Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Prof. Dr. Helmut Bölcskei
Chair for Mathematical Information Science

Examination on Mathematics of Information

August 12, 2024

- Do not turn this page before the official start of the exam.
- The problem statements consist of 7 pages including this page.
  Please verify that you have received all 7 pages.
- Throughout the problem statements there are references to definitions and theorems in the Handout, indicated by e.g. Definition H1 and Theorem H2.

Problem 1 (25 points)

We denote the standard orthonormal basis of the space $\ell^2(\mathbb{N})$ by $\{\delta_k\}_{k\in\mathbb{N}}$, i.e., for $k \in \mathbb{N}$ the sequence $\delta_k[\cdot] \in \ell^2(\mathbb{N})$ is given by

$$\delta_k[j] = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{otherwise} \end{cases}, \qquad j \in \mathbb{N}.$$

Throughout this problem we fix a Hilbert space $\mathcal{H}$. We will need the following definition.

Definition 1. *A set of elements $\{g_k\}_{k\in\mathbb{N}}$, $g_k \in \mathcal{H}$, is called a Riesz basis for $\mathcal{H}$ if there exists a bijective bounded linear operator $U : \ell^2(\mathbb{N}) \to \mathcal{H}$ such that $g_k = U\delta_k$, for $k \in \mathbb{N}$.*

(a) (13 points) Let $\{f_k\}_{k\in\mathbb{N}}$ be an exact frame for $\mathcal{H}$ and consider the linear operator T

$$T : \ell^2(\mathbb{N}) \to \mathcal{H}$$
$$T : c[\cdot] \to \sum_{k=1}^{\infty} c[k] f_k.$$

i. (5 points) Show that T is bijective.

ii. (5 points) Show that T is bounded, i.e., there is a constant M such that for all $c[\cdot] \in \ell^2(\mathbb{N})$ we have $\|Tc\|_{\mathcal{H}} \leq M \|c\|_{\ell^2(\mathbb{N})}$.

iii. (3 points) Show that $\{f_k\}_{k\in\mathbb{N}}$ is a Riesz basis for $\mathcal{H}$.

(b) (8 points) Let $\{f_k\}_{k\in\mathbb{N}}$ be a Riesz basis for $\mathcal{H}$. Show that $\{f_k\}_{k\in\mathbb{N}}$ is an exact frame for $\mathcal{H}$.

(c) (4 points) Let $\{f_k\}_{k\in\mathbb{N}}$ be a Riesz basis for $\mathcal{H}$. Show that there are constants A, B, with $0 < A \leq B < \infty$, such that for every collection of scalars $\{c_k\}_{k\in\mathcal{J}}$, with $\mathcal{J} \subset \mathbb{N}$ finite, it holds that

$$A \sum_{k\in\mathcal{J}} |c_k|^2 \leq \left\| \sum_{k\in\mathcal{J}} c_k f_k \right\|_{\mathcal{H}}^2 \leq B \sum_{k\in\mathcal{J}} |c_k|^2.$$

Problem 2 (25 points)

In this problem, we consider a collection $\mathcal{F}$ of probability distributions on $[0,1]$. For simplicity, we assume that $\mathcal{F}$ consists only of distributions that have a density and can be characterized by a cumulative distribution function (cdf) $f : [0,1] \to [0,1]$ which, by definition, is right-continuous ($\lim_{x\to a^+} f(x) = f(a)$, for all $a \in [0,1]$), monotonically non-decreasing, and satisfies $f(0) = 0$, $f(1) = 1$. We study the metric entropy of $\mathcal{F}$ under the L^2-norm, i.e., the L^2-distance between the corresponding cdfs given by

$$\|f-g\|_2 := \left(\int_0^1 |f(x)-g(x)|^2 dx\right)^{1/2}.$$

The purpose of this problem is to establish that[1]

$$\log_2 N(\epsilon; \mathcal{F}, \|\cdot\|_2) \asymp \epsilon^{-1}. \tag{1}$$

To this end, fix $\epsilon_0 = 1/36$ and assume, throughout the remainder of this problem, that $\epsilon \in (0, \epsilon_0)$. Now, set[2] $\epsilon' := (2\lceil \epsilon^{-1}/2\rceil - 3)^{-1}$, $n := 1/\epsilon' - 1 = 2\lceil \epsilon^{-1}/2\rceil - 4$, and construct the following collection of cdfs:

$$\mathcal{U} := \{f_\eta : [0,1] \to [0,1] \ | \ \eta = \{\eta_k\}_{k=1}^n \text{ with } \eta_k \in \{0,1\}, \text{ for } k = 1,2,\dots,n\}, \tag{2}$$

where f_η is a step function defined as (an example of f_η is given in Figure 1)

$$f_\eta(x) = \begin{cases} 0, & x \in [0, \epsilon'), \\ (k-1+\eta_k)\epsilon', & x \in [k\epsilon', (k+1)\epsilon'), \quad 1 \le k \le n, \\ 1, & x = (n+1)\epsilon' = 1. \end{cases} \tag{3}$$

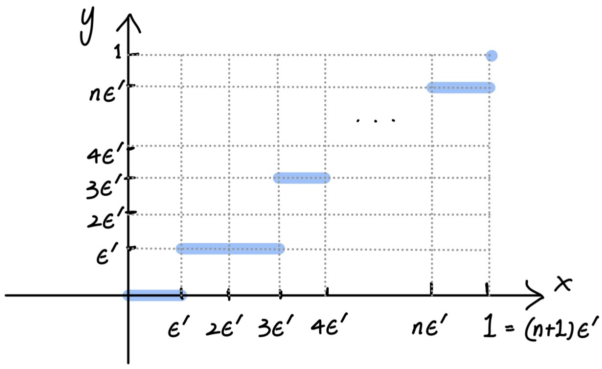

Figure 1: The blue lines represent $f_\eta(x)$.

[1]The notation $f(\epsilon) \asymp g(\epsilon)$ expresses that there exist c_1, c_2, with $0 < c_1 \le c_2$, and $\epsilon_0 > 0$ such that $c_1 g(\epsilon) \le f(\epsilon) \le c_2 g(\epsilon)$, for all $\epsilon \in (0, \epsilon_0)$.

[2]For every $x \in \mathbb{R}$, $\lceil x\rceil := \min\{n \in \mathbb{Z} \ | \ n \ge x\}$. You can use, without proof, that $x \le \lceil x\rceil < x+1$.

(a) (2 points) Verify that every function $f_\eta \in \mathcal{U}$ is right-continuous, monotonically non-decreasing, and satisfies $f_\eta(0) = 0$, $f_\eta(1) = 1$, i.e., f_η is, indeed, a cdf.

(b) (3 points) Given $f_\eta, f_{\eta'} \in \mathcal{U}$, define the Hamming distance[3]

$$d_h(f_\eta, f_{\eta'}) = \operatorname{card}\left(\{k : \eta_k \neq \eta'_k, 1 \leq k \leq n\}\right). \tag{4}$$

Prove that $d_h(f_\eta, f_{\eta'}) > \lceil n/32 \rceil$ implies $\|f_\eta - f_{\eta'}\|_2 > \epsilon'/8$.

Hint: Note that $\epsilon \in (0, \epsilon_0)$ implies $\epsilon' \in (0, 1/2)$.

(c) (3 points) For each $f_\eta \in \mathcal{U}$, prove that

$$\operatorname{card}\left(\{f_{\eta'} \in \mathcal{U} : d_h(f_\eta, f_{\eta'}) \leq \lceil n/32 \rceil\}\right) \leq 2^{n/2}. \tag{5}$$

Hint: Without proof, use the fact that $\sum_{k=0}^{d} \binom{n}{k} \leq \left(\frac{en}{d}\right)^d$, for $d \leq n$, and that $32e < 2^8$. In addition, note that $\epsilon \in (0, \epsilon_0)$ implies $n > 32$.

(d) (4 points) Prove that for all ϵ_1, ϵ_2, with $0 < \epsilon_1 < \epsilon_2$, it holds that

$$M(\epsilon_1; \mathcal{F}, \|\cdot\|_2) \geq M(\epsilon_2; \mathcal{F}, \|\cdot\|_2). \tag{6}$$

(e) (6 points) Combine the results from subproblems (a) to (d) to show that there exists a $c_1 > 0$ such that

$$\log_2 M(\epsilon; \mathcal{F}, \|\cdot\|_2) \geq c_1 \epsilon^{-1}. \tag{7}$$

Hint: Construct an $(\frac{\epsilon'}{8})$-packing based on (a)–(c), note that $\epsilon' = \left(2\lceil \epsilon^{-1}/2 \rceil - 3\right)^{-1} > \epsilon$, and apply the result in (d).

(f) (4 points) You can assume, without giving a proof, that

$$\exists\, p > 0, q > 1 \text{ such that } N(\epsilon; \mathcal{F}, \|\cdot\|_2) \leq 2^{p/\epsilon} \cdot N(q\epsilon; \mathcal{F}, \|\cdot\|_2), \quad \text{for all } \epsilon > 0. \tag{8}$$

Now, based on (8), show that there exists a $c_2 > 0$ such that

$$\log_2 N(\epsilon; \mathcal{F}, \|\cdot\|_2) \leq c_2 \epsilon^{-1}. \tag{9}$$

Hint: Prove that $N(\epsilon_1; \mathcal{F}, \|\cdot\|_2) = 1$, for all $\epsilon_1 \geq 1$.

(g) (3 points) Finally, combining the results in subproblems (e) and (f), derive the scaling behavior (1).

Hint: State the relation between $M(2\epsilon; \mathcal{F}, \|\cdot\|_2)$, $M(\epsilon; \mathcal{F}, \|\cdot\|_2)$, and $N(\epsilon; \mathcal{F}, \|\cdot\|_2)$ from class, and use this relation (without proof).

[3] $\operatorname{card}(A)$ denotes the cardinality of the set A.

Problem 3 (30 points)

Let $\mathbb{N}$ be the set of positive integers, and fix $n \in \mathbb{N}$. The empirical Rademacher complexity of a class $\mathcal{F}$ of functions $f\colon \mathcal{X} \subset \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\mathcal{R}\left(\mathcal{F}\left(x_1^n\right)/n\right) := \frac{1}{n}\mathbb{E}_\varepsilon\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(x_i)\right|\right],$$

where $x_1^n := \{x_1, \dots, x_n\} \subseteq \mathcal{X}$ is fixed and $\{\varepsilon_i\}_{i=1}^n$ is a sequence of Rademacher random variables, i.e., ε_i takes the values $+1$ and -1, each with probability $1/2$, for $i \in \{1, \dots, n\}$.

(a) (5 points) With the convex hull of $\mathcal{F}$ given by

$$\operatorname{conv}(\mathcal{F}) := \left\{x \mapsto \sum_{j=1}^N \alpha_j f_j(x)\colon N \in \mathbb{N}, \alpha_j \geq 0, f_j \in \mathcal{F}, j \in \{1, \dots, N\}, \sum_{j=1}^N \alpha_j = 1\right\},$$

show that $\mathcal{R}\left(\mathcal{F}\left(x_1^n\right)/n\right) = \mathcal{R}\left(\left(\operatorname{conv}(\mathcal{F})\right)\left(x_1^n\right)/n\right)$.
Hint: You may use the following fact without proof:

$$\sup_{(\alpha_1,\dots,\alpha_N)\in\Delta_N}\left|\sum_{j=1}^N \alpha_j v_j\right| = \max_{j\in\{1,\dots,N\}} |v_j|, \qquad \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} \in \mathbb{R}^N,$$

where $\Delta_N := \{(\alpha_1, \dots, \alpha_N) \in [0,1]^d\colon \sum_{j=1}^N \alpha_j = 1\}$.

(b) (8 points) Let now $\mathcal{X} := [-M, M]^d$, with $M > 0$, and consider the function class $\mathcal{F} := \{x \mapsto \langle x, w\rangle\colon w \in \mathbb{R}^d, \|w\|_1 \leq B\}$, where $B > 0$ is a constant. Show that

$$\mathcal{R}\left(\mathcal{F}\left(x_1^n\right)/n\right) \leq BM\sqrt{\frac{2\log(2d)}{n}}.$$

Here, and throughout Problem 3, $\log(\cdot)$ is to the base e.
Hint: Use the result in subproblem (a) together with Lemmata H4 and H8.

(c) (9 points) Let $\rho(y) := \max\{0, y\}$, $y \in \mathbb{R}$, $\mathcal{X} := [-M, M]^d$, $M > 0$, and consider the function class $\mathcal{F} := \{x \mapsto \sum_{j=1}^J u_j \rho(\langle x, v_j\rangle)\colon u_j \in \mathbb{R} \setminus \{0\}, v_j \in \mathbb{R}^d \setminus \{0\}, j \in \{1, \dots, J\}, \sum_{j=1}^J |u_j|\|v_j\|_1 \leq C\}$, where $J \in \mathbb{N}$ and $C > 0$. Show that

$$\mathcal{R}\left(\mathcal{F}\left(x_1^n\right)/n\right) \leq 2CM\sqrt{\frac{2\log(2d)}{n}}.$$

Hint: Use the result in subproblem (b) and Lemma H5.

(d) (8 points) Consider again the setup of subproblem (b), i.e., the function class $\mathcal{F} :=$

$\{x \mapsto \langle x, w\rangle \colon w \in \mathbb{R}^d, \|w\|_1 \le B\}$, $B > 0$, and $\mathcal{X} := [-M, M]^d$, $M > 0$. Further, let $\sigma \colon \mathbb{R} \to \mathbb{R}$ be L-Lipschitz with $\sigma(0) = 0$. For $i \in \mathbb{N}$, define recursively the function class

$$\mathcal{F}_i := \left\{ x \mapsto \sum_{j=1}^{J_i} w_j^i \, \sigma(f_j(x)) \colon f_j \in \mathcal{F}_{i-1}, j \in \{1, \ldots, J_i\}, w^i \in \mathbb{R}^{J_i}, \|w^i\|_1 \le B_i \right\},$$

where $J_i \in \mathbb{N}, B_i > 0$, and set $\mathcal{F}_0 := \mathcal{F}$. Show that

$$\mathcal{R}\left(\mathcal{F}_K\left(x_1^n\right)/n\right) \le BM\sqrt{\frac{2\log(2d)}{n}} \prod_{i=1}^{K} (2B_i L), \quad \text{for some } K \in \mathbb{N}.$$

Hint: You may use without proof that

$$\mathcal{R}\left(\left(\mathcal{G} + \mathcal{H}\right)\left(x_1^n\right)/n\right) \le \mathcal{R}\left(\mathcal{G}\left(x_1^n\right)/n\right) + \mathcal{R}\left(\mathcal{H}\left(x_1^n\right)/n\right),$$

where $\mathcal{G}$ and $\mathcal{H}$ are function classes and $\mathcal{G} + \mathcal{H} := \{g + h \colon g \in \mathcal{G}, h \in \mathcal{H}\}$. Also note that you will need to apply the result in subproblem (b).

Problem 4 (20 points)

Let $m, N \in \mathbb{N}$, $s, t \in \{1, \dots, N\}$, and consider the matrix $A \in \mathbb{C}^{m \times N}$. The s-th restricted isometry constant δ_s of A is the smallest $\delta \geq 0$ such that

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2,$$

for all s-sparse vectors $x \in \mathbb{C}^N$. A vector is said to be s-sparse if it has at most s nonzero entries. The (s,t)-restricted orthogonality constant $\theta_{s,t}$ of A is the smallest $\theta \geq 0$ such that

$$|\langle Au, Av \rangle| \leq \theta \|u\|_2 \|v\|_2,$$

for all disjointly supported s-sparse and t-sparse vectors $u, v \in \mathbb{C}^N$.

Prove that

$$\delta_{ns} \leq (n-1)\theta_{s,s} + \delta_s, \quad \text{for } n, s \in \mathbb{N}.$$