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 14, 2020

Please note:

- Exam duration: 180 minutes
- Maximum number of points: 100
- You are allowed to use any printed or handwritten material (i.e., books, lecture and discussion session notes, summaries), but no computers, tablets, smart phones or other electronic devices.
- Your solutions should be explained in detail and your handwriting needs to be clean and readable.
- Please do not use red or green pens. You may use pencils.
- Please note that the ETHZ "Disziplinarordnung RSETHZ 361.1" applies.

Before you start:

1. The problem statements consist of 7 pages including this page. Please verify that you have received all 7 pages.
2. Please fill in your name and your Legi-number below.
3. Please place an identification document on your desk so we can verify your identity.

During the exam:

4. For your solutions, please use only the empty sheets provided by us. Should you need more paper, please let us know.
5. Each problem consists of several subproblems. If you do not provide a solution to a subproblem, you may, whenever applicable, nonetheless assume its conclusion in the ensuing subproblems.

After the exam:

6. Please number all the sheets you want to hand in. Please specify below the number of *additional* sheets you want to hand in (excluding the sheets with the problem statements). All sheets containing problem statements must be handed in.

Family name: .... First name: ....

Legi-No.: ....

Number of additional sheets handed in: ....

Signature: ....

Problem 1

Let $\{g_k\}_{k=1}^N$ and $\{h_j\}_{j=1}^N$ be orthonormal bases for $\mathbb{C}^N$, and let $\mathcal{P}$ be a probability distribution on $[N] := \{1, 2, \ldots, N\}$ such that

$$p_n := \mathbb{P}_{\boldsymbol{t}\sim\mathcal{P}}(\boldsymbol{t} = n) > 0, \quad \text{for all } n \in [N],$$

but otherwise arbitrary, where $\boldsymbol{t}$ denotes a random variable taking values in $[N]$. Furthermore, set

$$\tilde{\mu}_j := \max_{k\in[N]} |\langle g_k, h_j\rangle|, \quad \text{for } j \in [N],$$

and

$$\kappa := \max_{j\in[N]} p_j^{-1} \tilde{\mu}_j^2.$$

For a set of scalar or vector quantities u_j indexed by $j \in [N]$, we write $u_{\boldsymbol{t}}$ for the random variable taking on the value u_n on the event $\{\boldsymbol{t} = n\}$, for each $n \in [N]$.

(a) Define a probability distribution $\widetilde{\mathcal{P}}$ on $[N]$ according to

$$\mathbb{P}_{\boldsymbol{t}\sim\widetilde{\mathcal{P}}}(\boldsymbol{t} = n) = \frac{\tilde{\mu}_n^2}{\tilde{\mu}_1^2 + \cdots + \tilde{\mu}_N^2}, \quad \text{for } n \in [N].$$

  (i) Show that $\sum_{j=1}^N \tilde{\mu}_j^2 \geq 1$.

  (ii) Show that $\kappa^{-1} \leq \mathbb{E}_{\boldsymbol{t}\sim\widetilde{\mathcal{P}}}\left[p_{\boldsymbol{t}}\, \tilde{\mu}_{\boldsymbol{t}}^{-2}\right]$.

  (iii) Use the inequality in (a)(ii) to show that $\kappa \geq \sum_{j=1}^N \tilde{\mu}_j^2$, with equality if and only if $\mathcal{P} = \widetilde{\mathcal{P}}$.

(b) Now, let $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_m$ be independent samples drawn from the distribution $\mathcal{P}$ and define a random matrix $A \in \mathbb{C}^{m\times N}$ according to $A_{\ell k} = (mp_{\boldsymbol{t}_\ell})^{-\frac{1}{2}} \langle g_k, h_{\boldsymbol{t}_\ell}\rangle$, for $\ell \in [m]$, $k \in [N]$. Furthermore, let $x \in \mathbb{C}^N$ be an arbitrary s-sparse vector with $\|x\|_2 = 1$ and, for $\ell \in [m]$, define

$$X_\ell = |\langle x, Y_\ell\rangle|^2 - \tfrac{1}{m},$$

where $Y_\ell = (\overline{A}_{\ell 1}\, \overline{A}_{\ell 2} \, \cdots \, \overline{A}_{\ell N})$ is the complex conjugate of the ℓ^{th} row of A.

  (i) Show that $\mathbb{E}[X_\ell] = 0$, $|X_\ell| \leq sm^{-1}\kappa$, and $\mathbb{E}[|X_\ell|^2] \leq sm^{-2}\kappa$.

  (ii) Use Bernstein's inequality together with the results of part (b)(i) to establish

$$\mathbb{P}\left(\left|\|Ax\|_2^2 - 1\right| \geq t\right) \leq 2\exp\left(-\frac{3t^2 m}{8s\kappa}\right), \quad \text{for } t \in (0, 1).$$

Problem 2

In this problem you may use—without proof—the following form of the Hölder inequality:

$$|\langle x, y\rangle| \le \|x\|_1 \|y\|_\infty, \quad \text{for all } x, y \in \mathbb{C}^N,$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product on $\mathbb{C}^N$.

For a vector $u \in \mathbb{C}^N$, a matrix $B \in \mathbb{C}^{m\times N}$, and a set $S \subset \{1, \ldots, N\}$, we define $u_S \in \mathbb{C}^{|S|}$ to be the vector obtained from u by keeping only the entries indexed by S, and similarly, we define $B_S \in \mathbb{C}^{m\times|S|}$ to be the matrix obtained from B by keeping only the columns indexed by S. We write $\mathrm{im}(B)$ and $\ker(B)$ for the column span and the nullspace of B, respectively. We additionally define the complex sign of u as the vector $\mathrm{sgn}(u) \in \mathbb{C}^N$ given by

$$\big(\mathrm{sgn}(u)\big)_k = \begin{cases} u_k/|u_k|, & \text{if } u_k \neq 0 \\ 0, & \text{else} \end{cases}.$$

In the remainder of the problem we fix an $x \in \mathbb{C}^N \setminus \{0\}$, let $S = \{k \in [N] : x_k \neq 0\}$ be its support, and define $\overline{S} = \{1, \ldots, N\} \setminus S$. We also fix a matrix $A \in \mathbb{C}^{m\times N}$ with columns $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_N \in \mathbb{C}^m$, assume that $A_S^{\mathsf{H}} A_S \in \mathbb{C}^{|S|\times|S|}$ is invertible (where A_S^{H} denotes the Hermitian transpose of A_S), and denote

$$\alpha := \big\|\big(A_S^{\mathsf{H}} A_S\big)^{-1}\big\|_2 \cdot \max_{k\in\overline{S}} \big\|A_S^{\mathsf{H}}\, \boldsymbol{a}_k\big\|_2,$$

where $\|\cdot\|_2$ in the first factor denotes the matrix operator norm with respect to the ℓ^2-norm on $\mathbb{C}^{|S|}$, i.e.,

$$\big\|\big(A_S^{\mathsf{H}} A_S\big)^{-1}\big\|_2 := \max_{\substack{u\in\mathbb{C}^{|S|} \\ \|u\|_2\le 1}} \big\|\big(A_S^{\mathsf{H}} A_S\big)^{-1} u\big\|_2.$$

(a) Let $v \in \mathbb{C}^N$ be arbitrary and write $z = x - v$.

  (i) Show that $\|x\|_1 = \langle x, \mathrm{sgn}(x)\rangle$.

  (ii) Establish

$$\|x\|_1 \le \|z\|_1 - \|v_{\overline{S}}\|_1 + |\langle v, \mathrm{sgn}(x)\rangle|.$$

  [Hint: Use $\|\mathrm{sgn}(x_S)\|_\infty = 1$ and $\|v_{\overline{S}}\|_1 = \|z_{\overline{S}}\|_1$.]

(b) Let $v \in \ker(A)$ be arbitrary.

  (i) Show that $A_S\, v_S + A_{\overline{S}}\, v_{\overline{S}} = 0$.

  (ii) Establish $\|v_S\|_2 \le \alpha\, \|v_{\overline{S}}\|_1$.

[Hint: Note that $v_S = \big(A_S^{\mathsf{H}} A_S\big)^{-1} A_S^{\mathsf{H}} A_S\, v_S$ and $A_{\overline{S}}\, v_{\overline{S}} = \sum_{k\in\overline{S}} \boldsymbol{a}_k v_k$.]

(c) Next, suppose that $v \in \ker(A)$ and $u \in \operatorname{im}(A^{\mathsf{H}})$. Show that

$$|\langle v, \operatorname{sgn}(x)\rangle| \leq \|u_S - \operatorname{sgn}(x_S)\|_2 \, \|v_S\|_2 \; + \; \|u_{\overline{S}}\|_\infty \, \|v_{\overline{S}}\|_1.$$

[Hint: Recall that $\operatorname{im}(A^{\mathsf{H}}) = (\ker(A))^\perp$.]

(d) Suppose that there exists a $u \in \operatorname{im}(A^{\mathsf{H}})$ satisfying the following condition:

$$\alpha \, \|u_S - \operatorname{sgn}(x_S)\|_2 + \|u_{\overline{S}}\|_\infty < 1.$$

(i) Use the results of parts (b) and (c) to show that

$$|\langle v, \operatorname{sgn}(x)\rangle| < \|v_{\overline{S}}\|_1, \quad \text{for all } v \in \ker(A) \setminus \{0\}.$$

(ii) Use (a)(ii) and (d)(i) to show that x is the unique solution to the following optimization problem:

$$\min_{z \in \mathbb{C}^N} \|z\|_1 \quad \text{subject to} \quad Az = Ax.$$

Problem 3

We consider a class $\mathcal{F}$ of functions $f\colon \mathbb{R}^d \to \mathbb{R}$ that are b-uniformly bounded, i.e.,

$$\|f\|_\infty \leq b, \quad \forall f \in \mathcal{F}.$$

Given an integer $n \geq 1$, recall the definition of empirical Rademacher complexity

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

where $x_1^n \coloneqq \{x_1, \ldots, x_n\}$ with $x_i \in \mathbb{R}^d$, $i = 1, \ldots, n$, 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 equiprobably, for $i = 1, \ldots, n$. The Rademacher complexity is then defined as

$$\mathcal{R}_n\left(\mathcal{F}\right) \coloneqq \mathbb{E}_X\left[\mathcal{R}_n\left(\mathcal{F}\left(X_1^n\right)/n\right)\right],$$

where $X_1^n \coloneqq \{X_1, \ldots, X_n\}$ and the X_i are i.i.d. random variables, for $i = 1, \ldots, n$.

(a) Show that

$$\mathcal{R}_n\left(\mathcal{F}\right) \leq \mathcal{R}_n\left(\mathcal{F}\left(X_1^n\right)/n\right) + \sqrt{\frac{2b^2\log(1/\delta)}{n}},$$

with probability at least $1-\delta$.

Hint: Show that the empirical Rademacher complexity satisfies the bounded difference property and use the one sided bounded difference inequality (provided in the handout).

(b) Consider now two general classes of functions, $\mathcal{F}_1$ and $\mathcal{F}_2$, and define

$$\mathcal{F}_{|\cdot|} \coloneqq \left\{|f_1 - f_2| \mid f_1 \in \mathcal{F}_1 \text{ and } f_2 \in \mathcal{F}_2\right\}.$$

Derive the following bound:

$$\mathcal{R}_n\left(\mathcal{F}_{|\cdot|}\left(x_1^n\right)/n\right) \leq 2\,\mathcal{R}_n\left(\mathcal{F}_1\left(x_1^n\right)/n\right) + 2\,\mathcal{R}_n\left(\mathcal{F}_2\left(x_1^n\right)/n\right). \tag{1}$$

Hint: Use the Ledoux-Talagrand contraction lemma provided in the handout.

(c) Define $\mathcal{F}_{\max} \coloneqq \{f\colon x \mapsto \max\{f_1(x), f_2(x)\} \mid f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$. Using (1), derive the following upper bound

$$\mathcal{R}_n\left(\mathcal{F}_{\max}\left(x_1^n\right)/n\right) \leq \frac{3}{2}\Big(\mathcal{R}_n\left(\mathcal{F}_1\left(x_1^n\right)/n\right) + \mathcal{R}_n\left(\mathcal{F}_2\left(x_1^n\right)/n\right)\Big).$$

Hint: First prove that $\max\{a,b\} = \frac{|a-b|+a+b}{2}$ *for all* $a, b \in \mathbb{R}$.

Problem 4

In the last discussion session, we used, without providing a proof, a relation between the metric entropy of a uniformly bounded class of functions and its VC dimension, in the case where the latter is finite. The present problem is concerned with a slightly weaker version of this result.

Namely, given a general set $\mathcal{X}$, we consider a class of subsets of $\mathcal{X}$, denoted by $\mathcal{S}$, and we assume that the corresponding class of indicator functions $\mathcal{F}_{\mathcal{S}} := \{\mathbb{1}_S \mid S \in \mathcal{S}\}$ has finite VC dimension ν. Given a probability measure $\mathbb{Q}$ on $\mathcal{X}$, we consider the distance associated with the $L_1(\mathbb{Q})$-norm on $\mathcal{F}_{\mathcal{S}}$, i.e., the distance that assigns to $\mathbb{1}_{S_1} \in \mathcal{F}_{\mathcal{S}}$ and $\mathbb{1}_{S_2} \in \mathcal{F}_{\mathcal{S}}$ the value

$$\|\mathbb{1}_{S_1} - \mathbb{1}_{S_2}\|_{L_1(\mathbb{Q})} := \mathbb{E}_{X\sim\mathbb{Q}}\left[|\mathbb{1}_{S_1}(X) - \mathbb{1}_{S_2}(X)|\right].$$

The aim of this problem is to prove that $\mathcal{F}_{\mathcal{S}}$ has metric entropy $N(\delta; \mathcal{F}_{\mathcal{S}}, L_1(\mathbb{Q}))$ satisfying

$$N(\delta; \mathcal{F}_{\mathcal{S}}, L_1(\mathbb{Q})) \leq (2\nu)^{2\nu-1}\left(\frac{3}{\delta}\right)^{2\nu}. \tag{2}$$

In order to establish (2), we fix $\delta > 0$ and take $\{\mathbb{1}_{S_1}, \ldots, \mathbb{1}_{S_M}\}$ to be a maximal δ-packing of size M of $\mathcal{F}_{\mathcal{S}}$ in the $L_1(\mathbb{Q})$-norm, that is

$$\|\mathbb{1}_{S_i} - \mathbb{1}_{S_j}\|_{L_1(\mathbb{Q})} > \delta, \quad \forall i \neq j.$$

We assume in what follows that $\nu \geq 2$ and δ is small enough for

$$3\log(M) > \delta(\nu + 1) \tag{3}$$

to hold.

(a) Prove that the metric entropy $N(\delta; \mathcal{F}_{\mathcal{S}}, L_1(\mathbb{Q}))$ is upper-bounded according to

$$N(\delta; \mathcal{F}_{\mathcal{S}}, L_1(\mathbb{Q})) \leq M.$$

(b) Given two distinct sets S_i and S_j, we define their symmetric difference as

$$S_i \triangle S_j := (S_i \cup S_j) \setminus (S_i \cap S_j).$$

Show that a random variable X drawn from $\mathbb{Q}$ satisfies

$$\mathbb{P}\left[X \notin (S_i \triangle S_j)\right] < 1 - \delta.$$

<u>Hint</u>: Note that $\mathbb{1}_{S_i \triangle S_j}(\cdot) = \left|\mathbb{1}_{S_i}(\cdot) - \mathbb{1}_{S_j}(\cdot)\right|$.

(c) Suppose that we are given n samples X_k, $k = 1, \ldots, n$, drawn i.i.d. from $\mathbb{Q}$. Show that, given distinct sets S_i and S_j, the sets $S_i \cap \{X_1, \ldots, X_n\}$ and $S_j \cap \{X_1, \ldots, X_n\}$ are distinct if and only if $X_k \in (S_i \bigtriangleup S_j)$ for at least one $k \in \{1, \ldots, n\}$.

(d) Using the results of subproblems (b) and (c), show that the probability of all the sets $S_i \cap \{X_1, \ldots, X_n\}$, for $i = 1, \ldots, M$, being distinct is at least $1 - \binom{M}{2}(1-\delta)^n$.

Hint: Note that, by subproblem (c), the complementary event of the sets $S_i \cap \{X_1, \ldots, X_n\}$, $i = 1, \ldots, M$, being distinct is equivalent to the existence of at least one pair of sets S_i and S_j among the $\binom{M}{2}$ possible pairs such that $X_k \notin (S_i \bigtriangleup S_j)$, for all $k \in \{1, \ldots, n\}$.

(e) From now on, we take $n = \frac{3\log(M)}{\delta} - 1$. Using the inequality $(1-\delta)^n \leq e^{-n\delta}$ together with the assumptions (3) and $\nu \geq 2$, show that

$$1 - \binom{M}{2}(1-\delta)^n \geq 1 - M^2 e^{-n\delta} > 0,$$

and deduce that there must exist a set of n points $\{x_1, \ldots, x_n\}$ from which $\mathcal{S}$ picks out at least M subsets, i.e.,

$$M \leq |\mathcal{S}(\{x_1, \ldots, x_n\})|,$$

where $\mathcal{S}(\{x_1, \ldots, x_n\}) := \{(\mathbb{1}_S(x_1), \ldots, \mathbb{1}_S(x_n)) \mid S \in \mathcal{S}\}$.

(f) By upper-bounding $|\mathcal{S}(\{x_1, \ldots, x_n\})|$, prove that

$$M \leq \left(\frac{3\log(M)}{\delta}\right)^{\nu}. \tag{4}$$

Hint: Use the Vapnik-Chervonenkis-Sauer-Shelah lemma (provided in the handout). You will also need assumption (3) and $n = \frac{3\log(M)}{\delta} - 1$.

(g) Conclude the proof of the desired result (2).

Hint: Using that $te^{-t} \leq 1$, for $t \geq 0$, first prove that

$$\sup_{t \geq 0}\left(t^{2\nu}e^{-t}\right) \leq (2\nu)^{2\nu-1},$$

and show that inequality (4) can be rewritten as

$$M \leq \frac{\log(M)^{2\nu}}{M}\left(\frac{3}{\delta}\right)^{2\nu}.$$

You will also need the result established in subproblem (a).