Sequential Learning

Final Examination

The duration of the exam is 2 hours. A single two-sided sheet of handwritten notes (with any content) is allowed. Answers can be written in French or English.

The exam is divided into **two separate parts**, each of which should be submitted on a **separate sheet**.

Part 1. Online Convex Optimization

1. Let $\Theta \subset \mathbb{R}^d$ be a convex compact set and $(\ell_t : \Theta \to \mathbb{R})_{1 \le t \le T}$ be a sequence of convex functions.
   (a) Provide the pseudocode of Online Gradient Descent (OGD).
   (b) Prove that OGD with step size η_t is equivalent to the update:
   $$\theta_{t+1} = \arg\min_{\theta \in \Theta} \left\{ \langle \nabla \ell_t(\theta_t), \theta \rangle + \lambda_t \|\theta - \theta_t\|^2 \right\}$$
   for some regularization parameter λ_t. Give the expression of λ_t as a function of η_t.
   (c) What is the connexion between OGD and Online Mirror Descent (OMD)? (Justify briefly)
2. Assume that ℓ_t are i.i.d.. Let $L = \mathbb{E}[\ell_t(\,\cdot\,)]$, which we assume μ-strongly convex, and $\theta_* \in \arg\min_{\theta \in \Theta} L(\theta)$. Let $R_T(\theta_*) := \sum_{t=1}^T \ell_t(\theta_t) - \ell_t(\theta_*)$ be the regret of some online algorithm.
   (a) Design a point $\bar\theta_T$ that controls its excess risk $\mathbb{E}\big[L(\bar\theta_T) - L(\theta_*)\big]$ as a function of $\mathbb{E}\big[R_T(\theta_*)\big]$.
   (b) Prove an upper bound on $\mathbb{E}\big[\|\bar\theta_T - \theta_*\|^2\big]$.

Problem: Optimistic Follow The Regularized Leader

Let $\Theta \subseteq \mathbb{R}^d$ such that $\theta_1 := 0 \in \Theta$. We assume that at each $t \ge 1$, the learner tries to guess the next gradient with some $\widehat{g}_{t+1} \in \mathbb{R}^d$ and updates

$$\theta_{t+1} \in \arg\min_{\theta \in \Theta} \Phi_t(\theta), \qquad \text{with} \quad \Phi_t(\theta) := \left\langle \theta, \textstyle\sum_{s=1}^t g_s + \widehat{g}_{t+1} \right\rangle + \frac{\lambda}{2}\|\theta\|^2$$

where $g_s = \nabla \ell_s(\theta_s)$ and $\lambda > 0$ is a regularization parameter. We assume $\widehat{g}_1 = 0$ and $\Phi_0(\theta) = \frac{\lambda}{2}\|\theta\|^2$.

3. Show that for all $t \ge 1$ and $\theta_* \in \Theta$, $\Phi_{t-1}(\theta_t) - \Phi_{t-1}(\theta_*) \le -\frac{\lambda}{2}\|\theta_t - \theta_*\|^2$.
4. Define $\tilde\theta_{t+1} \in \arg\min_{\theta \in \Theta} \left\{ \langle \theta, \sum_{s=1}^t g_s \rangle + \frac{\lambda}{2}\|\theta\|^2 \right\}$. Show that for all $t \ge 1$ and $\theta_* \in \Theta$
   $$\Phi_{t-1}(\theta_t) \le \langle \theta_*, \textstyle\sum_{s=1}^t g_s \rangle + \langle \tilde\theta_{t+1}, \widehat{g}_t - g_t \rangle - \frac{\lambda}{2}\|\tilde\theta_{t+1} - \theta_t\|^2 + \frac{\lambda}{2}\|\theta_*\|^2$$

5. Deduce by induction on t that for all $t \geq 1$ and $\theta_* \in \Theta$:
$$\sum_{s=1}^{t} \langle \theta_s - \tilde{\theta}_{s+1}, \widehat{g}_s \rangle + \langle \tilde{\theta}_{s+1}, g_s \rangle \leq \langle \theta_*, \textstyle\sum_{s=1}^t g_t \rangle - \frac{\lambda}{2} \sum_{s=1}^t \|\theta_t - \tilde{\theta}_{t+1}\|^2 + \frac{\lambda}{2}\|\theta_*\|^2 \,.$$

6. Deduce that for any $\theta_* \in \Theta$
$$\sum_{t=1}^{T} \langle \theta_t - \theta_*, g_t \rangle \leq \sum_{t=1}^{T} \langle \theta_t - \tilde{\theta}_{t+1}, g_t - \widehat{g}_t \rangle - \frac{\lambda}{2}\|\tilde{\theta}_{t+1} - \theta_t\|^2 + \frac{\lambda}{2}\|\theta_*\|^2 \,.$$

7. Conclude by proving a regret upper bound that depends on $V_T := \sum_{t=1}^T \|g_t - \widehat{g}_t\|^2$ for a well-optimized λ to be explicitly indicated.

Part 2. Stochastic bandits

8. Give an example of a stochastic bandit problem on which the Follow The Leader algorithm has linear expected regret. Prove that linear lower bound on the expected regret.

9. In fixed confidence best arm identification (BAI), an algorithm is δ-correct if it returns the best arm with probability at least $1 - \delta$.
   (a) If an algorithm is δ-correct on all bandits with Gaussian rewards, it satisfies a lower bound on its expected stopping time $\mathbb{E}[\tau_\delta]$. How does that lower bound depend on δ, for small δ?
   (b) Give an example of a δ-correct algorithm for BAI with Gaussian rewards with variance 1. Note that we are not asking for an algorithm with small sample complexity: any δ-correct algorithm suffices.

Problem: ε-greedy algorithm for stochastic bandits

We consider the stochastic bandit setting: an algorithm sequentially interacts with $K \in \mathbb{N}$ arms with $K > 1$, where each arm $k \in \{1, \dots, K\}$ is described by a distribution ν_k supported on $[0, 1]$ with mean μ_k. We suppose that $\mu_1 > \mu_k$ for all $k \in \{2, \dots, K\}$. We call $\Delta_k = \mu_1 - \mu_k$ the gap of arm k. When the algorithm pulls arm k_t at time t, it observes a reward X_{t,k_t} sampled from ν_{k_t}.

For $i \in \mathbb{N}$ with $i \geq 1$, we write $[i] = \{1, \dots, i\}$. For two propositions p_1 and p_2, the expression $p_1 \wedge p_2$ means p_1 and p_2.

The ε-greedy algorithm depends on a sequence of parameters $\varepsilon_1, \varepsilon_2, \dots$ in $[0, 1]$. First, the algorithm pulls each arm once. Then at time $t > K$, let $N_{t-1}^k = \sum_{s=1}^{t-1} \mathbb{1}_{\{k_s = k\}}$ be the number of times arm k was chosen up to time $t-1$ and let $\widehat{\mu}_{t-1}^k = \frac{1}{N_{t-1}^k} \sum_{s=1}^{t-1} X_{s,k_s} \mathbb{1}_{\{k_s = k\}}$. With probability $1 - \varepsilon_t$, the ε-greedy algorithm pulls the arm $k_t = \arg\max_k \widehat{\mu}_{t-1}^k$; with probability ε_t, it pulls an arm uniformly at random. Let Z_t be the Bernoulli random variable with expectation ε_t with value 1 if the arm is chosen uniformly and 0 otherwise.

Let the regret of the algorithm at time T be $R_T = T\mu_1 - \sum_{t=1}^T \mu_{k_t}$.

10. Prove that $\mathbb{E}[R_T] = \sum_{k=2}^K \Delta_k \mathbb{E}[N_T^k]$.

11. What is the expectation m_T of $\sum_{t=1}^T Z_t$? Give an upper bound on $\mathbb{P}\left\{\sum_{t=1}^T Z_t - m_T \geq Tx\right\}$ that is exponentially decreasing in T.

In the next questions, we suppose that $\varepsilon_t = \varepsilon \in [0,1]$ for all $t \in \mathbb{N}$.

We introduce the notation $N_{Z,t}^k = \sum_{s=1}^t \mathbb{1}_{\{Z_s=1 \,\wedge\, k_s=k\}}$, which corresponds to the number of pulls of arm k due to the uniform exploration. Let $\mathcal{E}_T$ be the event that for all $k \in [K]$ and $t \in [T]$, $\left|N_{Z,t}^k - t\frac{\varepsilon}{K}\right| \leq \sqrt{\frac{t}{2}\log(2KT^2)}$.

12. Prove that

$$\mathbb{E}[R_T] \leq T\mathbb{P}(\mathcal{E}_T^c)\max_k \Delta_k + \sum_{k=1}^K \Delta_k \sum_{t=1}^T \mathbb{P}\{\mathcal{E}_T \,\wedge\, Z_t = 0 \,\wedge\, k_t = k\} + \sum_{k=1}^K \Delta_k \mathbb{E}[N_{Z,T}^k \mathbb{1}_{\{\mathcal{E}_T\}}] \,. \quad (1)$$

13. (a) For $t \geq 1$, $k \in [K]$, what is the law of the random variable with value 1 if both $Z_t = 1$ and $k_t = k$, and value 0 otherwise?

    (b) Let $\delta \in (0,1)$, $t \in [T]$ and $k \in [K]$. By showing two concentration inequalities and doing an union bound, prove that with probability $1-\delta$,

$$\left|N_{Z,t}^k - t\frac{\varepsilon}{K}\right| \leq \sqrt{\frac{t}{2}\log\frac{2}{\delta}}\,. \quad (2)$$

    Deduce that with probability $1-\frac{1}{T}$, for all $k \in [K]$ and $t \in [T]$, $\left|N_{Z,t}^k - t\frac{\varepsilon}{K}\right| \leq \sqrt{\frac{t}{2}\log(2KT^2)}$. That is, $\mathbb{P}(\mathcal{E}_T) \geq 1 - \frac{1}{T}$.

14. (a) Suppose that $T \geq \frac{2K^2}{\varepsilon^2}\log(2KT^2)$. For $t \in [T]$ such that $t \geq \frac{2K^2}{\varepsilon^2}\log(2KT^2)$ and $k \neq 1$, show an upper bound on $\mathbb{P}\{\mathcal{E}_T \,\wedge\, \widehat{\mu}_t^k > \widehat{\mu}_t^1\}$ of the form $C_1 t\exp(-tC_2)$ where C_1 and C_2 may depend on the parameters of the problem but not on t.

    (b) Deduce an upper bound on $\sum_{t=1}^T \mathbb{P}\{\mathcal{E}_T \,\wedge\, Z_t = 0 \,\wedge\, k_t = k\}$ for $k \neq 1$. Your bound can be expressed as a function of the quantity $C_{\exp}(a) := \sum_{t=1}^\infty te^{-ta}$.

    (c) Prove that $\limsup_{T\to\infty} \frac{\mathbb{E}[R_T]}{T} \leq \frac{\varepsilon}{K}\sum_{k=2}^K \Delta_k$.

15. Prove that $\lim_{T\to\infty} \frac{\mathbb{E}[R_T]}{T} = \frac{\varepsilon}{K}\sum_{k=2}^K \Delta_k$.