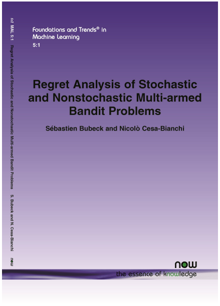
Lecture 1:
Introduction to regret analysis

Sébastien Bubeck
Machine Learning and Optimization group, MSR AI

Microsoft Research



Basic setting of online learning

Basic setting of online learning

Parameters: finite set of actions $[n]$ and number of rounds $T \geq n$.

Basic setting of online learning

Parameters: finite set of actions $[n]$ and number of rounds $T \geq n$.
Protocol: For each round $t \in [T]$, player chooses $i_t \in [n]$ and simultaneously adversary chooses a loss function $\ell_t : [n] \to [0, 1]$.

Basic setting of online learning

Parameters: finite set of actions $[n]$ and number of rounds $T \geq n$.
Protocol: For each round $t \in [T]$, player chooses $i_t \in [n]$ and simultaneously adversary chooses a loss function $\ell_t : [n] \to [0,1]$.

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 $\ell_t(i_t)$.

Basic setting of online learning

Parameters: finite set of actions $[n]$ and number of rounds $T \geq n$.
Protocol: For each round $t \in [T]$, player chooses $i_t \in [n]$ and simultaneously adversary chooses a loss function $\ell_t : [n] \to [0, 1]$.

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 $\ell_t(i_t)$.

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:

$$R_T := \mathbb{E} \sum_{t=1}^{T} \ell_t(i_t) - \min_{i \in [n]} \mathbb{E} \sum_{t=1}^{T} \ell_t(i) =: L_T - \min_{i \in [n]} L_{i,T} \,.$$

Basic setting of online learning

Parameters: finite set of actions $[n]$ and number of rounds $T \geq n$.
Protocol: For each round $t \in [T]$, player chooses $i_t \in [n]$ and simultaneously adversary chooses a loss function $\ell_t : [n] \to [0,1]$.

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 $\ell_t(i_t)$.

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:

$$R_T := \mathbb{E} \sum_{t=1}^{T} \ell_t(i_t) - \min_{i \in [n]} \mathbb{E} \sum_{t=1}^{T} \ell_t(i) =: L_T - \min_{i \in [n]} L_{i,T} .$$

What's it about? Full information game is about *hedging*, while bandit game also features the fundamental tension between *exploration* and *exploitation*.

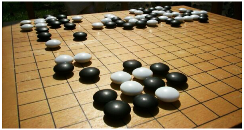
Applications

These challenges (scarce feedback, robustness to non i.i.d. data, exploration vs exploitation) are crucial components of many practical problems, hence the success of online learning and bandit theory!

Applications

These challenges (scarce feedback, robustness to non i.i.d. data, exploration vs exploitation) are crucial components of many practical problems, hence the success of online learning and bandit theory!

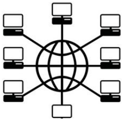
AI for games



Brain computer interface

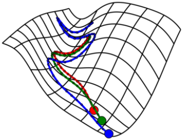
Medical trials

Packets routing



Ad placement

Hyperparameter opt



Hedging with multiplicative weights [Freund and Schapire 96, Littlestone and Warmuth 94, Vovk 90]

Assume for simplicity $\ell_t(i) \in \{0, 1\}$. MW keeps weights $w_{i,t}$ for each action, plays from normalized weights, and update as follows:

$$w_{i,t+1} = (1 - \eta \ell_t(i)) w_{i,t} \,.$$

Hedging with multiplicative weights [Freund and Schapire 96, Littlestone and Warmuth 94, Vovk 90]

Assume for simplicity $\ell_t(i) \in \{0, 1\}$. MW keeps weights $w_{i,t}$ for each action, plays from normalized weights, and update as follows:

$$w_{i,t+1} = (1 - \eta \ell_t(i)) w_{i,t} .$$

Key insight: if i^* does not make a mistake on round t then we get "closer" to δ_{i^*} (i.e., we learn), and otherwise we might get confused but i^* had to pay for it.

Hedging with multiplicative weights [Freund and Schapire 96, Littlestone and Warmuth 94, Vovk 90]

Assume for simplicity $\ell_t(i) \in \{0, 1\}$. MW keeps weights $w_{i,t}$ for each action, plays from normalized weights, and update as follows:

$$w_{i,t+1} = (1 - \eta \ell_t(i)) w_{i,t} \,.$$

Key insight: if i^* does not make a mistake on round t then we get "closer" to δ_{i^*} (i.e., we learn), and otherwise we might get confused but i^* had to pay for it.

Theorem

For any $\eta \in [0, 1/2]$ *and* $i \in [n]$,

$$L_T \leq (1 + \eta) L_{i,T} + \frac{\log(n)}{\eta} \,.$$

By optimizing η *one gets* $R_T \leq 2\sqrt{T \log(n)}$.

Hedging with multiplicative weights [Freund and Schapire 96, Littlestone and Warmuth 94, Vovk 90]

Assume for simplicity $\ell_t(i) \in \{0, 1\}$. MW keeps weights $w_{i,t}$ for each action, plays from normalized weights, and update as follows:

$$w_{i,t+1} = (1 - \eta \ell_t(i)) w_{i,t} .$$

Key insight: if i^* does not make a mistake on round t then we get "closer" to δ_{i^*} (i.e., we learn), and otherwise we might get confused but i^* had to pay for it.

Theorem

For any $\eta \in [0, 1/2]$ and $i \in [n]$,

$$L_T \leq (1 + \eta) L_{i,T} + \frac{\log(n)}{\eta} .$$

By optimizing η one gets $R_T \leq 2\sqrt{T \log(n)}$.

Note that $\Omega(\sqrt{T \log(n)})$ is the best one could hope for.

Potential based analysis

Define $\psi(t) = \sum_{i=1}^{n} w_{i,t}$. One has:

$$\psi(t+1) = \sum_{i=1}^{n} (1 - \eta \ell_t(i)) w_{i,t} = \psi(t)(1 - \eta \langle p_t, \ell_t \rangle),$$

Potential based analysis

Define $\psi(t) = \sum_{i=1}^{n} w_{i,t}$. One has:

$$\psi(t+1) = \sum_{i=1}^{n} (1 - \eta \ell_t(i)) w_{i,t} = \psi(t)(1 - \eta \langle p_t, \ell_t \rangle) \,,$$

so that (since $\psi(1) = n$):

$$\psi(T+1) = n \prod_{t=1}^{T} (1 - \eta \langle p_t, \ell_t \rangle) \leq n \exp(-\eta L_T) \,.$$

Potential based analysis

Define $\psi(t) = \sum_{i=1}^{n} w_{i,t}$. One has:

$$\psi(t+1) = \sum_{i=1}^{n} (1 - \eta \ell_t(i)) w_{i,t} = \psi(t)(1 - \eta \langle p_t, \ell_t \rangle) \,,$$

so that (since $\psi(1) = n$):

$$\psi(T+1) = n \prod_{t=1}^{T} (1 - \eta \langle p_t, \ell_t \rangle) \leq n \exp(-\eta L_T) \,.$$

On the other hand $\psi(T+1) \geq w_{i,T+1} = (1-\eta)^{L_{i,T}}$

Potential based analysis

Define $\psi(t) = \sum_{i=1}^{n} w_{i,t}$. One has:

$$\psi(t+1) = \sum_{i=1}^{n} (1 - \eta \ell_t(i)) w_{i,t} = \psi(t)(1 - \eta \langle p_t, \ell_t \rangle) \, ,$$

so that (since $\psi(1) = n$):

$$\psi(T+1) = n \prod_{t=1}^{T} (1 - \eta \langle p_t, \ell_t \rangle) \leq n \exp(-\eta L_T) \, .$$

On the other hand $\psi(T+1) \geq w_{i,T+1} = (1-\eta)^{L_{i,T}}$, and thus:

$$\eta L_T - \log \left(\frac{1}{1-\eta} \right) L_{i,T} \leq \log(n) \, ,$$

and the proof is concluded by $\log \left(\frac{1}{1-\eta} \right) \leq \eta + \eta^2$ for $\eta \in [0, 1/2]$.

Potential based analysis

Define $\psi(t) = \sum_{i=1}^{n} w_{i,t}$. One has:

$$\psi(t+1) = \sum_{i=1}^{n} (1 - \eta \ell_t(i)) w_{i,t} = \psi(t)(1 - \eta \langle p_t, \ell_t \rangle),$$

so that (since $\psi(1) = n$):

$$\psi(T+1) = n \prod_{t=1}^{T} (1 - \eta \langle p_t, \ell_t \rangle) \leq n \exp(-\eta L_T).$$

On the other hand $\psi(T+1) \geq w_{i,T+1} = (1-\eta)^{L_{i,T}}$, and thus:

$$\eta L_T - \log\left(\frac{1}{1-\eta}\right) L_{i,T} \leq \log(n),$$

and the proof is concluded by $\log\left(\frac{1}{1-\eta}\right) \leq \eta + \eta^2$ for $\eta \in [0, 1/2]$.

The mirror descent framework (Lec. 2) will give a principled approach to derive both the MW algorithm and its analysis.

A principled game-theoretic approach to regret analysis

[Abernethy, Warmuth, Yellin 2008; Rakhlin, Sridharan, Tewari 2010; B., Dekel, Koren, Peres 2015]

Let us focus on an oblivious adversary, that is he chooses $\ell_1, \dots, \ell_T \in \mathcal{L}$ at the beginning of the game.

A principled game-theoretic approach to regret analysis

[Abernethy, Warmuth, Yellin 2008; Rakhlin, Sridharan, Tewari 2010; B., Dekel, Koren, Peres 2015]

Let us focus on an oblivious adversary, that is he chooses $\ell_1, \dots, \ell_T \in \mathcal{L}$ at the beginning of the game.

A deterministic player's strategy is specified by a sequence of operators $a_1, \dots, a_T$, where in the full information case $a_s : ([0,1]^n)^{s-1} \to \mathcal{K}$, and in the bandit case $a_s : \mathbb{R}^{s-1} \to \mathcal{K}$. Denote $\mathcal{A}$ the set of such sequences of operators.

A principled game-theoretic approach to regret analysis

[Abernethy, Warmuth, Yellin 2008; Rakhlin, Sridharan, Tewari 2010; B., Dekel, Koren, Peres 2015]

Let us focus on an oblivious adversary, that is he chooses $\ell_1, \ldots, \ell_T \in \mathcal{L}$ at the beginning of the game.

A deterministic player's strategy is specified by a sequence of operators $a_1, \ldots, a_T$, where in the full information case $a_s : ([0,1]^n)^{s-1} \to \mathcal{K}$, and in the bandit case $a_s : \mathbb{R}^{s-1} \to \mathcal{K}$. Denote $\mathcal{A}$ the set of such sequences of operators.

Write $R_T(\mathbf{a}, \ell)$ for the regret of playing strategy $\mathbf{a} \in \mathcal{A}$ against loss sequence $\ell \in \mathcal{L}^T$.

A principled game-theoretic approach to regret analysis

[Abernethy, Warmuth, Yellin 2008; Rakhlin, Sridharan, Tewari 2010; B., Dekel, Koren, Peres 2015]

Let us focus on an oblivious adversary, that is he chooses $\ell_1, \dots, \ell_T \in \mathcal{L}$ at the beginning of the game.

A deterministic player's strategy is specified by a sequence of operators $a_1, \dots, a_T$, where in the full information case $a_s : ([0,1]^n)^{s-1} \to \mathcal{K}$, and in the bandit case $a_s : \mathbb{R}^{s-1} \to \mathcal{K}$. Denote $\mathcal{A}$ the set of such sequences of operators.

Write $R_T(\mathbf{a}, \ell)$ for the regret of playing strategy $\mathbf{a} \in \mathcal{A}$ against loss sequence $\ell \in \mathcal{L}^T$. Now we are interested in:

$$\inf_{\mu \in \Delta(\mathcal{A})} \sup_{\ell \in \mathcal{L}^T} \mathbb{E}_{\mathbf{a} \sim \mu} R_T(\mathbf{a}, \ell) = \sup_{\nu \in \Delta(\mathcal{L}^T)} \inf_{\mu \in \Delta(\mathcal{A})} \mathbb{E}_{\ell \sim \nu, \mathbf{a} \sim \mu} R_T(\mathbf{a}, \ell) \,,$$

where the swap of min and max comes from Sion's minimax theorem.

A principled game-theoretic approach to regret analysis

[Abernethy, Warmuth, Yellin 2008; Rakhlin, Sridharan, Tewari 2010; B., Dekel, Koren, Peres 2015]

Let us focus on an oblivious adversary, that is he chooses $\ell_1, \dots, \ell_T \in \mathcal{L}$ at the beginning of the game.

A deterministic player's strategy is specified by a sequence of operators $a_1, \dots, a_T$, where in the full information case $a_s : ([0,1]^n)^{s-1} \to \mathcal{K}$, and in the bandit case $a_s : \mathbb{R}^{s-1} \to \mathcal{K}$. Denote $\mathcal{A}$ the set of such sequences of operators.

Write $R_T(\mathbf{a}, \ell)$ for the regret of playing strategy $\mathbf{a} \in \mathcal{A}$ against loss sequence $\ell \in \mathcal{L}^T$. Now we are interested in:

$$\inf_{\mu \in \Delta(\mathcal{A})} \sup_{\ell \in \mathcal{L}^T} \mathbb{E}_{\mathbf{a} \sim \mu} R_T(\mathbf{a}, \ell) = \sup_{\nu \in \Delta(\mathcal{L}^T)} \inf_{\mu \in \Delta(\mathcal{A})} \mathbb{E}_{\ell \sim \nu, \mathbf{a} \sim \mu} R_T(\mathbf{a}, \ell),$$

where the swap of min and max comes from Sion's minimax theorem.

In other words we can study the minimax regret by designing a strategy for a *Bayesian* scenario where $\ell \sim \nu$ and ν is known.

A Doob strategy [B., Dekel, Koren, Peres 2015]

Since we known ν, we also know the *distribution* of i^*. In fact as we make observations, we can update our knowledge of i^* with the *posterior distribution*. Denote $\mathbb{E}_t$ for this posterior distribution (e.g., in full information $\mathbb{E}_t := \mathbb{E}[\cdot|\ell_1, \dots, \ell_{t-1}]$).

A Doob strategy [B., Dekel, Koren, Peres 2015]

Since we known ν, we also know the *distribution* of i^*. In fact as we make observations, we can update our knowledge of i^* with the *posterior distribution*. Denote $\mathbb{E}_t$ for this posterior distribution (e.g., in full information $\mathbb{E}_t := \mathbb{E}[\cdot|\ell_1, \dots, \ell_{t-1}]$).
By convexity of $\Delta([n]) =: \Delta_n$ it is natural to consider playing from:

$$p_t := \mathbb{E}_t \; \delta_{i^*} \, .$$

In other words we are playing from the posterior distribution of the optimum, a kind of "probability matching". This is also called Thompson Sampling.

A Doob strategy [B., Dekel, Koren, Peres 2015]

Since we known ν, we also know the *distribution* of i^*. In fact as we make observations, we can update our knowledge of i^* with the *posterior distribution*. Denote $\mathbb{E}_t$ for this posterior distribution (e.g., in full information $\mathbb{E}_t := \mathbb{E}[\cdot|\ell_1, \dots, \ell_{t-1}]$).
By convexity of $\Delta([n]) =: \Delta_n$ it is natural to consider playing from:

$$p_t := \mathbb{E}_t \ \delta_{i^*} \, .$$

In other words we are playing from the posterior distribution of the optimum, a kind of "probability matching". This is also called Thompson Sampling.
The regret of this strategy can be controlled via the *movement* of this Doob martingale (recall $\|\ell_t\|_\infty \leq 1$)

$$\mathbb{E} \sum_{t=1}^{T} \langle p_t - \delta_{i^*}, \ell_t \rangle = \mathbb{E} \sum_{t=1}^{T} \langle p_t - p_{t+1}, \ell_t \rangle \leq \mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1 \, .$$

How stable is a martingale?

Question: is a martingale in Δ_n "stable"? Following famous inequality is a possible answer (proof on the next slide):

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n) \, .$$

How stable is a martingale?

Question: is a martingale in Δ_n "stable"? Following famous inequality is a possible answer (proof on the next slide):

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n) \,.$$

This yields by Cauchy-Schwarz:

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1 \leq \sqrt{T \times \mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2} \leq \sqrt{2 T \log(n)} \,.$$

How stable is a martingale?

Question: is a martingale in Δ_n "stable"? Following famous inequality is a possible answer (proof on the next slide):

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n) \,.$$

This yields by Cauchy-Schwarz:

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1 \leq \sqrt{T \times \mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2} \leq \sqrt{2 T \log(n)} \,.$$

Thus we have recovered the regret bound of MW (in fact with an optimal constant) by a purely geometric argument!

Entropic proof of cotype for ℓ_1^n

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n) \,.$$

Entropic proof of cotype for ℓ_1^n

$$\mathbb{E}\sum_{t=1}^{T}\|p_t - p_{t+1}\|_1^2 \leq 2\log(n)\,.$$

By Pinsker's inequality:

$$\frac{1}{2}\|p_t - p_{t+1}\|_1^2 \leq \mathrm{Ent}(p_{t+1}; p_t) = \mathrm{Ent}_t(i^*|\ell_t; i^*)\,.$$

Entropic proof of cotype for ℓ_1^n

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2\log(n)\,.$$

By Pinsker's inequality:

$$\frac{1}{2}\|p_t - p_{t+1}\|_1^2 \leq \mathrm{Ent}(p_{t+1}; p_t) = \mathrm{Ent}_t(i^*|\ell_t; i^*)\,.$$

Now essentially by definition one has (recall that $I(X,Y) = H(X) - H(X|Y) = \mathbb{E}_Y \mathrm{Ent}(p_{X|Y}; p_X)$):

$$\mathbb{E}_{\ell_t} \mathrm{Ent}_t(i^*|\ell_t; i^*) = H_t(i^*) - H_{t+1}(i^*)\,.$$

Entropic proof of cotype for ℓ_1^n

$$\mathbb{E} \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1^2 \leq 2 \log(n) \,.$$

By Pinsker's inequality:

$$\frac{1}{2}\|p_t - p_{t+1}\|_1^2 \leq \mathrm{Ent}(p_{t+1}; p_t) = \mathrm{Ent}_t(i^*|\ell_t; i^*) \,.$$

Now essentially by definition one has (recall that $I(X, Y) = H(X) - H(X|Y) = \mathbb{E}_Y \mathrm{Ent}(p_{X|Y}; p_X)$):

$$\mathbb{E}_{\ell_t} \mathrm{Ent}_t(i^*|\ell_t; i^*) = H_t(i^*) - H_{t+1}(i^*) \,.$$

Proof concluded by telescopic sum and maximal entropy being $\log(n)$.

A more general story: M-cotype

Let us generalize the setting. In *online linear optimization*, the player plays $x_t \in K \subset \mathbb{R}^n$, and the adversary plays $\ell_t \in \mathcal{L} \subset \mathbb{R}^n$. We assume that there is a norm $\|\cdot\|$ such that $\|x_t\| \leq 1$ and $\|\ell_t\|^* \leq 1$.

A more general story: M-cotype

Let us generalize the setting. In *online linear optimization*, the player plays $x_t \in K \subset \mathbb{R}^n$, and the adversary plays $\ell_t \in \mathcal{L} \subset \mathbb{R}^n$. We assume that there is a norm $\|\cdot\|$ such that $\|x_t\| \leq 1$ and $\|\ell_t\|^* \leq 1$. The same game-theoretic argument goes through, and denoting $x^* = \mathrm{argmin}_{x \in K} \sum_{t=1}^T \langle \ell_t, x \rangle$, $x_t := \mathbb{E}_t x^*$, one has

$$\mathbb{E} \sum_{t=1}^{T} \langle \ell_t, x_t - x^* \rangle = \mathbb{E} \sum_{t=1}^{T} \langle \ell_t, x_t - x_{t+1} \rangle \leq \mathbb{E} \sum_{t=1}^{T} \|x_t - x_{t+1}\| .$$

A more general story: M-cotype

Let us generalize the setting. In *online linear optimization*, the player plays $x_t \in K \subset \mathbb{R}^n$, and the adversary plays $\ell_t \in \mathcal{L} \subset \mathbb{R}^n$. We assume that there is a norm $\|\cdot\|$ such that $\|x_t\| \leq 1$ and $\|\ell_t\|^* \leq 1$. The same game-theoretic argument goes through, and denoting $x^* = \mathrm{argmin}_{x \in K} \sum_{t=1}^T \langle \ell_t, x \rangle$, $x_t := \mathbb{E}_t x^*$, one has

$$\mathbb{E} \sum_{t=1}^T \langle \ell_t, x_t - x^* \rangle = \mathbb{E} \sum_{t=1}^T \langle \ell_t, x_t - x_{t+1} \rangle \leq \mathbb{E} \sum_{t=1}^T \|x_t - x_{t+1}\| .$$

The norm $\|\cdot\|$ has *M*-cotype (C, q) if for any martingale (x_t) one has:

$$\left(\mathbb{E} \sum_{t=1}^T \|x_t - x_{t+1}\|^q \right)^{1/q} \leq C \; \mathbb{E}\|x_{T+1}\| .$$

A more general story: M-cotype

Let us generalize the setting. In *online linear optimization*, the player plays $x_t \in K \subset \mathbb{R}^n$, and the adversary plays $\ell_t \in \mathcal{L} \subset \mathbb{R}^n$. We assume that there is a norm $\|\cdot\|$ such that $\|x_t\| \leq 1$ and $\|\ell_t\|^* \leq 1$. The same game-theoretic argument goes through, and denoting $x^* = \mathrm{argmin}_{x \in K} \sum_{t=1}^T \langle \ell_t, x \rangle$, $x_t := \mathbb{E}_t x^*$, one has

$$\mathbb{E} \sum_{t=1}^T \langle \ell_t, x_t - x^* \rangle = \mathbb{E} \sum_{t=1}^T \langle \ell_t, x_t - x_{t+1} \rangle \leq \mathbb{E} \sum_{t=1}^T \|x_t - x_{t+1}\| .$$

The norm $\|\cdot\|$ has M-cotype (C, q) if for any martingale (x_t) one has:

$$\left(\mathbb{E} \sum_{t=1}^T \|x_t - x_{t+1}\|^q \right)^{1/q} \leq C \ \mathbb{E} \|x_{T+1}\| .$$

In particular this gives a regret in $C \ T^{1-1/q}$.

A lower bound via M-type of the dual

Interestingly the analysis via cotype is tight in the following sense.

A lower bound via M-type of the dual

Interestingly the analysis via cotype is tight in the following sense. First if M-cotype (C, q) holds for $\|\cdot\|$, then so does M-type (C', p) for $\|\cdot\|_*$ (where p is the conjugate of q), i.e., for any martingale difference sequence (ℓ_t) one has

$$\mathbb{E}\left\|\sum_{t=1}^{T} \ell_t\right\|_* \leq C' \left(\mathbb{E}\sum_{t=1}^{T} \|\ell_t\|_*^p\right)^{1/p} .$$

A lower bound via M-type of the dual

Interestingly the analysis via cotype is tight in the following sense. First if M-cotype (C, q) holds for $\|\cdot\|$, then so does M-type (C', p) for $\|\cdot\|_*$ (where p is the conjugate of q), i.e., for any martingale difference sequence (ℓ_t) one has

$$\mathbb{E}\left\|\sum_{t=1}^{T} \ell_t\right\|_* \leq C' \left(\mathbb{E} \sum_{t=1}^{T} \|\ell_t\|_*^p\right)^{1/p} .$$

Moreover one can show that the violation of type/cotype can be witnessed by a martingale with unit norm increments. Thus if M-cotype (C, q) fails for $\|\cdot\|$, there must exist a martingale difference sequence (ℓ_t) with $\|\ell_t\|_* = 1$ that violates the above inequality.

A lower bound via M-type of the dual

Interestingly the analysis via cotype is tight in the following sense. First if M-cotype (C, q) holds for $\|\cdot\|$, then so does M-type (C', p) for $\|\cdot\|_*$ (where p is the conjugate of q), i.e., for any martingale difference sequence (ℓ_t) one has

$$\mathbb{E}\left\|\sum_{t=1}^{T}\ell_t\right\|_* \leq C'\left(\mathbb{E}\sum_{t=1}^{T}\|\ell_t\|_*^p\right)^{1/p} .$$

Moreover one can show that the violation of type/cotype can be witnessed by a martingale with unit norm increments. Thus if M-cotype (C, q) fails for $\|\cdot\|$, there must exist a martingale difference sequence (ℓ_t) with $\|\ell_t\|_* = 1$ that violates the above inequality. In particular:

$$\mathbb{E}\sum_{t=1}^{T}\langle \ell_t, x_t - x^*\rangle = \mathbb{E}\left\|\sum_{t=1}^{T}\ell_t\right\|_* \geq C' T^{1/p} = C' T^{1-1/q} .$$

A lower bound via M-type of the dual

Interestingly the analysis via cotype is tight in the following sense. First if M-cotype (C, q) holds for $\|\cdot\|$, then so does M-type (C', p) for $\|\cdot\|_*$ (where p is the conjugate of q), i.e., for any martingale difference sequence (ℓ_t) one has

$$\mathbb{E}\left\|\sum_{t=1}^{T}\ell_t\right\|_* \leq C'\left(\mathbb{E}\sum_{t=1}^{T}\|\ell_t\|_*^p\right)^{1/p} .$$

Moreover one can show that the violation of type/cotype can be witnessed by a martingale with unit norm increments. Thus if M-cotype (C, q) fails for $\|\cdot\|$, there must exist a martingale difference sequence (ℓ_t) with $\|\ell_t\|_* = 1$ that violates the above inequality. In particular:

$$\mathbb{E}\sum_{t=1}^{T}\langle\ell_t, x_t - x^*\rangle = \mathbb{E}\left\|\sum_{t=1}^{T}\ell_t\right\|_* \geq C' T^{1/p} = C' T^{1-1/q} .$$

Important: these are "dimension-free arguments", if one brings the dimension in the bounds then the story changes.

What about the bandit game? [Russo, Van Roy 2014]

So far we only talked about the *hedging* aspect of the problem. In particular for the full information game the "learning" part happens automatically. This is captured by the fact that the **variation in the posterior is lower bounded by the instantaneous regret**:

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \mathbb{E}_t \langle p_t - p_{t+1}, \ell_t \rangle \leq \mathbb{E}_t \| p_t - p_{t+1} \|_1 \, .$$

What about the bandit game? [Russo, Van Roy 2014]

So far we only talked about the *hedging* aspect of the problem. In particular for the full information game the "learning" part happens automatically. This is captured by the fact that the **variation in the posterior is lower bounded by the instantaneous regret**:

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \mathbb{E}_t \langle p_t - p_{t+1}, \ell_t \rangle \leq \mathbb{E}_t \| p_t - p_{t+1} \|_1 \,.$$

In the bandit game the first equality is not true anymore and thus the inequality does not hold a priori. In fact this is the whole difficulty: learning is now costly because of the tradeoff between exploration and exploitation.

What about the bandit game? [Russo, Van Roy 2014]

So far we only talked about the *hedging* aspect of the problem. In particular for the full information game the "learning" part happens automatically. This is captured by the fact that the **variation in the posterior is lower bounded by the instantaneous regret**:

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \mathbb{E}_t \langle p_t - p_{t+1}, \ell_t \rangle \leq \mathbb{E}_t \|p_t - p_{t+1}\|_1 .$$

In the bandit game the first equality is not true anymore and thus the inequality does not hold a priori. In fact this is the whole difficulty: learning is now costly because of the tradeoff between exploration and exploitation.

Importantly note that the cotype inequality for ℓ_1 is proved by relating the ℓ_1 variation squared to the mutual information between OPT and the feedback. Thus a weaker inequality that would suffice is:

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle \leq C \sqrt{I_t(i^*, (i_t, \ell_t(i_t)))},$$

which would lead to a regret in $C\sqrt{T \log(n)}$.

The Russo-Van Roy analysis

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

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \sum_i p_t(i)(\bar{\ell}_t(i) - \bar{\ell}_t(i,i)) \,,$$

and

$$I_t((i_t, \ell_t(i_t)), i^*) = \sum_{i,j} p_t(i) p_t(j) \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) \| \mathcal{L}_t(\ell_t(i)))$$

The Russo-Van Roy analysis

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

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \sum_i p_t(i)(\bar{\ell}_t(i) - \bar{\ell}_t(i,i)),$$

and

$$I_t((i_t, \ell_t(i_t)), i^*) = \sum_{i,j} p_t(i) p_t(j) \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) \| \mathcal{L}_t(\ell_t(i)))$$

Now using Cauchy-Schwarz the instantaneous regret is bounded by

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

The Russo-Van Roy analysis

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

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \sum_i p_t(i) (\overline{\ell}_t(i) - \overline{\ell}_t(i,i)) \,,$$

and

$$I_t((i_t, \ell_t(i_t)), i^*) = \sum_{i,j} p_t(i) p_t(j) \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) \| \mathcal{L}_t(\ell_t(i)))$$

Now using Cauchy-Schwarz the instantaneous regret is bounded by

$$\sqrt{n \sum_i p_t(i)^2 (\overline{\ell}_t(i) - \overline{\ell}_t(i,i))^2} \leq \sqrt{n \sum_{i,j} p_t(i) p_t(j) (\overline{\ell}_t(i) - \overline{\ell}_t(i,j))^2} \,.$$

Pinsker's inequality gives (using $\|\ell_t\|_\infty \leq 1$):

$$(\overline{\ell}_t(i) - \overline{\ell}_t(i,j))^2 \leq \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) \| \mathcal{L}_t(\ell_t(i))) \,,$$

The Russo-Van Roy analysis

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

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle = \sum_i p_t(i)(\bar{\ell}_t(i) - \bar{\ell}_t(i,i)) \,,$$

and

$$I_t((i_t, \ell_t(i_t)), i^*) = \sum_{i,j} p_t(i) p_t(j) \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) \| \mathcal{L}_t(\ell_t(i)))$$

Now using Cauchy-Schwarz the instantaneous regret is bounded by

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

Pinsker's inequality gives (using $\|\ell_t\|_\infty \leq 1$):

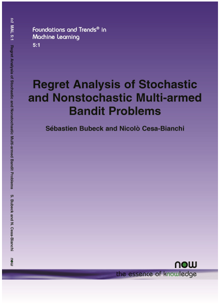
$$(\bar{\ell}_t(i) - \bar{\ell}_t(i,j))^2 \leq \mathrm{Ent}(\mathcal{L}_t(\ell_t(i) | i^* = j) \| \mathcal{L}_t(\ell_t(i))) \,,$$

Thus one obtains

$$\mathbb{E}_t \langle p_t - \delta_{i^*}, \ell_t \rangle \leq \sqrt{n \; I_t((i_t, \ell_t(i_t)), i^*)} \,.$$

Lecture 2:
Mirror descent and online decision making

Sébastien Bubeck
Machine Learning and Optimization group, MSR AI



Stability as an algorithmic guiding principle

Recall that we are looking for a rule to select $p_t \in \Delta_n$ based on $\ell_1, \ldots, \ell_{t-1} \in [-1, 1]^n$, such that we can control the regret with respect to any comparator $q \in \Delta_n$:

$$\sum_{t=1}^{T} \langle \ell_t, p_t - q \rangle .$$

Stability as an algorithmic guiding principle

Recall that we are looking for a rule to select $p_t \in \Delta_n$ based on $\ell_1, \ldots, \ell_{t-1} \in [-1, 1]^n$, such that we can control the regret with respect to any comparator $q \in \Delta_n$:

$$\sum_{t=1}^{T} \langle \ell_t, p_t - q \rangle \, .$$

In the game-theoretic approach we saw that the *movement* of the algorithm, $\sum_{t=1}^{T} \|p_t - p_{t+1}\|_1$, was the key quantity to control.

Stability as an algorithmic guiding principle

Recall that we are looking for a rule to select $p_t \in \Delta_n$ based on $\ell_1, \ldots, \ell_{t-1} \in [-1, 1]^n$, such that we can control the regret with respect to any comparator $q \in \Delta_n$:

$$\sum_{t=1}^{T} \langle \ell_t, p_t - q \rangle .$$

In the game-theoretic approach we saw that the *movement* of the algorithm, $\sum_{t=1}^{T} \|p_t - p_{t+1}\|_1$, was the key quantity to control. In fact the same is true in general up to an additional "1-lookahead" term:

$$\sum_{t=1}^{T} \langle \ell_t, p_t - q \rangle \leq \sum_{t=1}^{T} \langle \ell_t, p_{t+1} - q \rangle + \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1 .$$

Stability as an algorithmic guiding principle

Recall that we are looking for a rule to select $p_t \in \Delta_n$ based on $\ell_1, \dots, \ell_{t-1} \in [-1,1]^n$, such that we can control the regret with respect to any comparator $q \in \Delta_n$:

$$\sum_{t=1}^{T} \langle \ell_t, p_t - q \rangle .$$

In the game-theoretic approach we saw that the *movement* of the algorithm, $\sum_{t=1}^{T} \|p_t - p_{t+1}\|_1$, was the key quantity to control. In fact the same is true in general up to an additional "1-lookahead" term:

$$\sum_{t=1}^{T} \langle \ell_t, p_t - q \rangle \leq \sum_{t=1}^{T} \langle \ell_t, p_{t+1} - q \rangle + \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1 .$$

In other words p_{t+1} (which can depend on ℓ_t) is trading off being "good" for ℓ_t, while at the same time remaining close to p_t.

Metrical task systems [Borodin, Linial, Saks 1982]

This view of the problem is closely related to the following setting in *online algorithms*:

Metrical task systems [Borodin, Linial, Saks 1982]

This view of the problem is closely related to the following setting in *online algorithms*:

- At each time step t the algorithm maintains a *state* $i_t \in [n]$.

Metrical task systems [Borodin, Linial, Saks 1982]

This view of the problem is closely related to the following setting in *online algorithms*:

- At each time step t the algorithm maintains a *state* $i_t \in [n]$.
- Upon the observation of a loss function $\ell_t : [n] \to \mathbb{R}_+$ the algorithm can update the state to i_{t+1}.

Metrical task systems [Borodin, Linial, Saks 1982]

This view of the problem is closely related to the following setting in *online algorithms*:

- At each time step t the algorithm maintains a *state* $i_t \in [n]$.
- Upon the observation of a loss function $\ell_t : [n] \to \mathbb{R}_+$ the algorithm can update the state to i_{t+1}.
- The associated cost is composed of a service cost $\ell_t(i_{t+1})$ and a movement cost $d(i_t, i_{t+1})$ (d is some underlying metric on $[n]$).

Metrical task systems [Borodin, Linial, Saks 1982]

This view of the problem is closely related to the following setting in *online algorithms*:

- At each time step t the algorithm maintains a *state* $i_t \in [n]$.
- Upon the observation of a loss function $\ell_t : [n] \to \mathbb{R}_+$ the algorithm can update the state to i_{t+1}.
- The associated cost is composed of a service cost $\ell_t(i_{t+1})$ and a movement cost $d(i_t, i_{t+1})$ (d is some underlying metric on $[n]$).
- Typically interested in competitive ratio rather than regret.

Metrical task systems [Borodin, Linial, Saks 1982]

This view of the problem is closely related to the following setting in *online algorithms*:

- At each time step t the algorithm maintains a *state* $i_t \in [n]$.
- Upon the observation of a loss function $\ell_t : [n] \to \mathbb{R}_+$ the algorithm can update the state to i_{t+1}.
- The associated cost is composed of a service cost $\ell_t(i_{t+1})$ and a movement cost $d(i_t, i_{t+1})$ (d is some underlying metric on $[n]$).
- Typically interested in competitive ratio rather than regret.

Connection: If i_t is played at random from p_t, and consequent samplings are appropriately coupled, then the term we want to bound

$$\sum_{t=1}^{T} \langle \ell_t, p_{t+1} - q \rangle + \sum_{t=1}^{T} \|p_t - p_{t+1}\|_1 \,,$$

exactly corresponds to the sum of expected service cost and expected movement when the metric is trivial (i.e., $d \equiv 1$).

Gradient descent/regularization approach

A natural algorithm to consider is gradient descent:

$$x_{t+1} = x_t - \eta \ell_t \, ,$$

Gradient descent/regularization approach

A natural algorithm to consider is gradient descent:

$$x_{t+1} = x_t - \eta \ell_t \, ,$$

which can equivalently be viewed as

$$x_{t+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \langle x, \ell_t \rangle + \frac{1}{2\eta} \|x - x_t\|_2^2 \, .$$

Gradient descent/regularization approach

A natural algorithm to consider is gradient descent:

$$x_{t+1} = x_t - \eta \ell_t \,,$$

which can equivalently be viewed as

$$x_{t+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \langle x, \ell_t \rangle + \frac{1}{2\eta} \|x - x_t\|_2^2 \,.$$

This clearly does not seem adapted to our situation where we want to measure movement with respect to the ℓ_1-norm.

Gradient descent/regularization approach

A natural algorithm to consider is gradient descent:

$$x_{t+1} = x_t - \eta \ell_t ,$$

which can equivalently be viewed as

$$x_{t+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \langle x, \ell_t \rangle + \frac{1}{2\eta} \|x - x_t\|_2^2 .$$

This clearly does not seem adapted to our situation where we want to measure movement with respect to the ℓ_1-norm.

Side comment: another equivalent definition is as follows, say with $x_1 = 0$,

$$x_{t+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \langle x, \sum_{s \leq t} \ell_s \rangle + \frac{1}{2\eta} \|x\|_2^2 .$$

This view is called "Follow The Regularized Leader" (FTRL)

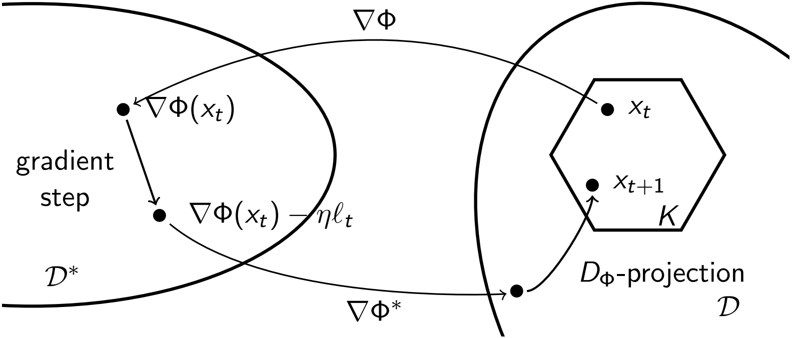
Mirror Descent (Nemirovski and Yudin 87)

Mirror Descent (Nemirovski and Yudin 87)

Mirror map/regularizer: convex function $\Phi : \mathcal{D} \supset K \to \mathbb{R}$.
Bregman divergence: $D_\Phi(x; y) = \Phi(x) - \Phi(y) - \nabla\Phi(y) \cdot (x - y)$.
Note that $\nabla_x D_\Phi(x; y) = \nabla\Phi(x) - \nabla\Phi(y)$.

Mirror Descent (Nemirovski and Yudin 87)

Mirror map/regularizer: convex function $\Phi : \mathcal{D} \supset K \to \mathbb{R}$.
Bregman divergence: $D_\Phi(x; y) = \Phi(x) - \Phi(y) - \nabla\Phi(y) \cdot (x - y)$.
Note that $\nabla_x D_\Phi(x; y) = \nabla\Phi(x) - \nabla\Phi(y)$.



Continuous-time mirror descent

Assume now a continuous time setting where the losses are revealed incrementally and the algorithm can respond instantaneously: the service cost is now $\int_{t \in \mathbb{R}_+} \ell(t) \cdot x(t) dt$ and the movement cost is $\int_{t \in \mathbb{R}_+} \|x'(t)\|_1 dt$.

Continuous-time mirror descent

Assume now a continuous time setting where the losses are revealed incrementally and the algorithm can respond instantaneously: the service cost is now $\int_{t\in\mathbb{R}_+} \ell(t)\cdot x(t)dt$ and the movement cost is $\int_{t\in\mathbb{R}_+} \|x'(t)\|_1 dt$.

Denote $N_K(x) = \{\theta : \theta\cdot(y-x)\rangle \leq 0,\ \forall y \in K\}$ and recall that

$$x^* \in \operatorname*{argmin}_{x\in K} f(x) \Leftrightarrow -\nabla f(x^*) \in N_K(x^*)$$

Continuous-time mirror descent

Assume now a continuous time setting where the losses are revealed incrementally and the algorithm can respond instantaneously: the service cost is now $\int_{t \in \mathbb{R}_+} \ell(t) \cdot x(t) dt$ and the movement cost is $\int_{t \in \mathbb{R}_+} \|x'(t)\|_1 dt$.

Denote $N_K(x) = \{\theta : \theta \cdot (y - x)\rangle \leq 0,\ \forall y \in K\}$ and recall that

$$x^* \in \operatorname*{argmin}_{x \in K} f(x) \Leftrightarrow -\nabla f(x^*) \in N_K(x^*)$$

$$x(t + \varepsilon) = \operatorname*{argmin}_{x \in K} D_\Phi(x, \nabla \Phi^*(\nabla \Phi(x(t)) - \varepsilon \eta \ell(t)))$$

Continuous-time mirror descent

Assume now a continuous time setting where the losses are revealed incrementally and the algorithm can respond instantaneously: the service cost is now $\int_{t\in\mathbb{R}_+} \ell(t)\cdot x(t)dt$ and the movement cost is $\int_{t\in\mathbb{R}_+} \|x'(t)\|_1 dt$.

Denote $N_K(x) = \{\theta : \theta\cdot(y-x)\rangle \leq 0,\ \forall y\in K\}$ and recall that

$$x^* \in \operatorname*{argmin}_{x\in K} f(x) \Leftrightarrow -\nabla f(x^*) \in N_K(x^*)$$

$$x(t+\varepsilon) = \operatorname*{argmin}_{x\in K} D_\Phi(x, \nabla\Phi^*(\nabla\Phi(x(t)) - \varepsilon\eta\ell(t)))$$
$$\Leftrightarrow \nabla\Phi(x(t+\varepsilon)) - \nabla\Phi(x(t)) + \varepsilon\eta\ell(t) \in -N_K(x(t+\varepsilon))$$

Continuous-time mirror descent

Assume now a continuous time setting where the losses are revealed incrementally and the algorithm can respond instantaneously: the service cost is now $\int_{t \in \mathbb{R}_+} \ell(t) \cdot x(t) dt$ and the movement cost is $\int_{t \in \mathbb{R}_+} \|x'(t)\|_1 dt$.

Denote $N_K(x) = \{\theta : \theta \cdot (y - x)\rangle \leq 0, \ \forall y \in K\}$ and recall that

$$x^* \in \operatorname*{argmin}_{x \in K} f(x) \Leftrightarrow -\nabla f(x^*) \in N_K(x^*)$$

$$\begin{aligned}
x(t+\varepsilon) &= \operatorname*{argmin}_{x \in K} D_\Phi(x, \nabla\Phi^*(\nabla\Phi(x(t)) - \varepsilon\eta\ell(t))) \\
&\Leftrightarrow \nabla\Phi(x(t+\varepsilon)) - \nabla\Phi(x(t)) + \varepsilon\eta\ell(t) \in -N_K(x(t+\varepsilon)) \\
&\Leftrightarrow \nabla^2\Phi(x(t))x'(t) \in -\eta\ell(t) - N_K(x(t))
\end{aligned}$$

Theorem (BCLLM17)

The above differential inclusion admits a (unique) solution $x : \mathbb{R}_+ \to \mathcal{X}$ provided that K is a compact convex set, Φ is strongly convex, and $\nabla^2\Phi$ and ℓ are Lipschitz.

The basic calculation

$$\nabla^2 \Phi(x(t)) x'(t) = -\eta \ell(t) - \lambda(t), \; \lambda(t) \in N_K(x(t))$$

The basic calculation

$$\nabla^2\Phi(x(t))x'(t) = -\eta\ell(t) - \lambda(t),\ \lambda(t) \in N_K(x(t))$$

Recall $D_\Phi(y;x) = \Phi(y) - \Phi(x) - \nabla\Phi(x) \cdot (y - x)$,

The basic calculation

$$\nabla^2\Phi(x(t))x'(t) = -\eta\ell(t) - \lambda(t),\ \lambda(t) \in N_K(x(t))$$

Recall $D_\Phi(y;x) = \Phi(y) - \Phi(x) - \nabla\Phi(x)\cdot(y-x)$,

$$\begin{aligned}
\partial_t D_\Phi(y;x(t)) &= -\nabla^2\Phi(x(t))x'(t)\cdot(y-x(t)) \\
&= (\eta\ell(t)+\lambda(t))\cdot(y-x(t)) \\
&\leq \eta\ell(t)\cdot(y-x(t))
\end{aligned}$$

The basic calculation

$$\nabla^2\Phi(x(t))x'(t) = -\eta\ell(t) - \lambda(t),\ \lambda(t) \in N_K(x(t))$$

Recall $D_\Phi(y;x) = \Phi(y) - \Phi(x) - \nabla\Phi(x)\cdot(y-x)$,

$$\begin{aligned}\partial_t D_\Phi(y;x(t)) &= -\nabla^2\Phi(x(t))x'(t)\cdot(y-x(t)) \\ &= (\eta\ell(t)+\lambda(t))\cdot(y-x(t)) \\ &\leq \eta\ell(t)\cdot(y-x(t))\end{aligned}$$

Lemma

The mirror descent path $(x(t))_{t\geq 0}$ satisfies for any comparator point y,

$$\int \ell(t)\cdot(x(t)-y)dt \leq \frac{D_\Phi(y;x(0))}{\eta}.$$

The basic calculation

$$\nabla^2 \Phi(x(t)) x'(t) = -\eta \ell(t) - \lambda(t), \ \lambda(t) \in N_K(x(t))$$

Recall $D_\Phi(y; x) = \Phi(y) - \Phi(x) - \nabla \Phi(x) \cdot (y - x)$,

$$\begin{aligned} \partial_t D_\Phi(y; x(t)) &= -\nabla^2 \Phi(x(t)) x'(t) \cdot (y - x(t)) \\ &= (\eta \ell(t) + \lambda(t)) \cdot (y - x(t)) \\ &\leq \eta \ell(t) \cdot (y - x(t)) \end{aligned}$$

Lemma

The mirror descent path $(x(t))_{t \geq 0}$ *satisfies for any comparator point* y*,*

$$\int \ell(t) \cdot (x(t) - y) dt \leq \frac{D_\Phi(y; x(0))}{\eta} .$$

Thus to control the regret it only remains to bound the movement cost $\int_{t \in \mathbb{R}_+} \|x'(t)\|_1 dt$ (recall that this continuous time setting is only valid for the 1-lookahead setting, i.e., MTS).

Controlling the movement and how the entropy arises

How to control $\|x'(t)\|_1 = \|(\nabla^2 \Phi(x(t)))^{-1}(\eta \ell(t) + \lambda(t))\|_1$? A particularly pleasant inequality would be to relate this to say $\eta \ell(t) \cdot x(t)$, in which case one would get a final regret bound of the form (up to a multiplicative factor $1/(1-\eta)$):

$$\frac{D_\Phi(y; x(0))}{\eta} + \eta L^* .$$

Controlling the movement and how the entropy arises

How to control $\|x'(t)\|_1 = \|(\nabla^2 \Phi(x(t)))^{-1}(\eta \ell(t) + \lambda(t))\|_1$? A particularly pleasant inequality would be to relate this to say $\eta \ell(t) \cdot x(t)$, in which case one would get a final regret bound of the form (up to a multiplicative factor $1/(1-\eta)$):

$$\frac{D_\Phi(y; x(0))}{\eta} + \eta L^* .$$

Ignore for a moment the Lagrange multiplier $\lambda(t)$ and assume that $\Phi(x) = \sum_{i=1}^n \varphi(x_i)$. We want to relate $\sum_{i=1}^n \ell_i(t)/\varphi''(x_i(t))$ to $\sum_{i=1}^n \ell_i(t) x_i(t)$.

Controlling the movement and how the entropy arises

How to control $\|x'(t)\|_1 = \|(\nabla^2\Phi(x(t)))^{-1}(\eta\ell(t) + \lambda(t))\|_1$? A particularly pleasant inequality would be to relate this to say $\eta\ell(t) \cdot x(t)$, in which case one would get a final regret bound of the form (up to a multiplicative factor $1/(1-\eta)$):

$$\frac{D_\Phi(y; x(0))}{\eta} + \eta L^* .$$

Ignore for a moment the Lagrange multiplier $\lambda(t)$ and assume that $\Phi(x) = \sum_{i=1}^n \varphi(x_i)$. We want to relate $\sum_{i=1}^n \ell_i(t)/\varphi''(x_i(t))$ to $\sum_{i=1}^n \ell_i(t) x_i(t)$. Making them equal gives $\Phi(x) = \sum_i x_i \log x_i$ with corresponding dynamics:

$$x_i'(t) = -\eta x_i(t)(\ell_i(t) + \mu(t)) .$$

In particular $\|x'(t)\|_1 \leq 2\eta\ell(t) \cdot x(t)$.

Controlling the movement and how the entropy arises

How to control $\|x'(t)\|_1 = \|(\nabla^2 \Phi(x(t)))^{-1}(\eta \ell(t) + \lambda(t))\|_1$? A particularly pleasant inequality would be to relate this to say $\eta \ell(t) \cdot x(t)$, in which case one would get a final regret bound of the form (up to a multiplicative factor $1/(1-\eta)$):

$$\frac{D_\Phi(y; x(0))}{\eta} + \eta L^* .$$

Ignore for a moment the Lagrange multiplier $\lambda(t)$ and assume that $\Phi(x) = \sum_{i=1}^n \varphi(x_i)$. We want to relate $\sum_{i=1}^n \ell_i(t)/\varphi''(x_i(t))$ to $\sum_{i=1}^n \ell_i(t) x_i(t)$. Making them equal gives $\Phi(x) = \sum_i x_i \log x_i$ with corresponding dynamics:

$$x_i'(t) = -\eta x_i(t)(\ell_i(t) + \mu(t)) .$$

In particular $\|x'(t)\|_1 \leq 2\eta \ell(t) \cdot x(t)$.
We note that this algorithm is exactly a continuous time version of the MW studied at the beginning of the first lecture.

The more classical discrete-time algorithm and analysis

Ignoring the Lagrangian and assuming $\ell'(t) = 0$ one has

$$\partial_t^2 D_\Phi(y; x(t)) = \nabla^2 \Phi(x(t))[x'(t), x'(t)] = \eta^2 (\nabla^2 \Phi(x(t)))^{-1} [\ell(t), \ell(t)] \,.$$

The more classical discrete-time algorithm and analysis

Ignoring the Lagrangian and assuming $\ell'(t) = 0$ one has

$$\partial_t^2 D_\Phi(y; x(t)) = \nabla^2 \Phi(x(t))[x'(t), x'(t)] = \eta^2 (\nabla^2 \Phi(x(t)))^{-1} [\ell(t), \ell(t)] \,.$$

Thus provided that the Hessian of Φ is well-conditioned on the scale of a mirror step, one expects a discrete time analysis to give a regret bound of the form (with the notation $\|h\|_x = \sqrt{\nabla^2 \Phi(x)[h, h]}$)

$$\frac{D_\Phi(y; x_1)}{\eta} + \eta \sum_{t=1}^{T} \|\ell_t\|_{x_t,*}^2 \,.$$

The more classical discrete-time algorithm and analysis

Ignoring the Lagrangian and assuming $\ell'(t) = 0$ one has

$$\partial_t^2 D_\Phi(y; x(t)) = \nabla^2 \Phi(x(t))[x'(t), x'(t)] = \eta^2 (\nabla^2 \Phi(x(t)))^{-1}[\ell(t), \ell(t)] \,.$$

Thus provided that the Hessian of Φ is well-conditioned on the scale of a mirror step, one expects a discrete time analysis to give a regret bound of the form (with the notation $\|h\|_x = \sqrt{\nabla^2 \Phi(x)[h,h]}$)

$$\frac{D_\Phi(y; x_1)}{\eta} + \eta \sum_{t=1}^{T} \|\ell_t\|_{x_t,*}^2 \,.$$

Theorem

The above is valid with a factor $2/c$ on the second term, provided that the following implication holds true for any $y_t \in \mathbb{R}^n$,

$$\nabla \Phi(y_t) \in [\nabla \Phi(x_t), \nabla \Phi(x_t) - \eta \ell_t] \Rightarrow \nabla^2 \Phi(y_t) \succeq c \nabla^2 \Phi(x_t) \,.$$

For FTRL one instead needs this for any $y_t \in [x_t, x_{t+1}]$.

MW is mirror descent with the negentropy

Let $\Phi(x) = \sum_{i=1}^{n} (x_i \log x_i - x_i)$ and $K = \Delta_n$. One has $\nabla \Phi(x) = \log(x_i)$ and thus the update step in the dual looks like:

$$\nabla \Phi(y_t) = \nabla \Phi(x_t) - \eta \ell_t \Leftrightarrow y_{i,t} = x_{i,t} \exp(-\eta \ell_t(i)) \,.$$

MW is mirror descent with the negentropy

Let $\Phi(x) = \sum_{i=1}^{n}(x_i \log x_i - x_i)$ and $K = \Delta_n$. One has $\nabla\Phi(x) = \log(x_i)$ and thus the update step in the dual looks like:

$$\nabla\Phi(y_t) = \nabla\Phi(x_t) - \eta\ell_t \Leftrightarrow y_{i,t} = x_{i,t}\exp(-\eta\ell_t(i)).$$

Furthermore the projection step to K amounts simply to a renormalization. Indeed $\nabla_x D_\Phi(x, y) = \sum_{i=1}^{n}\log(x_i/y_i)$ and thus

$$p = \operatorname*{argmin}_{x \in \Delta_n} D_\Phi(x, y) \Leftrightarrow \exists \mu \in \mathbb{R} : \log(p_i/y_i) = \mu, \forall i \in [n].$$

MW is mirror descent with the negentropy

Let $\Phi(x) = \sum_{i=1}^{n} (x_i \log x_i - x_i)$ and $K = \Delta_n$. One has $\nabla \Phi(x) = \log(x_i)$ and thus the update step in the dual looks like:

$$\nabla \Phi(y_t) = \nabla \Phi(x_t) - \eta \ell_t \Leftrightarrow y_{i,t} = x_{i,t} \exp(-\eta \ell_t(i)) \, .$$

Furthermore the projection step to K amounts simply to a renormalization. Indeed $\nabla_x D_\Phi(x, y) = \sum_{i=1}^{n} \log(x_i / y_i)$ and thus

$$p = \operatorname*{argmin}_{x \in \Delta_n} D_\Phi(x, y) \Leftrightarrow \exists \mu \in \mathbb{R} : \log(p_i / y_i) = \mu, \forall i \in [n] \, .$$

The analysis considers the potential $D_\Phi(i^*, p_t) = -\log(p_t(i^*))$, which in fact exactly corresponds to what we did in the second slide of the first lecture.

MW is mirror descent with the negentropy

Let $\Phi(x) = \sum_{i=1}^{n}(x_i \log x_i - x_i)$ and $K = \Delta_n$. One has $\nabla\Phi(x) = \log(x_i)$ and thus the update step in the dual looks like:

$$\nabla\Phi(y_t) = \nabla\Phi(x_t) - \eta\ell_t \Leftrightarrow y_{i,t} = x_{i,t}\exp(-\eta\ell_t(i))\,.$$

Furthermore the projection step to K amounts simply to a renormalization. Indeed $\nabla_x D_\Phi(x,y) = \sum_{i=1}^{n}\log(x_i/y_i)$ and thus

$$p = \operatorname*{argmin}_{x\in\Delta_n} D_\Phi(x,y) \Leftrightarrow \exists \mu \in \mathbb{R} : \log(p_i/y_i) = \mu, \forall i \in [n]\,.$$

The analysis considers the potential $D_\Phi(i^*, p_t) = -\log(p_t(i^*))$, which in fact exactly corresponds to what we did in the second slide of the first lecture.

Note also that the well-conditioning comes for free when $\ell_t(i) \geq 0$, and in general one just needs $\|\eta\ell_t\|_\infty$ to be $O(1)$.

Propensity score for the bandit game

Key idea: replace ℓ_t by $\widetilde{\ell}_t$ such that $\mathbb{E}_{i_t \sim p_t} \widetilde{\ell}_t = \ell_t$. The propensity score normalized estimator is defined by:

$$\widetilde{\ell}_t(i) = \frac{\ell_t(i_t)}{p_t(i)} \mathbb{1}\{i = i_t\} .$$

Propensity score for the bandit game

Key idea: replace ℓ_t by $\widetilde{\ell}_t$ such that $\mathbb{E}_{i_t \sim p_t} \widetilde{\ell}_t = \ell_t$. The propensity score normalized estimator is defined by:

$$\widetilde{\ell}_t(i) = \frac{\ell_t(i_t)}{p_t(i)} \mathbb{1}\{i = i_t\} .$$

The Exp3 strategy corresponds to doing MW with those estimators. Its regret is upper bounded by,

$$\mathbb{E} \sum_{t=1}^{T} \langle p_t - q, \ell_t \rangle = \mathbb{E} \sum_{t=1}^{T} \langle p_t - q, \widetilde{\ell}_t \rangle \leq \frac{\log(n)}{\eta} + \eta \mathbb{E} \sum_t \|\widetilde{\ell}_t\|_{p_t,*}^2 ,$$

where $\|h\|_{p,*}^2 = \sum_{i=1}^n p(i) h(i)^2$.

Propensity score for the bandit game

Key idea: replace ℓ_t by $\widetilde{\ell}_t$ such that $\mathbb{E}_{i_t \sim p_t} \widetilde{\ell}_t = \ell_t$. The propensity score normalized estimator is defined by:

$$\widetilde{\ell}_t(i) = \frac{\ell_t(i_t)}{p_t(i)} \mathbb{1}\{i = i_t\} .$$

The Exp3 strategy corresponds to doing MW with those estimators. Its regret is upper bounded by,

$$\mathbb{E} \sum_{t=1}^{T} \langle p_t - q, \ell_t \rangle = \mathbb{E} \sum_{t=1}^{T} \langle p_t - q, \widetilde{\ell}_t \rangle \leq \frac{\log(n)}{\eta} + \eta \mathbb{E} \sum_t \|\widetilde{\ell}_t\|_{p_t,*}^2 ,$$

where $\|h\|_{p,*}^2 = \sum_{i=1}^n p(i) h(i)^2$. Amazingly the variance term is automatically controlled:

$$\mathbb{E}_{i_t \sim p_t} \sum_{i=1}^{n} p_t(i) \widetilde{\ell}_t(i)^2 \leq \mathbb{E}_{i_t \sim p_t} \sum_{i=1}^{n} \frac{\mathbb{1}\{i = i_t\}}{p_t(i_t)} = n.$$

Propensity score for the bandit game

Key idea: replace ℓ_t by $\widetilde{\ell}_t$ such that $\mathbb{E}_{i_t \sim p_t} \widetilde{\ell}_t = \ell_t$. The propensity score normalized estimator is defined by:

$$\widetilde{\ell}_t(i) = \frac{\ell_t(i_t)}{p_t(i)} \mathbb{1}\{i = i_t\} .$$

The Exp3 strategy corresponds to doing MW with those estimators. Its regret is upper bounded by,

$$\mathbb{E} \sum_{t=1}^{T} \langle p_t - q, \ell_t \rangle = \mathbb{E} \sum_{t=1}^{T} \langle p_t - q, \widetilde{\ell}_t \rangle \leq \frac{\log(n)}{\eta} + \eta \mathbb{E} \sum_t \|\widetilde{\ell}_t\|_{p_t,*}^2 ,$$

where $\|h\|_{p,*}^2 = \sum_{i=1}^n p(i) h(i)^2$. Amazingly the variance term is automatically controlled:

$$\mathbb{E}_{i_t \sim p_t} \sum_{i=1}^{n} p_t(i) \widetilde{\ell}_t(i)^2 \leq \mathbb{E}_{i_t \sim p_t} \sum_{i=1}^{n} \frac{\mathbb{1}\{i = i_t\}}{p_t(i_t)} = n.$$

Thus with $\eta = \sqrt{n \log(n)/T}$ one gets $R_T \leq 2\sqrt{Tn \log(n)}$.

Simple extensions

- Removing the extraneous $\sqrt{\log(n)}$
- Contextual bandit
- Bandit with side information
- Different scaling per actions

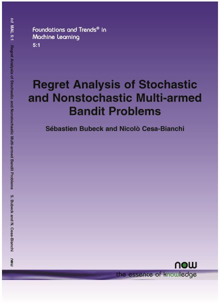
More subtle refinements

- Sparse bandit
- Variance bounds
- First order bounds
- Best of both worlds
- Impossibility of $\sqrt{T}$ with switching cost
- Impossibility of oracle models
- Knapsack bandits

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

Sébastien Bubeck
Machine Learning and Optimization group, MSR AI

Microsoft Research



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.

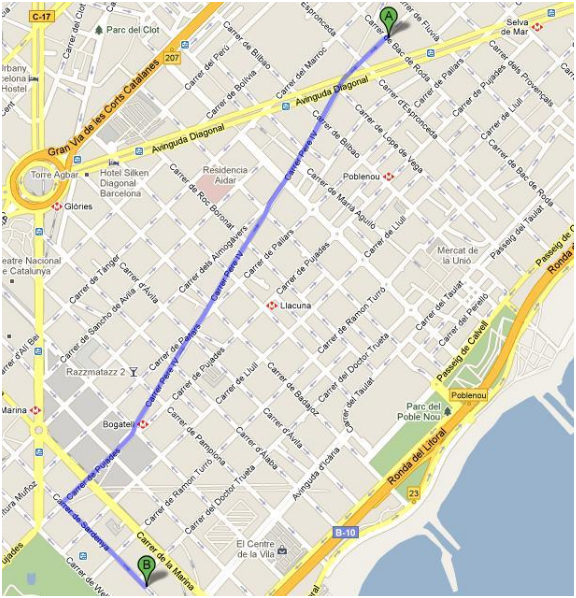
Protocol: For each round $t \in [T]$, player chooses $a_t \in \mathcal{A}$ and simultaneously adversary chooses a loss function $\ell_t \in [0,1]^n$. Loss suffered is $\ell_t \cdot a_t$.

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:

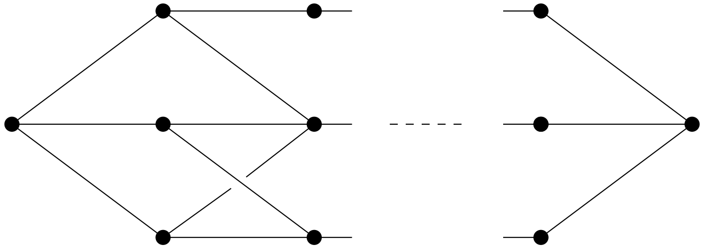
$$R_T := \mathbb{E} \sum_{t=1}^{T} \ell_t \cdot a_t - \min_{a \in \mathcal{A}} \mathbb{E} \sum_{t=1}^{T} \ell_t \cdot a .$$

Example: path planning



Example: path planning

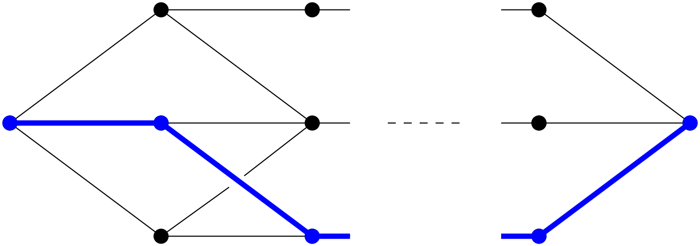
Adversary

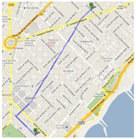


Player

Example: path planning

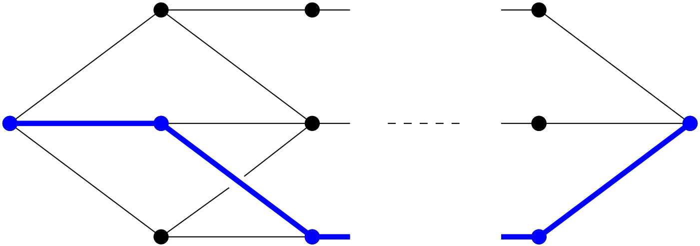
Adversary



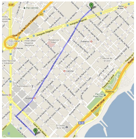
Player ⟶ 

Example: path planning

Adversary →

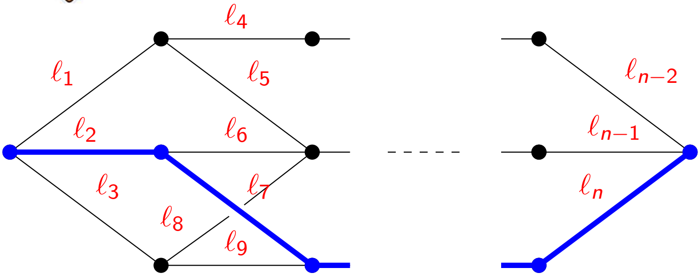


Player →

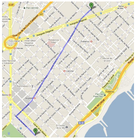


Example: path planning

Adversary ⟶

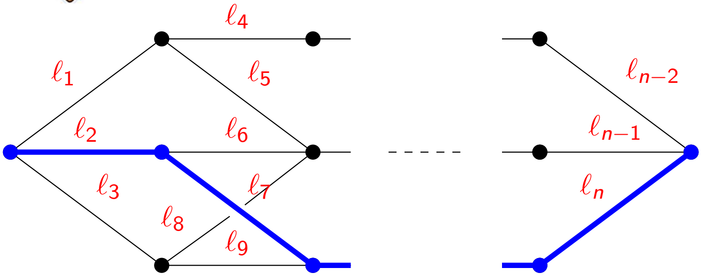


Player ⟶

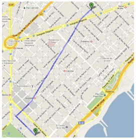


Example: path planning

Adversary ⟶



Player ⟶

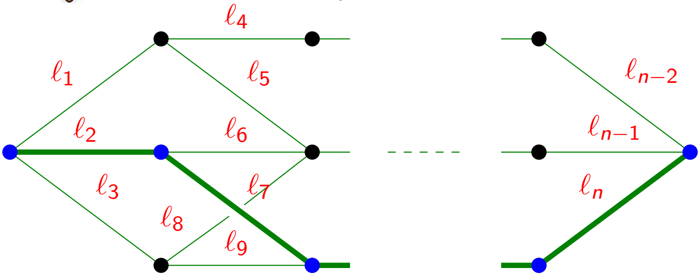


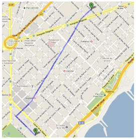
loss suffered: $\ell_2 + \ell_7 + \ldots + \ell_n$

Example: path planning

Adversary ⟶ 

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



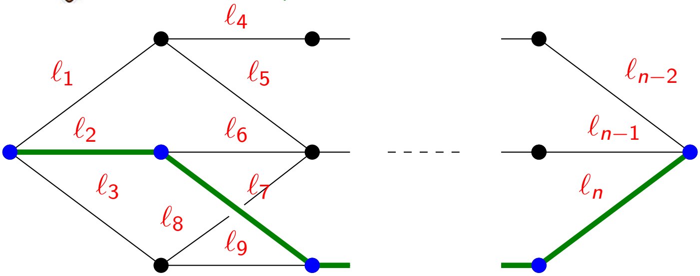
Player ⟶ 

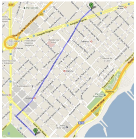
loss suffered: $\ell_2 + \ell_7 + \ldots + \ell_n$

Example: path planning

Adversary ⟶  Feedback:

- Full Info: $\ell_1, \ell_2, \ldots, \ell_n$
- Semi-Bandit: $\ell_2, \ell_7, \ldots, \ell_n$



Player ⟶ 

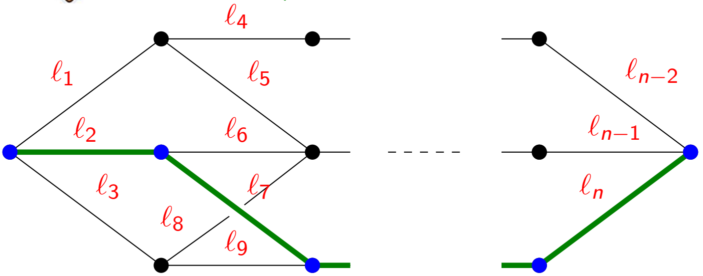
loss suffered: $\ell_2 + \ell_7 + \ldots + \ell_n$

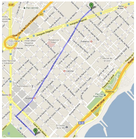
Example: path planning

Adversary ⟶ 

Feedback:

Full Info:	$\ell_1, \ell_2, \ldots, \ell_n$
Semi-Bandit:	$\ell_2, \ell_7, \ldots, \ell_n$
Bandit:	$\ell_2 + \ell_7 + \ldots + \ell_n$



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 $\mathcal{A} = \{a_1, \ldots, a_{|\mathcal{A}|}\}$. 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 $\bar{\ell}_t(i,j) = \mathbb{E}_t(\ell_t(i)|i^* = j)$.

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)

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

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)

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

Proof.

(i) self-concordance is invariant by Fenchel duality
(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) $\nabla^k \mathbb{e}^*(x) = \mathbb{E}_{X \sim p_{\theta(x)}}(X - \mathbb{E}X)^{\otimes k}$ 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 $\Rightarrow$ "sub-CLT" for $p_\theta \Rightarrow \nu$-s.c (bit more involved than (i)-(ii)-(iii)) □

(iv) in a nutshell

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

(iv) in a nutshell

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

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

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

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 $v'' \le -\frac{1}{n}(v')^2$ and so

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

(iv) in a nutshell

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

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)

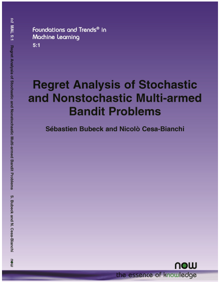
Let $f : K \to [0, +\infty)$ be convex and 1-Lipschitz, and $\varepsilon > 0$. There exists a probability measure μ on K such that the following holds true. For every $\alpha \in K$ and for every convex and 1-Lipschitz function $g : K \to \mathbb{R}$ satisfying $g(\alpha) < -\varepsilon$, one has

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

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.

Lecture 4:
Kernel-based methods for bandit convex optimization

Sébastien Bubeck
Machine Learning and Optimization group, MSR AI



Kernel-based methods

Notation: $\langle f, g \rangle := \int_{x \in \mathbb{R}^n} f(x) g(x) dx$. The expected regret with respect to point x can be written as $\sum_{t=1}^{T} \langle p_t - \delta_x, \ell_t \rangle$.

Kernel-based methods

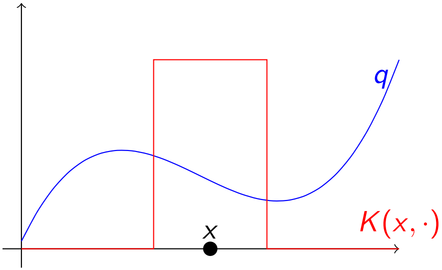
Notation: $\langle f, g \rangle := \int_{x \in \mathbb{R}^n} f(x) g(x) dx$. The expected regret with respect to point x can be written as $\sum_{t=1}^{T} \langle p_t - \delta_x, \ell_t \rangle$.

Kernel: $K : \mathcal{K} \times \mathcal{K} \to \mathbb{R}_+$ which we view as a linear operator over measures via $Kq(x) = \int K(x, y) q(y) dy$. The adjoint K^* acts on functions: $K^* f(y) = \int f(x) K(x, y) dx$ (since $\langle Kq, f \rangle = \langle q, K^* f \rangle$).

Kernel-based methods

Notation: $\langle f, g \rangle := \int_{x \in \mathbb{R}^n} f(x) g(x) dx$. The expected regret with respect to point x can be written as $\sum_{t=1}^{T} \langle p_t - \delta_x, \ell_t \rangle$.

Kernel: $K : \mathcal{K} \times \mathcal{K} \to \mathbb{R}_+$ which we view as a linear operator over measures via $Kq(x) = \int K(x, y) q(y) dy$. The adjoint K^* acts on functions: $K^* f(y) = \int f(x) K(x, y) dx$ (since $\langle Kq, f \rangle = \langle q, K^* f \rangle$).



Kernel-based methods

Notation: $\langle f, g \rangle := \int_{x \in \mathbb{R}^n} f(x) g(x) dx$. The expected regret with respect to point x can be written as $\sum_{t=1}^{T} \langle p_t - \delta_x, \ell_t \rangle$.

Kernel: $K : \mathcal{K} \times \mathcal{K} \to \mathbb{R}_+$ which we view as a linear operator over measures via $Kq(x) = \int K(x, y) q(y) dy$. The adjoint K^* acts on functions: $K^* f(y) = \int f(x) K(x, y) dx$ (since $\langle Kq, f \rangle = \langle q, K^* f \rangle$).

Key point: canonical estimator of $K^* f$ based on bandit feedback on f:

$$\mathbb{E}_{x \sim q} \frac{f(x) K(x, \cdot)}{q(x)} = K^* f$$

Kernel-based methods

Notation: $\langle f, g\rangle := \int_{x\in\mathbb{R}^n} f(x)g(x)dx$. The expected regret with respect to point x can be written as $\sum_{t=1}^{T}\langle p_t - \delta_x, \ell_t\rangle$.

Kernel: $K : \mathcal{K} \times \mathcal{K} \to \mathbb{R}_+$ which we view as a linear operator over measures via $Kq(x) = \int K(x,y)q(y)dy$. The adjoint K^* acts on functions: $K^*f(y) = \int f(x)K(x,y)dx$ (since $\langle Kq, f\rangle = \langle q, K^*f\rangle$).

Key point: canonical estimator of K^*f based on bandit feedback on f:

$$\mathbb{E}_{x\sim q}\frac{f(x)K(x,\cdot)}{q(x)} = K^*f$$

Kernelized regret? Say p_t is full info strat with $\widetilde{\ell}_t = \frac{\ell_t(x_t)K_t(x_t,\cdot)}{q_t(x_t)}$ and $x_t \sim q_t$.

Kernel-based methods

Notation: $\langle f, g \rangle := \int_{x \in \mathbb{R}^n} f(x) g(x) dx$. The expected regret with respect to point x can be written as $\sum_{t=1}^{T} \langle p_t - \delta_x, \ell_t \rangle$.

Kernel: $K : \mathcal{K} \times \mathcal{K} \to \mathbb{R}_+$ which we view as a linear operator over measures via $Kq(x) = \int K(x, y) q(y) dy$. The adjoint K^* acts on functions: $K^* f(y) = \int f(x) K(x, y) dx$ (since $\langle Kq, f \rangle = \langle q, K^* f \rangle$).

Key point: canonical estimator of $K^* f$ based on bandit feedback on f:

$$\mathbb{E}_{x \sim q} \frac{f(x) K(x, \cdot)}{q(x)} = K^* f$$

Kernelized regret? Say p_t is full info strat with $\widetilde{\ell}_t = \frac{\ell_t(x_t) K_t(x_t, \cdot)}{q_t(x_t)}$ and $x_t \sim q_t$. Then we can hope to control the regret with terms $\langle p_t - \delta_x, K_t^* \ell_t \rangle = \langle K_t(p_t - \delta_x), \ell_t \rangle$ while we want to control $\langle q_t - \delta_x, \ell_t \rangle$.

Kernel-based methods

Notation: $\langle f, g \rangle := \int_{x \in \mathbb{R}^n} f(x) g(x) dx$. The expected regret with respect to point x can be written as $\sum_{t=1}^{T} \langle p_t - \delta_x, \ell_t \rangle$.

Kernel: $K : \mathcal{K} \times \mathcal{K} \to \mathbb{R}_+$ which we view as a linear operator over measures via $Kq(x) = \int K(x, y) q(y) dy$. The adjoint K^* acts on functions: $K^* f(y) = \int f(x) K(x, y) dx$ (since $\langle Kq, f \rangle = \langle q, K^* f \rangle$).

Key point: canonical estimator of $K^* f$ based on bandit feedback on f:

$$\mathbb{E}_{x \sim q} \frac{f(x) K(x, \cdot)}{q(x)} = K^* f$$

Kernelized regret? Say p_t is full info strat with $\widetilde{\ell}_t = \frac{\ell_t(x_t) K_t(x_t, \cdot)}{q_t(x_t)}$ and $x_t \sim q_t$. Then we can hope to control the regret with terms $\langle p_t - \delta_x, K_t^* \ell_t \rangle = \langle K_t(p_t - \delta_x), \ell_t \rangle$ while we want to control $\langle q_t - \delta_x, \ell_t \rangle$. Seems reasonable to take $q_t := K_t p_t$ and then we want:

$$\langle K_t p_t - \delta_x, \ell_t \rangle \lesssim \langle K_t(p_t - \delta_x), \ell_t \rangle$$

A good kernel for convex losses

$$\langle K_t p_t - \delta_x, \ell_t \rangle \lesssim \langle K_t (p_t - \delta_x), \ell_t \rangle$$

A good kernel for convex losses

$$\langle K_t p_t - \delta_x, \ell_t \rangle \lesssim \langle K_t(p_t - \delta_x), \ell_t \rangle$$

Thus for a given p we want a kernel K such that $\forall x$ and f convex one has (for some $\lambda \in (0,1)$)

$$\langle Kp - \delta_x, f \rangle \leq \frac{1}{\lambda} \langle K(p - \delta_x), f \rangle \Leftrightarrow K^* f(x) \leq (1-\lambda)\langle Kp, f \rangle + \lambda f(x)$$

A good kernel for convex losses

$$\langle K_t p_t - \delta_x, \ell_t \rangle \lesssim \langle K_t(p_t - \delta_x), \ell_t \rangle$$

Thus for a given p we want a kernel K such that $\forall x$ and f convex one has (for some $\lambda \in (0,1)$)

$$\langle Kp - \delta_x, f \rangle \leq \frac{1}{\lambda} \langle K(p - \delta_x), f \rangle \Leftrightarrow K^* f(x) \leq (1-\lambda)\langle Kp, f \rangle + \lambda f(x)$$

Natural kernel: $K\delta_x$ is the distribution of $(1-\lambda)Z + \lambda x$ for some random variable Z to be defined. Indeed in this case one has

$$K^* f(x) = \mathbb{E} f((1-\lambda)Z + \lambda x) \leq (1-\lambda)\mathbb{E} f(Z) + \lambda f(x)$$

A good kernel for convex losses

$$\langle K_t p_t - \delta_x, \ell_t \rangle \lesssim \langle K_t(p_t - \delta_x), \ell_t \rangle$$

Thus for a given p we want a kernel K such that $\forall x$ and f convex one has (for some $\lambda \in (0,1)$)

$$\langle Kp - \delta_x, f \rangle \leq \frac{1}{\lambda} \langle K(p - \delta_x), f \rangle \Leftrightarrow K^* f(x) \leq (1-\lambda) \langle Kp, f \rangle + \lambda f(x)$$

Natural kernel: $K\delta_x$ is the distribution of $(1-\lambda)Z + \lambda x$ for some random variable Z to be defined. Indeed in this case one has

$$K^* f(x) = \mathbb{E} f((1-\lambda)Z + \lambda x) \leq (1-\lambda)\mathbb{E} f(Z) + \lambda f(x)$$

Thus we would like Z to be equal to Kp, that is Z satisfies the following distributional identity, where $X \sim p$,

$$Z \overset{D}{=} (1-\lambda)Z + \lambda X$$

Generalized Bernoulli convolutions

$$Z \stackrel{D}{=} (1-\lambda)Z + \lambda X$$

Generalized Bernoulli convolutions

$$Z \stackrel{D}{=} (1-\lambda)Z + \lambda X$$

We say that Z is the *core* of p. It satisfies $Z = \sum_{k=0}^{+\infty} \lambda(1-\lambda)^k X_k$ with (X_k) i.i.d. sequence from p. We need to understand the "smoothness" of Z (which will translate in smoothness of the corresponding kernel).

Generalized Bernoulli convolutions

$$Z \overset{D}{=} (1-\lambda)Z + \lambda X$$

We say that Z is the *core* of p. It satisfies $Z = \sum_{k=0}^{+\infty} \lambda(1-\lambda)^k X_k$ with (X_k) i.i.d. sequence from p. We need to understand the "smoothness" of Z (which will translate in smoothness of the corresponding kernel).

Consider the core ν_λ of a random sign (this is a distinguished object introduced in the 1930's known as a Bernoulli convolution):

Generalized Bernoulli convolutions

$$Z \overset{D}{=} (1-\lambda)Z + \lambda X$$

We say that Z is the *core* of p. It satisfies $Z = \sum_{k=0}^{+\infty} \lambda(1-\lambda)^k X_k$ with (X_k) i.i.d. sequence from p. We need to understand the "smoothness" of Z (which will translate in smoothness of the corresponding kernel).

Consider the core ν_λ of a random sign (this is a distinguished object introduced in the 1930's known as a Bernoulli convolution):

- Wintner 1935: ν_λ is either absolutely continuous or singular w.r.t. Lebesgue. For $\lambda \in (1/2, 1)$ is it singular, and for $\lambda = 1/2$ it is a.c.

Generalized Bernoulli convolutions

$$Z \overset{D}{=} (1-\lambda)Z + \lambda X$$

We say that Z is the *core* of p. It satisfies $Z = \sum_{k=0}^{+\infty} \lambda(1-\lambda)^k X_k$ with (X_k) i.i.d. sequence from p. We need to understand the "smoothness" of Z (which will translate in smoothness of the corresponding kernel).

Consider the core ν_λ of a random sign (this is a distinguished object introduced in the 1930's known as a Bernoulli convolution):

- Wintner 1935: ν_λ is either absolutely continuous or singular w.r.t. Lebesgue. For $\lambda \in (1/2, 1)$ is it singular, and for $\lambda = 1/2$ it is a.c.
- Erdős 1939: $\exists \infty$ of singular $\lambda \in (0, 1/2)$.

Generalized Bernoulli convolutions

$$Z \stackrel{D}{=} (1-\lambda)Z + \lambda X$$

We say that Z is the *core* of p. It satisfies $Z = \sum_{k=0}^{+\infty} \lambda(1-\lambda)^k X_k$ with (X_k) i.i.d. sequence from p. We need to understand the "smoothness" of Z (which will translate in smoothness of the corresponding kernel).

Consider the core ν_λ of a random sign (this is a distinguished object introduced in the 1930's known as a Bernoulli convolution):

- Wintner 1935: ν_λ is either absolutely continuous or singular w.r.t. Lebesgue. For $\lambda \in (1/2, 1)$ is it singular, and for $\lambda = 1/2$ it is a.c.
- Erdős 1939: $\exists \infty$ of singular $\lambda \in (0, 1/2)$.
- Erdős 1940, Solomyak 1996: a.e. $\lambda \in (0, 1/2)$ is a.c.

Generalized Bernoulli convolutions

$$Z \overset{D}{=} (1-\lambda)Z + \lambda X$$

We say that Z is the *core* of p. It satisfies $Z = \sum_{k=0}^{+\infty} \lambda(1-\lambda)^k X_k$ with (X_k) i.i.d. sequence from p. We need to understand the "smoothness" of Z (which will translate in smoothness of the corresponding kernel).

Consider the core ν_λ of a random sign (this is a distinguished object introduced in the 1930's known as a Bernoulli convolution):

- Wintner 1935: ν_λ is either absolutely continuous or singular w.r.t. Lebesgue. For $\lambda \in (1/2, 1)$ is it singular, and for $\lambda = 1/2$ it is a.c.
- Erdős 1939: $\exists \infty$ of singular $\lambda \in (0, 1/2)$.
- Erdős 1940, Solomyak 1996: a.e. $\lambda \in (0, 1/2)$ is a.c.
- For any $k \in \mathbb{N}$, $\exists \lambda_k \approx 1/k$ s.t. ν_{λ_k} has a C^k density.

What is left to do?

Summarizing the discussion so far, let us play from $K_t p_t$, where K_t is the kernel described above (i.e., it "mixes in" the core of p_t) and p_t is the continuous exponential weights strategy on the estimated losses $\widetilde{\ell}_s = \ell_s(x_s) \frac{K_s(x_s, \cdot)}{K_s p_s(x_s)}$ (that is $dp_t(x)/dx$ is proportional to $\exp(-\eta \sum_{s<t} \widetilde{\ell}_s(x)))$.

What is left to do?

Summarizing the discussion so far, let us play from $K_t p_t$, where K_t is the kernel described above (i.e., it "mixes in" the core of p_t) and p_t is the continuous exponential weights strategy on the estimated losses $\widetilde{\ell}_s = \ell_s(x_s) \frac{K_s(x_s,\cdot)}{K_s p_s(x_s)}$ (that is $dp_t(x)/dx$ is proportional to $\exp(-\eta \sum_{s<t} \widetilde{\ell}_s(x)))$.
Using the classical analysis of continuous exponential weights together with the previous slides we get for any q,

$$\begin{aligned}
\mathbb{E} \sum_{t=1}^{T} \langle K_t p_t - q, \ell_t \rangle &\leq \frac{1}{\lambda} \mathbb{E} \sum_{t=1}^{T} \langle K_t (p_t - q), \ell_t \rangle \\
&= \frac{1}{\lambda} \mathbb{E} \sum_{t=1}^{T} (\langle p_t - q, \widetilde{\ell}_t \rangle) \\
&\leq \frac{1}{\lambda} \mathbb{E} \left(\frac{\mathrm{Ent}(q \| p_1)}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \langle p_t, \left(\frac{K_t(x_t, \cdot)}{K_t p_t(x_t)} \right)^2 \rangle \right) .
\end{aligned}$$

Variance calculation

All that remains to be done is to control the variance term $\mathbb{E}_{x \sim Kp} \langle p, \widetilde{\ell}^2 \rangle$ where $\widetilde{\ell}(y) = \frac{K(x,y)}{Kp(x)} = \frac{K(x,y)}{\int K(x,y')p(y')dy}$. More precisely if this quantity is $O(1)$ then we obtain a regret of $\widetilde{O}\left(\frac{1}{\lambda}\sqrt{nT}\right)$.

Variance calculation

All that remains to be done is to control the variance term $\mathbb{E}_{x \sim Kp} \langle p, \widetilde{\ell}^2 \rangle$ where $\widetilde{\ell}(y) = \frac{K(x,y)}{Kp(x)} = \frac{K(x,y)}{\int K(x,y')p(y')dy}$. More precisely if this quantity is $O(1)$ then we obtain a regret of $\widetilde{O}\left(\frac{1}{\lambda}\sqrt{nT}\right)$.

It is sufficient to control from above $K(x,y)/K(x,y')$ for all y, y' in the support of p and all x in the support of Kp (in fact it is sufficient to have it with probability at least $1 - 1/T^{10}$ w.r.t. $x \sim Kp$).

Variance calculation

All that remains to be done is to control the variance term $\mathbb{E}_{x \sim Kp} \langle p, \widetilde{\ell}^2 \rangle$ where $\widetilde{\ell}(y) = \frac{K(x,y)}{Kp(x)} = \frac{K(x,y)}{\int K(x,y')p(y')dy}$. More precisely if this quantity is $O(1)$ then we obtain a regret of $\widetilde{O}\left(\frac{1}{\lambda}\sqrt{nT}\right)$.

It is sufficient to control from above $K(x,y)/K(x,y')$ for all y, y' in the support of p and all x in the support of Kp (in fact it is sufficient to have it with probability at least $1 - 1/T^{10}$ w.r.t. $x \sim Kp$).
Observe also that, with c denoting the core of p, one always has $K(x,y) = K\delta_y(x) = \mathrm{cst} \times c\left(\frac{x - \lambda y}{1 - \lambda}\right)$. Thus we want to bound w.h.p w.r.t. $x \sim Kp$,

$$\sup_{y, y' \in \mathrm{supp}(p)} c\left(\frac{x - \lambda y}{1 - \lambda}\right) / c\left(\frac{x - \lambda y'}{1 - \lambda}\right) .$$

Variance calculation heuristic

Control w.h.p w.r.t. $x \sim Kp$,

$$\sup_{y,y' \in \mathrm{supp}(p)} c\left(\frac{x-\lambda y}{1-\lambda}\right) \Big/ c\left(\frac{x-\lambda y'}{1-\lambda}\right) .$$

Variance calculation heuristic

Control w.h.p w.r.t. $x \sim Kp$,

$$\sup_{y,y' \in \mathrm{supp}(p)} c\left(\frac{x - \lambda y}{1 - \lambda}\right) / c\left(\frac{x - \lambda y'}{1 - \lambda}\right) .$$

Let us assume

1. $p = \mathcal{N}(0, \mathrm{I}_n)$ (its core is $c = \mathcal{N}(0, \frac{\lambda}{2-\lambda} I_n)$).

Variance calculation heuristic

Control w.h.p w.r.t. $x \sim Kp$,

$$\sup_{y, y' \in \mathrm{supp}(p)} c\left(\frac{x - \lambda y}{1 - \lambda}\right) / c\left(\frac{x - \lambda y'}{1 - \lambda}\right).$$

Let us assume

1. $p = \mathcal{N}(0, \mathrm{I}_n)$ (its core is $c = \mathcal{N}(0, \frac{\lambda}{2-\lambda} I_n)$).
2. $\mathrm{supp}(p) \subset \{y : |y| \leq R = \widetilde{O}(\sqrt{n})\}$

Variance calculation heuristic

Control w.h.p w.r.t. $x \sim Kp$,

$$\sup_{y,y' \in \mathrm{supp}(p)} c\left(\frac{x - \lambda y}{1-\lambda}\right) / c\left(\frac{x-\lambda y'}{1-\lambda}\right) .$$

Let us assume

1. $p = \mathcal{N}(0, \mathrm{I}_n)$ (its core is $c = \mathcal{N}(0, \frac{\lambda}{2-\lambda} I_n)$).
2. $\mathrm{supp}(p) \subset \{y : |y| \leq R = \widetilde{O}(\sqrt{n})\}$

Thus our quantity of interest is

$$\exp\left(\frac{2-\lambda}{2\lambda}\left(\left|\frac{x-\lambda y'}{1-\lambda}\right|^2 - \left|\frac{x-\lambda y}{1-\lambda}\right|^2\right)\right)$$
$$\leq \exp\left(\frac{1}{(1-\lambda)^2}(4R|x| + 2\lambda R^2)\right) .$$

Finally note that w.h.p. one has $|x| \lesssim \lambda R + \sqrt{\lambda n \log(T)}$, and thus with $\lambda = \widetilde{O}(1/n^2)$ we have a constant variance.

A reduction to the Gaussian case

We reduce to the Gaussian situation by observing that taking Z (in the definition of the kernel) to be the core of a measure convexly dominated by p is sufficient (instead of taking it to be directly the core of p), and furthermore one has:

A reduction to the Gaussian case

We reduce to the Gaussian situation by observing that taking Z (in the definition of the kernel) to be the core of a measure convexly dominated by p is sufficient (instead of taking it to be directly the core of p), and furthermore one has:

Lemma

Any isotropic log-concave measure p approximately convexly dominates a centered Gaussian with covariance $\widetilde{O}(\frac{1}{n})\mathrm{I}_n$.

A reduction to the Gaussian case

We reduce to the Gaussian situation by observing that taking Z (in the definition of the kernel) to be the core of a measure convexly dominated by p is sufficient (instead of taking it to be directly the core of p), and furthermore one has:

Lemma

Any isotropic log-concave measure p approximately convexly dominates a centered Gaussian with covariance $\widetilde{O}(\frac{1}{n})\mathrm{I}_n$.

Proof.

We show that p dominates any q supported on a small ball of cst radius. Pick a test function f, w.l.o.g. its minimum is 0 at 0 and the maximum on the ball is 1. By convexity f is above a linear function (maxed with 0) of constant slope. By light tails of log-concave, $\langle p, f \rangle$ is then at least a constant. □

A reduction to the Gaussian case

We reduce to the Gaussian situation by observing that taking Z (in the definition of the kernel) to be the core of a measure convexly dominated by p is sufficient (instead of taking it to be directly the core of p), and furthermore one has:

Lemma

Any isotropic log-concave measure p approximately convexly dominates a centered Gaussian with covariance $\widetilde{O}(\frac{1}{n})\mathrm{I}_n$.

Proof.

We show that p dominates any q supported on a small ball of cst radius. Pick a test function f, w.l.o.g. its minimum is 0 at 0 and the maximum on the ball is 1. By convexity f is above a linear function (maxed with 0) of constant slope. By light tails of log-concave, $\langle p, f \rangle$ is then at least a constant. □

What about assumption 2?

Restart and increasing learning rate

Unfortunately assumption 2 brings out a serious difficulty: it forces the algorithm to focus on smaller and smaller region of space. What if the adversary makes us focus on a region only to move the optimum far outside of it at a later time?

Restart and increasing learning rate

Unfortunately assumption 2 brings out a serious difficulty: it forces the algorithm to focus on smaller and smaller region of space. What if the adversary makes us focus on a region only to move the optimum far outside of it at a later time?

Idea: if the estimated optimum is too close to the boundary of the focus region then we restart the algorithm (similar idea appeared in Hazan and Li 2016).

Restart and increasing learning rate

Unfortunately assumption 2 brings out a serious difficulty: it forces the algorithm to focus on smaller and smaller region of space. What if the adversary makes us focus on a region only to move the optimum far outside of it at a later time?

Idea: if the estimated optimum is too close to the boundary of the focus region then we restart the algorithm (similar idea appeared in Hazan and Li 2016).

To be proved: negative regret at restart times (indeed the adversary must "pay" for making us focus and then move out the optimum). Technically this negative regret can come from a large relative entropy at some previous time.

Restart and increasing learning rate

Unfortunately assumption 2 brings out a serious difficulty: it forces the algorithm to focus on smaller and smaller region of space. What if the adversary makes us focus on a region only to move the optimum far outside of it at a later time?

Idea: if the estimated optimum is too close to the boundary of the focus region then we restart the algorithm (similar idea appeared in Hazan and Li 2016).

To be proved: negative regret at restart times (indeed the adversary must "pay" for making us focus and then move out the optimum). Technically this negative regret can come from a large relative entropy at some previous time.

Challenge: avoid the telescopic sum of entropies. For this we use a last idea: every time the focus region changes scale we also increase the learning rate.

Summary of the algorithm

- Compute the Gaussian N_t "inside" p_t, its associated core N'_t (when N_t is isotropic: $N'_t = \sqrt{\frac{\lambda}{2-\lambda}} N_t$), and the corresponding kernel: $K_t \delta_y = (1-\lambda) N'_t + \lambda y$ (i.e. $K_t(x, y) = N'_t(\frac{x-\lambda y}{1-\lambda}) \propto \exp(-\frac{n}{\lambda} \|x - \lambda y\|^2_{p_t})$).

Summary of the algorithm

- Compute the Gaussian N_t "inside" p_t, its associated core N'_t (when N_t is isotropic: $N'_t = \sqrt{\frac{\lambda}{2-\lambda}} N_t$), and the corresponding kernel: $K_t \delta_y = (1-\lambda) N'_t + \lambda y$ (i.e. $K_t(x,y) = N'_t(\frac{x-\lambda y}{1-\lambda}) \propto \exp(-\frac{n}{\lambda}\|x-\lambda y\|^2_{p_t})$).
- Sample $X_t \sim p_t$ and play $x_t = (1-\lambda)N'_t + \lambda X_t \sim K_t p_t$.

Summary of the algorithm

- Compute the Gaussian N_t "inside" p_t, its associated core N'_t (when N_t is isotropic: $N'_t = \sqrt{\frac{\lambda}{2-\lambda}} N_t$), and the corresponding kernel: $K_t\delta_y = (1-\lambda)N'_t + \lambda y$ (i.e. $K_t(x,y) = N'_t(\frac{x-\lambda y}{1-\lambda}) \propto \exp(-\frac{n}{\lambda}\|x-\lambda y\|^2_{p_t})$).
- Sample $X_t \sim p_t$ and play $x_t = (1-\lambda)N'_t + \lambda X_t \sim K_t p_t$.
- Update the exponential weights distribution: $p_{t+1}(y) \propto p_t(y)\exp(-\eta_t \widetilde{\ell}_t(y))$

Summary of the algorithm

- Compute the Gaussian N_t "inside" p_t, its associated core N_t' (when N_t is isotropic: $N_t' = \sqrt{\frac{\lambda}{2-\lambda}} N_t$), and the corresponding kernel: $K_t\delta_y = (1-\lambda)N_t' + \lambda y$ (i.e. $K_t(x,y) = N_t'(\frac{x-\lambda y}{1-\lambda}) \propto \exp(-\frac{n}{\lambda}\|x-\lambda y\|_{p_t}^2)$).
- Sample $X_t \sim p_t$ and play $x_t = (1-\lambda)N_t' + \lambda X_t \sim K_t p_t$.
- Update the exponential weights distribution: $p_{t+1}(y) \propto p_t(y)\exp(-\eta_t \widetilde{\ell}_t(y))$ where

$$\widetilde{\ell}_t(y) = \frac{\ell_t(x_t)}{K_t p_t(x_t)} K_t(x_t, y) \propto \exp(-n\lambda\|y - x_t/\lambda\|_{p_t}^2)$$

Summary of the algorithm

- Compute the Gaussian N_t "inside" p_t, its associated core N'_t (when N_t is isotropic: $N'_t = \sqrt{\frac{\lambda}{2-\lambda}} N_t$), and the corresponding kernel: $K_t \delta_y = (1-\lambda) N'_t + \lambda y$ (i.e. $K_t(x, y) = N'_t(\frac{x - \lambda y}{1-\lambda}) \propto \exp(-\frac{n}{\lambda} \|x - \lambda y\|^2_{p_t})$).
- Sample $X_t \sim p_t$ and play $x_t = (1-\lambda) N'_t + \lambda X_t \sim K_t p_t$.
- Update the exponential weights distribution: $p_{t+1}(y) \propto p_t(y) \exp(-\eta_t \widetilde{\ell}_t(y))$ where

$$\widetilde{\ell}_t(y) = \frac{\ell_t(x_t)}{K_t p_t(x_t)} K_t(x_t, y) \propto \exp(-n\lambda \|y - x_t/\lambda\|^2_{p_t})$$

 (note that $\|x_t/\lambda\| \approx 1/\sqrt{\lambda}$ and the standard deviation of the above Gaussian is $\approx 1/\sqrt{n\lambda}$).

Summary of the algorithm

- Compute the Gaussian N_t "inside" p_t, its associated core N'_t (when N_t is isotropic: $N'_t = \sqrt{\frac{\lambda}{2-\lambda} N_t}$), and the corresponding kernel: $K_t \delta_y = (1-\lambda) N'_t + \lambda y$ (i.e. $K_t(x,y) = N'_t(\frac{x-\lambda y}{1-\lambda}) \propto \exp(-\frac{n}{\lambda}\|x - \lambda y\|^2_{p_t})$).
- Sample $X_t \sim p_t$ and play $x_t = (1-\lambda)N'_t + \lambda X_t \sim K_t p_t$.
- Update the exponential weights distribution: $p_{t+1}(y) \propto p_t(y) \exp(-\eta_t \widetilde{\ell}_t(y))$ where

$$\widetilde{\ell}_t(y) = \frac{\ell_t(x_t)}{K_t p_t(x_t)} K_t(x_t, y) \propto \exp(-n\lambda \|y - x_t/\lambda\|^2_{p_t})$$

  (note that $\|x_t/\lambda\| \approx 1/\sqrt{\lambda}$ and the standard deviation of the above Gaussian is $\approx 1/\sqrt{n\lambda}$).
- Restart business: check if adversary is potentially moving out of focus region (if so restart the algorithm), check if updating the focus region would change the problem's scale (if so make the update and increase the learning rate multiplicatively by $(1 + \frac{1}{\widetilde{O}(\mathrm{poly}(n))})$).