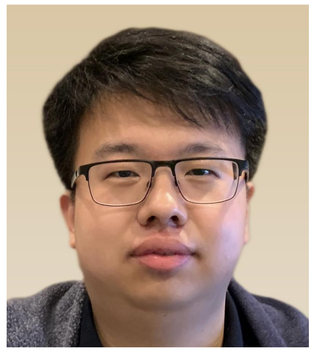
Personal website

Beyond task diversity: provable representation transfer for sequential multi-task linear bandits

Thang Duong
thangduong@arizona.edu



Dr. Zhi Wang

Prof. Chicheng Zhang

Problem: sequential multitask linear bandits

Protocol:

For task $n = 1, 2, \dots, N$

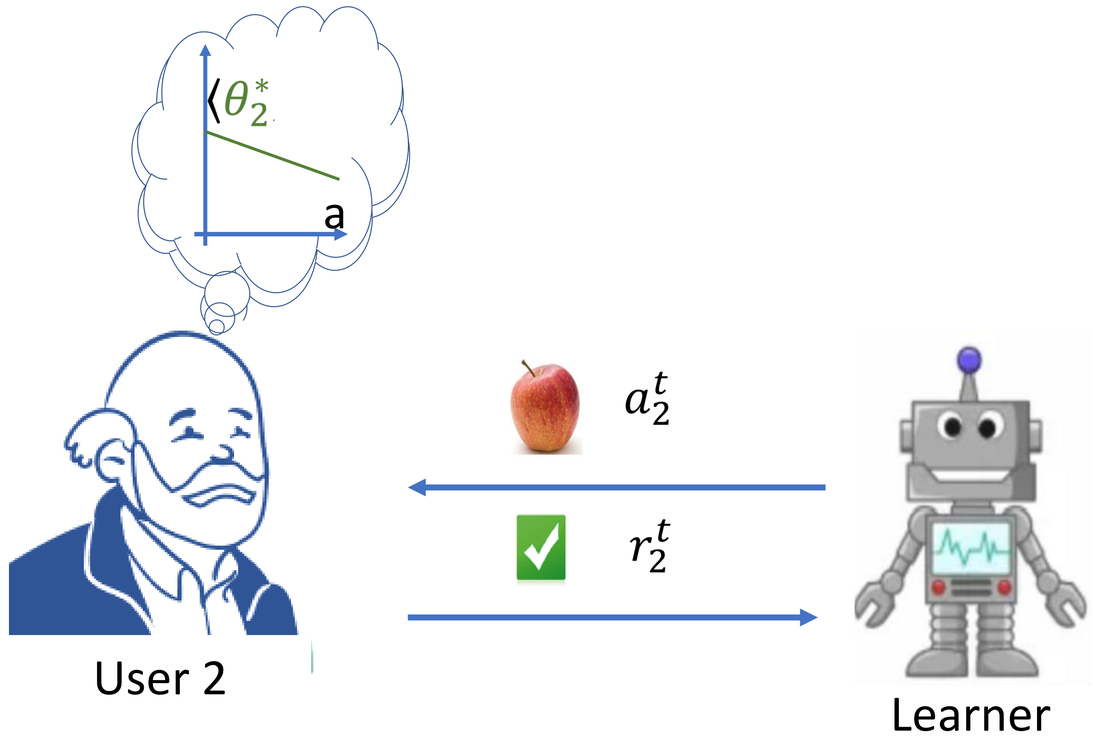
For time step $t = 1,2, \dots, \tau$:

Take action a_n^t from the action space $\mathcal{A}$

Receive reward $r_t^n = \langle \theta_n^*, a_n^t \rangle + \eta_n^t$

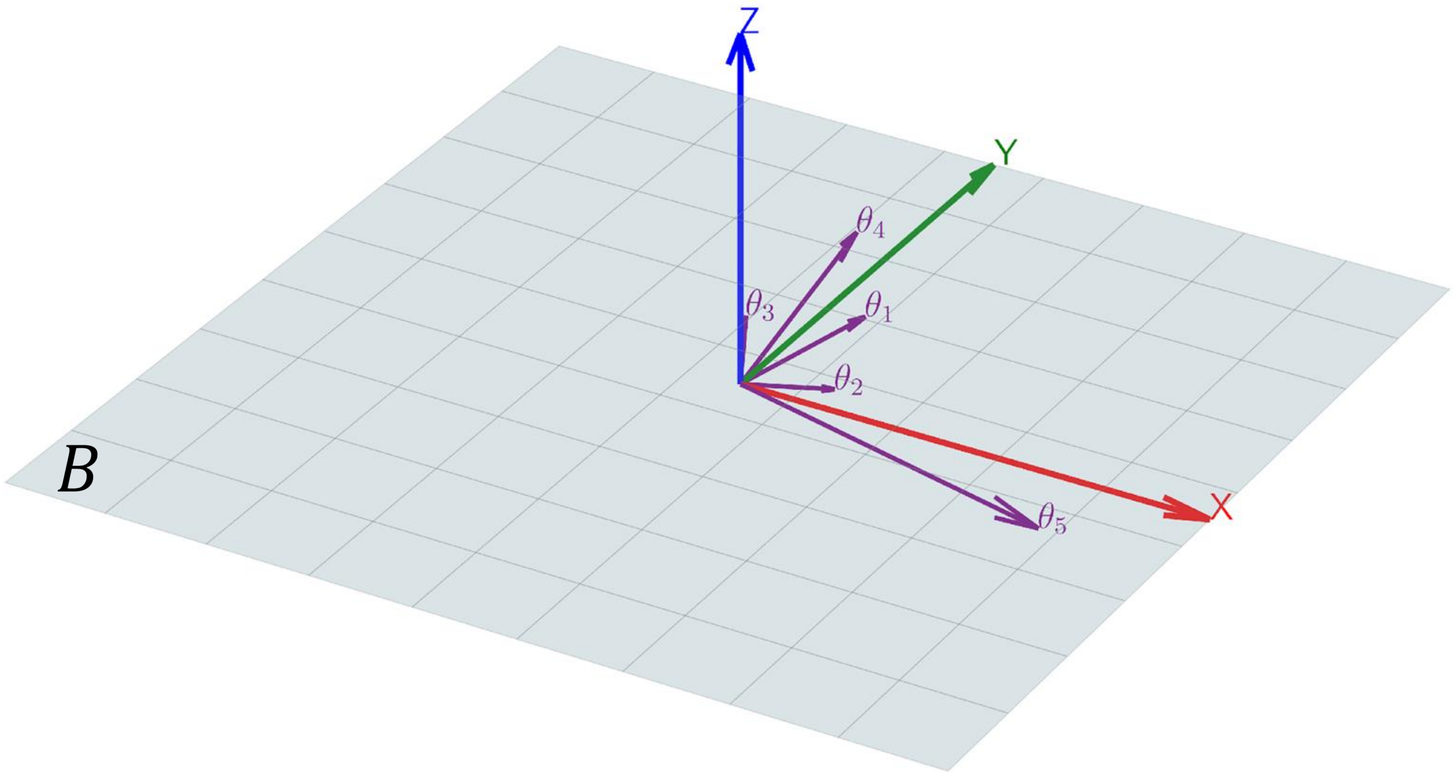
- Goal: minimize cumulative regret

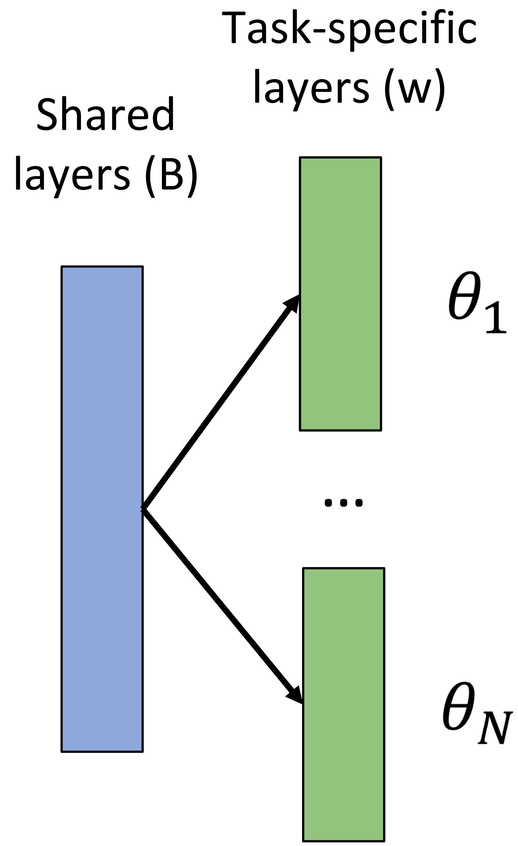
$$\sum_{n=1}^{N} \sum_{t=1}^{\tau} \max_{a_n \in \mathcal{A}} \langle \theta_n^*, a_n \rangle - \langle \theta_n^*, a_n^t \rangle$$



Problem: sequential multitask linear bandits

Shared structure: $(\theta_1^*, \ldots, \theta_N^*)$ lie on a subspace B of dimension m



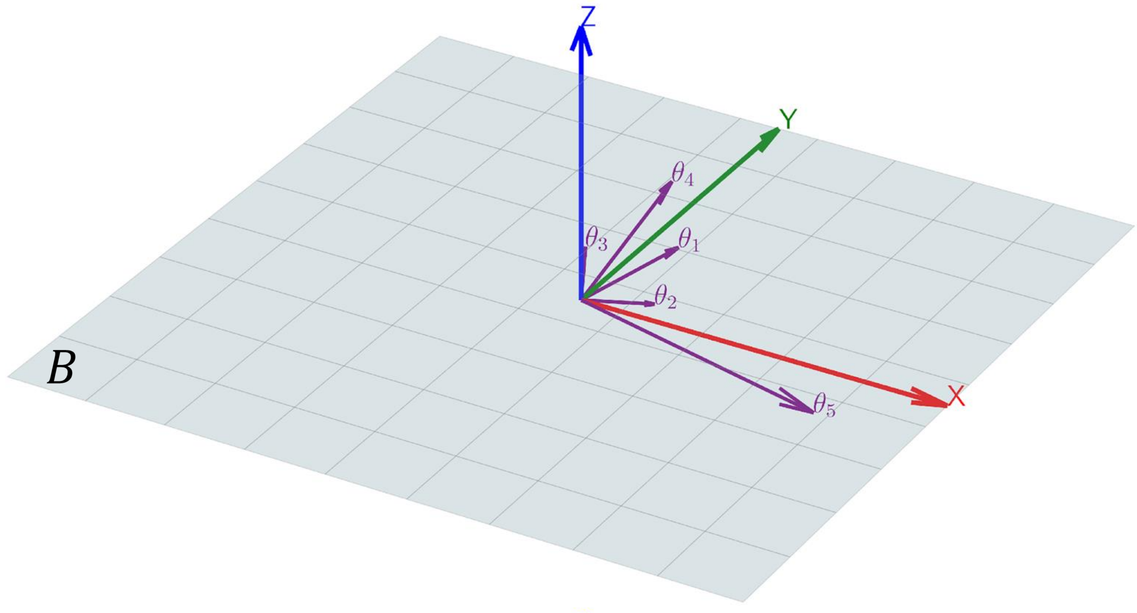


- Solving each task independently with PEGE[1]: $Nd\sqrt{\tau}$
- If B was known: $Nd\sqrt{\tau} \rightarrow Nm\sqrt{\tau}$

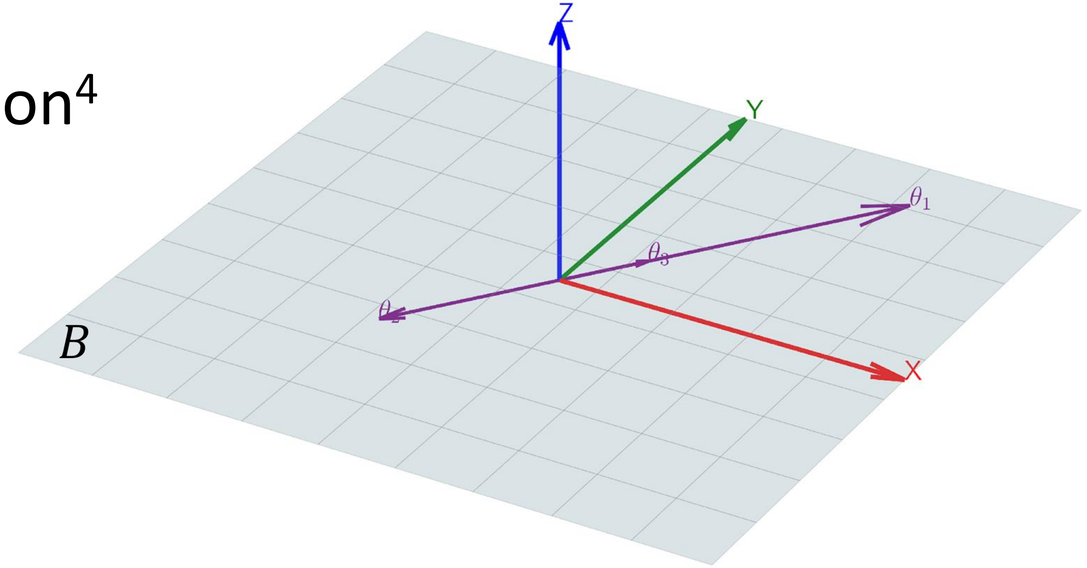
1: Paat Rusmevichientong and John N Tsitsiklis. Linearly parameterized bandits. Mathematics of Operations Research, 35(2):395–411, 2010.

Problem: sequential multitask linear bandits

Parallel setting[2] or Task Diversity assumption[3]

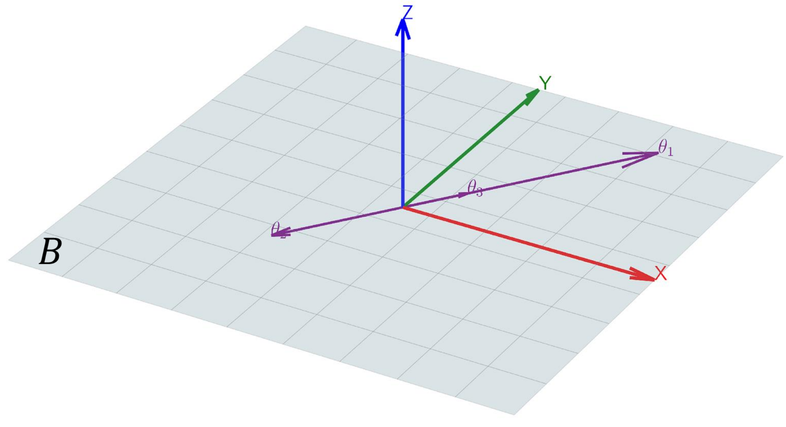


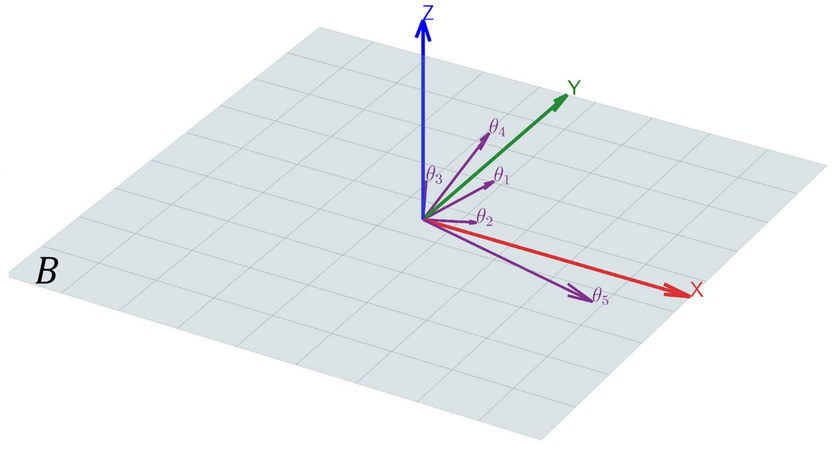
Sequential setting without Task Diversity assumption[4]
=> the environment can act adversarial by hiding some subspace's dimensions



2: Jiaqi Yang, Wei Hu, Jason D Lee, and Simon Shaolei Du. Impact of representation learning in linear bandits. In International Conference on Learning Representations, 2020.
3: Yuzhen Qin, Tommaso Menara, Samet Oymak, ShiNung Ching, and Fabio Pasqualetti. Non-stationary representation learning in sequential linear bandits. IEEE Open Journal of Control Systems, 1: 41–56, 2022.
4: Javad Azizi, Thang Duong, Yasin Abbasi-Yadkori, András György, Claire Vernade, and Mohammad Ghavamzadeh. Non-stationary bandits and meta-learning with a small set of optimal arms. Reinforcement Learning Journal, 5:2461–2491, 2024.

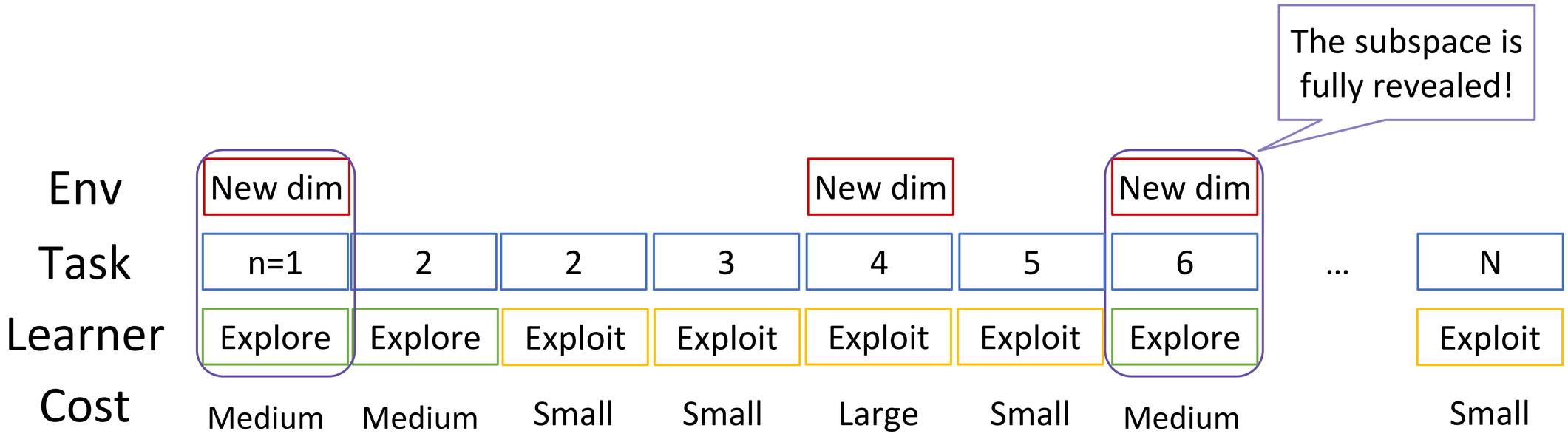
Approach





- Bilevel optimization:
  - High-level: Meta-exploration for learning B vs. meta-exploitation
  - Low-level: Using a variant of PEGE with the learned subspace B

- Key idea: Maintaining the uncertainty estimation over the learned subspace to prevent adversarial environments



Regret guarantee

Algorithm	Task Diversity	Regret
Oracle	No	$\tilde{O}(Nm\sqrt{\tau})$
Independent PEGE for each task	No	$\tilde{O}(Nd\sqrt{\tau})$
Qin et al. [2022]	Yes	$\tilde{O}(Nm\sqrt{\tau} + \sqrt{N\tau}dm)$
BOSS* (Ours)	**No**	$\tilde{O}\left(Nm\sqrt{\tau} + N^{\frac{2}{3}}\tau^{\frac{2}{3}}dm^{\frac{1}{3}}\right)$

*Informal guarantee

Summary

- We present the first nontrivial result for bandit sequential representation transfer without task diversity $\tilde{O}\left(Nm\sqrt{\tau} + N^{\frac{2}{3}}\tau^{\frac{2}{3}}dm^{\frac{1}{3}}\right)$
- Significantly improve upon the baseline of $\tilde{O}\left(Nd\sqrt{\tau}\right)$

Thank you