Advanced Optimization (2024 Fall)
Homework #2

Student ID, Name, Email

December 1, 2024

Evaluation: There is a problem section (in total 5 problems, 270pts) and a bonus section (5pts), and your score is the sum of the problem section and the bonus section. The scoring method for the problem section is as follows: Problem 1 (70pts) is asked to solve. Choose 3 of the remaining 4 problems (each with 50pts) to finish. There are two options for the final score evaluation of the problem section:

1. (**recommended**) If you choose 4 problems (Problem 1 + 3 selected ones, totally 220pts), you can obtain the full score (200pts) once you achieve at least 200pts.
2. If you choose 4 problems (totally 220pts) *and finish the remaining one (50pts)*:
   (a) If you haven't achieved 200pts on the chosen 4 problems, back to Case 1.
   (b) If you obtain $(245 + X)$pts ($X \geq 0$), the final score will be $(200 + X)$pts.

Attention: You are requested to indicate selected problem ids clearly.

My selected problem ids: 1,x,x,x.

```
% replace x,x,x by selected ids (e.g., 2,3,4,5)
% x,x,x = 2,3,4 by default if not explicitly specified
```

1 [70pts] OOMD for Game and Implementation

We consider a three-player game, where the strategies of three players are represented by $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$. We consider the game repeated T times. In round t, after all three players ***simultaneously*** submit their strategies $(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t)$, each player's individual ***cost*** is calculated using their own cost function. For example, $\mathbf{x}$-player' cost function is denoted as $\mathcal{G}^{\mathbf{x}} : (\mathbf{x}, \mathbf{y}, \mathbf{z}) \mapsto \mathbb{R}$, and $\mathcal{G}^{\mathbf{y}} : (\mathbf{x}, \mathbf{y}, \mathbf{z}) \mapsto \mathbb{R}$ for $\mathbf{y}$-player, $\mathcal{G}^{\mathbf{z}} : (\mathbf{x}, \mathbf{y}, \mathbf{z}) \mapsto \mathbb{R}$ for $\mathbf{z}$-player.

Let $\mathcal{G}(\cdot) \triangleq \mathcal{G}^{\mathbf{x}}(\cdot) + \mathcal{G}^{\mathbf{y}}(\cdot) + \mathcal{G}^{\mathbf{z}}(\cdot)$ denote the total cost for the three players. Ideally, we hope all players can cooperate to achieve the minimum total cost $\mathcal{G}^{\textsc{Min}} \triangleq \min_{\mathbf{x},\mathbf{y},\mathbf{z}} \mathcal{G}(\mathbf{x}, \mathbf{y}, \mathbf{z})$. However, a more likely scenario is that each player selfishly tries to minimize their own cost during the game. In this problem, we focus on a quantity: the average total cost of all players, i.e., $\overline{\mathcal{G}}_T \triangleq \frac{1}{T}\sum_{t=1}^{T} \mathcal{G}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t)$, and are concerned with the following question :

What condition can a game satisfy to ensure $\overline{\mathcal{G}}_T$ isn't much worse than $\mathcal{G}^{\textsc{Min}}$?

We focus on the *smooth games* defined as follows for simplicity.

Assumption 1 (Smooth Games)**.** For the game $\mathcal{G}$, it is called a (λ, μ)-smooth game with $\lambda > 0$ and $\mu < 1$, if there exists a strategy profile $(\mathbf{x}^\star, \mathbf{y}^\star, \mathbf{z}^\star)$ such that the following holds for any strategies $(\mathbf{x}, \mathbf{y}, \mathbf{z})$:

$$\mathcal{G}^{\mathbf{x}}(\mathbf{x}^\star, \mathbf{y}, \mathbf{z}) + \mathcal{G}^{\mathbf{y}}(\mathbf{x}, \mathbf{y}^\star, \mathbf{z}) + \mathcal{G}^{\mathbf{z}}(\mathbf{x}, \mathbf{y}, \mathbf{z}^\star) \le \lambda \cdot \mathcal{G}^{\textsc{Min}} + \mu \cdot \mathcal{G}(\mathbf{x}, \mathbf{y}, \mathbf{z}). \tag{1.1}$$

Intuitively, in smooth games, any player using her optimal strategy continues to do well, irrespective of other players' strategies.

In the following problems, define $\textsc{Reg}_T^{\mathbf{x}} \triangleq \max_{\mathbf{x}} \sum_{t=1}^{T} \big(\mathcal{G}^{\mathbf{x}}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t) - \mathcal{G}^{\mathbf{x}}(\mathbf{x}, \mathbf{y}_t, \mathbf{z}_t)\big)$, $\textsc{Reg}_T^{\mathbf{y}} \triangleq \max_{\mathbf{y}} \sum_{t=1}^{T} \big(\mathcal{G}^{\mathbf{y}}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t) - \mathcal{G}^{\mathbf{y}}(\mathbf{x}_t, \mathbf{y}, \mathbf{z}_t)\big)$, and $\textsc{Reg}_T^{\mathbf{z}}$ is similarly defined.

(1) **[10pts]** With (1.1), try to <u>prove</u> the following guarantees:

$$\frac{1}{T}\sum_{t=1}^{T} \mathcal{G}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t) \le \frac{\lambda}{1-\mu}\mathcal{G}^{\textsc{Min}} + \frac{1}{(1-\mu)T}\big(\textsc{Reg}_T^{\mathbf{x}} + \textsc{Reg}_T^{\mathbf{y}} + \textsc{Reg}_T^{\mathbf{z}}\big),$$

which means with sublinear regrets, we have the guarantee $\lim_{T\to\infty} \overline{\mathcal{G}}_T \le \frac{\lambda}{1-\mu}\mathcal{G}^{\textsc{Min}}$, thereby answering the question posed above.

(2) **[10pts]** In the class, we have learned that Optimistic Online Mirror Descent (OOMD) can lead to fast-rate convergence for two-player zero-sum games. We now consider the three-player game in this problem and assume that each player picks a mixed strategy from Δ_d. Each player has her own tensor to measure cost, that is, $\mathbf{x}$-player has $G^{\mathbf{x}} \in [0,1]^{d\times d\times d}$, $\mathbf{y}$-player has $G^{\mathbf{y}} \in [0,1]^{d\times d\times d}$, and $\mathbf{z}$-player has $G^{\mathbf{z}} \in [0,1]^{d\times d\times d}$. For the tensor $G \in \mathbb{R}^{d\times d\times d}$ and three vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^d$, We abbreviate $\sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{k=1}^{d} G_{i,j,k}\mathbf{x}_i\mathbf{y}_j\mathbf{z}_k$ as $G[\mathbf{x}, \mathbf{y}, \mathbf{z}]$. Then, the cost functions for the three players are specified as:

$$\mathcal{G}^{\mathbf{x}}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \triangleq G^{\mathbf{x}}[\mathbf{x}, \mathbf{y}, \mathbf{z}], \quad \mathcal{G}^{\mathbf{y}}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \triangleq G^{\mathbf{y}}[\mathbf{x}, \mathbf{y}, \mathbf{z}], \quad \mathcal{G}^{\mathbf{z}}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \triangleq G^{\mathbf{z}}[\mathbf{x}, \mathbf{y}, \mathbf{z}].$$

In round t, after the three players submit their strategies $\mathbf{x}_t, \mathbf{y}_t$ and $\mathbf{z}_t$, they can observe the gradient of their own cost functions, which is $\nabla_{\mathbf{x}}\mathcal{G}^{\mathbf{x}}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t)$ for $\mathbf{x}$-player, $\nabla_{\mathbf{y}}\mathcal{G}^{\mathbf{y}}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t)$ for $\mathbf{y}$-player, and $\nabla_{\mathbf{z}}\mathcal{G}^{\mathbf{z}}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t)$ for $\mathbf{z}$-player.

Design an OOMD algorithm with NE-entropy for each of the three players, prove that:

$$\textsc{Reg}_T^{\mathbf{x}} \lesssim \frac{1}{\eta^{\mathbf{x}}} + \eta^{\mathbf{x}} \sum_{t=2}^{T} \|\nabla_{\mathbf{x}}\mathcal{G}^{\mathbf{x}}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t) - \nabla_{\mathbf{x}}\mathcal{G}^{\mathbf{x}}(\mathbf{x}_{t-1}, \mathbf{y}_{t-1}, \mathbf{z}_{t-1})\|_\infty^2 - \frac{1}{\eta^{\mathbf{x}}} \sum_{t=2}^{T} \|\mathbf{x}_t - \mathbf{x}_{t-1}\|_1^2,$$

$$\textsc{Reg}_T^{\mathbf{y}} \lesssim \frac{1}{\eta^{\mathbf{y}}} + \eta^{\mathbf{y}} \sum_{t=2}^{T} \|\nabla_{\mathbf{y}}\mathcal{G}^{\mathbf{y}}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t) - \nabla_{\mathbf{y}}\mathcal{G}^{\mathbf{y}}(\mathbf{x}_{t-1}, \mathbf{y}_{t-1}, \mathbf{z}_{t-1})\|_\infty^2 - \frac{1}{\eta^{\mathbf{y}}} \sum_{t=2}^{T} \|\mathbf{y}_t - \mathbf{y}_{t-1}\|_1^2,$$

$$\textsc{Reg}_T^{\mathbf{z}} \lesssim \frac{1}{\eta^{\mathbf{z}}} + \eta^{\mathbf{z}} \sum_{t=2}^{T} \|\nabla_{\mathbf{z}}\mathcal{G}^{\mathbf{z}}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t) - \nabla_{\mathbf{z}}\mathcal{G}^{\mathbf{z}}(\mathbf{x}_{t-1}, \mathbf{y}_{t-1}, \mathbf{z}_{t-1})\|_\infty^2 - \frac{1}{\eta^{\mathbf{z}}} \sum_{t=2}^{T} \|\mathbf{z}_t - \mathbf{z}_{t-1}\|_1^2,$$

where $\eta^{\mathbf{x}}, \eta^{\mathbf{y}}$ and $\eta^{\mathbf{z}}$ are the constant step-sizes for each of the three algorithms. We use $\lesssim$ to denote "asymptotically smaller than" by dropping constant factors.

(3) **[10pts]** Prove the following inequality:

$$\|\nabla_{\mathbf{x}}\mathcal{G}^{\mathbf{x}}(\mathbf{x}_t, \mathbf{y}_t, \mathbf{z}_t) - \nabla_{\mathbf{x}}\mathcal{G}^{\mathbf{x}}(\mathbf{x}_t, \mathbf{y}_{t-1}, \mathbf{z}_t)\|_\infty^2 \leq \|\mathbf{y}_t - \mathbf{y}_{t-1}\|_1^2.$$

Then design the step-sizes $\eta^{\mathbf{x}}, \eta^{\mathbf{y}}, \eta^{\mathbf{z}}$, and prove the following guarantee:

$$\textsc{Reg}_T^{\mathbf{x}} + \textsc{Reg}_T^{\mathbf{y}} + \textsc{Reg}_T^{\mathbf{z}} \leq \mathcal{O}(1).$$

(4) **[40pts]** Implement the OMD and OOMD algorithms to solve the game mentioned above, and attach the figure comparing the average total cost curves of the two algorithms here. Detailed instructions are available in the `AOpt-Lab2/AOpt-Lab2.ipynb` jupyter notebook. Submit `AOpt-Lab2.ipynb` file along with your homework. Make sure the results can be checked.

Solution. Give your answers here. (中英文均可)

2 [50pts] Accelerated Composite Optimization

Consider the following composite optimization within a bounded domain:

$$\min_{\mathbf{x}\in\mathcal{X}} F(\mathbf{x}) \triangleq f(\mathbf{x}) + h(\mathbf{x}),$$

where both $f(\cdot)$ and $h(\cdot)$ are convex, and $f(\cdot)$ is L-smooth w.r.t. $\|\cdot\|_2$, whereas $h(\cdot)$ is not. We assume that the domain diameter is bounded by D, i.e., $\sup_{\mathbf{x},\mathbf{y}\in\mathcal{X}}\|\mathbf{x}-\mathbf{y}\|_2 \leq D$.

In class, we have learned a simple accelerated method for smooth convex optimization building on the general framework of optimistic online learning. Can the same approach be applied to the composite optimization?

More specifically, we consider the following weighted online-to-batch conversion:

$$\overline{\mathbf{x}}_t = \frac{1}{A_t}\sum_{s=1}^{t}\alpha_s\mathbf{x}_s, \text{ with } A_t = \sum_{s=1}^{t}\alpha_s \text{ and } \alpha_t > 0, \forall t \in [T]. \tag{2.1}$$

(1) **[10pts]** Try to prove that (2.1) ensures the following reduction:

$$F(\overline{\mathbf{x}}_T) - F(\mathbf{x}^\star) \leq \frac{\sum_{t=1}^{T}\big(\langle\alpha_t\nabla f(\overline{\mathbf{x}}_t), \mathbf{x}_t - \mathbf{x}^\star\rangle + \alpha_t h(\mathbf{x}_t) - \alpha_t h(\mathbf{x}^\star)\big)}{A_T}. \tag{2.2}$$

(2) **[20pts]** The inequality (2.2) allows us to reduce offline optimization as an online one. Define the online function as $F_t(\mathbf{x}) \triangleq f_t(\mathbf{x}) + h_t(\mathbf{x})$, where $f_t(\mathbf{x}) \triangleq \langle\alpha_t\nabla f(\overline{\mathbf{x}}_t), \mathbf{x}\rangle, h_t(\mathbf{x}) \triangleq \alpha_t h(\mathbf{x})$. To this end, we design the following optimistic online learning algorithm:

$$\mathbf{x}_t = \arg\min_{\mathbf{x}\in\mathcal{X}}\Big\{\eta\big(\langle M_t, \mathbf{x}\rangle + h_t(\mathbf{x})\big) + \frac{1}{2}\|\mathbf{x} - \widehat{\mathbf{x}}_t\|_2^2\Big\} \tag{2.3}$$

$$\widehat{\mathbf{x}}_{t+1} = \arg\min_{\mathbf{x}\in\mathcal{X}}\Big\{\eta\big(\langle\nabla f_t(\mathbf{x}_t), \mathbf{x}\rangle + h_t(\mathbf{x})\big) + \frac{1}{2}\|\mathbf{x} - \widehat{\mathbf{x}}_t\|_2^2\Big\} \tag{2.4}$$

(2.i) **[10pts]** Prove the stability property for the updates (2.3) and (2.4), that is

$$\|\mathbf{x}_t - \widehat{\mathbf{x}}_{t+1}\|_2 \leq \eta\|\nabla f_t(\mathbf{x}_t) - M_t\|_2.$$

(2.ii) **[10pts]** Prove the Bregman proximal inequality for the update (2.4):

$$\eta\langle\nabla f_t(\mathbf{x}_t) + \nabla h_t(\widehat{\mathbf{x}}_{t+1}), \widehat{\mathbf{x}}_{t+1} - \mathbf{x}^\star\rangle \leq \frac{1}{2}\|\mathbf{x}^\star - \widehat{\mathbf{x}}_t\|_2^2 - \frac{1}{2}\|\mathbf{x}^\star - \widehat{\mathbf{x}}_{t+1}\|_2^2 - \frac{1}{2}\|\widehat{\mathbf{x}}_{t+1} - \widehat{\mathbf{x}}_t\|_2^2.$$

(3) **[10pts]** Try to prove that, the algorithm using (2.3) and (2.4) satisfies:

$$\sum_{t=1}^{T}\big(F_t(\mathbf{x}_t) - F_t(\mathbf{x}^\star)\big) \leq \frac{\|\mathbf{x}^\star - \widehat{\mathbf{x}}_1\|_2^2}{2\eta} + \eta\sum_{t=1}^{T}\|\nabla f_t(\mathbf{x}_t) - M_t\|_2^2 - \frac{1}{4\eta}\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{x}_{t-1}\|_2^2.$$

(4) **[10pts]** Design the weights α_t, the step size η and the optimism M_t, prove that:

$$F(\overline{\mathbf{x}}_T) - F(\mathbf{x}^\star) \leq \mathcal{O}\left(L\cdot\frac{1}{T^2}\right).$$

Solution. Give your answers here. (中英文均可)

3 [50pts] Two-Point Bandit Convex Optimization

We consider Bandit Convex Optimization (BCO) with two-point feedback. At each round t, the online learner can query two points $\mathbf{x}_t^1, \mathbf{x}_t^2 \in \mathcal{X} \subseteq \mathbb{R}^d$, and observe the function values $f_t(\mathbf{x}_t^1)$ and $f_t(\mathbf{x}_t^2)$. The online functions $\{f_t\}_{t=1}^T$ are supposed to be G-Lipschitz. The objective is to minimize the following expected regret over T rounds:

$$\mathbb{E}\left[\mathrm{REG}_T\right] = \mathbb{E}\left[\sum_{t=1}^{T} \frac{1}{2}\Big(f_t\left(\mathbf{x}_t^1\right) + f_t\left(\mathbf{x}_t^2\right)\Big)\right] - \min_{\mathbf{x}\in\mathcal{X}} \sum_{t=1}^{T} f_t(\mathbf{x}) \tag{3.1}$$

Building on the two-point feedback, we aim to refine the Bandit Gradient Descent algorithm introduced in the course. At each round, we use the observed information to estimate a gradient $\widetilde{g}_t$ and then use it to perform gradient descent:

$$\mathbf{y}_{t+1} = \Pi_{(1-\alpha)\mathcal{X}}\left[\mathbf{y}_t - \eta\widetilde{\mathbf{g}}_t\right]$$

where $\Pi_{(1-\alpha)\mathcal{X}}$ denotes the projection onto the shrunk set $(1-\alpha)\mathcal{X}$.

(1) **[10pts]** A basic idea for gradient estimation is: first uniformly sample from a unit vector $\mathbf{s}_t \in \mathbb{S} \triangleq \left\{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\|_2 = 1\right\}$ at random and submit the following queries: $\mathbf{x}_t^1 = \mathbf{y}_t + \delta\mathbf{s}_t$ and $\mathbf{x}_t^2 = \mathbf{y}_t - \delta\mathbf{s}_t$; and then use the observed values $f_t(\mathbf{x}_t^1)$ and $f_t(\mathbf{x}_t^2)$ to construct the following gradient estimator

$$\widetilde{\mathbf{g}}_t = \frac{d}{2\delta}\left(f_t\left(\mathbf{y}_t + \delta\mathbf{s}_t\right) - f_t\left(\mathbf{y}_t - \delta\mathbf{s}_t\right)\right)\mathbf{s}_t,$$

(1.i) **[5pts]** Please prove that the gradient estimator still satisfies the unbiasedness condition:

$$\widehat{f}_t(\mathbf{y}_t) = \mathbb{E}_{\mathbf{v}\in\mathbb{B}}[f_t(\mathbf{y}_t + \delta\mathbf{v})], \quad \mathbb{E}_{\mathbf{s}\in\mathbb{S}}[\widetilde{\mathbf{g}}_t] = \nabla\widehat{f}(\mathbf{y}_t),$$

where $\mathbb{B} = \left\{\mathbf{x} \in \mathbb{R}^d \mid \|\mathbf{x}\|_2 \le 1\right\}$ is the unit ball and $\mathbb{S}$ is the unit sphere. Note that it is allowed to directly use Lemma 1 in Lecture 11 (unbiasedness of gradient estimator in one-point BCO).

(1.ii) **[5pts]**Please prove that, the gradient estimator has bounded norm: $\|\widetilde{\mathbf{g}}_t\|_2 \le Gd$.

(2) **[15pts]** Now we aim to analyze the regret of the refined BGD algorithm. Based on the analysis in question (1), we know that the gradient estimator satisfies $\mathbb{E}_{\mathbf{s}\in\mathbb{S}}[\widetilde{\mathbf{g}}_t] = \nabla\widehat{f}(\mathbf{y}_t)$. This implies that the refined BGD algorithm is performing online gradient descent (as if with full information) on the function $\widehat{f}_t$, restricted to the convex set $(1-\alpha)\mathcal{X}$. Thus, when analyzing the regret (3.1), we aim to relate it to the regret of OGD on $\widehat{f}_t$, namely,

$$\sum_{t=1}^{T} \widehat{f}_t(\mathbf{x}_t) - \sum_{t=1}^{T} \widehat{f}_t((1-\alpha)\mathbf{x}).$$

We first consider the single-round regret $\frac{1}{2}\left(f_t(\mathbf{x}_t^1)+f_t(\mathbf{x}_t^2)\right)-f_t(\mathbf{x})$, this regret can be decomposed into five components, each capturing a specific aspect of the algorithm:

$$\frac{1}{2}\left(f_t(\mathbf{x}_t^1)+f_t(\mathbf{x}_t^2)\right)-f_t(\mathbf{x})=\underbrace{\frac{1}{2}\left(f_t(\mathbf{x}_t^1)+f_t(\mathbf{x}_t^2)\right)-f_t(\mathbf{y}_t)}_{\text{TERM (A)}}+\underbrace{f_t(\mathbf{y}_t)-\widehat{f}_t(\mathbf{y}_t)}_{\text{TERM (B)}}$$
$$+\underbrace{\widehat{f}_t(\mathbf{y}_t)-\widehat{f}_t((1-\alpha)\mathbf{x})}_{\text{TERM (C)}}+\underbrace{\widehat{f}_t((1-\alpha)\mathbf{x})-f_t((1-\alpha)\mathbf{x})}_{\text{TERM (D)}}+\underbrace{f_t((1-\alpha)\mathbf{x})-f_t(\mathbf{x})}_{\text{TERM (E)}}.$$

(2.i) **[5pts]**Please explain the meaning of each of these 5 terms. What specific impact does each term represent?

(2.ii) **[10pts]** Given that $\|\mathbf{x}\|_2\le D$ and f_t is G-Lipschitz, use the above decomposition to prove that the following regret bound holds for all $\mathbf{x}\in\mathcal{X}$,

$$\sum_{t=1}^{T}\frac{1}{2}\left(f_t\left(\mathbf{x}_t^1\right)+f_t\left(\mathbf{x}_t^2\right)\right)-\sum_{t=1}^{T}f_t(\mathbf{x})\le\sum_{t=1}^{T}\widehat{f}_t\left(\mathbf{y}_t\right)-\sum_{t=1}^{T}\widehat{f}_t((1-\alpha)\mathbf{x})+3TG\delta+TGD\alpha.$$

(3) **[10pts]** We define $h_t(\mathbf{x})\triangleq\widehat{f}_t(\mathbf{x})+\left(\widetilde{\mathbf{g}}_t-\nabla\widehat{f}_t\left(\mathbf{y}_t\right)\right)^\top\mathbf{x}$, it is easily seen that $h_t(\mathbf{x})$ is also convex with $\nabla h_t(\mathbf{y}_t)=\widetilde{\mathbf{g}}_t$, which means the refined BGD algorithm is performing deterministic OGD on the function h_t restricted to the convex set $(1-\alpha)\mathcal{X}$.

(3.i) **[5pts]** Please prove that:

$$\sum_{t=1}^{T}h_t\left(\mathbf{y}_t\right)-\sum_{t=1}^{T}h_t((1-\alpha)\mathbf{x})\le\frac{D^2}{2}\frac{1}{\eta_T}+\frac{G^2d^2}{2}\sum_{t=1}^{T}\eta_t.$$

(3.ii) **[5pts]** Based on the results in (2.ii) and (3.i), please further prove that: (**Hint**: $\mathbb{E}[h_t(\mathbf{x})]=\widehat{f}_t(\mathbf{x})$)

$$\mathbb{E}\left[\sum_{t=1}^{T}\frac{1}{2}\left(f_t\left(\mathbf{x}_t^1\right)+f_t\left(\mathbf{x}_t^2\right)\right)-\sum_{t=1}^{T}f_t(\mathbf{x})\right]\le\frac{D^2}{2}\frac{1}{\eta_T}+\frac{G^2d^2}{2}\sum_{t=1}^{T}\eta_t+3TG\delta+TGD\alpha.$$

(4) **[10pts]** Assume $r\mathbb{B}\subset\mathcal{X}\subset D\mathbb{B}$, Please design the learning rate η_t, δ, α to make sure each step $\mathbf{x}_t^1,\mathbf{x}_t^2\in\mathcal{X}$ and prove the following regret bound.

$$\mathbb{E}\left[\sum_{t=1}^{T}\frac{1}{2}\left(f_t\left(\mathbf{x}_t^1\right)+f_t\left(\mathbf{x}_t^2\right)\right)-\sum_{t=1}^{T}f_t(\mathbf{x})\right]\le\frac{3DGd}{2}\sqrt{T}+3G+\frac{GD}{r}.$$

(5) **[5pts]** Notice that in the course, we introduced BCO with single-point feedback, which achieves a regret of $\mathcal{O}(T^{\frac{3}{4}})$. Please explain how two-point feedback provides advantages over single-point feedback, enabling us to achieve a regret of $\mathcal{O}(\sqrt{T})$.

Solution. Give your answers here. (中英文均可)

4 [50pts] Efficient Stochastic Logistic Bandits

We consider the Stochastic Logistic Bandits (LogB) problem. The reward satisfies $r_t = \mu(X_t^\top \theta_*) + \eta_t$, where $\mu(z) = (1 + \exp(-z))^{-1}$, and the noise η_t follows a Bernoulli distribution such that $\mathbb{P}(r_t = 1 \mid X_t) = \mu(X_t^\top \theta_*)$ and $\mathbb{P}(r_t = 0 \mid X_t) = 1 - \mu(X_t^\top \theta_*)$. The learner's goal is to minimize the regret:

$$\textsc{Reg}_T = \max_{\mathbf{x} \in \mathcal{X}} \sum_{t=1}^{T} \mu(\mathbf{x}^\top \theta_*) - \sum_{t=1}^{T} \mu(X_t^\top \theta_*).$$

To simplify the analysis, we assume that the feasible set and the unknown parameter are bounded: for all $X \in \mathcal{X} \subset \mathbb{R}^d$, $\|X\|_2 \leq 1$, and $\|\theta_*\|_2 \leq S$. Furthermore, $\mu(z)$ is L-Lipschitz on $z \in [-S, S]$, and its derivative satisfies $\inf_{z \in (-S,S)} \mu'(z) = \kappa$.

To estimate the unknown parameter θ_* in LogB, a common approach is to replace the least squares estimator used in LinUCB with the maximum likelihood estimator (MLE) or the following to minimize negative log-likelihood:

$$\widehat{\theta}_t = \arg\min_{\|\theta\| \leq S} \sum_{s=1}^{t} \ell_s(\theta),$$

where $-\ell_s(\theta) = r_s \log \mu\left(X_s^\top \theta\right) + (1 - r_s) \log\left(1 - \mu\left(X_s^\top \theta\right)\right)$.

However, MLE poses a significant computational challenge in this context, as it does not support online updates. As a result, each decision-making round requires costly re-computation using all past data, leading to scalability issues. To address this, recent advancements suggest modeling the problem as an OCO problem by incrementally feeding the loss functions $\{\ell_s(\theta)\}_{s=1}^t$ into an online learner $\mathcal{B}$. Based on the output θ_s from $\mathcal{B}$ at each round, we construct a virtual linear reward $z_s = X_s^\top \theta_s$. Using these, we compute the least squares estimator $\widehat{\theta}_t$ over the historical data pairs $\{X_s, z_s\}_{s=1}^t$. This approach allows for efficient online updates at every round. Next, we will prove that this method achieves reliable estimation error guarantees and a favorable regret bound.

(1) **[10pts]** Assume that the online learner $\mathcal{B}$ satisfies the following regret bound: $\forall \theta, \|\theta\|_2 \leq S, \forall t \geq 1, \sum_{s=1}^t \ell_s(\theta_s) - \ell_s(\theta) \leq B_t$. To construct the UCB for the online estimator, we need to relate the parameter estimation error with the regret B_t. Please prove the following result: (**Hint**: Taylor's theorem over $\ell_s(\theta)$.)

$$\sum_{s=1}^{t} \left(X_s^\top (\theta_s - \theta_*)\right)^2 \leq \frac{2}{\kappa} B_t + \frac{2}{\kappa} \sum_{s=1}^{t} \eta_s \left(X_s^\top (\theta_s - \theta_*)\right). \tag{4.1}$$

(2) **[15pts]** For Eq (4.1), the term $\sum_{s=1}^t \eta_s \left(X_s^\top (\theta_s - \theta^*)\right)$ exhibits a structure similar to the self-normalized concentration inequality introduced in the lecture. Hence, we aim to apply the self-normalized concentration inequality to handle this term.

(2.i) **[5pts]** Note that the self-normalized concentration inequality requires the noise to be sub-Gaussian. Please prove that the noise $\eta_t = r_t - \mu(X_t^\top \theta^*)$ is R-sub-Gaussian where $R \leq \frac{1}{2}$. (**Hint**: Use Hoeffding's lemma.)

(2.ii) **[5pts]** Now we know that the noise η_t is R-sub-Gaussian, try to prove that, with probability at least $1-\delta$, the following holds for any $t \in [T]$: (**Hint**: convert self-normalized concentration into 1-dimensional version.)

$$\begin{aligned}&\sum_{s=1}^{t} \eta_s \left(X_s^\top (\theta_s - \theta_*)\right) \\ &\leq R\sqrt{\left(2 + 2\sum_{s=1}^{t} \left(X_s^\top (\theta_s - \theta_*)\right)^2\right) \cdot \log\left(\frac{1}{\delta}\sqrt{1 + \sum_{s=1}^{t} \left(X_s^\top (\theta_s - \theta_*)\right)^2}\right)}.\end{aligned}$$

(2.iii) **[5pts]** Substitute the above inequality into (4.1), and further prove that, with probability at least $1-\delta$, the following holds for any $t \in [T]$ (**Hint**: define $q \triangleq \sqrt{1 + \sum_{s=1}^{t} \left(X_s^\top (\theta_s - \theta_*)\right)^2}$)

$$\sum_{s=1}^{t} \left(X_s^\top (\theta_s - \theta_*)\right)^2 \leq \beta_t' \triangleq 1 + \frac{4}{\kappa} B_t + \frac{8R^2}{\kappa^2} \log\left(\frac{1}{\delta}\sqrt{4 + \frac{8}{\kappa} B_t + \frac{64R^4}{\kappa^4 \cdot 4\delta^2}}\right). \quad (4.2)$$

(3) **[15pts]** Denote $z_s = X_s^\top \theta_s$ as the virtual reward at step s. Then we can compute the parameter estimator $\widehat{\theta}_t$ using least squares over the history data pairs $\{X_s, z_s\}_{s=1}^t$ as

$$\widehat{\theta}_t = \arg\min_{\theta} \lambda \|\theta\|_2^2 + \sum_{s=1}^{t} \left(z_s - X_s^\top \theta\right)^2. \quad (4.3)$$

Based on Eq (4.2) and estimator (4.3), let $V_t = \lambda I_d + \sum_{s=1}^{t} X_s X_s^\top$. Please prove that, with probability at least $1-\delta$, the following holds for any $t \in [T]$: (**Hint**: closed form of least square (4.3))

$$\|\theta_* - \widehat{\theta}_t\|_{V_t}^2 \leq \beta_t \triangleq \lambda S^2 + \beta_t' - \left(\lambda \|\widehat{\theta}_t\|_2^2 + \sum_{s=1}^{t} \left(z_s - X_s^\top \widehat{\theta}_t\right)^2\right). \quad (4.4)$$

(4) **[10pts]** Based on the UCB β_t in Eq (4.4),we design the UCB select criteria as follows,

$$X_t = \arg\max_{\mathbf{x} \in \mathcal{X}} \left\{\left\langle \mathbf{x}, \widehat{\theta}_{t-1} \right\rangle + \sqrt{\beta_{t-1}} \|\mathbf{x}\|_{V_{t-1}^{-1}}\right\}. \quad (4.5)$$

For the parameter estimator (4.3) and arm selection criteria (4.5), please prove that, with probability at least $1-2\delta$, the following regret bound holds: (**Hint**: $\mu(\mathbf{x}^\top \theta_*) - \mu(X_t^\top \theta_*) \leq L(\mathbf{x} - X_t)^\top \theta_*$)

$$\textsc{Reg}_T = \max_{\mathbf{x} \in \mathcal{X}} \sum_{t=1}^{T} \mu(\mathbf{x}^\top \theta_*) - \sum_{t=1}^{T} \mu(X_t^\top \theta_*) = \mathcal{O}\left(L\sqrt{\beta_T d T \log T}\right).$$

Solution. Give your answers here. (中英文均可)

5 [50pts] Online Regression with Available Information

This problem investigates how to incorporate the available side information to obtain an improved regret bound for online regression. At each round $t \in [T]$, the online learner submits a decision $\mathbf{x}_t \in \mathcal{X} \subseteq \mathbb{R}^d$, and the online function is defined as $f_t(\mathbf{x}_t) = \frac{1}{2}(\mathbf{x}_t^\top \boldsymbol{\psi}_t - y_t)^2$, where $\boldsymbol{\psi}_t \in \Psi \subseteq \mathbb{R}^d$ denotes the feature and $y_t \in \mathcal{Y} \subseteq \mathbb{R}$ denotes the corresponding label. Our goal is to minimize the regret for any $\mathbf{u} \in \mathcal{X}$:

$$\textsc{Reg}_T(\mathbf{u}) \triangleq \sum_{t=1}^{T} f_t(\mathbf{x}_t) - \sum_{t=1}^{T} f_t(\mathbf{u}).$$

We assume that $y_t \in [-Y, Y]$ holds for all $t \in [T]$.

(1) **[5pts]** Prove that online function $f_t(\mathbf{x})$ is α-exp-concave with $\alpha = \min_{\mathbf{x}\in\mathcal{X}} \left\{ \frac{1}{(\mathbf{x}^\top \boldsymbol{\psi}_t - y_t)^2} \right\}$.

Based on the above result, we know that an $\mathcal{O}\left(\max_{\mathbf{x}\in\mathcal{X}, t\in[T]}\{(\mathbf{x}^\top \boldsymbol{\psi}_t - y_t)^2\} \cdot d \log T\right)$ regret is attainable by employing Online Newton Step when assuming the boundedness of domain diameter and gradient norm. However, this regret may not be favorable when the domain $\mathcal{X}$ or the feature space Ψ is large (the exp-concave parameter α is very small).

Below, we will resolve the issue using the available information on this problem. Actually, in online regression, the feature $\boldsymbol{\psi}_t$ is available to the online learner *before* submitting the decision $\mathbf{x}_t$ (while the label y_t is definitely unknown now), which means the learner knows part of $f_t(\cdot)$'s information before updating. Hence, we use Optimistic Online Mirror Descent to leverage this available information by treating it as the "hint":

$$\begin{aligned} \mathbf{x}_t &= \arg\min_{\mathbf{x}\in\mathbb{R}^d} \left\{ \frac{1}{2}\left(\mathbf{x}^\top \boldsymbol{\psi}_t\right)^2 + \frac{1}{2}\|\mathbf{x} - \widehat{\mathbf{x}}_t\|^2_{\mathbf{A}_{t-1}} \right\}, \\ \widehat{\mathbf{x}}_{t+1} &= \arg\min_{\mathbf{x}\in\mathbb{R}^d} \left\{ \frac{1}{2}\left(\mathbf{x}^\top \boldsymbol{\psi}_t - y_t\right)^2 + \frac{1}{2}\|\mathbf{x} - \widehat{\mathbf{x}}_t\|^2_{\mathbf{A}_{t-1}} \right\}, \end{aligned} \tag{5.1}$$

where $\mathbf{A}_{t-1} = \lambda \mathbf{I} + \sum_{s=1}^{t-1} \boldsymbol{\psi}_s \boldsymbol{\psi}_s^\top$ is the regularized covariance matrix.

In (5.1), we have considered the difficult scenario that $\mathcal{X} = \mathbb{R}^d$ (recall that now α can approach 0). To simplify subsequent presentations, we denote by $h_t(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}^\top \boldsymbol{\psi}_t\right)^2$, serving as a "guess" of $f_t(\mathbf{x})$ by treating $y_t = 0$.

(2) **[20pts]** Prove a property of the online function $f_t(\mathbf{x})$:

$$f_t(\mathbf{x}) - f_t(\mathbf{y}) = \langle \nabla f_t(\mathbf{x}), \mathbf{x} - \mathbf{y} \rangle - \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2_{\boldsymbol{\psi}_t \boldsymbol{\psi}_t^\top}.$$

With this property, try to prove an intermediate result for the algorithm in (5.1):

$$\begin{aligned} \textsc{Reg}_T(\mathbf{u}) \leq & \sum_{t=1}^{T} \frac{1}{2}\Big(\|\mathbf{u} - \widehat{\mathbf{x}}_t\|^2_{\mathbf{A}_{t-1}} - \|\mathbf{u} - \widehat{\mathbf{x}}_{t+1}\|^2_{\mathbf{A}_{t-1}} - \|\mathbf{u} - \widehat{\mathbf{x}}_{t+1}\|^2_{\boldsymbol{\psi}_t \boldsymbol{\psi}_t^\top} \Big) \\ & + \sum_{t=1}^{T} \Big(f_t(\mathbf{x}_t) - f_t(\widehat{\mathbf{x}}_{t+1}) + h_t(\widehat{\mathbf{x}}_{t+1}) - h_t(\mathbf{x}_t) \Big). \end{aligned}$$

(3) **[10pts]** Prove the following technical result:

$$f_t(\mathbf{x}_t) - f_t(\widehat{\mathbf{x}}_{t+1}) + h_t(\widehat{\mathbf{x}}_{t+1}) - h_t(\mathbf{x}_t) = y_t^2 \boldsymbol{\psi}_t^\top \mathbf{A}_t^{-1} \boldsymbol{\psi}_t.$$

(4) **[15pts]** Prove the final regret bound of the algorithm in (5.1):

$$\textsc{Reg}_T(\mathbf{u}) \leq \mathcal{O}\left(\lambda \|\mathbf{u}\|_2^2 + dY^2 \log(T)\right).$$

Compared to $\mathcal{O}\left(\max_{\mathbf{x}\in\mathcal{X}, t\in[T]}\{(\mathbf{x}^\top \boldsymbol{\psi}_t - y_t)^2\} \cdot d \log T\right)$, the refined regret bound depends on Y^2 rather than a potentially large quantity $\max_{\mathbf{x}\in\mathcal{X}, t\in[T]}\{(\mathbf{x}^\top \boldsymbol{\psi}_t - y_t)^2\}$, hence demonstrating the value of employing this side information in online regression.

Solution. Give your answers here. (中英文均可)

6 [5pts] Bonus (Lecture Slides 8-12)

You can earn bonus points by pointing out errors in the lecture slides 8-12 on the course website. Specifically, consider the following three types of errors:

(A) Technical errors (e.g., incorrect coefficients in formulas), 1pts each.

(B) Serious typo in presentation (e.g., AB but actually $A^\top B$, $\mathbf{x}A$ but actually $\mathbf{x}^\top A$), 0.5pts each.

(C) Typos in formula/statement (e.g., writing vector $\mathbf{x}_t$ as x_t; grammar/spelling errors), 0.25pts each.

(D) Other suggestions: like how to better organize the proofs or alternative simplified proofs..., up to 1.5pts each.

List the errors in lecture slides 8-12 and state the way to correct. Please clearly indicate which type each error belongs to, with a total score not exceeding 5pts.

For example,

(1) [**(A) Technical errors**] Lecture X. Page2. xxx

(2) [**(B) Serious typo in presentation**] Lecture Y. Page4. yyy → zzz

(3) [**(C) Typos in formula/statement**] Lecture W. Page6. www → vvv

(4) [**(D) Other suggestions**] Lecture V. Page8. It would be better...

Solution. Give your answers here. (中英文均可)

Acknowledgements

The homework bearing your name must represent your individual contribution. While discussions during the completion of the assignment are permissible, they are conditioned upon the fact that none of the participating individuals have completed the discussed topics. We emphasize that the implementation of key ideas within the assignment must be done independently. **You should extend your acknowledgments to those individuals who have participated in the discussions here**.

This course adopts a zero-tolerance policy toward plagiarism. The grades of students found to have engaged in plagiarism without providing proper citations or acknowledgments will be **annulled**. In cases of mutual plagiarism, the grades of **both** the plagiarizer and the plagiarized will be **annulled**.