Embedded Bandits for Large-Scale Black-Box Optimization

Abdullah Al-Dujaili, S. Suresh

School of Computer Science & Engineering, Nanyang Technological University

Large-Scale Black-Box Optimization

$$\text{minimize} \quad f(\mathbf{x})$$
$$\text{subject to } \mathbf{x} \in \mathcal{X}$$

- $f : \mathcal{X} = [-1, 1]^n \rightarrow \mathbb{R}$
- $n \gg 10^2$
- $\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) = f(\mathbf{x}^*) = f^*$
- High-order information (e.g., derivatives) are unavailable.

Related Work

- Algorithmic work has been based on either *decomposition* or *embedding* techniques.
- Embedding algorithms exploit the assumption/empirical observation of *low effective dimensionality*
- Recent works presented *Random Embedding* (RE) techniques based on the random matrix theory and provided probabilistic theoretical guarantees [3, 2, 1].
- Multiple runs are employed for RE to substantiate the probabilistic theoretical performance.

Motivation

Breaking away from the *multiple-run* framework and follow the *optimism in the face of uncertainty* principle via *stochastic hierarchical bandits* over a low-dimensional search space $\mathcal{Y}$.

Notation

- $\mathcal{N}$ denotes the Gaussian distribution with zero mean and $1/n$ variance.
- $\{A_p\}_p \subseteq \mathbb{R}^{n \times d}$, with $d \ll n$, is a sequence of realization matrices of the random matrix $\mathbf{A}$ whose entries are sampled independently from $\mathcal{N}$.
- The Euclidean random projection of the ith coordinate $[\mathbf{y}]_i$ to $[\mathcal{X}]_i$ is defined as follows.

$$[\mathcal{P}_{\mathcal{X}}(A\mathbf{y})]_i = \begin{cases} 1, & \text{if } [A\mathbf{y}]_i \geq 1; \\ -1, & \text{if } [A\mathbf{y}]_i \leq -1; \\ [A\mathbf{y}]_i & \text{otherwise.} \end{cases}$$

- $g_P(\mathbf{y})$ is a random (stochastic) function such that $g_P(\mathbf{y}) \stackrel{\text{def}}{=} f(\mathcal{P}_{\mathcal{X}}(\mathbf{Ay}))$ and $g_p(\mathbf{y}) = f(\mathcal{P}_{\mathcal{X}}(A_p\mathbf{y}))$ is a realization (deterministic) function, where $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^d$.

Contribution I

- The mean variation in the objective value for a point y in the low-dimensional space $\mathcal{Y} \subseteq \mathbb{R}^d$ projected randomly into the decision space $\mathcal{X}$ of Lipschitz-continuous problems is *bounded*.
- Mathematically, $\forall y \in \mathcal{Y} \subseteq \mathbb{R}^d$, we have

$$E[|g_p(\mathbf{y}) - g_q(\mathbf{y})|] \leq \sqrt{8} \cdot L \cdot ||\mathbf{y}|| \; .$$

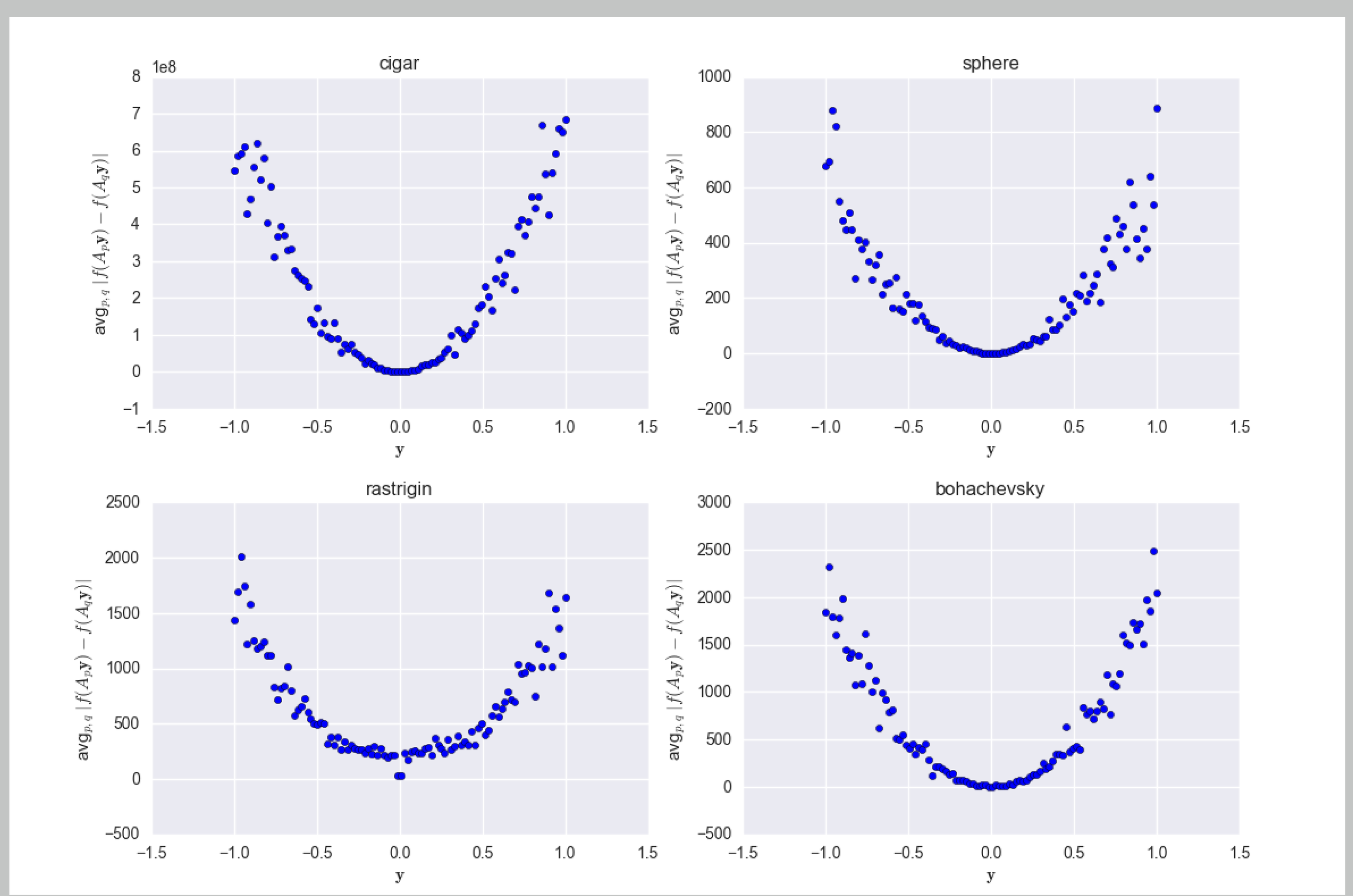


Figure 1: Numerical bound validation

Contribution II

- EMBEDDEDHUNTER is a $\mathcal{Y}$-partitioning tree-search algorithm.
- The partitioning is represented by a K-ary tree $\mathcal{T}$, where nodes of the same depth h correspond to a partition of K^h subspaces / cells.
- For each node (h, i), f is evaluated at the center point $\mathbf{y}_{h,i}$ of its cell $\mathcal{Y}_{h,i}$ once or more times with different projections based on $||\mathbf{y}_{h,i}||$.

Convergence Analysis

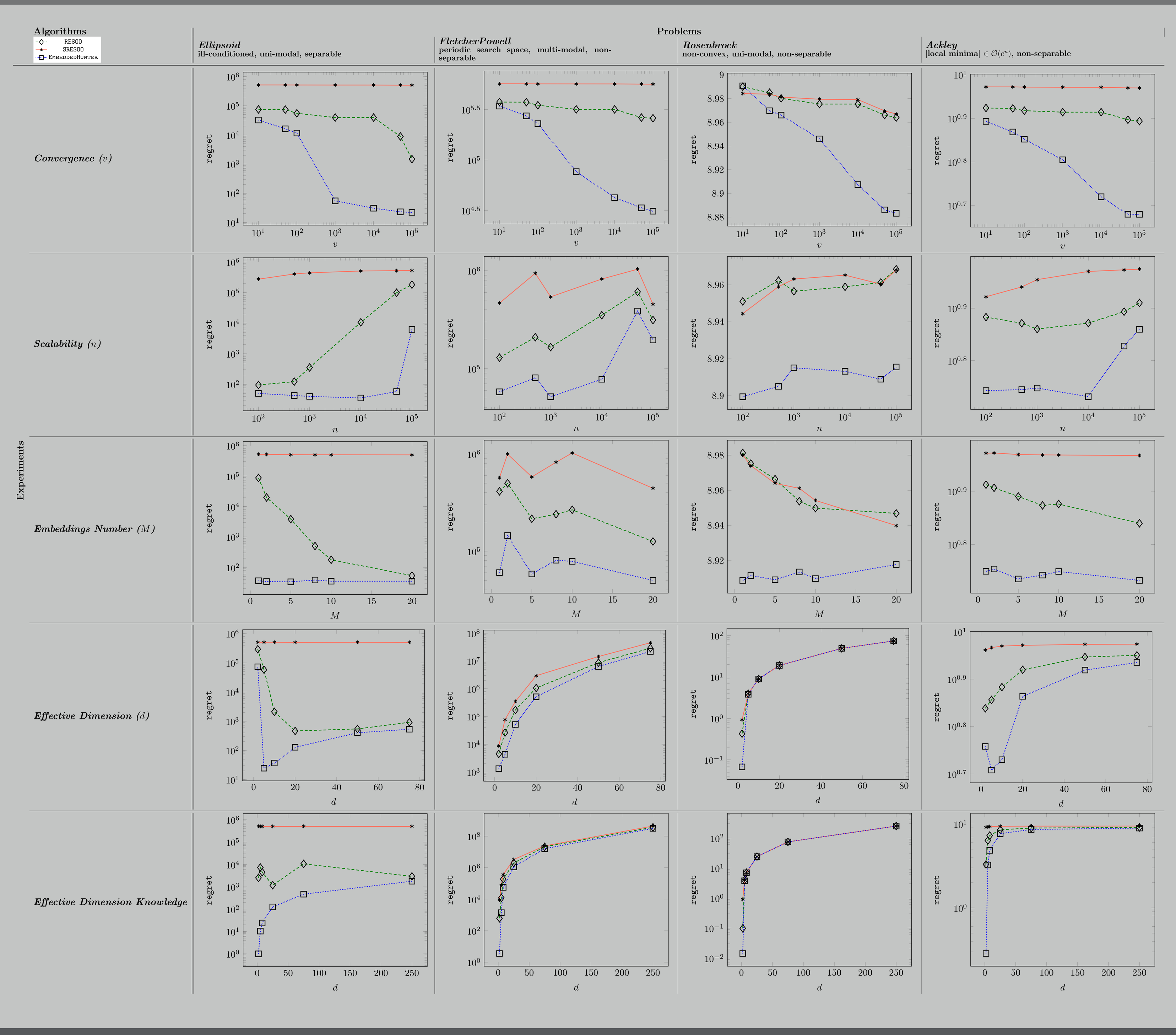
- Define $h(t)$ as the smallest $h \geq 0$ such that:

$$C h_{max} \sum_{l=0}^{h(t)} (\hat{m}\delta(l))^{-\hat{d}} \geq t \; ,$$

where t is the number of iterations. Then EMBEDDEDHUNTER's regret is bounded as

$$r(t) \leq \min_{h \leq \min(h(t), h_{max}+1)} \tau(h) + \delta(h) \; .$$

Performance Evaluation



Conclusion

- EMBEDDEDHUNTER builds a stochastic tree over a low-dimensional search space $\mathcal{Y}$, where stochasticity has shown to be proportional on average with the norm of the nodes' base points.
- Besides its theoretically-proven performance, numerical experiments have validated EMBEDDEDHUNTER's in comparison with recent random-embedding methods.

References

[1] H. Qian, Y.-Q. Hu, and Y. Yu.
Derivative-free optimization of high-dimensional non-convex functions by sequential random embeddings.
In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI'16)*, 2016.

[2] H. Qian and Y. Yu.
Scaling simultaneous optimistic optimization for high-dimensional non-convex functions with low effective dimensions.
In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16)*, 2016.

[3] Z. Wang, M. Zoghi, F. Hutter, D. Matheson, N. Freitas, et al.
Bayesian optimization in high dimensions via random embeddings.
In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13)*, pages 1778–1784. AAAI Press/International Joint Conferences on Artificial Intelligence, 2013.

Acknowledgments

- The research was partially supported by the ST Engineering - NTU Corporate Lab, Singapore, through the NRF corporate lab@universty scheme.

aldujail001@e.ntu.edu.sg