

# Optimisation convexe en ligne

Regrets logarithmiques

Samuel Givois - Fatou Sall

Apprentissage en ligne  
Ensaë

Vendredi 6 avril 2018

1. Problème

2. Hypothèses

3. OGD et EW00

4. « Online Newton Step »

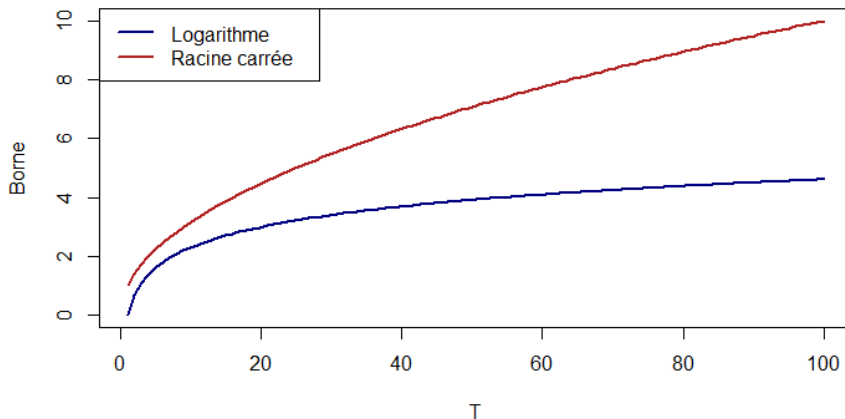
5. Implémentation

# Problème

Hazan, E., Kalai, A., Kale, S., Agarwal, A. (2006, June). Logarithmic regret algorithms for online convex optimization. *International Conference on Computational Learning Theory* (pp. 499-513). Springer, Berlin, Heidelberg.

- ▶ Problème d'optimisation convexe en ligne.
- ▶ Cadre introduit par Zinkevich en 2003.
  - ▶ Descente de gradient en ligne en  $O(\sqrt{T})$
- ▶ Algorithmes en  $O(\log(T))$  sous certaines hypothèses

## Bornes logarithmique et en racine carrée



- ▶ Ensemble convexe  $\mathbf{K} \in \mathbb{R}^n$
- ▶  $T$  périodes
- ▶ Décisions  $\mathbf{x}_1, \dots, \mathbf{x}_T \in \mathbf{K}$
- ▶ Fonctions de perte convexes  $f_1, \dots, f_T : \mathbf{K} \rightarrow \mathbb{R}$

$$\text{Regret} : \mathbf{R}_T = \sum_{t=1}^T f_t(\mathbf{x}_t) - \min_{\mathbf{x} \in \mathbf{K}} \sum_{t=1}^T f_t(\mathbf{x})$$

On veut garantir  $\mathbf{R}_T$  faible

# Hypothèses

Hypothèse forte :

- ▶  $G \geq 0 : \exists G \geq 0$  tq  $\forall t \leq T, \forall x \in K, \|\nabla f_t(x)\| \leq G$
- ▶  $H \geq 0 : \exists H \geq 0$  tq  $\forall t \leq T, \forall x \in K, \nabla^2 f_t(x) \geq H I_n$

Hypothèse faible :

$\exists \alpha \geq 0$  tq  $\forall t \leq T, h_t : x \mapsto e^{-\alpha f_t(x)}$  est concave sur  $K$

La première implique la seconde.



Démonstration en dimension 1 :

$$h_t''(x) = ((\alpha f_t'(x))^2 - \alpha f_t''(x)) e^{-\alpha f_t(x)} \leq 0$$

$$\iff$$

$$\alpha \leq \frac{f_t''(x)}{(f_t'(x))^2}$$

A remplir

# OGD et EWOO

# Descente de gradient en ligne (OGD)

---



Descente de gradient dans le cadre « en ligne ».

Pas :  $\eta_1, \dots, \eta_T$ .

Mise à jour :

$$\mathbf{x}_{t+1} = \Pi_K(\mathbf{x}_t - \eta_{t+1} \nabla f_t(\mathbf{x}_t))$$

Projection :

$$\Pi_K(\mathbf{y}) = \operatorname{argmin}_{\mathbf{x} \in K} \|\mathbf{x} - \mathbf{y}\|$$

## Théorème

Supposons que  $\exists G, H \geq 0$  tel que  $\forall t \leq T, \forall x \in K$ ,  
 $\|\nabla f_t(x)\| \leq G$  et  $\nabla^2 f_t(x) \geq H I_n$ .

Alors, en prenant  $\forall t \leq T, \eta_t = \frac{1}{Ht}$ , on a :

$$R_T \leq \frac{G^2}{2H} (1 + \log(T))$$

1. Inégalité de Taylor-Lagrange :

$$f_t(\mathbf{x}_t) - f_t(\mathbf{x}^*) \leq \nabla_t^T(\mathbf{x}_t - \mathbf{x}^*) - \frac{H}{2} \|\mathbf{x}^* - \mathbf{x}_t\|^2$$

2. Propriété des projections sur un ensemble convexe :  
borne pour  $\nabla_t^T(\mathbf{x}_t - \mathbf{x}^*)$

3. Simplification grâce au choix des  $\eta_t$

Généralisation de l'algorithme **EW**A au cas continu.

Poids exponentiels :

$$w_{t+1}(x) = e^{-\alpha \sum_{\tau=1}^t f_{\tau}(x)} = \prod_{\tau=1}^t h_{\tau}(x)$$

Mise à jour :

$$x_{t+1} = \frac{\int_K x w_{t+1}(x) dx}{\int_K w_{t+1}(x) dx}$$

## Théorème

Supposons que  $\exists \alpha$  tel que  $\forall t \leq T$ ,  $h_t : x \mapsto e^{-\alpha f_t(x)}$  est concave sur  $\mathbf{K}$ .

Alors, on a :

$$R_T \leq \frac{n}{\alpha} (1 + \log(1 + T))$$



1. Concavité de  $h_t$  et inégalité de Jensen, puis produit

téléscopique : 
$$\prod_{\tau=1}^t h_{\tau}(x_{\tau}) \geq \frac{\int_K \prod_{\tau=1}^t h_{\tau}(x) dx}{\text{vol}(K)}$$

2. Borne sur ensemble  $S$  construit à partir de  $x^*$  et  $K$  (concavité et positivité de  $h_t$ ).
3. Volume de  $S$  à partir de celui de  $K$  (mise à l'échelle en dimension  $n$ ).

# « Online Newton Step »

# Changement de fonction de perte



$\mathbf{x}_1, \dots, \mathbf{x}_{T+1}$  : décisions « follow the leader ». On a :

$$R_T = \sum_{t=1}^T f_t(\mathbf{x}_t) - \min_{\mathbf{x} \in K} \sum_{t=1}^T f_t(\mathbf{x}) \leq \sum_{t=1}^T (f_t(\mathbf{x}_t) - f_t(\mathbf{x}_{t+1}))$$

Soit  $\tilde{f}_t$  tel que :  $\forall t, \tilde{f}_t(\mathbf{x}_t) = f_t(\mathbf{x}_t)$  et  $\forall \mathbf{x}, \tilde{f}_t(\mathbf{x}) \leq f(\mathbf{x}_t)$ .

Alors,

$$R_T \leq \tilde{R}_T \leq \sum_{t=1}^T (\tilde{f}_t(\tilde{\mathbf{x}}_t) - \tilde{f}_t(\tilde{\mathbf{x}}_{t+1}))$$

## Paramètre $\beta$

- ▶  $\mathbf{x}_1$  arbitraire
- ▶ Pour  $t \geq 2$ ,  $\mathbf{x}_t = \operatorname{argmin}_{\mathbf{x} \in K} \sum_{\tau=1}^{t-1} \tilde{f}_t(\mathbf{x})$
- ▶  $\tilde{f}_t(\mathbf{x}) =$   
$$f_t(\mathbf{x}_t) + \nabla_t^T (\mathbf{x} - \mathbf{x}_t) + \frac{\beta}{2} (\mathbf{x} - \mathbf{x}_t)^T \nabla_t \nabla_t^T (\mathbf{x} - \mathbf{x}_t)$$

Soit  $f : \mathbf{K} \mapsto \mathbb{R}$  tel que  $e^{-\alpha f(\mathbf{x})}$  est concave, de gradient borné  $\|\nabla f\| \leq G$ , et  $\beta = \frac{1}{2} \min(\frac{1}{4GD}, \alpha)$ .  $\forall \mathbf{x}, \mathbf{y} \in \mathbf{K}$  :

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}) + \frac{\beta}{2} (\mathbf{x} - \mathbf{y})^\top \nabla f(\mathbf{y}) \nabla f(\mathbf{y})^\top (\mathbf{x} - \mathbf{y})$$

Sous les mêmes hypothèses :

$$\sum_{t=1}^T (\tilde{f}_t(\tilde{\mathbf{x}}_t) - \tilde{f}_t(\tilde{\mathbf{x}}_{t+1})) \leq 3\left(\frac{1}{\alpha} + 4GD\right)n \log(T)$$

« Online Newton Step » : application de l'algorithme de Newton-Raphson au cas en ligne.  
Equivalence entre ONS et FTAL.

# Implémentation

- ▶ Application au problème de la sélection de portefeuille.
- ▶ Jeu de données : indices boursiers de 490 actions S&P entre 2001 et 2005 (source : [ici](#))
- ▶ Cotation sur 1000 jours



- ▶ A chaque période, choix d'un portefeuille  $\mathbf{x}_t$ .
- ▶ Hypothèse :  $\mathbf{x}_t^T \mathbf{u} = 1$ , donc  $\mathbf{K}$  est le simplexe  
 $\mathbf{K} = \{\mathbf{y} \in \mathbb{R}^n | \mathbf{y}^T \mathbf{u} = 1 \text{ et } \forall k, y_k \geq 0\}$
- ▶ A chaque période, prix de marché  $\mathbf{c}_t = \frac{a_t}{a_{t-1}}$
- ▶ Fonction de perte logarithmique :  $f_t(\mathbf{x}_t) = -\log(\mathbf{x}_t^T \mathbf{c}_t)$

- ▶ Gradient :

$$\nabla f_t(\mathbf{x}_t) = \frac{-1}{\mathbf{x}_t^T \mathbf{c}_t} \mathbf{c}_t$$

- ▶ Hessienne :

$$\nabla^2 f_t(\mathbf{x}_t) = \frac{1}{(\mathbf{x}_t^T \mathbf{c}_t)^2} (\mathbf{c}_{t,i} \mathbf{c}_{t,j})_{i,j}$$