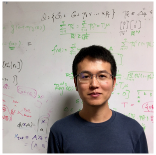
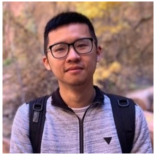
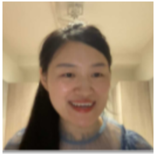
Achieving Near Instance-Optimality and Minimax-Optimality in Stochastic and Adversarial Linear Bandits Simultaneously

Mengxiao Zhang

joint with **Chung-Wei Lee**, **Haipeng Luo**, **Chen-Yu Wei** and **Xiaojin Zhang**







USC Viterbi
School of Engineering

Bandits Problem

Bandits Problem

Multi-Armed Bandits (MAB)

Bandits Problem

Multi-Armed Bandits (MAB)

- d arms/actions available

Bandits Problem

Multi-Armed Bandits (MAB)

- d arms/actions available
- environment decides the losses for each arm

Bandits Problem

Multi-Armed Bandits (MAB)

- d arms/actions available
- environment decides the losses for each arm
- learner sequentially pulls an arm and observes its loss

Bandits Problem

Multi-Armed Bandits (MAB)

- d arms/actions available
- environment decides the losses for each arm
- learner sequentially pulls an arm and observes its loss
- goal: be competitive with the best fixed arm

Bandits Problem

Multi-Armed Bandits (MAB)

- d arms/actions available
- environment decides the losses for each arm
- learner sequentially pulls an arm and observes its loss
- goal: be competitive with the best fixed arm

Linear Bandits (LB)

Bandits Problem

Multi-Armed Bandits (MAB)

- d arms/actions available
- environment decides the losses for each arm
- learner sequentially pulls an arm and observes its loss
- goal: be competitive with the best fixed arm

Linear Bandits (LB)

- a finite set of actions $\mathcal{X} \subset \mathbb{R}^d$

Bandits Problem

Multi-Armed Bandits (MAB)

- d arms/actions available
- environment decides the losses for each arm
- learner sequentially pulls an arm and observes its loss
- goal: be competitive with the best fixed arm

Linear Bandits (LB)

- a finite set of actions $\mathcal{X} \subset \mathbb{R}^d$
- environment decides the loss vectors $\theta_t \in \mathbb{R}^d$

Bandits Problem

Multi-Armed Bandits (MAB)

- d arms/actions available
- environment decides the losses for each arm
- learner sequentially pulls an arm and observes its loss
- goal: be competitive with the best fixed arm

Linear Bandits (LB)

- a finite set of actions $\mathcal{X} \subset \mathbb{R}^d$
- environment decides the loss vectors $\theta_t \in \mathbb{R}^d$
- learner sequentially chooses an action x_t from $\mathcal{X}$ and observes its loss, which is $\langle x_t, \theta_t \rangle + \epsilon_t$, ϵ_t is zero-mean random noise

Bandits Problem

Multi-Armed Bandits (MAB)

- d arms/actions available
- environment decides the losses for each arm
- learner sequentially pulls an arm and observes its loss
- goal: be competitive with the best fixed arm

Linear Bandits (LB)

- a finite set of actions $\mathcal{X} \subset \mathbb{R}^d$
- environment decides the loss vectors $\theta_t \in \mathbb{R}^d$
- learner sequentially chooses an action x_t from $\mathcal{X}$ and observes its loss, which is $\langle x_t, \theta_t \rangle + \epsilon_t$, ϵ_t is zero-mean random noise
- goal: be competitive with the best fixed action in $\mathcal{X}$

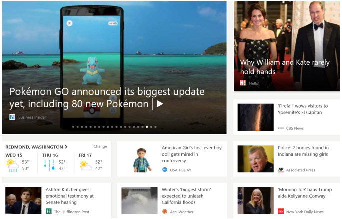
Bandits Problem

Multi-Armed Bandits (MAB)

- d arms/actions available
- environment decides the losses for each arm
- learner sequentially pulls an arm and observes its loss
- goal: be competitive with the best fixed arm

Linear Bandits (LB) (e.g. news recommendation)

- a finite set of actions $\mathcal{X} \subset \mathbb{R}^d$
- environment decides the loss vectors $\theta_t \in \mathbb{R}^d$
- learner sequentially chooses an action x_t from $\mathcal{X}$ and observes its loss, which is $\langle x_t, \theta_t \rangle + \epsilon_t$, ϵ_t is zero-mean random noise
- goal: be competitive with the best fixed action in $\mathcal{X}$



Linear Bandits with Different Environments

- stochastic linear bandits
- stochastic linear bandits with corruptions
- adversarial linear bandits

Stochastic Linear Bandits

Stochastic Linear Bandits

Stochastic environment: the loss vectors over rounds $\theta_t = \theta$ are the same

Stochastic Linear Bandits

Stochastic environment: the loss vectors over rounds $\theta_t = \theta$ are the same

- LinUCB: expected regret bound $\mathcal{O}\left(\frac{\log^2 T + d\log T + d^2 \log\log T}{\Delta_{\min}}\right)$ [APS11]

Stochastic Linear Bandits

Stochastic environment: the loss vectors over rounds $\theta_t = \theta$ are the same

- $\textsc{LinUCB}$: expected regret bound $\mathcal{O}\left(\frac{\log^2 T + d \log T + d^2 \log\log T}{\Delta_{\min}}\right)$ [APS11]
  - $\Delta_{\min}$: minimum sub-optimality gap

Stochastic Linear Bandits

Stochastic environment: the loss vectors over rounds $\theta_t = \theta$ are the same

- $\textsc{LinUCB}$: expected regret bound $\mathcal{O}\left(\frac{\log^2 T + d\log T + d^2 \log\log T}{\Delta_{\min}}\right)$ [APS11]
  - $\Delta_{\min}$: minimum sub-optimality gap
  - NOT instance-optimal; can be arbitrarily worse than the optimal bound [LS16]

Stochastic Linear Bandits

Stochastic environment: the loss vectors over rounds $\theta_t = \theta$ are the same

- $\textsc{LinUCB}$: expected regret bound $\mathcal{O}\left(\frac{\log^2 T + d \log T + d^2 \log\log T}{\Delta_{\min}}\right)$ [APS11]
  - $\Delta_{\min}$: minimum sub-optimality gap
  - NOT instance-optimal; can be arbitrarily worse than the optimal bound [LS16]
- a line of recent works achieve instance-optimality [LS16,CMP17,HL20]

Stochastic Linear Bandits

Stochastic environment: the loss vectors over rounds $\theta_t = \theta$ are the same

- LinUCB: expected regret bound $\mathcal{O}\left(\frac{\log^2 T + d \log T + d^2 \log\log T}{\Delta_{\min}}\right)$ [APS11]
  - $\Delta_{\min}$: minimum sub-optimality gap
  - NOT instance-optimal; can be arbitrarily worse than the optimal bound [LS16]
- a line of recent works achieve instance-optimality [LS16,CMP17,HL20]
  - instance-optimal expected regret bound $c(\mathcal{X}, \theta) \log T$

Stochastic Linear Bandits

Stochastic environment: the loss vectors over rounds $\theta_t = \theta$ are the same

- LinUCB: expected regret bound $\mathcal{O}\left(\frac{\log^2 T + d \log T + d^2 \log\log T}{\Delta_{\min}}\right)$ [APS11]
  - $\Delta_{\min}$: minimum sub-optimality gap
  - NOT instance-optimal; can be arbitrarily worse than the optimal bound [LS16]
- a line of recent works achieve instance-optimality [LS16,CMP17,HL20]
  - instance-optimal expected regret bound $c(\mathcal{X}, \theta) \log T$
  - $c(\mathcal{X}, \theta)$ is an instance dependent constant

Stochastic Linear Bandits with Corruptions

Stochastic Linear Bandits with Corruptions

Stochastic environment with corruption: environment can corrupt the total loss by C

Stochastic Linear Bandits with Corruptions

Stochastic environment with corruption: environment can corrupt the total loss by C

- $C = 0$ recovers the stochastic setting

Stochastic Linear Bandits with Corruptions

Stochastic environment with corruption: environment can corrupt the total loss by C

- $C = 0$ recovers the stochastic setting
- optimal $\mathcal{O}\left(\sum_{i \neq i^*} \frac{\log T}{\Delta_i} + C\right)$ regret in MAB [GKT19,ZS19]

Stochastic Linear Bandits with Corruptions

Stochastic environment with corruption: environment can corrupt the total loss by C

- $C = 0$ recovers the stochastic setting
- optimal $\mathcal{O}\left(\sum_{i \neq i^*} \frac{\log T}{\Delta_i} + C\right)$ regret in MAB [GKT19,ZS19]
- sub-optimal $\mathcal{O}\left(\frac{d^6 \log^2 T}{\Delta_{\min}^2} + \frac{d^{2.5} C \log T}{\Delta_{\min}}\right)$ regret in LB [LLS19]

Stochastic Linear Bandits with Corruptions

Stochastic environment with corruption: environment can corrupt the total loss by C

- $C = 0$ recovers the stochastic setting
- optimal $\mathcal{O}\left(\sum_{i \neq i^*} \frac{\log T}{\Delta_i} + C\right)$ regret in MAB [GKT19,ZS19]
- sub-optimal $\mathcal{O}\left(\frac{d^6 \log^2 T}{\Delta_{\min}^2} + \frac{d^{2.5} C \log T}{\Delta_{\min}}\right)$ regret in LB [LLS19]

Question 1: whether instance-optimal regret bound with optimal $\mathcal{O}(C)$ overhead is achievable in stochastic LB with corruptions?

Adversarial Linear Bandits

Adversarial environment: the loss vectors θ_t over rounds are arbitrary.

Adversarial Linear Bandits

Adversarial environment: the loss vectors θ_t over rounds are arbitrary.

Expected regret:

Adversarial Linear Bandits

Adversarial environment: the loss vectors θ_t over rounds are arbitrary.

Expected regret:

- SCRIBLE, GEOMETRICHEDGE: $\tilde{\mathcal{O}}\left(\sqrt{T}\right)$ [AHR12, DHK08]

Adversarial Linear Bandits

Adversarial environment: the loss vectors θ_t over rounds are arbitrary.

Expected regret:

- $\textsc{SCRiBLe}$, $\textsc{GeometricHedge}$: $\tilde{\mathcal{O}}\left(\sqrt{T}\right)$ [AHR12, DHK08]

High probability regret: $\tilde{\mathcal{O}}\left(\sqrt{T}\right)$ [BDHKRT08,LLWZ20]

Best-of-Three-Worlds

Best-of-Three-Worlds

Question 2: can we achieve the best of three worlds simultaneously?

Best-of-Three-Worlds

Question 2: can we achieve the best of three worlds simultaneously?

Recent works made progress for different problems:

Best-of-Three-Worlds

> Question 2: can we achieve the best of three worlds simultaneously?

Recent works made progress for different problems:

- MAB: Tsallis-INF [ZS19]

Best-of-Three-Worlds

Question 2: can we achieve the best of three worlds simultaneously?

Recent works made progress for different problems:

- MAB: Tsallis-INF [ZS19]
- Combinatorial semi-bandits (for stochastic and adversarial environments) [ZLW19]

Best-of-Three-Worlds

Question 2: can we achieve the best of three worlds simultaneously?

Recent works made progress for different problems:

- MAB: Tsallis-INF [ZS19]
- Combinatorial semi-bandits (for stochastic and adversarial environments) [ZLW19]
- Markov Decision Processes [JL20, JHK21]

Questions:

1. instance-optimal regret + optimal $\mathcal{O}(C)$ overhead in corrupted LB?
2. achieve the best of three worlds simultaneously in LB?

Questions:

1. instance-optimal regret $+$ optimal $\mathcal{O}(C)$ overhead in corrupted LB?
2. achieve the best of three worlds simultaneously in LB?

This work provides positive answers:

Questions:

1. instance-optimal regret + optimal $\mathcal{O}(C)$ overhead in corrupted LB?
2. achieve the best of three worlds simultaneously in LB?

This work provides positive answers:

- a new algorithm $\mathcal{A}$ with $\mathcal{O}(c(\mathcal{X},\theta)\log^2 T + C)$ h.p. regret

Questions:

1. instance-optimal regret + optimal $\mathcal{O}(C)$ overhead in corrupted LB?
2. achieve the best of three worlds simultaneously in LB?

This work provides positive answers:

- a new algorithm $\mathcal{A}$ with $\mathcal{O}(c(\mathcal{X},\theta)\log^2 T + C)$ h.p. regret
  - design a novel optimization problem to construct the strategy based on robust loss estimators

Questions:

1. instance-optimal regret $+$ optimal $\mathcal{O}(C)$ overhead in corrupted LB?
2. achieve the best of three worlds simultaneously in LB?

This work provides positive answers:

- a new algorithm $\mathcal{A}$ with $\mathcal{O}(c(\mathcal{X},\theta)\log^2 T + C)$ h.p. regret
  - design a novel optimization problem to construct the strategy based on robust loss estimators
- another algorithm achieves the following h.p. bounds simultaneously

Questions:

1. instance-optimal regret + optimal $\mathcal{O}(C)$ overhead in corrupted LB?
2. achieve the best of three worlds simultaneously in LB?

This work provides positive answers:

- a new algorithm $\mathcal{A}$ with $\mathcal{O}(c(\mathcal{X}, \theta) \log^2 T + C)$ h.p. regret
  - design a novel optimization problem to construct the strategy based on robust loss estimators
- another algorithm achieves the following h.p. bounds simultaneously
  - $\mathcal{O}(c(\mathcal{X}, \theta) \log^2 T)$ in the stochastic environment;

Questions:

1. instance-optimal regret + optimal $\mathcal{O}(C)$ overhead in corrupted LB?
2. achieve the best of three worlds simultaneously in LB?

This work provides positive answers:

- a new algorithm $\mathcal{A}$ with $\mathcal{O}(c(\mathcal{X},\theta)\log^2 T + C)$ h.p. regret
  - design a novel optimization problem to construct the strategy based on robust loss estimators
- another algorithm achieves the following h.p. bounds simultaneously
  - $\mathcal{O}(c(\mathcal{X},\theta)\log^2 T)$ in the stochastic environment;
  - $\mathcal{O}(\frac{d\log^2 T}{\Delta_{\min}} + C)$ in the stochastic environment with corruptions;

Questions:

1. instance-optimal regret + optimal $\mathcal{O}(C)$ overhead in corrupted LB?
2. achieve the best of three worlds simultaneously in LB?

This work provides positive answers:

- a new algorithm $\mathcal{A}$ with $\mathcal{O}(c(\mathcal{X}, \theta) \log^2 T + C)$ h.p. regret
  - design a novel optimization problem to construct the strategy based on robust loss estimators
- another algorithm achieves the following h.p. bounds simultaneously
  - $\mathcal{O}(c(\mathcal{X}, \theta) \log^2 T)$ in the stochastic environment;
  - $\mathcal{O}(\frac{d \log^2 T}{\Delta_{\min}} + C)$ in the stochastic environment with corruptions;
  - $\tilde{\mathcal{O}}(\sqrt{T})$ in the adversarial environment;

Questions:

1. instance-optimal regret + optimal $\mathcal{O}(C)$ overhead in corrupted LB?
2. achieve the best of three worlds simultaneously in LB?

This work provides positive answers:

- a new algorithm $\mathcal{A}$ with $\mathcal{O}(c(\mathcal{X},\theta)\log^2 T + C)$ h.p. regret
  - design a novel optimization problem to construct the strategy based on robust loss estimators
- another algorithm achieves the following h.p. bounds simultaneously
  - $\mathcal{O}(c(\mathcal{X},\theta)\log^2 T)$ in the stochastic environment;
  - $\mathcal{O}(\frac{d\log^2 T}{\Delta_{\min}} + C)$ in the stochastic environment with corruptions;
  - $\tilde{\mathcal{O}}(\sqrt{T})$ in the adversarial environment;

  a two-phase algorithm

Questions:

1. instance-optimal regret + optimal $\mathcal{O}(C)$ overhead in corrupted LB?
2. achieve the best of three worlds simultaneously in LB?

This work provides positive answers:

- a new algorithm $\mathcal{A}$ with $\mathcal{O}(c(\mathcal{X},\theta)\log^2 T + C)$ h.p. regret
  - design a novel optimization problem to construct the strategy based on robust loss estimators
- another algorithm achieves the following h.p. bounds simultaneously
  - $\mathcal{O}(c(\mathcal{X},\theta)\log^2 T)$ in the stochastic environment;
  - $\mathcal{O}(\frac{d\log^2 T}{\Delta_{\min}} + C)$ in the stochastic environment with corruptions;
  - $\tilde{\mathcal{O}}(\sqrt{T})$ in the adversarial environment;

  a two-phase algorithm
  - Phase 1: a h.p. adversarial LB algorithm

Questions:

1. instance-optimal regret + optimal $\mathcal{O}(C)$ overhead in corrupted LB?
2. achieve the best of three worlds simultaneously in LB?

This work provides positive answers:

- a new algorithm $\mathcal{A}$ with $\mathcal{O}(c(\mathcal{X},\theta)\log^2 T + C)$ h.p. regret
  - design a novel optimization problem to construct the strategy based on robust loss estimators
- another algorithm achieves the following h.p. bounds simultaneously
  - $\mathcal{O}(c(\mathcal{X},\theta)\log^2 T)$ in the stochastic environment;
  - $\mathcal{O}(\frac{d\log^2 T}{\Delta_{\min}} + C)$ in the stochastic environment with corruptions;
  - $\tilde{\mathcal{O}}(\sqrt{T})$ in the adversarial environment;

  a two-phase algorithm
  - Phase 1: a h.p. adversarial LB algorithm
  - Phase 2: a modified $\mathcal{A}$ with a stationarity check