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 12, 2024

Problem 1

(a) i. Let $\{\widetilde{f}_k\}_{k\in\mathbb{N}}$ be the canonical dual frame of $\{f_k\}_{k\in\mathbb{N}}$. By Theorem 1.22 in the lecture notes we have that every $x \in \mathcal{H}$ can be represented by

$$x = \sum_{k=1}^{\infty} c[k] f_k,$$

where $c[k] := \langle x, \widetilde{f}_k\rangle$. As $\{\widetilde{f}_k\}_{k\in\mathbb{N}}$ is also a frame, the sequence $c[\cdot]$ is in $\ell^2(\mathbb{N})$. Thus, T is surjective.

Furthermore, by Theorem 1.37 in the lecture notes, the sequence $c[\cdot]$ is unique. Thus, T is also injective and hence bijective.

ii. Let A and B, with $0 < A \leq B < \infty$, be the frame bounds of $\{f_k\}_{k\in\mathbb{N}}$ and arbitrarily fix $c[\cdot] \in \ell^2(\mathbb{N})$. Now, we calculate

$$\begin{aligned}
\|Tc\|_{\mathcal{H}} &= \left\| \sum_{k=1}^{\infty} c[k] f_k \right\|_{\mathcal{H}} \\
&\overset{\text{Lemma H1}}{=} \sup_{\|g\|_{\mathcal{H}}=1} \left| \langle g, \sum_{k=1}^{\infty} c[k] f_k \rangle \right| \\
&= \sup_{\|g\|_{\mathcal{H}}=1} \left| \sum_{k=1}^{\infty} c[k] \langle g, f_k \rangle \right| \\
&\leq \sup_{\|g\|_{\mathcal{H}}=1} \sum_{k=1}^{\infty} |c[k]| |\langle g, f_k\rangle| \\
&\overset{\text{C.S.}}{\leq} \sup_{\|g\|_{\mathcal{H}}=1} \|c\|_{\ell^2} \left(\sum_{k=1}^{\infty} |\langle g, f_k\rangle|^2 \right)^{1/2} \\
&\overset{\text{frame bound}}{\leq} \sup_{\|g\|_{\mathcal{H}}=1} \|c\|_{\ell^2} \sqrt{B} \|g\|_{\mathcal{H}} = \sqrt{B} \|c\|_{\ell^2}
\end{aligned} \tag{1}$$

As $B < \infty$, we can hence conclude that T is a bounded operator.

iii. In the previous subtasks, we established that $T : \ell^2(\mathbb{N}) \to \mathcal{H}$ is a bijective bounded linear operator. Furthermore, we have

$$T\delta_j = \sum_{k=1}^{\infty} \delta_j[k] f_k = f_j, \quad \forall j \in \mathbb{N}.$$

Therefore, T satisfies the requirements of Definition 1 and $\{f_k\}_{k\in\mathbb{N}}$ is thus a Riesz basis for $\mathcal{H}$.

(b) As $\{f_k\}_{k\in\mathbb{N}}$ is a Riesz basis, there exists a bijective bounded linear operator $U : \ell^2(\mathbb{N}) \to \mathcal{H}$ such that $f_k = U\delta_k$ for $k \in \mathbb{N}$. We first show that $\{f_k\}_{k\in\mathbb{N}}$ is a frame for $\mathcal{H}$. To this end, fix $x \in \mathcal{H}$ arbitrarily and compute

$$\sum_{k=1}^{\infty} |\langle x, f_k\rangle|^2 = \sum_{k=1}^{\infty} |\langle x, U\delta_k\rangle|^2 = \sum_{k=1}^{\infty} |\langle U^* x, \delta_k\rangle|^2 = \|U^* x\|_{\ell^2}^2, \tag{2}$$

where in the last equality we used Parseval's identity. We now bound, using Lemma H2,

$$\|U^* x\|_{\ell^2}^2 \leq |||U^*|||^2 \|x\|_{\mathcal{H}}^2 \tag{3}$$

and

$$\|U^* x\|_{\ell^2}^2 = \frac{|||(U^*)^{-1}|||^2}{|||(U^*)^{-1}|||^2} \|U^* x\|_{\ell^2}^2 \geq \frac{1}{|||(U^*)^{-1}|||^2} \|(U^*)^{-1} U^* x\|_{\mathcal{H}}^2 = \frac{1}{|||(U^*)^{-1}|||^2} \|x\|_{\mathcal{H}}^2. \tag{4}$$

As x was arbitrary, it follows from (3) and (4) that $\{f_k\}_{k\in\mathbb{N}}$ is a frame with frame bounds $A = \frac{1}{|||(U^*)^{-1}|||^2}$ and $B = |||U^*|||^2$. The fact that $|||U^*|||$ and $|||(U^*)^{-1}|||$ are finite follows from Lemma H3 together with Lemma H2.

We next show that $\{f_k\}_{k\in\mathbb{N}}$ is exact. To this end, we fix $m \in \mathbb{N}$ arbitrarily and show that $\{f_k\}_{k\neq m}$ is incomplete. First, note that by Lemma H3, $(U^{-1})^*$ is a bijective bounded linear operator. Since $\delta_m \neq 0$ we thus have $x := (U^{-1})^*\delta_m \neq 0$.

Now we compute

$$\langle x, f_k\rangle = \langle (U^{-1})^*\delta_m, f_k\rangle = \langle \delta_m, U^{-1} f_k\rangle = \langle \delta_m, \delta_k\rangle = 0, \quad \text{for all } k \neq m, \tag{5}$$

which establishes that $\{f_k\}_{k\neq m}$ is incomplete and hence $\{f_k\}_{k\in\mathbb{N}}$ must be exact.

(c) We compute

$$\begin{aligned}
\frac{1}{|||U^{-1}|||^2}\sum_{k\in\mathcal{J}}|c_k|^2 &= \frac{1}{|||U^{-1}|||^2}\left\|\sum_{k\in\mathcal{J}}c_k\delta_k\right\|_{\ell^2}^2 \\
&= \frac{1}{|||U^{-1}|||^2}\left\|U^{-1}U\sum_{k\in\mathcal{J}}c_k\delta_k\right\|_{\ell^2}^2 \\
&\leq \frac{|||U^{-1}|||^2}{|||U^{-1}|||^2}\left\|\sum_{k\in\mathcal{J}}c_kU\delta_k\right\|_{\mathcal{H}}^2 \\
&= \left\|\sum_{k\in\mathcal{J}}c_kf_k\right\|_{\mathcal{H}}^2 \\
&= \left\|U\sum_{k\in\mathcal{J}}c_k\delta_k\right\|_{\mathcal{H}}^2 \\
&\leq |||U|||^2\left\|\sum_{k\in\mathcal{J}}c_k\delta_k\right\|_{\ell^2}^2 = |||U|||^2\sum_{k\in\mathcal{J}}|c_k|^2.
\end{aligned}$$

Thus, the equation from the problem statement is satisfied with $A = \frac{1}{|||U^{-1}|||^2}$ and $B = |||U|||^2$.

Problem 2

(a) It follows directly from the definition of f_η that f_η is right-continuous and satisfies $f_\eta(0) = 0,\ f_\eta(1) = 1$. To verify that f_η is monotonically non-decreasing, we check the following

- $(1 - 1 + \eta_k)\epsilon' \geq 0$;
- $(k + 1 - 1 + \eta_{k+1})\epsilon' = (k - 1 + \eta_k)\epsilon' + (1 + \eta_{k+1} - \eta_k)\epsilon' \geq (k - 1 + \eta_k)\epsilon'$;
- $(n - 1 + \eta_n)\epsilon' \leq n\epsilon' < (n + 1)\epsilon' = 1$.

(b) $d_h(f_\eta, f_{\eta'}) > \lceil n/32 \rceil$ implies that $|f_\eta(x) - f_{\eta'}(x)| = \epsilon'$ on more than $\lceil n/32 \rceil$ disjoint intervals of length ϵ', which yields

$$\begin{aligned} \|f_\eta - f_{\eta'}\|_2^2 &= \int_0^1 |f_\eta(x) - f_{\eta'}(x)|^2 dx \\ &= \sum_{k=0}^{n} \int_{k\epsilon'}^{(k+1)\epsilon'} |f_\eta(x) - f_{\eta'}(x)|^2 dx \\ &> \lceil n/32 \rceil \cdot (\epsilon')^2 \cdot \epsilon' \\ &\geq \frac{(\epsilon')^{-1} - 1}{32} \cdot (\epsilon')^2 \cdot \epsilon' \\ &> \frac{(\epsilon')^2}{64}, \end{aligned}$$

where the last inequality follows from $(\epsilon')^{-1} - 1 > (\epsilon')^{-1}/2$ as $\epsilon' \in (0, 1/2)$.

(c) We note that $f_{\eta_1} \in \{f_{\eta'} \in \mathcal{U} : d_h(f_\eta, f_{\eta'}) \leq \lceil n/32 \rceil\} =: P_\eta$ if and only if η_1 differs from η in at most $\lceil n/32 \rceil$ out of n entries, which means that P_η contains no more than

$$\begin{aligned} \sum_{k=0}^{\lceil n/32 \rceil} \binom{n}{k} &\leq \left(\frac{en}{\lceil n/32 \rceil}\right)^{\lceil n/32 \rceil} \\ &\leq (32e)^{n/32+1} \\ &\leq \left(2^8\right)^{n/32+1} \\ &= 2^{n/4+8} \\ &< 2^{n/2} \end{aligned}$$

elements, where the last inequality follows from $n > 32$.

(d) Pick a maximal ϵ'-packing $\left\{f_1, f_2, \ldots, f_{M(\epsilon'; \mathcal{F}, \|\cdot\|_2)}\right\}$. Then, we have,

$$\text{for all } i, j \text{ s.t. } 1 \leq i < j \leq M(\epsilon'; \mathcal{F}, \|\cdot\|_2), \text{ it holds that } \|f_i - f_j\|_2 > \epsilon' > \epsilon,$$

which implies that $\left\{f_1, f_2, \ldots, f_{M(\epsilon'; \mathcal{F}, \|\cdot\|_2)}\right\}$ is also an ϵ-packing. Based on the definition of the packing number, we can conclude that

$$M(\epsilon;\mathcal{F},\|\cdot\|_2) \geq M(\epsilon';\mathcal{F},\|\cdot\|_2).$$

(e) Arbitrarily pick a cdf f_1 from $\mathcal{U}_1 := \mathcal{U}$. Based on subproblem (c), the set

$$P_1 := \{f_{\eta'} \in \mathcal{U} : d_h(f_1, f_{\eta'}) \leq \lceil n/32 \rceil\}$$

contains no more than $2^{n/2}$ elements. We then arbitrarily pick a cumulative distribution function (cdf) f_2 from

$$\mathcal{U}_2 := \mathcal{U}_1 \setminus P_1 = \mathcal{U} \setminus P_1$$

and note that $\operatorname{card}(\mathcal{U}_2) \geq 2^n - 2^{n/2}$. It is guaranteed that $d_h(f_2, f_1) > \lceil n/32 \rceil$. Now, denote $P_2 := \{f_{\eta'} \in \mathcal{U} : d_h(f_2, f_{\eta'}) \leq \lceil n/32 \rceil\}$. Again, we arbitrarily pick a cdf f_3 from

$$\mathcal{U}_3 := \mathcal{U}_2 \setminus P_2 = \mathcal{U} \setminus (P_1 \cup P_2),$$

which ensures $d_h(f_3, f_1) > \lceil n/32 \rceil$ and $d_h(f_3, f_2) > \lceil n/32 \rceil$. Moreover, thanks to subproblem (c), we have $\operatorname{card}(\mathcal{U}_3) \geq \operatorname{card}(\mathcal{U}_2) - 2^{n/2} \geq 2^n - 2 \cdot 2^{n/2}$. Then, we can iteratively define

$$P_m = \{f_{\eta'} \in \mathcal{U} : d_h(f_m, f_{\eta'}) \leq \lceil n/32 \rceil\}$$

and pick f_{m+1} from

$$\mathcal{U}_{m+1} := \mathcal{U}_m \setminus P_m = \mathcal{U} \setminus \left(\bigcup_{k=1}^{m} P_m \right),$$

which ensures that

$$\text{for all } i,j \text{ s.t. } 1 \leq i < j \leq m+1, \text{ it holds that } d_h(f_i, f_j) > \lceil n/32 \rceil.$$

Note that $\operatorname{card}(P_m) \leq 2^{n/2}$ and $\mathcal{U}_m \neq \emptyset$, for all $m = 1, 2, \ldots, M$, with $M = \operatorname{card}(\mathcal{U})/2^{n/2} = 2^n/2^{n/2} = 2^{n/2}$ being a positive integer. We finally pick a set of cdfs $\{f_1, f_2, \ldots, f_M\}$, such that

$$\text{for all } i,j \text{ s.t. } 1 \leq i < j \leq M, \text{ we have } d_h(f_i, f_j) > \lceil n/32 \rceil.$$

Now, applying the result in subproblem (b), we obtain

$$\text{for all } i,j \text{ s.t. } 1 \leq i < j \leq M, \text{ it holds that } \|f_i - f_j\|_2 > \frac{\epsilon'}{8},$$

i.e., $\{f_1, f_2, \ldots, f_M\}$ constitutes an $(\frac{\epsilon'}{8})$-packing. Based on the definition of the packing number and the result in subproblem (d), we have

$$
\begin{aligned}
M(\epsilon/8;\mathcal{F},\|\cdot\|_2) &\geq M(\epsilon'/8;\mathcal{F},\|\cdot\|_2) &(6)\\
&\geq M = 2^{n/2} &(7)\\
&= 2^{\lceil \epsilon^{-1}/2\rceil -2} &(8)\\
&\geq 2^{\epsilon^{-1}/2-2} &(9)\\
&\geq 2^{\epsilon^{-1}/4}. &(10)
\end{aligned}
$$

where (10) follows from $\epsilon^{-1}/2 - 2 \geq \epsilon^{-1}/4$, which is thanks to $\epsilon \in (0, \epsilon_0)$ with $\epsilon_0 = 1/36$. Replacing $\epsilon/8$ by ϵ in (6) and (10), and taking the logarithm, concludes the proof with $c_1 = 1/32$.

(f) For every $\epsilon > 0$, we can find a positive integer L_ϵ such that $q^{L_\epsilon}\epsilon \geq 1$. Now, iteratively applying relation (8) from the problem statement with ϵ chosen as $\epsilon, q\epsilon, \ldots, q^{L_\epsilon - 1}\epsilon$ and multiplying the resulting inequalities, we obtain

$$
\begin{aligned}
N(\epsilon;\mathcal{F},\|\cdot\|_2) &\leq \left(\prod_{k=0}^{L_\epsilon-1} 2^{p/(q^k\epsilon)}\right)\cdot N(q^{L_\epsilon}\epsilon;\mathcal{F},\|\cdot\|_2)\\
&= 2^{(p/\epsilon)\sum_{k=0}^{L_\epsilon-1} q^{-k}}\cdot N(q^{L_\epsilon}\epsilon;\mathcal{F},\|\cdot\|_2)\\
&\leq 2^{(p/\epsilon)\sum_{k=0}^{\infty} q^{-k}}\cdot N(q^{L_\epsilon}\epsilon;\mathcal{F},\|\cdot\|_2)\\
&= 2^{\frac{p}{1-q^{-1}}\epsilon^{-1}}\cdot N(q^{L_\epsilon}\epsilon;\mathcal{F},\|\cdot\|_2)\\
&= 2^{\frac{p}{1-q^{-1}}\epsilon^{-1}},
\end{aligned}
$$

where the last equality follows from $N(q^{L_\epsilon}\epsilon;\mathcal{F},\|\cdot\|_2) = 1$ owing to the fact that for all $f, g \in \mathcal{F}$,

$$
\begin{aligned}
\|f-g\|_2 &= \left(\int_0^1 |f(x) - g(x)|^2 dx\right)^{1/2}\\
&\leq \left(\int_0^1 1 dx\right)^{1/2}\\
&= 1 \leq q^{L_\epsilon}\epsilon,
\end{aligned}
$$

and thus every function in $\mathcal{F}$ constitutes a $(q^{L_\epsilon}\epsilon)$-covering of $\mathcal{F}$. Finally, taking the logarithm, concludes the proof with $c_2 = \frac{p}{1-q^{-1}}$.

(g) The statement follows immediately from

$$
M(2\epsilon;\mathcal{F},\|\cdot\|_2) \leq N(\epsilon;\mathcal{F},\|\cdot\|_2) \leq M(\epsilon;\mathcal{F},\|\cdot\|_2),
$$

and the results in subproblems (e) and (f).

Problem 3

(a) First note that $\mathcal{F} \subseteq \operatorname{conv}(\mathcal{F})$, so that

$$\mathcal{R}\left(\mathcal{F}\left(x_1^n\right)/n\right) \leq \mathcal{R}\left(\left(\operatorname{conv}(\mathcal{F})\right)\left(x_1^n\right)/n\right). \tag{11}$$

Let $\Delta_N := \{(\alpha_1, \ldots, \alpha_N) \in [0,1]^d \colon \sum_{j=1}^N \alpha_j = 1\}$ and $\mathcal{F}^N := \underbrace{\mathcal{F} \times \cdots \times \mathcal{F}}_{N \text{ times}}$. We have

$$\begin{aligned}
\mathbb{E}_\varepsilon\left[\sup_{f \in \operatorname{conv}(\mathcal{F})}\left|\sum_{i=1}^n \varepsilon_i f(x_i)\right|\right] &= \mathbb{E}_\varepsilon\left[\sup_{\substack{N \in \mathbb{N} \\ (\alpha_1,\ldots,\alpha_N) \in \Delta_N \\ (f_1,\ldots,f_N) \in \mathcal{F}^N}}\left|\sum_{i=1}^n \varepsilon_i \sum_{j=1}^N \alpha_j f_j(x_i)\right|\right] \\
&= \mathbb{E}_\varepsilon\left[\sup_{\substack{N \in \mathbb{N} \\ (f_1,\ldots,f_N) \in \mathcal{F}^N}} \sup_{(\alpha_1,\ldots,\alpha_N) \in \Delta_N}\left|\sum_{j=1}^N \alpha_j \sum_{i=1}^n \varepsilon_i f_j(x_i)\right|\right] \\
&\overset{(*)}{=} \mathbb{E}_\varepsilon\left[\sup_{\substack{N \in \mathbb{N} \\ (f_1,\ldots,f_N) \in \mathcal{F}^N}}\left|\sum_{i=1}^n \varepsilon_i f_{j*}(x_i)\right|\right] \\
&\leq \mathbb{E}_\varepsilon\left[\sup_{\substack{N \in \mathbb{N} \\ (f_1,\ldots,f_N) \in \mathcal{F}^N}} \sup_{f \in \mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(x_i)\right|\right] \\
&= \mathbb{E}_\varepsilon\left[\sup_{f \in \mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(x_i)\right|\right],
\end{aligned} \tag{12}$$

where in $(*)$ we used the fact from the hint, namely that

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

and $j^* \in \{1, \ldots, N\}$ is such that $\left|\sum_{i=1}^n \varepsilon_i f_{j*}(x_i)\right| = \max_{j \in \{1,\ldots,N\}}\left|\sum_{i=1}^n \varepsilon_i f_j(x_i)\right|$. Combining (11) and (12) yields $\mathcal{R}\left(\mathcal{F}\left(x_1^n\right)/n\right) = \mathcal{R}\left(\left(\operatorname{conv}(\mathcal{F})\right)\left(x_1^n\right)/n\right)$, as desired.

(b) Let $\mathcal{W} := \bigcup_{k=1}^d \{e_k, -e_k\}$, where $\{e_k\}_{k=1}^d$ denotes the standard basis of $\mathbb{R}^d$, i.e.,

$$(e_k)_j = \begin{cases} 1, & \text{if } k = j, \\ 0, & \text{otherwise,} \end{cases} \qquad j, k \in \{1, \ldots, d\}.$$

Consider the function class $\mathcal{F}' := \{x \mapsto \langle x, w \rangle \colon w \in \mathcal{W}\}$. It follows from Lemma H8 in the Handout and the properties of the inner product (Definition H9 in the

Handout) that $\operatorname{conv}(\mathcal{F}') = B^{-1}\mathcal{F}$, where $B^{-1}\mathcal{F} := \{B^{-1}f \colon f \in \mathcal{F}\}$. Application of the result in subproblem (a) now yields

$$\mathcal{R}\left(\left(B^{-1}\mathcal{F}\right)\left(x_1^n\right)/n\right) = \mathcal{R}\left(\mathcal{F}'\left(x_1^n\right)/n\right). \tag{13}$$

Using Hölder's inequality (Lemma H7 in the Handout), we get, for $w \in \mathcal{W}$,

$$\frac{1}{n}\sum_{i=1}^{n}\langle x_i, w\rangle^2 \le \frac{1}{n}\sum_{i=1}^{n}\|x_i\|_\infty^2\|w\|_1^2 \le M^2,$$

where we used that $x_1^n \subseteq \mathcal{X} = [-M, M]^d$. Moreover, since $\mathcal{F}'$ is finite with $|\mathcal{F}'| = 2d$, we can apply Massart's lemma (Lemma H4 in the Handout), as suggested by the hint, to obtain

$$\mathcal{R}\left(\mathcal{F}'\left(x_1^n\right)/n\right) \le M\sqrt{\frac{2\log(2d)}{n}}. \tag{14}$$

By definition of the empirical Rademacher complexity, we have

$$\mathcal{R}\left(\left(B^{-1}\mathcal{F}\right)\left(x_1^n\right)/n\right) = B^{-1}\mathcal{R}\left(\mathcal{F}\left(x_1^n\right)/n\right). \tag{15}$$

Finally, we obtain the desired result according to

$$\mathcal{R}\left(\mathcal{F}\left(x_1^n\right)/n\right) \overset{(15)}{=} B\mathcal{R}\left(\left(B^{-1}\mathcal{F}\right)\left(x_1^n\right)/n\right) \overset{(13)}{=} B\mathcal{R}\left(\mathcal{F}'\left(x_1^n\right)/n\right) \overset{(14)}{\le} BM\sqrt{\frac{2\log(2d)}{n}}.$$

(c) For notational convenience, we introduce

$$\Theta := \left\{(u_1, \dots, u_J, v_1, \dots, v_J) \in (\mathbb{R}\setminus\{0\})^J \times (\mathbb{R}^d\setminus\{0\})^J \colon \sum_{j=1}^{J}|u_j|\|v_j\|_1 \le C\right\}.$$

Using the positive homogeneity of ρ, we compute

$$\begin{aligned}
\mathcal{R}\left(\mathcal{F}\left(x_1^n\right)/n\right) &= \frac{1}{n}\mathbb{E}_\varepsilon\left[\sup_{\theta\in\Theta}\left|\sum_{i=1}^{n}\varepsilon_i\sum_{j=1}^{J}u_j\rho(\langle x_i, v_j\rangle)\right|\right] \\
&= \frac{1}{n}\mathbb{E}_\varepsilon\left[\sup_{\theta\in\Theta}\left|\sum_{j=1}^{J}u_j\|v_j\|_1\sum_{i=1}^{n}\varepsilon_i\rho\left(\left\langle x_i, \frac{v_j}{\|v_j\|_1}\right\rangle\right)\right|\right] \\
&\le \frac{C}{n}\mathbb{E}_\varepsilon\left[\sup_{\theta\in\Theta}\max_{j\in\{1,\dots,J\}}\left|\sum_{i=1}^{n}\varepsilon_i\rho\left(\left\langle x_i, \frac{v_j}{\|v_j\|_1}\right\rangle\right)\right|\right] \\
&\le \frac{C}{n}\mathbb{E}_\varepsilon\left[\sup_{w\in\mathbb{R}^d, \|w\|_1\le 1}\left|\sum_{i=1}^{n}\varepsilon_i\rho\left(\langle x_i, w\rangle\right)\right|\right]
\end{aligned}$$

$$= C\mathcal{R}\left(\left(\rho \circ \widetilde{\mathcal{F}}\right)(x_1^n)/n\right),$$

where $\widetilde{\mathcal{F}} := \{x \mapsto \langle x, w\rangle : w \in \mathbb{R}^d, \|w\|_1 \le 1\}$. Note that ρ is 1-Lipschitz with $\rho(0) = 0$. As suggested by the hint, we can thus apply the Ledoux–Talagrand contraction lemma (Lemma H5 in the Handout) to conclude that

$$\begin{aligned}\mathcal{R}\left(\mathcal{F}(x_1^n)/n\right) &\le 2C\mathcal{R}\left(\widetilde{\mathcal{F}}(x_1^n)/n\right)\\ &\le 2CM\sqrt{\frac{2\log(2d)}{n}},\end{aligned}$$

as desired, where the last inequality follows from the result of subproblem (b), particularized to the function class $\widetilde{\mathcal{F}}$, with the constant B in subproblem (b) accordingly set to 1.

(d) Let $i \in \mathbb{N}$. Using the result from the hint, we can conclude that

$$\mathcal{R}\left(\mathcal{F}_i(x_1^n)/n\right) \le \sum_{j=1}^{J_i} \mathcal{R}\big((\sigma_j^i \circ \mathcal{F}_{i-1})(x_1^n)/n\big),$$

where $\sigma_j^i(\cdot) := w_j^i \sigma(\cdot)$, $j \in \{1, \ldots, J_i\}$. Note that σ_j^i is $(|w_j^i|L)$-Lipschitz with $\sigma_j^i(0) = 0$. We can thus apply the Ledoux–Talagrand contraction lemma (Lemma H5 in the Handout) to get

$$\mathcal{R}\left(\mathcal{F}_i(x_1^n)/n\right) \le \sum_{j=1}^{J_i} 2|w_j^i|L\mathcal{R}\left(\mathcal{F}_{i-1}(x_1^n)/n\right) \le 2B_iL\mathcal{R}\left(\mathcal{F}_{i-1}(x_1^n)/n\right), \tag{16}$$

where we used in the last inequality that $\|w^i\|_1 \le B_i$. It finally follows by repeated application of (16) that

$$\mathcal{R}\left(\mathcal{F}_K(x_1^n)/n\right) \le \mathcal{R}\left(\mathcal{F}_0(x_1^n)/n\right)\prod_{i=1}^{K}(2B_iL) \le BM\sqrt{\frac{2\log(2d)}{n}}\prod_{i=1}^{K}(2B_iL),$$

where the last inequality is by the result in subproblem (b).

Problem 4

Rewriting the definition of the (ns)-th restricted isometry constant, we get that δ_{ns} is the smallest $\delta \geq 0$ such that

$$\left| \|Av\|_2^2 - \|v\|_2^2 \right| \leq \delta \|v\|_2^2, \quad \text{for all } (ns)\text{-sparse vectors } v \in \mathbb{C}^N.$$

It thus suffices to show that

$$\left| \|Av\|_2^2 - \|v\|_2^2 \right| \leq \left((n-1)\theta_{s,s} + \delta_s \right) \|v\|_2^2,$$

for all (ns)-sparse vectors $v \in \mathbb{C}^N$. Denote by $S \coloneqq \{j_1, \dots, j_{ns}\}$ the support set of v, i.e., $S = \{j \in \{1, \dots, N\} \colon v_j \neq 0\}$, and decompose S into the subsets $S_1, \dots, S_n$, where $S_i \coloneqq \{j_{(i-1)s+1}, \dots, j_{is}\}$, $i \in \{1, \dots, n\}$. We can write

$$v = \sum_{i=1}^{n} v_{S_i}. \tag{17}$$

Note that v_{S_i} and v_{S_j}, $i \neq j$, $i, j \in \{1, \dots, n\}$, are s-sparse and disjointly supported. The latter property together with (17) implies

$$\|v\|_2^2 = \sum_{i=1}^{n} \|v_{S_i}\|_2^2. \tag{18}$$

We compute

$$\begin{aligned}
\left| \|Av\|_2^2 - \|v\|_2^2 \right| &= \left| \langle (A^{\mathsf{H}}A - I)v, v \rangle \right| \\
&\leq \sum_{i=1}^{n} \sum_{j=1}^{n} \left| \langle (A^{\mathsf{H}}A - I)v_{S_i}, v_{S_j} \rangle \right| \\
&= \sum_{i=1}^{n} \left| \langle (A^{\mathsf{H}}A - I)v_{S_i}, v_{S_i} \rangle \right| + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \left| \langle (A^{\mathsf{H}}A - I)v_{S_i}, v_{S_j} \rangle \right| \\
&\overset{\text{(a)}}{=} \sum_{i=1}^{n} \left| \langle (A_{S_i}^{\mathsf{H}}A_{S_i} - I)v_{S_i}, v_{S_i} \rangle \right| + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \left| \langle A_{S_j}^{\mathsf{H}}A_{S_i}v_{S_i}, v_{S_j} \rangle \right| \\
&\overset{\text{(b)}}{\leq} \sum_{i=1}^{n} \delta_s \|v_{S_i}\|_2^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \theta_{s,s} \|v_{S_i}\|_2 \|v_{S_j}\|_2 \\
&= \theta_{s,s} \left(\sum_{i=1}^{n} \|v_{S_i}\|_2 \right)^2 - (\theta_{s,s} - \delta_s) \sum_{i=1}^{n} \|v_{S_i}\|_2^2 \\
&\overset{\text{(c)}}{\leq} \theta_{s,s} n \sum_{i=1}^{n} \|v_{S_i}\|_2^2 - (\theta_{s,s} - \delta_s) \sum_{i=1}^{n} \|v_{S_i}\|_2^2
\end{aligned}$$

$$\stackrel{\text{(d)}}{=} ((n-1)\theta_{s,s} + \delta_s) \|v\|_2^2,$$

where (a) follows as v_{S_i} and v_{S_j}, $i \neq j$, $i, j \in \{1, \ldots, n\}$, are disjointly supported and A_{S_i} denotes the matrix obtained from A by retaining the columns indexed by S_i, (b) holds by definition of δ_s and $\theta_{s,s}$, in (c) we used the Cauchy–Schwarz inequality, and (d) is a consequence of (18).