# On efficient optimal algorithms for bandit problems


Sasha Rakhlin

University of Pennsylvania


February 13, 2009

Joint work with Jacob Abernethy

# The Problem

Given $\mathcal{K}, \mathcal{F} \subset \mathbb{R}^n$, compact, convex.

At each time step $t = 1$ to $T$,

- Player chooses $\mathbf{x}_t \in \mathcal{K}$
- Adversary independently chooses $\mathbf{f}_t \in \mathcal{F}$
- Player observes the cost $\mathbf{f}_t^\top \mathbf{x}_t$

Goal: minimize *regret* against any "comparator" $\mathbf{u} \in \mathcal{K}$

$$\mathrm{Reg}_T(\mathbf{u}) := \sum_{t=1}^{T} \mathbf{f}_t^\top \mathbf{x}_t - \sum_{t=1}^{T} \mathbf{f}_t^\top \mathbf{u}.$$

# The Problem

Given $\mathcal{K}, \mathcal{F} \subset \mathbb{R}^n$, compact, convex.

At each time step $t = 1$ to $T$,

- Player chooses $\mathbf{x}_t \in \mathcal{K}$
- Adversary independently chooses $\mathbf{f}_t \in \mathcal{F}$
- Player observes the cost $\mathbf{f}_t^\top \mathbf{x}_t$

Goal: minimize *regret* against any "comparator" $\mathbf{u} \in \mathcal{K}$

$$\mathrm{Reg}_T(\mathbf{u}) := \sum_{t=1}^{T} \mathbf{f}_t^\top \mathbf{x}_t - \sum_{t=1}^{T} \mathbf{f}_t^\top \mathbf{u}.$$

Holy grail: $\mathrm{Reg}_T(\mathbf{u}) = \tilde{O}(\sqrt{T})$ for all $\mathbf{u} \in \mathcal{K}$

# Previous results

| | Opt? | Efficient? | Set $\mathcal{K}$ | Adptve? | HighProb? |
|---|---|---|---|---|---|
| Auer Cesa-Bianchi Freund Schapire 02 | ✓ | ✓ | Simplex | ✓ | ✓ |
| McMahan and Blum 04 | 🚫 | ✓ | Any | ✓ | 🚫 |
| Awerbuch and Kleinberg 04 | 🚫 | ✓ | Flows | ✓ 🚫 | 🚫 |
| Flaxman, Kalai, McMahan 05 | 🚫 | ✓ | Any | ✓ | 🚫 |
| György et al 07 | 🚫 | ✓ | Flows | ✓ | ✓ |
| Dani, Hayes, Kakade 07 | ✓ | 🚫 | Any | 🚫 | 🚫 |
| Bartlett et al 08 | ✓ | 🚫 | Any | ✓ | ✓ |
| Abernethy, Hazan, Rakhlin 08 | ✓ | ✓ | Any | 🚫 | 🚫 |
| Abernethy and Rakhlin 09 | ✓ | ✓ | Smplx, Sphere, + | ✓ | ✓ |

# Full Information

Given $\mathcal{K}, \mathcal{F} \subset \mathbb{R}^n$, compact, convex.

At each time step $t = 1$ to $T$,

- Player chooses $\mathbf{x}_t \in \mathcal{K}$
- Adversary independently chooses $\mathbf{f}_t \in \mathcal{F}$
- Player observes the cost $\mathbf{f}_t^\mathsf{T} \mathbf{x}_t$

# Full Information

Given $\mathcal{K}, \mathcal{F} \subset \mathbb{R}^n$, compact, convex.

At each time step $t = 1$ to $T$,

- Player chooses $\mathbf{x}_t \in \mathcal{K}$
- Adversary independently chooses $\mathbf{f}_t \in \mathcal{F}$
- Player observes the cost ~~$\mathbf{f}_t^\mathsf{T} \mathbf{x}_t$~~ $\mathbf{f}_t$

# Full Information

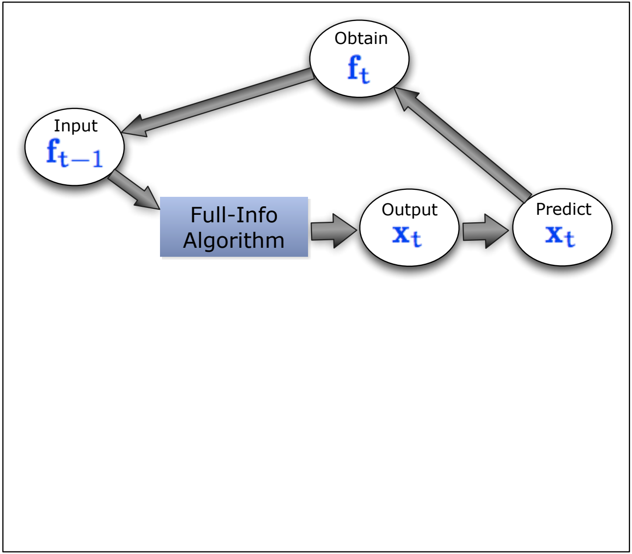
Given $\mathcal{K}, \mathcal{F} \subset \mathbb{R}^n$, compact, convex.

At each time step $t = 1$ to $T$,

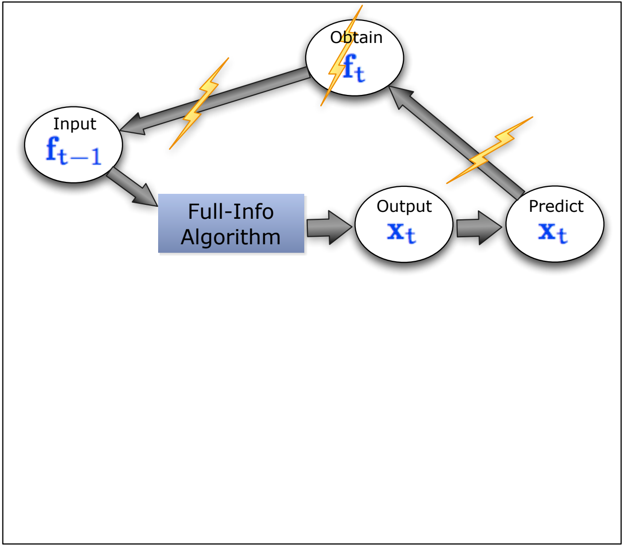
- Player chooses $\mathbf{x}_t \in \mathcal{K}$
- Adversary independently chooses $\mathbf{f}_t \in \mathcal{F}$
- Player observes the cost ~~$\mathbf{f}_t^T \mathbf{x}_t$~~ $\mathbf{f}_t$

Known algorithms achieve $\text{Reg}_T = O(\sqrt{T})$

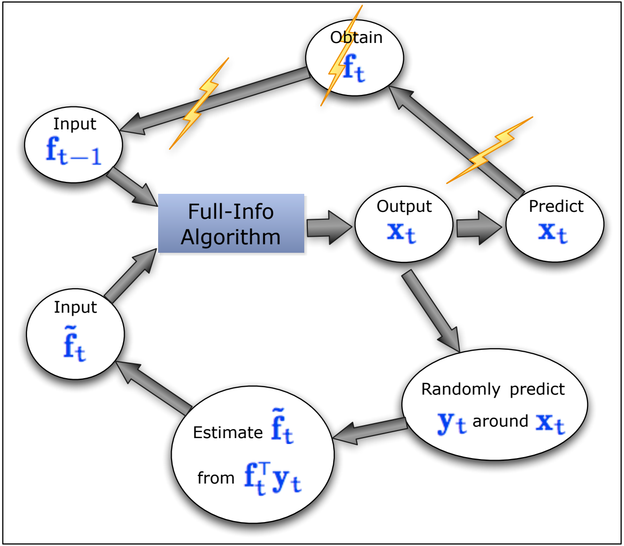
# Black-box reduction



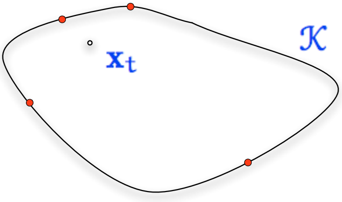
# Black-box reduction



# Black-box reduction



# Three ingredients




- Which black box to use?
- How to sample $\mathbf{y}_t$?
- How to construct $\tilde{\mathbf{f}}_t$?

# Which black-box?

Family of Follow the Regularized Leader algorithms:

$$\mathbf{x}_{t+1} := \arg\min_{\mathbf{x} \in \mathcal{K}} \left[ \eta \sum_{s=1}^{t} \tilde{\mathbf{f}}_s^\mathsf{T} \mathbf{x} + \mathcal{R}(\mathbf{x}) \right].$$

Known:

$$\mathrm{Reg}_T(\mathbf{u}) \lesssim \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \|\tilde{\mathbf{f}}_t\|^{*2}$$

if $\mathcal{R}$ is strongly convex in a norm $\|\cdot\|$.

# Which black-box?

Family of Follow the Regularized Leader algorithms:

$$\mathbf{x}_{t+1} := \arg\min_{\mathbf{x}\in\mathcal{K}} \left[ \eta \sum_{s=1}^{t} \tilde{\mathbf{f}}_s^{\mathsf{T}} \mathbf{x} + \mathcal{R}(\mathbf{x}) \right].$$

Known:

$$\mathrm{Reg}_T(\mathbf{u}) \lesssim \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \|\tilde{\mathbf{f}}_t\|^{*2}$$

if $\mathcal{R}$ is strongly convex in a norm $\|\cdot\|$.

Strong convexity is not good enough !

# Which black-box?

Family of Follow the Regularized Leader algorithms:

$$\mathbf{x}_{t+1} := \arg\min_{\mathbf{x}\in\mathcal{K}} \left[ \eta \sum_{s=1}^{t} \tilde{\mathbf{f}}_s^\top \mathbf{x} + \mathcal{R}(\mathbf{x}) \right].$$

Pick $\mathcal{R}$ such that

$$\mathrm{Reg}_T(\mathbf{u}) \lesssim \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \|\tilde{\mathbf{f}}_t\|_t^{*2}$$

where

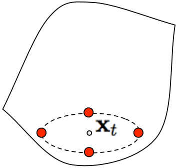
$$\|\mathbf{h}\|_t = \sqrt{\mathbf{h}^\top \nabla^2 \mathcal{R}(\mathbf{x}_t) \mathbf{h}}$$

# First ingredient: Regularizer

- Negative entropy for simplex
- Self-concordant barriers for general bodies

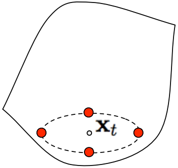
# Second ingredient: Sampling Scheme

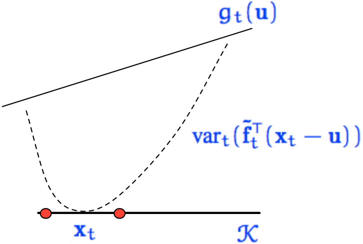
The scheme from (Abernethy, Hazan, Rakhlin 08) does not work for high probability:



# Second ingredient: Sampling Scheme

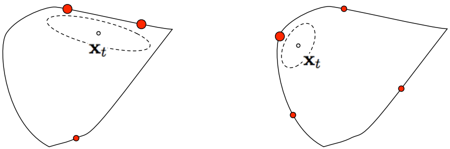
The scheme from (Abernethy, Hazan, Rakhlin 08) does not work for high probability:



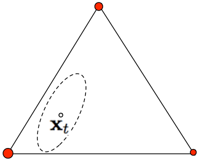



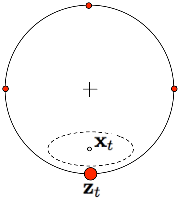
# Second ingredient: Sampling Scheme

Reduce variance:



# Second ingredient: Sampling Scheme



# 3rd ingredient: unbiased estimate and upper bound on variance

Construct

$$\mathbb{E}\tilde{\mathbf{f}}_t = \mathbf{f}_t$$

and a linear function $g_t(\mathbf{u})$ such that

$$(\mathbf{x}_t - \mathbf{u})^\top \mathbb{E}_t \tilde{\mathbf{f}}_t \tilde{\mathbf{f}}_t^\top (\mathbf{x}_t - \mathbf{u}) \leqslant g_t(\mathbf{u}) \quad \forall \mathbf{u} \in \mathcal{K}$$

and

$$g_t(\mathbf{x}_t) = O(1)$$

# Theorem

If have all three ingredients and $\approx 4$ more conditions are satisfied, then with high probability, for all $\mathbf{u} \in \mathcal{K}$

$$\mathrm{Reg}_{\mathsf{T}}(\mathbf{u}) = \tilde{\mathsf{O}}(\sqrt{\mathsf{T}}).$$

# Theorem

Suppose $\mathbf{f}_t \in B_p$ for all $t$ and $\mathcal{K} \subseteq B_q$, where $p$ and $q$ are dual. Let $\alpha = \sqrt{\frac{\log(2\log(T)/\delta')}{nT}}$. Suppose all of the following hold:

(A) The black-box full information algorithm enjoys a regret bound of the form

$$\mathrm{Reg}_T(\mathbf{u}) \leqslant c_1 \eta \sum_{t=1}^{T} [\|\mathbf{f}_t\|_t^*]^2 + \eta^{-1}\mathcal{R}(\mathbf{u})$$

(B) $\|\mathbb{E}_t \mathbf{y}_t - \mathbf{x}_t\|_q \leqslant c_2 \sqrt{\frac{n}{T}}$.

(C) $|\tilde{\mathbf{f}}_t^{\mathsf{T}} \mathbf{u}| \leqslant c_3 \sqrt{nT}$ for all $\mathbf{u} \in \mathcal{K}$.

(D) We can construct a linear function $g_t(\mathbf{u}) = \tilde{\mathbf{g}}_t^{\mathsf{T}} \mathbf{u} + \mu_t$ such that

$$\mathrm{var}_t(\tilde{\mathbf{f}}_t^{\mathsf{T}}(\mathbf{x}_t - \mathbf{u})) \leqslant g_t(\mathbf{u}) \quad \forall \mathbf{u} \in \mathcal{K} \quad \text{and} \quad g_t(\mathbf{x}_t) \leqslant c_4 n.$$

(E) Construction satisfies $\left[\left\|\tilde{\mathbf{f}}_t - \alpha\tilde{\mathbf{g}}_t\right\|_t^*\right]^2 \leqslant c_5\sqrt{T}$, $\mathbb{E}_t\left[\left\|\tilde{\mathbf{f}}_t - \alpha\tilde{\mathbf{g}}_t\right\|_t^*\right]^2 \leqslant c_6$.

Then, for any fixed $\mathbf{u} \in \mathcal{K}$, with probability at least $1 - (\delta + \delta' + \delta'')$

$$\sum_{t=1}^{T} \mathbf{f}_t^{\mathsf{T}}(\mathbf{y}_t - \mathbf{u}) \leqslant \eta^{-1}\mathcal{R}(\mathbf{u}) + \eta T A_1 + \sqrt{T} A_2$$

# Contributions

- First efficient optimal algorithm for bandit optimization against adaptive adversary over sphere.
- General framework for achieving similar results for arbitrary convex bodies.
- Showed "local norm" bounds for entropy and self-concordant barrier regularization, unifying these.
- Decoupled sampling scheme and estimation from the black-box optimization method.
- Replaced $\sqrt{T \log T}$ in the bound of (Auer et al 2002) for multiarmed bandit by $\sqrt{T} \log \log T$ (lower bound is $\Omega(\sqrt{T})$)