Descente miroir séquentielle
Application aux bandits linéaires combinatoires

I/ Cadre : décisions séquentielles dans un compact convexe de $\mathbb{R}^K$

C compact convexe de $\mathbb{R}^K$

Jeu de prédiction :

- L'environnement choisit des f° convexes sous-diff $\ell_t : C \to \mathbb{R}$, $t \geq 1$.
- A chaque $t \geq 1$,
  - * le stat choisit $x_t \in C$ en fonction du passé et encourt la perte $\ell_t(x_t)$
  - * La fonction ℓ_t (ou une partie de cette info) est révélée au statisticien.

  ℓ_t : info parfaite ; une partie de cette info : $\ell_t(x_t)$: bandits

But : minimiser le regret

$$\sum_{t=1}^{T} \ell_t(x_t) - \inf_{x \in C} \sum_{t=1}^{T} \ell_t(x)$$

Applications :

- pb usuel : choix randomisé d'une action parmi K actions : $(e_1, \dots, e_K)$ base canonique de $\mathbb{R}^K$

  $C = \mathrm{conv}(\{e_1, \dots, e_K\}) = K$-simplexe $=$ ensemble des probas sur les K actions.

  $\ell_t(q) = \langle \underline{\ell}_t, q \rangle = \sum_{i=1}^{K} q_i \ell_{i,t} = \mathbb{E}_{I_t \sim q}[\ell_{I_t,t}]$

  convexification $\leftrightarrow$ randomisation

- randomisation sur une partie $\mathcal{C} \subset \mathbb{R}^K$ de nature combinatoire

  Ex : $\mathcal{C} = \left\{ x \in \{0,1\}^K : \sum_{i=1}^{K} x_i = m \right\}$ où $m \in \mathbb{N}^*$, puis $C = \mathrm{conv}(\mathcal{C})$.

  → choix simultané de m bras (multiple plays : Uchiya et al. '10)
  Par ex : choix de pubs sur une bannière web.

  Fonction de gain : $g_t(x) = \langle \underline{g}_t, x \rangle = \sum_{i=1}^{K} x_i g_{i,t}$

  de sorte que : $\forall x \in \mathcal{C}$, $g_t(x) = \sum_{i \in S_x} g_{i,t}$, où $S_x = \{i : x_i = 1\}$.

  $=$ nb de clics total associé au choix de pubs $S_x \subset \{1, \dots, K\}$ si on considère $g_{i,t} = \mathbb{1}_{\{\text{l'internaute a cliqué sur la pub } i \text{ à l'instant } t\}}$.

  [Ici le but est de rendre le nb total cumulé de clics $\sum_{t=1}^{T} \sum_{i \in \hat{S}_t} g_{i,t}$ le plus grand possible, et notamment proche de $\max_{x \in \mathcal{C}} \sum_{t=1}^{T} \sum_{i \in S_x} g_{i,t}$.]

  → Autre appli : choix d'un chemin de longueur m de A vers B ds le graphe :

  A  B

  Ici, $\mathcal{C} \subset \left\{ x \in \{0,1\}^K : \sum_{i=1}^{K} x_i = m \right\}$ et

  $\forall x \in \mathcal{C}$, $\ell_t(x) = \langle \underline{\ell}_t, x \rangle =$ tps de trajet total du chemin x si $\ell_{i,t} =$ tps de trajet de l'arête i

II - Descente miroir séquentielle en information parfaite

Refs : Nemirovski '79, Nemirovski & Yudin '83, Beck & Teboulle '03 ...
- Online : Herbster & Warmuth '98 ...
- Stats : Juditsky et al. '05 ...

<u>Déf</u> : Soit $\mathcal{D} \subset \mathbb{R}^K$ ouvert convexe et $F : \overline{\mathcal{D}} \to \mathbb{R}$.

F est dite "de Legendre" ssi
- F strictement convexe sur $\mathcal{D}$
- $\lim_{n \to +\infty} \|\nabla f(x_n)\| = +\infty$ pour tout $x_n \xrightarrow[n \to +\infty]{} \partial \mathcal{D}$.

<u>Algo</u> : <u>descente miroir séquentielle</u>

- Paramètres : C compact convexe de $\mathbb{R}^K$
  $F : \overline{\mathcal{D}} \to \mathbb{R}$ Legendre, avec $\mathcal{D}$ ouvert convexe tq $\overline{\mathcal{D}} \supset C$ et $C \cap \mathcal{D} \neq \emptyset$
  $\eta > 0$ (paramètre d'apprentissage)
- Initialisation : $x_1 \in C \cap \mathcal{D}$. (typiquement : $x_1 \in \operatorname{argmin}_{x \in C} F(x)$)
- A chaque date $t \geq 1$,
  (1) Jouer x_t et observer $\ell_t : C \to \mathbb{R}$.
  (2) $w_{t+1} = \nabla(F^*)\big(\nabla F(x_t) - \eta \nabla \ell_t(x_t)\big)$ ← descente du gradient sur $\mathcal{D}^*$ ramenée sur $\mathcal{D}$ via ∇F^*
  (3) $x_{t+1} = \operatorname{argmin}_{y \in C} D_F(y, w_{t+1})$ (<u>N.B.</u> : $x_{t+1} \in C \cap \mathcal{D}$ car F Legendre)
  projection de Bregman sur C

où
- $F^*(y) = \sup_{x \in \overline{\mathcal{D}}} \{ \langle x, y \rangle - F(x) \}$ (transformée de Fenchel-Legendre)
- $\mathcal{D}^* = \nabla F(\mathcal{D})$ (<u>NB</u> : sur $\mathcal{D}^*$, $\nabla(F^*) = (\nabla F)^{-1}$, cf [Rockafellar])
- $D_F(y, x) = F(y) - F(x) - \langle \nabla F(x), y - x \rangle$ pour tout $(y, x) \in \overline{\mathcal{D}} \times \mathcal{D}$.
  divergence de Bregman
  ≥ 0 et strictement convexe. [cf illustration page ④]

<u>Rem</u> :

- L'étape (2) n'est bien définie que si $\nabla F(x_t) - \eta \nabla \ell_t(x_t) \in \mathcal{D}^*$ (<u>hyp</u>)
- L'étape (3) est une projection de $w_{t+1} \in \mathcal{D}$ sur $C \subset \overline{\mathcal{D}}$.
- <u>Exemples</u> :
  * $\mathcal{D} = \mathbb{R}^K$, $F(x) = \frac{1}{2}\|x\|_2^2$ ⇝ $\nabla F = id = \nabla F^*$
    $D_F(y, x) = \frac{1}{2}\|y - x\|_2^2$
    étape (2) : descente de gradient usuelle, étape (3) vide.
  * $\mathcal{D} =]0, +\infty[^K$, $F(x) = \sum_{i=1}^K x_i \ln x_i - \sum_{i=1}^K x_i$ (entropie négative)
    Alors $\nabla F(x) = (\ln x_i)_{1 \leq i \leq K}$ et $\nabla F^*(y) = (\nabla F)^{-1}(y) = (e^{y_i})_{1 \leq i \leq K}$

et $D_F(y,x) = \sum_{i=1}^{K} y_i \ln \frac{y_i}{x_i} - \sum_{i=1}^{K} (y_i - x_i)$

Dans le cas où $C = \Delta(K) := \{ q \in \mathbb{R}_+^K : \sum_{i=1}^{K} q_i = 1 \}$, on a : $\forall x \in \mathbb{R}_+^{*K}$,

$$\underset{y \in C}{\operatorname{argmin}}\, D_F(y,x) = \underset{y \in \Delta(K)}{\operatorname{argmin}} \left\{ \sum_{i=1}^{K} y_i \ln \frac{y_i}{x_i / \|x\|_1} \right\} = \underset{y \in \Delta(K)}{\operatorname{argmin}}\, KL\left(y, \frac{x}{\|x\|_1}\right) = \frac{x}{\|x\|_1}.$$

D'où : étape (2) : $w_{i,t+1} = w_{i,t}\, e^{-\eta \nabla_i l_t(x_t)}$

étape (3) : $x_{t+1} = \frac{w_{t+1}}{\|w_{t+1}\|_1}$.

On retrouve donc les poids exponentiels : $w_{i,t+1} = \frac{\exp\left(-\eta \sum_{s=1}^{t} \nabla_i l_s(x_s)\right)}{\sum_{j=1}^{K} \exp\left(-\eta \sum_{s=1}^{t} \cdots\right)}$

<u>Formulation proximale</u> : l'étape (2) est équivalente à

$$w_{t+1} \in \underset{w \in \mathcal{D}}{\operatorname{argmin}} \left\{ \underbrace{l_t(x_t) + \langle \nabla l_t(x_t), w - x_t \rangle}_{\text{approx } 1^{\text{er}} \text{ ordre de } l_t} + \frac{D_F(w, x_t)}{\eta} \right\}$$

<u>Théorème 1</u>: majoration du regret de l'algo de descente miroir séquentielle

Soit C un compact convexe de $\mathbb{R}^K$
$F : \bar{\mathcal{D}} \to \mathbb{R}$ Legendre, avec $\mathcal{D}$ ouvert convexe tq $\bar{\mathcal{D}} \supset C$ et $C \cap \mathcal{D} \neq \emptyset$.
$\eta > 0$.

Alors, l'algorithme de descente miroir séquentielle calibré avec $\eta > 0$ vérifie : $\forall x \in C$,

$$\sum_{t=1}^{T} l_t(x_t) - \sum_{t=1}^{T} l_t(x) \leq \frac{D_F(x, x_1)}{\eta} + \frac{1}{\eta} \sum_{t=1}^{T} D_{F^*}\left(\nabla F(x_t) - \eta \nabla l_t(x_t), \nabla F(x_t)\right)$$

Preuve : par convexité,

$$\sum_{t=1}^{T} \left(l_t(x_t) - l_t(x) \right) \leq \sum_{t=1}^{T} \langle \underbrace{\nabla l_t(x_t)}_{\text{(sous-gradient)}}, x_t - x \rangle$$

Or, d'après l'étape (2) et $\nabla F^* = (\nabla F)^{-1}$ sur $\mathcal{D}^*$, on a

$$\nabla F(w_{t+1}) = \nabla F(x_t) - \eta \nabla l_t(x_t)$$

d'où
$$\langle \eta \nabla l_t(x_t), x_t - x \rangle = \langle \nabla F(x_t) - \nabla F(w_{t+1}), x_t - x \rangle = D_F(x, x_t) + D_F(x_t, w_{t+1}) - D_F(x, w_{t+1}) \quad (*)$$

car $\forall (x,y,z) \in \bar{\mathcal{D}} \times \mathcal{D}^2$, $D_F(x,y) = D_F(x,z) + D_F(z,y) - \langle \nabla F(z) - \nabla F(y), z - x \rangle$
(appliquée avec $z = x_t$, $y = w_{t+1}$ et $x = x$)

Or $x_{t+1} \in \underset{y \in C}{\operatorname{argmin}} D_F(y, w_{t+1})$ = project° de Bregman sur C

donc (inégalité de Pythagore généralisée) :

$$\forall x \in C, \quad D_F(x, w_{t+1}) \geq D_F(x, x_{t+1}) + D_F(x_{t+1}, w_{t+1})$$

En injectant dans $(*)$, il vient :

$$\langle \eta \nabla \ell_t(x_t), x_t - x \rangle \leq D_F(x, x_t) + D_F(x_t, w_{t+1}) - D_F(x, x_{t+1}) - \underbrace{D_F(x_{t+1}, w_{t+1})}_{\geq 0}$$

télescopique

$$\text{D'où} \quad \sum_{t=1}^{T} \langle \eta \nabla \ell_t(x_t), x_t - x \rangle \leq \sum_{t=1}^{T} \left(D_F(x, x_t) - D_F(x, x_{t+1}) \right) + \sum_{t=1}^{T} D_F(x_t, w_{t+1})$$

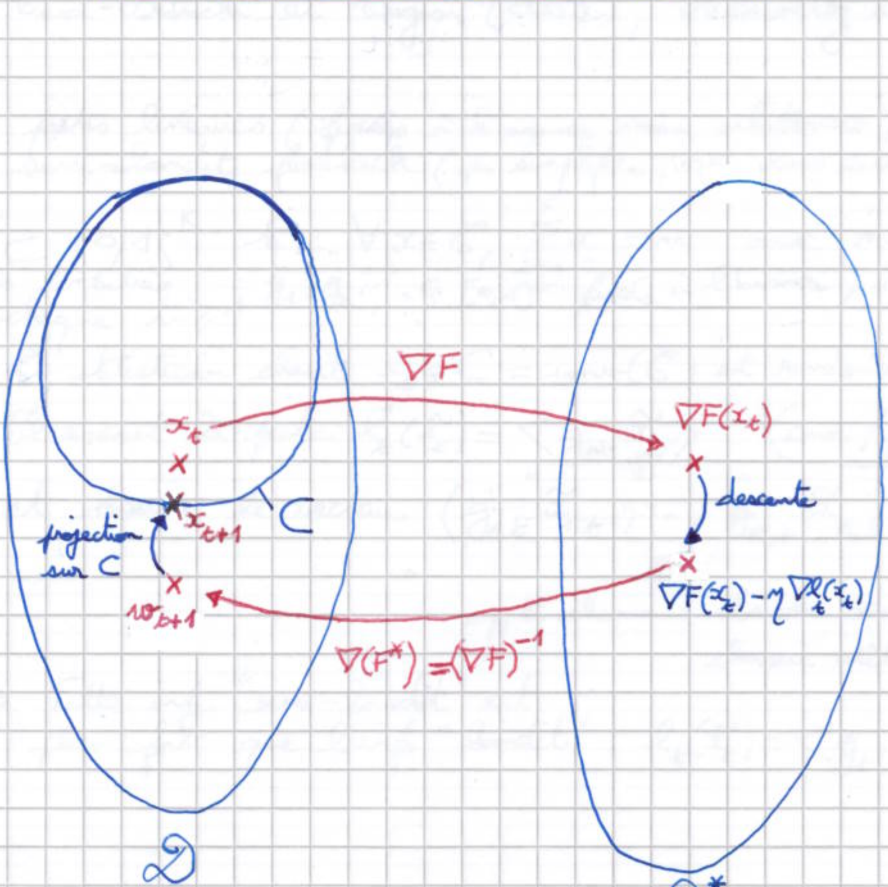
$$= D_F(x, x_1) - \underbrace{D_F(x, x_{T+1})}_{\geq 0} + \sum_{t=1}^{T} D_F(x_t, w_{t+1})$$

On conclut en remarquant que

$$D_F(x_t, w_{t+1}) = D_{F^*}\left(\nabla F(w_{t+1}), \nabla F(x_t)\right)$$

$$= D_{F^*}\left(\nabla F(x_t) - \eta \nabla \ell_t(x_t), \nabla F(x_t)\right) \text{ d'après étape (2)} \blacksquare$$

Illustration de la descente miroir séquentielle



III - Application aux bandits linéaires combinatoires

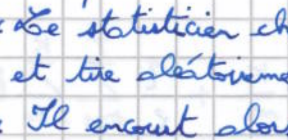
Réfs : Cesa-Bianchi & Lugosi (2009), Abernethy et al. (2009).

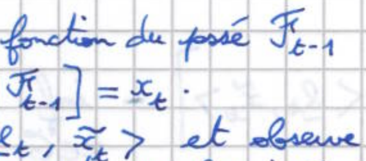
Jeu de prévision :

- Paramètre : $\mathcal{C} \subset \{0,1\}^K$ tq $\forall x \in \mathcal{C}$, $\sum_{i=1}^{K} x_i = m$ avec $m \in \mathbb{N}^*$ connu.
- Initialisation : l'environnement choisit à l'avance des vecteurs de pertes $\underline{l}_t \in [0,1]^K$ (on posera $l_t(x) = \langle \underline{l}_t, x \rangle$ pour tout $x \in \mathbb{R}^K$), $t \geq 1$.
- A chaque date $t \geq 1$,
  - * Le statisticien choisit $x_t \in C = \mathrm{conv}(\mathcal{C})$ en fonction du passé $\mathcal{F}_{t-1}$ et tire aléatoirement $\tilde{x}_t \in \mathcal{C}$ tel que $\mathbb{E}\left[\tilde{x}_t \mid \mathcal{F}_{t-1}\right] = x_t$.
  - * Il encourt alors la perte $l_t(\tilde{x}_t) = \langle \underline{l}_t, \tilde{x}_t \rangle$ et observe les valeurs $l_{i,t}$ pour tout i tq $\tilde{x}_{i,t} = 1$ (correspond aux bras joués à l'instant t); autrement dit, le statisticien observe le vecteur $(l_{i,t}\,\tilde{x}_{i,t})_{1 \leq i \leq K} \in [0,1]^K$.

Rem : l'observation du vecteur $(l_{i,t}\,\tilde{x}_{i,t})_{1 \leq i \leq K}$ est qualifiée d'information "semi-bandits". Cette inf est plus forte que la seule observation de $\sum_{i=1}^{K} l_{i,t}\,\tilde{x}_{i,t} = \langle \underline{l}_t, \tilde{x}_t \rangle$ qui correspond au cas des bandits linéaires.

Ex : SEMI-BANDITS versus BANDITS





A chaque date $t \geq 1$, on observe les temps de trajets $l_{i,t}$ des arêtes i parcourues ($\Leftrightarrow \tilde{x}_{i,t} = 1$)

Sur l'ex : durée $(A \to A_1)$, durée $(A_1 \to A_2)$ durée $(A_2 \to A_3)$ et durée $(A_3 \to B)$

A chaque date $t \geq 1$, on observe le temps de trajet total $\sum_{i : \tilde{x}_{i,t} = 1} l_{i,t}$ du chemin $\tilde{x}_t$ parcouru.

Sur l'ex : durée $(A \to A_1 \to A_2 \to A_3 \to B)$.

On ne traitera ici que le cas "semi-bandits" linéaires, mais le pb des bandits linéaires est également abordable.

Comment se ramener au cas de l'information parfaite ? $\rightsquigarrow$ On estime le gradient $\nabla l_t(x_t) = \underline{l}_t$.

On estime $l_{i,t}$ par
$$\tilde{l}_{i,t} = \frac{l_{i,t}\,\tilde{x}_{i,t}}{x_{i,t}} = \begin{cases} \dfrac{l_{i,t}}{x_{i,t}} & \text{si } \tilde{x}_{i,t} = 1 \text{ (bras } i \text{ joué à l'instant } t) \\ 0 & \text{sinon} \end{cases}$$

On vérifie immédiatement que
$$\mathbb{E}\Big[\tilde{l}_{i,t} \mid \underset{\sigma(\tilde{x}_1, \dots, \tilde{x}_{t-1})}{\mathcal{F}_{t-1}}\Big] = \frac{l_{i,t}}{x_{i,t}} \underbrace{\mathbb{E}\left[\tilde{x}_{i,t} \mid \mathcal{F}_{t-1}\right]}_{= x_{i,t}} = l_{i,t}$$
$\rightsquigarrow$ estimateur sans biais.

L'algorithme de descente miroir séquentielle vu en section II a été étudié pour toute suite $\ell_1, \ell_2, \ell_3, \dots$ de fonctions de pertes convexes et sous-différentiables.

Nous allons donc pouvoir appliquer trajectoriellement (pour tout tirage $\omega \in \Omega$) l'algo de descente miroir séquentielle avec les fonctions de pertes

$$\langle \underline{\tilde{\ell}}_1, \cdot \rangle, \ \langle \underline{\tilde{\ell}}_2, \cdot \rangle, \ \langle \underline{\tilde{\ell}}_3, \cdot \rangle, \ \dots$$

Théorème 2: l'algorithme de descente miroir séquentielle appliqué aux fonctions de pertes $\langle \underline{\tilde{\ell}}_t, \cdot \rangle$, $t \geq 1$, avec $\mathcal{C} = \left\{ x \in \{0,1\}^K : \sum_{i=1}^K x_i = m \right\}$, $x_1 = \left(\frac{m}{K}, \dots, \frac{m}{K}\right)$ et $F : \mathbb{R}_+^K \to \mathbb{R}$, $x \mapsto \sum_{i=1}^K x_i \ln x_i - \sum_{i=1}^K x_i$ vérifie :

$$\mathbb{E}\left[\sum_{t=1}^T \langle \underline{\ell}_t, \tilde{x}_t \rangle\right] - \inf_{x \in \mathcal{C}} \sum_{t=1}^T \langle \underline{\ell}_t, x \rangle \leq \frac{m}{\eta} \ln \frac{K}{m} + \frac{\eta}{2} \sum_{t=1}^T \sum_{i=1}^K \mathbb{E}\left[x_{i,t} \tilde{\ell}_{i,t}^2\right]$$
$$\leq \sqrt{2TKm \ln \frac{K}{m}}$$

où la 2ème inégalité est vraie avec $\eta = \sqrt{\frac{2m}{TK} \ln \frac{K}{m}}$.

Rem: on retrouve la borne en $\sqrt{TK \ln K}$ des bandits usuels (correspond à $m = 1$ et $\mathcal{C} = \{e_1, \dots, e_K\}$ = sommets du simplexe).

Le terme de complexité $m \ln \frac{K}{m}$ correspond à $\ln \binom{K}{m} \leq \ln\left[\left(\frac{eK}{m}\right)^m\right] \leq m \ln\left(\frac{eK}{m}\right)$.

En fait, le terme $\sqrt{\ln \frac{K}{m}}$ n'est pas nécessaire, car d'autres choix de la fonction F conduisent à la borne $\sqrt{TKm}$; cf. le survey de Bubeck et Cesa-Bianchi (p. 79-81).

Preuve: D'après le thm 1 et la remarque qui précède le thm 2, on a : $\forall x \in \mathcal{C}$,

$$(*) \quad \text{p.s.}, \ \sum_{t=1}^T \langle \underline{\tilde{\ell}}_t, x_t \rangle \leq \sum_{t=1}^T \langle \underline{\tilde{\ell}}_t, x \rangle + \frac{D_F(x, x_1)}{\eta} + \frac{1}{\eta} \sum_{t=1}^T D_{F^*}\left(\nabla F(x_t) - \eta \underline{\tilde{\ell}}_t, \nabla F(x_t)\right)$$

Majorons les deux derniers termes. Après quelques manipulations élémentaires sur $F(x) = \sum_{i=1}^K x_i (\ln x_i - 1)$, on obtient : $\nabla F(x) = (\ln x_i)_{1 \leq i \leq K}$, $\nabla(F^*)(y) = (\nabla F)^{-1}(y) = (e^{y_i})_{1 \leq i \leq K}$

$$\forall y, z \in \mathbb{R}^K, \quad F^*(y) = \sum_{i=1}^K e^{y_i} \text{ et } D_{F^*}(z, y) = \sum_{i=1}^K (e^{z_i} - e^{y_i}) - \sum_{i=1}^K e^{y_i}(z_i - y_i)$$

D'où :

$$\bullet \ D_F(x, x_1) = \sum_{i=1}^K x_i \ln \frac{x_i}{x_{i,1}} - \underbrace{\sum_{i=1}^K (x_i - \overset{= \frac{m}{K}}{x_{i,1}})}_{= m - m = 0} = \sum_{i=1}^K x_i \big(\underbrace{\ln x_i}_{\leq 0} + \ln \frac{K}{m}\big) \underset{\sum_{i=1}^K x_i = m}{\leq} m \ln \frac{K}{m}$$

• $D_{F^*}\left(\nabla F(x_t) - \eta\, \tilde{\underline{\ell}}_t, \nabla F(x_t)\right) = D_{F^*}\left(\overrightarrow{\ln}\, x_t - \eta \tilde{\underline{\ell}}_t, \overrightarrow{\ln}\, x_t\right)$

$$= \sum_{i=1}^K \left(x_{i,t} e^{-\eta \tilde{\ell}_{i,t}} - x_{i,t}\right) - \sum_{i=1}^K x_{i,t}\left(-\eta \tilde{\ell}_{i,t}\right)$$

$$= \sum_{i=1}^K x_{i,t} \underbrace{\left(e^{-\eta \tilde{\ell}_{i,t}} - (-\eta \tilde{\ell}_{i,t}) - 1\right)}_{\leq \frac{\eta^2 \tilde{\ell}_{i,t}^2}{2} \text{ car } e^x - x - 1 \leq \frac{x^2}{2}}$$

pour tout $x \leq 0$.

En injectant les 2 majorations ↑ dans (*), puis en prenant l'espérance, il vient : $\forall x \in C$,

$$\sum_{t=1}^T \mathbb{E}\left[\langle \tilde{\underline{\ell}}_t, x_t \rangle\right] \leq \sum_{t=1}^T \langle \mathbb{E}[\tilde{\underline{\ell}}_t], x \rangle + \frac{m}{\eta} \ln \frac{K}{m} + \frac{\eta}{2} \sum_{t=1}^T \sum_{i=1}^K \mathbb{E}\left[x_{i,t} \tilde{\ell}_{i,t}^2\right].$$

Pour conclure, il suffit de remarquer que

• $\mathbb{E}\left[\langle \tilde{\underline{\ell}}_t, \underset{\mathcal{F}_{t-1}\text{-mes}}{x_t} \rangle \mid \mathcal{F}_{t-1}\right] = \langle \mathbb{E}[\tilde{\underline{\ell}}_t \mid \mathcal{F}_{t-1}], x_t \rangle = \langle \underline{\ell}_t, x_t \rangle = \langle \underline{\ell}_t, \mathbb{E}[\tilde{x}_t \mid \mathcal{F}_{t-1}] \rangle$

$$= \mathbb{E}\left[\langle \underline{\ell}_t, \tilde{x}_t \rangle \mid \mathcal{F}_{t-1}\right]$$

• $\mathbb{E}[\tilde{\underline{\ell}}_t] = \mathbb{E}\left[\mathbb{E}[\tilde{\underline{\ell}}_t \mid \mathcal{F}_{t-1}]\right] = \underline{\ell}_t$. D'où la 1ère inégalité du thm.

Or, $\mathbb{E}\left[x_{i,t} \tilde{\ell}_{i,t}^2 \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[x_{i,t} \frac{\ell_{i,t}^2 \tilde{x}_{i,t}^2}{x_{i,t}^2} \mid \mathcal{F}_{t-1}\right] \overset{\tilde{x}_{i,t} \leq 1}{\leq} \frac{\ell_{i,t}^2}{x_{i,t}} \mathbb{E}\left[\tilde{x}_{i,t} \mid \mathcal{F}_{t-1}\right]$

$$= \ell_{i,t}^2 \leq 1.$$

D'où :

$$\mathbb{E}\left[\sum_{t=1}^T \langle \underline{\ell}_t, \tilde{x}_t \rangle\right] - \inf_{x \in C} \sum_{t=1}^T \langle \underline{\ell}_t, x \rangle \leq \frac{m}{\eta} \ln \frac{K}{m} + \frac{\eta T K}{2}$$

$$= \sqrt{2TKm \ln \frac{K}{m}}$$

en prenant $\eta = \sqrt{\frac{2m}{TK} \ln \frac{K}{m}}$ ∎

<u>Rem</u> : la même étude est valable si les vecteurs de perte $\underline{\ell}_t$ sont choisis par un adversaire antagoniste (i.e. en fonction du passé $\mathcal{F}_{t-1} = \sigma(\tilde{x}_1, \dots, \tilde{x}_{t-1})$), auquel cas les $\underline{\ell}_t$ sont aléatoires (il faut ajouter une espérance autour de $\langle \underline{\ell}_t, x \rangle$ ds le thm 2).

<u>Efficacité algorithmique</u> ?

• Même si $C = \text{conv}(\mathcal{C}) = \left\{ q \in [0,1]^K : \sum_{i=1}^K q_i = m \right\}$ n'est pas un K-simplexe, on peut aisément vérifier que l'algo de descente miroir séquentielle avec la fonction entropie $F(x) = \sum_{i=1}^{K} x_i \ln x_i - \sum_{i=1}^{K} x_i$ donne lieu aux poids exponentiels :

$$x_{i,t} = m \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \tilde{\ell}_{i,s}\right)}{\sum_{j=1}^{K} \exp\left(-\eta \sum_{s=1}^{t-1} \tilde{\ell}_{j,s}\right)} \quad , \quad 1 \leq i \leq K, \quad t \geq 1.$$

ou encore : $$x_{i,t} = m \frac{x_{i,t-1} e^{-\eta \tilde{\ell}_{i,t-1}}}{\sum_{j=1}^{K} x_{j,t-1} e^{-\eta \tilde{\ell}_{j,t-1}}},$$

donc le calcul itératif des $x_t \in C$ ne requiert que $O(K)$ opérations élémentaires à chaque itération t.

- Il existe par ailleurs un moyen efficace de tirer aléatoirement $\tilde{x}_t \in \mathcal{C}$ pour que $\mathbb{E}[\tilde{x}_t \mid \mathcal{F}_{t-1}] = x_t$.

(<u>Première idée naïve</u> : puisque $x_t \in C = \mathrm{conv}(\mathcal{C})$, il existe une proba $q = (q_x)_{x \in \mathcal{C}}$ sur $\mathcal{C}$ telle que $x_t = \sum_{x \in \mathcal{C}} q_x x$. On tire alors $\tilde{x}_t \sim q$. → mauvaise complexité algo car $\mathcal{C}$ est gros !)

<u>Implémentation efficace du tirage de $\tilde{x}_t$</u> : [cf Uchiya et al '10 et Gandhi et al '06]

- On part de $q_0 = x_t \in C = \mathrm{conv}(\mathcal{C})$.
- On construit itérativement des vecteurs $q_1, q_2, \ldots, q_{k_{final}} \in C$ tels que $\mathbb{E}[q_k \mid q_0, \ldots, q_{k-1}] = q_{k-1}$ et $q_{k_{final}} \in \mathcal{C}$ p.s.. <u>Méthode</u> :

(a) $q := q_0$; $k = 0$

(b) TANT QUE $\exists i \in \{1, \ldots, K\} : 0 < q_i < 1$ FAIRE

- Choisir $i \neq j$ tq $0 < q_i < 1$ et $0 < q_j < 1$ (existe car $\sum_{\ell=1}^{K} q_\ell = m \in \mathbb{N}$)
- Poser $\alpha = \min\{1 - q_i, q_j\}$ et $\beta = \min\{q_i, 1 - q_j\}$.
- Mettre à jour q_i et q_j de la façon suivante :

$$(q_i, q_j) \longleftarrow \begin{cases} (q_i + \alpha, q_j - \alpha) & \text{avec proba } \frac{\beta}{\alpha + \beta} \\ (q_i - \beta, q_j + \beta) & \text{avec proba } \frac{\alpha}{\alpha + \beta} \end{cases}$$

(• Poser $k \leftarrow k+1$ et $q_k \leftarrow q$)

FIN TANT QUE

(c) $k_{final} \leftarrow k$.

- On renvoit $\tilde{x}_t = q_{k_{final}} \in \mathcal{C}$. Cette randomisation est efficace d'un point de vue algo car le nombre d'itérations k_{final} vérifie $k_{final} \leq K$ (puisque l'une des coordonnées q_i ou q_j devient 0 ou 1 à chaque itération).

BILAN : à chaque date $t \geq 1$, l'obtention de x_t et $\tilde{x}_t$ ne requiert que $O(K)$ opérations élémentaires.

Extensions du problème (cf survey de Bubeck & Cesa-Bianchi)

- information de type "bandits" au lieu de "semi-bandits"
  → algo "John's exploration" par ex
- autres compacts convexes $C \subset \mathbb{R}^K$ (ex: boule euclidienne)
- fonctions de pertes non-linéaires
- etc