

On the Complexity of Bandit Linear Optimization

Ohad Shamir

Weizmann Institute of Science



COLT 2015

Bandit Linear Optimization

For $t = 1, 2, \dots, T$

- Player chooses $\mathbf{w}_t \in \mathcal{W} \subset \mathbb{R}^d$
- Simultaneously, adversary chooses $\mathbf{x}_t \in \mathcal{X} \subset \mathbb{R}^d$
- Player observes and suffers loss $\langle \mathbf{w}_t, \mathbf{x}_t \rangle$ (\mathbf{x}_t not revealed)

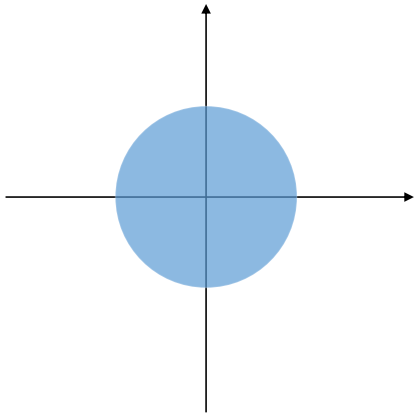
Goal: Minimize regret

$$\sum_{t=1}^T \langle \mathbf{w}_t, \mathbf{x}_t \rangle - \min_{\mathbf{w} \in \mathcal{W}} \sum_{t=1}^T \langle \mathbf{w}, \mathbf{x}_t \rangle$$

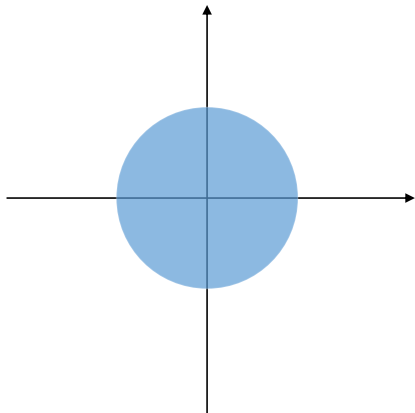
Question

How does regret depend on $\mathcal{W}, \mathcal{X}, d, T$?

Domain \mathcal{W}



Domain \mathcal{W}

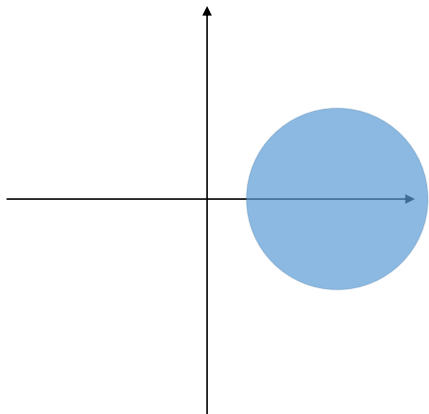


Full-inf. regret: \sqrt{T}

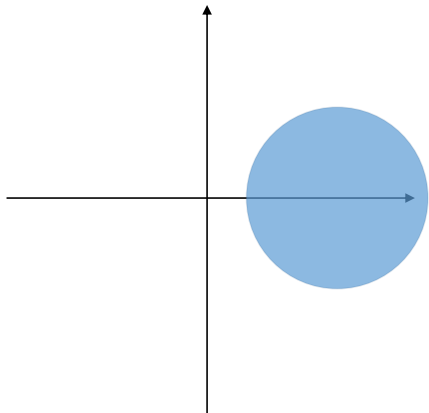
Bandit regret: \sqrt{dT}

(Bubeck, Cesa-Bianchi, Kakade 2012)

Domain \mathcal{W}

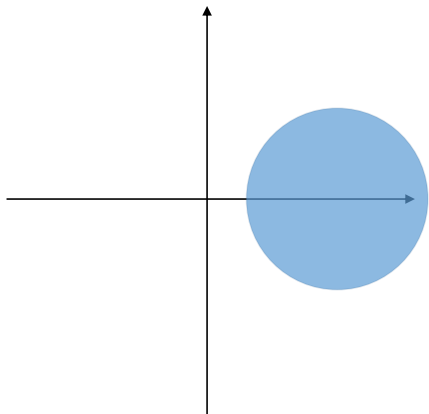


Domain \mathcal{W}



Full-inf. regret: \sqrt{T}

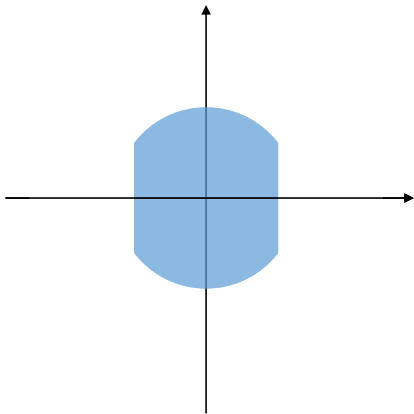
Domain \mathcal{W}



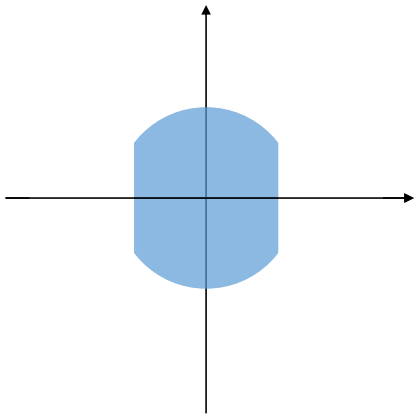
Full-inf. regret: \sqrt{T}

Bandit regret: $d\sqrt{T}$

Domain \mathcal{W}



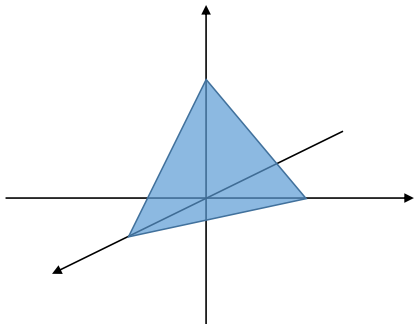
Domain \mathcal{W}



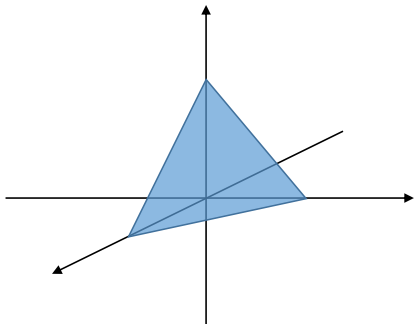
Full-inf. regret: \sqrt{T}

Bandit regret: $d\sqrt{T}$

Domain \mathcal{W}

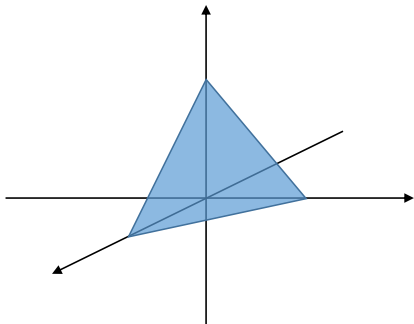


Domain \mathcal{W}



Full-inf. regret: $\sqrt{\log(\mathbf{d})T}$

Domain \mathcal{W}



Full-inf. regret: $\sqrt{\log(\mathbf{d})\mathbf{T}}$

Bandit regret: $\sqrt{\log(\mathbf{d})\mathbf{T}}$

See Poster for details