

# Online Learning with Limited Feedback

Sasha Rakhlin

UC Berkeley, UPenn Stats

September 26, 2008

Joint work with Jacob Abernethy and Elad Hazan

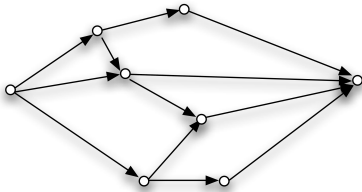
# Outline

- 1 The Problem
- 2 Difficulties
- 3 Solution
- 4 Conclusions

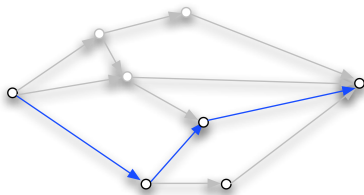
# Outline

- 1 The Problem
- 2 Difficulties
- 3 Solution
- 4 Conclusions

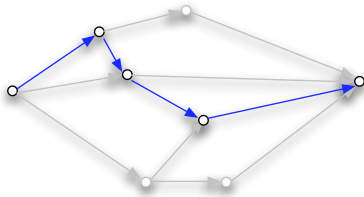
# Driving to Work



# Driving to Work



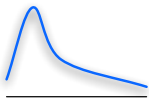
# Driving to Work



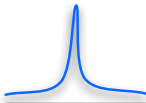
# Multiarmed bandit



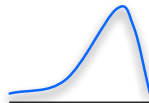
\$ -2, 2, -1, 0



\$ 1, 1, 1



\$ 4, 5, -1, 4



# Adversarial continuum-armed bandit

$\mathcal{K} \subset \mathbb{R}^K$  some compact convex set.

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

- Player plays  $\mathbf{x}_t \in \mathcal{K}$
- Adversary (simultaneously) chooses  $\mathbf{f}_t \in [0, 1]^K$
- Player suffers loss  $\mathbf{f}_t^\top \mathbf{x}_t$
- Only  $\mathbf{f}_t^\top \mathbf{x}_t$  is revealed

Regret is

$$\mathcal{R}_T = \sum_{t=1}^T \mathbf{f}_t^\top \mathbf{x}_t - \min_{\mathbf{x} \in \mathcal{K}} \mathbb{E} \sum_{t=1}^T \mathbf{f}_t^\top \mathbf{x}$$

# Adversarial continuum-armed bandit

$\mathcal{K} \subset \mathbb{R}^K$  some compact convex set.

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

- Player chooses a distribution over  $\mathcal{K}$ , draws  $\mathbf{x}_t \in \mathcal{K}$
- Adversary (simultaneously) chooses  $\mathbf{f}_t \in [0, 1]^K$
- Player suffers loss  $\mathbf{f}_t^\top \mathbf{x}_t$
- Only  $\mathbf{f}_t^\top \mathbf{x}_t$  is revealed

Regret is

$$\mathcal{R}_T = \sum_{t=1}^T \mathbf{f}_t^\top \mathbf{x}_t - \min_{\mathbf{x} \in \mathcal{K}} \mathbb{E} \sum_{t=1}^T \mathbf{f}_t^\top \mathbf{x}$$

# Adversarial continuum-armed bandit

$\mathcal{K} \subset \mathbb{R}^K$  some compact convex set.

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

- Player chooses a distribution over  $\mathcal{K}$ , draws  $\mathbf{x}_t \in \mathcal{K}$
- Adversary (simultaneously) chooses  $\mathbf{f}_t \in [0, 1]^K$
- Player suffers loss  $\mathbf{f}_t^\top \mathbf{x}_t$
- Only  $\mathbf{f}_t^\top \mathbf{x}_t$  is revealed

Regret is

$$\mathcal{R}_T = \sum_{t=1}^T \mathbf{f}_t^\top \mathbf{x}_t - \min_{\mathbf{x} \in \mathcal{K}} \mathbb{E} \sum_{t=1}^T \mathbf{f}_t^\top \mathbf{x}$$

Case of  $\mathcal{K}$  being the  $K$ -probability simplex was solved by (Auer, Cesa-Bianchi, Freund, & Schapire 94);  $\mathcal{R}_T = O(\sqrt{T \log T})$

# Previous work on adversarial continuum-armed bandit

- $O(T^{2/3})$  Awerbuch and Kleinberg, 2004
- $O(T^{3/4})$  McMahan and Blum, 2004
- $O(T^{3/4})$  Flaxman, Kalai, and McMahan, 2005
- $O(T^{2/3})$  Dani and Hayes, 2005
- $O(T^{2/3})$  György, Linder, Lugosi, and Ottucsák, 2007

Finally,

- $O(\sqrt{T})$  (in expectation) Dani, Hayes, and Kakade, 2007
- $O(\sqrt{T})$  (with high probability) Bartlett, Dani, Hayes, Kakade, Rakhlin, Tewari 2008)

by a reduction to  $K$ -armed bandit. **Exponential running time...**

# Outline

- 1 The Problem
- 2 Difficulties**
- 3 Solution
- 4 Conclusions

# Simpler problem: Online Convex Optimization

Let  $\mathcal{K} \subset \mathbb{R}^k$  be some compact convex set.

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

- Player chooses  $\mathbf{x}_t \in \mathcal{K}$
- Adversary chooses convex  $f_t(\mathbf{x}) : \mathbb{R}^k \rightarrow \mathbb{R}$
- Player suffers loss  $f_t(\mathbf{x}_t)$  and **observes**  $f_t$

Regret is

$$\mathcal{R}_T = \sum_{t=1}^T f_t(\mathbf{x}_t) - \min_{\mathbf{x} \in \mathcal{K}} \sum_{t=1}^T f_t(\mathbf{x})$$

# Simpler problem: Online Convex Optimization

Let  $\mathcal{K} \subset \mathbb{R}^k$  be some compact convex set.

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

- Player chooses  $\mathbf{x}_t \in \mathcal{K}$
- Adversary chooses convex  $f_t(\mathbf{x}) : \mathbb{R}^k \rightarrow \mathbb{R}$
- Player suffers loss  $f_t(\mathbf{x}_t)$  and **observes**  $f_t$

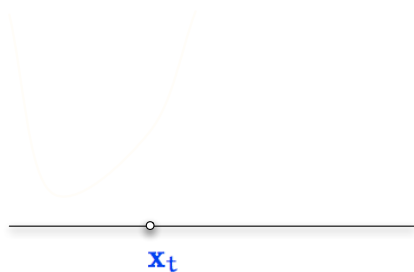
Regret is

$$\mathcal{R}_T = \sum_{t=1}^T f_t(\mathbf{x}_t) - \min_{\mathbf{x} \in \mathcal{K}} \sum_{t=1}^T f_t(\mathbf{x})$$

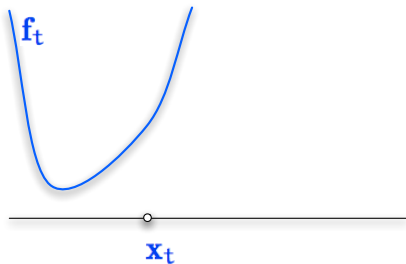
**Looks similar but, in fact, much easier!**

Note: this would be the Full-Info Driving to Work problem.

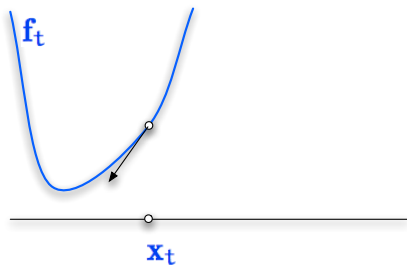
# Online Gradient Descent



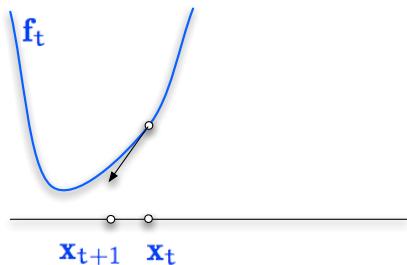
# Online Gradient Descent



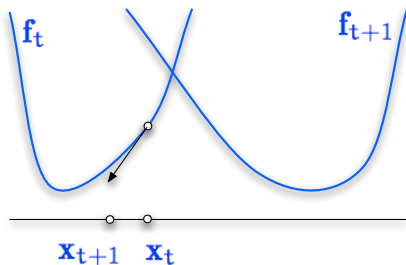
# Online Gradient Descent



# Online Gradient Descent



# Online Gradient Descent



# Online Convex Optimization is well-understood

If functions  $f_t$  are convex (linear), the regret is  $O(\sqrt{T})$ .

If functions  $f_t$  are strongly convex, the regret is  $O(\log T)$ .

Depending on the curvature of  $f_t$ 's, we can get all the rates between  $\log T$  and  $\sqrt{T}$  (Bartlett, Hazan, Rakhlin 2007)

The above rates are minimax optimal (up to constant) (Abernethy, Bartlett, Rakhlin, Tewari 2008)

# Online Linear Optimization with full information

Let  $\mathcal{K} \subset \mathbb{R}^k$  be some compact convex set.

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

- Player chooses  $\mathbf{x}_t \in \mathcal{K}$
- Adversary chooses  $\mathbf{f}_t \in \mathbb{R}^k$
- Player suffers loss  $\mathbf{f}_t^\top \mathbf{x}_t$  and **observes**  $\mathbf{f}_t$

Regret is

$$\mathcal{R}_T = GD\sqrt{T}$$

where  $G$  is the largest  $\|\mathbf{f}_t\|$  and  $D$  is the diameter of  $\mathcal{K}$

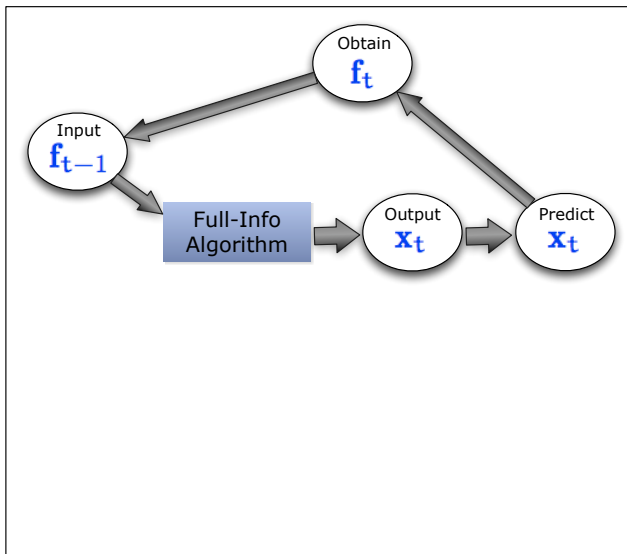
# Bandit setting



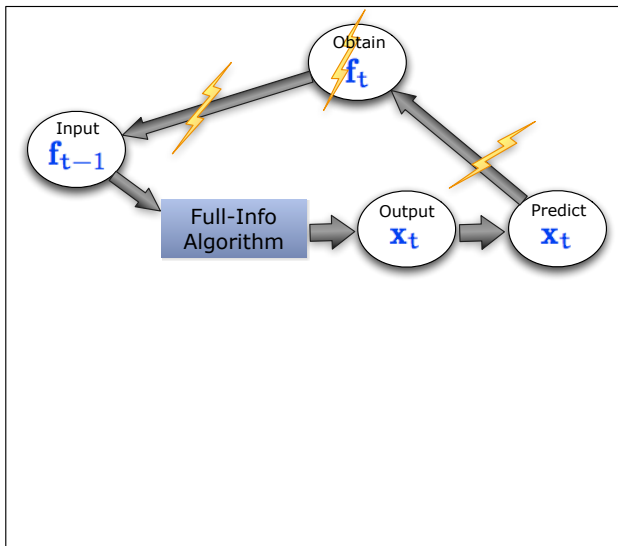
... unfortunately, we only observe  $\mathbf{f}_t^\top \mathbf{x}_t$  and therefore cannot take the gradient step...

Idea: estimate  $\mathbf{f}_t$  from a single sample  $\mathbf{f}_t^\top \mathbf{x}_t$  and then proceed as if in the full-information setting.

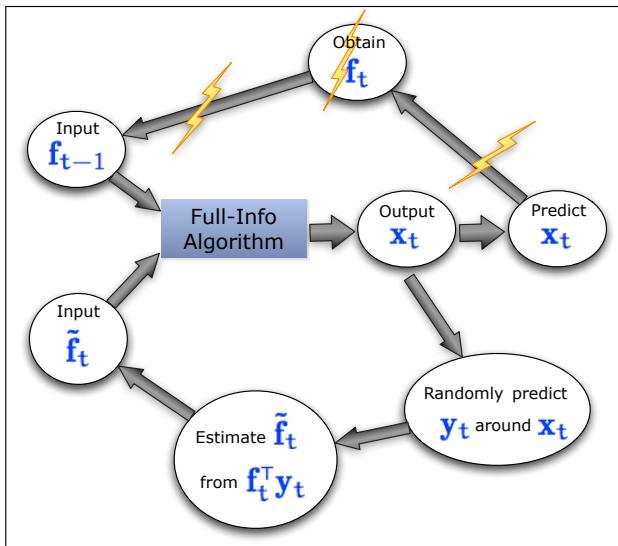
# Black-box reduction



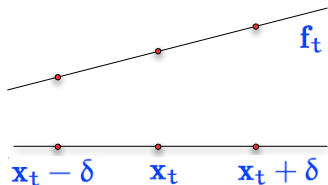
# Black-box reduction



# Black-box reduction



# Estimating the slope



Set  $\tilde{f}_t = \pm(f_t \cdot y_t)/\delta$ .

Unbiased:

$$\mathbb{E}\tilde{f}_t = \frac{1}{2} \frac{f_t \cdot (x_t + \delta)}{\delta} - \frac{1}{2} \frac{f_t \cdot (x_t - \delta)}{\delta} = f_t$$

and

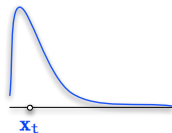
$$\mathbb{E}y_t = x_t$$

Predict randomly  $y_t = x_t \pm \delta$ .

Dilemma: estimation of  $f_t$  (exploration) is at odds with predicting  $x_t$  (exploitation)

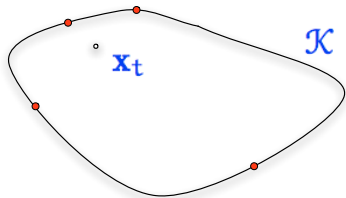
# Forces at play

We require  $\mathbb{E}\tilde{\mathbf{f}}_t = \mathbf{f}_t$  and  $\mathbb{E}\mathbf{y}_t = \mathbf{x}_t$  at the same time.



- want to increase the variance of the distribution (sample far away) in order to decrease the variance of  $\tilde{\mathbf{f}}_t$ , **YET**
- want to decrease the variance of the distribution (sample close) in order to follow the optimization procedure.
- want to shift  $\mathbf{x}_t$ , the center of the distribution, away from the boundary in order to have more room to sample, **YET**
- want to be close to the boundary, as the optimum has to occur at the edge.

# Sampling for general sets



Any estimator will scale as inverse distance to the boundary.

High variance cannot be avoided...

No way to get  $\sqrt{T}$  regret with vanilla optimization techniques.

# Outline

- 1 The Problem
- 2 Difficulties
- 3 Solution**
- 4 Conclusions

# The Curse of High Variance and The Blessing of Regularization

Idea: *regularize* to avoid problems near the boundary.

What if we solve

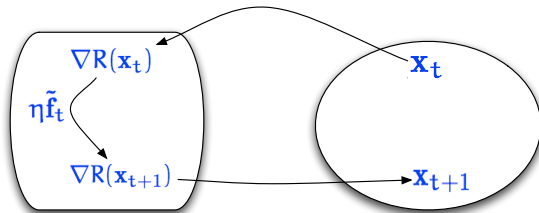
$$\mathbf{x}_{t+1} := \arg \min_{\mathbf{x} \in \mathcal{K}} \left[ \eta \sum_{s=1}^t \tilde{\mathbf{f}}_s^\top \mathbf{x} + \mathcal{R}(\mathbf{x}) \right].$$

Motivation: regularization in Statistical Learning.

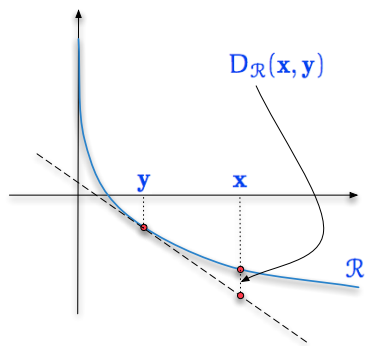
# Mirror Descent

Dual space

Primal space



# Bregman divergences



*Bregman divergence*

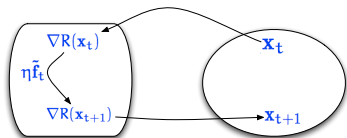
$$D_{\mathcal{R}}(\mathbf{x}, \mathbf{y}) = \mathcal{R}(\mathbf{x}) - \mathcal{R}(\mathbf{y}) - \nabla \mathcal{R}(\mathbf{y})(\mathbf{x} - \mathbf{y})$$

# Mirror Descent

Regularization solutions

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

satisfy



and

$$\sum_{t=1}^T \tilde{\mathbf{f}}_t^\top \mathbf{x}_t - \min_{\mathbf{x} \in \mathcal{K}} \left( \sum_{t=1}^T \tilde{\mathbf{f}}_t^\top \mathbf{x} + \eta^{-1} D_{\mathcal{R}}(\mathbf{x}, \mathbf{x}_1) \right) \leq \eta^{-1} \sum_{t=1}^T D_{\mathcal{R}}(\mathbf{x}_t, \mathbf{x}_{t+1})$$

# Self-concordant functions

A careful analysis of necessary properties for  $\mathcal{R}$  leads to a relation between 2nd and 3rd derivative.

Surprisingly, in Optimization this type of function, called *self-concordant*, is well-studied.

Central object of Interior Point methods.

The problem of *high variance* led us to a disparate field.

# Self-concordant barrier

## Definition

A *self-concordant function*  $\mathcal{R} : \text{int } \mathcal{K} \rightarrow \mathbb{R}$  is a  $C^3$  convex function such that

$$|\mathcal{D}^3 \mathcal{R}(\mathbf{x})[\mathbf{h}, \mathbf{h}, \mathbf{h}]| \leq 2 (\mathcal{D}^2 \mathcal{R}(\mathbf{x})[\mathbf{h}, \mathbf{h}])^{3/2}.$$

Here, the third-order differential is defined as

$$\mathcal{D}^3 \mathcal{R}(\mathbf{x})[\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3] := \frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} \Big|_{t_1=t_2=t_3=0} \mathcal{R}(\mathbf{x} + t_1 \mathbf{h}_1 + t_2 \mathbf{h}_2 + t_3 \mathbf{h}_3).$$

## Definition

A  *$\vartheta$ -self-concordant barrier*  $\mathcal{R}$  is a self-concordant function with

$$|\mathcal{D} \mathcal{R}(\mathbf{x})[\mathbf{h}]| \leq \vartheta^{1/2} [\mathcal{D}^2 \mathcal{R}(\mathbf{x})[\mathbf{h}, \mathbf{h}]]^{1/2}.$$

# Barriers

- 1 The generality of interior-point methods comes from the fact that any arbitrary  $K$ -dimensional closed convex set admits an  $O(K)$ -self-concordant barrier. Hence,  $\vartheta = O(K)$  (furthermore,  $\vartheta = 1$  for the sphere).

Through the Hessian of a self-concordant  $\mathcal{R}$ , we get

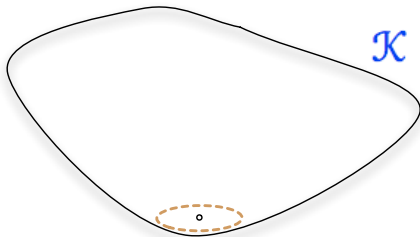
- 2 A handle on the regret:

$$D_{\mathcal{R}}(\mathbf{x}_t, \mathbf{x}_{t+1}) \approx \tilde{\mathbf{f}}_t^\top (\nabla^2 \mathcal{R}(\mathbf{x}_t))^{-1} \tilde{\mathbf{f}}_t$$

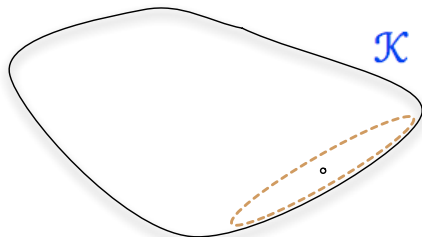
- 3 Local Euclidean geometry:

$$\|\mathbf{z}\|_{\mathbf{x}_t}^2 = (\mathbf{z} - \mathbf{x}_t)^\top \nabla^2 \mathcal{R}(\mathbf{x}_t) (\mathbf{z} - \mathbf{x}_t)$$

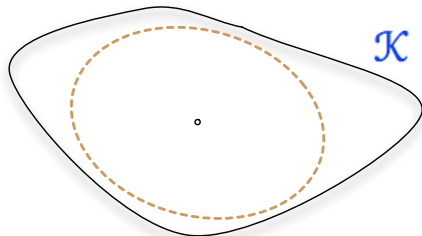
# Dikin ellipsoid – always contained in the set!



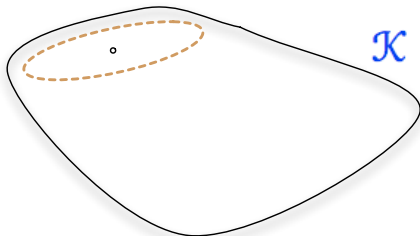
# Dikin ellipsoid – always contained in the set!



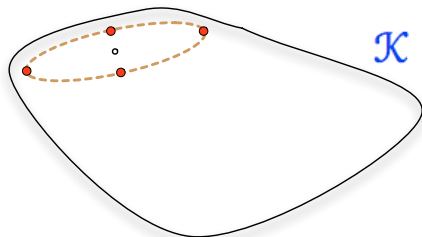
# Dikin ellipsoid – always contained in the set!



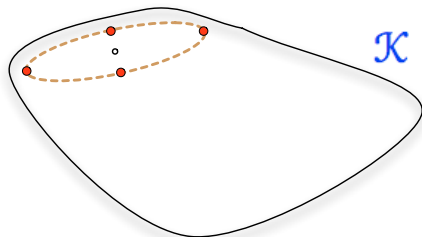
# Dikin ellipsoid – always contained in the set!



# Dikin ellipsoid – always contained in the set!



# Dikin ellipsoid – always contained in the set!



Hessian of  $\mathcal{R}$  gives us local geometry for estimating  $\tilde{\mathbf{f}}_t$  and controls regret through Bregman divergences  $D_{\mathcal{R}}$ .

# Estimates with the ellipsoid

Recall

$$D_{\mathcal{R}}(\mathbf{x}_t, \mathbf{x}_{t+1}) \propto \tilde{\mathbf{f}}_t^\top (\nabla^2 \mathcal{R}(\mathbf{x}_t))^{-1} \tilde{\mathbf{f}}_t.$$

Defining

$$\tilde{\mathbf{f}}_t \propto \sqrt{\lambda_i} \mathbf{e}_i,$$

large estimates are annihilated in exactly the directions we need.

Here  $\{\mathbf{e}_i\}, \{\lambda_i\}$  are eigenvectors and eigenvalues of  $\nabla^2 \mathcal{R}(\mathbf{x}_t)$ .

# Algorithm and Regret

Input:  $\eta > 0$ ,  $\vartheta$ -self-concordant  $\mathcal{R}$

Let  $\mathbf{x}_1 = \arg \min_{\mathbf{x} \in \mathcal{K}} [\mathcal{R}(\mathbf{x})]$ .

**for**  $t = 1$  to  $T$  **do**

Let  $\{\mathbf{e}_1, \dots, \mathbf{e}_K\}$  and  $\{\lambda_1, \dots, \lambda_K\}$  be the set of eigenvectors and eigenvalues of  $\nabla^2 \mathcal{R}(\mathbf{x}_t)$ .

Choose  $i_t$  uniformly at random from  $\{1, \dots, K\}$  and  $\varepsilon_t = \pm 1$  with prob.  $1/2$ .

Predict  $\mathbf{y}_t = \mathbf{x}_t + \varepsilon_t \lambda_{i_t}^{-1/2} \mathbf{e}_{i_t}$ .

Observe the loss  $\mathbf{f}_t^\top \mathbf{y}_t \in \mathbb{R}$ .

Define

$$\tilde{\mathbf{f}}_t := K (\mathbf{f}_t^\top \mathbf{y}_t) \varepsilon_t \lambda_{i_t}^{1/2} \cdot \mathbf{e}_{i_t}.$$

Update

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

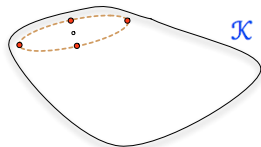
Theorem (Abernethy, Hazan, Rakhlin, 2008)

Let  $\mathbf{u}$  be any vector in  $\mathcal{K}$ . Suppose  $|\mathbf{f}_t^\top \mathbf{x}| \leq 1$  for any  $\mathbf{x} \in \mathcal{K}$ .

Setting  $\eta = \frac{\sqrt{\vartheta \log T}}{4K\sqrt{T}}$ , the regret is bounded as

$$\mathbb{E} \sum_{t=1}^T \mathbf{f}_t^\top \mathbf{y}_t - \min_{\mathbf{u} \in \mathcal{K}} \mathbb{E} \left( \sum_{t=1}^T \mathbf{f}_t^\top \mathbf{u} \right) \leq 16K\sqrt{\vartheta T \log T}$$

whenever  $T > 8\vartheta \log T$ .



	finite-armed	continuum-armed
stochastic	$O(\sqrt{KT \log T})$	$O(K\sqrt{T \log T})$
adversarial	$O(\sqrt{KT \log T})$	$O(K\sqrt{\vartheta T \log T})$

# Outline

- 1 The Problem
- 2 Difficulties
- 3 Solution
- 4 Conclusions**

# Conclusions

- Solution to the adversarial continuum-armed bandit problem
  - Novel connections between sequential prediction and Interior Point methods
  - Principled way to estimate missing information
  - Sequential view of Regularized Empirical Risk minimization
  - Solution to the Driving to Work problem
  - Applications: Industrial, Financial, Dynamic Treatment Strategies
- 
- Phenomenon: regret does not depend on whether Nature is stochastic or adversarial
  - Understanding worst-case scenario is important for understanding the statistical assumptions