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.*

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 $f(x) \leq \min_{y \in \mathcal{K}} f(y) + \varepsilon$.*

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.

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* $x \in (0, +\infty)^d$

---

[1] The leader at time t is a choice x_{t+1} i.e. optimal choice of the next round.
[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* $x \in \Delta_d$.

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* $\nabla R^* = (\nabla R)^{-1}$.

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) $R(x) = \frac{1}{2} x^T Q x$ *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]*

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