Adaptive Monte Carlo Optimization: Ultra Fast Medoid Identification via Correlated Sequential Halving

Tavor Z. Baharav, David N. Tse
Stanford University

Problem Formulation

$$x_1, \ldots, x_n \in \mathbb{R}^d \qquad i^* = \arg\min_{i\in[n]} \theta_i \qquad \theta_i \triangleq \frac{1}{n}\sum_{j=1}^{n} d(x_i, x_j)$$

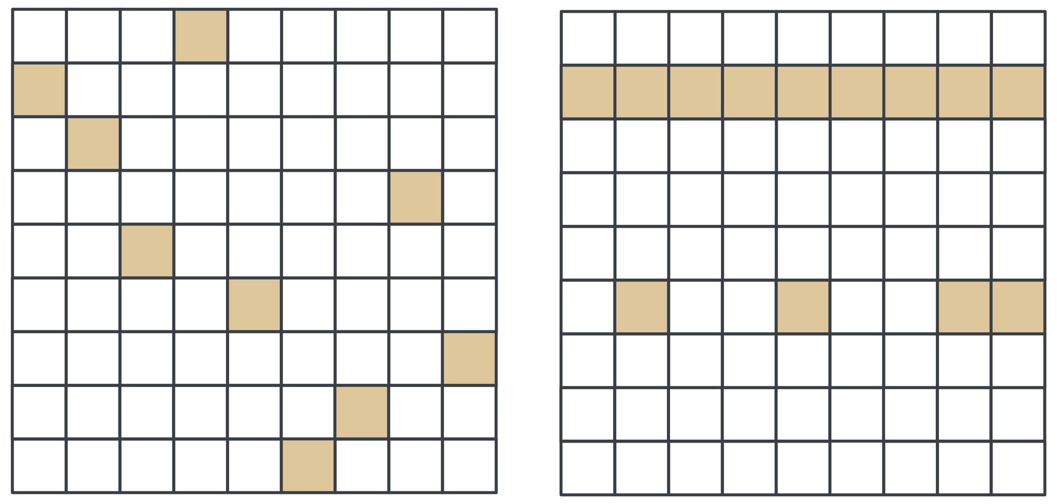
- Generalization of median
- Unlike mean, medoid is inside dataset
- Special instance of AMO framework, also works for k-NN

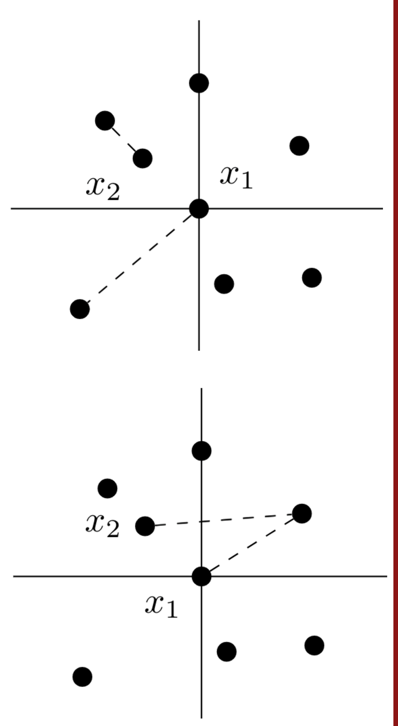
Computation → Estimation

- Estimate θ_i via random sampling
  $\mathbb{E}\{d(x_i, x_J)\} = \theta_i \qquad J \sim \mathrm{Unif}([n])$
- RAND: estimate each θ_i to same degree of accuracy
  $$\hat{\theta}_i = \frac{1}{|\mathcal{J}_i|}\sum_{j\in\mathcal{J}_i} d(x_i, x_j)$$
- Medoid Bandit (Med-dit): sample adaptively (UCB) [1]

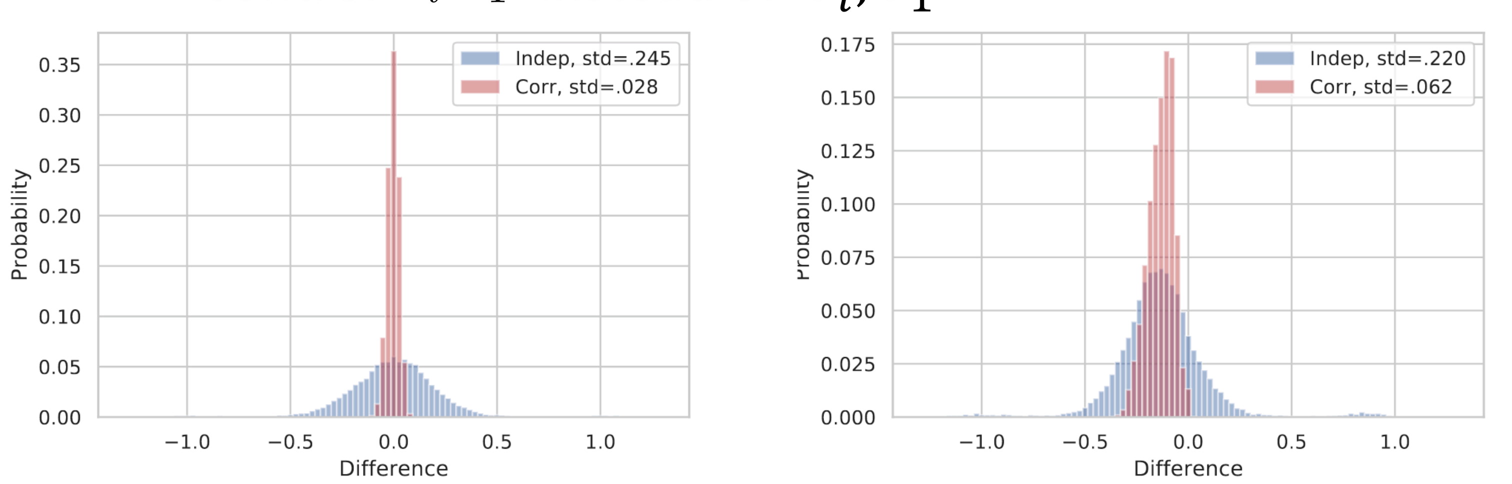
Our contributions [4]

- Naïve bandit reduction to statistical estimation ignores structure of problem

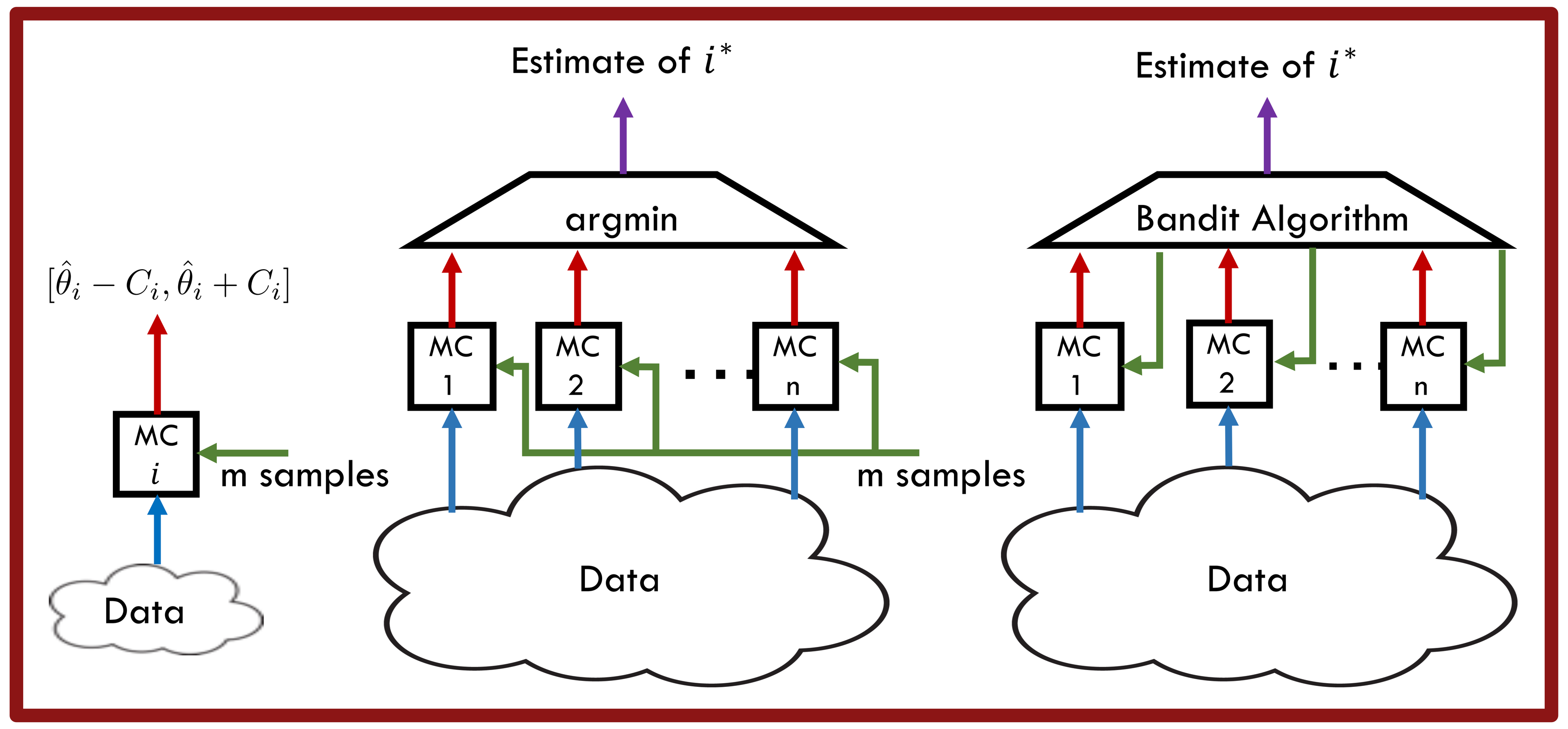




- Can overcome via *correlating* our sampling
  - Sample rows of D, $D_{i,j} = d(x_i, x_j)$
- Need to prove $\widehat{\theta_i} < \widehat{\theta_1}$
  - Control $\widehat{\theta_i\text{-}\theta_1}$ instead of $\widehat{\theta_i}, \widehat{\theta_1}$

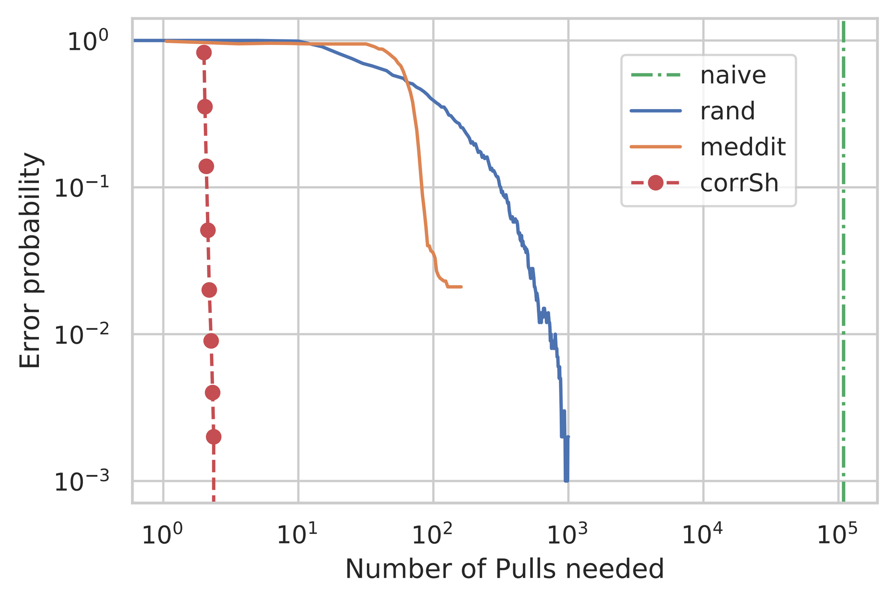


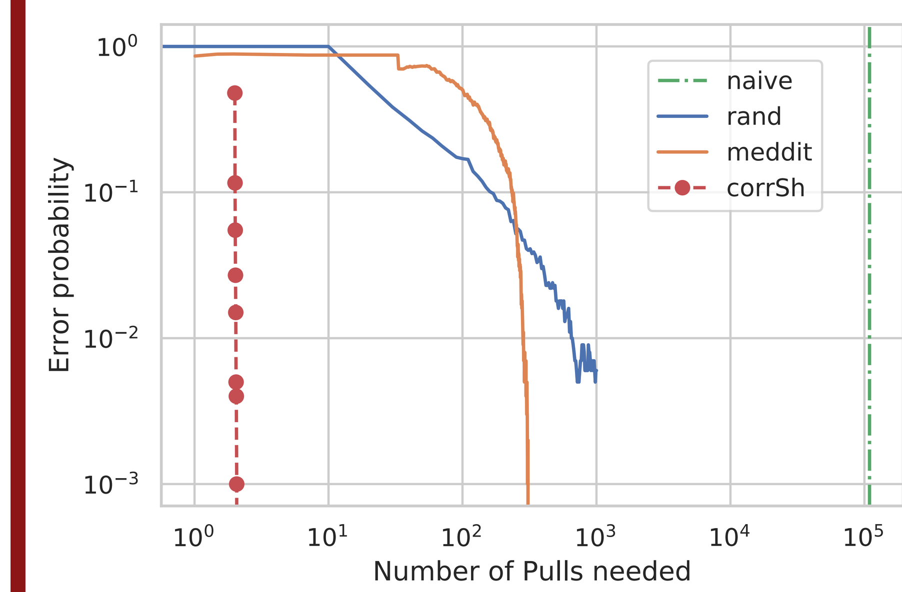
(a) Comparison of top 2 points (i=2) (b) Comparison of top and mid (i=10000)

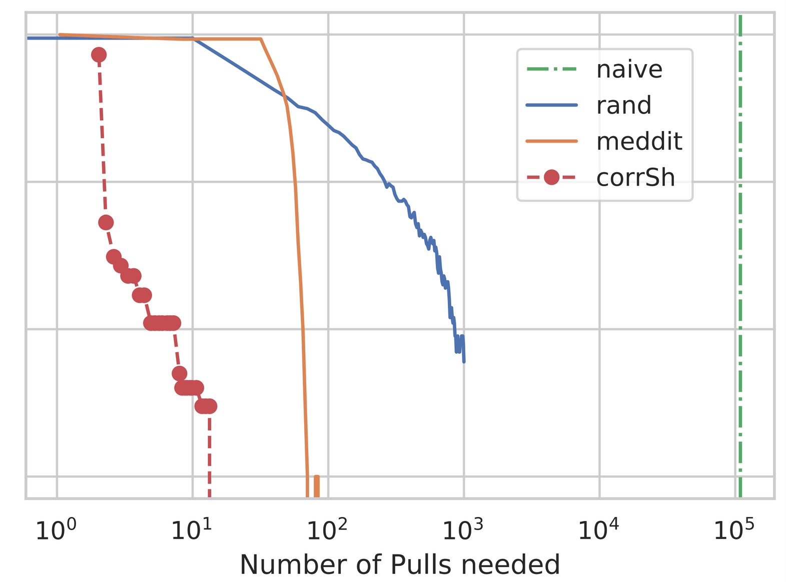


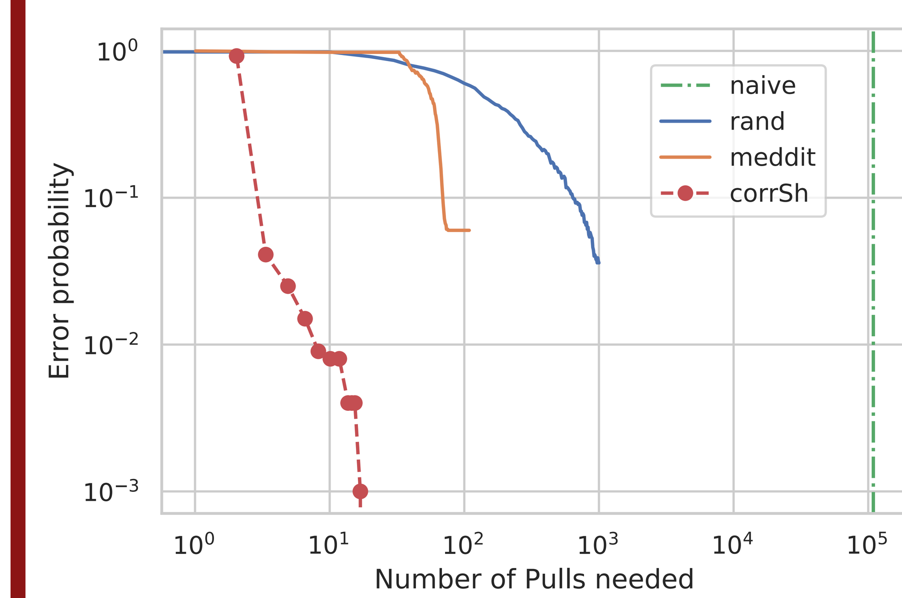
Simulation Results

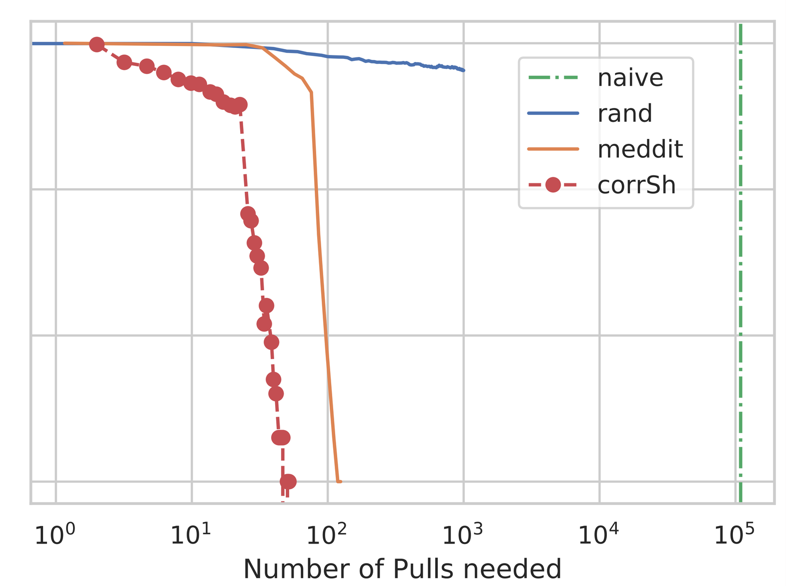
Dataset, Metric	n, d
RNA-Seq 20k, ℓ_1	20k, 28k
RNA-Seq 100k, ℓ_1	109k, 28k
Netflix 20k, cos	20k, 18k
Netflix 100k, cos	100k, 18k
MNIST Zeros, ℓ_2	6424, 784











- Figures arranged top to bottom, left to right, following the table

Theorem Statement

- **Assumption:** $d(x_1, x_J) - d(x_i, x_J)$ is $\sigma\rho_i$–subgaussian
- **Theorem:** corrSH [4] identifies the medoid with probability at least 1-δ after computing
  $$T = O\left(\sigma^2 \log n \log\left(\frac{\log n}{\delta}\right) \max_{i \geq \frac{T}{n\log_2 n}} \left[\frac{i\rho_{(i)}^2}{\Delta_{(i)}^2}\right]\right)$$
  distance evaluations

Another Application: k-NN

$$\theta_i \triangleq \frac{1}{d}\sum_{\ell=1}^{d} |x_i[\ell] - x_0[\ell]|^p \qquad \mathbb{E}\{|x_i[J] - x_0[J]|^p\} = \theta_i$$

- Works for any ℓ_p (separable) distance
- Randomly rotate data for better subgaussian constant
- $O(n^2 \log^2 nd)$ time under distributional assumption
- 100x gain in theory on ImageNet, 25x wall clock speedup

Summary

- Convert computational problem to statistical estimation
- Fast randomized algorithms
- Incorporating structure of the computational problem in this reduction can yield massive gains

Paper

Code

References

[1] V. Bagaria, G. Kamath, V. Ntranos, M. Zhang, and D. Tse, “Almost-linear time via multi-armed bandits,” in Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics, pp. 500–509, 2018.
[2] V. Bagaria, G. M. Kamath, and D. N. Tse, “Adaptive monte-carlo optimization,” arXiv preprint arXiv:1805.08321, 2018.
[3] Z. Karnin, T. Koren, and O. Somekh, “Almost optimal exploration in multi-armed bandits,” in International Conference on Machine Learning, pp. 1238–1246, 2013.
[4] Baharav, Tavor Z., and David N. Tse. "Ultra Fast Medoid Identification via Correlated Sequential Halving." accepted to NeurIPS 2019.