Anytime Model Selection in Linear Bandits

Parnian Kassraie, Aldo Pacchiano, Nicolas Emmenegger, Andreas Krause

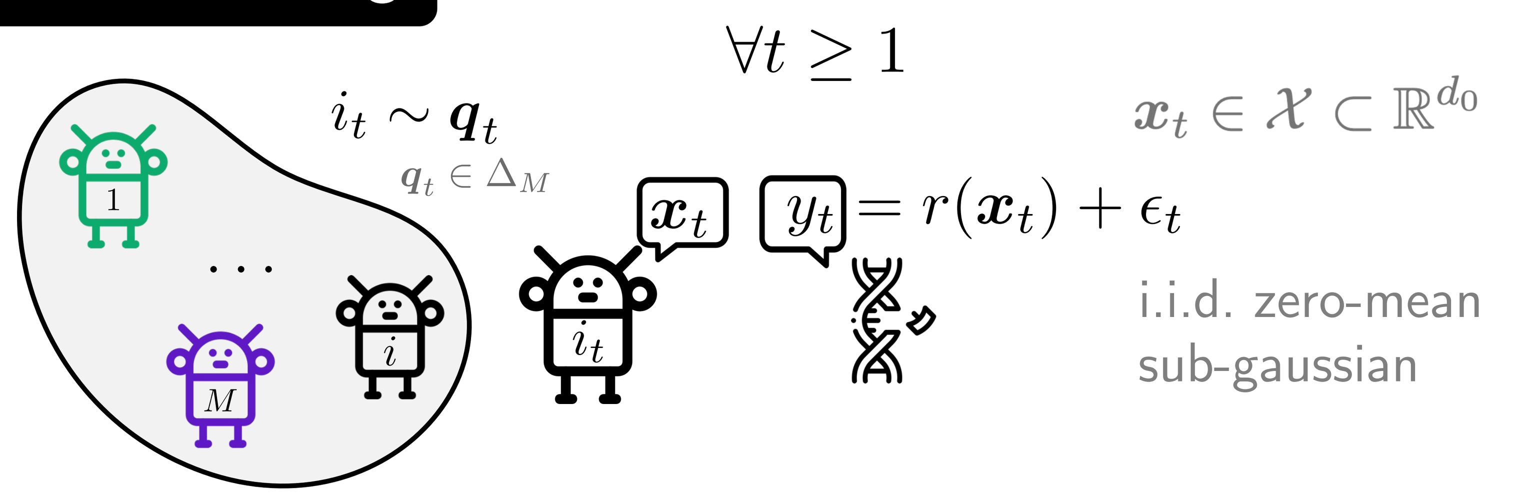
ETH zürich

Can we perform adaptive model selection, while simultaneously optimizing for a reward? Can we be sample-efficient & anytime?

- Solving a Bandit/BO problem:
  1. Commit to a reward model (a priori)
  2. Interact with the environment accordingly
- There are many ways to model the reward

  $M \gg n$ n: horizon/stopping time
- Not known a priori which agent is going to be the best (e.g. in terms of sample efficiency, or regret)
- We can select the model based on emprical evidence.

Problem Setting



- Model Class

  $\{\boldsymbol{\phi}_j : \mathbb{R}^{d_0} \to \mathbb{R}^d, j = 1, \dots, M\}$ + typical regularity assumptions

  $\exists\, j^\star \in [M]$ s.t. $r(\cdot) = \boldsymbol{\theta}_{j^\star}^\top \boldsymbol{\phi}_{j^\star}(\cdot)$
- Agents

  Agent j only uses $\boldsymbol{\phi}_j$ to model the reward

  Update its action selection policy $p_{t,j} \in \mathcal{M}(\mathcal{X})$

  Using the full history $H_{t-1} = \{(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_{t-1}, y_{t-1})\}$
- Goal $R(n) = \sum_{t=1}^{n} r(\boldsymbol{x}^*) - r(\boldsymbol{x}_t)$ n unknown

Ingredient I: Exponential Weights Updates

Adjust the probability of selecting each agent wrt the reward it has obtained so far

- known to yield $\log M$ regret in full-info setting

  η_t sensitivity of updates

$$q_{t,j} = \frac{\exp(\eta_t \sum_{s=1}^{t-1} \hat{r}_{s,j})}{\sum_{i=1}^{M} \exp(\eta_t \sum_{s=1}^{t-1} \hat{r}_{s,i})}$$

- we don't observe the reward of every agent, so:

$$\hat{r}_{t,j} = \mathbb{E}_{\boldsymbol{x} \sim p_{t,j}} \hat{\boldsymbol{\theta}}_t^\top \boldsymbol{\phi}(\boldsymbol{x})$$

- Regret will depend on bias and variance of $\hat{\boldsymbol{\theta}}_t$
- Typical online regression oracles are $\sqrt{M}$ $\to$ poly$M$ regret

Ingredient II: Sparse Online Regression Oracle

Turn lasso into a sparse online regression oracle

$$\hat{\boldsymbol{\theta}}_t = \arg\min \frac{1}{t} \|\boldsymbol{y}_t - \Phi_t \boldsymbol{\theta}\|_2^2 + \lambda_t \sum_{j=1}^{M} \|\boldsymbol{\theta}_j\|_2$$

Bias and variance are both $\log M$

Theorem (Anytime Conf. Seq.)

If for all $t \geq 1$

cost of going 'time uniform'

$$\lambda_t \geq \frac{c_1}{\sqrt{t}} \sqrt{\log(M/\delta) + \sqrt{d\,(\log(M/\delta) + \boxed{\log\log d})} + 1}$$

then, c_1 and c_2 made exact in the paper

$$\mathbb{P}\left(\forall t \geq 1 : \left\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_t\right\|_2 \leq \frac{c_2 \lambda_t}{\kappa^2(\Phi_t, 2)}\right) \geq 1 - \delta$$

Restricted Eigenvalue property [check paper]

Simultaneous Model Selection and Optimization

- Putting it together we get Anytime Exponential weighting algorithm with Lasso reward estimates (ALEXP)

Algorithm 1 ALEXP

Inputs: γ_t, η_t, λ_t for $t \geq 1$

for $t \geq 1$ **do** mix with exploration

  Draw $\boldsymbol{x}_t \sim (1 - \gamma_t) \sum_{j=1}^{M} q_{t,j} p_{t,j} + \gamma_t \mathrm{Unif}(\mathcal{X})$

  Observe $y_t = r(\boldsymbol{x}_t) + \epsilon_t$.

  Append history $H_t = H_{t-1} \cup \{(\boldsymbol{x}_t, y_t)\}$.

  Update agents $p_{t,j}$ for $j = 1, \dots, M$.

  Calculate $\hat{\boldsymbol{\theta}}_t \leftarrow$ Lasso(H_t, λ_t) and estimate

$$\hat{r}_{t,j} \leftarrow \mathbb{E}_{\boldsymbol{x} \sim p_{t+1,j}}[\hat{\boldsymbol{\theta}}_t^\top \phi(\boldsymbol{x})]$$

  Update selection distribution

$$q_{t+1,j} \leftarrow \frac{\exp(\eta_t \sum_{s=1}^{t} \hat{r}_{s,j})}{\sum_{i=1}^{M} \exp(\eta_t \sum_{s=1}^{t} \hat{r}_{s,i})}$$

Theorem (Regret - Informal)

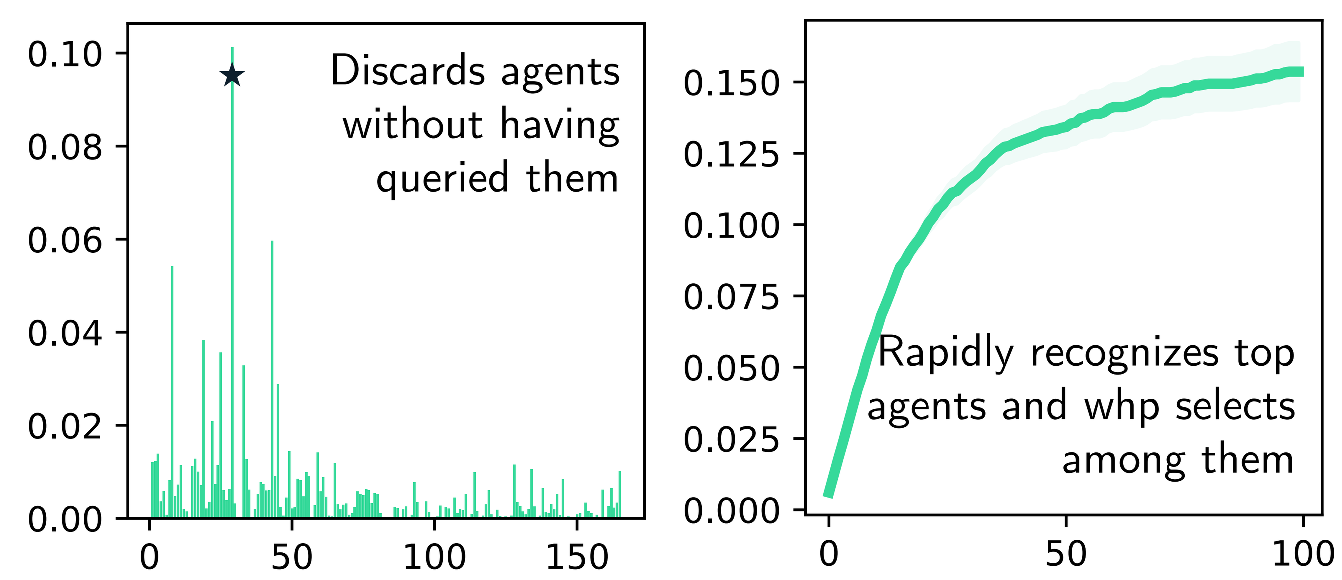
For appropriate choices of $(\gamma_t, \lambda_t, \eta_t)$, prescribed in the paper

$$R(n) = \mathcal{O}\left(C(M, \delta, d)\left(\sqrt{n}\log M + n^{3/4}\right)\right)$$

with probability greater than $1 - \delta$, *simultaneously for all* $n \geq 1$.

$C(M, \delta, d) = \mathcal{O}\left(\sqrt{d \log M/\delta} + \sqrt{d \log M/\delta}\right)$

Model Selection Dynamics



Comparison to prior work

	technique	$\log M$ regret	MS guarantee	adaptive & anytime
Sparse Linear Bandits	Lasso	✓	✗	✗
MS for Black-Box Bandits	OMD with bandit info	✗	✓	✗
MS for Linear Bandits (Ours)	EXP4 with full info	✓	✓	✓