Bandit Algorithms for Online Optimization

Nicolò Cesa-Bianchi

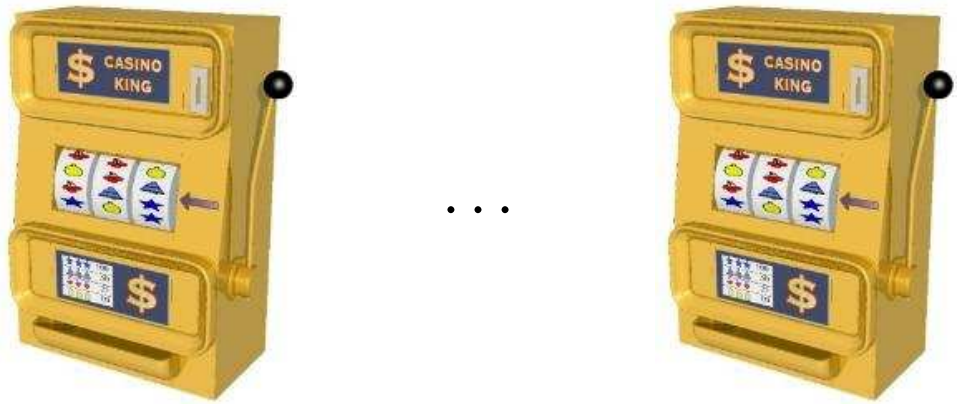
Università degli Studi di Milano

Multiarmed Bandits

Overview

1. The multiarmed bandit problem
    1. Algorithm `Exp3`
2. Combinatorial bandits (online linear optimization in $\{0,1\}^d$)
    1. Spanners and `GeometricHedge`
    2. A more specific approach: `ComBand`
    3. Self-concordant functions
3. Discussion

The multiarmed bandit problem



. . .

N slot machines

- Each play of a slot machine (action) returns a **payoff** (gain or loss)
- Design a strategy of **repeated play** to maximize gain / minimize loss
- A classical problem in **sequential design of experiments** [Robbins, 1952]
- Modern applications: ad placement, dynamic pricing, routing in networks, active model selection, . . .

Main ingredients

- Partial feedback: only one action is played at each time step

Main ingredients

- Partial feedback: only one action is played at each time step
- Exploration: find out the single best action
- Exploitation: play the best action as often as possible

Main ingredients

- Partial feedback: only one action is played at each time step
- Exploration: find out the single best action
- Exploitation: play the best action as often as possible
- Dilemma: when to explore? And how much?

Main ingredients

- Partial feedback: only one action is played at each time step
- Exploration: find out the single best action
- Exploitation: play the best action as often as possible
- Dilemma: when to explore? And how much?

Payoff generation

- Stochastic/nonstochastic assumptions
- We focus on the nonstochastic case

Nonstochastic bandit problem [Baños, 1968; Megiddo, 1980]

Problem description

- A fixed but unknown sequence $x_1, x_2, \ldots$ of **loss functions**

$$x_t : \{1, \ldots, N\} \to [0, 1]$$

- $x_t(i)$ is loss of action $i = 1, \ldots, N$ if played at time t

Nonstochastic bandit problem [Baños, 1968; Megiddo, 1980]

Problem description

- A fixed but unknown sequence $x_1, x_2, \ldots$ of **loss functions**

$$x_t : \{1, \ldots, N\} \to [0, 1]$$

- $x_t(i)$ is loss of action $i = 1, \ldots, N$ if played at time t

For $t = 1, 2, \ldots$

1. Select action $I_t \in \{1, \ldots, N\}$
2. Observe loss $x_t(I_t)$ (losses $x_t(i)$ for $i \neq I_t$ remains hidden)

Nonstochastic bandit problem [Baños, 1968; Megiddo, 1980]

Problem description

- A fixed but unknown sequence $x_1, x_2, \ldots$ of **loss functions**

$$x_t : \{1, \ldots, N\} \to [0, 1]$$

- $x_t(i)$ is loss of action $i = 1, \ldots, N$ if played at time t

For $t = 1, 2, \ldots$

1. Select action $I_t \in \{1, \ldots, N\}$
2. Observe loss $x_t(I_t)$ (losses $x_t(i)$ for $i \neq I_t$ remains hidden)

Regret for play sequence $I_1, I_2, \ldots, I_T$

How much we lose by not consistently playing the **single best** action

$$\sum_{t=1}^{T} x_t(I_t) - \min_{i=1,\ldots,N} \sum_{t=1}^{T} x_t(i)$$

A playing strategy: Exp3 [Auer, C-B, Freund and Schapire, 1995]

- Randomization: At time t, action i is played with probability

$$p_t(i) = (1-\gamma)\underbrace{\frac{w_t(i)}{\sum_j w_t(j)}}_{\text{exploitation}} + \underbrace{\frac{\gamma}{N}}_{\text{exploration}} \qquad (0 \leqslant \gamma \leqslant 1)$$

A playing strategy: Exp3 [Auer, C-B, Freund and Schapire, 1995]

- Randomization: At time t, action i is played with probability

$$p_t(i) = (1-\gamma)\underbrace{\frac{w_t(i)}{\sum_j w_t(j)}}_{\text{exploitation}} + \underbrace{\frac{\gamma}{N}}_{\text{exploration}} \qquad (0 \leqslant \gamma \leqslant 1)$$

- Weights: $$w_t(i) = \exp\left(-\frac{\gamma}{N}\underbrace{\sum_{s=1}^{t-1}\widehat{x}_s(i)}_{\text{estimated past loss}}\right)$$

A playing strategy: Exp3 [Auer, C-B, Freund and Schapire, 1995]

- Randomization: At time t, action i is played with probability

$$p_t(i) = (1-\gamma)\underbrace{\frac{w_t(i)}{\sum_j w_t(j)}}_{\text{exploitation}} + \underbrace{\frac{\gamma}{N}}_{\text{exploration}} \qquad (0 \leqslant \gamma \leqslant 1)$$

- Weights: $$w_t(i) = \exp\left(-\frac{\gamma}{N}\underbrace{\sum_{s=1}^{t-1}\widehat{x}_s(i)}_{\text{estimated past loss}}\right)$$

- Loss vector estimate: $$\widehat{x}_t(i) = \frac{x_t(i)}{p_t(i)}\mathbb{I}_{\{I_t=i\}}$$

A playing strategy: Exp3 [Auer, C-B, Freund and Schapire, 1995]

- Randomization: At time t, action i is played with probability

$$p_t(i) = (1-\gamma)\underbrace{\frac{w_t(i)}{\sum_j w_t(j)}}_{\text{exploitation}} + \underbrace{\frac{\gamma}{N}}_{\text{exploration}} \qquad (0 \leqslant \gamma \leqslant 1)$$

- Weights: $$w_t(i) = \exp\left(-\frac{\gamma}{N}\underbrace{\sum_{s=1}^{t-1}\widehat{x}_s(i)}_{\text{estimated past loss}}\right)$$

- Loss vector estimate: $$\widehat{x}_t(i) = \frac{x_t(i)}{p_t(i)}\mathbb{I}_{\{I_t=i\}}$$

- Key properties: $$\mathbb{E}_t\left[\widehat{x}_t(i)\right] = x_t(i) \qquad \left|\widehat{x}_t(i)\right| \leqslant \frac{N}{\gamma}$$

A playing strategy: Exp3 [Auer, C-B, Freund and Schapire, 1995]

- **Randomization:** At time t, action i is played with probability

$$p_t(i) = (1-\gamma)\underbrace{\frac{w_t(i)}{\sum_j w_t(j)}}_{\text{exploitation}} + \underbrace{\frac{\gamma}{N}}_{\text{exploration}} \qquad (0 \leqslant \gamma \leqslant 1)$$

- **Weights:**

$$w_t(i) = \exp\left(-\frac{\gamma}{N}\underbrace{\sum_{s=1}^{t-1}\widehat{x}_s(i)}_{\text{estimated past loss}}\right)$$

- **Loss vector estimate:**

$$\widehat{x}_t(i) = \frac{x_t(i)}{p_t(i)}\mathbb{I}_{\{I_t=i\}}$$

- **Key properties:**

$$\mathbb{E}_t\left[\widehat{x}_t(i)\right] = x_t(i) \qquad \boxed{\left|\widehat{x}_t(i)\right| \leqslant \frac{N}{\gamma}}$$

Regret bound

[Auer, C-B, Freund and Schapire, 2001]

Theorem

For any sequence $x_1, x_2, \ldots$ *of loss functions*

$$\sum_{t=1}^{T} \mathbb{E}\left[x_t(I_t)\right] - \min_{i=1,\ldots,N} \sum_{t=1}^{T} x_t(i) \leqslant c\sqrt{TN \ln N}$$

Regret bound

[Auer, C-B, Freund and Schapire, 2001]

Theorem

For any sequence $x_1, x_2, \ldots$ *of loss functions*

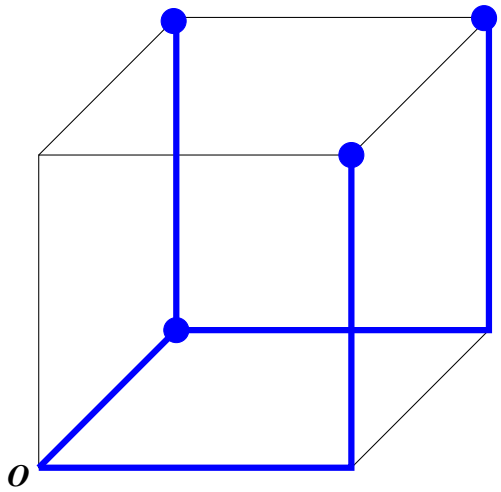
$$\sum_{t=1}^{T} \mathbb{E}\left[x_t(I_t)\right] - \min_{i=1,\ldots,N} \sum_{t=1}^{T} x_t(i) \leqslant c\sqrt{TN \ln N}$$

Remarks

- Factor $\sqrt{T \ln N}$ is necessary even when $x_t(i)$ is known for $i \neq I_t$
- Factor $\sqrt{N}$ is due to **range of estimates** $\left|\widehat{x}_t(i)\right| \leqslant \dfrac{N}{\gamma}$

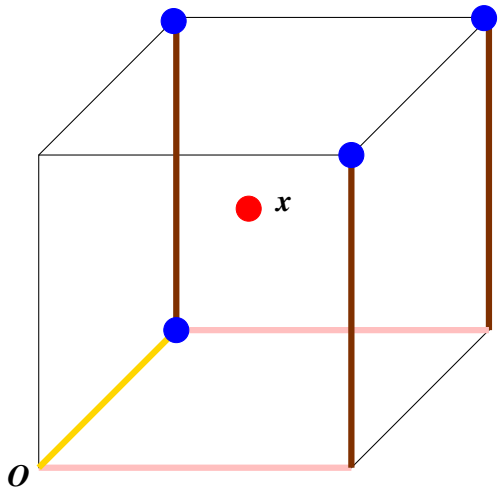
Combinatorial bandits

- Actions are **boolean vectors**: $\boldsymbol{v}(i) \subseteq \{0,1\}^d \qquad i = 1, \ldots, N$



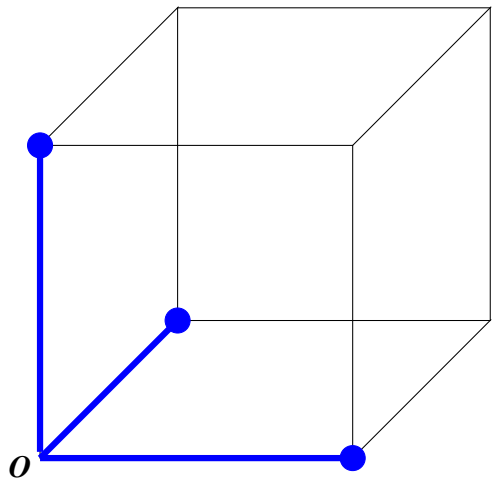
Combinatorial bandits

- Actions are **boolean vectors**: $v(i) \subseteq \{0,1\}^d \qquad i = 1, \ldots, N$
- Losses $x_1, x_2, \cdots \in \mathbb{R}^d$ are **linear**: $v(i)^\top x_t$



Combinatorial bandits

- Actions are **boolean vectors**: $v(i) \subseteq \{0,1\}^d \qquad i = 1, \ldots, N$
- Losses $x_1, x_2, \cdots \in \mathbb{R}^d$ are **linear**: $v(i)^\top x_t$
- Classical bandits: $v(1), \ldots, v(N)$ are the canonical basis of $\mathbb{R}^d$



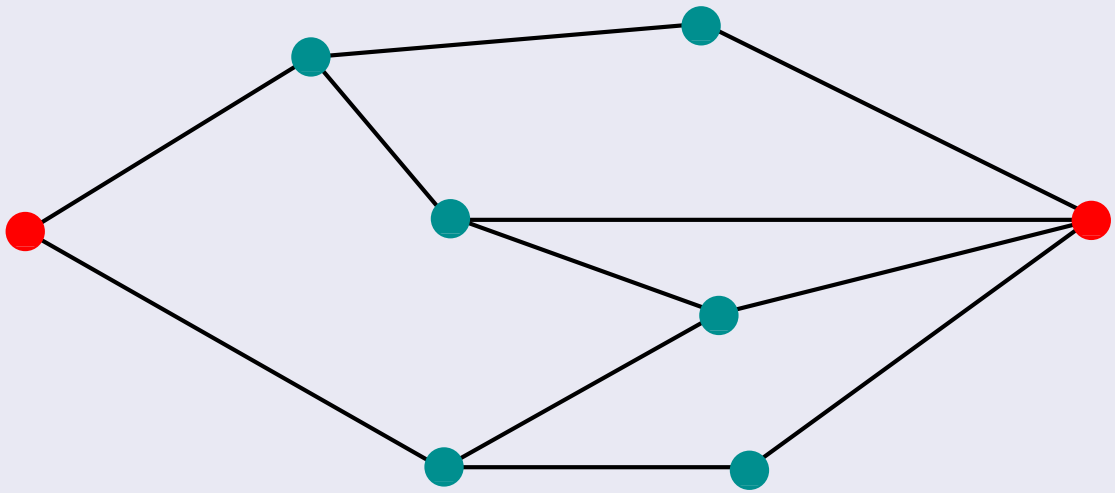
Remarks

- Loss vector estimation problem is now d-dimensional
- Expect regret bound $\sqrt{T d \ln N}$ instead of $\sqrt{TN \ln N}$

Application to multidimensional allocation problems

Multitask bandits $d = 6$

Path selection problems on networks $d = 11$

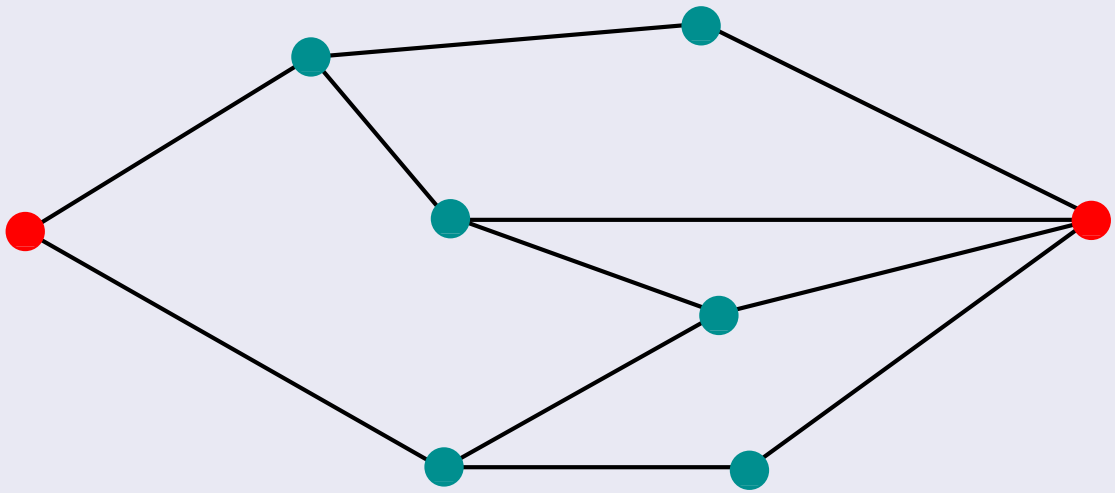


Application to multidimensional allocation problems

Multitask bandits $d = 6$

$$\boldsymbol{v}(I_t) = (1,0,0,\ 0,0,1)$$

Path selection problems on networks $d = 11$

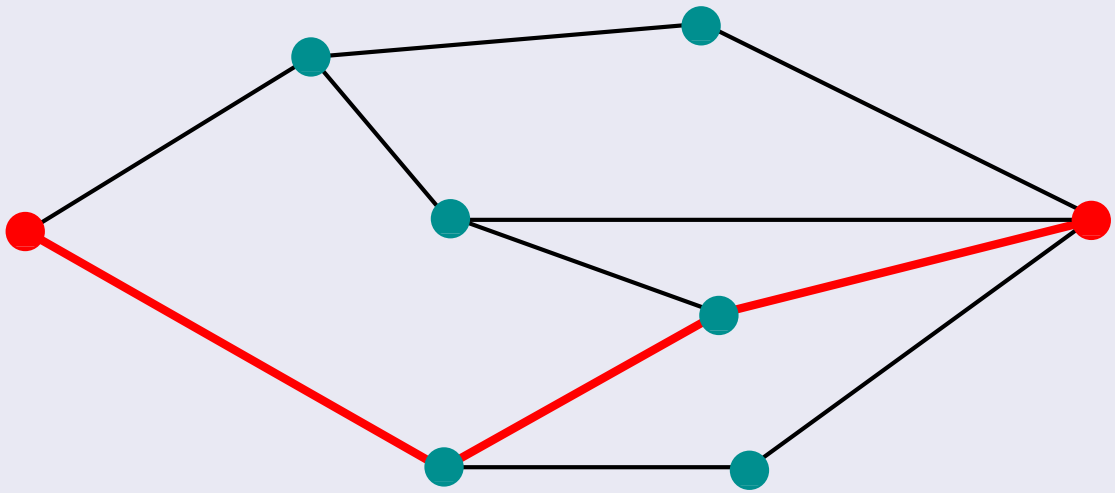


Application to multidimensional allocation problems

Multitask bandits $d = 6$

$$\boldsymbol{v}(I_t) = (1,0,0,\ 0,0,1)$$

Path selection problems on networks $d = 11$

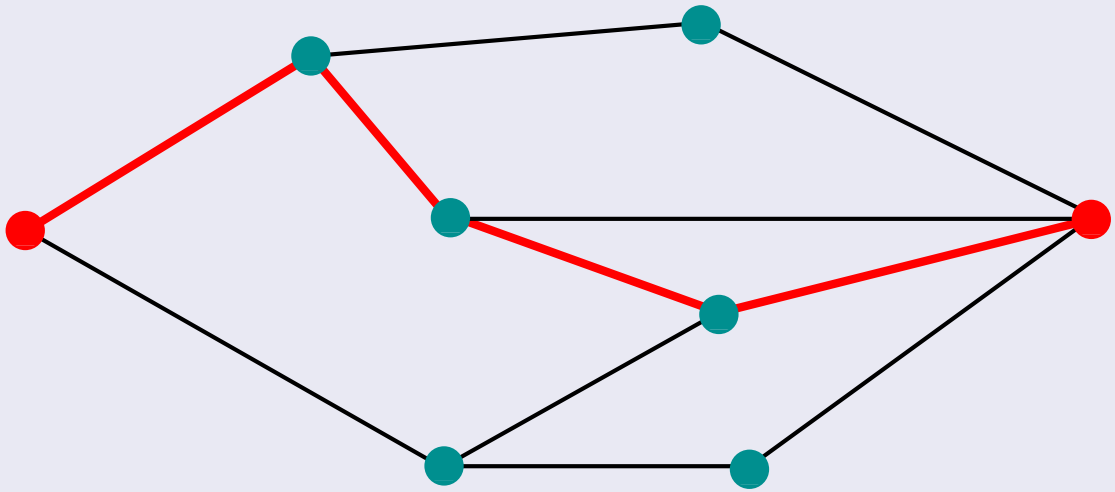


Application to multidimensional allocation problems

Multitask bandits $d = 6$

$$\boldsymbol{v}(I_t) = (1,0,0,\ 0,0,1)$$

Path selection problems on networks $d = 11$



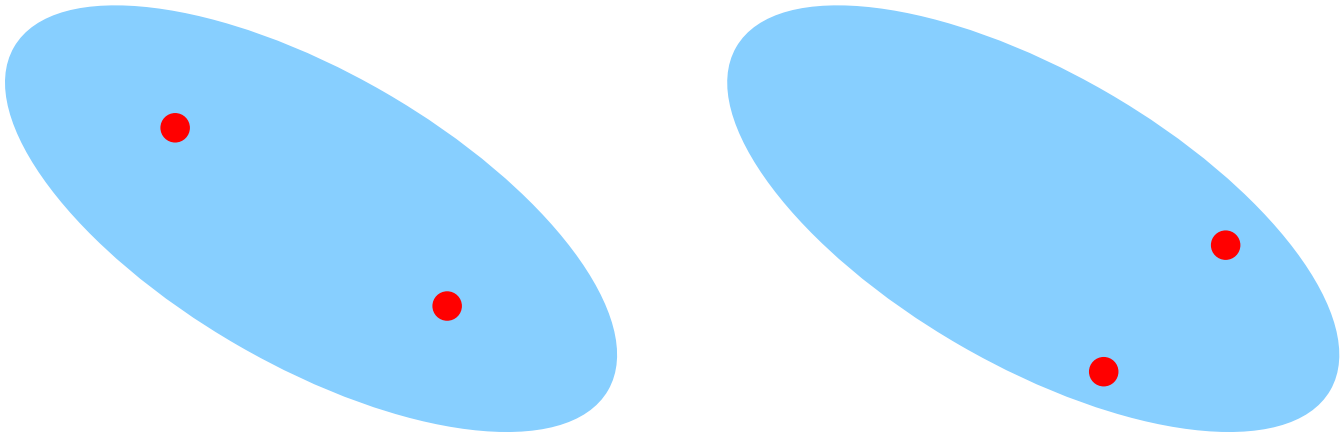
More examples

Actions are combinatorial objects represented by their incidence vectors $\boldsymbol{v}(i) \subseteq \{0,1\}^d$

K-sized subsets of d elements	$\{0,1\}^d$	$N = \binom{d}{K}$
Permutations of K objects	$\{0,1\}^{K^2}$	$N = K!$
Spanning trees of a K-clique	$\{0,1\}^{\binom{K}{2}}$	$N = K^{K-2}$
Balanced cuts of a $2K$-clique	$\{0,1\}^{\binom{2K}{2}}$	$N = \binom{2K}{K}$
Hamiltonian cycles in a K-clique	$\{0,1\}^{\binom{K}{2}}$	$N = \frac{(K-1)!}{2}$

Barycentric spanners

- Regret of `Exp3` is $\sqrt{TN \ln N}$ . . . bad when N is large
- Factor $\sqrt{N}$ comes from range of estimates N/γ controlled by exploration term – we are not exploiting overlappings
- Idea: explore on a subset of the action set and estimate losses of remaining actions by interpolation



A key result for spanners [Awerbuch and Kleinberg, 2004]

Theorem

For any $\mathcal{S} \subseteq \mathbb{R}^d$ there exists a (generally nonorthogonal) basis for the span of $\mathcal{S}$ such that $\|\boldsymbol{v}\|_\infty \leqslant 1$ for all $\boldsymbol{v} \in \mathcal{S}$

A key result for spanners [Awerbuch and Kleinberg, 2004]

Theorem

For any $S \subseteq \mathbb{R}^d$ there exists a (generally nonorthogonal) basis for the span of S such that $\|v\|_\infty \leqslant 1$ for all $v \in S$

Coordinate-free scaling assumption

For any $v \in S$ $\qquad\qquad$ $v^\top x_t \leqslant B$

Algorithm GeometricHedge [Dani, Hayes and Kakade, 2008]

- Draw $\boldsymbol{v}(I_t) \in \{\boldsymbol{v}(1), \ldots, \boldsymbol{v}(N)\}$ with probability

$$p_t(i) = (1-\gamma) \underbrace{\frac{w_t(i)}{\sum_j w_t(j)}}_{\text{exploitation}} + \frac{\gamma}{d} \underbrace{\mathbb{I}_{\{\boldsymbol{v}(i) \text{ is spanner}\}}}_{\text{exploration}} \qquad (0 \leqslant \gamma \leqslant 1)$$

Algorithm GeometricHedge [Dani, Hayes and Kakade, 2008]

- Draw $\boldsymbol{v}(I_t) \in \{\boldsymbol{v}(1), \ldots, \boldsymbol{v}(N)\}$ with probability

$$p_t(i) = (1-\gamma) \underbrace{\frac{w_t(i)}{\sum_j w_t(j)}}_{\text{exploitation}} + \frac{\gamma}{d} \underbrace{\mathbb{I}_{\{\boldsymbol{v}(i) \text{ is spanner}\}}}_{\text{exploration}} \qquad (0 \leqslant \gamma \leqslant 1)$$

- Observe loss $\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t$

Algorithm GeometricHedge [Dani, Hayes and Kakade, 2008]

- Draw $\boldsymbol{v}(I_t) \in \{\boldsymbol{v}(1), \ldots, \boldsymbol{v}(N)\}$ with probability

$$p_t(i) = (1-\gamma) \underbrace{\frac{w_t(i)}{\sum_j w_t(j)}}_{\text{exploitation}} + \frac{\gamma}{d} \underbrace{\mathbb{I}_{\{\boldsymbol{v}(i) \text{ is spanner}\}}}_{\text{exploration}} \qquad (0 \leqslant \gamma \leqslant 1)$$

- Observe loss $\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t$
- Loss vector estimate (least squares):

$$\widehat{\boldsymbol{x}}_t = C_t^+ \boldsymbol{v}(I_t) \boldsymbol{v}(I_t)^\top \boldsymbol{x}_t \qquad \text{where} \qquad C_t = \underbrace{\mathbb{E}_t\left[\boldsymbol{v}(I_t) \boldsymbol{v}(I_t)^\top\right]}_{\text{correlation matrix}}$$

Algorithm GeometricHedge [Dani, Hayes and Kakade, 2008]

- Draw $\boldsymbol{v}(I_t) \in \{\boldsymbol{v}(1), \ldots, \boldsymbol{v}(N)\}$ with probability

$$p_t(i) = (1-\gamma) \underbrace{\frac{w_t(i)}{\sum_j w_t(j)}}_{\text{exploitation}} + \frac{\gamma}{d} \underbrace{\mathbb{I}_{\{\boldsymbol{v}(i) \text{ is spanner}\}}}_{\text{exploration}} \qquad (0 \leqslant \gamma \leqslant 1)$$

- Observe loss $\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t$
- Loss vector estimate (least squares):

$$\widehat{\boldsymbol{x}}_t = C_t^+ \boldsymbol{v}(I_t) \boldsymbol{v}(I_t)^\top \boldsymbol{x}_t \qquad \text{where} \qquad C_t = \underbrace{\mathbb{E}_t\left[\boldsymbol{v}(I_t) \boldsymbol{v}(I_t)^\top\right]}_{\text{correlation matrix}}$$

- Weights: $$w_t(i) = \exp\left(-\frac{\gamma}{N} \sum_{s=1}^{t-1} \widehat{\boldsymbol{x}}_s^\top \boldsymbol{v}(i)\right)$$

GeometricHedge — Properties

- Loss is unbiased

$$\mathbb{E}_t\left[\widehat{\boldsymbol{x}}_t\right] = \mathbb{E}_t\left[C_t^+ \boldsymbol{v}(I_t)\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t\right] = C_t^+ \,\mathbb{E}_t\left[\boldsymbol{v}(I_t)\boldsymbol{v}(I_t)^\top\right]\boldsymbol{x}_t = \boldsymbol{x}_t$$

GeometricHedge — Properties

- Loss is unbiased

$$\mathbb{E}_t\left[\widehat{x}_t\right] = \mathbb{E}_t\left[C_t^+ v(I_t) v(I_t)^\top x_t\right] = C_t^+ \, \mathbb{E}_t\left[v(I_t) v(I_t)^\top\right] x_t = x_t$$

- Size of estimated losses

$$\left|\widehat{x}_t^\top v(i)\right| = \underbrace{v(I_t)^\top x_t}_{} \left|v(I_t)^\top C_t^+ v(i)\right| \leqslant \underbrace{\left\|C_t^+\right\|}_{} \underbrace{\left(\max_{v \in S} \|v\|^2\right)}_{}$$

GeometricHedge — Properties

- Loss is unbiased

$$\mathbb{E}_t\left[\widehat{x}_t\right] = \mathbb{E}_t\left[C_t^+ v(I_t) v(I_t)^\top x_t\right] = C_t^+ \, \mathbb{E}_t\left[v(I_t) v(I_t)^\top\right] x_t = x_t$$

- Size of estimated losses

$$\left|\widehat{x}_t^\top v(i)\right| = \underbrace{v(I_t)^\top x_t}_{\leqslant B} \left|v(I_t)^\top C_t^+ v(i)\right| \leqslant \underbrace{\left\|C_t^+\right\|}\, \underbrace{\left(\max_{v \in S} \|v\|^2\right)}$$

- Using scaling assumption $\quad v^\top x_t \leqslant B$

GeometricHedge — Properties

- Loss is unbiased

$$\mathbb{E}_t\left[\widehat{x}_t\right] = \mathbb{E}_t\left[C_t^+ v(I_t) v(I_t)^\top x_t\right] = C_t^+ \, \mathbb{E}_t\left[v(I_t) v(I_t)^\top\right] x_t = x_t$$

- Size of estimated losses

$$\left|\widehat{x}_t^\top v(i)\right| = \underbrace{v(I_t)^\top x_t}_{\leqslant B} \left|v(I_t)^\top C_t^+ v(i)\right| \leqslant \underbrace{\left\|C_t^+\right\|}_{\leqslant d/\gamma} \underbrace{\left(\max_{v \in \mathcal{S}} \|v\|^2\right)}_{\leqslant d}$$

- Using scaling assumption $v^\top x_t \leqslant B$
- Using spanner property $\|v\|_\infty \leqslant 1$

GeometricHedge — Properties

- Loss is unbiased

$$\mathbb{E}_t\left[\widehat{x}_t\right] = \mathbb{E}_t\left[C_t^+ v(I_t) v(I_t)^\top x_t\right] = C_t^+ \, \mathbb{E}_t\left[v(I_t) v(I_t)^\top\right] x_t = x_t$$

- Size of estimated losses

$$\left|\widehat{x}_t^\top v(i)\right| = \underbrace{v(I_t)^\top x_t}_{\leqslant B} \left|v(I_t)^\top C_t^+ v(i)\right| \leqslant \underbrace{\left\|C_t^+\right\|}_{\leqslant d/\gamma} \underbrace{\left(\max_{v \in S} \|v\|^2\right)}_{\leqslant d}$$

- Using scaling assumption $v^\top x_t \leqslant B$
- Using spanner property $\|v\|_\infty \leqslant 1$

Final regret bound

$$B\, d \sqrt{T \ln N}$$

Algorithm ComBand

[C-B and Lugosi, 2009]

- Draw $\boldsymbol{v}(I_t) \in \{\boldsymbol{v}(1), \ldots, \boldsymbol{v}(N)\}$ with probability

$$p_t(i) = (1-\gamma) \underbrace{\frac{w_t(i)}{\sum_j w_t(j)}}_{\text{exploitation}} + \frac{\gamma}{d} \underbrace{\mathbb{I}_{\{\boldsymbol{v}(i) \text{ is spanner}\}}}_{\text{exploration}} \qquad (0 \leqslant \gamma \leqslant 1)$$

- Observe loss $\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t$
- Loss vector estimate:

$$\widehat{\boldsymbol{x}}_t = C_t^+ \boldsymbol{v}(I_t) \boldsymbol{v}(I_t)^\top \boldsymbol{x}_t \qquad \text{where} \qquad C_t = \mathbb{E}_t\left[\boldsymbol{v}(I_t) \boldsymbol{v}(I_t)^\top\right]$$

- Weights: $$w_t(i) = \exp\left(-\frac{\gamma}{N} \sum_{s=1}^{t-1} \widehat{\boldsymbol{x}}_s^\top \boldsymbol{v}(i)\right)$$

Algorithm ComBand

[C-B and Lugosi, 2009]

- Draw $\boldsymbol{v}(I_t) \in \{\boldsymbol{v}(1), \ldots, \boldsymbol{v}(N)\}$ with probability

$$p_t(i) = (1-\gamma) \underbrace{\frac{w_t(i)}{\sum_j w_t(j)}}_{\text{exploitation}} + \underbrace{\frac{\gamma}{N}}_{\text{exploration}} \qquad (0 \leqslant \gamma \leqslant 1)$$

- Observe loss $\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t$
- Loss vector estimate:

$$\widehat{\boldsymbol{x}}_t = C_t^+ \boldsymbol{v}(I_t)\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t \qquad \text{where} \qquad C_t = \mathbb{E}_t\left[\boldsymbol{v}(I_t)\boldsymbol{v}(I_t)^\top\right]$$

- Weights: $$w_t(i) = \exp\left(-\frac{\gamma}{N}\sum_{s=1}^{t-1} \widehat{\boldsymbol{x}}_s^\top \boldsymbol{v}(i)\right)$$

ComBand — Properties

Coordinate-dependent scaling assumption

For $S \subseteq \{0,1\}^d$ $\quad \|\boldsymbol{v}\|_1 \leqslant B \quad \|\boldsymbol{x}_t\|_\infty \leqslant 1$

ComBand — Properties

Coordinate-dependent scaling assumption

For $\mathcal{S} \subseteq \{0,1\}^d$ $\qquad \|\boldsymbol{v}\|_1 \leqslant B \qquad \|\boldsymbol{x}_t\|_\infty \leqslant 1$

Size of estimated losses

$$\left|\widehat{\boldsymbol{x}}_t^\top \boldsymbol{v}(i)\right| = \underbrace{\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t}_{} \quad \left|\boldsymbol{v}(I_t)^\top C_t^+ \boldsymbol{v}(i)\right| \leqslant B \quad \underbrace{\left\|C_t^+\right\|}_{} \quad \underbrace{\left(\max_{\boldsymbol{v}\in\mathcal{S}} \|\boldsymbol{v}\|^2\right)}_{}$$

ComBand — Properties

Coordinate-dependent scaling assumption

For $S \subseteq \{0,1\}^d$ $\qquad \|v\|_1 \leqslant B \qquad \|x_t\|_\infty \leqslant 1$

Size of estimated losses

$$\left|\widehat{x}_t^\top v(i)\right| = \underbrace{v(I_t)^\top x_t}_{\leqslant B} \left|v(I_t)^\top C_t^+ v(i)\right| \leqslant B \quad \underbrace{\left\|C_t^+\right\|}_{} \quad \underbrace{\left(\max_{v \in S} \|v\|^2\right)}_{}$$

- Using scaling assumption $\|v\|_1 \leqslant B$ and $\|x_t\|_\infty \leqslant 1$

ComBand — Properties

Coordinate-dependent scaling assumption

For $S \subseteq \{0,1\}^d$ $\quad \|v\|_1 \leqslant B \quad \|x_t\|_\infty \leqslant 1$

Size of estimated losses

$$\left|\widehat{x}_t^\top v(i)\right| = \underbrace{v(I_t)^\top x_t}_{\leqslant B} \left|v(I_t)^\top C_t^+ v(i)\right| \leqslant B \underbrace{\left\|C_t^+\right\|}_{\leqslant 1/(\gamma\lambda_{\min})} \underbrace{\left(\max_{v\in S} \|v\|^2\right)}$$

- Using scaling assumption $\|v\|_1 \leqslant B$ and $\|x_t\|_\infty \leqslant 1$
- $\lambda_{\min}$ = smallest eigenval of $\mathbb{E}\left[v(I_t)v(I_t)^\top\right]$ w.r.t. the uniform dist.

ComBand — Properties

Coordinate-dependent scaling assumption

For $\mathcal{S} \subseteq \{0,1\}^d$ $\qquad \|\boldsymbol{v}\|_1 \leqslant B \qquad \|\boldsymbol{x}_t\|_\infty \leqslant 1$

Size of estimated losses

$$\left|\widehat{\boldsymbol{x}}_t^\top \boldsymbol{v}(i)\right| = \underbrace{\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t}_{\leqslant B} \left|\boldsymbol{v}(I_t)^\top C_t^+ \boldsymbol{v}(i)\right| \leqslant B \quad \underbrace{\left\|C_t^+\right\|}_{\leqslant 1/\left(\gamma \lambda_{\min}\right)} \underbrace{\left(\max_{\boldsymbol{v}\in\mathcal{S}} \|\boldsymbol{v}\|^2\right)}_{\leqslant B}$$

- Using scaling assumption $\|\boldsymbol{v}\|_1 \leqslant B$ and $\|\boldsymbol{x}_t\|_\infty \leqslant 1$
- $\lambda_{\min} =$ smallest eigenval of $\mathbb{E}\left[\boldsymbol{v}(I_t)\boldsymbol{v}(I_t)^\top\right]$ w.r.t. the uniform dist.
- $\boldsymbol{v} \in \{0,1\}^d \quad \rightarrow \quad \|\boldsymbol{v}\|^2 \leqslant \|\boldsymbol{v}\|_1 \leqslant B$

Optimality

Regret bound

$$B \sqrt{\left(1 + \frac{B}{d\,\lambda_{\min}}\right) T d \ln N}$$

Optimality

Regret bound

$$B \sqrt{\left(1 + \frac{B}{d\,\lambda_{\min}}\right) T d \ln N}$$

Theorem

In all previous examples* of combinatorial bandits the condition $\lambda_{\min} = \Omega(B/d)$ holds

---

* but paths! ☹

Optimality

Regret bound

$$B \sqrt{\left(1 + \frac{B}{d\,\lambda_{\min}}\right) T d \ln N}$$

Theorem

In all previous examples* of combinatorial bandits the condition $\lambda_{\min} = \Omega(B/d)$ holds

---

* but paths! ☹

Corollary

- In these examples the regret becomes $O\left(B\sqrt{T d \ln N}\right)$
- This is not improvable in general for $\mathcal{S} \subseteq \{0,1\}^d$

How many parameters?

Recall

Loss vector estimates $\widehat{\boldsymbol{x}}_t = C_t^+ \boldsymbol{v}(I_t)\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t$

- Maintain d weights over coordinates $j = 1, \dots, d$

$$w_{j,t} = \exp\left(-\frac{\gamma}{N}\sum_{s=1}^{t-1} \widehat{x}_{j,t}\right)$$

How many parameters?

Recall

Loss vector estimates $\widehat{\boldsymbol{x}}_t = C_t^+ \boldsymbol{v}(I_t)\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t$

- Maintain d weights over coordinates $j = 1, \ldots, d$

$$w_{j,t} = \exp\left(-\frac{\gamma}{N}\sum_{s=1}^{t-1} \widehat{x}_{j,t}\right)$$

- . . . translate them into weights over actions $i = 1, \ldots, N$

$$w_t(i) = \prod_{j\,:\,v_j(i)=1} w_{j,t}$$

How many parameters?

Recall

Loss vector estimates $\widehat{\boldsymbol{x}}_t = C_t^+ \boldsymbol{v}(I_t)\boldsymbol{v}(I_t)^\top \boldsymbol{x}_t$

- Maintain d weights over coordinates $j = 1, \dots, d$

$$w_{j,t} = \exp\left(-\frac{\gamma}{N}\sum_{s=1}^{t-1} \widehat{x}_{j,t}\right)$$

- . . . translate them into weights over actions $i = 1, \dots, N$

$$w_t(i) = \prod_{j\,:\,v_j(i)=1} w_{j,t}$$

Problem

How to efficiently sample from $w_t(1), \dots, w_t(N)$?

Efficient sampling

K bandit problems	☺	independent draws from each bandit
K-sized subsets	☺	compute conditionals using dynamic programming
Permutations of K objects	☺	random sampling of perfect matchings
Spanning trees of a K-clique	☹	OK for uniform, unknown for weighted
Balanced cuts of a $2K$-clique	☺	sampling ferromagnetic Ising model
Hamiltonian cycles in a K-clique	☹	notoriously hard

A more general framework

Reduction to online gradient descent [Zinkevich, 2003]

1. Pick $\mathbf{p}_t$ from the convex hull of $\mathcal{S}$ and use it to draw action $I_t \in \mathcal{S}$
2. Suffer loss $\boldsymbol{\nu}(I_t)^\top \mathbf{x}_t$ and observe loss $\mathbf{p}_t^\top \mathbf{x}_t$
3. Compute loss vector estimate $\widehat{\mathbf{x}}_t$
4. Perform gradient update: $\mathbf{p}_{t+1} = \mathbf{p}_t - \eta \widehat{\mathbf{x}}_t$

A more general framework

Reduction to online gradient descent [Zinkevich, 2003]

1. Pick $\mathbf{p}_t$ from the convex hull of $\mathcal{S}$ and use it to draw action $I_t \in \mathcal{S}$
2. Suffer loss $\boldsymbol{\nu}(I_t)^\top \mathbf{x}_t$ and observe loss $\mathbf{p}_t^\top \mathbf{x}_t$
3. Compute loss vector estimate $\widehat{\mathbf{x}}_t$
4. Perform gradient update: $\mathbf{p}_{t+1} = \mathbf{p}_t - \eta \widehat{\mathbf{x}}_t$

Issues

Sampling problem: need to draw I_t so that $\mathbb{E}\left[\boldsymbol{\nu}(I_t)^\top \mathbf{x}_t\right] = \mathbf{p}_t^\top \mathbf{x}_t$

Point $\mathbf{p}_t$ must simultaneously

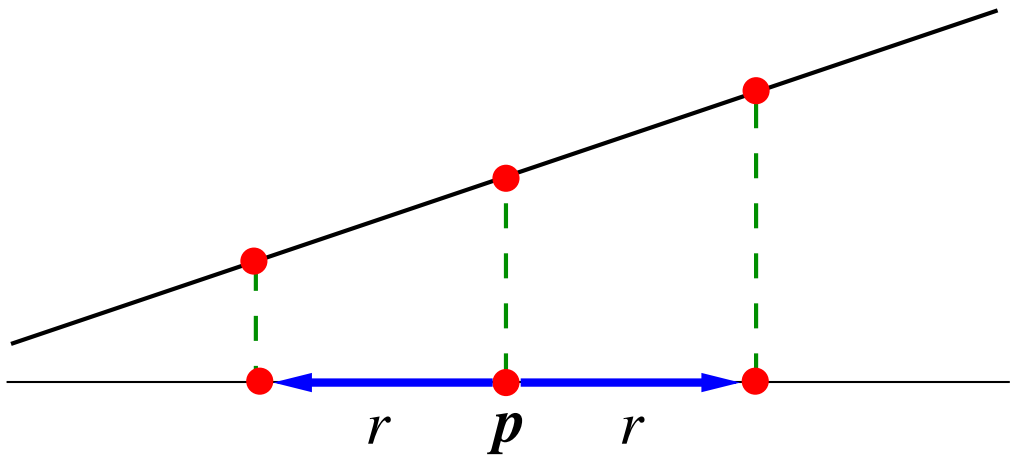
1. have small loss $\mathbf{p}_t^\top \mathbf{x}_t$ (exploitation)
2. lead to a good estimate $\widehat{\mathbf{x}}_t$ (exploration)

Gradient descent without a gradient

[Flaxman, Kalai and McMahan, 2004]

- Play $\mathbf{p}_t + r\,\mathbf{u}$ where $\mathbf{u}$ is a random unit vector and $r > 0$
- Loss vector estimate: $\widehat{\mathbf{x}}_t = d\big(\mathbf{p}_t + r\,\mathbf{u}\big)^\top \mathbf{x}_t \, \dfrac{\mathbf{u}}{r}$
- In one dimension

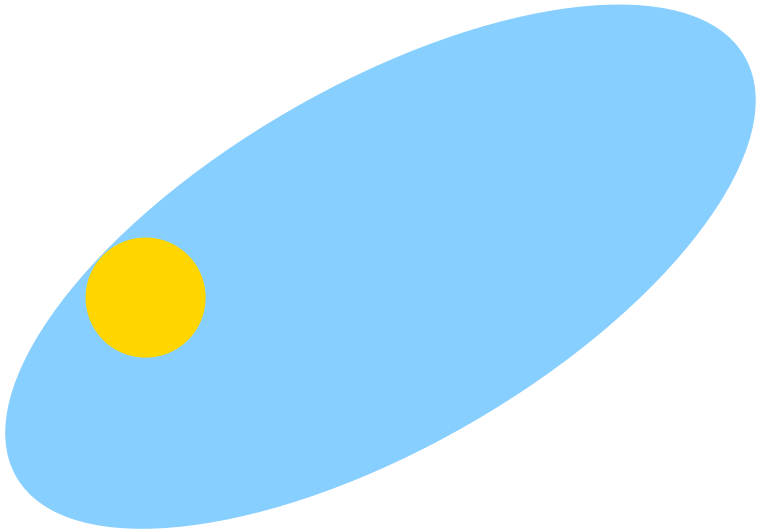
$$\mathbb{E}\big[\widehat{\mathbf{x}}_t\big] = \frac{1}{2}\frac{\big(\mathbf{p}_t + r\big)\mathbf{x}_t}{r} + \frac{1}{2}\frac{\big(\mathbf{p}_t - r\big)\mathbf{x}_t}{r} = \mathbf{x}_t$$



Issues

- $r \in \mathbb{R}$ in $\mathbf{p}_t + r\,\mathbf{u}$ determines the sizes of estimated losses
- r cannot be too big when $\mathbf{p}_t$ is close to a border
- r must be big enough to control the estimates

$$\widehat{\mathbf{x}}_t = d\big(\mathbf{p}_t + r\,\mathbf{u}\big)^\top \mathbf{x}_t \,\frac{\mathbf{u}}{r}$$

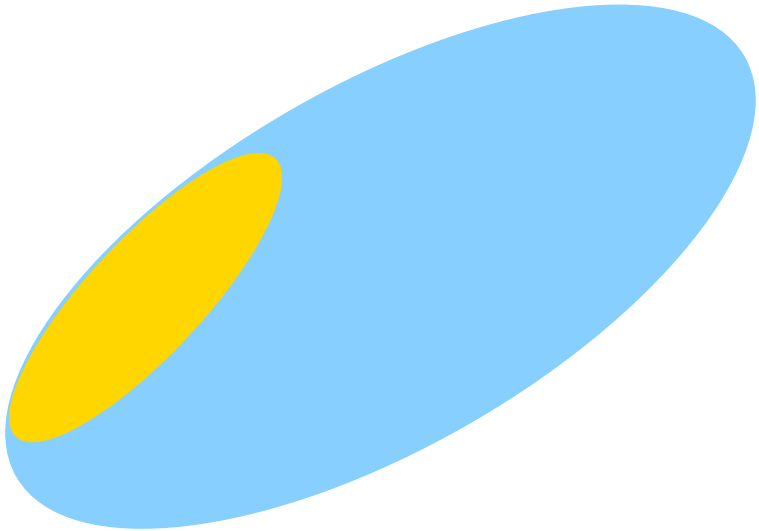


Issues

- $r \in \mathbb{R}$ in $\mathbf{p}_t + r\,\mathbf{u}$ determines the sizes of estimated losses
- r cannot be too big when $\mathbf{p}_t$ is close to a border
- r must be big enough to control the estimates

$$\widehat{\mathbf{x}}_t = d\big(\mathbf{p}_t + r\,\mathbf{u}\big)^\top \mathbf{x}_t \, \frac{\mathbf{u}}{r}$$

- We need to exploit the **local geometry**

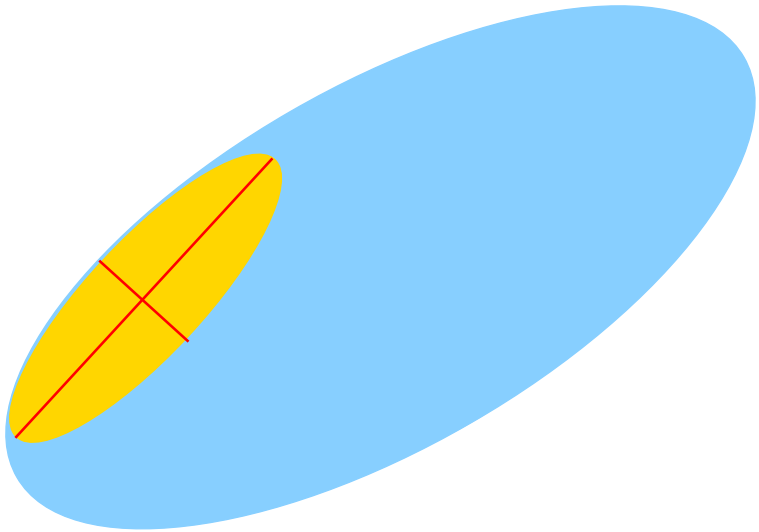


Self-concordant functions [Abernethy, Hazan and Rakhlin, 2008]

- Barrier functions in interior point optimization
- Fact: any convex closed set $\mathcal{K}$ admits a self-concordant function F (may be hard to find)

Self-concordant functions [Abernethy, Hazan and Rakhlin, 2008]

- Barrier functions in interior point optimization
- Fact: any convex closed set $\mathcal{K}$ admits a self-concordant function F (may be hard to find)
- The Hessian matrix $\nabla^2 F$ of a self-concordant function F defines the Dikin ellipsoid which is always contained in $\mathcal{K}$ (our action space)
- Loss vector estimate: $\widehat{\mathbf{x}}_t = \mathbf{p}_t \pm \mathbf{e}_i \sqrt{\lambda_i}$
  where $\{\mathbf{e}_i, \lambda_i\}$ is a random eigenvector-eigenvalue pair of $\nabla^2 F$



Summary for combinatorial bandits

Regrets

GeometricHedge	ComBand	SelfConcordant
$Bd\sqrt{T \ln N}$	$B\sqrt{Td \ln N}$	$Bd^{3/2}\sqrt{T \ln T}$

Summary for combinatorial bandits

Regrets

GeometricHedge	ComBand	SelfConcordant
$Bd\sqrt{T \ln N}$	$B\sqrt{Td \ln N}$	$Bd^{3/2}\sqrt{T \ln T}$

Efficiency

- All: efficiency of the associated sampling problem
- SelfConcordant efficiency of computation of the self-concordant function

Summary for combinatorial bandits

Regrets

GeometricHedge	ComBand	SelfConcordant
$B d \sqrt{T \ln N}$	$B \sqrt{T d \ln N}$	$B d^{3/2} \sqrt{T \ln T}$

Efficiency

- All: efficiency of the associated sampling problem
- SelfConcordant efficiency of computation of the self-concordant function

Remark

Typically: $\ln N = \mathcal{O}\left(\sqrt{d} \ln d\right)$

Summary for combinatorial bandits

Regrets

GeometricHedge	ComBand	SelfConcordant
$B d^{5/4} \sqrt{T \ln d}$	$B d^{3/4} \sqrt{T \ln d}$	$B d^{3/2} \sqrt{T \ln T}$

Efficiency

- All: efficiency of the associated sampling problem
- SelfConcordant efficiency of computation of the self-concordant function

Remark

Typically: $\ln N = \mathcal{O}\left(\sqrt{d} \ln d\right)$

Conclusions

- When $\mathcal{S} \subseteq \{0,1\}^d$, ComBand improves on GeometricHedge and SelfConcordant by exploiting structure of $\mathcal{S}$
- In problems like path selection $\lambda_{\min}$ may get too small
  $\rightarrow$ CombBand is suboptimal
- Fix in progress: Perform exploration using a distribution that maximizes $\lambda_{\min}$
- Is this optimal in general?