Improved Regret Guarantees for OCO in Bandit Setting

Ankan Saha ¹  Ambuj Tewari ²

AISTATS 2011

¹University of Chicago
²University of Texas at Austin
Introduction and Motivation

- Sequential decision making: important question in machine learning, Economics, Operations Research and related fields.

- Modeled as a sequential game between learner and adversary.
Setting

At every time step $t$: 

- Player plays a point $x_t \in K \subseteq \mathbb{R}^d$, where $K$ is convex and compact.
- Adversary responds with a function $f_t \in F$.
- Player suffers a loss $f_t(x_t)$.

Online Convex Optimization: $f_t$ are convex functions.

Full information vs. Bandit Setting.
At every time step $t$:

- Player plays a point $x_t \in \mathcal{K} \subseteq \mathbb{R}^d$, $\mathcal{K}$ is convex and compact.
Setting

At every time step $t$:
- Player plays a point $x_t \in \mathcal{K} \subseteq \mathbb{R}^d$, $\mathcal{K}$ is convex and compact.
- Adversary responds with a function $f_t \in \mathcal{F}$.
- Player suffers a loss $f_t(x_t)$. 
Setting

- At every time step $t$:
- Player plays a point $x_t \in \mathcal{K} \subseteq \mathbb{R}^d$, $\mathcal{K}$ is convex and compact.
- Adversary responds with a function $f_t \in \mathcal{F}$.
- Player suffers a loss $f_t(x_t)$.
- *Online Convex Optimization*: $f_t$ are convex functions.
Setting

- At every time step $t$:
  - Player plays a point $x_t \in \mathcal{K} \subseteq \mathbb{R}^d$, $\mathcal{K}$ is convex and compact.
  - Adversary responds with a function $f_t \in \mathcal{F}$.
  - Player suffers a loss $f_t(x_t)$.
  - **Online Convex Optimization**: $f_t$ are convex functions.
  - Full information vs. Bandit Setting.
Goal: To minimize the **Regret** i.e. the player’s performance with respect to the best performance in hindsight.

\[
R_T = \sum_{i=1}^{T} f_t(x_t) - \min_{x^* \in \mathcal{K}} \sum_{t=1}^{T} f_t(x^*)
\]
Goal: To minimize the **Regret** i.e. the player’s performance with respect to the best performance in hindsight.

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

Full information Setting: Zinkevich’s algorithm.

\[
x_{t+1} = \text{Proj}_\mathcal{K} \{x_t - \eta_t \nabla f_t(x_t)\}
\]
Goal: To minimize the **Regret** i.e. the player’s performance with respect to the best performance in hindsight.

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

- Full information Setting: Zinkevich’s algorithm.

\[ x_{t+1} = \text{Proj}_{\mathcal{K}} \{ x_t - \eta_t \nabla f_t(x_t) \} \]

- Projected gradient descent strategy incurs \( O(\sqrt{T}) \) regret.

- Zinkevich’s algorithm just requires information about the gradients of \( f_t \) at every time step.
Zinkevich’s Algorithm

- Key lies in the fact that we minimize the regret and not the actual loss.

**Theorem**

*If the updates are given by $x_{t+1} = \text{Proj}_K(x_t - \eta_t \nabla f_t(x_t))$, choosing $\eta_t = t^{-1/2}$ gives the following bound on regret after $T$ steps*

$$R_T \leq \frac{\text{diam}(K)^2 \sqrt{T}}{2} + (\sqrt{T} - \frac{1}{2}) \|\nabla f\|^2$$

*where $\|\nabla f\| = \max_{x \in K, t \in [T]} \|\nabla f_t(x)\|$*
Zinkevich’s Algorithm

- Key lies in the fact that we minimize the regret and not the actual loss.

**Theorem**

If the updates are given by \( x_{t+1} = \text{Proj}_K (x_t - \eta_t \nabla f_t(x_t)) \), choosing \( \eta_t = t^{-1/2} \) gives the following bound on regret after \( T \) steps

\[
R_T \leq \frac{\text{diam}(K)^2 \sqrt{T}}{2} + (\sqrt{T} - \frac{1}{2}) \| \nabla f \|^2
\]

where \( \| \nabla f \| = \max_{x \in K, t \in [T]} \| \nabla f_t(x) \| \)

- Convergence rate depends on \( \| \nabla f \| \).
Proxy for “Missing Information”

Full Information Oracle

\[ x_t \rightarrow \text{Bandit Online Algorithms} \rightarrow x_{t+1} \]
Missing information $\Rightarrow$ Gradient of the function.
Missing information $\rightarrow$ Gradient of the function.

How to evaluate the gradient from a single point evaluation?
Unbiased gradient estimation

- [FKM05] provides such a scheme.
- Introduce randomness.

\[ \hat{f}(x) = \mathbb{E}_{v \in \mathbb{B}^d} [f(x + \delta v)] \]
Unbiased gradient estimation

- [FKM05] provides such a scheme.
- Introduce randomness.

\[ \hat{f}(x) = \mathbb{E}_{v \in \mathcal{B}^d}[f(x + \delta v)] \]

- Then

\[ \nabla \hat{f}(x) = \mathbb{E}_{u \in \mathcal{S}^d} \left[ \frac{d}{\delta} f(x + \delta u) \cdot u \right] \]
Problems with existing approach

- Key trick in [FKM05] is to choose the size of the ball for evaluating the unbiased gradient estimate.
Problems with existing approach

- Key trick in [FKM05] is to choose the size of the ball for evaluating the unbiased gradient estimate.
- Too small $\delta$ blows up the $\|\nabla f\|$, depending on the location of $x_t$ and weakens the bounds.
Problems with existing approach

- Key trick in [FKM05] is to choose the size of the ball for evaluating the unbiased gradient estimate.
- Too small $\delta$ blows up the $\|\nabla f\|$, depending on the location of $x_t$ and weakens the bounds.
- Too large $\delta \implies$ Gradient estimates are not very accurate.
Problems with existing approach

- Key trick in [FKM05] is to choose the size of the ball for evaluating the unbiased gradient estimate.
- Too small $\delta$ blows up the $||\nabla f||$, depending on the location of $x_t$ and weakens the bounds.
- Too large $\delta \implies$ Gradient estimates are not very accurate.
- Trading off the ball size along with Zinkevich’s scheme gives $O(T^{3/4})$ regret in the OCO bandit setting.
Convex Analysis Basics

Convex functions: Lower bounded by linear approximation.

\[ f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \]

Strongly Convex functions: Lower bounded by a quadratic.

\[ f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \sigma^2 \|x - y\|^2 \]

l.c.g functions: Upper bounded by a quadratic.

\[ f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + L_2 \|x - y\|^2 \]
Convex Analysis Basics

- Convex functions: Lower Bounded by linear approximation.

\[ f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \]

- Strongly Convex functions: Lower Bounded by a quadratic.
  
  \[ f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \sigma^2 \|x - y\|^2 \]

- l.c.g functions: Upper Bounded by a quadratic.

\[ f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \|x - y\|^2 \]
Convex Analysis Basics

- Convex functions: Lower Bounded by linear approximation.
  \[ f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \]

- Strongly Convex functions: Lower Bounded by a quadratic. Eg. 2-norm squared.
  \[ f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{\sigma}{2} \|x - y\|^2 \]
Convex Analysis Basics

- Convex functions: Lower Bounded by linear approximation.
  \[ f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \]

- Strongly Convex functions: Lower Bounded by a quadratic.
  Eg. 2-norm squared.
  \[ f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{\sigma}{2} \|x - y\|^2 \]

- \textit{l.c.g} functions: Upper Bounded by a quadratic.
  \[ f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2} \|x - y\|^2 \]
Figure: \( l.c.g \) functions
Results for Subclasses of Convex functions (Full Information Setting)

<table>
<thead>
<tr>
<th>Function Class</th>
<th>Best Rates on Regret</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{F}_{\text{Lip}} )</td>
<td>( \Theta(T^{1/2}) ) [Zin03]</td>
</tr>
</tbody>
</table>
Results for Subclasses of Convex functions (Full Information Setting)

<table>
<thead>
<tr>
<th>Function Class</th>
<th>Best Rates on Regret</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipschitz ($\mathcal{F}_{Lip}$)</td>
<td>$\Theta(T^{1/2})$ [Zin03]</td>
</tr>
<tr>
<td>Smooth ($\mathcal{F}_{Smth}$)</td>
<td>$\Theta(T^{1/2})$ [Zin03]</td>
</tr>
</tbody>
</table>
Results for Subclasses of Convex functions (Full Information Setting)

<table>
<thead>
<tr>
<th>Function Class</th>
<th>Best Rates on Regret</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipschitz ($\mathcal{F}_{\text{Lip}}$)</td>
<td>$\Theta(T^{1/2})$ [Zin03]</td>
</tr>
<tr>
<td>Smooth ($\mathcal{F}_{\text{Smth}}$)</td>
<td>$\Theta(T^{1/2})$ [Zin03]</td>
</tr>
<tr>
<td>Lipschitz and Strongly Convex ($S_{\text{Lip}}$)</td>
<td>$\Theta(\log T)$ [HAK07]</td>
</tr>
</tbody>
</table>
Results for Subclasses of Convex functions (Full Information Setting)

<table>
<thead>
<tr>
<th>Function Class</th>
<th>Best Rates on Regret</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipschitz ($\mathcal{F}_{Lip}$)</td>
<td>$\Theta(T^{1/2})$ [Zin03]</td>
</tr>
<tr>
<td>Smooth ($\mathcal{F}_{Smth}$)</td>
<td>$\Theta(T^{1/2})$ [Zin03]</td>
</tr>
<tr>
<td>Lipschitz and Strongly Convex ($\mathcal{S}_{Lip}$)</td>
<td>$\Theta(\log T)$ [HAK07]</td>
</tr>
<tr>
<td>Smooth and Strongly Convex ($\mathcal{S}_{Smth}$)</td>
<td>$\Theta(\log T)$ [HAK07]</td>
</tr>
</tbody>
</table>
Results for Bandit Setting

<table>
<thead>
<tr>
<th>Function Class</th>
<th>Best Rates on Regret</th>
<th>Lower Bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipschitz ($\mathcal{F}_{\text{Lip}}$)</td>
<td>$O(T^{3/4})$ [FKM05]</td>
<td>$\Omega(T^{1/2})$</td>
</tr>
</tbody>
</table>
## Results for Bandit Setting

<table>
<thead>
<tr>
<th>Function Class</th>
<th>Best Rates on Regret</th>
<th>Lower Bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mathcal{F}_{\text{Lip}})</td>
<td>(O(T^{3/4})) [FKM05]</td>
<td>(\Omega(T^{1/2}))</td>
</tr>
<tr>
<td>(\mathcal{F}_{\text{Smth}})</td>
<td>(O^*(T^{2/3})) [This paper]</td>
<td>(\Omega(T^{1/2}))</td>
</tr>
</tbody>
</table>
Results for Bandit Setting

<table>
<thead>
<tr>
<th>Function Class</th>
<th>Best Rates on Regret</th>
<th>Lower Bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipschitz ($\mathcal{F}_{\text{Lip}}$)</td>
<td>$O(T^{3/4})$ [FKM05]</td>
<td>$\Omega(T^{1/2})$</td>
</tr>
<tr>
<td>Smooth ($\mathcal{F}_{\text{Smth}}$)</td>
<td>$O^*(T^{2/3})$ [This paper]</td>
<td>$\Omega(T^{1/2})$</td>
</tr>
<tr>
<td>Lipschitz and S.C. ($S_{\text{Lip}}$)</td>
<td>$O^*(T^{2/3})$ [ADX10]</td>
<td>$\Omega(\log T)$</td>
</tr>
</tbody>
</table>
## Results for Bandit Setting

![Diagram](image)

<table>
<thead>
<tr>
<th>Function Class</th>
<th>Best Rates on Regret</th>
<th>Lower Bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipschitz ((\mathcal{F}_{\text{Lip}}))</td>
<td>(O(T^{3/4})) [FKM05]</td>
<td>(\Omega(T^{1/2}))</td>
</tr>
<tr>
<td><strong>Smooth</strong> ((\mathcal{F}_{\text{Smth}}))</td>
<td>(O^*(T^{2/3})) [This paper]</td>
<td>(\Omega(T^{1/2}))</td>
</tr>
<tr>
<td>Lipschitz and S.C. ((S_{\text{Lip}}))</td>
<td>(O^*(T^{2/3})) [ADX10]</td>
<td>(\Omega(\log T))</td>
</tr>
<tr>
<td>Smooth and S.C. ((S_{\text{Smth}}))</td>
<td>(O^*(T^{2/3})) [ADX10]</td>
<td>(\Omega(\log T))</td>
</tr>
</tbody>
</table>
New Approach

- Tighter bounds obtained by calculating \(\|\nabla f\|\) with respect to a changing local norm, demonstrated for Bandit OLO [AHR08].

**Definition**

A function \(R: \mathcal{K} \to \mathbb{R}\) is a self-concordant barrier if

a) \(R \to \infty\) near the boundary of \(\mathcal{K}\) and

b) \(R\) and \(\nabla^2 R\) are Lipschitz continuous with respect to the local norm \(\|\cdot\|_{R,w}\), given by \(\|f\|_{R,w} = \sqrt{\langle f, \nabla^2 R(w)f \rangle}\).
New Approach

- Tighter bounds obtained by calculating $\|\nabla f\|$ with respect to a changing local norm, demonstrated for Bandit OLO [AHR08].

**Definition**

A function $R : \mathcal{K} \to \mathbb{R}$ is a self-concordant barrier if

a) $R \to \infty$ near the boundary of $\mathcal{K}$ and

b) $R$ and $\nabla^2 R$ are Lipschitz continuous with respect to the local norm $\|\cdot\|_{R,w}$, given by $\|f\|_{R,w} = \sqrt{\langle f, \nabla^2 R(w) f \rangle}$.

- Given a self concordant barrier $R$, for every $w \in \mathcal{K}$, the Dikin ellipsoid centered at $w$

$$D_w = \left\{ w' : \|w - w'\|_{R,w} \leq 1 \right\}$$

is always contained in $\mathcal{K}$. 
The Dikin ellipsoid always lies inside $\mathcal{K}$. It needs to curve strongly near the boundary.
Dikin ellipsoid always lies inside $\mathcal{K}$.
• Dikin ellipsoid always lies inside $\mathcal{K}$.
• $R$ needs to curve strongly near the boundary.
If $A_t^2 = \{\nabla^2 R(x_t)\}^{-1}$

Then $y_t = x_t + A_t u_t$ lies on $D_{x_t}$

- Dikin ellipsoid always lies inside $\mathcal{K}$.
- $R$ needs to curve strongly near the boundary.
Combining Ideas

- Combine single point gradient estimate with the idea of self concordant barriers.
Combining Ideas

- Combine single point gradient estimate with the idea of self concordant barriers.
- Removes the need for projections.
Combining Ideas

- Combine single point gradient estimate with the idea of self concordant barriers.
- Removes the need for projections.
- Generate gradient estimate $g_t$ at step $t$.
- Feed it to the full information algorithm blackbox from [AHR08] to obtain

$$x_{t+1} = \arg\min_{x \in \mathcal{K}} \sum_{i=1}^{t} \eta \langle g_i, x \rangle + R(x)$$
Algorithm

- Given $R$: a self concordant barrier for $\mathcal{K}$. 

$\text{Algorithm}$ 

1. Pick $x_1 \in \mathcal{K}$.
2. At every step $t = 1, 2, 3, \ldots$,
   - Evaluate $A_t = (\nabla^2 R(x_t))^{-1}/2$.
   - Sample $u_t \in S_d$.
   - $y_t = x_t + \delta A_t u_t$.
   - Player plays $y_t$ and receives loss $f_t(y_t)$.
   - $g_t = d\delta f_t(y_t) A_t u_t$. 

Ankan Saha, Ambuj Tewari 

Improved Regret Guarantees for OCO in Bandit Setting
Algorithm

- Given $R$: a self concordant barrier for $\mathcal{K}$.
- Pick $x_1 \in \mathcal{K}$.

At every step $t = 1, 2, 3, \ldots$, evaluate $A_t = \left( \nabla^2 R(x_t) \right) - \frac{1}{2}$.

Sample $u_t \in S_d$.

$y_t = x_t + \delta A_t u_t$.

Player plays $y_t$ and receives loss $f_t(y_t)$.

$g_t = d \delta f_t(y_t) A_t^{-1} u_t$. 

$\mathcal{K}$
Algorithm

- Given $R$: a self concordant barrier for $\mathcal{K}$.
- Pick $x_1 \in \mathcal{K}$.

At every step $t = 1, 2, 3, \ldots$,

- Evaluate $A_t = (\nabla^2 R(x_t))^{-1/2}$.
Given $R$: a self concordant barrier for $\mathcal{K}$.

Pick $x_1 \in \mathcal{K}$.

At every step $t = 1, 2, 3, \ldots$,

Evaluate $A_t = \left(\nabla^2 R(x_t)\right)^{-1/2}$.

Sample $u_t \in S^d$.

$y_t = x_t + \delta A_t u_t$. 

$\mathcal{K}$
Algorithm

- Given $R$: a self concordant barrier for $\mathcal{K}$.
- Pick $x_1 \in \mathcal{K}$.

At every step $t = 1, 2, 3, \ldots$,

- Evaluate $A_t = (\nabla^2 R(x_t))^{-1/2}$.
- Sample $u_t \in \mathcal{S}^d$.
  - $y_t = x_t + \delta A_t u_t$.
- Player plays $y_t$ and receives loss $f_t(y_t)$. 
Given $R$: a self concordant barrier for $\mathcal{K}$.

Pick $x_1 \in \mathcal{K}$.

At every step $t = 1, 2, 3, \ldots$,

- Evaluate $A_t = (\nabla^2 R(x_t))^{-1/2}$.
- Sample $u_t \in S^d$.
  
  $y_t = x_t + \delta A_t u_t$.

- Player plays $y_t$ and receives loss $f_t(y_t)$.
- $g_t = \frac{d}{\delta} f_t(y_t) A_t^{-1} u_t$. 

\begin{itemize}
  \item Algorithm
  \item Ankan Saha, Ambuj Tewari
  \item Improved Regret Guarantees for OCO in Bandit Setting
\end{itemize}
Given $R$: a self concordant barrier for $\mathcal{K}$.

Pick $x_1 \in \mathcal{K}$.

At every step $t = 1, 2, 3, \ldots$,

Evaluate $A_t = \left(\nabla^2 R(x_t)\right)^{-1/2}$.

Sample $u_t \in S^d$.

$y_t = x_t + \delta A_t u_t$.

Player plays $y_t$ and receives loss $f_t(y_t)$.

$g_t = \frac{d}{\delta} f_t(y_t) A_t^{-1} u_t$.

Full Information
Oracle

[ACHR08]

Ankan Saha, Ambuj Tewari

Improved Regret Guarantees for OCO in Bandit Setting
Proof Sketch

- Define the approximation

\[ \hat{f}_t(x) = \mathbb{E}_{v \in \mathbb{B}^d} \left[ f_t(x + \delta A_t v_t) \right] \]

- Regret written as a telescopic sum.
- Using approximation + Randomization incurs \( O(T\delta^2) \) cost.
- [AHR08] blackbox incurs \( O^*(\sqrt{T}/\delta) \).
- Trading off \( \delta \) gives \( O^*(T^{2/3}) \) regret.
Conclusions and Future Work

- Improve bounds on regret for bandit OCO from $O(T^{3/4})$ to $O^*(T^{2/3})$, when $f_t$ are smooth.
Conclusions and Future Work

- Improve bounds on regret for bandit OCO from $O(T^{3/4})$ to $O^*(T^{2/3})$, when $f_t$ are smooth.
- Lower Bounds still correspond to full information setting.
Conclusions and Future Work

- Improve bounds on regret for bandit OCO from $O(T^{3/4})$ to $O^*(T^{2/3})$, when $f_t$ are smooth.
- Lower Bounds still correspond to full information setting.
- How to bridge the gap?
Conclusions and Future Work

- Improve bounds on regret for bandit OCO from $O(T^{3/4})$ to $O^*(T^{2/3})$, when $f_t$ are smooth.
- Lower Bounds still correspond to full information setting.
- How to bridge the gap?
- Better bounds on “price of bandit information” for more specific function classes?
Conclusions and Future Work

- Improve bounds on regret for bandit OCO from $O(T^{3/4})$ to $O^*(T^{2/3})$, when $f_t$ are smooth.
- Lower Bounds still correspond to full information setting.
- How to bridge the gap?
- Better bounds on “price of bandit information” for more specific function classes?

Thank You!


M. Zinkevich.
Online convex programming and generalised infinitesimal gradient ascent.