Un regret unifié pour l'optimisation convexe en ligne

Pierre Gaillard [1] Nicolò Cesa-Bianchi [2] Gàbor Lugosi [3] Gilles Stoltz [4]

[1] *École Normale Supérieure, 45 rue d'Ulm, 75005 Paris*
EDF R&D, 1 av du Général de Gaulle, 92140 Clamart
pierre.gaillard@ens.fr
[2] *Università degli studi di Milano*
nicolo.cesa-bianchi@unimi.it
[3] *Universitat Pompeu Fabra, Barcelone*
gabor.lugosi@gmail.com
[4] *École Normale Supérieure, CNRS, 45 rue d'Ulm, 75005 Paris*
gilles.stoltz@ens.fr

Résumé. Le problème d'optimisation convexe en ligne consiste à prévoir séquentiellement les valeurs d'une certaine suite à valeur dans un ensemble convexe. L'objectif du joueur est de s'assurer une performance similaire à la meilleure stratégie (oracle) d'un ensemble de stratégies de référence. Plus l'ensemble de référence est grand, plus l'erreur de l'oracle est potentiellement faible, mais plus le joueur a du mal à s'en rapprocher et plus son *regret* sera grand. C'est le compromis entre erreur d'estimation et erreur d'approximation, que l'on retrouve régulièrement en statistiques. Dans la littérature, l'objectif initial n'est pas toujours la minimisation de la perte cumulée, de plus le compromis biais-variance est géré de façons variées et différents ensembles de références sont considérés. Cela a mené à diverses notions de regret, comme le *regret en ruptures*, le *regret adaptatif*, ou le *regret escompté*. Cet exposé propose une notion de regret plus puissante, qui tend à unifier les trois précédentes. Nous en profiterons pour montrer que des algorithmes comme celui d'Herbster et Warmuth (1998) ou de Zinkevich (2003) permettent d'obtenir des bornes satisfaisantes sur ce regret généralisé.

Mots-clés. Apprentissage séquentiel, suites individuelles

Abstract. Online convex optimization consists in forecasting, in a sequential fashion, the values of an unknown sequence of a convex set. The player's goal is to ensure in worst case a performance slightly worse than that of the best strategy (the oracle) of a set of reference strategies. The bigger the reference set, the lower is the oracle's error, but the harder it is for the player to get closer and the greater the regret will be. This is the famous estimation-approximation trade-off. In the literature, the initial target is not always the minimization of the cumulative loss, the bias-variance tradeoff is furthermore managed in various ways and different sets of references are considered. This has led to various notions of regret, such as *shifting regret*, *adaptive regret*, or *discounted regret*. This

talk suggests a new notion of regret more powerful, which tends to unify the previous three. We additionally take the opportunity to show that algorithms such as *fixed-share* or *Mirror-descent* obtain satisfactory bounds on this generalized regret.

Keywords. On-line learning, individual sequences

1 Préliminaires

1.1 Cadre et présentation du problème

Le cadre. L'optimisation convexe en ligne correspond au problème de prévision séquentielle, dans lequel à chaque instant le joueur choisit un élément $\widehat{x}_t$ d'un ensemble convexe S, avant d'avoir accès à une fonction de perte ℓ_t sur S, qui définit sa perte $\ell_t(\widehat{x}_t)$. De nombreux problèmes comme la prévision à l'aide d'experts, l'investissement séquentiel, ou la régression/classification en ligne peuvent être vus comme des cas particuliers de ce cadre général.

L'objectif. Le joueur peut se fixer différents objectifs. Le plus courant consiste à minimiser sa perte cumulée $\sum_{t=1}^{T} \ell_t(\widehat{x}_t)$. Il peut aussi être plus ambitieux et tenter d'assurer une faible perte cumulée pour tout sous intervalle de temps de l'expérience $\sum_{t=r}^{s} \ell_t(\widehat{x}_t)$, où $1 \leqslant r < s \leqslant T$. Enfin, on peut imaginer que certains instants sont plus ou moins importants et tenter d'assurer une faible perte cumulée $\sum_{t=1}^{T} \gamma_t \, \ell_t(\widehat{x}_t)$, où l'escompte $\gamma_t \geqslant 0$ mesure l'importance relative des pertes récentes comparativement aux pertes plus anciennes.

Le regret. Afin d'atteindre ces différents objectifs, le joueur considère un ensemble de stratégies de référence $\mathcal{A}$ et suppose que la meilleure stratégie de cet ensemble, l'*oracle*, atteint une performance convenable. Il tente alors de s'approcher de cette performance. Suivant l'objectif fixé, nous pouvons décomposer la perte du joueur selon le compromis biais-variance. Par exemple, si l'objectif est une faible perte cumulée, nous avons

$$\underbrace{\sum_{t=1}^{T} \ell_t(\widehat{x}_t)}_{\text{Perte du joueur}} = \underbrace{\inf_{a \in \mathcal{A}} \sum_{t=1}^{T} \ell_t(a_t)}_{\text{Erreur d'approximation}} + \underbrace{R_T}_{\text{Erreur d'estimation}}$$

où $a_t \in S$ correspond au choix de la stratégie $a \in \mathcal{A}$ à l'instant t. Le joueur tente alors d'assurer dans le pire des cas un regret R_T aussi faible que possible. L'objectif minimal étant un regret sous-linéaire

$$\sup_{\ell_1, \ldots, \ell_T} R_T = o(T) \, .$$

Un ensemble de stratégies de référence habituel est l'ensemble des stratégies constantes, ou par abus de notations $\mathcal{A} = S$, $a \in S$, et $a_t = a$ à chaque instant. Un autre ensemble, plus grand, et donc bien plus dur à approcher, consiste en toutes les suites arbitraires $(a_t) \in S^T$. Les bornes de regret sont alors exprimées en fonction de la régularité des suites (a_t).

Dans la littérature, les différents objectifs initiaux, ainsi que les différents ensembles de références ont mené à de nombreuses notions de regrets, comme le *regret en ruptures*, le *regret adaptatif* ou le *regret escompté*. Nous proposons ici un regret généralisé qui permet une analyse unifiée.

Simplification. Afin de simplifier l'exposé, sans perte de généralité, nous dérivons nos résultats d'un cadre restreint. Nous supposons tout d'abord que l'ensemble convexe S est le simplexe Δ_d de dimension d. De plus, nous supposons que l'environnement choisit des fonctions de pertes linéaires $\boldsymbol{\ell}_t \in [0,1]^d$. Le cadre simplifié est donc le jeu répété présenté Algorithme 1 entre le joueur et l'environnement.

Pour chaque instant $t = 1, \dots, T$,

1. Le joueur choisit $\widehat{\boldsymbol{p}}_t = (\widehat{p}_{1,t}, \dots, \widehat{p}_{d,t}) \in \Delta_d$
2. L'environnement choisit un vecteur de perte $\boldsymbol{\ell}_t = (\ell_{1,t}, \dots, \ell_{d,t}) \in [0,1]^d$
3. Le joueur subit la perte $\widehat{\boldsymbol{p}}_t^\top \boldsymbol{\ell}_t$

Algorithme 1: Optimisation linéaire en ligne dans le simplexe.

Il s'agit maintenant du problème classique de la prévision séquentielle avec avis d'experts. Nous allons maintenant présenter un algorithme permettant d'obtenir des bornes de regret sous-linéaires.

1.2 L'algorithme

L'Algorithme 2, ou algorithme fixed-share, a été introduit par Herbster et Warmuth (1998) pour la prévision avec avis d'experts. Par souci de simplicité, nous n'énoncerons que les résultats obtenus par cet algorithme. Nos résultats sont cependant plus généraux et restent valables, à de légères modifications près, pour des algorithmes tels que la descente miroir de Zinkevich (2003) ou l'algorithme de Bousquet et Warmuth (2001). Ces algorithmes correspondent à différentes variantes de la mise à jour de partage, lors de l'étape (4) de l'Algorithme 2.

Paramètres : $\eta > 0$ vitesse d'apprentissage et $0 < \alpha < 1$ proportion de mélange

Initialisation : $\widehat{\boldsymbol{p}}_1 = \widehat{\boldsymbol{v}}_1 = (1/d, \ldots, 1/d)$

Pour chaque instant $t = 1, \ldots, T$,

1. Prévoir $\widehat{\boldsymbol{p}}_t$;
2. Observer les pertes $\boldsymbol{\ell}_t \in [0,1]^d$
3. [Mise à jour des poids selon les pertes] $\widehat{\boldsymbol{v}}_{t+1}$ tq
$$\widehat{v}_{j,t+1} = \frac{\widehat{p}_{j,t}\, e^{-\eta\, \ell_{j,t}}}{\sum_{i=1}^{d} \widehat{p}_{i,t}\, e^{-\eta\, \ell_{i,t}}}$$
4. [Mise à jour de partage] $\widehat{\boldsymbol{p}}_{t+1} = (1-\alpha)\widehat{\boldsymbol{v}}_{t+1} + \alpha\, \widehat{\boldsymbol{v}}_1$

Algorithme 2: L'algorithme fixed-share d'Herbster et Warmuth (1998).

2 Résultat

2.1 Un regret généralisé

Nous introduisons maintenant le regret généralisé qui unifie les notions de regret escompté, de Cesa-Bianchi et Lugosi (2006), de regret d'Hazan et Seshadhri (2009) et du regret en ruptures de Herbster et Warmuth (2001). Pour un horizon T fixé, une suite de facteurs d'escomptes $\gamma_1^T = \gamma_1, \ldots, \gamma_T \in \mathbb{R}_+$ attribue des importances différentes aux pertes subies aux instants $t = 1, \ldots, T$. Nous considérons comme ensemble de stratégie de référence l'ensemble des suites de vecteurs $\boldsymbol{q}_1^T = \boldsymbol{q}_1, \ldots, \boldsymbol{q}_t$ du simplexe Δ_d. Notre but est de majorer le regret,

$$\sum_{t=1}^{T} \gamma_t\, \widehat{\boldsymbol{p}}_t^{\top} \boldsymbol{\ell}_t - \sum_{t=1}^{T} \gamma_t\, \boldsymbol{q}_t^{\top} \boldsymbol{\ell}_t \tag{1}$$

selon la régularité de la suite de comparaison $\boldsymbol{q}_1^T = \boldsymbol{q}_1, \ldots, \boldsymbol{q}_T$ et des variations des facteurs γ_t. Nous introduisons la mesure de régularité suivante

$$m(\gamma_1^T, \boldsymbol{q}_1^T) = \sum_{t=1}^{T} d_{TV}(\gamma_t \boldsymbol{q}_t, \gamma_{t-1} \boldsymbol{q}_{t-1}) \tag{2}$$

où pour $\boldsymbol{x} = (x_1, \ldots, x_d), \boldsymbol{y} = (y_1, \ldots, y_d) \in \mathbb{R}_+^d$, nous notons $d_{\mathrm{TV}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{x_i \geqslant y_i} (x_i - y_i)$. Remarquons que quand $\boldsymbol{x}, \boldsymbol{y} \in \Delta_d$, $d_{\mathrm{TV}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2} \left\| \boldsymbol{x} - \boldsymbol{y} \right\|_1$ et nous retrouvons la distance en variation totale. Dans le cas général $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}_+^d$, la quantité $d_{\mathrm{TV}}(\boldsymbol{x}, \boldsymbol{y})$ n'est pas nécessairement symétrique mais toujours bornée par $\left\| \boldsymbol{x} - \boldsymbol{y} \right\|_1$.

2.2 Borne sur le regret

Présentons maintenant notre résultat principal qui borne le regret de l'Algorithme 2.

Théorème 1 *Pour tout $T \geqslant 1$, pour toute suite de pertes $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_T \in [0,1]^d$, pour toute suite de comparaison $\boldsymbol{q}_1, \dots, \boldsymbol{q}_T \in \mathbb{R}_+^d$, et toute suite d'escomptes $\gamma_1, \dots, \gamma_t \in \mathbb{R}$,*

$$\sum_{t=1}^{T} \gamma_t \, \boldsymbol{p}_t^\top \boldsymbol{\ell}_t - \sum_{t=1}^{T} \gamma_t \boldsymbol{q}_t^\top \boldsymbol{\ell}_t \leqslant \frac{\gamma_1 \ln d}{\eta} + \frac{\eta}{8} \sum_{t=1}^{T} \gamma_t + \frac{m(\gamma_1^T, \boldsymbol{q}_1^T)}{\eta} \ln \frac{d}{\alpha} + \frac{\sum_{t=2}^{T} \gamma_t - m(\gamma_1^T, \boldsymbol{q}_1^T)}{\eta} \ln \frac{1}{1-\alpha} \, .$$

L'Algorithme 2 ne nécessite aucune connaissance à l'avance sur la suite (γ_t) pour que la borne ci-dessus soit valide. Bien sûr, afin de minimiser la borne supérieure obtenue, les paramètres α, η doivent être calibrés. Leurs valeurs optimales dépendant de $m(\gamma_1^T, \boldsymbol{q}_1^T)$ et $\sum_{t=1}^{T} \gamma_t$ pour les suites que nous souhaitons concurrencer. Ceci est illustré dans le corollaire suivant.

Corollaire 1 *Soit $m_0 > 0$ and $U_0 > 0$. Pour tout $T \geqslant 1$, pour toutes pertes $\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_T \in [0,1]^d$, toute suite $\boldsymbol{q}_1, \dots, \boldsymbol{q}_T \in \mathbb{R}_+^d$, et toute suite d'escomptes $\gamma_1, \dots, \gamma_T \in \mathbb{R}$ telles que $\gamma_1 + m(\gamma_1^T, \boldsymbol{q}_1^T) \leqslant m_0$ et $\sum_{t=1}^{T} \gamma_t \leqslant U_0$,*

$$\sum_{t=1}^{T} \gamma_t \, \boldsymbol{p}_t^\top \boldsymbol{\ell}_t - \sum_{t=1}^{T} \gamma_t \boldsymbol{q}_t^\top \boldsymbol{\ell}_t \leqslant \sqrt{\frac{U_0 \, m_0}{2} \ln \left(\frac{d \, e \, U_0}{m_0} \right)}$$

dès que les paramètres η et α sont choisis de façon optimale en fonction de m_0 et de U_0.

Le regret en ruptures a été introduit par Herbster et Warmuth (2001) afin de s'adapter à un environnement pouvant évoluer. Il se déduit immédiatement de (1) lorsque $\gamma_t = 1$ pour tout t.

Le regret adaptatif a été introduit par Hazan et Seshadhri (2008) et propose une alternative au regret en ruptures pour faire face à un environnement changeant. Pour $\tau_0 \in \{1, \dots, T\}$, le *$\tau_0$-regret adaptatif* d'un joueur est défini par

$$\max_{\substack{[r,s] \subset [1,T] \\ s+1-r \leqslant \tau_0}} \left\{ \sum_{t=r}^{s} \boldsymbol{p}_t^\top \boldsymbol{\ell}_t - \min_{\boldsymbol{q} \in \Delta_d} \sum_{t=r}^{s} \boldsymbol{q}^\top \boldsymbol{\ell}_t \right\} . \tag{3}$$

En appliquant le Corollaire 1 à des suites d'escomptes γ_t telles que $\gamma_t = 1$ si $r \leqslant t \leqslant s$ et 0 sinon, et à des suites de comparaison constantes $\boldsymbol{q}_t = \boldsymbol{q} \in \Delta_d$ pour tout t, nous déduisons immédiatement la majoration du regret adaptatif de l'Algorithme 2 par $\sqrt{\tau_0 \, \ln(ed\tau_0)/2}$.

Le regret escompté a été introduit par Cesa-Bianchi et Lugosi (2006) et est défini par

$$\max_{\boldsymbol{q}\in\Delta_d} \sum_{t=1}^{T} \gamma_t \left(\boldsymbol{p}_t^\top \boldsymbol{\ell}_t - \boldsymbol{q}^\top \boldsymbol{\ell}_t\right) . \tag{4}$$

Les facteurs d'escomptes γ_t mesurent l'importance relative des pertes plus ou moins récentes, ou dans un cadre de théorie des jeux celle des pertes à échéances plus ou moins proches. Nous ne considérons que les suites d'escomptes monotones et sans perte de généralité à valeur dans $[0, 1]$. Dans ce cas particulier, nous pouvons remarquer que le critère de régularité (2) vérifie $m(\gamma_1^T, \boldsymbol{q}) = \max\{\gamma_1, \gamma_T\} \leqslant 1$. Nous majorons donc par le Corollaire 1 le regret escompté de l'Algorithme 2 par $\sqrt{U_0 \ln(d\, e\, U_0)/2}$, dès que ses paramètres α et η sont optimisés en fonction de $U_0 = \sum_{t=1}^{T} \gamma_t$.

Bibliographie

[1] Bousquet, O et Warmuth, M. (2002), *Tracking a small set of experts by mixing past posteriors*, Journal of Machine Learning Research.

[2] Cesa-Bianchi, N. Gaillard, P. Lugosi, G. Stoltz, G. (2012) *Mirror descent meets fixed-share (and feels no regret)*, Proceedings of the 26th Conference on Neural Information Processing Systems (NIPS).

[3] Cesa-Bianchi, N. et Lugosi, G. (2006), *Prediction, learning, and games*, Cambridge University Press.

[4] Hazan, E. et Seshadhri, C. (2009), *Efficient learning algorithms for changing environments*, Proceedings of the 26th International Conference of Machine Learning (ICML).

[5] Herbster, M. et Warmuth, M. (2001), *Tracking the best linear predictor*, Journal of Machine Learning Research.

[6] Zinkevich, M. (2003), *Online convex programming and generalized infinitesimal gradient ascent*, Proceedings of the 20th International Conference on Machine Leaning (ICML).