ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Prof. Dr. Helmut Bölcskei
Chair for Mathematical Information Science

Handout
Examination on Mathematics of Information
August 4, 2025

Definition H1. *The spark of the matrix $D \in \mathbb{C}^{M\times N}$, denoted by $\operatorname{spark}(D)$, is defined as the cardinality of the smallest set of linearly dependent columns of D.*

Definition H2. *The 1-operator norm of the matrix $D \in \mathbb{C}^{M\times N}$ is defined as*

$$\|D\|_{1,1} = \max_{v\in\mathbb{C}^N} \frac{\|Dv\|_1}{\|v\|_1}.$$

Lemma H3. *For $D \in \mathbb{C}^{M\times N}$ and $v \in \mathbb{C}^N$, we have*

$$\|Dv\|_1 \le \|D\|_{1,1}\|v\|_1.$$

Lemma H4. *For the matrix $D \in \mathbb{C}^{M\times N}$, we have*

$$\|D\|_{1,1} = \max_{\ell\in\{1,\dots,N\}} \sum_{j=1}^{M} |D_{j,\ell}| = \max_{\ell\in\{1,\dots,N\}} \|d_\ell\|_1,$$

with d_ℓ denoting the ℓ-th column of D.

Lemma H5. *Let $G \in \mathbb{C}^{M\times M}$ be such that $\|G\|_{1,1} < 1$. Then, we have*

$$\|(\mathbb{I}+G)^{-1}\|_{1,1} \le \frac{1}{1-\|G\|_{1,1}},$$

where $\mathbb{I}$ denotes the identity matrix of size $M \times M$.

Definition H6. *The recovery problem (P1) for the matrix $D \in \mathbb{C}^{M\times N}$ and the vector $x \in \mathbb{C}^N$ is defined by*

$$\arg\min_{\widehat{x}} \|\widehat{x}\|_1 \quad \text{subject to } D\widehat{x} = Dx. \tag{P1}$$

We say that (P1) uniquely recovers x iff the minimizer of (P1) is unique and equal to x.

Theorem H7 (Homework sheet 5, Problem 3b)**.** *Consider (P1) for the matrix $D \in \mathbb{C}^{M\times N}$ and the vector $x \in \mathbb{C}^N$ and let $S \subset \{1,\dots,N\}$ be the support set of x. (P1) uniquely recovers x if the following sufficient condition holds:*

$$\mathcal{N}(D_S) = \{0\} \qquad \text{and} \qquad \left|\langle (D_S)^\dagger d_k, \operatorname{sgn}(x_S)\rangle\right| < 1,\ \forall k \in S^c,$$

where d_j denotes the j-th column of D and $(D_S)^\dagger := ((D_S)^H D_S)^{-1}(D_S)^H$ is the pseudo-inverse of the matrix D_S, which is obtained from D by keeping only the columns indexed by S. Furthermore, $\operatorname{sgn}(x_S)$ is the vector obtained from x by collecting $\operatorname{sgn}(x_j) := \frac{x_j}{|x_j|}$, $j \in S$, in a $|S|$-dimensional vector.

Lemma H8. *Let $d \in \mathbb{N}$ and $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. It holds that*

$$|\langle x, y\rangle| \leq \|x\|_1 \|y\|_\infty, \quad x, y \in \mathbb{K}^d.$$

Definition H9 (Covering number). *Let (X, ρ) be a metric space and let $\mathcal{C} \subseteq X$ be compact. The set $\{x_1, x_2, \ldots, x_N\} \subseteq \mathcal{C}$ is an ϵ-covering for $(\mathcal{C}, \rho)$ if, for each $x \in \mathcal{C}$, there exists an $i \in \{1, 2, \ldots, N\}$ so that $\rho(x, x_i) \leq \epsilon$. The ϵ-covering number $N(\epsilon; \mathcal{C}, \rho)$ is the cardinality of a smallest ϵ-covering for $(\mathcal{C}, \rho)$.*

Definition H10 (Packing number). *Let (X, ρ) be a metric space and let $\mathcal{C} \subseteq X$ be compact. The set $\{x_1, x_2, \ldots, x_M\} \subseteq \mathcal{C}$ is an ϵ-packing for $(\mathcal{C}, \rho)$ if, for every pair $i, j \in \{1, 2, \ldots, M\}$ with $i \neq j$, we have $\rho(x_i, x_j) > \epsilon$. The ϵ-packing number $M(\epsilon; \mathcal{C}, \rho)$ is the cardinality of a largest ϵ-packing for $(\mathcal{C}, \rho)$.*

Definition H11 (Isometric isomorphism). *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed vector spaces and let $T : X \to Y$ be a linear mapping. We say that T is an isometric isomorphism if and only if T is bijective and an isometry, i.e., for every $x \in X$, we have $\|x\|_X = \|T(x)\|_Y$.*

Lemma H12. *Let (X, ρ_X) and (Y, ρ_Y) be metric spaces and consider the compact sets $\mathcal{C}_X \subseteq X$ and $\mathcal{C}_Y \subseteq Y$. Assume that there exists an isometric isomorphism $f : \mathcal{C}_X \to \mathcal{C}_Y$, i.e., f is bijective and for every pair $a, b \in \mathcal{C}_X$, one has $\rho_Y(f(a), f(b)) = \rho_X(a, b)$. Then,*

$$N(\epsilon; \mathcal{C}_X, \rho_X) = N(\epsilon; \mathcal{C}_Y, \rho_Y) \quad \textit{and} \quad M(\epsilon; \mathcal{C}_X, \rho_X) = M(\epsilon; \mathcal{C}_Y, \rho_Y). \tag{1}$$

Lemma H13. *Let (X, ρ_1) be a metric space and consider a compact set $\mathcal{C}_X \subseteq X$. Assume that for another metric ρ_2 on X, there exist constants $C_1 \geq C_2 > 0$ such that for all $a, b \in \mathcal{C}_X$,*

$$C_2 \rho_1(a, b) \leq \rho_2(a, b) \leq C_1 \rho_1(a, b), \tag{2}$$

then $\mathcal{C}_X$ is also compact under ρ_2. Moreover, we have

$$N(\epsilon/C_2; \mathcal{C}_X, \rho_1) \leq N(\epsilon; \mathcal{C}_X, \rho_2) \leq N(\epsilon/C_1; \mathcal{C}_X, \rho_1). \tag{3}$$

Definition H14. *Let $\mathbb{N}$ be the set of positive integers and fix $n \in \mathbb{N}$. The* empirical Rademacher complexity *of the class $\mathcal{F}$ of functions $f \colon \mathcal{X} \to \mathbb{R}$ is defined as*

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

where $\{x_i\}_{i=1}^n \subseteq \mathcal{X}$ *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* $i \in \{1, \dots, n\}$*. Given a collection of random variables* $\{X_i\}_{i=1}^n$*, the* Rademacher complexity *of* $\mathcal{F}$ *is given by*

$$\mathcal{R}_n(\mathcal{F}) = \mathbb{E}\left[\mathcal{R}\left(\mathcal{F}\left(\{X_i\}_{i=1}^n\right)/n\right)\right].$$

Lemma H15 (Jensen's inequality)**.** *Let* X *be an integrable real-valued random variable, and suppose that* $\varphi\colon \mathbb{R} \to \mathbb{R}$ *is convex. Then,*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

Lemma H16. *Let* $\phi\colon \mathbb{R} \to \mathbb{R}$ *be an* L*-Lipschitz function and* $\mathcal{F}$ *a class of functions. Let* $\phi \circ \mathcal{F} \coloneqq \{\phi \circ f \mid f \in \mathcal{F}\}$*. Then, we have*

$$\mathcal{R}\left(\left(\phi \circ \mathcal{F}\right)\left(\{x_i\}_{i=1}^n\right)/n\right) \leq L\mathcal{R}\left(\mathcal{F}\left(\{x_i\}_{i=1}^n\right)/n\right).$$

Theorem H17. *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* $\mathcal{Z}$*. For* $\delta' > 0$*, with probability* $\geq 1 - \delta'$*,*

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

Here, $\log(\cdot)$ *denotes the logarithm to base* e*.*

Lemma H18 (One-sided bounded difference inequality)**.** *Let* $\mathcal{Z}$ *be a non-empty set, and suppose that* $f\colon \mathcal{Z}^n \to \mathbb{R}$ *is such that*

$$|f(z_1, \dots, z_{i-1}, z_i, z_{i+1}, \dots, z_n) - f(z_1, \dots, z_{i-1}, y, z_{i+1}, \dots, z_n)| \leq L, \tag{4}$$

for every $i \in \{1, \dots, n\}$*, every* $(z_1, \dots, z_n) \in \mathcal{Z}^n$*, and every* $y \in \mathcal{Z}$*. Also suppose that the random vector* $Z = (Z_1, \dots, Z_n)$ *has i.i.d. components. Then, we have*

$$\mathbb{P}\big[\mathbb{E}[f(Z)] - f(Z) > \epsilon\big] \leq e^{-\frac{2\epsilon^2}{nL^2}}, \quad \forall \epsilon \geq 0.$$