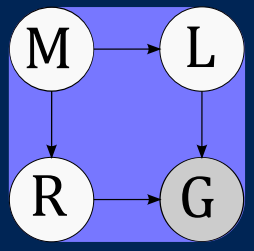


Machine Learning
Research Group

UNIVERSITY OF
OXFORD

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)

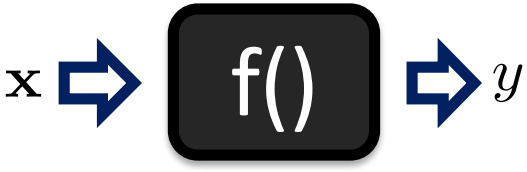
- To solve the global optimisation

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$$

- The objective function

1. Bayesian Optimisation (BO)

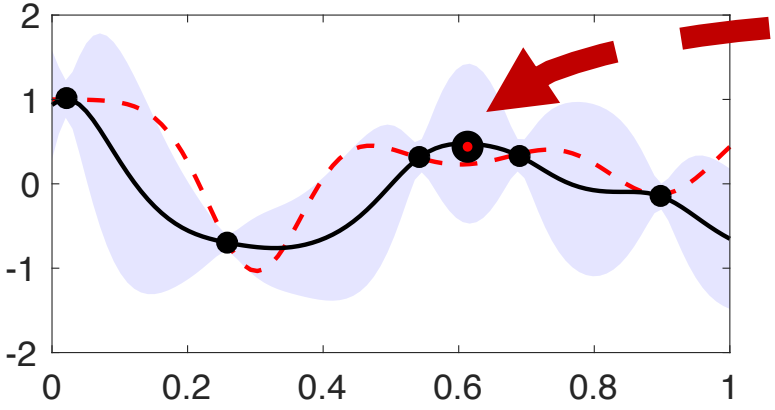
$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$$

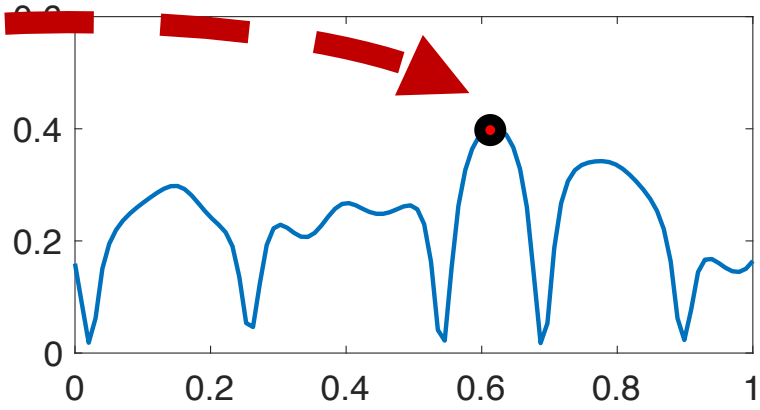


$$f \sim \mathcal{GP}(\mu_t, K_t)$$

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x})$$

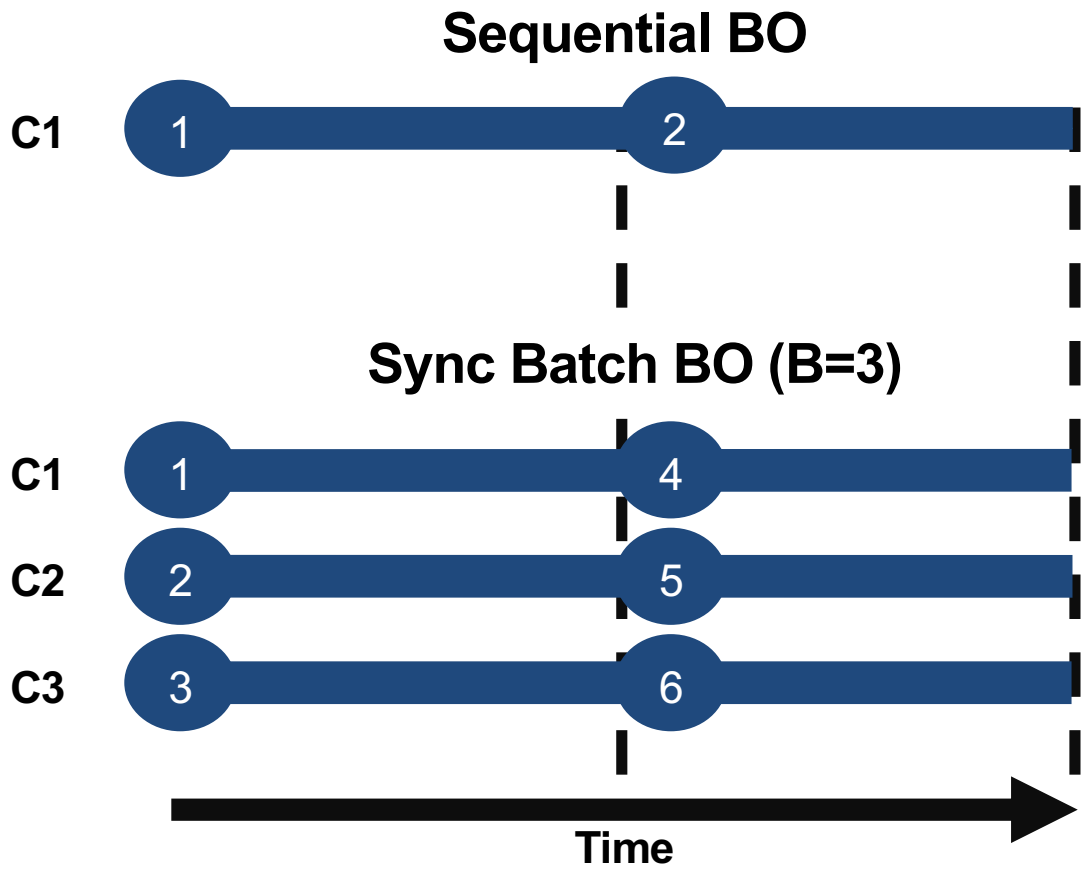
$$\mathbf{x}_{t+1} = \arg\max_{\mathbf{x}\in\mathcal{X}} \alpha_t(\mathbf{x})$$





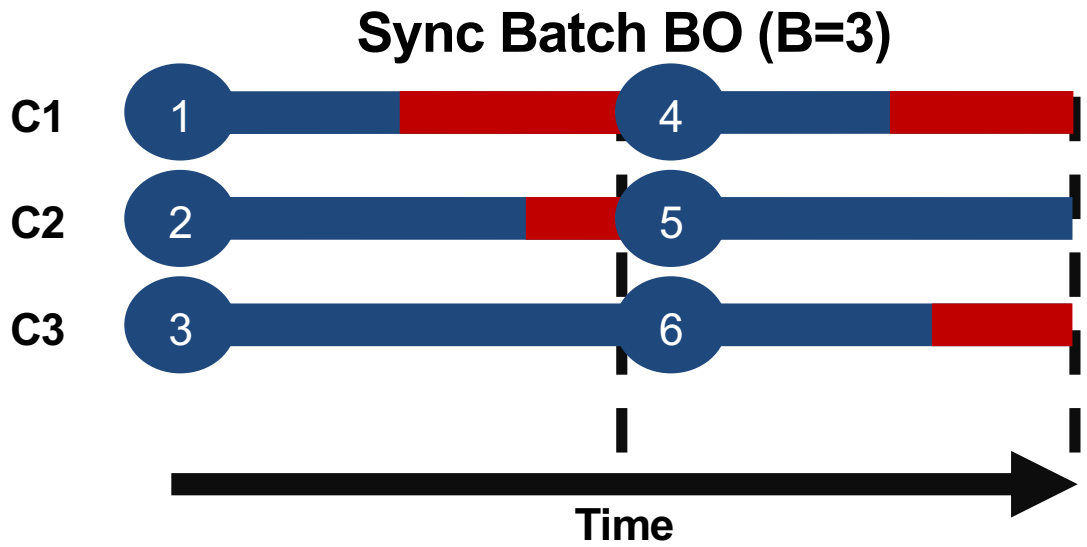
2. Synchronous Batch BO

- Enable *multiple* evaluations in parallel



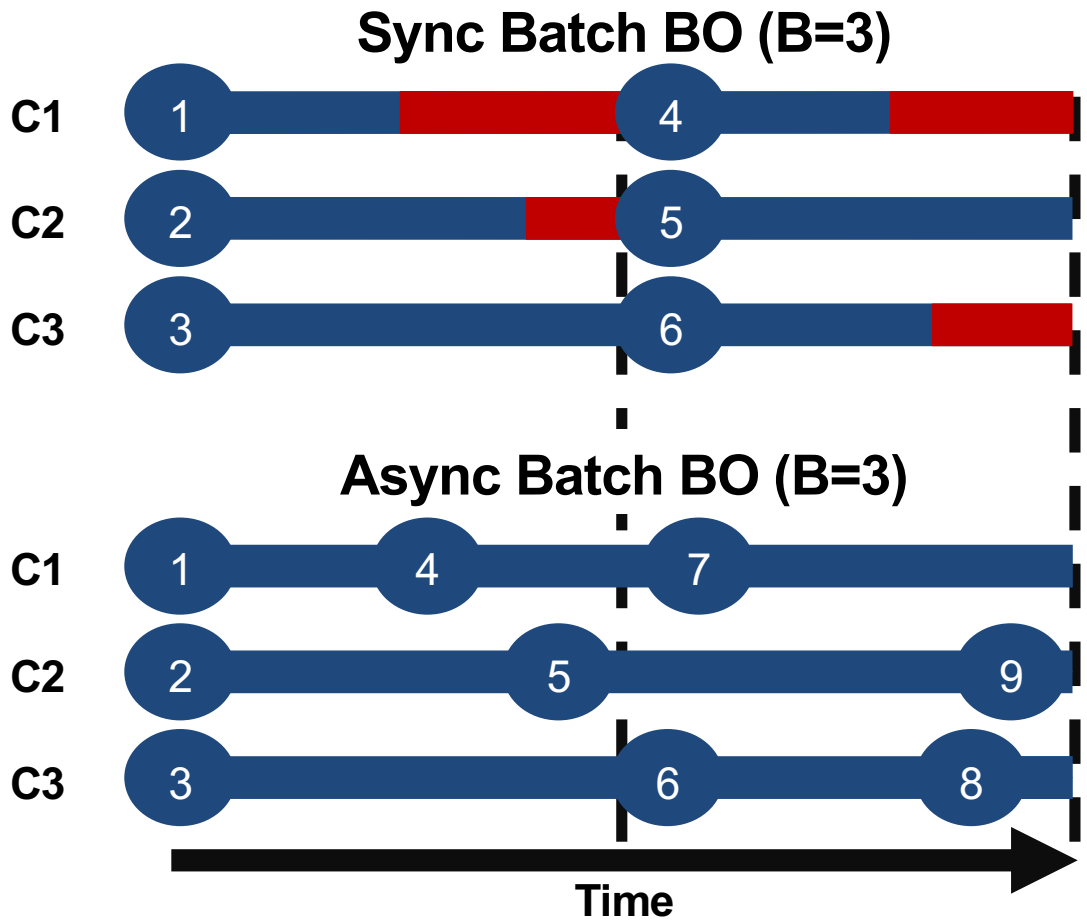
2. Asynchronous Batch BO

- Maximise utilisation of parallel workers



2. Asynchronous Batch BO

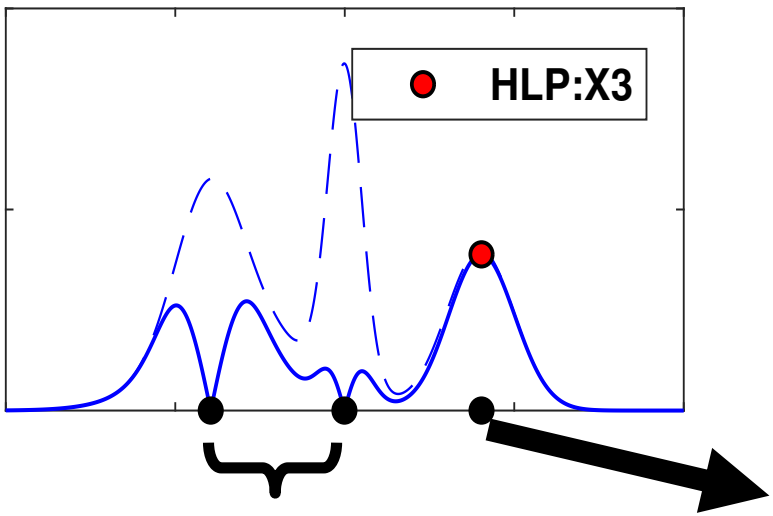
- Maximise utilisation of parallel workers



3. Our Method

- A new async batch BO: ***P**enalising **L**ocally for **A**sy**n**chronous **B**ayesian **O**ptimisation **o**n* **k** *Workers* (**PLAyBOOK**)

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



busy locations

(points under evaluation at busy workers)

new point assigned to the free worker

3. Our Method

- ***P**enalising **L**ocally for **A**s**y**nchronous **B**ayesian **O**ptimisation **o**n* **k** *Workers* (**PLAyBOOK**)

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

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

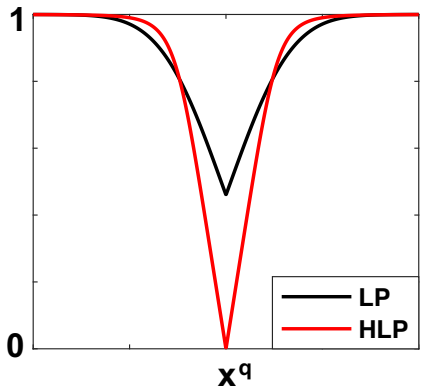
4. Penaliser design

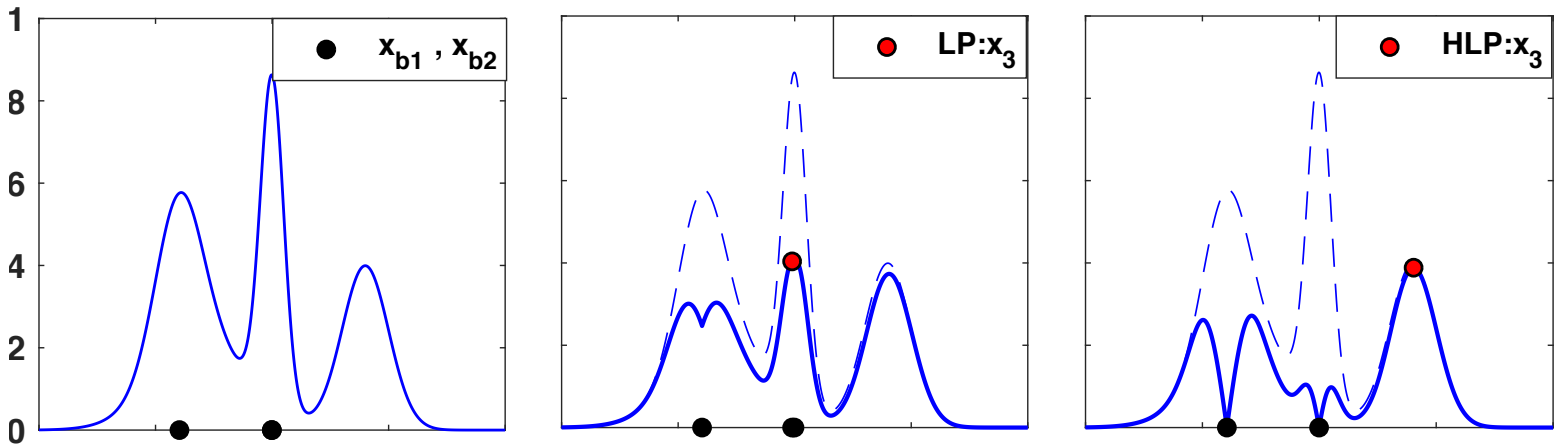
- Our hard penaliser (HLP):

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

- LP (Gonzalez et al., 2016) :

$$\psi_{LP}(\mathbf{x}|\mathbf{x}^q) = \Phi\left(\frac{\hat{L}\|\mathbf{x}-\mathbf{x}^q\|-|\mu(\mathbf{x}^q)-M|}{\sigma(\mathbf{x}^q)}\right)$$



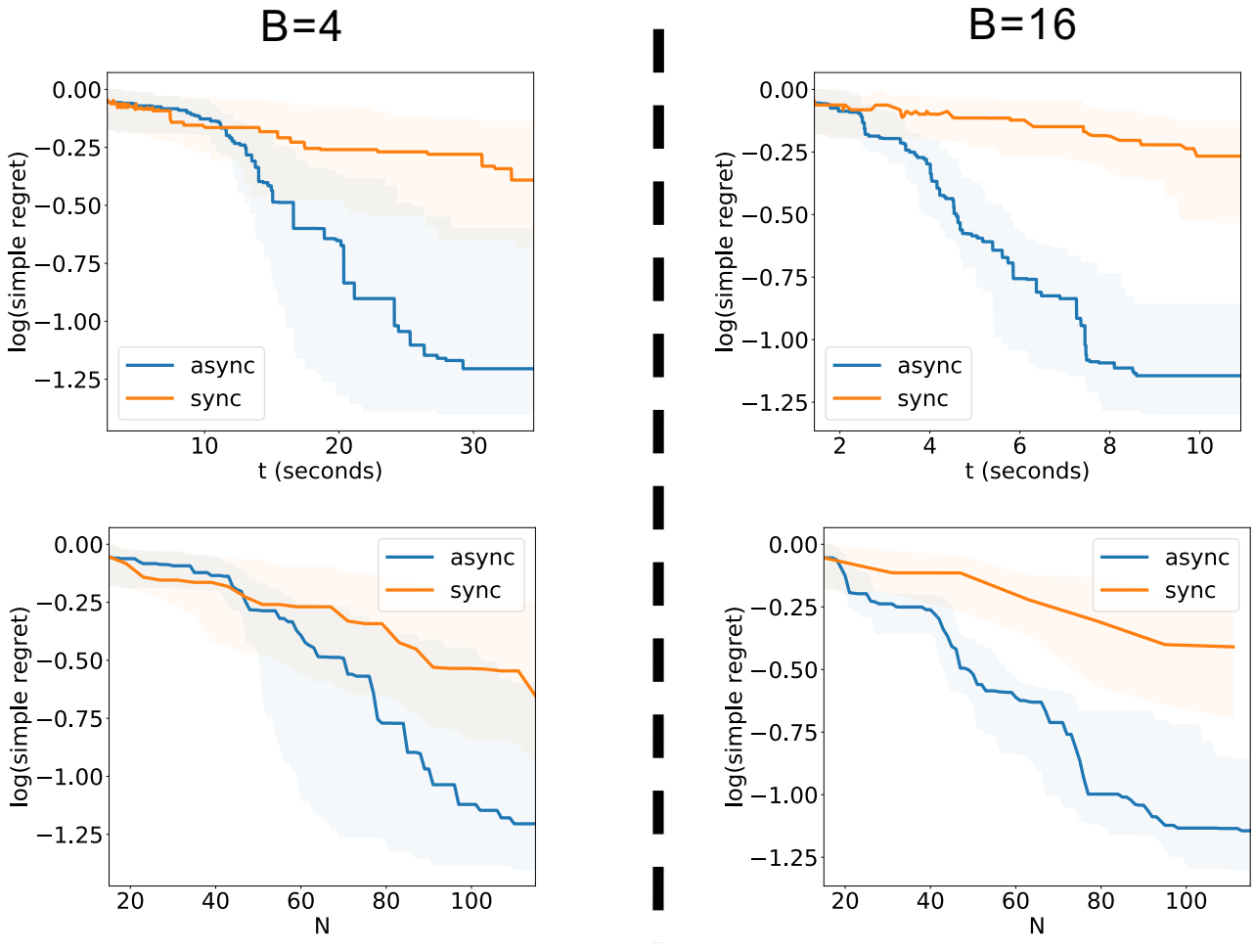


5. Empirical Results: Async. vs. Sync.

- PLAyBOOK-HL: Ackley 5-D: B=4 and B=16

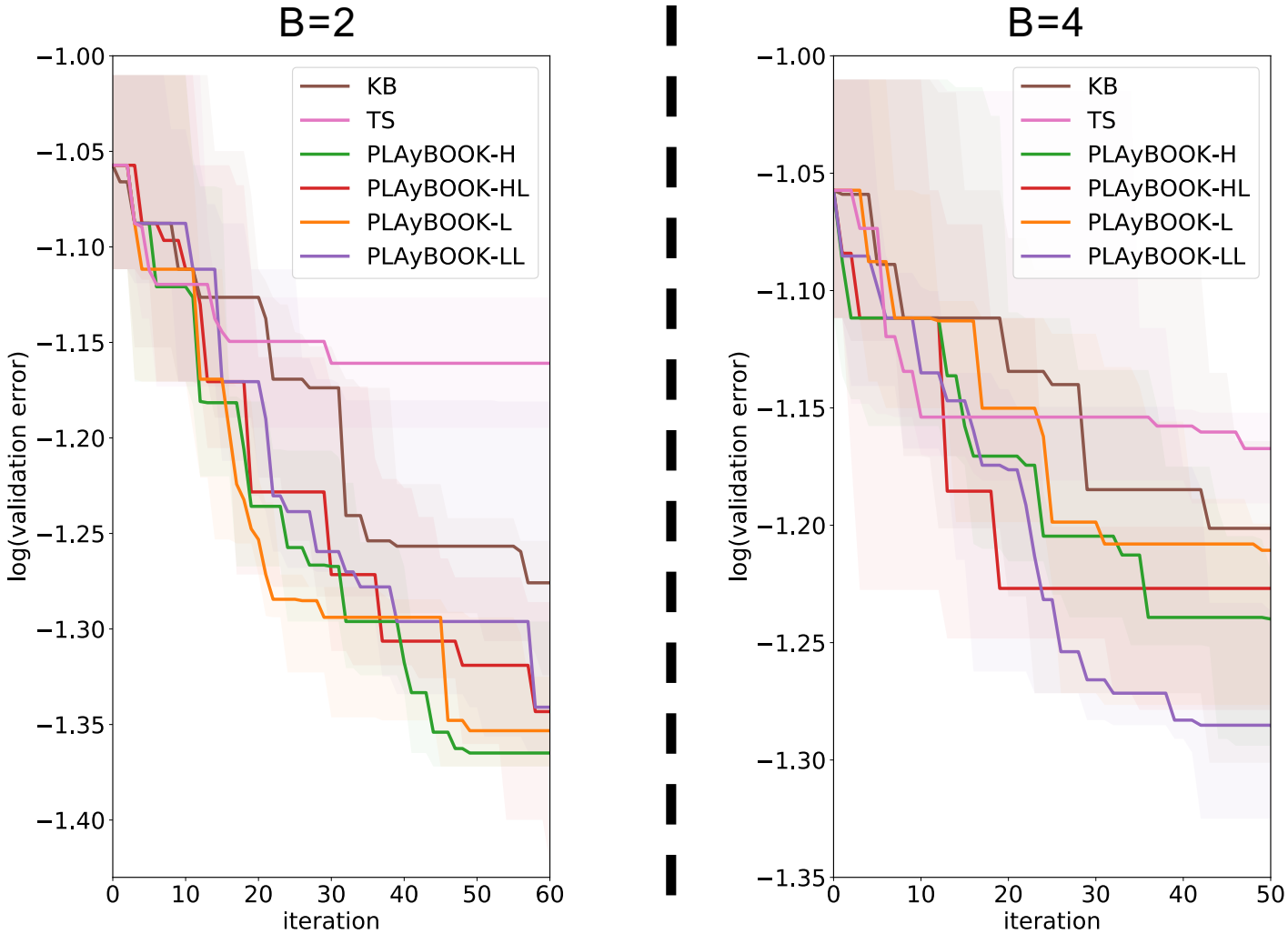
X-axis:
Run Time

X-axis:
No. of Func.
Evaluations



5. Empirical Results: Async. methods

- Tuning 9 hyperparameters of a CNN for CIFAR-10



Thank you!

Meet us at poster #213!