Optimal Algorithms for Online Convex Optimization with Multi-Point Bandit Feedback

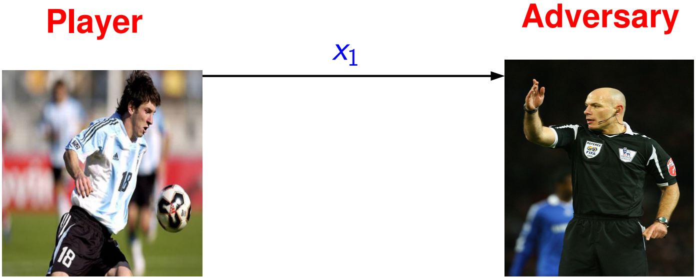
Alekh Agarwal Ofer Dekel Lin Xiao
UC Berkeley Microsoft Research

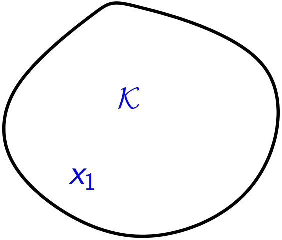
Online Convex Optimization (Full-Info)

Player

Adversary

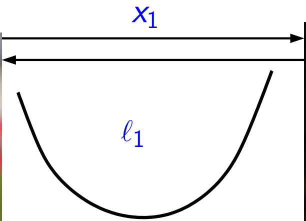
Online Convex Optimization (Full-Info)

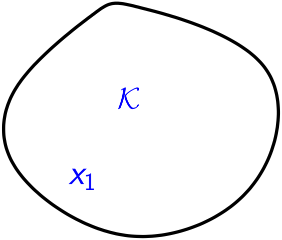




Online Convex Optimization (Full-Info)

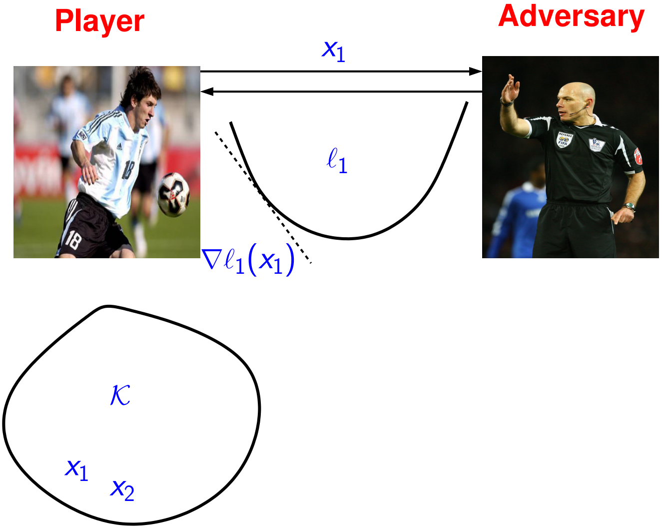
Player **Adversary**



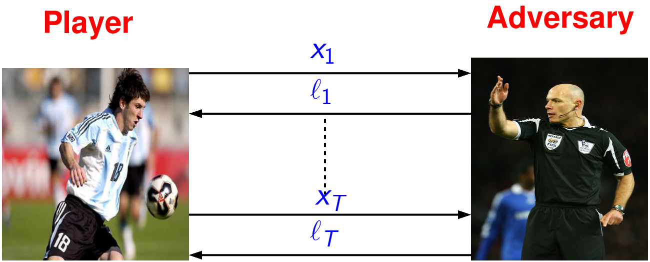


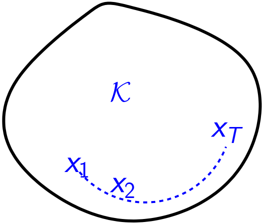
Online Convex Optimization (Full-Info)

- Player updates $x_{t+1} = \Pi_{\mathcal{K}}(x_t - \eta \nabla \ell_t(x_t))$.



Online Convex Optimization (Full-Info)





- Minimize regret: $R_T = \sum_{t=1}^{T} \ell_t(x_t) - \min_{x\in\mathcal{K}} \sum_{t=1}^{T} \ell_t(x)$.

Bandit Convex Optimization

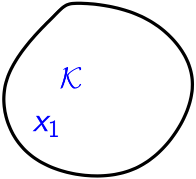
Player

Adversary

Player

Adversary

$\xrightarrow{x_1}$

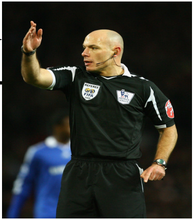


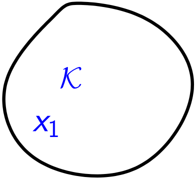
Bandit Convex Optimization

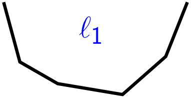
Player **Adversary**

x_1

$\ell_1(x_1)$







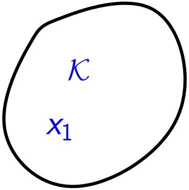
Bandit Gradient Descent [FKM'05]

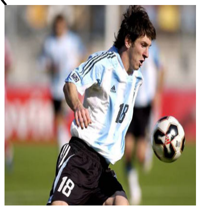
Player

Adversary

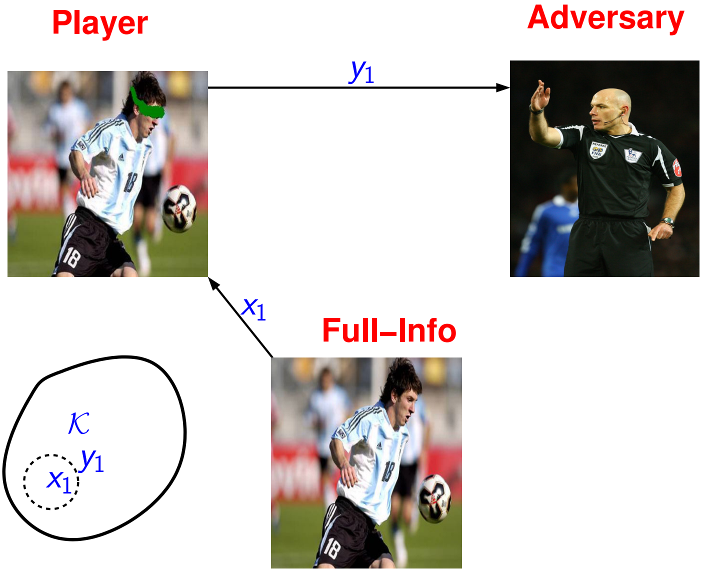
x_1

Full–Info

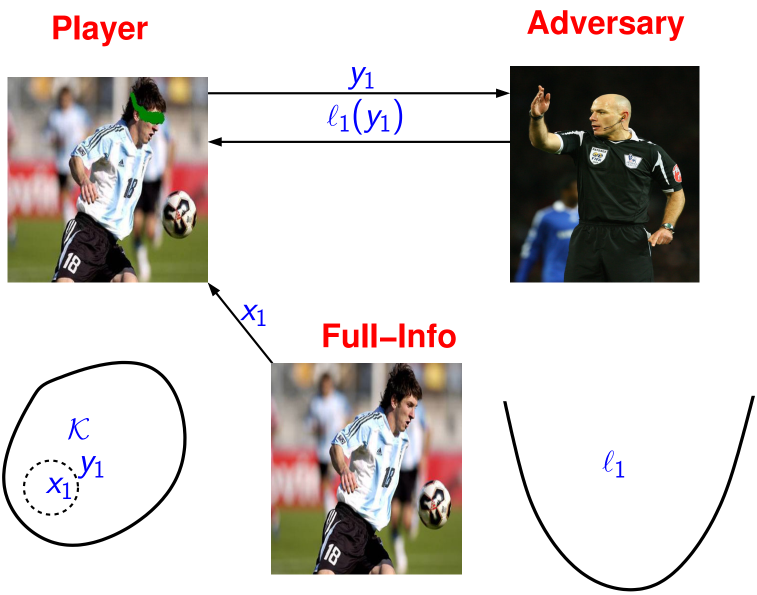




Bandit Gradient Descent [FKM'05]

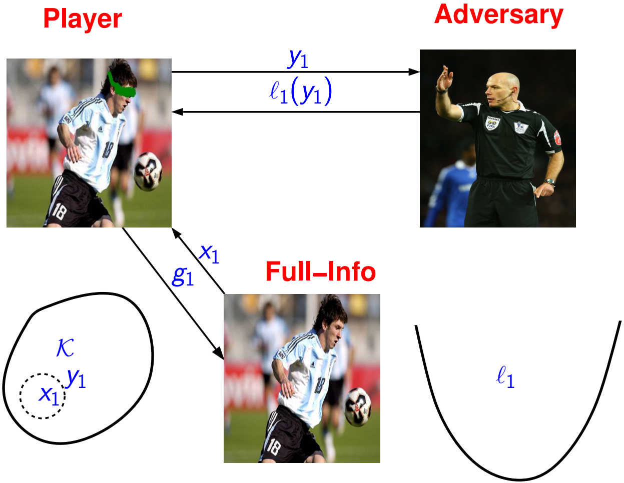


Bandit Gradient Descent [FKM'05]



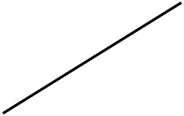
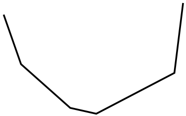
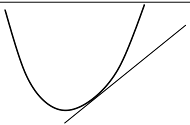
Bandit Gradient Descent [FKM'05]

- Updates $x_{t+1} = \Pi_{(1-\xi)\mathcal{K}}(x_t - \eta_t g_t)$.



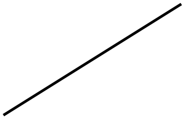
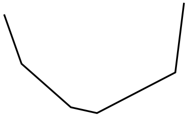
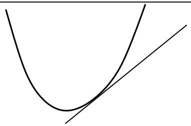
- Minimize regret: $R_T = \sum_{t=1}^{T} \ell_t(y_t) - \min_{x \in \mathcal{K}} \sum_{t=1}^{T} \ell_t(x)$.

A survey of known regret bounds

	Linear		Convex		Strongly Convex	
						
	Upper	Lower	Upper	Lower	Upper	Lower
Full-Info	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\log T)$	$\mathcal{O}(\log T)$

- Deterministic results against completely adaptive adversaries in Full-Info.

A survey of known regret bounds

	Linear		Convex		Strongly Convex	
						
	Upper	Lower	Upper	Lower	Upper	Lower
Full-Info	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\log T)$	$\mathcal{O}(\log T)$
Bandit	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(T^{3/4})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(T^{2/3})$	$\mathcal{O}(\sqrt{T})$?

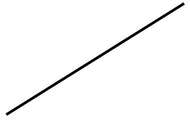
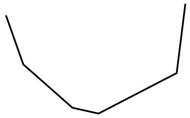
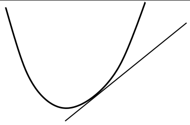
- Deterministic results against completely adaptive adversaries in Full-Info.
- High probability results against adaptive adversaries for Bandit.

The Multi-Point (MP) feedback setup

- Want to interpolate between bandit and full information.
- Player allowed several queries per round.
- Adversary reveals value of ℓ_t at all points picked.
- Average regret on points played:

$$R_T = \sum_{t=1}^{T} \frac{1}{k} \sum_{i=1}^{k} \ell_t(y_{t,i}) - \min_{x \in \mathcal{K}} \ell_t(x).$$

A survey of known regret bounds

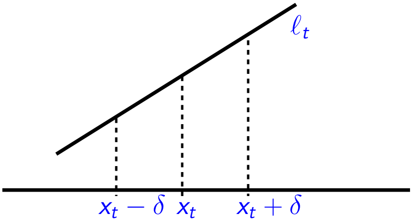
	Linear		Convex		Strongly Convex	
						
	Upper	Lower	Upper	Lower	Upper	Lower
Full-Info	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\log T)$	$\mathcal{O}(\log T)$
Bandit	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(T^{3/4})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(T^{2/3})$	$\mathcal{O}(\sqrt{T})$?
MP Bandit	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\sqrt{T})$	$\mathcal{O}(\log T)$	$\mathcal{O}(\log T)$

- Deterministic results against completely adaptive adversaries in Full-Info.
- High probability results against adaptive adversaries for Bandit.

Properties of gradient estimator g_t [FKM'05]

$$g_t = \frac{d}{\delta} \ell_t(x_t + \delta u_t) u_t.$$

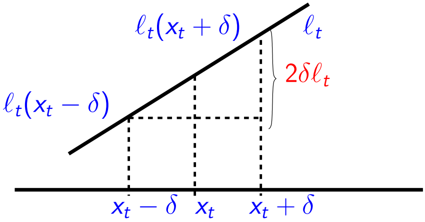
- Unbiased for linear functions.
- Nearly unbiased for general convex functions.



Properties of gradient estimator g_t [FKM'05]

$$g_t = \frac{d}{\delta}\ell_t(x_t + \delta u_t)u_t.$$

- Unbiased for linear functions.
- Nearly unbiased for general convex functions.



- Regret bounds scale with $\|g_t\|$.
- $\|g_t\|$ grows as $1/\delta$.

Gradient Descent Algorithm with two queries per round (GD2P)

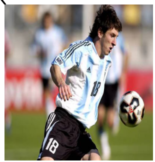
- Estimates gradient $\tilde{g}_t = \frac{d}{2\delta}(\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t))u_t$.
- Updates $x_{t+1} = \Pi_{(1-\xi)\mathcal{K}}(x_t - \eta \tilde{g}_t)$.

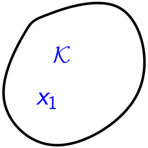
Player

Adversary

x_1

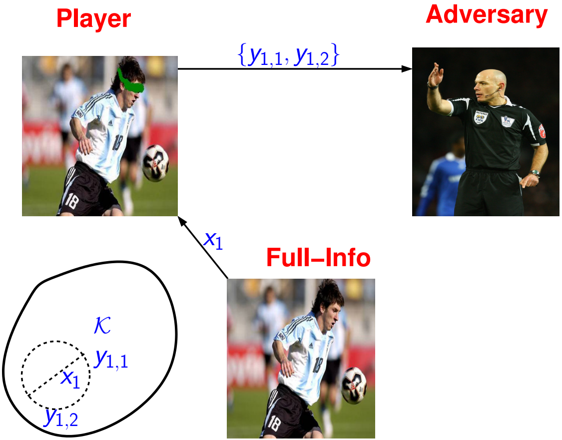
Full–Info





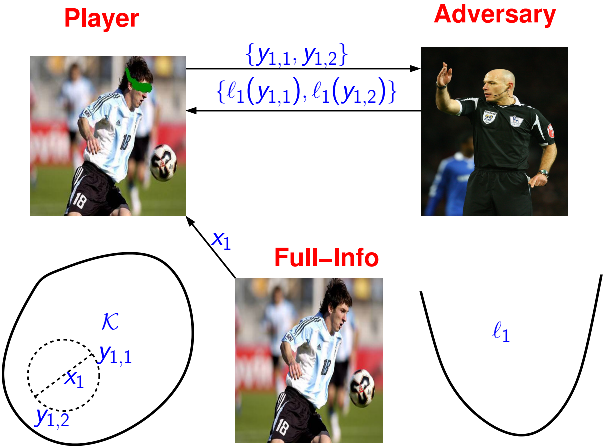
Gradient Descent Algorithm with two queries per round (GD2P)

- Estimates gradient $\tilde{g}_t = \frac{d}{2\delta}(\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t))u_t$.
- Updates $x_{t+1} = \Pi_{(1-\xi)\mathcal{K}}(x_t - \eta\tilde{g}_t)$.



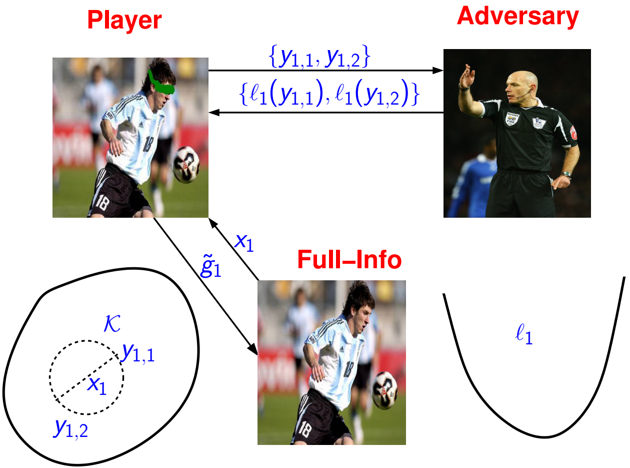
Gradient Descent Algorithm with two queries per round (GD2P)

- Estimates gradient $\tilde{g}_t = \frac{d}{2\delta}(\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t))u_t$.
- Updates $x_{t+1} = \Pi_{(1-\xi)\mathcal{K}}(x_t - \eta \tilde{g}_t)$.



Gradient Descent Algorithm with two queries per round (GD2P)

- Estimates gradient $\tilde{g}_t = \frac{d}{2\delta}(\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t))u_t$.
- Updates $x_{t+1} = \Pi_{(1-\xi)\mathcal{K}}(x_t - \eta \tilde{g}_t)$.



Properties of the gradient estimator $\tilde{g}_t$

$$g_t = \frac{d}{\delta}\ell_t(x_t + \delta u_t)u_t, \quad \tilde{g}_t = \frac{d}{2\delta}(\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t))u_t.$$

- Identical to g_t in expectation, $\mathbb{E}\tilde{g}_t = \mathbb{E}g_t$.
- Bounded norm $\|\tilde{g}_t\| \leq dG$.

$$\|\tilde{g}_t\| = \frac{d}{2\delta}\|(\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t))u_t\|$$

Properties of the gradient estimator $\tilde{g}_t$

$$g_t = \frac{d}{\delta}\ell_t(x_t + \delta u_t)u_t, \quad \tilde{g}_t = \frac{d}{2\delta}(\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t))u_t.$$

- Identical to g_t in expectation, $\mathbb{E}\tilde{g}_t = \mathbb{E}g_t$.
- Bounded norm $\|\tilde{g}_t\| \leq dG$.

$$\begin{aligned}
\|\tilde{g}_t\| &= \frac{d}{2\delta}\|(\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t))u_t\| \\
&= \frac{d}{2\delta}|\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t)|
\end{aligned}$$

Properties of the gradient estimator $\tilde{g}_t$

$$g_t = \frac{d}{\delta}\ell_t(x_t + \delta u_t)u_t, \quad \tilde{g}_t = \frac{d}{2\delta}(\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t))u_t.$$

- Identical to g_t in expectation, $\mathbb{E}\tilde{g}_t = \mathbb{E}g_t$.
- Bounded norm $\|\tilde{g}_t\| \leq dG$.

$$\begin{aligned}
\|\tilde{g}_t\| &= \frac{d}{2\delta}\|(\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t))u_t\| \\
&= \frac{d}{2\delta}|\ell_t(x_t + \delta u_t) - \ell_t(x_t - \delta u_t)| \\
&\leq \frac{dG}{2\delta}\|2\delta u_t\| = Gd.
\end{aligned}$$

Regret analysis for gradient descent with two queries

- Bounded non-empty set: $r\mathcal{B} \subseteq \mathcal{K} \subseteq D\mathcal{B}$.
- Lipschitz loss functions:
  $|\ell_t(x) - \ell_t(y)| \ \leq \ G\|x - y\|, \quad \forall x, y \in \mathcal{K}, \ \forall \ t.$
- σ_t-strong convexity:

$$\ell_t(y) \geq \ell_t(x) + \langle \nabla \ell_t(x), y - x \rangle + \frac{\sigma_t}{2}\|x - y\|^2.$$

Theorem

Under above assumptions, let $\sigma_1 > 0$. If the GD2P algorithm is run with $\eta_t = \frac{1}{\sigma_{1:t}}$, $\delta = \frac{\log T}{T}$ and $\xi = \frac{\delta}{r}$, then for any $x \in \mathcal{K}$,

$$\mathbb{E}\sum_{t=1}^{T} \frac{1}{2}(\ell_t(y_{t,1}) + \ell_t(y_{t,2})) - \mathbb{E}\sum_{t=1}^{T} \ell_t(x) \ \leq$$

$$\frac{d^2 G^2}{2} \sum_{t=1}^{T} \frac{1}{\sigma_{1:t}} + G\log(T)\left(3 + \frac{D}{r}\right).$$

Regret bound for convex, Lipschitz functions

Corollary

Suppose the set $\mathcal{K}$ is bounded and non-empty, and ℓ_t is convex, G Lipschitz for all t. If the GD2P algorithm is run with $\eta_t = \frac{1}{\sqrt{T}}$, $\delta = \frac{\log T}{T}$ and $\xi = \frac{\delta}{r}$, then

$$\mathbb{E}\sum_{t=1}^{T}\frac{1}{2}(\ell_t(y_{t,1}) + \ell_t(y_{t,2})) - \min_{x\in\mathcal{K}}\mathbb{E}\sum_{t=1}^{T}\ell_t(x) \;\leq\; (d^2G^2 + D^2)\sqrt{T} + G\log(T)\left(3 + \frac{D}{r}\right).$$

- Optimal due to matching lower bound in full-information setup.
- Bound also holds with high probability for adaptive adversaries.

Regret bound for strongly convex, Lipschitz functions

Corollary

Suppose the set $\mathcal{K}$ *is bounded and non-empty, and* ℓ_t *is* σ*-strongly convex,* G *Lipschitz for all* t*. If the GD2P algorithm is run with* $\eta_t = \frac{1}{\sigma t}$, $\delta = \frac{\log T}{T}$ *and* $\xi = \frac{\delta}{r}$*, then*

$$\mathbb{E} \sum_{t=1}^{T} \frac{1}{2}(\ell_t(y_{t,1}) + \ell_t(y_{t,2})) - \min_{x \in \mathcal{K}} \mathbb{E} \sum_{t=1}^{T} \ell_t(x) \le G \log(T) \left(\frac{d^2 G}{\sigma} + 3 + \frac{D}{r} \right).$$

- Optimal due to matching lower bound in full-information setup.

Extension to other gradient estimators

- Bounded exploration (BE): $\|x_t - y_{i,t}\| \le \delta$.
- Bounded gradient estimator (BG): $\|\tilde{g}_t\| \le G_1$.
- Approximately unbiased (AU): $\|\mathbb{E}_t \tilde{g}_t - \nabla \ell_t(x_t)\| \le c\delta$.

Theorem

Let $\mathcal{K}$ be bounded, non-empty and ℓ_t be σ_t-strongly convex with for $\sigma_1 > 0$. For any gradient estimator satisfying above conditions, the regret of GD2P algorithm is bounded as:

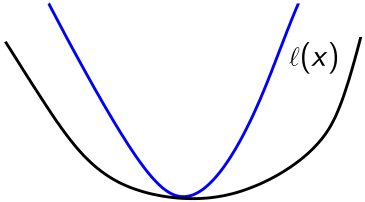
$$\mathbb{E} \sum_{t=1}^{T} \frac{1}{2}(\ell_t(y_{t,1}) + \ell_t(y_{t,2})) - \mathbb{E} \sum_{t=1}^{T} \ell_t(x) \le \frac{G_1^2}{2} \sum_{t=1}^{T} \frac{1}{\sigma_{1:t}} + G \log(T) \left(1 + 2c + \frac{D}{r}\right).$$

Analysis of other estimators for smooth functions

- Need to establish conditions (BE), (BG) and (AU).
- Smoothness assumption:

$$\ell_t(y) \leq \ell_t(x) + \langle \nabla \ell_t(x), y - x \rangle + \frac{L}{2}\|x - y\|^2.$$

- Examples:
  - Squared ℓ_p norm $\|x - \theta\|_p^2$ for $p \geq 2$.
  - Quadratic loss $(y - w^T x)^2$ for bounded x.
  - Logistic loss $\log(1 + \exp(-w^T x))$.



A Randomized Co-ordinate Descent algorithm

- Pick a co-ordinate $i_t \in \{i, \ldots, d\}$ u.a.r.
- Play $y_{t,1} = x_t + \delta e_{i_t}$, $y_{t,2} = x_t - \delta e_{i_t}$.
- Set $\tilde{g}_t = \frac{d}{2\delta}(\ell_t(y_{t,1}) - \ell_t(y_{t,2}))e_{i_t}$.

A Randomized Co-ordinate Descent algorithm

- Pick a co-ordinate $i_t \in \{i, \ldots, d\}$ u.a.r.
- Play $y_{t,1} = x_t + \delta e_{i_t}$, $y_{t,2} = x_t - \delta e_{i_t}$.
- Set $\tilde{g}_t = \frac{d}{2\delta}(\ell_t(y_{t,1}) - \ell_t(y_{t,2}))e_{i_t}$.
- (AU) holds: $\|\mathbb{E}_t \tilde{g}_t - \nabla \ell_t(x_t)\| \leq \frac{\sqrt{d} L\delta}{4}$.
- Same regret bound as before, with 1-dimensional gradient updates.

Extension to completely adaptive adversaries

- Previously needed ℓ_t independent of x_t.
- Randomization futile if ℓ_t depends on x_t.
- Can satisfy (AU) deterministically with $d+1$ queries.
- Deterministic first and second-order algorithms for smooth loss functions.
- Play the points $x_t, x_t + \delta e_i$ for $i = 1, \ldots, d$.
- Set $\tilde{g}_t = \frac{1}{\delta}\sum_{i=1}^{d}(\ell_t(x_t + \delta e_i) - \ell_t(x_t))e_i$.
- Satisfies (BE), (BG) and (AU):
  $\|\tilde{g}_t\| \leq dG$, $\|\tilde{g}_t \nabla \ell_t(x_t)\| \leq \frac{\sqrt{d}L\delta}{2}$.

Regret bounds for $d + 1$ queries

- $\mathcal{O}(\sqrt{T})$ regret for smooth, convex functions.
- $\mathcal{O}(\log T)$ regret for smooth, strongly convex functions.
- $\mathcal{O}(\log T)$ regret for smooth, exp-concave functions using quasi-Newton variant.
- Matches lower bounds from full-information setup.
- Regret bounds hold for completely adaptive adversaries.

Conclusion

- Introduce the multi-point feedback model for partial information.
- Optimal regret with high probability against adaptive adveraries using just 2 queries per round.
- Completely adaptive adversaries using $d + 1$ queries.
- Open questions:
  - One-point bandit feedback.
  - $\sqrt{T}$ lower bound for bandit strongly convex.
  - Distribution over number of queries.
  - High probability $\log(T)$ for strongly convex.

Thank You