Understanding the Role of Feedback in Online Learning with Switching Costs

Duo Cheng[1], Xingyu Zhou[2], Bo Ji[1]
[1]Department of Computer Science, Virginia Tech
[2]Department of Electrical and Computer Engineering, Wayne State University

Online Learning over a Finite Set of Actions

Online Learning over a Finite Set of Actions

T-round repeated game between a learner and an adversary

Online Learning over a Finite Set of Actions

T-round repeated game between a learner and an adversary

For round $t = 1, \ldots, T$:

Online Learning over a Finite Set of Actions

T-round repeated game between a learner and an adversary

For round $t = 1, \ldots, T$:

1. The learner chooses (or plays) one of the K actions, denoted by X_t

Online Learning over a Finite Set of Actions

T-round repeated game between a learner and an adversary

For round $t = 1, \ldots, T$:

1. The learner chooses (or plays) one of the K actions, denoted by X_t
2. The learner suffers the loss of the chosen action, which is determined by the (oblivious) adversary

Online Learning over a Finite Set of Actions

T-round repeated game between a learner and an adversary

For round $t = 1, \ldots, T$:

1. The learner chooses (or plays) one of the K actions, denoted by X_t
2. The learner suffers the loss of the chosen action, which is determined by the (oblivious) adversary
3. The learner receives some feedback associated with the losses at this round

Online Learning over a Finite Set of Actions

T-round repeated game between a learner and an adversary

For round $t = 1,\ldots,T$:

1. The learner chooses (or plays) one of the K actions, denoted by X_t
2. The learner suffers the loss of the chosen action, which is determined by the (oblivious) adversary
3. The learner receives some feedback associated with the losses at this round
4. The learner uses the feedback to update her policy

Two Typical Types of Feedback

Two Typical Types of Feedback

Full-information feedback: The losses of all actions

Two Typical Types of Feedback

Full-information feedback: The losses of all actions

Bandit feedback: The loss of the chosen action only

Two Typical Types of Feedback

Full-information feedback: The losses of all actions

Bandit feedback: The loss of the chosen action only

Minimax regret

- Full-information feedback: $\Theta(\sqrt{T \ln K})$ [Cesa-Bianchi & Lugosi, 06]

Two Typical Types of Feedback

Full-information feedback: The losses of all actions

Bandit feedback: The loss of the chosen action only

Minimax regret

- Full-information feedback: $\Theta(\sqrt{T \ln K})$ [Cesa-Bianchi & Lugosi, 06]
- Bandit feedback: $\Theta(\sqrt{TK})$ [Auer et al., 95] [Audibert & Bubeck, 09]

Two Typical Types of Feedback

Full-information feedback: The losses of all actions

Bandit feedback: The loss of the chosen action only

Minimax regret

- Full-information feedback: $\Theta(\sqrt{T \ln K})$ [Cesa-Bianchi & Lugosi, 06]
- Bandit feedback: $\Theta(\sqrt{TK})$ [Auer et al., 95] [Audibert & Bubeck, 09]

$\Theta(T^{1/2})$ vs. $\Theta(T^{1/2})$

Online Learning with Switching Costs

For round $t = 1, \ldots, T$:

1. The learner chooses (or plays) one of the K actions, denoted by X_t

2. The learner suffers the loss of the chosen action, which is determined by the (oblivious) adversary; **The learner additionally suffers one unit of loss (i.e., switching cost) if $X_t \neq X_{t-1}$**

3. The learner receives some feedback associated with the losses at this round

4. The learner uses the feedback to update her policy

A Strong Separation with Switching Costs

A Strong Separation with Switching Costs

Minimax regret (with switching costs)

- Full-information feedback: $\Theta(\sqrt{T \ln K})$ [Geulen et al., 10] [Devroye et al., 10]

A Strong Separation with Switching Costs

Minimax regret (with switching costs)

- Full-information feedback: $\Theta(\sqrt{T \ln K})$ [Geulen et al., 10] [Devroye et al., 10]
- Bandit feedback: $\tilde{\Theta}(T^{2/3}K^{1/3})$ [Arora et al., 12] [Dekel et al., 13]

A Strong Separation with Switching Costs

Minimax regret (with switching costs)

- Full-information feedback: $\Theta(\sqrt{T \ln K})$ [Geulen et al., 10] [Devroye et al., 10]
- Bandit feedback: $\tilde{\Theta}(T^{2/3}K^{1/3})$ [Arora et al., 12] [Dekel et al., 13]

$$\tilde{\Theta}(T^{2/3}) \text{ vs. } \Theta(T^{1/2})$$

A Strong Separation with Switching Costs

Minimax regret (with switching costs)

- Full-information feedback: $\Theta(\sqrt{T \ln K})$ [Geulen et al., 10] [Devroye et al., 10]
- Bandit feedback: $\tilde{\Theta}(T^{2/3}K^{1/3})$ [Arora et al., 12] [Dekel et al., 13]

Feedback	Bandit		Full-information
Minimax Regret	$\tilde{\Theta}(T^{2/3})$	?	$\Theta(T^{1/2})$

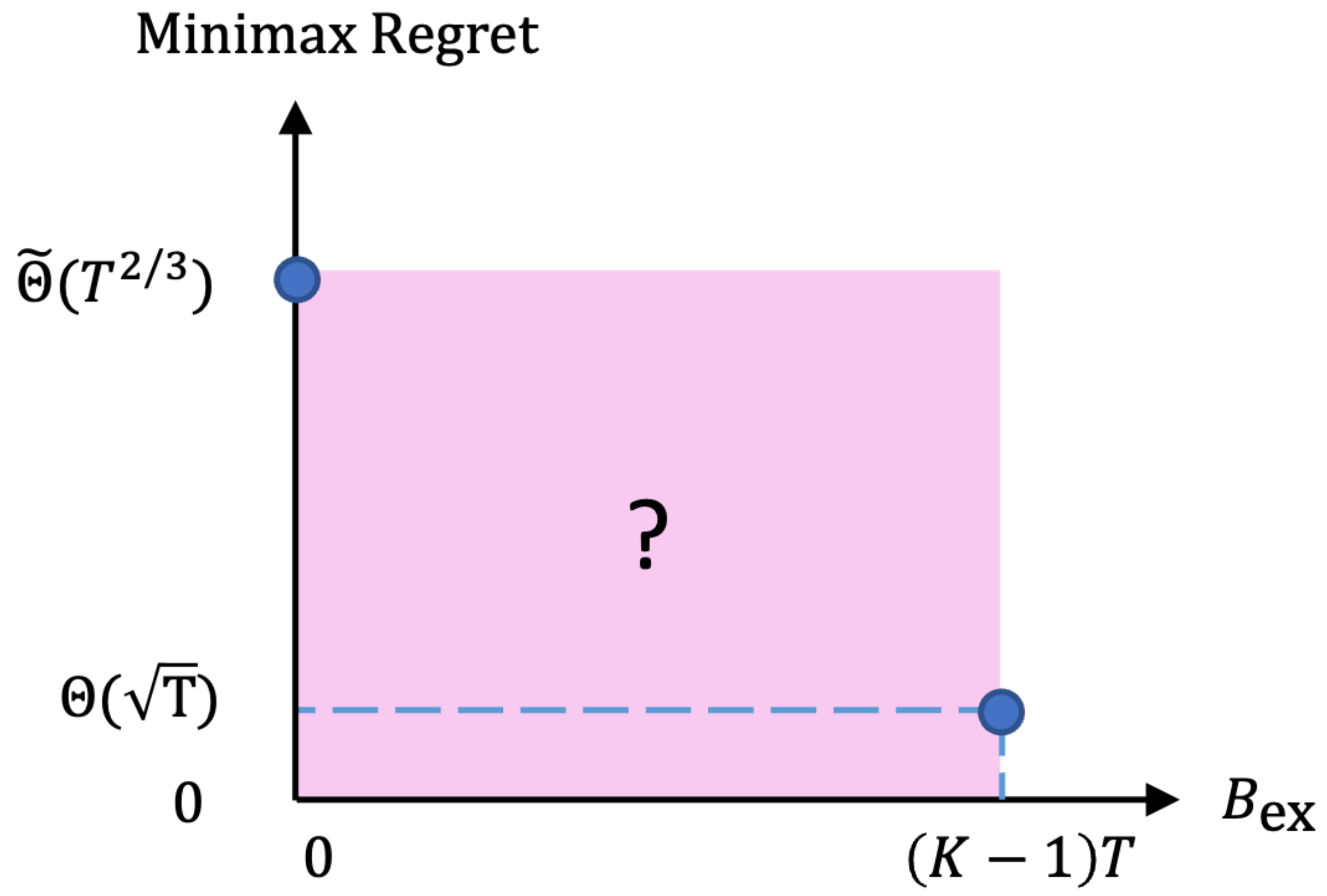
Learning with Bandit Feedback under Extra Observation Budget

Feedback: Bandit + B_{ex} losses in total

Learning with Bandit Feedback under Extra Observation Budget

Feedback: Bandit + B_{ex} losses in total

Extra Observations	$B_{\text{ex}} = 0$ (Bandit)		$B_{\text{ex}} = (K-1)T$ (Full-information)
Minimax Regret	$\tilde{\Theta}(T^{2/3})$	?	$\Theta(T^{1/2})$

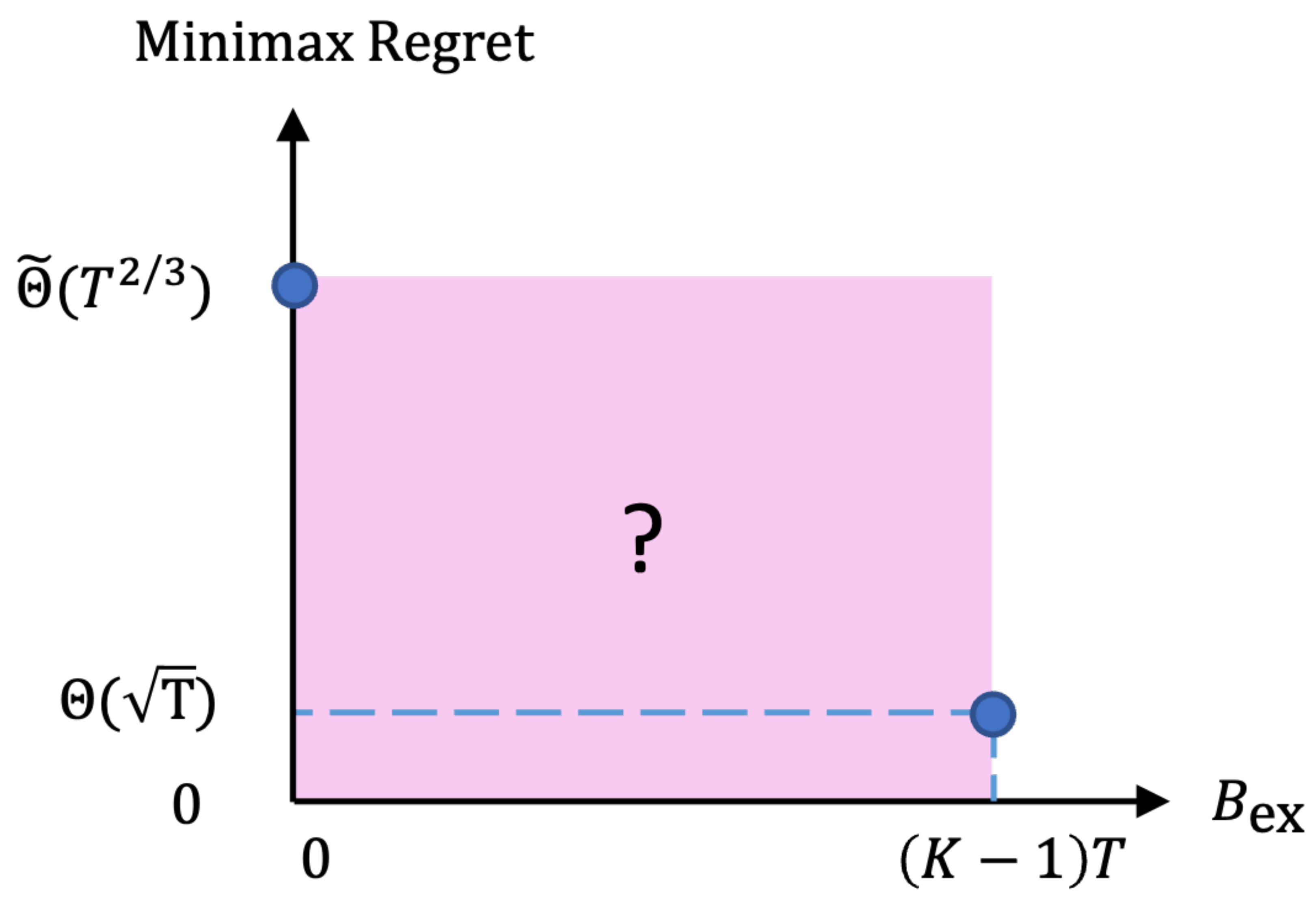


Learning with Bandit Feedback under Extra Observation Budget

Feedback: Bandit + B_{ex} losses in total

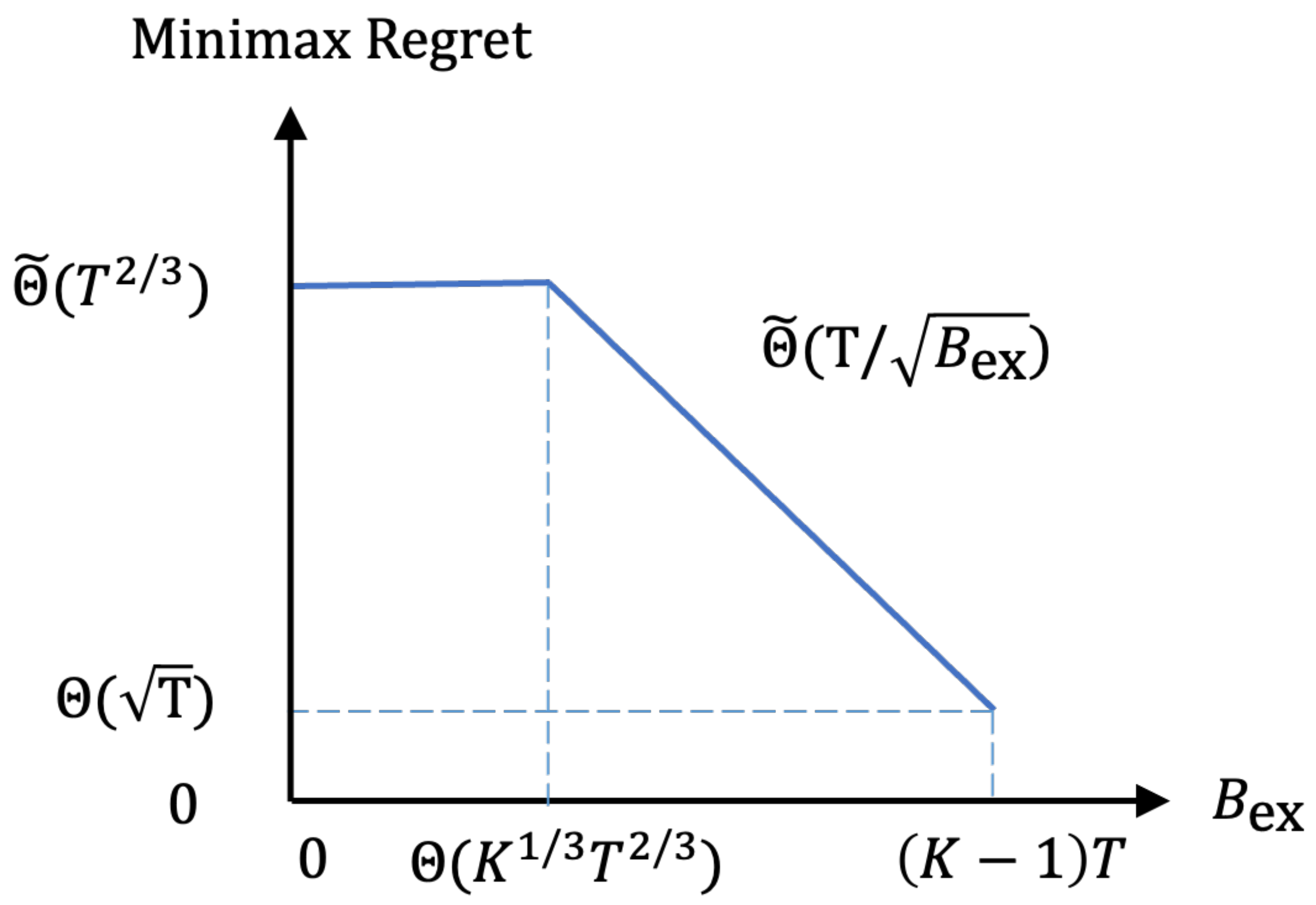
Extra Observations	$B_{\text{ex}} = 0$ (Bandit)		$B_{\text{ex}} = (K-1)T$ (Full-information)
Minimax Regret	$\tilde{\Theta}(T^{2/3})$	?	$\Theta(T^{1/2})$

Key Question:
How do extra observations help improve the regret in general?



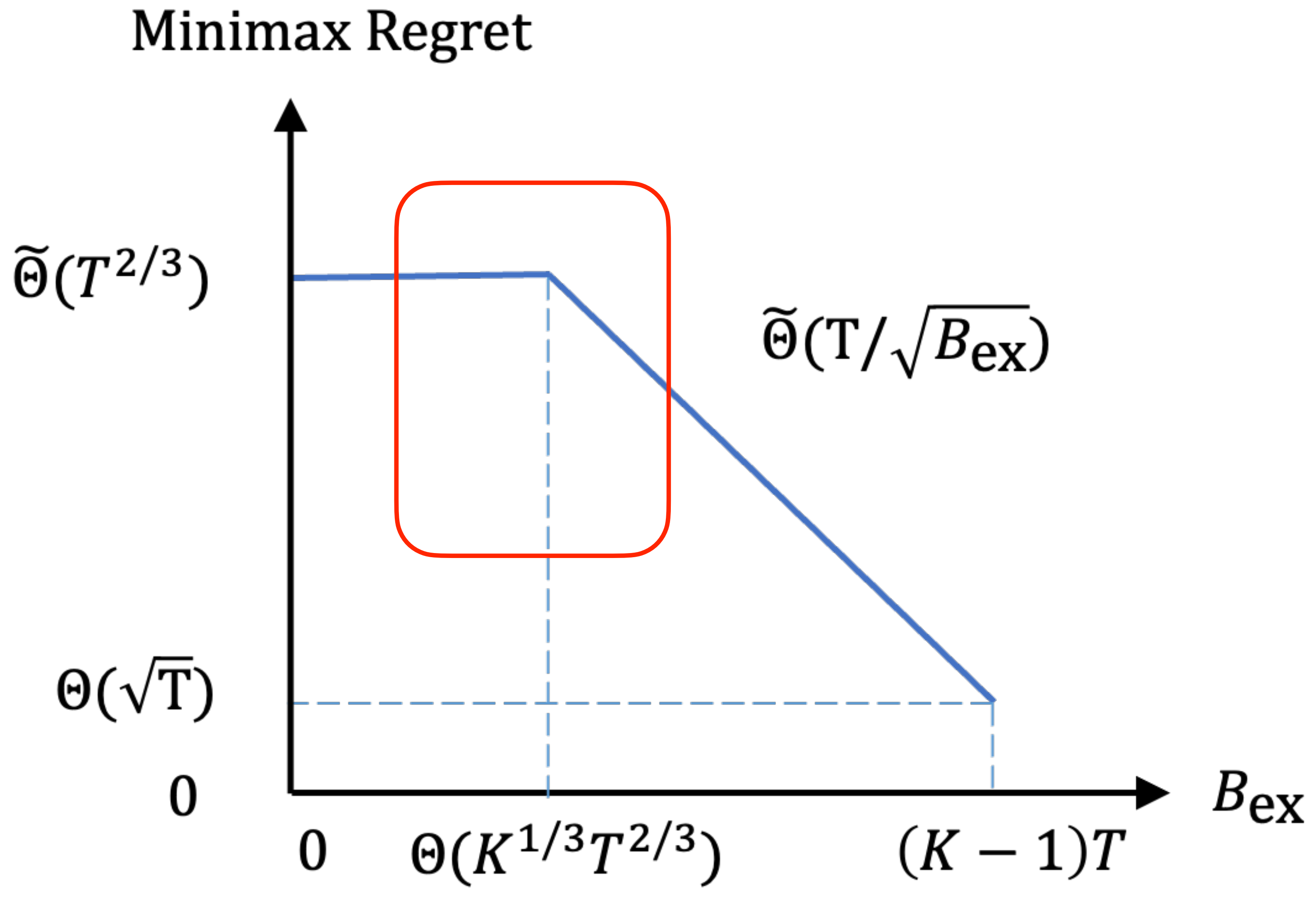
An Interesting Phase Transition

Extra Observations	$B_{\text{ex}} = 0$ (Bandit)	$B_{\text{ex}} = O(T^{2/3}K^{1/3})$	$B_{\text{ex}} = \Omega(T^{2/3}K^{1/3})$	$B_{\text{ex}} = (K-1)T$ (Full-information)
Minimax Regret	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T/\sqrt{B_{\text{ex}}})$	$\Theta(T^{1/2})$



An Interesting Phase Transition

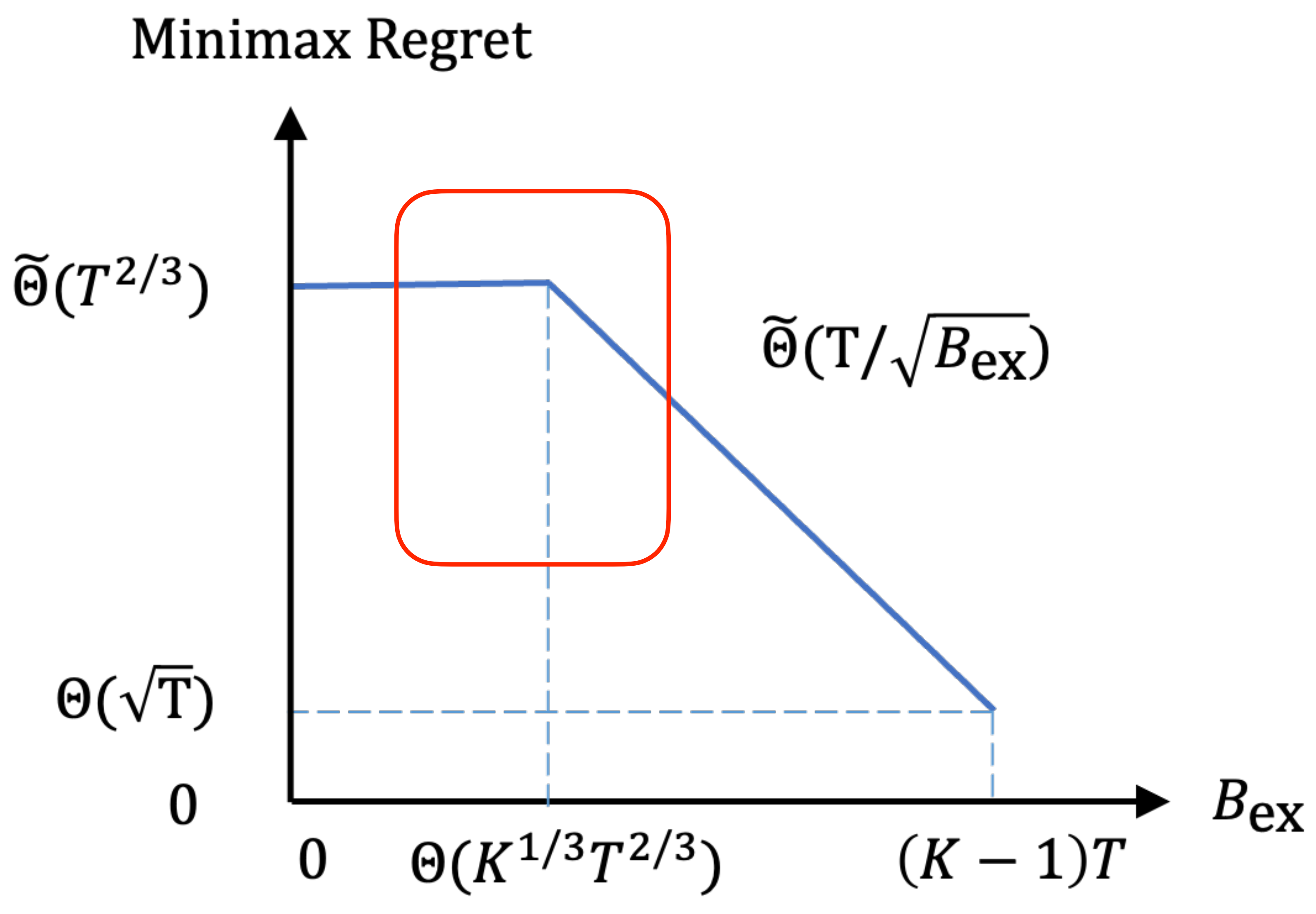
Extra Observations	$B_{\text{ex}} = 0$ (Bandit)	$B_{\text{ex}} = O(T^{2/3}K^{1/3})$	$B_{\text{ex}} = \Omega(T^{2/3}K^{1/3})$	$B_{\text{ex}} = (K-1)T$ (Full-information)
Minimax Regret	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T/\sqrt{B_{\text{ex}}})$	$\Theta(T^{1/2})$



An Interesting Phase Transition

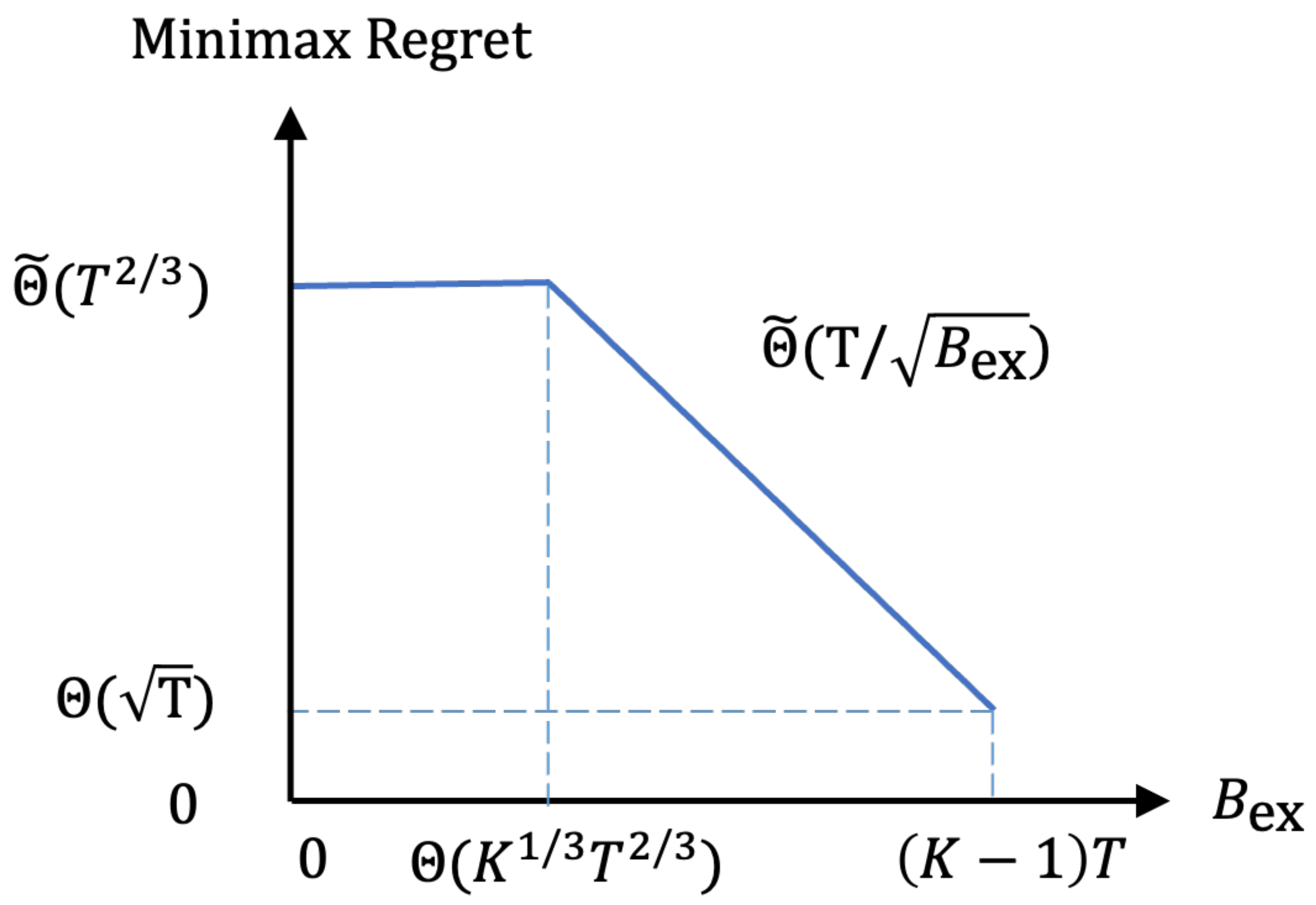
Extra Observations	$B_{\text{ex}} = 0$ (Bandit)	$B_{\text{ex}} = O(T^{2/3}K^{1/3})$	$B_{\text{ex}} = \Omega(T^{2/3}K^{1/3})$	$B_{\text{ex}} = (K-1)T$ (Full-information)
Minimax Regret	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T/\sqrt{B_{\text{ex}}})$	$\Theta(T^{1/2})$

Extra observations do not help until the amount is large enough



An Interesting Phase Transition

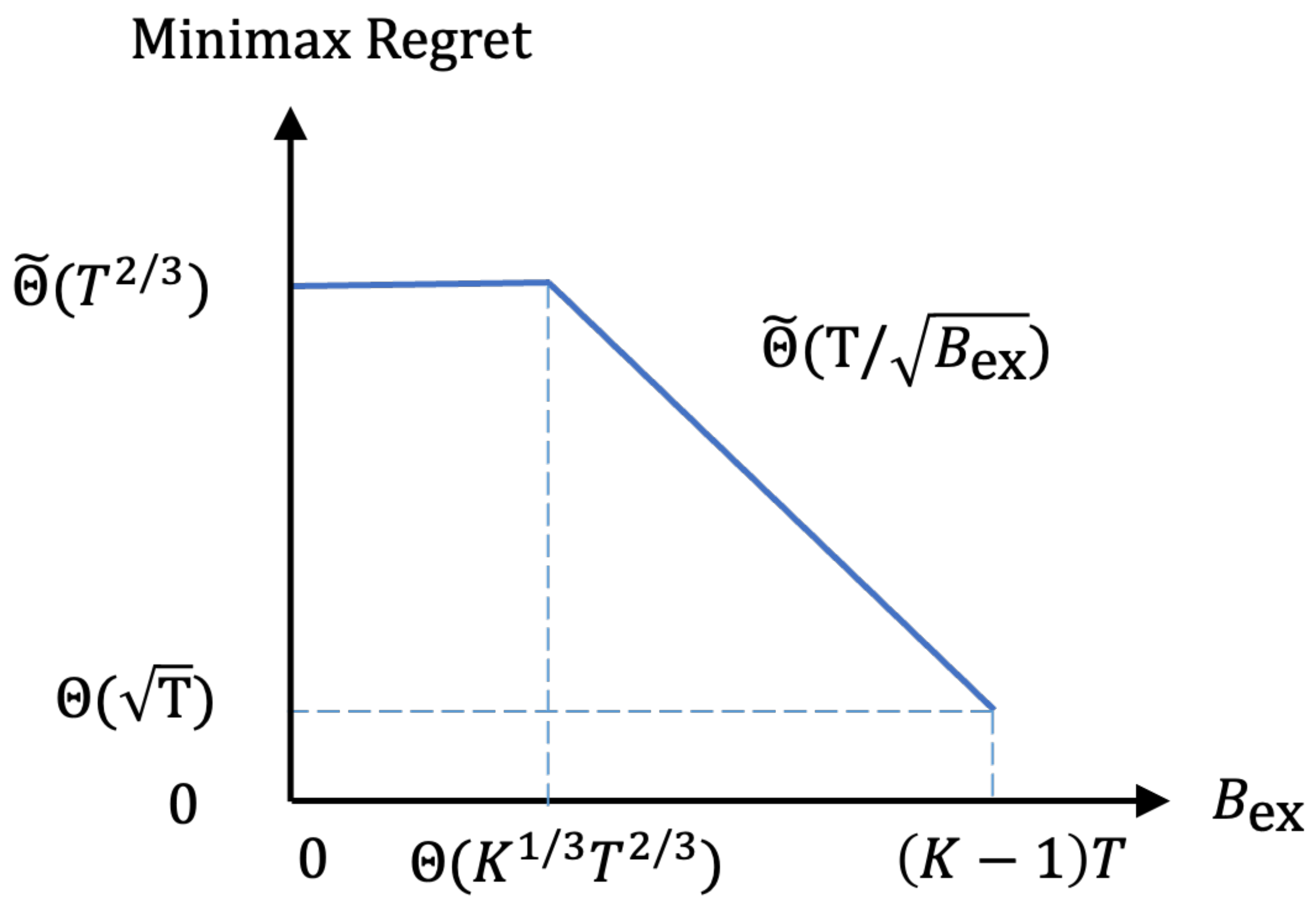
Extra Observations	$B_{\text{ex}} = 0$ (Bandit)	$B_{\text{ex}} = O(T^{2/3}K^{1/3})$	$B_{\text{ex}} = \Omega(T^{2/3}K^{1/3})$	$B_{\text{ex}} = (K-1)T$ (Full-information)
Minimax Regret	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T/\sqrt{B_{\text{ex}}})$	$\Theta(T^{1/2})$



Lower bound: Multi-scale random walk [Dekel et al., 13]

An Interesting Phase Transition

Extra Observations	$B_{\text{ex}} = 0$ (Bandit)	$B_{\text{ex}} = O(T^{2/3}K^{1/3})$	$B_{\text{ex}} = \Omega(T^{2/3}K^{1/3})$	$B_{\text{ex}} = (K-1)T$ (Full-information)
Minimax Regret	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T/\sqrt{B_{\text{ex}}})$	$\Theta(T^{1/2})$



Lower bound: Multi-scale random walk [Dekel et al., 13]

Upper bound: Instructive to study a different setup (to be introduced)

Learning under Total Observation Budget

Feedback: ~~Bandit~~ + $\cancel{B_{\text{ex}}}$ B losses in total

Learning under Total Observation Budget

Feedback: ~~Bandit~~ + $\cancel{B_{\text{ex}}}$ B losses in total

Total Observations	$B \in [K, KT]$	
	Without Switching Costs	With Switching Costs
Lower Bound	$\Omega(T/\sqrt{B})$ [Seldin et al., 14]	
Upper Bound	$\tilde{O}(T/\sqrt{B})$ [Seldin et al., 14]	

Learning under Total Observation Budget

Feedback: ~~Bandit~~ + ~~B_{ex}~~ B losses in total

Total Observations	$B \in [K, KT]$	
	Without Switching Costs	With Switching Costs
Lower Bound	$\Omega(T/\sqrt{B})$ [Seldin et al., 14]	$\Omega(T/\sqrt{B})$
Upper Bound	$\tilde{O}(T/\sqrt{B})$ [Seldin et al., 14]	

Key Questions:

1. Is this lower bound tight with switching costs?

Learning under Total Observation Budget

Feedback: ~~Bandit~~ + $\cancel{B_{\text{ex}}}$ B losses in total

Total Observations	$B \in [K, KT]$	
	Without Switching Costs	With Switching Costs
Lower Bound	$\Omega(T/\sqrt{B})$ [Seldin et al., 14]	$\Omega(T/\sqrt{B})$
Upper Bound	$\tilde{O}(T/\sqrt{B})$ [Seldin et al., 14]	?

Key Questions:

1. Is this lower bound tight with switching costs?
2. What is the best upper bound we can achieve?

The Same Minimax Regret Rate

Total Observations	$B \in [K, KT]$	
	Without Switching Costs	With Switching Costs
Lower Bound	$\Omega(T/\sqrt{B})$	$\Omega(T/\sqrt{B})$
Upper Bound	$\tilde{O}(T/\sqrt{B})$	$\tilde{O}(T/\sqrt{B})$

The Same Minimax Regret Rate

Total Observations	$B \in [K, KT]$	
	Without Switching Costs	With Switching Costs
Lower Bound	$\Omega(T/\sqrt{B})$	$\Omega(T/\sqrt{B})$
Upper Bound	$\tilde{O}(T/\sqrt{B})$	$\tilde{O}(T/\sqrt{B})$
Minimax Regret	$\tilde{\Theta}(T/\sqrt{B})$	$\tilde{\Theta}(T/\sqrt{B})$

The Same Minimax Regret Rate

Total Observations	$B \in [K, KT]$	
	Without Switching Costs	With Switching Costs
Lower Bound	$\Omega(T/\sqrt{B})$	$\Omega(T/\sqrt{B})$
Upper Bound	$\tilde{O}(T/\sqrt{B})$	$\tilde{O}(T/\sqrt{B})$
Minimax Regret	$\tilde{\Theta}(T/\sqrt{B})$	$\tilde{\Theta}(T/\sqrt{B})$

Adding switching costs does not increase the minimax regret rate

The Feedback Type Matters

Feedback Type	Minimax Regret $B \in [K, KT]$	
	Without Switching Costs	With Switching Costs
Full-information		
Bandit ($B = \mathcal{O}(T^{2/3}K^{1/3})$)		
Bandit ($B = \Omega(T^{2/3}K^{1/3})$)		

The Feedback Type Matters

<table>
<tr><th rowspan="2">Feedback Type</th><th colspan="2">Minimax Regret
$B \in [K, KT]$</th></tr>
<tr><th>Without
Switching Costs</th><th>With
Switching Costs</th></tr>
<tr><td>Full-information</td><td rowspan="3">$\widetilde{\Theta}(T/\sqrt{B})$</td><td rowspan="2"></td></tr>
<tr><td>Bandit
($B = \mathcal{O}(T^{2/3}K^{1/3})$)</td></tr>
<tr><td>Bandit
($B = \Omega(T^{2/3}K^{1/3})$)</td><td></td></tr>
</table>

The Feedback Type Matters

Feedback Type	Minimax Regret $B \in [K, KT]$	
	Without Switching Costs	With Switching Costs
Full-information	$\widetilde{\Theta}(T/\sqrt{B})$	$\widetilde{\Theta}(T/\sqrt{B})$
Bandit ($B = \mathcal{O}(T^{2/3}K^{1/3})$)		
Bandit ($B = \Omega(T^{2/3}K^{1/3})$)		

The Feedback Type Matters

<table>
<tr><th rowspan="2">Feedback Type</th><th colspan="2">Minimax Regret
$B \in [K, KT]$</th></tr>
<tr><th>Without
Switching Costs</th><th>With
Switching Costs</th></tr>
<tr><td>Full-information</td><td rowspan="3">$\widetilde{\Theta}(T/\sqrt{B})$</td><td rowspan="2">$\widetilde{\Theta}(T/\sqrt{B})$</td></tr>
<tr><td>Bandit
($B = \mathcal{O}(T^{2/3}K^{1/3})$)</td></tr>
<tr><td>Bandit
($B = \Omega(T^{2/3}K^{1/3})$)</td><td>$\widetilde{\Theta}(T^{2/3})$</td></tr>
</table>

Reference

- [Cesa-Bianchi & Lugosi, 06] Cesa-Bianchi, Nicolò and Gábor Lugosi. “Prediction, learning, and games.” (2006).
- [Auer et al., 95] Auer, Peter et al. “Gambling in a rigged casino: The adversarial multi-armed bandit problem.” *Proceedings of IEEE 36th Annual Foundations of Computer Science* (1995): 322-331.
- [Audibert & Bubeck, 09] Audibert, Jean-Yves and Sébastien Bubeck. “Minimax Policies for Adversarial and Stochastic Bandits.” Annual Conference Computational Learning Theory (2009).
- [Geulen et al., 10] Geulen, Sascha et al. “Regret Minimization for Online Buffering Problems Using the Weighted Majority Algorithm.” Annual Conference Computational Learning Theory (2010).
- [Devroye et al., 10] Devroye, Luc et al. “Prediction by random-walk perturbation.” Annual Conference Computational Learning Theory (2013).
- [Arora et al., 12] Arora, Raman et al. “Online Bandit Learning against an Adaptive Adversary: from Regret to Policy Regret.” International Conference on Machine Learning (2012).
- [Dekel et al., 13] Dekel, Ofer, et al. "Bandits with Switching Costs: T^{2/3} Regret." arXiv preprint arXiv:1310.2997 (2013).
- [Seldin et al., 14] Seldin, Yevgeny et al. “Prediction with Limited Advice and Multiarmed Bandits with Paid Observations.” International Conference on Machine Learning (2014).

Thank you!