Mirror Descent Meets Fixed Share (and feels no regret)

Nicolò Cesa-Bianchi, Pierre Gaillard, Gábor Lugosi, Gilles Stoltz

Summary

Unified analysis for several algorithms in online linear optimization.

1. **New** notion of regret that unifies shifting, adaptive, and discounted regret.
2. **Notion of regularity** for changing environments.

Setting

Online linear optimization in the simplex.
For each round $t = 1, \dots, T$,

1. Forecaster chooses in the simplex
   $\widehat{\boldsymbol{p}}_t = (\widehat{p}_{1,t}, \dots, \widehat{p}_{d,t}) \in \Delta_d$
2. Environment chooses a loss vector
   $\boldsymbol{\ell}_t = (\ell_{1,t}, \dots, \ell_{d,t}) \in [0,1]^d$
3. Forecaster suffers loss $\widehat{\boldsymbol{p}}_t^\top \boldsymbol{\ell}_t$.

Algorithm

Parameters: $\eta > 0$ (learning rate) and ψ_t (mixing functions)

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

For each round $t = 1, \dots, T$,

1. Predict $\widehat{\boldsymbol{p}}_t$;
2. Observe losses $\boldsymbol{\ell}_t \in [0,1]^d$;
3. [loss update]
   $$v_{j,t+1} = \frac{\widehat{p}_{j,t}\, e^{-\eta\, \ell_{j,t}}}{\sum_{i=1}^d \widehat{p}_{i,t}\, e^{-\eta\, \ell_{i,t}}}$$
4. [shared update] $\widehat{\boldsymbol{p}}_{t+1} = \psi_{t+1}\left([v_{i,t}]_{i,s} \right)$

Updates

[Zinkevich, 03] [Bous. & War., 01] [Herb. & War., 98]

Projected update. $\Delta_d^\alpha = [\alpha/d, 1]^d \cap \Delta_d$,
$\widehat{\boldsymbol{p}}_{t+1} = \Pi_{\Delta_d^\alpha}(\boldsymbol{v}_{t+1})$.

Fixed-share update.
$\widehat{\boldsymbol{p}}_{t+1} = (1-\alpha)\boldsymbol{v}_{t+1} + \alpha \boldsymbol{v}_1$.

Bousquet-Warmuth update.
$$p_{j,t+1} = (1-\alpha) v_{j,t+1} + \alpha \, \frac{\max_{s \leqslant t} v_{j,s}}{\sum_{i=1}^N \max_{s \leqslant t} v_{i,s}} .$$

Generalized regret

Aim: To bound the regret
$$\sum_{t=1}^T \|\boldsymbol{u}_t\|_1 \, \widehat{\boldsymbol{p}}_t^\top \boldsymbol{\ell}_t - \sum_{t=1}^T \boldsymbol{u}_t^\top \boldsymbol{\ell}_t$$
in terms of the regularity of the comparison sequence $\boldsymbol{u}_1^T = \boldsymbol{u}_1, \dots, \boldsymbol{u}_T \in \mathbb{R}_+^d$.

The regularity of a sequence $\boldsymbol{u}_1^T$ is usually measured by the number of times $\boldsymbol{u}_t \neq \boldsymbol{u}_{t-1}$ or by the cumulative difference of norms $\sum_t \|\boldsymbol{u}_t - \boldsymbol{u}_{t-1}\|_1$.

We introduce the following measure of regularity
$$m(\boldsymbol{u}_1^T) = \sum_{t=2}^T D_{TV}(\boldsymbol{u}_t, \boldsymbol{u}_{t-1})$$
where $D_{TV}(\boldsymbol{x}, \boldsymbol{y}) = \sum_i (x_i - y_i)_+$ is a "generalized" total variation distance. Note that $m(\boldsymbol{u}_1^T) \leqslant \sum_t \|\boldsymbol{u}_t - \boldsymbol{u}_{t-1}\|_1$.

Main Result – Fixed-share update

Theorem. *Let $m_0 > 0$ and $U_0 > 0$. For all $T \geqslant 1$, whenever η and α are optimally chosen in terms of m_0 and U_0,*
$$\max_{\boldsymbol{\ell}_1^T} \max_{\boldsymbol{u}_1^T \text{ st. } (1)} \sum_{t=1}^T \|\boldsymbol{u}_t\|_1 \, \widehat{\boldsymbol{p}}_t^\top \boldsymbol{\ell}_t - \sum_{t=1}^T \boldsymbol{u}_t^\top \boldsymbol{\ell}_t \leqslant \sqrt{\frac{U_0\, m_0}{2} \ln\left(\frac{e\, d\, U_0}{m_0}\right)}$$
where (1) controls the regularity of $\boldsymbol{u}_1^T$ by $\|\boldsymbol{u}_1\|_1 + m(\boldsymbol{u}_1^T) \leqslant m_0$ and $\sum_{t=1}^T \|\boldsymbol{u}_t\|_1 \leqslant U_0$.

Shifting regret

[Herbster & Warmuth, 01]

Regret against a convex comparison sequence $\boldsymbol{q}_1, \dots, \boldsymbol{q}_T \in \Delta_d$.
$$\sum_{t=1}^T \widehat{\boldsymbol{p}}_t^\top \boldsymbol{\ell}_t - \min_{\substack{\boldsymbol{q}_1^T \text{ st.} \\ \frac{1}{2}\sum_{t=1}^T \|\boldsymbol{q}_t - \boldsymbol{q}_{t-1}\|_1 \leqslant m_0'}} \sum_{t=1}^T \boldsymbol{q}_t^\top \boldsymbol{\ell}_t .$$

We get regret bound of order $\sqrt{m_0' T \ln \frac{dT}{m_0'}}$. In particular, the regret is sublinear only if $m_0' \ll T$.

Proof. Restriction to convex comparison sequences $\boldsymbol{u}_t = \boldsymbol{q}_t \in \Delta_d$ such that $m(\boldsymbol{q}_1^T) \leqslant m_0'$. The theorem is applied with $m_0 = m_0' + 1$ and $U_0 = T$.

Adaptive regret $\mathcal{R}_T^{\tau_0\text{-adapt}}$

[Hazan & Seshadhri, 09]

Regret against a constant $\boldsymbol{q} \in \Delta_d$ during sub-intervals of time.
$$\max_{\substack{[r,s] \subset [1,T] \\ s+1-r \leqslant \tau_0}} \left\{ \sum_{t=r}^s \widehat{\boldsymbol{p}}_t^\top \boldsymbol{\ell}_t - \min_{\boldsymbol{q} \in \Delta_d} \sum_{t=r}^s \boldsymbol{q}^\top \boldsymbol{\ell}_t \right\} .$$

We get essentially $\square\sqrt{\tau_0 \ln (d\tau_0)}$ as a regret bound.

Proof. Restriction to comparison sequences of the form
$\boldsymbol{u}_1^T = 0, .., 0, \boldsymbol{q}, .., \boldsymbol{q}, 0, .., 0$
with $\boldsymbol{q} \in \Delta_d$. The theorem is applied with $m_0 = 1$ and $U_0 = \tau_0$.

Discounted regret

[Cesa-Bianchi & Lugosi, 06]

Discounts factors $\beta_{t,T}$ measure the relative importance of more recent rounds,
$$\max_{\boldsymbol{q} \in \Delta_d} \sum_{t=1}^T \beta_{t,T} \left(\widehat{\boldsymbol{p}}_t^\top \boldsymbol{\ell}_t - \boldsymbol{q}^\top \boldsymbol{\ell}_t \right) .$$

By considering discounts $\beta_{t,T}$ monotonic in t, we get regret bounds of order $\sqrt{U_0 \beta_{\max} \ln \frac{dU_0}{\beta_{\max}}}$, where $U_0 = \sum_{t \leqslant T} \beta_{t,T}$ and $\beta_{\max} = \max\{\beta_{1,T}, \beta_{T,T}\}$.

Example. $\beta_{t,T} = (T-t)^{-1}$. We get $U_0 = \sum_{t=1}^T (T-t)^{-1} \leqslant \ln T$ and the regret is bounded by $\square\sqrt{\ln T \ln (d \ln T)}$.

Remark. Only U_0 needs here to be known in advance, in contrast to earlier discounted regret algorithms that need to know beforehand all the discounts $\beta_{t,T}$. The calibration of the parameters is thus much easier.

Proof. Restriction to comparison sequences of the form
$\boldsymbol{u}_1^T = \beta_{1,T}\boldsymbol{q}, \dots, \beta_{T,T}\boldsymbol{q}$
with monotonic discounts $\beta_{t,T} \geqslant 0$ in t and $\boldsymbol{q} \in \Delta_d$. The theorem is applied with $m_0 = \max\{\beta_{1,T}, \beta_{T,T}\}$ and $U_0 = \sum_{t \leqslant T} \beta_{t,T}$.