AI4010: Online Learning
Second Midterm Exam
Oct 2025

Instructions:

- *The total number of marks is 20.*
- *The total duration of the exam is 90 minutes. No electronic aids are allowed. You can keep a maximum of one sheet of paper with formulas/notes.*
- *All questions are mandatory. A yes/no answer without proper proof or justification will be given zero marks even if it is correct.*
- *Any plagiarism case, if detected, will attract F grade in the course irrespective of overall performance.*
- *Use $0\log(0) = 0$ wherever required.*

Problem 0.1 (2 marks). *Consider that the loss functions are bounded in range $[0, 1]$. Show that the regret of FTL algorithm is upper bounded by the number of times the* leader[1] *is changed during the sequence of plays.*

Solution. From FTL-BTL lemma we have,

$$\mathcal{R}_T(FTL) \leq \sum_{t=1}^{T}(f_t(x_t) - f_t(x_{t+1})) = \sum_{t:x_t \neq x_{t+1}} (f_t(x_t) - f_t(x_{t+1}))$$

$$\leq \sum_{t:x_t \neq x_{t+1}} 1 = \text{\# times the leader changes}$$

The last inequality follows from the fact that loss function is bounded $[0, 1]$. ■

Problem 0.2 (3 marks). *(Batch vs Online convex optimization) Consider a standard batch convex optimization problem: you want to find a minimum of the convex function $f : \mathbb{R} \to \mathbb{R}_+ \setminus \{\infty\}$ over a bounded convex set $\mathcal{K} \subset \mathbb{R}$ to within an accuracy $\varepsilon > 0$. In other words, you must output $x \in \mathcal{K}$ satisfying $f(x) \leq \min_{y \in \mathcal{K}} f(y) + \varepsilon$.*

[1] The leader at time t is a choice x_{t+1} i.e. optimal choice of the next round.

You are given an online algorithm ALG *with the following property. For any number of rounds $t \geq 1$ and any sequence of non-negative loss functions $\{f_s\}_{s=1}^t$ from a family of convex functions $\mathcal{F}$, the algorithm's regret over* [2] *a convex set $\mathcal{K}$ is at t^α for some $\alpha \in [0, 1)$; i.e.* ALG *satisfies sublinear regret guarantee. How will you accomplish batch optimization objective using* ALG*? Show your work.*

Hint: Use Jensen's inequality.

Solution.

We know from the regret guarantee of ALG that the below equation holds for every $t \geq 1$

$$\sum_{s=1}^{t} f_s(w_s) - \sum_{s=1}^{t} f_s(w^*) \leq t^\alpha$$

Here $w^* = \arg\min_{w \in \mathcal{K}} \sum_{s=1}^{t} f_s(w^*)$.

Now feed the convex optimization algorithm ALG, $f_s = f$ for all s and observe the sequence of choices of ALG, $\{w_s\}_{s \geq 1}$. The regret guarantee then becomes

$$\sum_{s=1}^{t} f(w_s) - t \cdot f(w^*) \leq t^\alpha$$

Let $w^t := \frac{1}{t}\sum_{s=1}^{t} w_s$ and notice that since $\mathcal{K}$ is convex, $w^t \in \mathcal{K}$. Furthermore we have from Jensens inequality that $f(w^t) = f(\frac{1}{t}\sum_{s=1}^{t} w_s) \leq \frac{1}{t}\sum_{s=1}^{t} f(w_s)$. Hence for any t

$$f(w^t) \leq f(w^*) + t^{\alpha-1}$$

Notice that the above inequality holds for any $t \geq 1$. Set the stopping time $T = \lceil (1/\varepsilon)^{1-\alpha} \rceil$ and output the solution $x = w^T = \frac{1}{T}\sum_{s=1}^{T} w_s$. ■

Problem 0.3 (2+2+2 = 6 marks). *Compute the Bregman Divergence of*

1. $R(x) = \frac{1}{2}\|x\|_2^2$ *for* $x \in \mathbb{R}^d$
2. $R(x) = -2\sum_{i=1}^{d} \sqrt{x_i}$ *for* $x \in (0, +\infty)^d$

[2] The regret is computed with respect to any single point in $\mathcal{K}$.

3. $R(x) = \sum_{i=1}^{d} x_i(\log(x_i) - 1)$ *for* $x \in \Delta_d$.

Solution.

1.

$$B_R(x||y) = \frac{1}{2}||x||_2^2 - \frac{1}{2}||y||_2^2 - \langle y, x - y\rangle$$
$$= \frac{1}{2}||x||_2^2 + \frac{1}{2}||y||_2^2 - \langle y, x\rangle = \frac{1}{2}||x - y||_2^2$$

2.

$$B_R(x||y) = -2\sum_{i=1}^{d}\sqrt{x_i} + 2\sum_{i=1}^{d}\sqrt{y_i} + \sum_{i=1}^{d}\frac{x_i - y_i}{\sqrt{y_i}}$$
$$= \sum_{i=1}^{d}\frac{x_i + y_i - 2\sqrt{x_i y_i}}{\sqrt{y_i}} = \sum_{i=1}^{d}\frac{(\sqrt{x_i} - \sqrt{y_i})^2}{\sqrt{y_i}}$$

3.

$$B_R(x||y) = \sum_{i=1}^{d} x_i\log(x_i) - \sum_{i=1}^{d} x_i - \sum_{i=1}^{d} y_i\log(y_i) + \sum_{i=1}^{d} y_i - \sum_{i=1}^{d}\log(y_i)(x_i - y_i)$$
$$= \sum_{i=1}^{d} x_i\log(x_i) - \sum_{i=1}^{d} x_i\log(y_i) = \sum_{i=1}^{d} x_i\log(\frac{x_i}{y_i}) = KL(x||y).$$

∎

Problem 0.4 (Dual Function). *In class we saw that the OMD algorithm*[3] *performs the update in the dual space defined by the gradients. This can be interpreted (slightly more elaborately) as follows. OMD first maps the point* x_t *from primal space (i.e. x-space) to the dual space defined by* $\nabla R(.)$ *i.e. computes* $\theta_t = \nabla R(x_t)$, *performs the (gradient) update in gradient space i.e. computes* $\theta_{t+1} = \theta_t - \eta_t\nabla_t$, *maps the updated point using* $(\nabla R)^{-1}$ *back to the original space i.e. computes* y_{t+1} *such that* $\nabla R(y_{t+1}) = \theta_{t+1}$ *i.e.* $y_{t+1} = (\nabla R)^{-1}(\theta_{t+1})$ *and then take the Bregman projection onto the convex set i.e. compute* $x_{t+1} = \arg\min_{x\in\mathcal{K}} B_R(x||y_{t+1})$.

Define a function $R^*(\theta) = \sup_{x\in\mathbb{R}^d}(\langle x, \theta\rangle - R(x))$. *The function* $R^*(.)$ *is called as the Fenchel dual/conjugate of* R. *In this question we will show that* $\nabla R^* = (\nabla R)^{-1}$.

[3]Here, we will consider all assumptions we made for OMD.

1. *Write the Fenchel conjugates of below functions. [2+2 = 4 marks]*

   (a) $R(x) = \log(\sum_{i=1}^{d} e^{x_i})$ *(use* $0\log(0) = 0$*)*

   (b) $R(x) = \frac{1}{2}x^T Qx$ *where* $Q \in \mathbb{R}^{n\times n}$ *is a positive definite matrix.*

2. *Show that the following two conditions are equivalent. [2 marks]*

   (a) $R(u) + R^*(v) = \langle u, v\rangle$

   (b) $v = \nabla R(u)$

3. *Finally, using the above result and also using the fact that* $R^{**} = R$, *show that* $u = \nabla R^*(\nabla R(u))$ *and* $u = \nabla R(\nabla R^*(u))$. *[3 marks]*

Solution.

1. a First we determine the value of θ for which the supremum over x $\langle\theta, x\rangle - R(x)$ is attained. This gives

$$\theta_i = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{0.1}$$

   The above equations are simultenously solvable (for all i) for x iff $\theta_i \geq 0$ for all i and $\sum_i \theta_i = 1$. By substitution we obtain $R^*(\theta) = \sum_i y_i \log(y_i)$ (interpret $0\log(0) = 0$)

   b Notice that Q^{-1} exists since Q is positive definite. Observe that $\frac{d}{dx}(\langle x, \theta\rangle - \frac{1}{2}x^T Qx) = 0 \implies x^* = Q^{-1}\theta$. We obtain $R^*(\theta) = \frac{1}{2}\theta^T Q^{-1}\theta$.

2. Fix u, v and let $h(u) = \langle u, v\rangle - R(u)$. From condition 1 we have

$$R^*(v) = h(u) \tag{0.2}$$

   and by definition, we have $R^*(v) = \sup_u h(u)$. From the fact that R is strongly convex we have that h is strictly concave. Hence, $\sup_u h(u)$ is unique and condition 1 holds iff $\nabla h(u) = 0$. This is equivalent to $v - \nabla R(u) = 0$ i.e. the second condition.

3. Let u and $u' = \nabla R(u)$. By the above result we have $R(u) + R^*(u') = \langle u, u'\rangle$). Since $R^{**} = R$, we have

$$R^{**}(u) + R^*(u') = \langle u, u'\rangle$$

Applying the previous lemma to R^* in place of R, we have

$$u = \nabla R^*(u')$$

This shows that $u = \nabla R^*(\nabla R(u))$. In other words, ∇R^* is the right inverse of ∇R. Since ∇R is a surjection , ∇R^* must be also its left inverse.

■