Follow the Perturbed Leader

Nicolò Cesa-Bianchi

Notes for the PhD Course — October 2, 2017

The set of actions is $\mathcal{S} \subseteq \{0,1\}^d$. We assume there is a fixed but unknown sequence of loss vectors $\ell_1, \ell_2, \ldots \in [0,1]^d$. Let also

$$L_t = \sum_{s=1}^{t} \ell_s \ .$$

Let $Z_1, Z_2, \dots \in \mathbb{R}^d$ be i.i.d. random vectors with Laplace density $f(z) = \frac{\eta}{2}\exp\bigl(-\eta\|z\|_1\bigr)$. Let $X_1, X_2, \ldots \in \mathcal{S}$ be the sequence of actions chosen by FPL, where

$$X_t = \operatorname*{argmin}_{x \in \mathcal{S}} x^\top \bigl(L_{t-1} + Z_t\bigr) \ .$$

We start with a preliminary lemma.

Lemma 1 (FTL-BTL). *Let $\ell_1, \ell_2, \ldots$ an arbitrary sequence of losses and let*

$$\widehat{x}_t = \operatorname*{argmin}_{x \in \mathcal{S}} x^\top L_t$$

Then

$$\sum_{t=1}^{T} \ell_t^\top \widehat{x}_t \le \sum_{t=1}^{T} \ell_t^\top \widehat{x}_T = \min_{x \in \mathcal{S}} x^\top L_T$$

Proof. The statement is proven by induction on T. The case $T = 1$ is obvious. Assume now that

$$\sum_{t=1}^{T-1} \ell_t^\top \widehat{x}_t \le \sum_{t=1}^{T-1} \ell_t^\top \widehat{x}_{T-1}$$

Since by definition

$$\sum_{t=1}^{T-1} \ell_t^\top \widehat{x}_{T-1} \le \sum_{t=1}^{T-1} \ell_t^\top \widehat{x}_T$$

the inductive assumption implies

$$\sum_{t=1}^{T-1} \ell_t^\top \widehat{x}_t \le \sum_{t=1}^{T-1} \ell_t^\top \widehat{x}_T$$

Adding $\ell_T^\top \widehat{x}_T$ on both sides gives the result. □

We now prove a bound on the regret of FPL, defined by

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \ell_t^\top X_t\right] - \min_{x\in\mathcal{S}} L_T^\top x \ .$$

Theorem 2. *The regret of FPL is bounded by* $R_T \le 4D\sqrt{dT}$ *where* $D = \max_{x\in\mathcal{S}} \|x\|_1$.

Proof. Introduce $\widehat{X}_t = \operatorname*{argmin}_{x\in\mathcal{S}} x^\top\big(L_t + Z_t\big)$. We have

$$\mathbb{E}\left[\sum_{t=1}^{T} \ell_t^\top X_t\right] = \mathbb{E}\left[\sum_{t=1}^{T}\big(\ell_t^\top X_t - \ell_t^\top \widehat{X}_t\big)\right] + \mathbb{E}\left[\sum_{t=1}^{T} \ell_t^\top \widehat{X}_t\right] \tag{1}$$

The second term in (1) is bounded as follows. Note that

$$\widehat{X}_t = \operatorname*{argmin}_{x\in\mathcal{S}} x^\top\big(L_t + Z_t\big) = \operatorname*{argmin}_{x\in\mathcal{S}} \sum_{s=1}^{t} x^\top\big(\ell_s + Z_s - Z_{s-1}\big)$$

where $Z_0 = (0,\dots,0)$. Now apply the FTL-BTL lemma to the losses $\ell_t' = \ell_t + Z_t - Z_{t-1}$ and obtain

$$\begin{aligned}\sum_{t=1}^{T} \widehat{X}_t^\top\big(\ell_t + Z_t - Z_{t-1}\big) &\le \min_{x\in\mathcal{S}} \sum_{t=1}^{T} x^\top\big(\ell_t + Z_t - Z_{t-1}\big)\\ &= \min_{x\in\mathcal{S}} x^\top\big(L_T + Z_T\big)\\ &\le \min_{x\in\mathcal{S}} x^\top L_T + x_0^\top Z_T\end{aligned}$$

where $x_0 = \operatorname*{argmin}_{x\in\mathcal{S}} x^\top L_T$. Hence,

$$\begin{aligned}\sum_{t=1}^{T} \widehat{X}_t^\top \ell_t &\le \min_{x\in\mathcal{S}} x^\top L_T + x_0^\top Z_T + \sum_{t=1}^{T} \widehat{X}_t^\top\big(Z_{t-1} - Z_t\big)\\ &\le \min_{x\in\mathcal{S}} x^\top L_T + \max_{x\in\mathcal{S}} \|x\|_1 \left(\|Z_T\|_\infty + \sum_{t=1}^{T} \|Z_{t-1} - Z_t\|_\infty\right)\end{aligned}$$

By letting $X_t^* = \operatorname*{argmin}_{x\in\mathcal{S}} x^\top\big(L_t + Z^*\big)$, where Z^* has the same distribution as each Z_t,

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T} \ell_t^\top \widehat{X}_t\right] &= \sum_{t=1}^{T} \mathbb{E}\big[\ell_t^\top \widehat{X}_t\big]\\ &= \sum_{t=1}^{T} \mathbb{E}\big[\ell_t^\top X_t^*\big]\\ &\le \min_{x\in\mathcal{S}} x^\top L_T + \max_{x\in\mathcal{S}} \|x\|_1\, \mathbb{E}\Big[\|Z^*\|_\infty + \|Z^*\|_\infty\Big]\end{aligned}$$

where we used

$$\sum_{t=1}^{T} \|Z_{t-1} - Z_t\|_\infty = \|Z^*\|_\infty$$

when $Z_0 = (0, \dots, 0)$ and $Z_1 = \cdots = Z_T = Z^*$. This gives

$$\mathbb{E}\left[\sum_{t=1}^{T} \ell_t^\top \widehat{X}_t\right] \le \min_{x \in \mathcal{S}} x^\top L_T + 2D\,\mathbb{E}\big[\|Z^*\|_\infty\big] \ .$$

Using standard probability facts, we further bound

$$\mathbb{E}\big[\|Z^*\|_\infty\big] \le \frac{2}{\eta}(1 + \ln d) \ .$$

Therefore,

$$\mathbb{E}\left[\sum_{t=1}^{T} \ell_t^\top \widehat{X}_t\right] \le \min_{x \in \mathcal{S}} x^\top L_T + \frac{4D}{\eta}(1 + \ln d) \ .$$

In order to bound the first term in (1), introduce the function

$$F_t(z) = x_t(z)^\top \ell_t \qquad \text{where} \qquad x_t(z) = \operatorname*{argmin}_{x \in \mathcal{S}} x^\top \big(L_{t-1} + z\big) \ .$$

This allows us to write

$$\begin{aligned} \mathbb{E}\big[\ell_t^\top X_t\big] &= \int_{z \in \mathbb{R}^d} \big(x_t(z)^\top \ell_t\big) f(z) dz = \int_{z \in \mathbb{R}^d} F_t(z) f(z) dz \\ \mathbb{E}\big[\ell_t^\top \widehat{X}_t\big] &= \int_{z \in \mathbb{R}^d} F_t(\ell_t + z) f(z) dz = \int_{z \in \mathbb{R}^d} F_t(z') f(z' - \ell_t) dz' \end{aligned}$$

where in the last step we performed the change of variable $z' = \ell_t + z$. This gives us

$$\mathbb{E}\left[\sum_{t=1}^{T} \Big(\ell_t^\top X_t - \ell_t^\top \widehat{X}_t\Big)\right] = \sum_{t=1}^{T} \int_{z' \in \mathbb{R}^d} F_t(z) \Big(f(z) - f(z - \ell_t)\Big) dz \ .$$

We start by bounding the difference in the integral

$$\begin{aligned} f(z) - f(z - \ell_t) &= f(z)\left(1 - \frac{f(z - \ell_t)}{f(z)}\right) \\ &= f(z)\Big[1 - e^{-\eta\big(\|z - \ell_t\|_1 - \|z\|_1\big)}\Big] \\ &\le f(z)\Big[1 - \big(1 - \eta\|z - \ell_t\|_1 + \eta\|z\|_1\big)\Big] \\ &\le f(z)\eta\Big(\|z - \ell_t\|_1 - \|z\|_1\Big) \\ &\le f(z)\eta\|\ell_t\|_1 \\ &\le f(z)\eta d \end{aligned}$$

where we used $e^{-x} \geq 1 - x$ and the triangular inequality $\|z - \ell_t\|_1 \leq \|z\|_1 + \|\ell_t\|_t$. Using the positivity of $F_t(z)$, we can thus write

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}\left(\ell_t^\top X_t - \ell_t^\top \widehat{X}_t\right)\right] &\leq \eta d\, \mathbb{E}\left[\sum_{t=1}^{T}\int F_t(z) f(z)\right] \\ &= \eta d\, \mathbb{E}\left[\sum_{t=1}^{T}\ell_t^\top X_t\right] \\ &\leq \eta d D T \ .\end{aligned}$$

Summarizing, we have

$$\mathbb{E}\left[\sum_{t=1}^{T}\ell_t^\top X_t\right] \leq \min_{x \in \mathcal{S}} x^\top L_T + \frac{4D}{\eta}(1 + \ln d) + \eta d D T \ .$$

□