Instructions.

The only reference you can use during the exam is a hand-written double-sided A4-sized cheat sheet . If you use any other reference during the exam, then you will get zero point. You can use any result in the problem statements even when you do not know how to prove it. There are in total 105 points.

Problem 1

In this problem, we will develop a learning-based algorithm for compressive magnetic resonance imaging (MRI).

Let $A \in \mathbb{R}^{p \times p}$ be a unitary matrix, and $A_\Omega \in \mathbb{R}^{n \times p}$ be the sub-matrix of A consisting of rows indexed by $\Omega \subseteq \{1, \dots, p\}$. For example, $A_{\{1,3,5\}}$ is the $3 \times p$ matrix consisting of the first, third, and fifth rows of A. The standard theory of compressive MRI considers the linear model:

$$y = A_\Omega x,$$

where $y \in \mathbb{R}^n$ denotes the measurement outcome, $A_\Omega \in \mathbb{R}^{n \times p}$ encodes the measurement set-up, and $x \in \mathbb{R}^p$ represents an image of some part of a human body. The unitary matrix A is fixed in MRI machines. Therefore, for any $n \in \mathbb{N}$, our goal is to find some Ω of n elements, such that we can accurately estimate x given y and A_Ω.

We consider the *linear decoder*, which outputs

$$\hat{x}(y, \Omega) := A_\Omega^{\mathrm{T}} \left(A_\Omega A_\Omega^{\mathrm{T}} \right)^{-1} y$$

as an estimate of x.

1. (5 points) Consider the loss function

$$\lambda(x, \Omega) := \frac{\| x - \hat{x}(A_\Omega x, \Omega) \|_2^2}{\| x \|_2^2}, \quad \forall x \in \mathbb{R}^p, \Omega \subseteq \{1, \dots, p\}.$$

   Show that $\lambda(x, \Omega) \in [0, 1]$ **for every** $x \in \mathbb{R}^p$ **and** $\Omega \subseteq \{1, \dots, p\}$**.**

Problem 1 continued on next page...

2. (20 points) Suppose we want to find an Ω suitable for imaging human brains, and we have data—a collection of typical human brain images $x_1, \ldots, x_m \in \mathbb{R}^p$, modeled as independent and identically distributed random vectors. Consider the empirical risk minimization approach to learning a good Ω:

$$\hat{\Omega}_m \in \operatorname*{argmin}_{\Omega \subseteq \{1,\ldots,p\}, |\Omega| = n} \frac{1}{m} \sum_{1 \le i \le m} \lambda(x_i, \Omega),$$

where $|\Omega|$ denotes the number of elements in Ω. Define the risk minimizer as

$$\Omega^\star \in \operatorname*{argmin}_{\Omega \subseteq \{1,\ldots,p\}, |\Omega| = n} \mathsf{E}\, \lambda(x_1, \Omega).$$

Show that for any $\delta \in]0,1[$**,**

$$\mathsf{P}\left(\mathsf{E}\, \lambda(x, \hat{\Omega}_m) \le \mathsf{E}\, \lambda(x, \Omega^\star) + O\left(\sqrt{\frac{\log \binom{p}{n} + \log(1/\delta)}{m}} \right) \right) \ge 1 - \delta.$$

Problem 2

In this problem, we will study the regret of a variant of the aggregating algorithm that addresses possibly non-mixable losses.

Consider the following standard learning with expert advice protocol. Let T and n be integers larger than 2. Let Ω and Γ be two sets. For $t = 1, 2, \dots,$

1. EXPERT-i announces $\gamma_t(i) \in \Gamma$, $1 \leq i \leq n$.
2. LEARNER announces $\gamma_t \in \Gamma$.
3. REALITY announces $\omega_t \in \Omega$.

Denote the loss function as $\lambda : \Omega \times \Gamma \to \mathbb{R}$. We define the regret as

$$R_T(i) := \sum_{t=1}^{T} \lambda(\omega_t, \gamma_t) - \sum_{t=1}^{T} \lambda(\omega_t, \gamma_t(i)), \quad \forall 1 \leq i \leq n.$$

Consider the following algorithm. Let η be a positive number. Let $(\pi_1(i))_{1 \leq i \leq n}$ be a probability vector. For every $1 \leq t \leq T$, LEARNER announces some $\gamma_t \in \Gamma$ such that

$$\lambda(\omega, \gamma_t) \leq \sum_{1 \leq i \leq n} \left[\pi_t(i) \lambda(\omega, \gamma_t(i)) \right], \quad \forall \omega \in \Omega.$$

and then compute

$$\pi_{t+1}(i) := \frac{\pi_t(i) \mathrm{e}^{-\eta \lambda(\omega_t, \gamma_t(i))}}{\sum_{1 \leq i \leq n} \pi_t(i) \mathrm{e}^{-\eta \lambda(\omega_t, \gamma_t(i))}}, \quad \forall 1 \leq i \leq n.$$

Assume that γ_t are well-defined.

1. (5 points) **Show that γ_t are well-defined, if Γ is a convex set and $\lambda(\omega, \cdot)$ is a convex function for all $\omega \in \Omega$.**

Problem 2 continued on next page...

2. (20 points) **Show that**

$$R_T(i) \le \frac{1}{\eta} \log \frac{1}{\pi_1(i)} + \sum_{1 \le t \le T} \delta_t, \quad \forall 1 \le i \le n,$$

where

$$\delta_t := \sum_{1 \le i \le n} \pi_t(i) \lambda(\omega_t, \gamma_t(i)) + \left(\frac{1}{\eta}\right) \log \left[\sum_{1 \le i \le n} \pi_t(i) \mathrm{e}^{-\eta \lambda(\omega_t, \gamma_t(i))} \right], \quad \forall 1 \le t \le T.$$

Problem 2 continued on next page…

3. (10 points) Suppose $\lambda(\omega, \gamma) \in [0, L]$ for all $(\omega, \gamma) \in \Omega \times \Gamma$, for some $L > 0$. Set

$$\pi_1(i) = \frac{1}{n}, \quad \forall 1 \leq i \leq n.$$

Show that

$$R_T(i) = O\left(L\sqrt{T \log n}\right), \quad \forall 1 \leq i \leq n.$$

Problem 3

In this problem, we will reduce a convex optimization problem to individual sequence prediction on a binary sequence.

Consider the following online convex optimization problem. Let $T \in \mathbb{N}$. For every $1 \leq t \leq T$, the following happen sequentially.

1. LEARNER announces $\gamma_t \in \mathbb{R}$.
2. REALITY announces $\omega_t \in \mathbb{R}$.

We define the regret as

$$R_T(\gamma) := \sum_{t=1}^{T} \lambda(\omega_t, \gamma_t) - \sum_{t=1}^{T} \lambda(\omega_t, \gamma), \quad \forall \gamma \in \mathbb{R},$$

where

$$\lambda(\omega, \gamma) := |\omega - \gamma|, \quad \forall (\omega, \gamma) \in \mathbb{R}^2.$$

1. (5 points) Define

$$g(\omega, \gamma) := \begin{cases} 1 & , \text{if } \omega \leq \gamma, \\ -1 & , \text{otherwise.} \end{cases}$$

   Show that for every $\omega \in \mathbb{R}$,

$$\lambda(\omega, \gamma_2) \geq \lambda(\omega, \gamma_1) + g(\omega, \gamma_1)(\gamma_2 - \gamma_1), \quad \forall (\gamma_1, \gamma_2) \in \mathbb{R}^2,$$

   and

$$R_T(\gamma) \leq \sum_{t=1}^{T} g_t \gamma_t - \sum_{t=1}^{T} g_t \gamma, \quad \forall \gamma \in \mathbb{R},$$

   where $g_t := g(\omega_t, \gamma_t)$.

Problem 3 continued on next page. . .

2. (5 points) Suppose that

$$\sum_{1\le t\le T} g_t\gamma_t \le -f\left(-\sum_{1\le t\le T} g_t\right),$$

for some function $f:\mathbb{R}\to\mathbb{R}$. **Show that**

$$R_T(\gamma)\le f^*(\gamma),\quad \forall\gamma\in\mathbb{R},$$

where

$$f^*(y) := \max_{x\in\mathbb{R}}\left[yx - f(x)\right].$$

That is, an upper bound on the cumulative linear loss (summation of $g_t\gamma_t$) yields an upper bound on the regret

3. (5 points) Define $W_0 := 1$, and

$$W_t := 1-\sum_{1\le\tau\le t} g_\tau\gamma_\tau,\quad \beta_t := \frac{\gamma_t}{W_{t-1}},\quad \forall 1\le t\le T.$$

Show that

$$W_T = \prod_{t=1}^{T}(1-g_t\beta_t).$$

Problem 3 continued on next page. . .

4. (20 points) Consider the following individual sequence prediction game. For every $1 \le t \le T$, the following happen sequentially.

   (a) LEARNER announces $p_t := \frac{(1+\beta_t)}{2} \in [0,1]$.

   (b) REALITY announces $a_t := \frac{1-g_t}{2} \in \{0,1\}$.

   The symbols β_t and g_t are exactly those appearing in the online convex optimization protocol. Consider the logarithmic loss given by

$$\lambda(a,p) := -(1-a)\log(1-p) - a\log p, \quad \forall (a,p) \in \{0,1\} \times [0,1].$$

   Denote the corresponding regret by $\tilde{R}_T$, i.e.,

$$\tilde{R}_T := \sum_{1 \le t \le T} \lambda(a_t, p_t) - \min_{p \in [0,1]} \sum_{1 \le t \le T} \lambda(a_t, p).$$

   Show that

$$W_T \ge \exp\left\{ T \left[\log 2 - H_b\left(\frac{1}{2} - \frac{\sum_{1 \le t \le T} g_t}{2T} \right) \right] - \tilde{R}_T \right\},$$

   where H_b denotes the binary entropy function, i.e.,

$$H_b(p) := -(1-p)\log(1-p) - p\log p, \quad \forall p \in [0,1].$$

   We define $0 \log 0 := 0$. Notice then we have an upper bound on $\sum_{1 \le t \le T} g_t \gamma_t$.

Problem 3 continued on next page. . .

5. (10 points) Suppose we use the Krichevsky-Trofimov mixture forecaster for the binary sequence prediction protocol. **Write down the resulting algorithm for the original online convex optimization problem.**

   Remark. Following the approach in this problem, Orabona and Pál proved the resulting algorithm satisfies

$$R_T(\gamma) \leq |\gamma| \sqrt{T \log\left(1 + 24 T^2 |\gamma|^2\right)} + o(1), \quad \forall \gamma \in \mathbb{R}.$$