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
Driving to Work
Driving to Work

The Problem

Difficulties

Solution

Conclusions

Online Learning with Limited Feedback
Driving to Work

Sasha Rakhlin

Online Learning with Limited Feedback
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 \( x_t \in \mathcal{K} \)
- Adversary (simultaneously) chooses \( f_t \in [0, 1]^K \)
- Player suffers loss \( f_t^T x_t \)
- Only \( f_t^T x_t \) is revealed

Regret is

\[
\mathcal{R}_T = \sum_{t=1}^{T} f_t^T x_t - \min_{x \in \mathcal{K}} \mathbb{E} \sum_{t=1}^{T} f_t^T x
\]
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 \( x_t \in \mathcal{K} \)
- Adversary (simultaneously) chooses \( f_t \in [0, 1]^K \)
- Player suffers loss \( f_t^T x_t \)
- Only \( f_t^T x_t \) is revealed

Regret is

\[
R_T = \sum_{t=1}^{T} f_t^T x_t - \min_{x \in \mathcal{K}} \mathbb{E} \sum_{t=1}^{T} f_t^T x
\]
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 \( x_t \in \mathcal{K} \)
- Adversary (simultaneously) chooses \( f_t \in [0, 1]^K \)
- Player suffers loss \( f_t^T x_t \)
- Only \( f_t^T x_t \) is revealed

Regret is

\[
R_T = \sum_{t=1}^{T} f_t^T x_t - \min_{x \in \mathcal{K}} \mathbb{E} \sum_{t=1}^{T} f_t^T x
\]

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

- $O(\sqrt{T})$ (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 $x_t \in \mathcal{K}$
- Adversary chooses convex $f_t(x) : \mathbb{R}^K \to \mathbb{R}$
- Player suffers loss $f_t(x_t)$ and observes $f_t$

Regret is

$$ R_T = \sum_{t=1}^{T} f_t(x_t) - \min_{x \in \mathcal{K}} \sum_{t=1}^{T} f_t(x) $$
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 $x_t \in \mathcal{K}$
- Adversary chooses convex $f_t(x) : \mathbb{R}^K \to \mathbb{R}$
- Player suffers loss $f_t(x_t)$ and observes $f_t$

Regret is

$$R_T = \sum_{t=1}^{T} f_t(x_t) - \min_{x \in \mathcal{K}} \sum_{t=1}^{T} f_t(x)$$

Looks similar but, in fact, much easier!

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

\[ X^*_t \]
Online Gradient Descent

\[ f_t \]

\[ x_t \]
Online Gradient Descent

Online Learning with Limited Feedback

Sasha Rakhlin
Online Gradient Descent

\[ \mathbf{f}_t \]

\[ \mathbf{x}_{t+1} \quad \mathbf{x}_t \]
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).
Let $\mathcal{K} \subset \mathbb{R}^K$ be some compact convex set.

At each time step $t = 1$ to $T$,
- Player chooses $x_t \in \mathcal{K}$
- Adversary chooses $f_t \in \mathbb{R}^K$
- Player suffers loss $f_t^T x_t$ and observes $f_t$

Regret is

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

where $G$ is the largest $\|f_t\|$ and $D$ is the diameter of $\mathcal{K}$
... unfortunately, we only observe $f_t^T x_t$ and therefore cannot take the gradient step...

Idea: estimate $f_t$ from a single sample $f_t^T x_t$ and then proceed as if in the full-information setting.
Black-box reduction

The Problem
Difficulties
Solution
Conclusions

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reduction

Black-box reductio
Black-box reduction

- Input \( f_{t-1} \)
- Full-Info Algorithm
- Output \( X_t \)
- Predict \( X_t \)
- Obtain \( f_t \)
Black-box reduction

The Problem
Difficulties
Solution
Conclusions

<table>
<thead>
<tr>
<th>Black-box reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-Info Algorithm</td>
</tr>
<tr>
<td>Input $f_{t-1}$</td>
</tr>
<tr>
<td>Obtain $f_t$</td>
</tr>
<tr>
<td>Full-Info Algorithm</td>
</tr>
<tr>
<td>Output $X_t$</td>
</tr>
<tr>
<td>Predict $X_t$</td>
</tr>
<tr>
<td>Input $\tilde{f}_t$</td>
</tr>
<tr>
<td>Estimate $\mathbf{y}_t$ around $X_t$</td>
</tr>
<tr>
<td>Randomly predict $y_t$ from $f^T_t y_t$</td>
</tr>
</tbody>
</table>

- Sasha Rakhlin
- Online Learning with Limited Feedback
Estimating the slope

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

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

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

$$\mathbb{E}y_t = x_t$$
We require $\mathbb{E}\tilde{f}_t = f_t$ and $\mathbb{E}y_t = x_t$ at the same time.

- want to increase the variance of the distribution (sample far away) in order to decrease the variance of $\tilde{f}_t$, YET
- want to decrease the variance of the distribution (sample close) in order to follow the optimization procedure.

- want to shift $x_t$, 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.
The Curse of High Variance and
The Blessing of Regularization

Idea: regularize to avoid problems near the boundary.

What if we solve

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

Motivation: regularization in Statistical Learning.
Mirror Descent

Dual space

Primal space

\[ \nabla R(x_t) \]

\[ \eta \tilde{f}_t \]

\[ \nabla R(x_{t+1}) \]

\[ x_t \]

\[ x_{t+1} \]
Bregman divergences

Bregman divergence

\[ D_\mathcal{R}(x, y) = \mathcal{R}(x) - \mathcal{R}(y) - \nabla \mathcal{R}(y)(x - y) \]
Mirror Descent

Regularization solutions

\[ x_{t+1} := \arg \min_{x \in \mathcal{X}} \left[ \eta \sum_{s=1}^{t} \tilde{f}_s^T x + \mathcal{R}(x) \right] \]

satisfy

\[ \sum_{t=1}^{T} \tilde{f}_t^T x_t - \min_{x \in \mathcal{X}} \left( \sum_{t=1}^{T} \tilde{f}_t^T x + \eta^{-1} \mathcal{D}_\mathcal{R}(x, x_1) \right) \leq \eta^{-1} \sum_{t=1}^{T} \mathcal{D}_\mathcal{R}(x_t, x_{t+1}) \]
Self-concordant functions

A careful analysis of necessary properties for $\mathbb{R}$ 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.
A *self-concordant function* $R : \text{int } K \rightarrow \mathbb{R}$ is a $C^3$ convex function such that

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

Here, the third-order differential is defined as

$$D^3 R(x)[h_1, h_2, h_3] := \frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3}|_{t_1=t_2=t_3=0} R(x + t_1 h_1 + t_2 h_2 + t_3 h_3).$$

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

$$|D R(x)[h]| \leq \vartheta^{1/2} \left[D^2 R(x)[h, h]\right]^{1/2}.$$
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 $R$, we get

$$D_R(x_t, x_{t+1}) \approx \tilde{f}_t^T (\nabla^2 R(x_t))^{-1} \tilde{f}_t$$

Local Euclidean geometry:

$$\|z\|_{x_t}^2 = (z - x_t) \nabla^2 R(x_t) (z - x_t)$$
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!
Hessian of $\mathcal{R}$ gives us local geometry for estimating $\tilde{f}_t$ and controls regret through Bregman divergences $D_{\mathcal{R}}$. 

Dikin ellipsoid – always contained in the set!
Estimates with the ellipsoid

Recall

\[ D_{\mathcal{R}}(x_t, x_{t+1}) \propto \tilde{f}_t^T (\nabla^2 \mathcal{R}(x_t))^{-1} \tilde{f}_t. \]

Defining

\[ \tilde{f}_t \propto \sqrt{\lambda_i} e_i, \]

large estimates are annihilated in exactly the directions we need.

Here \( \{e_i\}, \{\lambda_i\} \) are eigenvectors and eigenvalues of \( \nabla^2 \mathcal{R}(x_t) \).
**Algorithm and Regret**

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

Let $x_1 = \arg\min_{x \in \mathcal{K}} [\mathcal{R}(x)]$.

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

Let $\{e_1, \ldots, e_K\}$ and $\{\lambda_1, \ldots, \lambda_K\}$ be the set of eigenvectors and eigenvalues of $\nabla^2 \mathcal{R}(x_t)$.

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

Predict $y_t = x_t + \varepsilon_t \lambda_{i_t}^{-1/2} e_{i_t}$.

Observe the loss $f_t^T y_t \in \mathbb{R}$.

Define

$$\tilde{f}_t := K (f_t^T y_t) \varepsilon_t \lambda_{i_t}^{1/2} e_{i_t}.$$ 

Update

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

**Theorem** (*Abernethy, Hazan, Rakhlin, 2008*)

Let $u$ be any vector in $\mathcal{K}$. Suppose $|f_t^T x| \leq 1$ for any $x \in \mathcal{K}$.

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

$$\mathbb{E} \sum_{t=1}^{T} f_t^T y_t - \min_{u \in \mathcal{K}} \mathbb{E} \left( \sum_{t=1}^{T} f_t^T u \right) \leq 16K \sqrt{\vartheta T \log T}$$

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

*Sasha Rakhlin*  
*Online Learning with Limited Feedback*
<table>
<thead>
<tr>
<th></th>
<th>finite-armed</th>
<th>continuum-armed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>stochastic</strong></td>
<td>$O(\sqrt{KT \log T})$</td>
<td>$O(K\sqrt{T \log T})$</td>
</tr>
<tr>
<td><strong>adversarial</strong></td>
<td>$O(\sqrt{KT \log T})$</td>
<td>$O(K\sqrt{\vartheta T \log T})$</td>
</tr>
</tbody>
</table>
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