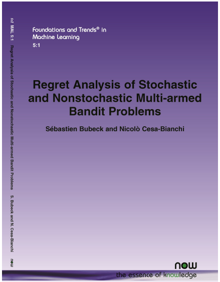
Lecture 3: Online combinatorial optimization, bandit linear optimization, and self-concordant barriers

Sébastien Bubeck

Machine Learning and Optimization group, MSR AI

Microsoft® Research



Foundations and Trends® in Machine Learning
5:1

Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems

Sébastien Bubeck and Nicolò Cesa-Bianchi

now
the essence of knowledge

Online combinatorial optimization

Parameters: action set $\mathcal{A} \subset \{a \in \{0,1\}^n : \|a\|_1 = m\}$, number of rounds T.

Online combinatorial optimization

Parameters: action set $\mathcal{A} \subset \{a \in \{0,1\}^n : \|a\|_1 = m\}$, number of rounds T.

Protocol: For each round $t \in [T]$, player chooses $a_t \in \mathcal{A}$ and simultaneously adversary chooses a loss function $\ell_t \in [0,1]^n$. Loss suffered is $\ell_t \cdot a_t$.

Online combinatorial optimization

Parameters: action set $\mathcal{A} \subset \{a \in \{0,1\}^n : \|a\|_1 = m\}$, number of rounds T.

Protocol: For each round $t \in [T]$, player chooses $a_t \in \mathcal{A}$ and simultaneously adversary chooses a loss function $\ell_t \in [0,1]^n$. Loss suffered is $\ell_t \cdot a_t$.

Feedback model: In the *full information* game the player observes the complete loss function ℓ_t. In the *bandit* game the player only observes her own loss $\ell_t \cdot a_t$. In the *semi-bandit* game one observes $a_t \odot \ell_t$.

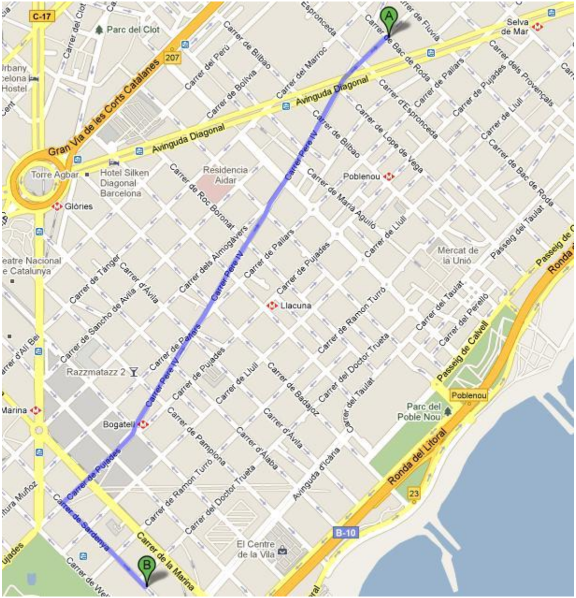
Online combinatorial optimization

Parameters: action set $\mathcal{A} \subset \{a \in \{0,1\}^n : \|a\|_1 = m\}$, number of rounds T.
Protocol: For each round $t \in [T]$, player chooses $a_t \in \mathcal{A}$ and simultaneously adversary chooses a loss function $\ell_t \in [0,1]^n$. Loss suffered is $\ell_t \cdot a_t$.
Feedback model: In the *full information* game the player observes the complete loss function ℓ_t. In the *bandit* game the player only observes her own loss $\ell_t \cdot a_t$. In the *semi-bandit* game one observes $a_t \odot \ell_t$.

Performance measure: The regret is the difference between the player's accumulated loss and the minimum loss she could have obtained had she known all the adversary's choices:

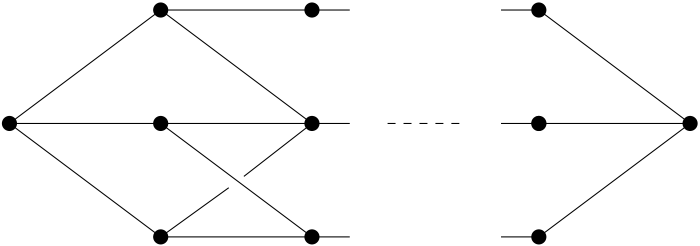
$$R_T := \mathbb{E} \sum_{t=1}^{T} \ell_t \cdot a_t - \min_{a \in \mathcal{A}} \mathbb{E} \sum_{t=1}^{T} \ell_t \cdot a \,.$$

Example: path planning



Example: path planning

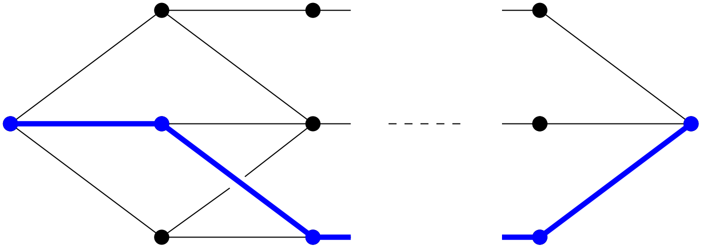
Adversary



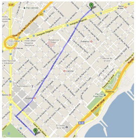
Player

Example: path planning

Adversary

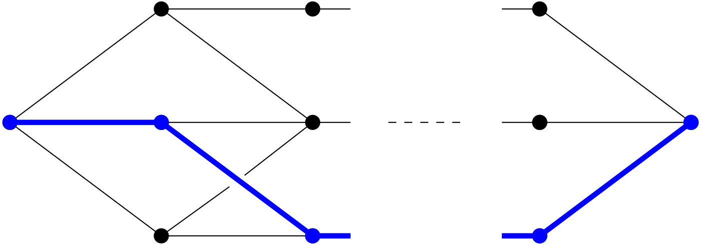


Player →

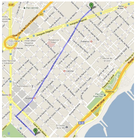


Example: path planning

Adversary ⟶

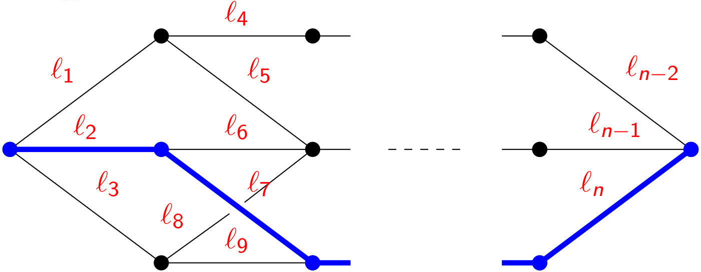


Player ⟶

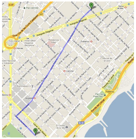


Example: path planning

Adversary →

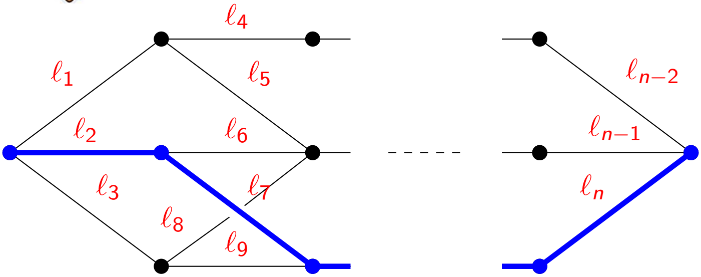


Player →

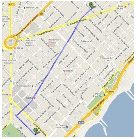


Example: path planning

Adversary ⟶



Player ⟶

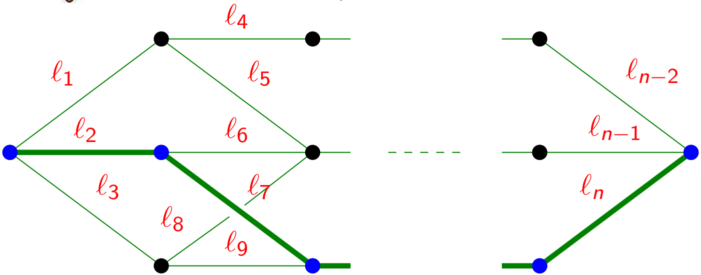


loss suffered: $\ell_2 + \ell_7 + \ldots + \ell_n$

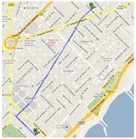
Example: path planning

Adversary ⟶

Feedback: Full Info: $\ell_1, \ell_2, \ldots, \ell_n$



Player ⟶

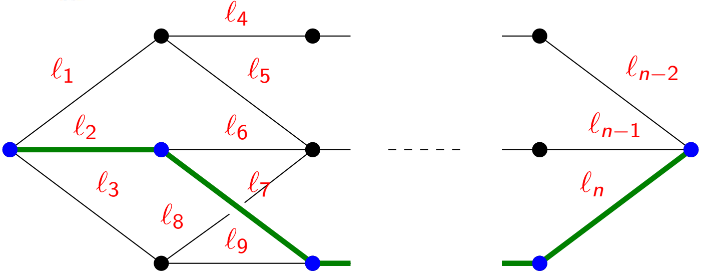


loss suffered: $\ell_2 + \ell_7 + \ldots + \ell_n$

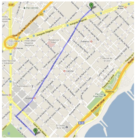
Example: path planning

Adversary ⟶

$$\text{Feedback:}\ \begin{cases} \text{Full Info:} & \ell_1, \ell_2, \ldots, \ell_n \\ \text{Semi-Bandit:} & \ell_2, \ell_7, \ldots, \ell_n \end{cases}$$



Player ⟶

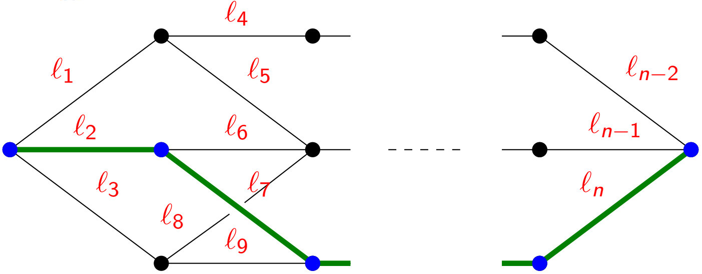


loss suffered: $\ell_2 + \ell_7 + \ldots + \ell_n$

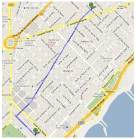
Example: path planning

Adversary ⟶

Feedback: $\begin{cases} \text{Full Info:} & \ell_1, \ell_2, \ldots, \ell_n \\ \text{Semi-Bandit:} & \ell_2, \ell_7, \ldots, \ell_n \\ \text{Bandit:} & \ell_2 + \ell_7 + \ldots + \ell_n \end{cases}$



Player ⟶



loss suffered: $\ell_2 + \ell_7 + \ldots + \ell_n$

Mirror descent and MW are now different!

Playing MW on $\mathcal{A}$ and accounting for the scale of the losses and the size of the action set one gets a $O(m\sqrt{m\log(n/m)T}) = \widetilde{O}(m^{3/2}\sqrt{T})$-regret.

Mirror descent and MW are now different!

Playing MW on $\mathcal{A}$ and accounting for the scale of the losses and the size of the action set one gets a $O(m\sqrt{m\log(n/m)T}) = \widetilde{O}(m^{3/2}\sqrt{T})$-regret.

However playing mirror descent with the negentropy regularizer on the set $\mathrm{conv}(\mathcal{A})$ gives a better bound! Indeed the variance term is controlled by m, while one can easily check that the radius term is controlled by $m\log(n/m)$, and thus one obtains a $\widetilde{O}(m\sqrt{T})$-regret.

Mirror descent and MW are now different!

Playing MW on $\mathcal{A}$ and accounting for the scale of the losses and the size of the action set one gets a $O(m\sqrt{m \log(n/m) T}) = \widetilde{O}(m^{3/2}\sqrt{T})$-regret.

However playing mirror descent with the negentropy regularizer on the set $\mathrm{conv}(\mathcal{A})$ gives a better bound! Indeed the variance term is controlled by m, while one can easily check that the radius term is controlled by $m \log(n/m)$, and thus one obtains a $\widetilde{O}(m\sqrt{T})$-regret.

This was first noticed in [Koolen, Warmuth, Kivinen 2010], and both phenomenon were shown to be "inherent" in [Audibert, B., Lugosi 2011] (in the sense that there is a lower bound of $\Omega(m^{3/2}\sqrt{T})$ for MW with *any* learning rate, and that $\Omega(m\sqrt{T})$ is a lower bound for all algorithms).

Semi-bandit [Audibert, B., Lugosi 2011, 2014]

Denote $v_t = \mathbb{E}_t a_t \in \mathrm{conv}(\mathcal{A})$. A natural unbiased estimator in this context is given by:

$$\widetilde{\ell}_t(i) = \frac{\ell_t(i) a_t(i)}{v_t(i)} \,.$$

Semi-bandit [Audibert, B., Lugosi 2011, 2014]

Denote $v_t = \mathbb{E}_t a_t \in \mathrm{conv}(\mathcal{A})$. A natural unbiased estimator in this context is given by:

$$\widetilde{\ell}_t(i) = \frac{\ell_t(i) a_t(i)}{v_t(i)} .$$

It is an easy exercise to show that the variance term for this estimator is $\leq n$, which leads to an overall regret of $\widetilde{O}(\sqrt{nmT})$. Notice that the gap between full information and semi-bandit is $\sqrt{n/m}$, which makes sense (and is optimal).

A tentative bandit estimator [Dani, Hayes, Kakade 2008]

DHK08 proposed the following (beautiful) unbiased estimator with bandit information:

$$\widetilde{\ell}_t = \Sigma_t^{-1} a_t a_t^\top \ell_t \text{ where } \Sigma_t = \mathbb{E}_{a \sim p_t}(a a^\top).$$

A tentative bandit estimator [Dani, Hayes, Kakade 2008]

DHK08 proposed the following (beautiful) unbiased estimator with bandit information:

$$\widetilde{\ell}_t = \Sigma_t^{-1} a_t a_t^\top \ell_t \text{ where } \Sigma_t = \mathbb{E}_{a \sim p_t}(a a^\top).$$

Amazingly, the variance in MW is automatically controlled:

$$\mathbb{E}(\mathbb{E}_{a \sim p_t}(\widetilde{\ell}_t^\top a)^2) = \mathbb{E}\widetilde{\ell}_t^\top \Sigma_t \widetilde{\ell}_t \leq m^2 \mathbb{E} a_t^\top \Sigma_t^{-1} a_t = m^2 \mathbb{E}\mathrm{Tr}(\Sigma_t^{-1} a_t a_t) = m^2 n\,.$$

This suggests a regret in $\widetilde{O}(m\sqrt{nmT})$, which is in fact optimal ([Koren et al 2017]). Note that this extra factor m suggests that for bandit it is enough to consider the normalization $\ell_t \cdot a_t \leq 1$, and we focus now on this case.

A tentative bandit estimator [Dani, Hayes, Kakade 2008]

DHK08 proposed the following (beautiful) unbiased estimator with bandit information:

$$\widetilde{\ell}_t = \Sigma_t^{-1} a_t a_t^\top \ell_t \text{ where } \Sigma_t = \mathbb{E}_{a \sim p_t}(a a^\top).$$

Amazingly, the variance in MW is automatically controlled:

$$\mathbb{E}(\mathbb{E}_{a \sim p_t}(\widetilde{\ell}_t^\top a)^2) = \mathbb{E}\widetilde{\ell}_t^\top \Sigma_t \widetilde{\ell}_t \leq m^2 \mathbb{E} a_t^\top \Sigma_t^{-1} a_t = m^2 \mathbb{E}\mathrm{Tr}(\Sigma_t^{-1} a_t a_t) = m^2 n\,.$$

This suggests a regret in $\widetilde{O}(m\sqrt{nmT})$, which is in fact optimal ([Koren et al 2017]). Note that this extra factor m suggests that for bandit it is enough to consider the normalization $\ell_t \cdot a_t \leq 1$, and we focus now on this case.

However there is one small issue: this estimator can take negative values, and thus the "well-conditionning" property of the entropic regularizer is not automatically verified! Resolving this issue will take us in the territory of self-concordant barriers. But first, can we gain some confidence that the claimed bound $O(\sqrt{n \log(|\mathcal{A}|) T})$ is correct?

Back to the information theoretic argument

Assume $\mathcal{A} = \{a_1, \ldots, a_{|\mathcal{A}|}\}$. Recall from Lecture 1 that Thompson Sampling satisfies

$$\sum_i p_t(i)(\bar{\ell}_t(i) - \bar{\ell}_t(i,i)) \leq \sqrt{C \sum_{i,j} p_t(i)p_t(j)(\bar{\ell}_t(i,j) - \bar{\ell}_t(i))^2}$$

$$\Rightarrow R_T \leq \sqrt{C \;\; T \;\; \log(|\mathcal{A}|)/2},$$

where $\bar{\ell}_t(i) = \mathbb{E}_t \ell_t(i)$ and $\bar{\ell}_t(i,j) = \mathbb{E}_t(\ell_t(i)|i^* = j)$.

Back to the information theoretic argument

Assume $\mathcal{A} = \{a_1, \dots, a_{|\mathcal{A}|}\}$. Recall from Lecture 1 that Thompson Sampling satisfies

$$\sum_i p_t(i)(\bar{\ell}_t(i) - \bar{\ell}_t(i,i)) \leq \sqrt{C \sum_{i,j} p_t(i)p_t(j)(\bar{\ell}_t(i,j) - \bar{\ell}_t(i))^2}$$

$$\Rightarrow R_T \leq \sqrt{C \ T \ \log(|\mathcal{A}|)/2},$$

where $\bar{\ell}_t(i) = \mathbb{E}_t \ell_t(i)$ and $\bar{\ell}_t(i,j) = \mathbb{E}_t(\ell_t(i)|i^* = j)$.

Writing $\bar{\ell}_t(i) = a_i^\top \bar{\ell}_t$, $\bar{\ell}_t(i,j) = a_i^\top \bar{\ell}_t^j$, and $(M_{i,j}) = \left(\sqrt{p_t(i)p_t(j)} a_i^\top (\bar{\ell}_t - \bar{\ell}_t^j)\right)$ we want to show that

$$\mathrm{Tr}(M) \leq \sqrt{C} \|M\|_F.$$

Back to the information theoretic argument

Assume $\mathcal{A} = \{a_1, \dots, a_{|\mathcal{A}|}\}$. Recall from Lecture 1 that Thompson Sampling satisfies

$$\sum_i p_t(i)(\bar\ell_t(i) - \bar\ell_t(i,i)) \leq \sqrt{C \ \sum_{i,j} p_t(i) p_t(j) (\bar\ell_t(i,j) - \bar\ell_t(i))^2}$$

$$\Rightarrow R_T \leq \sqrt{C \ \ T \ \ \log(|\mathcal{A}|)/2},$$

where $\bar\ell_t(i) = \mathbb{E}_t \ell_t(i)$ and $\bar\ell_t(i,j) = \mathbb{E}_t(\ell_t(i) | i^* = j)$.

Writing $\bar\ell_t(i) = a_i^\top \bar\ell_t$, $\bar\ell_t(i,j) = a_i^\top \bar\ell_t^j$, and $(M_{i,j}) = \left(\sqrt{p_t(i)p_t(j)} a_i^\top (\bar\ell_t - \bar\ell_t^j)\right)$ we want to show that

$$\mathrm{Tr}(M) \leq \sqrt{C} \|M\|_F.$$

Using the eigenvalue formula for the trace and the Frobenius norm one can see that $\mathrm{Tr}(M)^2 \leq \mathrm{rank}(M) \|M\|_F^2$.

Back to the information theoretic argument

Assume $\mathcal{A} = \{a_1, \dots, a_{|\mathcal{A}|}\}$. Recall from Lecture 1 that Thompson Sampling satisfies

$$\sum_i p_t(i)(\bar{\ell}_t(i) - \bar{\ell}_t(i,i)) \leq \sqrt{C \sum_{i,j} p_t(i)p_t(j)(\bar{\ell}_t(i,j) - \bar{\ell}_t(i))^2}$$

$$\Rightarrow R_T \leq \sqrt{C \;\; T \;\; \log(|\mathcal{A}|)/2},$$

where $\bar{\ell}_t(i) = \mathbb{E}_t \ell_t(i)$ and $\bar{\ell}_t(i,j) = \mathbb{E}_t(\ell_t(i)|i^* = j)$.

Writing $\bar{\ell}_t(i) = a_i^\top \bar{\ell}_t$, $\bar{\ell}_t(i,j) = a_i^\top \bar{\ell}_t^j$, and $(M_{i,j}) = \left(\sqrt{p_t(i)p_t(j)} a_i^\top (\bar{\ell}_t - \bar{\ell}_t^j)\right)$ we want to show that

$$\mathrm{Tr}(M) \leq \sqrt{C} \|M\|_F.$$

Using the eigenvalue formula for the trace and the Frobenius norm one can see that $\mathrm{Tr}(M)^2 \leq \mathrm{rank}(M) \|M\|_F^2$. Moreover the rank of M is at most n since $M = UV^\top$ where $U, V \in \mathbb{R}^{|\mathcal{A}| \times n}$ (the i^{th} row of U is $\sqrt{p_t(i)} a_i$ and for V it is $\sqrt{p_t(i)}(\bar{\ell}_t - \bar{\ell}_t^i)$).

Bandit linear optimization

We now come back to the general online linear optimization setting: the player plays in a convex body $K \subset \mathbb{R}^n$ and the adversary plays in $K^\circ = \{\ell : |\ell \cdot x| \leq 1, \forall x \in K\}$. An important point we have ignored so far but which matters for bandit feedback is the sampling scheme: this is a map $p : K \to \Delta(K)$ such that if MD recommends $x \in K$ then one plays at random from $p(x)$.

Bandit linear optimization

We now come back to the general online linear optimization setting: the player plays in a convex body $K \subset \mathbb{R}^n$ and the adversary plays in $K^\circ = \{\ell : |\ell \cdot x| \leq 1, \forall x \in K\}$. An important point we have ignored so far but which matters for bandit feedback is the sampling scheme: this is a map $p : K \to \Delta(K)$ such that if MD recommends $x \in K$ then one plays at random from $p(x)$. Observe that the MD-variance term for $\widetilde{\ell}_t = \Sigma_t^{-1}(a_t - x_t)a_t^\top \ell_t$ is:

$$\begin{aligned}
\mathbb{E}[(\|\widetilde{\ell}_t\|_{x_t}^*)^2] &\leq \mathbb{E}[(\|\Sigma_t^{-1}(a_t - x_t)\|_{x_t}^*)^2] \\
&= \mathbb{E}(a_t - x_t)^\top \Sigma_t^{-1} \nabla^2 \Phi(x_t)^{-1} \Sigma_t^{-1}(a_t - x_t) \\
&= \mathbb{E}\ \mathrm{Tr}(\nabla^2 \Phi(x_t)^{-1} \Sigma_t^{-1}),
\end{aligned}$$

where the last equality follows from using cyclic invariance of the trace and $\mathbb{E}[(a_t - x_t)(a_t - x_t)^\top | x_t] = \Sigma(x_t)$.

Bandit linear optimization

We now come back to the general online linear optimization setting: the player plays in a convex body $K \subset \mathbb{R}^n$ and the adversary plays in $K^\circ = \{\ell : |\ell \cdot x| \leq 1, \forall x \in K\}$. An important point we have ignored so far but which matters for bandit feedback is the sampling scheme: this is a map $p : K \to \Delta(K)$ such that if MD recommends $x \in K$ then one plays at random from $p(x)$. Observe that the MD-variance term for $\widetilde{\ell}_t = \Sigma_t^{-1}(a_t - x_t)a_t^\top \ell_t$ is:

$$\begin{aligned}
\mathbb{E}[(\|\widetilde{\ell}_t\|_{x_t}^*)^2] &\leq \mathbb{E}[(\|\Sigma_t^{-1}(a_t - x_t)\|_{x_t}^*)^2] \\
&= \mathbb{E}(a_t - x_t)^\top \Sigma_t^{-1} \nabla^2 \Phi(x_t)^{-1} \Sigma_t^{-1}(a_t - x_t) \\
&= \mathbb{E} \ \mathrm{Tr}(\nabla^2 \Phi(x_t)^{-1} \Sigma_t^{-1}) \,,
\end{aligned}$$

where the last equality follows from using cyclic invariance of the trace and $\mathbb{E}[(a_t - x_t)(a_t - x_t)^\top | x_t] = \Sigma(x_t)$.

Notice that Σ_t^{-1} has to explode when x_t tends to an extremal point of K, and thus in turns $\nabla^2 \Phi(x_t)$ would also have to explode to hope to compensate in the variance. This makes the well-conditionning problem more acute.

A small detour: Interior Point Methods

Barrier method: given $\Phi : \mathrm{int}(K) \to \mathbb{R}$ such that $\Phi(x) \to +\infty$ as $x \to \partial K$,

$$x(t) = \operatorname*{argmin}_{x \in \mathbb{R}^n} tc \cdot x + \Phi(x), \quad t \geq 0$$

A small detour: Interior Point Methods

Barrier method: given $\Phi : \mathrm{int}(K) \to \mathbb{R}$ such that $\Phi(x) \to +\infty$ as $x \to \partial K$,

$$x(t) = \operatorname*{argmin}_{x \in \mathbb{R}^n} tc \cdot x + \Phi(x), \quad t \geq 0$$

Interior point method: From $x(t)$ to $x(t')$, $t' > t$, via Newton's method (Karmakar 1984).

A small detour: Interior Point Methods

Barrier method: given $\Phi : \mathrm{int}(K) \to \mathbb{R}$ such that $\Phi(x) \to +\infty$ as $x \to \partial K$,

$$x(t) = \operatorname*{argmin}_{x \in \mathbb{R}^n} tc \cdot x + \Phi(x), \quad t \geq 0$$

Interior point method: From $x(t)$ to $x(t')$, $t' > t$, via Newton's method (Karmakar 1984). Smoothness of barrier's Hessian is critical, Nesterov and Nemirovski introduced the notion of a self-concordant function:

A small detour: Interior Point Methods

Barrier method: given $\Phi : \mathrm{int}(K) \to \mathbb{R}$ such that $\Phi(x) \to +\infty$ as $x \to \partial K$,

$$x(t) = \operatorname*{argmin}_{x \in \mathbb{R}^n} tc \cdot x + \Phi(x), \quad t \geq 0$$

Interior point method: From $x(t)$ to $x(t')$, $t' > t$, via Newton's method (Karmakar 1984). Smoothness of barrier's Hessian is critical, Nesterov and Nemirovski introduced the notion of a self-concordant function:

$$\nabla^3 \Phi(x)[h, h, h] \leq 2(\nabla^2 \Phi(x)[h, h])^{3/2}. \tag{1}$$

A small detour: Interior Point Methods

Barrier method: given $\Phi : \mathrm{int}(K) \to \mathbb{R}$ such that $\Phi(x) \to +\infty$ as $x \to \partial K$,

$$x(t) = \operatorname*{argmin}_{x \in \mathbb{R}^n} tc \cdot x + \Phi(x), \quad t \geq 0$$

Interior point method: From $x(t)$ to $x(t')$, $t' > t$, via Newton's method (Karmakar 1984). Smoothness of barrier's Hessian is critical, Nesterov and Nemirovski introduced the notion of a self-concordant function:

$$\nabla^3 \Phi(x)[h, h, h] \leq 2(\nabla^2 \Phi(x)[h, h])^{3/2}. \tag{1}$$

To control the rate at which t can be increased, one needs ν-self concordance:

A small detour: Interior Point Methods

Barrier method: given $\Phi : \mathrm{int}(K) \to \mathbb{R}$ such that $\Phi(x) \to +\infty$ as $x \to \partial K$,

$$x(t) = \operatorname*{argmin}_{x \in \mathbb{R}^n} tc \cdot x + \Phi(x), \quad t \geq 0$$

Interior point method: From $x(t)$ to $x(t')$, $t' > t$, via Newton's method (Karmakar 1984). Smoothness of barrier's Hessian is critical, Nesterov and Nemirovski introduced the notion of a self-concordant function:

$$\nabla^3 \Phi(x)[h, h, h] \leq 2(\nabla^2 \Phi(x)[h, h])^{3/2}. \tag{1}$$

To control the rate at which t can be increased, one needs ν-self concordance:

$$\nabla \Phi(x)[h] \leq \sqrt{\nu \cdot \nabla^2 \Phi(x)[h, h]}. \tag{2}$$

A small detour: Interior Point Methods

Barrier method: given $\Phi : \mathrm{int}(K) \to \mathbb{R}$ such that $\Phi(x) \to +\infty$ as $x \to \partial K$,

$$x(t) = \operatorname*{argmin}_{x \in \mathbb{R}^n} tc \cdot x + \Phi(x), \quad t \geq 0$$

Interior point method: From $x(t)$ to $x(t')$, $t' > t$, via Newton's method (Karmakar 1984). Smoothness of barrier's Hessian is critical, Nesterov and Nemirovski introduced the notion of a self-concordant function:

$$\nabla^3 \Phi(x)[h, h, h] \leq 2(\nabla^2 \Phi(x)[h, h])^{3/2}. \tag{1}$$

To control the rate at which t can be increased, one needs ν-self concordance:

$$\nabla \Phi(x)[h] \leq \sqrt{\nu \cdot \nabla^2 \Phi(x)[h, h]}. \tag{2}$$

Theorem (Nesterov and Nemirovski 1989)

$\exists$ a $O(n)$-s.c.b. For $K = [-1, 1]^n$ any ν-s.c.b. satisfies $\nu \geq n$.

Basic properties of self-concordant barriers

Theorem

1. *If Φ is ν-self-concordant then for any $x, y \in \mathrm{int}(K)$,*

$$\Phi(y) - \Phi(x) \leq \nu \log \left(\frac{1}{1 - \pi_x(y)} \right),$$

where $\pi_x(y)$ is the Minkowski gauge, i.e.,
$\pi_x(y) = \inf\{t > 0 : x + \frac{1}{t}(y - x) \in K\}$*.*

2. *Φ is self-concordant if and only if Φ^* is self-concordant.*
3. *If Φ is self-concordant then for any $x \in \mathrm{int}(\mathcal{K})$ and h such that $\|h\|_x < 1$ and $x + h \in \mathrm{int}(K)$,*

$$D_\Phi(x + h, x) \leq \frac{\|h\|_x^2}{1 - \|h\|_x}.$$

4. *If Φ is a self-concordant barrier then for any $x \in \mathrm{int}(K)$,*
$\{x + h : \|h\|_x \leq 1\} \subset K$*.*

Abernethy-Hazan-Rakhlin sampling scheme

Given a point $x \in \mathrm{int}(\mathcal{K})$ let $p(x)$ be uniform on the boundary of the Dikin ellipsoid $\{x + h : \|h\|_x \leq 1\}$ (this is valid by property 4).

Abernethy-Hazan-Rakhlin sampling scheme

Given a point $x \in \mathrm{int}(\mathcal{K})$ let $p(x)$ be uniform on the boundary of the Dikin ellipsoid $\{x + h : \|h\|_x \leq 1\}$ (this is valid by property 4). Another description of p is as follows: let U be uniform on the $n-1$ dimensional sphere $\{u \in \mathbb{R}^n : |u| = 1\}$ and $X = x + \nabla^2 \Phi(x)^{-1/2} U$, then X has law $p(x)$. In particular with this description we readily see that $\Sigma(x) = \frac{1}{n} \nabla^2 \Phi(x)^{-1}$ (since $\mathbb{E}\ UU^\top = \frac{1}{n} I_n$).

Abernethy-Hazan-Rakhlin sampling scheme

Given a point $x \in \mathrm{int}(\mathcal{K})$ let $p(x)$ be uniform on the boundary of the Dikin ellipsoid $\{x + h : \|h\|_x \leq 1\}$ (this is valid by property 4). Another description of p is as follows: let U be uniform on the $n-1$ dimensional sphere $\{u \in \mathbb{R}^n : |u| = 1\}$ and $X = x + \nabla^2 \Phi(x)^{-1/2} U$, then X has law $p(x)$. In particular with this description we readily see that $\Sigma(x) = \frac{1}{n} \nabla^2 \Phi(x)^{-1}$ (since $\mathbb{E}\ UU^\top = \frac{1}{n} I_n$).

We can now bound (almost surely) the dual local norm of the loss estimator as follows (we write $a_t = x_t + \nabla^2 \Phi(x)^{-1/2} u_t$)

$$\|\widetilde{\ell}_t\|_{x_t}^* \leq \|\Sigma(x_t)^{-1}(a_t - x_t)\|_{x_t}^* = n \|\nabla^2 \Phi(x_t)^{1/2} u_t\|_{x_t}^* = n|u_t| = n.$$

Abernethy-Hazan-Rakhlin sampling scheme

Given a point $x \in \mathrm{int}(\mathcal{K})$ let $p(x)$ be uniform on the boundary of the Dikin ellipsoid $\{x + h : \|h\|_x \leq 1\}$ (this is valid by property 4). Another description of p is as follows: let U be uniform on the $n-1$ dimensional sphere $\{u \in \mathbb{R}^n : |u| = 1\}$ and $X = x + \nabla^2 \Phi(x)^{-1/2} U$, then X has law $p(x)$. In particular with this description we readily see that $\Sigma(x) = \frac{1}{n} \nabla^2 \Phi(x)^{-1}$ (since $\mathbb{E}\ UU^\top = \frac{1}{n} I_n$).
We can now bound (almost surely) the dual local norm of the loss estimator as follows (we write $a_t = x_t + \nabla^2 \Phi(x)^{-1/2} u_t$)

$$\|\widetilde{\ell}_t\|_{x_t}^* \leq \|\Sigma(x_t)^{-1}(a_t - x_t)\|_{x_t}^* = n\|\nabla^2 \Phi(x_t)^{1/2} u_t\|_{x_t}^* = n|u_t| = n.$$

In particular we get the well-conditioning as soon as $\eta \leq 1/n$ (by property 3), and the regret bound is of the form (using property 1) $\nu \log(T)/\eta + n^2 \eta$, that is $\widetilde{O}(n\sqrt{\nu T})$.

The entropic barrier

Canonical exponential family on K: $\{p_\theta, \theta \in \mathbb{R}^n\}$ where

$$\frac{dp_\theta}{dx} = \frac{1}{Z(\theta)} \exp(\langle \theta, x \rangle) \mathbb{1}\{x \in K\}.$$

The entropic barrier

Canonical exponential family on K: $\{p_\theta, \theta \in \mathbb{R}^n\}$ where

$$\frac{dp_\theta}{dx} = \frac{1}{Z(\theta)} \exp(\langle \theta, x \rangle) \mathbb{1}\{x \in K\}.$$

For $x \in \mathrm{int}(K)$ let $\theta(x)$ be such that $\mathbb{E}_{X \sim p_{\theta(x)}} X = x$.

The entropic barrier

Canonical exponential family on K: $\{p_\theta, \theta \in \mathbb{R}^n\}$ where

$$\frac{dp_\theta}{dx} = \frac{1}{Z(\theta)} \exp(\langle \theta, x \rangle) \mathbb{1}\{x \in K\}.$$

For $x \in \mathrm{int}(K)$ let $\theta(x)$ be such that $\mathbb{E}_{X \sim p_{\theta(x)}} X = x$.

Theorem (B. and Eldan 2015)

$\mathbbm{e} : x \mapsto -H(p_{\theta(x)})$ is a $(1 + o(1))n$-s.c.b.
Moreover it gives a regret for BLO in $\widetilde{O}(n\sqrt{T})$.

The entropic barrier

Canonical exponential family on K: $\{p_\theta, \theta \in \mathbb{R}^n\}$ where

$$\frac{dp_\theta}{dx} = \frac{1}{Z(\theta)} \exp(\langle \theta, x \rangle) \mathbb{1}\{x \in K\}.$$

For $x \in \mathrm{int}(K)$ let $\theta(x)$ be such that $\mathbb{E}_{X \sim p_{\theta(x)}} X = x$.

Theorem (B. and Eldan 2015)

$\mathfrak{e} : x \mapsto -H(p_{\theta(x)})$ is a $(1 + o(1))n$-s.c.b.
Moreover it gives a regret for BLO in $\widetilde{O}(n\sqrt{T})$.

Proof.

(i)
(ii)
(iii)

(iv)

□

The entropic barrier

Canonical exponential family on K: $\{p_\theta, \theta \in \mathbb{R}^n\}$ where

$$\frac{dp_\theta}{dx} = \frac{1}{Z(\theta)} \exp(\langle \theta, x \rangle) \mathbb{1}\{x \in K\}.$$

For $x \in \mathrm{int}(K)$ let $\theta(x)$ be such that $\mathbb{E}_{X \sim p_{\theta(x)}} X = x$.

Theorem (B. and Eldan 2015)

$\mathbb{e} : x \mapsto -H(p_{\theta(x)})$ is a $(1 + o(1))n$-s.c.b.
Moreover it gives a regret for BLO in $\widetilde{O}(n\sqrt{T})$.

Proof.

(i) self-concordance is invariant by Fenchel duality
(ii)
(iii)

(iv)

$\square$

The entropic barrier

Canonical exponential family on K: $\{p_\theta, \theta \in \mathbb{R}^n\}$ where

$$\frac{dp_\theta}{dx} = \frac{1}{Z(\theta)} \exp(\langle \theta, x \rangle) \mathbb{1}\{x \in K\}.$$

For $x \in \mathrm{int}(K)$ let $\theta(x)$ be such that $\mathbb{E}_{X \sim p_{\theta(x)}} X = x$.

Theorem (B. and Eldan 2015)

$\mathbb{e} : x \mapsto -H(p_{\theta(x)})$ is a $(1 + o(1))n$-s.c.b.
Moreover it gives a regret for BLO in $\widetilde{O}(n\sqrt{T})$.

Proof.

(i) self-concordance is invariant by Fenchel duality
(ii) $\nabla^k \mathbb{e}^*(x) = \mathbb{E}_{X \sim p_{\theta(x)}} (X - \mathbb{E}X)^{\otimes k}$ for $k \in \{1, 2, 3\}$.
(iii)

(iv)

□

The entropic barrier

Canonical exponential family on K: $\{p_\theta, \theta \in \mathbb{R}^n\}$ where

$$\frac{dp_\theta}{dx} = \frac{1}{Z(\theta)} \exp(\langle \theta, x \rangle) \mathbb{1}\{x \in K\}.$$

For $x \in \mathrm{int}(K)$ let $\theta(x)$ be such that $\mathbb{E}_{X \sim p_{\theta(x)}} X = x$.

Theorem (B. and Eldan 2015)

$\mathbb{e} : x \mapsto -H(p_{\theta(x)})$ *is a* $(1 + o(1))n$*-s.c.b.*
Moreover it gives a regret for BLO in $\widetilde{O}(n\sqrt{T})$*.*

Proof.

(i) self-concordance is invariant by Fenchel duality
(ii) $\nabla^k \mathbb{e}^*(x) = \mathbb{E}_{X \sim p_{\theta(x)}} (X - \mathbb{E}X)^{\otimes k}$ for $k \in \{1, 2, 3\}$.
(iii) X log-concave
$\Rightarrow \mathbb{E}(X - \mathbb{E}X)^{\otimes 3}[h, h, h] \leq 2 \left(\mathbb{E}(X - \mathbb{E}X)^{\otimes 2}[h, h] \right)^{3/2}$
(iv)

□

The entropic barrier

Canonical exponential family on K: $\{p_\theta, \theta \in \mathbb{R}^n\}$ where

$$\frac{dp_\theta}{dx} = \frac{1}{Z(\theta)} \exp(\langle \theta, x \rangle) \mathbb{1}\{x \in K\}.$$

For $x \in \mathrm{int}(K)$ let $\theta(x)$ be such that $\mathbb{E}_{X \sim p_{\theta(x)}} X = x$.

Theorem (B. and Eldan 2015)

$\mathbb{e} : x \mapsto -H(p_{\theta(x)})$ *is a* $(1 + o(1))n$*-s.c.b.*
Moreover it gives a regret for BLO in $\widetilde{O}(n\sqrt{T})$.

Proof.

(i) self-concordance is invariant by Fenchel duality
(ii) $\nabla^k \mathbb{e}^*(x) = \mathbb{E}_{X \sim p_{\theta(x)}} (X - \mathbb{E}X)^{\otimes k}$ for $k \in \{1, 2, 3\}$.
(iii) X log-concave
$\Rightarrow \mathbb{E}(X - \mathbb{E}X)^{\otimes 3}[h, h, h] \leq 2 \left(\mathbb{E}(X - \mathbb{E}X)^{\otimes 2}[h, h]\right)^{3/2}$
(iv) Brunn-Minkowski ⇒ "sub-CLT" for p_θ ⇒ ν-s.c (bit more involved than (i)-(ii)-(iii)) $\square$

(iv) in a nutshell

$$\begin{aligned}
&\nabla \mathbb{e}(x)[h] \leq \sqrt{\nu \cdot \nabla^2 \mathbb{e}(x)[h,h]} \\
&\Leftrightarrow [\nabla^2 \mathbb{e}(x)]^{-1}[\nabla \mathbb{e}(x), \nabla \mathbb{e}(x)] \leq \nu \\
&\Leftrightarrow \operatorname{Cov}(p_\theta)[\theta, \theta] \leq \nu \\
&\Leftrightarrow \operatorname{Var}(Y) \leq \frac{\nu}{|\theta|^2} \text{ where } Y = \langle X, \theta/|\theta| \rangle, X \sim p_\theta
\end{aligned}$$

(iv) in a nutshell

$$\begin{aligned}
&\nabla \mathbb{e}(x)[h] \leq \sqrt{\nu \cdot \nabla^2 \mathbb{e}(x)[h,h]} \\
&\Leftrightarrow [\nabla^2 \mathbb{e}(x)]^{-1}[\nabla \mathbb{e}(x), \nabla \mathbb{e}(x)] \leq \nu \\
&\Leftrightarrow \mathrm{Cov}(p_\theta)[\theta,\theta] \leq \nu \\
&\Leftrightarrow \mathrm{Var}(Y) \leq \frac{\nu}{|\theta|^2} \text{ where } Y = \langle X, \theta/|\theta| \rangle, X \sim p_\theta
\end{aligned}$$

Let u be the log-density of Y and v the log-marginal of the uniform measure on K in the direction $\theta/|\theta|$, that is $u(y) = v(y) + y|\theta| + cst$.

(iv) in a nutshell

$$\begin{aligned}
&\nabla \mathbb{e}(x)[h] \leq \sqrt{\nu \cdot \nabla^2 \mathbb{e}(x)[h,h]} \\
&\Leftrightarrow [\nabla^2 \mathbb{e}(x)]^{-1}[\nabla \mathbb{e}(x), \nabla \mathbb{e}(x)] \leq \nu \\
&\Leftrightarrow \mathrm{Cov}(p_\theta)[\theta,\theta] \leq \nu \\
&\Leftrightarrow \mathrm{Var}(Y) \leq \frac{\nu}{|\theta|^2} \text{ where } Y = \langle X, \theta/|\theta| \rangle, X \sim p_\theta
\end{aligned}$$

Let u be the log-density of Y and v the log-marginal of the uniform measure on K in the direction $\theta/|\theta|$, that is
$u(y) = v(y) + y|\theta| + cst$.
By Brunn-Minkowski $v'' \leq -\frac{1}{n}(v')^2$

(iv) in a nutshell

$$\begin{aligned}
&\nabla \mathbb{e}(x)[h] \leq \sqrt{\nu \cdot \nabla^2 \mathbb{e}(x)[h,h]} \\
&\Leftrightarrow [\nabla^2 \mathbb{e}(x)]^{-1}[\nabla \mathbb{e}(x), \nabla \mathbb{e}(x)] \leq \nu \\
&\Leftrightarrow \mathrm{Cov}(p_\theta)[\theta,\theta] \leq \nu \\
&\Leftrightarrow \mathrm{Var}(Y) \leq \frac{\nu}{|\theta|^2} \text{ where } Y = \langle X, \theta/|\theta| \rangle, X \sim p_\theta
\end{aligned}$$

Let u be the log-density of Y and v the log-marginal of the uniform measure on K in the direction $\theta/|\theta|$, that is $u(y) = v(y) + y|\theta| + cst$.
By Brunn-Minkowski $v'' \leq -\frac{1}{n}(v')^2$ and so

$$u'' \leq -\frac{1}{n}(u' - |\theta|)^2,$$

(iv) in a nutshell

$$
\begin{aligned}
&\nabla \mathbb{e}(x)[h] \leq \sqrt{\nu \cdot \nabla^2 \mathbb{e}(x)[h, h]} \\
&\Leftrightarrow [\nabla^2 \mathbb{e}(x)]^{-1}[\nabla \mathbb{e}(x), \nabla \mathbb{e}(x)] \leq \nu \\
&\Leftrightarrow \mathrm{Cov}(p_\theta)[\theta, \theta] \leq \nu \\
&\Leftrightarrow \mathrm{Var}(Y) \leq \frac{\nu}{|\theta|^2} \text{ where } Y = \langle X, \theta / |\theta| \rangle, X \sim p_\theta
\end{aligned}
$$

Let u be the log-density of Y and v the log-marginal of the uniform measure on K in the direction $\theta/|\theta|$, that is
$u(y) = v(y) + y|\theta| + cst$.
By Brunn-Minkowski $v'' \leq -\frac{1}{n}(v')^2$ and so

$$u'' \leq -\frac{1}{n}(u' - |\theta|)^2,$$

which implies for any y close enough to the maximum y_0 of u,

$$u(y) \leq -\frac{|y - y_0|^2}{2n/|\theta|^2} + cst.$$

Beyond BLO: Bandit Convex Optimization [Flaxman, Kalai, McMahan 2004; Kleinberg 2004]

We now assume that the adversary plays a Lipschitz *convex function* $\ell_t : K \to [0, 1]$.

Beyond BLO: Bandit Convex Optimization [Flaxman, Kalai, McMahan 2004; Kleinberg 2004]

We now assume that the adversary plays a Lipschitz *convex function* $\ell_t : K \to [0,1]$.

It turns out that we might as well assume that the adversary plays the linear function $\nabla \ell_t(x_t)$ in the sense that:

$$\ell_t(x_t) - \ell_t(x) \leq \nabla \ell_t(x_t) \cdot (x_t - x) .$$

In particular online convex optimization with full information simply reduces to online linear optimization.

Beyond BLO: Bandit Convex Optimization [Flaxman, Kalai, McMahan 2004; Kleinberg 2004]

We now assume that the adversary plays a Lipschitz *convex function* $\ell_t : K \to [0, 1]$.

It turns out that we might as well assume that the adversary plays the linear function $\nabla \ell_t(x_t)$ in the sense that:

$$\ell_t(x_t) - \ell_t(x) \leq \nabla \ell_t(x_t) \cdot (x_t - x) \,.$$

In particular online convex optimization with full information simply reduces to online linear optimization.

However with bandit feedback the scenario becomes different: given access to a value of the function, can we give an unbiased estimator with low variance of the *gradient*?

BCO via small perturbations

Say that given $\ell_t(a_t)$ with $a_t \sim p_t(x_t)$ we obtain $\widetilde{g}_t$ such that $\mathbb{E}_t \widetilde{g}_t = \nabla \ell_t(x_t)$, then we have:

$$\begin{aligned}
\mathbb{E} \sum_{t=1}^{T} (\ell_t(a_t) - \ell_t(x)) &\leq \mathbb{E} \sum_{t=1}^{T} (\ell_t(x_t) - \ell_t(x) + \|a_t - x_t\|) \\
&\leq \mathbb{E} \sum_{t=1}^{T} (\nabla \ell_t(x_t) \cdot (x_t - x) + \|a_t - x_t\|) \\
&\leq \mathbb{E} \sum_{t=1}^{T} (\widetilde{g}_t \cdot (x_t - x) + \|a_t - x_t\|) \, .
\end{aligned}$$

BCO via small perturbations

Say that given $\ell_t(a_t)$ with $a_t \sim p_t(x_t)$ we obtain $\widetilde{g}_t$ such that $\mathbb{E}_t \widetilde{g}_t = \nabla \ell_t(x_t)$, then we have:

$$
\begin{aligned}
\mathbb{E} \sum_{t=1}^{T} (\ell_t(a_t) - \ell_t(x)) & \leq \mathbb{E} \sum_{t=1}^{T} (\ell_t(x_t) - \ell_t(x) + \|a_t - x_t\|) \\
& \leq \mathbb{E} \sum_{t=1}^{T} (\nabla \ell_t(x_t) \cdot (x_t - x) + \|a_t - x_t\|) \\
& \leq \mathbb{E} \sum_{t=1}^{T} (\widetilde{g}_t \cdot (x_t - x) + \|a_t - x_t\|) \, .
\end{aligned}
$$

Using mirror descent on $\widetilde{g}_t$ we are left with controlling $\mathbb{E} \|\widetilde{g}_t\|^2$.

BCO via small perturbations

Say that given $\ell_t(a_t)$ with $a_t \sim p_t(x_t)$ we obtain $\widetilde{g}_t$ such that $\mathbb{E}_t \widetilde{g}_t = \nabla \ell_t(x_t)$, then we have:

$$\begin{aligned}
\mathbb{E} \sum_{t=1}^{T} (\ell_t(a_t) - \ell_t(x)) & \leq \mathbb{E} \sum_{t=1}^{T} (\ell_t(x_t) - \ell_t(x) + \|a_t - x_t\|) \\
& \leq \mathbb{E} \sum_{t=1}^{T} (\nabla \ell_t(x_t) \cdot (x_t - x) + \|a_t - x_t\|) \\
& \leq \mathbb{E} \sum_{t=1}^{T} (\widetilde{g}_t \cdot (x_t - x) + \|a_t - x_t\|) \, .
\end{aligned}$$

Using mirror descent on $\widetilde{g}_t$ we are left with controlling $\mathbb{E}\|\widetilde{g}_t\|^2$.

Question: how to get a gradient estimate at a point x with a value function estimate at a small perturbation of x? Answer: divergence theorem!

One-point gradient estimator

Lemma

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, B the unit ball in $\mathbb{R}^n$, and σ the normalized Haar measure on the sphere ∂B. Then one has

$$\nabla \int_B f(u)du = n \int_{\partial B} f(u)u \, d\sigma(u) \, .$$

One-point gradient estimator

Lemma

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, B the unit ball in $\mathbb{R}^n$, and σ the normalized Haar measure on the sphere ∂B. Then one has

$$\nabla \int_B f(u)du = n \int_{\partial B} f(u)u \, d\sigma(u) \, .$$

In particular define $\bar{\ell}_t(x) = \ell_t(x + \varepsilon u)$ where u is uniform in B. Then one has $\nabla \bar{\ell}_t(x) = \frac{n}{\varepsilon} \mathbb{E} \, \ell_t(x + \varepsilon v)v$ with $v = u/\|u\|$.

One-point gradient estimator

Lemma

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, B the unit ball in $\mathbb{R}^n$, and σ the normalized Haar measure on the sphere ∂B. Then one has

$$\nabla \int_B f(u) du = n \int_{\partial B} f(u) u \, d\sigma(u) .$$

In particular define $\bar{\ell}_t(x) = \ell_t(x + \varepsilon u)$ where u is uniform in B. Then one has $\nabla \bar{\ell}_t(x) = \frac{n}{\varepsilon} \mathbb{E} \, \ell_t(x + \varepsilon v) v$ with $v = u / \|u\|$.

Playing $a_t = x_t + \varepsilon v_t$ and setting $\widetilde{g}_t = \frac{n}{\varepsilon} \ell_t(a_t) v_t$ one obtains a regret in

$$O \left(\varepsilon T + \eta T \frac{n^2}{\varepsilon^2} + \frac{1}{\eta} \right) .$$

One-point gradient estimator

Lemma

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, B the unit ball in $\mathbb{R}^n$, and σ the normalized Haar measure on the sphere ∂B. Then one has

$$\nabla \int_B f(u)du = n \int_{\partial B} f(u)u \, d\sigma(u) \, .$$

In particular define $\overline{\ell}_t(x) = \ell_t(x + \varepsilon u)$ where u is uniform in B. Then one has $\nabla \overline{\ell}_t(x) = \frac{n}{\varepsilon} \mathbb{E} \, \ell_t(x + \varepsilon v)v$ with $v = u/\|u\|$.

Playing $a_t = x_t + \varepsilon v_t$ and setting $\widetilde{g}_t = \frac{n}{\varepsilon} \ell_t(a_t)v_t$ one obtains a regret in

$$O\left(\varepsilon T + \eta T \frac{n^2}{\varepsilon^2} + \frac{1}{\eta}\right) \, .$$

Optimizing the parameters yields a regret in $O(n^{1/2}T^{3/4})$.

The quest for $\sqrt{T}$-BCO

For a decade the $T^{3/4}$ remained the state of the art, despite many attempts by the community. Some partial progress on the way was obtained by making further assumptions (smoothness, strong convexity, dimension 1). The first proof that $\sqrt{T}$ is achievable was via the information theoretic argument and the following geometric theorem:

The quest for $\sqrt{T}$-BCO

For a decade the $T^{3/4}$ remained the state of the art, despite many attempts by the community. Some partial progress on the way was obtained by making further assumptions (smoothness, strong convexity, dimension 1). The first proof that $\sqrt{T}$ is achievable was via the information theoretic argument and the following geometric theorem:

Theorem (B. and Eldan 2015)

Let $f : K \to [0, +\infty)$ be convex and 1-Lipschitz, and $\varepsilon > 0$. There exists a probability measure μ on K such that the following holds true. For every $\alpha \in K$ and for every convex and 1-Lipschitz function $g : K \to \mathbb{R}$ satisfying $g(\alpha) < -\varepsilon$, one has

$$\mu\left(\left\{x \in K : |f(x) - g(x)| > \widetilde{O}\left(\frac{\varepsilon}{n^{7.5}}\right)\right\}\right) > \widetilde{O}\left(\frac{1}{n^3}\right).$$

The quest for $\sqrt{T}$-BCO

For a decade the $T^{3/4}$ remained the state of the art, despite many attempts by the community. Some partial progress on the way was obtained by making further assumptions (smoothness, strong convexity, dimension 1). The first proof that $\sqrt{T}$ is achievable was via the information theoretic argument and the following geometric theorem:

Theorem (B. and Eldan 2015)

Let $f : K \to [0, +\infty)$ be convex and 1-Lipschitz, and $\varepsilon > 0$. There exists a probability measure μ on K such that the following holds true. For every $\alpha \in K$ and for every convex and 1-Lipschitz function $g : K \to \mathbb{R}$ satisfying $g(\alpha) < -\varepsilon$, one has

$$\mu\left(\left\{x \in K : |f(x) - g(x)| > \widetilde{O}\left(\frac{\varepsilon}{n^{7.5}}\right)\right\}\right) > \widetilde{O}\left(\frac{1}{n^3}\right).$$

Later Hazan and Li provided an algorithm with regret in $\exp(\mathrm{poly}(n))\sqrt{T}$. In the final lecture we will discuss the efficient algorithm by B., Eldan and Lee which obtains $\widetilde{O}(n^{9.5}\sqrt{T})$ regret.