $\min(\mathbf{y})$ can be replaced within the integrals since within each P_j it holds that $\min(\mathbf{y}) = y_j$. Rewriting the integral in terms of indicator functions we have that

$$\sum_{j=1}^{n} \int_{P_j} y_j \mathcal{N}(\mathbf{y}; \mu, \Sigma) \mathrm{d}\mathbf{y} = \sum_{j=1}^{n} \int_{\mathbb{R}^n} y_j \prod_{i=1}^{n} t_{j,i}(\mathbf{y}) \mathcal{N}(\mathbf{y}; \mu, \Sigma) \mathrm{d}\mathbf{y} \tag{S.1}$$

where $t_{j,i}(y) = \mathbb{I}\{y_i \leq \eta\}$ if $j = i$ and $t_{j,i}(y) = \mathbb{I}\{y_j \leq y_i\}$ otherwise.

The second term can be written as

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

where $h_i(\mathbf{y}) = \mathbb{I}\{y_i > \eta\}$. Merge (S.1) and (S2) 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.

Name	Function
SinCos	$f(x) = x\sin(x) + x\cos(2x)$
Alpine2-q	$f(\mathbf{x}) = \prod_{i=1}^{q} \sqrt{x_i}\sin(x_i)$
Cosines	$f(\mathbf{x}) = 1 - \sum_{i=1}^{2}(g(x_i) - r(x_i))$ with $g(x_i) = (1.6x_i - 0.5)^2$ and $r(x_i) = 0.3\cos(3\pi(1.6x_i - 0.5))$.

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.

	1D	2D
GP-LCB	-1.90	-1.28
MPI	-2.09	-1.15
EL	-2.35	-1.34
GLASSES	-2.38	-1.37

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

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

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.