Asynchronous Batch Bayesian Optimisation with Improved Local Penalisation

Ahsan Alvi
Binxin Ru, Jan-Peter Calliess, Stephen Roberts, Michael A. Osborne

ICML 2019
Talk Overview

• Bayesian optimisation (BO) recap
• Synchronous vs asynchronous BO
• Our Method
  – Design of penaliser
  – Locally estimated Lipschitz constant
• Empirical results
1. Bayesian Optimisation (BO)

- To solve the global optimisation

\[ x^* = \arg \min_{x \in \mathcal{X}} f(x) \]

- The objective function

\[ f(\ ) \]

Non-convex
Expensive
Noisy
1. Bayesian Optimisation (BO)

\[ x^* = \arg\min_{x \in \mathcal{X}} f(x) \]

\[ f \sim \mathcal{GP}(\mu_t, K_t) \]

\[ x_t+1 = \arg\max_{x \in \mathcal{X}} \alpha_t(x) \]
2. Synchronous Batch BO

- Enable *multiple* evaluations in parallel
2. Asynchronous Batch BO

- Maximise utilisation of parallel workers

![Sync Batch BO (B=3)]

<table>
<thead>
<tr>
<th>C1</th>
<th>1</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>C3</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>
2. Asynchronous Batch BO

- Maximise utilisation of parallel workers

Sync Batch BO (B=3)

- C1
  - 1
  - 4

- C2
  - 2
  - 5

- C3
  - 3
  - 6

Async Batch BO (B=3)

- C1
  - 1
  - 4
  - 7

- C2
  - 2
  - 5
  - 9

- C3
  - 3
  - 6
  - 8

Time
3. Our Method

- A new async batch BO: **Penalising Locally for Asynchronous Bayesian Optimisation on k Workers (PLAyBOOK)**

\[
x^q = \arg \max_{x \in \mathcal{X}} \left\{ \alpha(x) \prod_{i=1}^{q-1} \psi(x | x^i) \right\}
\]
3. Our Method

- **Penalising Locally for Asynchronous Bayesian Optimisation on k Workers (PLAyBOOK)**

  \[
  x^q = \arg \max_{x \in \mathcal{X}} \left\{ \alpha(x) \prod_{i=1}^{q-1} \psi(x|\mathbf{x}^i) \right\}
  \]

- Empirically show: PLAyBOOK outperforms
  - other async BO methods
  - its sync. variants in both **time and sample** efficiency
4. Penaliser design

• Our hard penaliser (HLP):

$$\psi_{HLP}(x|x^q) = \min \left\{ \frac{\hat{L}||x-x^q||}{|\mu(x^q)-M|+\sigma(x^q)}, 1 \right\}$$

• LP (Gonzalez et al., 2016):

$$\psi_{LP}(x|x^q) = \Phi \left( \frac{\hat{L}||x-x^q||-|\mu(x^q)-M|}{\sigma(x^q)} \right)$$
5. Empirical Results: Async. vs. Sync.

- **PLAyBOOK-HL: Ackley 5-D: B=4 and B=16**
5. Empirical Results: Async. methods

- Tuning 9 hyperparameters of a CNN for CIFAR-10
Thank you!

Meet us at poster #213!