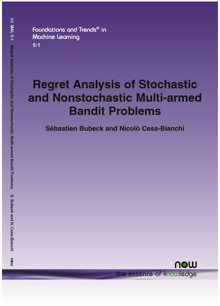
Lecture 2:
Mirror descent and online decision making

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 $\ell_1, \dots, \ell_{t-1} \in [-1, 1]^n$, 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:

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

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:

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

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

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

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

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

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 $x_1 = 0$,

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

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

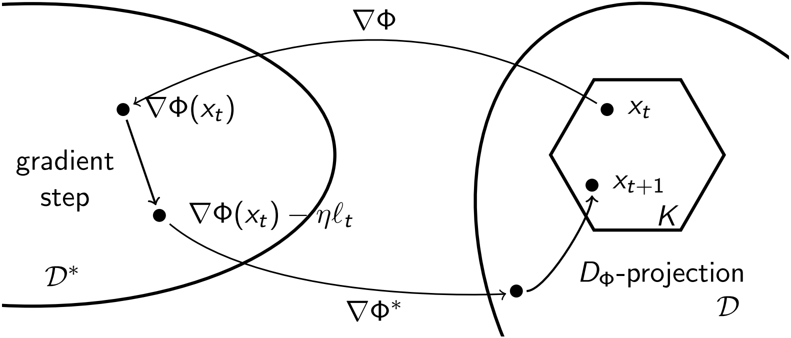
Mirror Descent (Nemirovski and Yudin 87)

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