Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Prof. Dr. Helmut Bölcskei
Chair for Mathematical Information Science

Solutions to the
Examination on Mathematics of Information
August 4, 2025

Problem 1

(a) We first show that $\{\delta_x : x \in \{0,1\}^n\}$ is a linearly independent set. Let $\{\lambda_x \in \mathbb{C} : x \in \{0,1\}^n\}$ be such that

$$\sum_{x\in\{0,1\}^n} \lambda_x \delta_x(y) = 0, \quad \forall y \in \{0,1\}^n.$$

Then, for all $y \in \{0,1\}^n$, we have

$$0 = \sum_{x\in\{0,1\}^n} \lambda_x \delta_x(y) = \lambda_y. \tag{1}$$

Therefore, $\lambda_y = 0$ for all $y \in \{0,1\}^n$, which implies that $\{\delta_x : x \in \{0,1\}^n\}$ is a linearly independent set. Now, we show that $\{\delta_x : x \in \{0,1\}^n\}$ spans F_n by noting that every $f \in F_n$ satisfies

$$f(y) = \sum_{x\in\{0,1\}^n} f(x)\delta_x(y), \quad \forall y \in \{0,1\}^n. \tag{2}$$

Since $\{\delta_x : x \in \{0,1\}^n\}$ is a linearly independent set and spans F_n, it is a basis for F_n.

(b) Let $S, T \in \mathcal{P}(n)$ and let $x \in \{0,1\}^n$. Then,

$$\begin{aligned}
\chi_S(x)\chi_T(x) &= \prod_{i\in S}(-1)^{x_i}\prod_{j\in T}(-1)^{x_j} \\
&= \prod_{i\in(S\setminus T)\cup(S\cap T)}(-1)^{x_i}\prod_{j\in(T\setminus S)\cup(S\cap T)}(-1)^{x_j} \\
&= \prod_{i\in S\setminus T}(-1)^{x_i}\prod_{i\in S\cap T}(-1)^{x_i}\prod_{i\in T\setminus S}(-1)^{x_i}\prod_{i\in S\cap T}(-1)^{x_i} \\
&= \prod_{i\in(S\setminus T)\cup(T\setminus S)}(-1)^{x_i}\left(\prod_{i\in S\cap T}(-1)^{x_i}\right)^2 \\
&\overset{(a)}{=} \prod_{i\in(S\setminus T)\cup(T\setminus S)}(-1)^{x_i} \\
&= \chi_{(S\setminus T)\cup(T\setminus S)}(x), \qquad (3)
\end{aligned}$$

where (a) is by $\prod_{i\in S\cap T}(-1)^{x_i} \in \{-1,1\}$.

(c) For fixed $S \neq \varnothing$, fix an $i \in \{1,\ldots,n\}$ such that $i \in S$. Note that $\chi_S(x) = (-1)^{x_i}\prod_{j\in S\setminus\{i\}}(-1)^{x_j}$, for all $x \in \{0,1\}^n$. Then,

$$\sum_{x\in\{0,1\}^n}\chi_S(x)\chi_\varnothing(x) \overset{(a)}{=} \sum_{x\in\{0,1\}^n}\chi_S(x) = \sum_{x\in\{0,1\}^n}\left((-1)^{x_i}\prod_{j\in S\setminus\{i\}}(-1)^{x_j}\right) \qquad (4)$$

$$\overset{(b)}{=} \sum_{x_i\in\{0,1\}}\sum_{(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)\in\{0,1\}^{n-1}}\left((-1)^{x_i}\prod_{j\in S\setminus\{i\}}(-1)^{x_j}\right) \qquad (5)$$

$$= \sum_{x_i\in\{0,1\}}(-1)^{x_i}\cdot I = (1-1)\cdot I = 0, \qquad (6)$$

with

$$I = \sum_{(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)\in\{0,1\}^{n-1}}\prod_{j\in S\setminus\{i\}}(-1)^{x_j},$$

where (a) follows from $\chi_\varnothing(x) = 1$, for all $x \in \{0,1\}^n$, and (b) is obtained by reordering the sum $\sum_{x\in\{0,1\}^n}$. In summary, we can conclude that

$$\langle\chi_S,\chi_\varnothing\rangle = \frac{1}{2^n}\sum_{x\in\{0,1\}^n}\chi_S(x)\chi_\varnothing(x) = \frac{1}{2^n}\cdot 0 = 0. \qquad (7)$$

(d) Consider $S, T \in \mathcal{P}(n)$, with $S \neq T$. From subproblem (b), we have

$$\begin{aligned}\langle \chi_S, \chi_T \rangle &= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \chi_S(x)\overline{\chi_T(x)} = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \chi_S(x)\chi_T(x) \\ &= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \chi_{(S\setminus T)\cup(T\setminus S)}(x) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \chi_{(S\setminus T)\cup(T\setminus S)}(x) \cdot 1 \\ &= \langle \chi_{(S\setminus T)\cup(T\setminus S)}, \chi_\varnothing \rangle. \end{aligned} \tag{8}$$

We now observe that $(S\setminus T)\cup(T\setminus S) \neq \varnothing$ as otherwise $S\setminus T = T\setminus S = \varnothing$, i.e., $S = T$, which stands in contradiction to the assumption $S \neq T$. From subproblem (c), we hence get

$$\langle \chi_S, \chi_T \rangle = \langle \chi_{(S\setminus T)\cup(T\setminus S)}, \chi_\varnothing \rangle = 0. \tag{9}$$

Moreover,

$$\langle \chi_S, \chi_S \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} \chi_S(x)\overline{\chi_S(x)} = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} (\chi_S(x))^2 = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} 1 = 1, \tag{10}$$

which establishes that $\mathcal{G}$ is an orthonormal set of functions. As $\#\mathcal{G} = 2^n = \#\{\delta_x : x \in \{0,1\}^n\}$ and all orthonormal bases of a finite-dimensional vector space have the same cardinality, it follows that $\mathcal{G}$ is an orthonormal basis for F_n.

(e) Let $S \in \mathcal{P}(n)$ and let $x \in \{0,1\}^n$. Then, we have

$$\widetilde{\chi}_S(x) = \prod_{i \in S} (\omega_1)^{x_i} = \prod_{i \in S} (\exp(i\pi))^{x_i} = \prod_{i \in S} (-1)^{x_i} = \chi_S(x), \tag{11}$$

which shows that $\mathcal{H}_1 = \mathcal{G}$. Therefore, $\mathcal{H}_1$ is an orthonormal basis by subproblem (d) and hence $\mathcal{H}_1$ is a tight frame with frame bounds $A = B = 1$.

(f) Let $k \geq 2$. The set of eigenvalues of $\mathbb{T}$ is given by

$$\{\lambda_S := h(|S|) : S \in \mathcal{P}(n)\}, \tag{12}$$

where

$$h(\ell) = 2^{n(k-1)}(1 + C_k)^\ell (1 - C_k)^{n-\ell}, \quad \ell \in \{0, \dots, n\}, \tag{13}$$

with $C_k = \left(\cos\left(\frac{\pi}{k+1}\right)\right)^{k+1}$. Following the hint in the problem statement, we will show that $\lambda_\varnothing$ and $\lambda_{\{1,\dots,n\}}$ are positive and are respectively the minimum and maximum eigenvalues. Noting that thanks to $k \geq 2$, $C_k \in (0,1)$, it follows that $1 - C_k, 1 + C_k > 0$, so that $\lambda_S > 0$, for all $S \in \mathcal{P}(n)$. In particular, the positivity of $\lambda_\varnothing$ and $\lambda_{\{1,\dots,n\}}$ is established. Now, $C_k \in (0,1)$ also implies that $1 + C_k > 1 - C_k$, and hence $h(\ell)$ is a strictly increasing function as

$$\frac{h(\ell)}{h(\ell-1)} = \frac{1+C_k}{1-C_k} > 1, \quad \ell \in \{1,\ldots,n\}. \tag{14}$$

Therefore, the minimum and maximum of $h(\ell)$ are attained at $\ell = 0$ and $\ell = n$, respectively. In summary, $\lambda_\varnothing = h(0)$ and $\lambda_{\{1,\ldots,n\}} = h(n)$ are the minimum and maximum eigenvalues of $\mathbb{T}$, respectively, and they are both positive. We hence get

$$\inf_{f\in F_n\setminus\{0\}} \frac{\langle \mathbb{T}f, f\rangle}{\|f\|^2} = \lambda_\varnothing > 0, \tag{15}$$

$$\sup_{f\in F_n\setminus\{0\}} \frac{\langle \mathbb{T}f, f\rangle}{\|f\|^2} = \lambda_{\{1,\ldots,n\}} > 0. \tag{16}$$

In summary, we can conclude that

$$\lambda_\varnothing \|f\|^2 \le \langle \mathbb{T}f, f\rangle = \sum_{\mathbf{S}\in\mathcal{P}(n)^k} |\langle f, \widetilde{\chi}_{\mathbf{S}}\rangle|^2 \le \lambda_{\{1,\ldots,n\}}\|f\|^2, \quad \forall f \in F_n, \tag{17}$$

so $\mathcal{H}_k$ is a frame for F_n with frame bounds $\lambda_\varnothing$ and $\lambda_{\{1,\ldots,n\}}$. As $h(\ell)$ is strictly increasing, the frame bounds are not equal, so that $\mathcal{H}_k$ is not a tight frame.

Problem 2

(a) Arbitrarily fix $S \subset \{1, \ldots, N\}$ with $|S| \leq m$. Let $\ell^* = \arg\max_{\ell \in S} \sum_{j \in S \setminus \{\ell\}} |\langle d_\ell, d_j \rangle|$ and define $\widehat{S} = S \setminus \{\ell^*\}$. Then, we have

$$\begin{aligned}
\max_{\ell \in S} \sum_{j \in S \setminus \{\ell\}} |\langle d_\ell, d_j \rangle| &= \sum_{j \in \widehat{S}} |\langle d_{\ell^*}, d_j \rangle| \\
&\leq \max_{\ell' \in \widehat{S}^c} \sum_{j \in \widehat{S}} |\langle d_{\ell'}, d_j \rangle| && \text{(since } \ell^* \in \widehat{S}^c\text{)} \\
&\leq \max_{|S'| \leq m-1} \max_{\ell' \in S'^c} \sum_{j \in S'} |\langle d_{\ell'}, d_j \rangle| && \text{(since } |\widehat{S}| = |S| - 1 \leq m-1\text{)} \\
&= \mu_{m-1}(D),
\end{aligned}$$

where the maximum in the last inequality is over all subsets $S' \subseteq \{1, \ldots, N\}$ with cardinality less than or equal to $m-1$. As S was arbitrary, this completes the proof.

(b) Towards a contradiction assume that $\operatorname{spark}(D) \leq m$. Then, there exists a set $J \subseteq \{1, \ldots, N\}$ with $|J| \leq m$ such that the corresponding columns of D are linearly dependent. Hence, there are $\alpha_j \in \mathbb{C}, j \in J$, not all equal to zero such that

$$\sum_{j \in J} \alpha_j d_j = 0.$$

Let $j^* \in J$ be the index of the coefficient with maximal absolute value. We can rewrite the above equation as

$$d_{j^*} = - \sum_{j \in J \setminus \{j^*\}} \frac{\alpha_j}{\alpha_{j^*}} d_j,$$

which, upon taking the inner product with d_{j^*} on both sides and using $\|d_{j^*}\|_2 = 1$, yields

$$1 = - \sum_{j \in J \setminus \{j^*\}} \frac{\alpha_j}{\alpha_{j^*}} \langle d_{j^*}, d_j \rangle.$$

We thus obtain the contradiction as follows

$$\begin{aligned}
1 &= \left| - \sum_{j \in J \setminus \{j^*\}} \frac{\alpha_j}{\alpha_{j^*}} \langle d_{j^*}, d_j \rangle \right| \\
&\leq \sum_{j \in J \setminus \{j^*\}} \left| \frac{\alpha_j}{\alpha_{j^*}} \right| |\langle d_{j^*}, d_j \rangle| && \text{(triangle inequality)} \\
&\leq \sum_{j \in J \setminus \{j^*\}} |\langle d_{j^*}, d_j \rangle| && \text{(since } |\alpha_j| \leq |\alpha_{j^*}|\text{)} \\
&\leq \max_{j' \in J} \sum_{j \in J \setminus \{j'\}} |\langle d_{j'}, d_j \rangle| && \text{(since } j^* \in J\text{)} \\
&\leq \mu_{m-1}(D) && \text{(previous subproblem)} \\
&\leq \mu_{m-1}(D) + \mu_m(D) && \text{(since } \mu_m(D) \geq 0\text{)} \\
&< 1. && \text{(by (4) in the problem statement)}
\end{aligned}$$

As S was arbitrary, this completes the proof.

(c) We rewrite

$$\begin{aligned}
& & \mu_{m-1}(D) + \mu_m(D) < 1 \\
\Leftrightarrow & & \mu_m(D) < 1 - \mu_{m-1}(D) \\
\Leftrightarrow & & \frac{\mu_m(D)}{1 - \mu_{m-1}(D)} < 1,
\end{aligned}$$

where we used $0 \leq \mu_m(D) < 1 - \mu_{m-1}(D)$. Next, arbitrarily fix $S \subset \{1, \ldots, N\}$ with $|S| \leq m$. We have

$$\max_{\ell \in S^c} \sum_{j \in S} |\langle d_\ell, d_j \rangle| \overset{(*)}{\leq} \max_{|S'| \leq m} \max_{\ell \in S'^c} \sum_{j \in S'} |\langle d_\ell, d_j \rangle| = \mu_m(D),$$

where in $(*)$ we take the maximum over all subsets $S' \subseteq \{1, \ldots, N\}$ with cardinality less than or equal to m. Furthermore, by subproblem (a),

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

Putting everything together, we obtain

$$\frac{\max_{\ell \in S^c} \sum_{j \in S} |\langle d_\ell, d_j \rangle|}{1 - \max_{\ell \in S} \sum_{j \in S \setminus \{\ell\}} |\langle d_\ell, d_j \rangle|} \leq \frac{\mu_m(D)}{1 - \mu_{m-1}(D)} < 1.$$

As S was arbitrary, this completes the proof.

(d) Arbitrarily fix $S \subset \{1, \ldots, N\}$ with $|S| \leq m$, and define the matrix $G \in \mathbb{C}^{|S| \times |S|}$

according to

$$G_{j,\ell} = \begin{cases} \langle d_\ell, d_j \rangle, & \text{if } j \neq \ell, \\ 0, & \text{if } j = \ell \end{cases}.$$

Next, note that, by Lemma H4, $\|G\|_{1,1} = \max_{\ell \in S} \sum_{j \in S \setminus \{\ell\}} |\langle d_\ell, d_j \rangle|$ and thus, by subproblem (a) and (4) in the problem statement, we have

$$\|G\|_{1,1} \leq \mu_{m-1}(D) < 1.$$

Furthermore, we write

$$((D_S)^H D_S) = \mathbb{I} + G.$$

and apply Lemma H5 to get

$$\begin{aligned} \|((D_S)^H D_S)^{-1}\|_{1,1} &= \|(\mathbb{I} + G)^{-1}\|_{1,1} \\ &\leq \frac{1}{1 - \|G\|_{1,1}} \\ &= \frac{1}{1 - \max_{\ell \in S} \sum_{j \in S \setminus \{\ell\}} |\langle d_\ell, d_j \rangle|}. \end{aligned} \tag{18}$$

Now we compute

$$\begin{aligned} \max_{\ell \in S^c} \|(D_S)^\dagger d_\ell\|_1 &= \max_{\ell \in S^c} \|((D_S)^H D_S)^{-1} (D_S)^H d_\ell\|_1 \\ &\leq \max_{\ell \in S^c} \|((D_S)^H D_S)^{-1}\|_{1,1} \, \|(D_S)^H d_\ell\|_1 && \text{by Lemma H3} \\ &= \|((D_S)^H D_S)^{-1}\|_{1,1} \, \max_{\ell \in S^c} \|(D_S)^H d_\ell\|_1 \\ &= \|((D_S)^H D_S)^{-1}\|_{1,1} \, \max_{\ell \in S^c} \sum_{j \in S} |\langle d_\ell, d_j \rangle| \\ &\leq \frac{\max_{\ell \in S^c} \sum_{j \in S} |\langle d_\ell, d_j \rangle|}{1 - \max_{\ell \in S} \sum_{j \in S \setminus \{\ell\}} |\langle d_\ell, d_j \rangle|} && \text{by (18)} \\ &< 1. && \text{by subproblem (c)} \end{aligned}$$

As S was arbitrary, this completes the proof.

(e) Let S with $|S| \leq m$ be the support set of x. We have, for all $\ell \in S^c$, that

$$\begin{aligned} |\langle (D_S)^\dagger d_\ell, \operatorname{sgn}(x_S) \rangle| &\leq \|(D_S)^\dagger d_\ell\|_1 \|\operatorname{sgn}(x_S)\|_\infty && \text{by Lemma H8} \\ &= \|(D_S)^\dagger d_\ell\|_1 \\ &< 1. && \text{by subproblem (d)} \end{aligned}$$

Furthermore, $\mathcal{N}(D_S) = \{0\}$, as by subproblem (b) every set of m columns of D must be linearly independent. We can thus apply Theorem H7 to conclude that

(P1) uniquely recovers x.

Problem 3

(a) We have

$$M(2\epsilon; \mathbb{M}_K^{n\times n}, \|\cdot\|_2) \leq N(\epsilon; \mathbb{M}_K^{n\times n}, \|\cdot\|_2) \leq M(\epsilon; \mathbb{M}_K^{n\times n}, \|\cdot\|_2).$$

(b) Surjectivity of V follows from the definition of $V(\mathbb{M}_K^{n\times n})$. To verify that V is also an injection, arbitrarily pick two elements $A, B \in \mathbb{M}_K^{n\times n}$. When $V(A) = V(B)$, we have

$$A_{i,j} = (V(A))_{(i-1)n+j} = (V(B))_{(i-1)n+j} = B_{i,j}, \quad \text{for } i, j \in \{1, 2, \dots, n\},$$

which implies $A = B$. Finally, we note that, for every $A \in \mathbb{M}_K^{n\times n}$, it holds that

$$\|V(A)\|_2 = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}\left((V(A))_{(i-1)n+j}\right)^2} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(A_{i,j})^2} = \|A\|_F\,,$$

which establishes that V is norm-preserving and hence an isometric isomorphism.

(c) First, we prove the left-hand side of the inequality by noting that

$$\begin{aligned}
\|A\|_2 &= \sup_{x\in\mathbb{R}^n, \|x\|_2=1} \|Ax\|_2 \\
&= \sup_{x\in\mathbb{R}^n, \|x\|_2=1} \sqrt{\sum_{i=1}^{n}\left(\sum_{j=1}^{n} A_{i,j}x_j\right)^2} \\
&\overset{\text{Cauchy inequality}}{\leq} \sup_{x\in\mathbb{R}^n, \|x\|_2=1} \sqrt{\sum_{i=1}^{n}\left(\sum_{j=1}^{n}(A_{i,j})^2\right)\left(\sum_{j=1}^{n}(x_j)^2\right)} \\
&= \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(A_{i,j})^2} \\
&= \|A\|_F\,.
\end{aligned}$$

To establish the inequality $\|A\|_F \leq \sqrt{n}\,\|A\|_2$, we pick x^* with $\|x^*\|_2 = 1$, as the eigenvector of A^TA corresponding to the largest eigenvalue $\lambda_1(A^TA)$. Then, we have

$$\begin{aligned}
\|A\|_2 &= \sup_{x\in\mathbb{R}^n, \|x\|_2=1} \|Ax\|_2 = \sup_{x\in\mathbb{R}^n, \|x\|_2=1} \sqrt{x^TA^TAx} \\
&\geq \sqrt{(x^*)^TA^TAx^*} = \sqrt{\lambda_1(A^TA)}\,\|x^*\|_2 \\
&\geq \sqrt{\frac{\sum_{i=1}^{n}\lambda_i(A^TA)}{n}} = \sqrt{\frac{\mathrm{tr}(A^TA)}{n}} = \frac{1}{\sqrt{n}}\,\|A\|_F\,.
\end{aligned}$$

(d) On the one hand, given $x \in V(\mathbb{M}_K^{n\times n})$, there exists $A \in \mathbb{M}_K^{n\times n}$, such that $x = V(A)$. We have

$$\|x\|_2 = \|V(A)\|_2 \overset{\text{(b)}}{=} \|A\|_F \overset{\text{(c)}}{\leq} \sqrt{n}\,\|A\|_2 \leq \sqrt{n}K,$$

which implies $V(\mathbb{M}_K^{n\times n}) \subseteq B_{\sqrt{n}K}^{n^2}$. On the other hand, given $x \in B_K^{n^2}$, define the matrix $A \in \mathbb{R}^{n\times n}$ according to

$$A_{i,j} = x_{(i-1)n+j}, \qquad \text{for } i,j \in \{1,2,\dots,n\}.$$

Then, $x = V(A)$ and

$$\|A\|_2 \overset{\text{(c)}}{\leq} \|A\|_F \overset{\text{(b)}}{=} \|V(A)\|_2 = \|x\|_2 \leq K,$$

which shows that $A \in \mathbb{M}_K^{n\times n}$ and hence $B_K^{n^2} \subseteq V(\mathbb{M}_K^{n\times n})$.

(e) Assume that $\{x_1, x_2, \dots, x_M\}$ is an ϵ-packing of $\mathcal{C}_1$ with $M = M(\epsilon; \mathcal{C}_1, \rho_X)$, i.e.,

$$\{x_1, x_2, \dots, x_M\} \subseteq \mathcal{C}_1$$

and, for every $i \neq j$, we have $\rho_X(x_i, x_j) > \epsilon$. Then, $\{x_1, x_2, \dots, x_M\} \subseteq \mathcal{C}_2$ and $\{x_1, x_2, \dots, x_M\}$ is trivially also an ϵ-packing of $\mathcal{C}_2$, which yields

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

(f) We note that

$$\begin{aligned}
n^2 \log\left(\epsilon^{-1}\right) &\overset{\text{(8) in the problem statement}}{\asymp} \log M(2\epsilon; B_K^{n^2}, \|\cdot\|_2) \\
&\overset{\text{(d),(e)}}{\leq} \log M(2\epsilon; V(\mathbb{M}_K^{n\times n}), \|\cdot\|_2) \\
&\overset{\text{(a)}}{\leq} \log N(\epsilon; V(\mathbb{M}_K^{n\times n}), \|\cdot\|_2) \\
&\overset{\text{(b), Lemma H12}}{=} \log N(\epsilon; \mathbb{M}_K^{n\times n}, \|\cdot\|_F) \\
&= \log N(\epsilon; V(\mathbb{M}_K^{n\times n}), \|\cdot\|_2) \\
&\overset{\text{(a)}}{\leq} \log M(\epsilon; V(\mathbb{M}_K^{n\times n}), \|\cdot\|_2) \\
&\overset{\text{(d),(e)}}{\leq} \log M(\epsilon; B_{\sqrt{n}K}^{n^2}, \|\cdot\|_2) \\
&\overset{\text{(8)}}{\asymp} n^2 \log\left(\epsilon^{-1}\right),
\end{aligned}$$

i.e., $\log N(\epsilon; \mathbb{M}_K^{n\times n}, \|\cdot\|_F) \asymp n^2 \log(\epsilon^{-1})$. Finally, from subproblem (c), we know that $\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\,\|A\|_2$, for all $A \in \mathbb{R}^{n\times n}$, which, by Lemma H13, yields

$$\log N(\epsilon; \mathbb{M}_K^{n\times n}, \|\cdot\|_2) \asymp \log N(\epsilon; \mathbb{M}_K^{n\times n}, \|\cdot\|_F) \asymp n^2 \log\left(\epsilon^{-1}\right).$$

Problem 4

(a) By Definition H14, we have

$$
\begin{aligned}
\mathcal{R}\left(\mathcal{F}_b\left(\{x_i\}_{i=1}^n\right)/n\right) &= \frac{1}{n}\mathbb{E}_\varepsilon\left[\sup_{f\in\mathcal{F}_b}\left|\sum_{i=1}^n \varepsilon_i f(x_i)\right|\right] \\
&\overset{\text{(a)}}{=} \frac{1}{n}\mathbb{E}_\varepsilon\left[\sup_{f\in\mathcal{F}_b}\left|\sum_{i=1}^n \varepsilon_i \langle f, \phi_{x_i}\rangle\right|\right] \\
&= \frac{1}{n}\mathbb{E}_\varepsilon\left[\sup_{f\in\mathcal{F}_b}\left|\left\langle f, \sum_{i=1}^n \varepsilon_i \phi_{x_i}\right\rangle\right|\right] \\
&\overset{\text{(b)}}{\leq} \frac{1}{n}\mathbb{E}_\varepsilon\left[\sup_{f\in\mathcal{F}_b}\|f\|\left\|\sum_{i=1}^n \varepsilon_i \phi_{x_i}\right\|\right] \\
&= \frac{b}{n}\mathbb{E}_\varepsilon\left[\left\|\sum_{i=1}^n \varepsilon_i \phi_{x_i}\right\|\right] \\
&= \frac{b}{n}\mathbb{E}_\varepsilon\left[\sqrt{\left\langle \sum_{i=1}^n \varepsilon_i \phi_{x_i}, \sum_{j=1}^n \varepsilon_j \phi_{x_j}\right\rangle}\right] \\
&\overset{\text{(c)}}{\leq} \frac{b}{n}\sqrt{\mathbb{E}_\varepsilon\left[\left\langle \sum_{i=1}^n \varepsilon_i \phi_{x_i}, \sum_{j=1}^n \varepsilon_j \phi_{x_j}\right\rangle\right]} \\
&= \frac{b}{n}\sqrt{\sum_{i=1}^n\sum_{j=1}^n \mathbb{E}_\varepsilon[\varepsilon_i\varepsilon_j]\left\langle \phi_{x_i}, \phi_{x_j}\right\rangle} \\
&\overset{\text{(d)}}{=} \frac{b}{n}\sqrt{\sum_{i=1}^n \langle \phi_{x_i}, \phi_{x_i}\rangle} \\
&\overset{\text{(e)}}{=} \frac{b}{n}\sqrt{\sum_{i=1}^n \phi_{x_i}(x_i)} \\
&= \frac{b}{n}\sqrt{\sum_{i=1}^n k(x_i, x_i)} \\
&= \frac{b}{n}\sqrt{\operatorname{tr}(K)},
\end{aligned}
$$

where (a) is by the reproducing property, (b) follows from the Cauchy–Schwarz inequality, and (c) is by Jensen's inequality (Lemma H15). In (d), we used the fact that $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. Rademacher random variables so that $\mathbb{E}[\varepsilon_i\varepsilon_j] = 0$, for $i \neq j$, and $\mathbb{E}[\varepsilon_i^2] = 1$, for $i, j \in \{1, \ldots, n\}$. Finally, (e) is again due to the reproducing property.

(b) Let $\mathcal{H}' := \{u \colon \mathcal{X} \times \{-1,1\} \to \mathbb{R} \colon u(x,y) = -yf(x), \forall (x,y) \in \mathcal{X} \times \{-1,1\}, f \in \mathcal{F}_b\}$. As the random variables $\{\varepsilon_i y_i\}_{i=1}^n$ are i.i.d. Rademacher whenever $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. Rademacher, we have

$$\begin{aligned}
\mathcal{R}\left(\mathcal{H}'\left(\{(x_i,y_i)\}_{i=1}^n\right)/n\right) &= \frac{1}{n}\mathbb{E}_\varepsilon\left[\sup_{u\in\mathcal{H}'}\left|\sum_{i=1}^n \varepsilon_i u(x_i,y_i)\right|\right] \\
&= \frac{1}{n}\mathbb{E}_\varepsilon\left[\sup_{u\in\mathcal{H}'}\left|\sum_{i=1}^n -\varepsilon_i y_i f(x_i)\right|\right] \\
&= \frac{1}{n}\mathbb{E}_\varepsilon\left[\sup_{f\in\mathcal{F}_b}\left|\sum_{i=1}^n \varepsilon_i f(x_i)\right|\right] \\
&= \mathcal{R}\left(\mathcal{F}_b\left(\{x_i\}_{i=1}^n\right)/n\right). \qquad (19)
\end{aligned}$$

Upon noting that ρ_γ is $(1/\gamma)$-Lipschitz, application of Lemma H16 yields

$$\mathcal{R}\left(\mathcal{H}\left(\{(x_i,y_i)\}_{i=1}^n\right)/n\right) \overset{\text{Lemma H16}}{\leq} \frac{1}{\gamma}\mathcal{R}\left(\mathcal{H}'\left(\{x_i\}_{i=1}^n\right)/n\right) \overset{(19)}{=} \frac{1}{\gamma}\mathcal{R}\left(\mathcal{F}_b\left(\{x_i\}_{i=1}^n\right)/n\right).$$

(c) We follow the hint and note that (4) in Lemma H18 holds with $L = 2/n$. Let $\delta' > 0$ and consider the event

$$\mathcal{E}_1 := \left\{\mathcal{R}_n(\mathcal{G}) - \mathcal{R}\left(\mathcal{G}\left(\{Z_i\}_{i=1}^n\right)/n\right) > \sqrt{\frac{2\log(1/\delta')}{n}}\right\}.$$

Setting $\epsilon := \sqrt{\frac{2\log(1/\delta')}{n}}$, application of Lemma H18 yields

$$\mathbb{P}\left[\mathcal{E}_1\right] \leq e^{-\frac{2n}{4}\frac{2\log(1/\delta')}{n}} = \delta', \qquad (20)$$

where we used that $\mathcal{R}_n(\mathcal{G}) = \mathbb{E}[\mathcal{R}\left(\mathcal{G}\left(\{Z_i\}_{i=1}^n\right)/n\right)]$. Furthermore, for the event

$$\mathcal{E}_2 := \left\{\psi(Z_1^n) > 2\mathcal{R}_n(\mathcal{G}) + \sqrt{\frac{2\log(1/\delta')}{n}}\right\}$$

with $\psi(Z_1^n) := \sup_{g\in\mathcal{G}}\left(\mathbb{E}[g(Z)] - \frac{1}{n}\sum_{i=1}^n g(Z_i)\right)$, we have, by Theorem H17,

$$\mathbb{P}\left[\mathcal{E}_2\right] \leq \delta'. \qquad (21)$$

We then obtain

$$\begin{aligned}
\mathbb{P}\left[\psi(Z_1^n) \leq 2\mathcal{R}\left(\mathcal{G}\left(\{Z_i\}_{i=1}^n\right)/n\right) + 3\sqrt{\frac{2\log(1/\delta')}{n}}\right] &\geq \mathbb{P}[\mathcal{E}_1^c \cap \mathcal{E}_2^c] \\
&= \mathbb{P}[(\mathcal{E}_1 \cup \mathcal{E}_2)^c]
\end{aligned}$$

$$
\begin{aligned}
&= 1 - \mathbb{P}[\mathcal{E}_1 \cup \mathcal{E}_2] \\
&\overset{\text{(a)}}{\geq} 1 - (\mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_2]) \\
&\overset{\text{(b)}}{\geq} 1 - 2\delta',
\end{aligned}
$$

where (a) is by a union bound and (b) follows from (20) and (21). Here, the superscript c denotes the complement of an event. Setting $\delta' = \delta/2$ yields the desired result.

(d) Note that $\rho_\gamma(u) \geq \mathbb{1}_{\{u \geq 0\}}$, $u \in \mathbb{R}$, where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function. We thus have

$$
\mathbb{P}(\text{sign}(f(X)) \neq Y) = \mathbb{E}[\mathbb{1}_{\{\text{sign}(f(X)) \neq Y\}}] \leq \mathbb{E}[\mathbb{1}_{\{-Yf(X) \geq 0\}}] \leq \mathbb{E}[\rho_\gamma(-Yf(X))]. \quad (22)
$$

Further, with probability $\geq 1 - \delta$, it holds that

$$
\begin{aligned}
\mathbb{E}[\rho_\gamma(-Yf(X))] &= \frac{1}{n}\sum_{i=1}^n \rho_\gamma(-Y_i f(X_i)) + \mathbb{E}[\rho_\gamma(-Yf(X))] - \frac{1}{n}\sum_{i=1}^n \rho_\gamma(-Y_i f(X_i)) \\
&\leq \frac{1}{n}\sum_{i=1}^n \rho_\gamma(-Y_i f(X_i)) + \sup_{h \in \mathcal{H}} \left(\mathbb{E}[h(X,Y)] - \frac{1}{n}\sum_{i=1}^n h(X_i, Y_i) \right) \\
&\overset{\text{(a)}}{\leq} \frac{1}{n}\sum_{i=1}^n \rho_\gamma(-Y_i f(X_i)) + 2\mathcal{R}\left(\mathcal{H}\left(\{(X_i,Y_i)\}_{i=1}^n\right)/n\right) + 3\sqrt{\frac{2\log(2/\delta)}{n}} \\
&\overset{\text{(b)}}{\leq} \frac{1}{n}\sum_{i=1}^n \rho_\gamma(-Y_i f(X_i)) + \frac{2b}{n\gamma}\sqrt{\text{tr}(\mathsf{K})} + 3\sqrt{\frac{2\log(2/\delta)}{n}}, \qquad (23)
\end{aligned}
$$

where (a) follows from the result of subproblem (c) and (b) is by the results of subproblems (b) and (a). Substituting (23) into (22) yields the desired expression.