The many faces of Exponential Weights in Online Learning

Dirk van der Hoeven, Tim van Erven, and Wojciech Kotłowski

06-2018

Universiteit Leiden
The Netherlands

Motivation: Football prediction

- Every football match t there are K models that predict the outcome of the match.
- Form probability distribution $p_t(k)$ and play the weighted average of the K predictions: $\mathbb{E}_{p_t(k)}[\boldsymbol{w}_k]$.
- After match: observe losses $f_t(\boldsymbol{w}_k)$ and suffer loss $f_t(\mathbb{E}_{p_t(k)}[\boldsymbol{w}_k])$.
- Improve probability distribution $p_t(k) \to p_{t+1}(k)$ based on observed losses.

Organization

Online Convex Optimization

Parameters $\boldsymbol{w}$ take values in convex domain $\mathcal{W}$.

for $t = 1, 2, \ldots, T$ **do**
    Learner predicts $\boldsymbol{w}_t \in \mathcal{W}$
    Observe convex loss function $f_t : \mathcal{W} \mapsto \mathbb{R}$
    Learner suffers loss $f_t(\boldsymbol{w}_t)$

Online Convex Optimization

Parameters $\boldsymbol{w}$ take values in convex domain $\mathcal{W}$.

for $t = 1, 2, \ldots, T$ **do**

    Learner predicts $\boldsymbol{w}_t \in \mathcal{W}$

    Observe convex loss function $f_t : \mathcal{W} \mapsto \mathbb{R}$

    Learner suffers loss $f_t(\boldsymbol{w}_t)$

Objective: minimize regret w.r.t. oracle parameter $\boldsymbol{u} \in \mathcal{W}$

$$\mathcal{R}_T(\boldsymbol{u}) = \sum_{t=1}^{T} f_t(\boldsymbol{w}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{u}). \tag{1}$$

How to control Regret?

- Convex loss function: Mirror Descent.
- Strongly convex loss functions: Gradient Descent.
- Exp-concave loss functions: Online Newton Step.
- Adaptive expert algorithms: Iprod, Squint, Coin Betting.
- Linear bandits: Mirror Descent with self concordant barrier regularizer.

We all live an in exponentially weighted world

- Convex loss function: Mirror Descent.
- Strongly convex loss functions: Gradient Descent.
- Exp-concave loss functions: Online Newton Step.
- Adaptive expert algorithms: Iprod, Squint, Coin Betting.
- Linear bandits: Mirror Descent with self concordant barrier regularizer.

We say:

$$\left.\begin{array}{l}\text{Mirror Descent}\\ \text{Gradient Descent}\\ \text{Online Newton Step}\\ \text{Squint, Coin Betting}\\ \text{MD self concordant}\end{array}\right\} = \text{Exponential Weights}$$

Organization

Exponential Weights

Input: a convex set of distributions $\mathcal{P}$ over $\boldsymbol{w}$, a prior $P_1 \in \mathcal{P}$ and learning rates $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_T > 0$

Lazy:

1. update step:
   $\tilde{P}_{t+1} = \arg\min_P \; \mathbb{E}_P \left[\sum_{s=1}^t f_s(\boldsymbol{w}) \right] + \frac{1}{\eta_t} \mathrm{KL}(P \| P_1)$
2. projection step:
   $P_{t+1} = \arg\min_{P \in \mathcal{P}} \; \mathrm{KL}(P \| \tilde{P}_{t+1})$

Greedy:

1. update step:
   $\tilde{P}_{t+1} = \arg\min_P \; \mathbb{E}_P[f_t(\boldsymbol{w})] + \frac{1}{\eta_t} \mathrm{KL}(P \| P_t)$
2. projection step:
   $P_{t+1} = \arg\min_{P \in \mathcal{P}} \; \mathrm{KL}(P \| \tilde{P}_{t+1})$

Exponential Weights

The algorithm gets its name from the distributions $\tilde{P}_t$, whose densities have the following exponential forms:

$$\mathrm{d}\tilde{P}_{t+1}(\boldsymbol{w}) = \frac{e^{-\eta_t \sum_{s=1}^{t} f_s(\boldsymbol{w})} \, \mathrm{d}P_1(\boldsymbol{w})}{\int e^{-\eta_t \sum_{s=1}^{t} f_s(\boldsymbol{w})} \, \mathrm{d}P_1(\boldsymbol{w})} \qquad \text{(lazy EW)}$$

$$\mathrm{d}\tilde{P}_{t+1}(\boldsymbol{w}) = \frac{e^{-\eta_t f_t(\boldsymbol{w})} \, \mathrm{d}P_t(\boldsymbol{w})}{\int e^{-\eta_t f_t(\boldsymbol{w})} \, \mathrm{d}P_t(\boldsymbol{w})} \qquad \text{(greedy EW).}$$

Regret Exponential Weights

Let $Q \in \mathcal{P}$ be any comparator distribution such that $\mathrm{KL}(Q \| \tilde{P}_t) < \infty$ for all t, let $\{\boldsymbol{w}_t \in \mathcal{W}\}_{t=1}^T$ be the actions of any learner, and define $\eta_0 \stackrel{\text{def}}{=} \eta_1$. Then lazy EW satisfies

$$\mathop{\mathbb{E}}_{\boldsymbol{u} \sim Q}[\mathcal{R}(\boldsymbol{u})] \leq \frac{1}{\eta_T} \mathrm{KL}(Q \| P_1) + \sum_{t=1}^T \Big\{ \underbrace{f_t(\boldsymbol{w}_t) + \tfrac{1}{\eta_{t-1}} \ln \mathop{\mathbb{E}}_{P_t(\boldsymbol{w})} \Big[e^{-\eta_{t-1} f_t(\boldsymbol{w})} \Big]}_{\text{“mixability gap”}} \Big\}.$$

Greedy EW satisfies:

$$\mathop{\mathbb{E}}_{\boldsymbol{u} \sim Q}[\mathcal{R}(\boldsymbol{u})] \leq \frac{1}{\eta_1} \mathrm{KL}(Q \| P_1) + \left(\frac{1}{\eta_T} - \frac{1}{\eta_1} \right) \max_{t=2,\dots,T} \mathrm{KL}(Q \| P_t) + \sum_{t=1}^T \Big\{ \underbrace{f_t(\boldsymbol{w}_t) + \tfrac{1}{\eta_t} \ln \mathop{\mathbb{E}}_{P_t(\boldsymbol{w})} \Big[e^{-\eta_t f_t(\boldsymbol{w})} \Big]}_{\text{“mixability gap”}} \Big\}.$$

Regret Exponential Weights

Proof structure in most settings

- Bound the mixability gap
- Find Q for which the expected loss under Q together with $\mathrm{KL}(Q \| P1)$ can be related to the loss of a deterministic comparator.

Unless specified otherwise

- $\boldsymbol{w}_t = \mathbb{E}_{P_t}[\boldsymbol{w}]$
- $\mathcal{P} = \{P : \mathbb{E}_P[\boldsymbol{w}] \in \mathcal{W}\}$

Organization

Convex losses

A standard approach in OCO is to lower-bound the convex losses f_t by their tangent at $\boldsymbol{w}_t$:

$$\sum_{t=1}^{T} \left(f_t(\boldsymbol{w}_t) - f_t(\boldsymbol{u})\right) \leq \sum_{t=1}^{T} \left(\langle \boldsymbol{w}_t, \boldsymbol{g}_t \rangle - \langle \boldsymbol{u}, \boldsymbol{g}_t \rangle\right),$$

where $\boldsymbol{g}_t = \nabla f_t(\boldsymbol{w}_t)$

Standard approach: Mirror Descent

- $B_{F^*}(\boldsymbol{u}\|\boldsymbol{w}) = F^*(\boldsymbol{u}) - F^*(\boldsymbol{w}) - \nabla F^*(\boldsymbol{w})^\intercal(\boldsymbol{u}-\boldsymbol{w})$ denotes the Bregman divergence generated by F^*.
- $F^*(\boldsymbol{w}) = \sup_{\boldsymbol{\theta}}\langle \boldsymbol{w}, \boldsymbol{\theta}\rangle - F(\boldsymbol{\theta})$ denotes the convex conjugate of F.

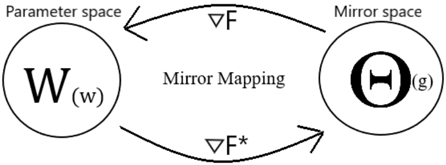
Lazy Mirror Descent:

$$\begin{aligned}\tilde{\boldsymbol{w}}_{t+1} &= \operatorname*{arg\,min}_{\boldsymbol{w}} \textstyle\sum_{s=1}^{t}\langle \boldsymbol{w}, \boldsymbol{g}_s\rangle + \frac{1}{\eta_t} B_{F^*}(\boldsymbol{w}\|\boldsymbol{w}_1)\\ \boldsymbol{w}_{t+1} &= \operatorname*{arg\,min}_{\boldsymbol{w}\in\mathcal{W}} B_{F^*}(\boldsymbol{w}\|\tilde{\boldsymbol{w}}_{t+1}).\end{aligned}$$

Greedy Mirror Descent:

$$\begin{aligned}\tilde{\boldsymbol{w}}_{t+1} &= \operatorname*{arg\,min}_{\boldsymbol{w}} \langle \boldsymbol{w}, \boldsymbol{g}_t\rangle + \frac{1}{\eta_t} B_{F^*}(\boldsymbol{w}\|\boldsymbol{w}_t)\\ \boldsymbol{w}_{t+1} &= \operatorname*{arg\,min}_{\boldsymbol{w}\in\mathcal{W}} B_{F^*}(\boldsymbol{w}\|\tilde{\boldsymbol{w}}_{t+1}).\end{aligned}$$

Interpretation Mirror Descent



Common choices for F:

- Gradient Descent: $F(\boldsymbol{\theta}) = \frac{1}{2}\|\boldsymbol{\theta}\|_2^2$
- Unnormalized Relative entropy: $F(\boldsymbol{\theta}) = \sum_{i=1}^{d} e^{\theta_i}$
- Exponentiated Gradient ($\pm$): $F(\boldsymbol{\theta}) = \log(\sum_{i=1}^{d} e^{\theta_i})$

Mirror Descent as Exponential Weights

We consider prior from exponential families:

- Have form $\mathcal{E} = \{P_{\boldsymbol{\theta}} \mid \mathrm{d}P_{\boldsymbol{\theta}}(\boldsymbol{w}) = e^{\langle \boldsymbol{\theta}, \boldsymbol{w} \rangle - F(\boldsymbol{\theta})} \mathrm{d}K(\boldsymbol{w}), \boldsymbol{\theta} \in \Theta\}$
- Nonnegative *carrier measure* K
- Cumulant generating function $F(\boldsymbol{\theta}) = \ln \int e^{\langle \boldsymbol{\theta}, \boldsymbol{w} \rangle} \mathrm{d}K(\boldsymbol{w})$
- Parameter space $\Theta = \{\boldsymbol{\theta} \mid F(\boldsymbol{\theta}) < \infty\} \subset \mathbb{R}^d$
- Called *regular* if Θ is an open set

Mirror Descent as Exponential Weights

Theorem

Suppose F is the cumulant generating function of a regular exponential family $\mathcal{E}$. Then the lazy and greedy versions of MD predict with the means $\boldsymbol{w}_t = \mathbb{E}_{P_t}[\boldsymbol{w}]$ of lazy and greedy EW on the linearized losses with the same η_t, prior $P_{\boldsymbol{\theta}_1}$ for $\boldsymbol{\theta}_1 = \nabla F^*(\boldsymbol{w}_1)$ and $\mathcal{P} = \{P : \mathbb{E}_P[\boldsymbol{w}] \in \mathcal{W}\}$.

Lazy EW:

$$\mathop{\mathbb{E}}_{P_{t+1}}[\boldsymbol{w}] = \boldsymbol{w}_{t+1} = \operatorname*{arg\,min}_{\boldsymbol{w}\in\mathcal{W}} \textstyle\sum_{s=1}^{t} \langle \boldsymbol{w}, \boldsymbol{g}_s \rangle + \frac{1}{\eta_t} B_{F^*}(\boldsymbol{w} \| \boldsymbol{w}_1)$$

Greedy EW:

$$\mathop{\mathbb{E}}_{P_{t+1}}[\boldsymbol{w}] = \boldsymbol{w}_{t+1} = \operatorname*{arg\,min}_{\boldsymbol{w}\in\mathcal{W}} \langle \boldsymbol{w}, \boldsymbol{g}_t \rangle + \frac{1}{\eta_t} B_{F^*}(\boldsymbol{w} \| \boldsymbol{w}_t)$$

Greedy MD as Greedy EW proof

We can restrict P to an exponential family:

$$\begin{aligned}
&\min_{P\in\mathcal{P}}\left\{\mathbb{E}_{P}[\langle \boldsymbol{w},\boldsymbol{g}_t\rangle]+\frac{1}{\eta_t}\mathsf{KL}(P\|P_t)\right\}\\
&=\min_{\boldsymbol{\mu}\in\mathcal{W}}\ \min_{P\,:\,\mathbb{E}_P[\boldsymbol{w}]=\boldsymbol{\mu}}\left\{\mathbb{E}_{P}[\langle \boldsymbol{w},\boldsymbol{g}_t\rangle]+\frac{1}{\eta_t}\mathsf{KL}(P\|P_t)\right\}\\
&=\min_{\boldsymbol{\mu}\in\mathcal{W}}\ \min_{P\in\mathcal{E}\,:\,\mathbb{E}_P[\boldsymbol{w}]=\boldsymbol{\mu}}\left\{\langle \boldsymbol{\mu},\boldsymbol{g}_t\rangle+\frac{1}{\eta_t}\mathsf{KL}(P\|P_t)\right\},
\end{aligned}$$

where the last equality is due to Theorem 3.1.4 in Ihara (1993).

Greedy MD as Greedy EW proof

To finish the proof we use a result from Banerjee et al. (2005); Nielsen and Nock (2010). Let $\boldsymbol{\mu}_P = \mathbb{E}_P[\boldsymbol{w}]$. For $Q, P \in \mathcal{E}$:

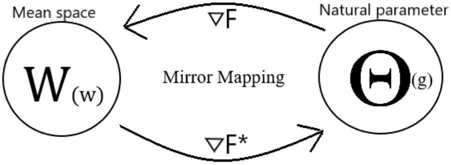
$$\mathrm{KL}(P \| Q) = B_F(\boldsymbol{\theta}_Q \| \boldsymbol{\theta}_P) = B_{F^*}(\boldsymbol{\mu}_P \| \boldsymbol{\mu}_Q).$$

We now have:

$$\begin{aligned} P_{t+1} &= \operatorname*{arg\,min}_{P \in \mathcal{E} : \boldsymbol{\mu}_P \in \mathcal{W}} \left\{ \langle \boldsymbol{\mu}_P, \boldsymbol{g}_t \rangle + \frac{1}{\eta_t} \mathrm{KL}(P \| P_t) \right\} \\ &= \operatorname*{arg\,min}_{P \in \mathcal{E} : \boldsymbol{\mu}_P \in \mathcal{W}} \left\{ \langle \boldsymbol{\mu}_P, \boldsymbol{g}_t \rangle + \frac{1}{\eta_t} B_{F^*}(\boldsymbol{\mu}_P \| \boldsymbol{\mu}_{P_t}) \right\}, \end{aligned}$$

which coincides with the definition of greedy Mirror Descent.

MD as EW interpretation



Before we had:

- Gradient Descent: $F(\boldsymbol{\theta}) = \frac{1}{2}\|\boldsymbol{\theta}\|_2^2$
- Unnormalized Relative entropy: $F(\boldsymbol{\theta}) = \sum_{i=1}^{d} e^{\theta_i}$
- Exponentiated Gradient $(\pm)$: $F(\boldsymbol{\theta}) = \log(\sum_{i=1}^{d} e^{\theta_i})$

Now we have:

- Gradient Descent: Gaussian Prior
- Unnormalized Relative entropy: Poisson Prior
- Exponentiated Gradient $(\pm)$: Multinomial Prior (1 trial)

Organization

Quadratic losses

We assume that the losses f_t satisfy quadratic lower bounds:

$$f_t(\boldsymbol{w}) - f_t(\boldsymbol{w}_t) \geq \langle \boldsymbol{w} - \boldsymbol{w}_t, \boldsymbol{g}_t \rangle + \frac{1}{2}(\boldsymbol{w} - \boldsymbol{w}_t)^\intercal \boldsymbol{M}_t(\boldsymbol{w} - \boldsymbol{w}_t) =: \ell_t(\boldsymbol{w}),$$

where $\boldsymbol{M}_t$ is a positive semi-definite matrix.

We treat two cases:

- α-strongly convex loss functions: $M_t = \alpha \boldsymbol{I}$
- α-exp concave loss functions: $\boldsymbol{M}_t = \beta \boldsymbol{g}_t \boldsymbol{g}_t^T$, where $\beta = \frac{1}{2}\min\{\frac{1}{4GB}, \alpha\}$, assuming $\|\boldsymbol{g}_t\|_2 \leq G$ and $B = \max_{\boldsymbol{w},\boldsymbol{u} \in \mathcal{W}} \|\boldsymbol{w} - \boldsymbol{u}\|_2$

Quadratic losses: Gaussian prior

Theorem

Let $P_1 = \mathcal{N}(\boldsymbol{w}_1, \Sigma_1)$. Both versions of the Exponential Weights algorithm, run on ℓ_t with learning rate η and $\mathcal{P} = \{P : \mathbb{E}_P[\boldsymbol{w}] \in \mathcal{W}\}$, yield a multivariate normal distribution $P_{t+1} = \mathcal{N}(\boldsymbol{w}_{t+1}, \Sigma_{t+1})$. Furthermore for all $\boldsymbol{u} \in \mathbb{R}^d$ both versions of EW satisfy:

$$\mathcal{R}_T(\boldsymbol{u}) \leq \frac{1}{2\eta}(\boldsymbol{w}_1 - \boldsymbol{u})^\intercal \Sigma_1^{-1}(\boldsymbol{w}_1 - \boldsymbol{u}) + \frac{\eta}{2}\sum_{t=1}^{T} \boldsymbol{g}_t^\intercal \Sigma_{t+1} \boldsymbol{g}_t. \qquad (2)$$

Quadratic losses: Gaussian prior

Lazy EW Gaussian prior:

$$\begin{aligned}
\Sigma_{t+1}^{-1} &= \Sigma_t^{-1} + \eta \boldsymbol{M}_t \\
\tilde{\boldsymbol{w}}_{t+1} &= \tilde{\boldsymbol{w}}_t - \eta \Sigma_{t+1} \boldsymbol{g}_t \\
\boldsymbol{w}_{t+1} &= \operatorname*{arg\,min}_{\boldsymbol{w} \in \mathcal{W}} (\boldsymbol{w} - \tilde{\boldsymbol{w}}_{t+1})^\intercal \Sigma_{t+1}^{-1} (\boldsymbol{w} - \tilde{\boldsymbol{w}}_{t+1})
\end{aligned}$$

Greedy EW Gaussian prior:

$$\begin{aligned}
\Sigma_{t+1}^{-1} &= \Sigma_t^{-1} + \eta \boldsymbol{M}_t \\
\tilde{\boldsymbol{w}}_{t+1} &= \boldsymbol{w}_t - \eta \Sigma_{t+1} \boldsymbol{g}_t \\
\boldsymbol{w}_{t+1} &= \operatorname*{arg\,min}_{\boldsymbol{w} \in \mathcal{W}} (\boldsymbol{w} - \tilde{\boldsymbol{w}}_{t+1})^\intercal \Sigma_{t+1}^{-1} (\boldsymbol{w} - \tilde{\boldsymbol{w}}_{t+1})
\end{aligned}$$

Strongly convex loss functions

- For α-strongly convex loss functions the standard approach is to use Greedy Gradient Descent with learning rates $\eta_t = 1/(\alpha t)$ (Hazan et al., 2007)
- Greedy EW on $\ell_t(\boldsymbol{w})$ with *fixed* learning rate η and Gaussian prior $P_1 = \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$ yields greedy GD with $\eta_t = 1/(\frac{1}{\eta\sigma^2} + \alpha t)$
- Regret EW: $\mathcal{R}_T(\boldsymbol{u}) \leq \frac{G^2}{2\alpha} \ln \left(\frac{\frac{1}{\eta\sigma^2} + \alpha T}{\frac{1}{\eta\sigma^2} + \alpha} \right) + \frac{G^2}{\frac{2}{\eta\sigma^2} + 2\alpha} + \frac{D^2}{2\eta\sigma^2}$
- The standard learning rate and corresponding regret bound for GD (Hazan et al., 2007) correspond to the limiting case $\eta\sigma^2 \to \infty$

Exp-concave loss functions

- For α-exp-concave loss functions the standard approach is to use the Online Newton Step algorithm (Hazan et al., 2007)
- Exponential Weights on $\ell_t(\boldsymbol{w})$ with Gaussian prior $\mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I})$ leads to the Online Newton Step algorithm
- Regret EW: $\mathcal{R}_T(\boldsymbol{u}) \leq \frac{d}{2\beta} \ln\left(1 + \frac{\eta\sigma^2\beta G^2 T}{d}\right) + \frac{D^2}{2\eta\sigma^2}$
- To obtain the standard regret bound set $\eta\sigma^2 = \beta D^2$

Organization

Expert setting

- Linear losses $f_t(\boldsymbol{w}) = \langle \boldsymbol{w}, \boldsymbol{g}_t \rangle$ over the simplex $\mathcal{W} = \{\boldsymbol{w} : w_i \geq 0, \sum_{i=1}^{d} w_i = 1\}$, with $g_{t,i} \in [0, 1]$
- Instantaneous regret in round t with respect to expert i is $r_t(i) = f_t(\boldsymbol{w}_t) - f_t(\boldsymbol{e}_i)$
- Total Regret $\mathcal{R}_T(i) = \sum_{t=1}^{T} r_t(i)$

Adaptive algorithms

- Standard algorithms requires the learner to specify η
- η usually specified to guard against worst case, algorithm may be slow to converge
- To specify η one may require knowledge of unknown quantities

Solution: learn the optimal η.

Adaptive algorithms regret

Define $\mathcal{V}_T(i) = \sum_{t=1}^{T} r_t(i)^2$.

- Iprod, Squint:

$$\mathop{\mathbb{E}}_{\hat{\pi}}\left[\mathcal{R}_T(i)\right] = O\left(\sqrt{\mathop{\mathbb{E}}_{\hat{\pi}}[\mathcal{V}_T(i)]\Big(\mathrm{KL}(\hat{\pi}\|\pi) + \ln\ln T\Big)}\right)$$

- Coin Betting:

$$\mathop{\mathbb{E}}_{\hat{\pi}}[\mathcal{R}_T(i)] \leq \sqrt{3T\left(\mathrm{KL}(\hat{\pi}\|\pi) + 3\right)}$$

Surrogate Task

- Surrogate loss: $\ell_t(\eta, i) = -\ln(1 + \eta r_t(i))$
- Predictions take the form of joint distributions P_t on (η, i) for $\eta \in [0, 1]$
- Map back to predictions in the original task via $\boldsymbol{w}_t = \frac{\mathbb{E}_{P_t}[\eta \boldsymbol{e}_i]}{\mathbb{E}_{P_t}[\eta]}$
- Aim: achieve small *mix-regret* with respect to any comparator distribution Q on (η, i)

Mix regret

Mix regret:

$$S(Q) = \sum_{t=1}^{T} - \ln \mathop{\mathbb{E}}_{P_t} \left[e^{-\ell_t(\eta,i)} \right] - \mathop{\mathbb{E}}_{Q} \left[\sum_{t=1}^{T} \ell_t(\eta, i) \right].$$

If the learner can guarantee

$$0 \leq \sum_{t=1}^{T} \mathop{\mathbb{E}}_{Q} [\ell_t(\eta, i)] + S(Q)$$

then use $-\ln(1+x) \leq -x + x^2$ for $|x| \leq \frac{1}{2}$ to obtain:

$$0 \leq \sum_{t=1}^{T} \mathop{\mathbb{E}}_{Q} [-\eta r_t(i) + \eta^2 r_t(i)^2] + S(Q)$$

EW is the solution

How to guarantee $0 \leq \sum_{t=1}^{T} \mathbb{E}_Q[\ell_t(\eta, i)] + S(Q)$?

Use Exponential weights on ℓ_t with predictions $\boldsymbol{w}_t = \frac{\mathbb{E}_{P_t}[\eta \boldsymbol{e}_i]}{\mathbb{E}_{P_t}[\eta]}$:

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}_{Q}[\ell_t(\eta, i)] + S(Q) &= \prod_{t=1}^{T} \mathbb{E}_{P_t}\left[e^{-\ell_t(\eta, i)}\right] \\
&= \prod_{t=1}^{T-1} \mathbb{E}_{P_t}\left[e^{-\ell_t(\eta, i)}\right] \mathbb{E}_{P_T}\left[1 + \eta r_T(i)\right] = 0
\end{aligned}$$

EW is the solution

- Since the surrogate loss is 1-exp-concave there is no mixability gap to pay
- Running EW with constant learning rate 1 on ℓ_t achieves $S(Q) \leq \mathrm{KL}(Q \| P_1)$ for any Q

This gives

$$\sum_{t=1}^{T} \mathop{\mathbb{E}}_{Q}[\eta r_t(i)] \leq \sum_{t=1}^{T} \mathop{\mathbb{E}}_{Q}[\eta^2 r_t(i)^2] + \mathrm{KL}(Q || P1)$$

Regret Iprod

Theorem

If we use EW in the surrogate OCO task with learning rate 1 *and any product prior* $P_1 = \gamma \times \pi$ *for* γ *a distribution on* $\eta \in [0, \frac{1}{2}]$ *and* π *a distribution on* i*, and we take as comparator* $Q = \gamma(\eta \mid \eta \in [\hat{\eta}/2, \hat{\eta}]) \times \hat{\pi}$ *for any* $\hat{\eta} \in [0, \frac{1}{2}]$ *and distribution* $\hat{\pi}$ *on* i *that can both depend on all the losses, then*

$$\mathop{\mathbb{E}}_{\hat{\pi}}\left[\mathcal{R}_T(i)\right] \leq 2\hat{\eta} \mathop{\mathbb{E}}_{\hat{\pi}}[\mathcal{V}_T(i)] + \frac{2}{\hat{\eta}}\Big(\mathrm{KL}(\hat{\pi} \| \pi) - \ln \gamma([\hat{\eta}/2, \hat{\eta}])\Big). \quad (3)$$

After optimizing $\hat{\eta}$, this leads to an adaptive regret bound of

$$\mathop{\mathbb{E}}_{\hat{\pi}}\left[\mathcal{R}_T(i)\right] = O\left(\sqrt{\mathop{\mathbb{E}}_{\hat{\pi}}[\mathcal{V}_T(i)]\Big(\mathrm{KL}(\hat{\pi} \| \pi) + \ln \ln T\Big)}\right) \qquad \text{for all } \hat{\pi}$$

Squint

- EW with a continuous prior on η for the iProd surrogate losses requires evaluating a t-degree polynomial in η in every round: $O(T^2)$ total running time
- By choosing the slightly larger surrogate loss $\ell_t(\eta, i) = -\eta r_t(i) + \eta^2 r_t(i)^2$ we turn Iprod into Squint: $O(T)$ total running time
- Exactly the same regret guarantees as iProd:

$$\mathop{\mathbb{E}}_{\hat{\pi}}\left[\mathcal{R}_T(i)\right] = O\left(\sqrt{\mathop{\mathbb{E}}_{\hat{\pi}}[\mathcal{V}_T(i)]\Big(\mathrm{KL}(\hat{\pi}\|\pi) + \ln\ln T\Big)}\right) \qquad \text{for all } \hat{\pi}$$

Coin betting

We study a variant of the Coin Betting algorithm for experts of Orabona and Pál (2016)

- Idea: Split the learning of $\eta \in [0, 1]$ and i into separate steps
- restrict $P_t(\eta \mid i)$ to be a point mass on some η_t^i
- Choose η_t^i to achieve small regret for the surrogate loss

$$\ell_t^i(\eta) = -\frac{1 + r_t(i)}{2} \ln \frac{1+\eta}{2} - \frac{1 - r_t(i)}{2} \ln \frac{1-\eta}{2} - \ln 2$$

- Learn i for the surrogate losses $\tilde{\ell}_t(i) = -\ln(1 + \eta_t^i r_t(i))$

Mix-regret Coin Betting

For $\eta \in [0,1]$ and $\hat{\pi}$ a distribution on i, let

$$S_T^i(\eta) = \sum_{t=1}^T \ell_t^i(\eta_t^i) - \sum_{t=1}^T \ell_t^i(\eta) \quad \text{(regret log loss)}$$

$$\tilde{S}_T(\hat{\pi}) = \sum_{t=1}^T - \ln \mathop{\mathbb{E}}_{i \sim P_t}\left[e^{-\tilde{\ell}_t(i)}\right] - \mathop{\mathbb{E}}_{\hat{\pi}}\left[\sum_{t=1}^T \tilde{\ell}_t(i)\right]$$

be the mix-regret in the two surrogate OCO tasks.

Regret Coin Betting

Theorem

If we use EW with learning rate 1 *and prior* π *on* i *for the losses* $\tilde{\ell}_t$*, and for the losses* ℓ_t^i *we let* η_t^i *be the mean of lazy EW with learning rate* 1 *and with prior on* $\eta \in [-1, +1]$ *such that* $\frac{1+\eta}{2}$ *has a beta-distribution* $\beta(a, a)$ *with* $a = \frac{T}{4} + \frac{1}{2}$ *and with projections onto* $\mathcal{P} = \{P \mid \mathbb{E}_P[\eta] \in [0, 1]\}$*, then*

$$\mathop{\mathbb{E}}_{\hat{\pi}}[\mathcal{R}_T(i)] \leq \sqrt{3T\,(\mathrm{KL}(\hat{\pi}\|\pi) + 3)} \qquad \text{for any } \hat{\pi} \text{ on } i.$$

Resulting algorithm

- EW on ℓ_t^i with the (conjugate) $\beta(a, a)$ prior is a generalization of the Krichevsky-Trofimov estimator with mean $\frac{\mathcal{R}_{t-1}(i)}{t-1+2a}$
- Lazily projecting onto $\eta \in [0, 1]$ simply amounts to clipping at 0
- Combining the above we get $\eta_t^i = \max\left\{\frac{\mathcal{R}_{t-1}(i)}{t-1+2a}, 0\right\}$
- Predict with $\boldsymbol{w}_t = \frac{\mathbb{E}_{P_t(i)}[\eta_t^i e_i]}{\mathbb{E}_{P_t(i)}[\eta_t^i]}$

Interpretation Regret Coin Betting

We can now explain three design choices by Orabona and Pál (2016):

- δ-shifted KT-Potential function: naturally arises in our proof when we bound the regret $S^i_T(\mathcal{R}^+_T(i)/T)$ for EW
- The choice for δ, which is simply specifying a prior with most mass in a region of order $1/\sqrt{T}$ around $\eta = 0$
- The clipping of the unnormalized weights $\tilde{p}_t(i)\eta^i_t$ to 0 when $\mathcal{R}_{t-1}(i) < 0$, which in our presentation happens automatically

Organization

Online Linear Optimization with Bandit Feedback

- Linear losses $f_t(\boldsymbol{w}) = \langle \boldsymbol{w}, \boldsymbol{g}_t \rangle \in [-1, +1]$
- Instead of seeing the vectors $\boldsymbol{g}_t$ we only observe the loss $f_t(\boldsymbol{w}_t)$ for the algorithm's choice $\boldsymbol{w}_t$
- The algorithm is allowed to randomize its choice $\boldsymbol{w}_t$
- Goal: minimize the expected regret $\mathbb{E}[\mathcal{R}_T(\boldsymbol{u})]$, where the expectation is with respect to the algorithm's randomness

Standard solution SCRiBLe (Abernethy et al., 2012):

- Mirror Descent with self concordant barrier function F^*
- Sample based on the spectrum of the Hessian of F^*

Sampling from EW distribution

We consider the EW algorithm with fixed learning rate η and uniform prior distribution P_1 over $\mathcal{W}$. Let R be a fixed "exploration" distribution chosen to be *John's exploration*.

for $t = 1, 2, \ldots, T$ **do**

  Sample $\boldsymbol{w}_t \sim Q_t$, where $Q_t = (1 - \gamma) P_t + \gamma R$

  Observe $f_t(\boldsymbol{w}_t) = \langle \boldsymbol{w}_t, \boldsymbol{g}_t \rangle$

  Constructs a random unbiased estimate $\tilde{\boldsymbol{g}}_t$

  Update P_t to P_{t+1} based on $\tilde{f}_t(\boldsymbol{w}_t) = \langle \boldsymbol{w}_t, \tilde{\boldsymbol{g}}_t \rangle$

Linear Bandits with EW

When η and γ are appropriately chosen, this algorithm achieves expected regret of order $O(d\sqrt{T \ln T})$, which is the best known expected regret.

Compared to SCRiBLe:

- Instead of sampling from the spectrum of the Hessian we sample from the EW distribution
- We achieve a regret bound that is a factor $O(\sqrt{d})$ better
- A proof outline of this fact was given by Bubeck and Eldan (2015); to complete our story of EW we spell out the proof details.

Concluding Remarks

On average, we all live in an exponentially weighted world.