ETH
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 4, 2025

- Do not turn this page before the official start of the exam.
- The problem statements consist of 8 pages including this page.
  Please verify that you have received all 8 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)

Let $n \in \mathbb{N}$ and consider the set of functions $F_n := \{f : \{0,1\}^n \to \mathbb{C}\}$. For $x \in \{0,1\}^n$, we let $\delta_x \in F_n$ be such that

$$\delta_x(y) := \begin{cases} 1, & \text{if } y = x, \\ 0, & \text{if } y \neq x, \end{cases} \qquad \forall y \in \{0,1\}^n. \tag{1}$$

(a) (4 points) Show that $\{\delta_x : x \in \{0,1\}^n\}$ is a basis for F_n.

We define the inner product of two functions $f, g \in F_n$ as

$$\langle f, g \rangle := \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)\overline{g(x)}. \tag{2}$$

We also define $\mathcal{P}(n)$ to be the set of all subsets of $\{1, \dots, n\}$, including the empty set $\varnothing$ and the set $\{1, \dots, n\}$ itself, and we consider the family of functions

$$\mathcal{G} := \{\chi_S : x \mapsto \prod_{i \in S} (-1)^{x_i} : S \in \mathcal{P}(n)\},$$

where $\chi_\varnothing : x \mapsto 1$ by convention. We will show that $\mathcal{G}$ is an orthonormal basis for F_n.

(b) (3 points) Show that for all $S, T \in \mathcal{P}(n)$, it holds that

$$\chi_S(x)\chi_T(x) = \chi_{(S \setminus T) \cup (T \setminus S)}(x), \quad \forall x \in \{0,1\}^n.$$

(c) (4 points) Show that for every $S \neq \varnothing$, χ_S is orthogonal to $\chi_\varnothing$.

(d) (5 points) Show that $\mathcal{G}$ is an orthonormal basis for F_n.

We now study frames for F_n. Namely, for $k \in \mathbb{N}$, define $\omega_k := \exp\left(\frac{2\pi i}{k+1}\right)$ and let

$$\mathcal{H}_k := \left\{ \widetilde{\chi}_{\mathbf{S}}(x) = \prod_{j=1}^{k} \prod_{\ell \in S_j} \omega_k^{x_\ell}, \ \mathbf{S} \in \mathcal{P}(n)^k \right\}, \tag{3}$$

where S_j is the j-th component of $\mathbf{S}$.

(e) (3 points) Show that $\mathcal{H}_1 = \mathcal{G}$. Is $\mathcal{H}_1$ a frame for F_n? If so, specify the frame bounds and determine whether the frame is tight.

It can be shown that for $k \geq 2$, the set of eigenvectors of the operator $\mathbb{T}$ defined by

$$\mathbb{T}f = \sum_{\mathbf{S} \in \mathcal{P}(n)^k} \langle f, \widetilde{\chi}_{\mathbf{S}} \rangle \widetilde{\chi}_{\mathbf{S}}, \quad f \in F_n,$$

is exactly $\mathcal{G}$. Therefore, every eigenvalue (with multiplicities) of $\mathbb{T}$ is associated with some $\chi_S \in \mathcal{G}$, where $S \in \mathcal{P}(n)$. We denote by λ_S the eigenvalue associated with χ_S. It can be shown that

$$\lambda_S = 2^{n(k-1)}(1 + C_k)^{|S|}(1 - C_k)^{n-|S|},\ \forall S \in \mathcal{P}(n),$$

where $|\varnothing| = 0$ and

$$C_k := \left(\cos\left(\frac{\pi}{k+1}\right)\right)^{k+1}.$$

(f) (6 points) Prove that for $k \geq 2$, $\mathcal{H}_k$ is a frame for F_n, specify the frame bounds, and determine whether the frame is tight.
Hint: You might show that $\lambda_\varnothing$ and $\lambda_{\{1,\ldots,n\}}$ are positive and are the minimum and maximum eigenvalues of $\mathbb{T}$, respectively.

Problem 2 (25 points)

Notation: For a matrix $D \in \mathbb{C}^{M\times N}$, and a set $S \subseteq \{1,\dots,N\}$, we define $D_S \in \mathbb{C}^{M\times|S|}$ to be the matrix obtained from D by keeping only the columns indexed by S. Further, $S^c := \{1,\dots,N\}\backslash S$ denotes the complement of the set S in $\{1,\dots,N\}$. D^H stands for the conjugate transpose of the matrix D. For vectors $x, y \in \mathbb{C}^N$, the inner product is given by $\langle x, y\rangle = \sum_{j=1}^N x_j \overline{y}_j = y^H x$ and $\|x\|_0$ denotes the number of non-zero entries of x.

We fix a matrix $D \in \mathbb{C}^{M\times N}$ such that

$$\|d_\ell\|_2 = 1, \qquad \text{for } \ell = 1,\dots,N,$$

where d_ℓ denotes the ℓ-th column of D. Furthermore, we define the *cumulative coherence* of D as

$$\mu_n(D) = \max_{|S|\le n} \max_{\ell\in S^c} \sum_{j\in S} |\langle d_\ell, d_j\rangle|, \qquad \text{for } n = 1,\dots,N,$$

where the first maximum is taken over all subsets $S \subseteq \{1,\dots,N\}$ with cardinality less than or equal to n. Next, we fix $m \in \{2,\dots,N\}$ and assume that

$$\mu_{m-1}(D) + \mu_m(D) < 1. \tag{4}$$

(a) (4 points) Show that for every $S \subseteq \{1,\dots,N\}$ with $|S| \le m$, it holds that

$$\max_{\ell\in S} \sum_{j\in S\backslash\{\ell\}} |\langle d_\ell, d_j\rangle| \le \mu_{m-1}(D). \tag{5}$$

(b) (7 points) Show that $\operatorname{spark}(D) > m$.

(c) (5 points) Show that

$$\frac{\max_{\ell\in S^c} \sum_{j\in S} |\langle d_\ell, d_j\rangle|}{1 - \max_{\ell\in S} \sum_{j\in S\backslash\{\ell\}} |\langle d_\ell, d_j\rangle|} < 1, \qquad \text{for all } S \subseteq \{1,\dots,N\} \text{ with } |S| \le m. \tag{6}$$

(d) (6 points) Show that

$$\max_{\ell\in S^c} \|(D_S)^\dagger d_\ell\|_1 < 1, \qquad \text{for all } S \subseteq \{1,\dots,N\} \text{ with } |S| \le m, \tag{7}$$

where $(D_S)^\dagger = ((D_S^H)D_S)^{-1}(D_S)^H$ denotes the Moore-Penrose pseudo-inverse of the matrix D_S.
Hint: Subproblem (c), Definition H2 and Lemmata H3, H4, and H5 might be useful.

(e) (3 points) Let $x \in \mathbb{C}^N$ be such that $\|x\|_0 \le m$ and consider the recovery problem (P1) for D and x (Definition H6). Show that (P1) uniquely recovers x.
Hint: Use Theorem H7 and the result in subproblem (b).

Problem 3 (25 points)

In the lecture, we learned that the metric entropy of the ball

$$B_C^d := \{x \in \mathbb{R}^d \mid \|x\|_2 \le C\}$$

satisfies[1]

$$\log N(\epsilon; B_C^d, \|\cdot\|_2) \asymp d \log\left(\epsilon^{-1}\right). \tag{8}$$

In this problem, we consider a set of matrices in $\mathbb{R}^{n\times n}$ and the 2-norm

$$\|A\|_2 := \sup_{x\in\mathbb{R}^n, \|x\|_2=1} \|Ax\|_2 .$$

The goal is to analyze the metric entropy of a ball in $\mathbb{R}^{n\times n}$ defined as

$$\mathbb{M}_K^{n\times n} := \{A \in \mathbb{R}^{n\times n} \mid \|A\|_2 \le K\}.$$

(a) (2 points) State the double-inequality linking the quantities $M(2\epsilon; \mathbb{M}_K^{n\times n}, \|\cdot\|_2)$, $N(\epsilon; \mathbb{M}_K^{n\times n}, \|\cdot\|_2)$, and $M(\epsilon; \mathbb{M}_K^{n\times n}, \|\cdot\|_2)$.

(b) (3 points) Consider the mapping $V : \mathbb{M}_K^{n\times n} \to \mathbb{R}^{n^2}$ such that for $A \in \mathbb{M}_K^{n\times n}$, $V(A) \in \mathbb{R}^{n^2}$ is the vector with[2]

$$(V(A))_{(i-1)n+j} = A_{i,j}, \qquad \text{for } i, j \in \{1, 2, \ldots, n\}.$$

Prove that V is an isometric isomorphism between $\left(\mathbb{M}_K^{n\times n}, \|\cdot\|_F\right)$ and $\left(V(\mathbb{M}_K^{n\times n}), \|\cdot\|_2\right)$.

(c) (4 points) For $A \in \mathbb{R}^{n\times n}$, prove the following relations between 2-norm and Frobenius norm[3],

$$\|A\|_2 \le \|A\|_F \le \sqrt{n}\, \|A\|_2 .$$

Hint: You may use, without proof, that $\operatorname{tr}(A^TA) = \sum_{i=1}^n \lambda_i(A^TA)$, *where* $\lambda_1(A^TA) \ge \lambda_2(A^TA) \ge \cdots \ge \lambda_n(A^TA) \ge 0$ *are the eigenvalues of* A^TA.

(d) (4 points) Prove that

$$B_K^{n^2} \subseteq V(\mathbb{M}_K^{n\times n}) \subseteq B_{\sqrt{n}K}^{n^2}.$$

(e) (4 points) Let (X, ρ_X) be a metric space and consider the compact sets $\mathcal{C}_1, \mathcal{C}_2$ with

[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 the vector $x \in \mathbb{R}^d$, we denote its j-th entry by x_j or $(x)_j$. For the matrix $A \in \mathbb{R}^{n\times n}$, we denote its entry in row i and column j by $A_{i,j}$.

[3] For the matrix $A \in \mathbb{R}^{n\times n}$, we denote its Frobenius norm by $\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^n (A_{i,j})^2}$, and its trace by $\operatorname{tr}(A) = \sum_{i=1}^n A_{i,i}$

$\mathcal{C}_1 \subseteq \mathcal{C}_2 \subseteq X$. Prove that

$$M(\epsilon; \mathcal{C}_1, \rho_X) \leq M(\epsilon; \mathcal{C}_2, \rho_X).$$

(f) (8 points) Derive the scaling behavior of $\log N(\epsilon; \mathbb{M}_K^{n\times n}, \|\cdot\|_2)$. *Hint: Use results from subproblems (a)-(e), and Lemmata H12 and H13 from the Handout.*

Problem 4 (25 points)

Let $(\mathcal{F}, \langle\cdot,\cdot\rangle)$ be a Hilbert space of functions mapping from the non-empty set $\mathcal{X} \subseteq \mathbb{R}^d$ into $\mathbb{R}$. Suppose that there exists a function $k\colon \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ such that $\phi_x(\cdot) := k(x,\cdot) \in \mathcal{F}$ for all $x \in \mathcal{X}$, and the so-called *reproducing property*

$$f(x) = \langle f, \phi_x\rangle$$

holds for all $f \in \mathcal{F}$ and all $x \in \mathcal{X}$. Consider the function class $\mathcal{F}_b := \{f \in \mathcal{F}\colon \|f\| \leq b\}$ for some $b > 0$, where $\|f\| := \sqrt{\langle f, f\rangle}$, $f \in \mathcal{F}$. In this problem, we study the misclassification probability of classifiers of the form $x \mapsto \operatorname{sign}(f(x))$, $x \in \mathcal{X}$, with $f \in \mathcal{F}_b$. Here, for $a \in \mathbb{R}$, $\operatorname{sign}(a) = -1$ if $a < 0$, and $\operatorname{sign}(a) = 1$ if $a \geq 0$.

(a) (8 points) With the empirical Rademacher complexity according to Definition H14, show that for $\{x_i\}_{i=1}^n \subseteq \mathcal{X}$,

$$\mathcal{R}\left(\mathcal{F}_b\left(\{x_i\}_{i=1}^n\right)/n\right) \leq \frac{b}{n}\sqrt{\operatorname{tr}(K)},$$

where $K := (k(x_i, x_j))_{i,j=1}^n \in \mathbb{R}^{n\times n}$.

(b) (5 points) For $\gamma > 0$, let

$$\rho_\gamma(u) := \begin{cases} 0, & \text{if } u \leq -\gamma, \\ 1 + u/\gamma, & \text{if } -\gamma < u \leq 0, \\ 1, & \text{if } u > 0, \end{cases} \qquad u \in \mathbb{R}.$$

Consider the function class $\mathcal{H} := \{h\colon \mathcal{X} \times \{-1,1\} \to \mathbb{R}\colon h(x,y) = \rho_\gamma(-yf(x)), \forall (x,y) \in \mathcal{X}\times\{-1,1\}, f \in \mathcal{F}_b\}$. Show that for $\{(x_i,y_i)\}_{i=1}^n \subseteq \mathcal{X}\times\{-1,1\}$,

$$\mathcal{R}\left(\mathcal{H}\left(\{(x_i,y_i)\}_{i=1}^n\right)/n\right) \leq \frac{1}{\gamma}\mathcal{R}\left(\mathcal{F}_b\left(\{x_i\}_{i=1}^n\right)/n\right).$$

Hint: Apply Lemma H16.

(c) (7 points) Let $\mathcal{G}$ be a class of real-valued functions on the non-empty set $\mathcal{Z}$ taking values in $[0,1]$, and let $\{Z_i\}_{i=1}^n$ be i.i.d. random variables taking values in $\mathcal{Z}$. Deduce from Theorem H17 that for $\delta > 0$, with probability $\geq 1-\delta$,

$$\sup_{g\in\mathcal{G}}\left(\mathbb{E}[g(Z)] - \frac{1}{n}\sum_{i=1}^n g(Z_i)\right) \leq 2\mathcal{R}\left(\mathcal{G}\left(\{Z_i\}_{i=1}^n\right)/n\right) + 3\sqrt{\frac{2\log(2/\delta)}{n}}.$$

Here, and throughout this problem, $\log(\cdot)$ denotes the logarithm to base e.
Hint: Apply Lemma H18 and recall from the lecture that $\mathcal{Z}^n \to \mathbb{R}, (z_1,\dots,z_n) \mapsto$

$\mathcal{R}\left(\mathcal{G}\left(\{z_i\}_{i=1}^n\right)/n\right)$ satisfies (4) in Lemma H18 with $L = 2/n$. You may use this result without proof.

(d) (5 points) Given a collection of n i.i.d. samples $\{(X_i, Y_i)\}_{i=1}^n \subseteq \mathcal{X} \times \{-1, 1\}$, show that for $\delta > 0$, with probability $\geq 1 - \delta$, for all $f \in \mathcal{F}_b$,

$$\mathbb{P}(\mathrm{sign}(f(X)) \neq Y) \leq \frac{1}{n}\sum_{i=1}^n \rho_\gamma(-Y_i f(X_i)) + \frac{2b}{n\gamma}\sqrt{\mathrm{tr}(\mathsf{K})} + 3\sqrt{\frac{2\log(2/\delta)}{n}},$$

where $\mathsf{K} := (k(X_i, X_j))_{i,j=1}^n \in \mathbb{R}^{n \times n}$.
Hint: Use the results of subproblems (a)-(c).