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 16, 2023

Problem 1

(a) Let $n \in \{1, \dots, N\}$. Then,

$$|\langle x, e_n \rangle| \overset{(a)}{\leq} \|x\|_2 \|e_n\|_2 \overset{(b)}{=} \|x\|_2 \overset{(c)}{\leq} \sqrt{\sum_{i=1}^{d} \left(\frac{1}{\sqrt{d}} K \right)^2} = K, \tag{1}$$

where (a) is by Theorem H5, (b) is because $\mathcal{F}$ is normalized, and (c) follows from the assumption $x \in \left[-\frac{1}{\sqrt{d}}K, \frac{1}{\sqrt{d}}K \right]^d$.

(b) The proof will be effected by induction, starting from $u_0 = 0 \leq 1/2$. Let $n \in \{1, \dots, N\}$ and assume that $|u_{n-1}| \leq \frac{1}{2}$. By subproblem (a), we have $x_n^{\mathcal{F}} \in [-K, K]$, so $u_{n-1} + x_n^{\mathcal{F}} \in [-K - 1/2, K + 1/2]$. Then,

$$|u_n| \overset{(a)}{=} |u_{n-1} + x_n^{\mathcal{F}} - q_n| \overset{(b)}{=} |u_{n-1} + x_n^{\mathcal{F}} - Q(u_{n-1} + x_n^{\mathcal{F}})| \overset{(c)}{\leq} 1/2, \tag{2}$$

where (a) is by (5) in the problem statement, (b) is by (4) in the problem statement, and (c) is a consequence of (2) in the problem statement, with $t = u_{n-1} + x_n^{\mathcal{F}}$.

(c)

$$u_N \overset{(a)}{=} u_0 + \sum_{n=1}^{N} x_n^{\mathcal{F}} - \sum_{n=1}^{N} q_n \overset{(b)}{=} \sum_{n=1}^{N} x_n^{\mathcal{F}} - \sum_{n=1}^{N} q_n, \tag{3}$$

where (a) follows by iterating (5) in the problem statement and (b) is a consequence of $u_0 = 0$. Next, note that as $\mathcal{F}$ has the zero-sum property, one gets

$$\sum_{n=1}^{N} x_n^{\mathcal{F}} = \sum_{n=1}^{N} \langle x, e_n \rangle = \left\langle x, \sum_{n=1}^{N} e_n \right\rangle = 0, \quad \text{for all } x \in \mathbb{R}^d. \tag{4}$$

Therefore, we have

$$u_N = -\sum_{n=1}^{N} q_n. \tag{5}$$

Using (1) and (4) in the problem statement, it follows that $q_n \in D \subseteq \mathbb{Z}$, for all $n \in \{1, \dots, N\}$, so that

$$u_N = -\sum_{n=1}^{N} q_n \in \mathbb{Z}. \tag{6}$$

As by the result in subproblem (b) we have $|u_N| \leq 1/2$, it follows that $u_N = 0$.

(d)

$$x - \tilde{x}^{\mathcal{F}} \overset{(a)}{=} \sum_{n=1}^{N} x_n^{\mathcal{F}} \mathbb{S}^{-1} e_n - \tilde{x}^{\mathcal{F}} \overset{(b)}{=} \sum_{n=1}^{N} x_n^{\mathcal{F}} \mathbb{S}^{-1} e_n - \sum_{n=1}^{N} q_n \mathbb{S}^{-1} e_n \tag{7}$$

$$= \sum_{n=1}^{N} (x_n^{\mathcal{F}} - q_n) \mathbb{S}^{-1} e_n \tag{8}$$

$$\overset{(c)}{=} \sum_{n=1}^{N} (u_n - u_{n-1}) \mathbb{S}^{-1} e_n \tag{9}$$

$$\overset{(d)}{=} \sum_{n=1}^{N-1} u_n \mathbb{S}^{-1}(e_n - e_{n+1}) - u_0 \mathbb{S}^{-1} e_1 + u_N \mathbb{S}^{-1} e_N \tag{10}$$

$$\overset{(e)}{=} \sum_{n=1}^{N-1} u_n \mathbb{S}^{-1}(e_n - e_{n+1}), \tag{11}$$

where (a) is by Theorem H12, (b) follows from (6) in the problem statement, (c) is by (5) in the problem statement, (d) is obtained by reorganization of terms and (e) uses the fact that $u_0 = 0$ (see (3) in the problem statement), and $u_N = 0$ (see subproblem (c)). Then, we get

$$\|x - \tilde{x}^{\mathcal{F}}\|_2 \overset{(a)}{=} \left\| \sum_{i=1}^{N-1} u_n \, \mathbb{S}^{-1} \left(e_n - e_{n+1}\right) \right\|_2 \leq \sum_{i=1}^{N-1} \left\| u_n \, \mathbb{S}^{-1} \left(e_n - e_{n+1}\right) \right\|_2 \tag{12}$$

$$\leq \sum_{i=1}^{N-1} |u_n| \, \left\| \mathbb{S}^{-1} \left(e_n - e_{n+1}\right) \right\|_2 \tag{13}$$

$$\overset{(b)}{\leq} \sum_{i=1}^{N-1} (1/2) \, \left|\left|\left| \mathbb{S}^{-1} \right|\right|\right|_2 \, \|e_n - e_{n+1}\|_2 \tag{14}$$

$$= (1/2) \, \left|\left|\left| \mathbb{S}^{-1} \right|\right|\right|_2 \sum_{i=1}^{N-1} \|e_n - e_{n+1}\|_2 \tag{15}$$

$$\overset{(c)}{=} (1/2) \, \left|\left|\left| \mathbb{S}^{-1} \right|\right|\right|_2 \sigma(\mathcal{F}), \tag{16}$$

$$\overset{(d)}{=} \frac{\sigma(\mathcal{F})}{2A}, \tag{17}$$

where (a) follows from (11), b) is by $|u_n| \leq \frac{1}{2}$ (see subproblem (b)) and Theorem H7, (c) is by Definition H9, and (d) is by Theorem H11.

Problem 2

(a) We have

$$\begin{aligned}\|P_{\mathcal{T}^c}(B)x\|_2 &= \|BD_{\mathcal{T}^c}B^{\mathsf{H}}x\|_2 \\ &\overset{\text{(a)}}{=} \|D_{\mathcal{T}^c}B^{\mathsf{H}}x\|_2 \\ &= \|D_{\mathcal{T}^c}A^{\mathsf{H}}AB^{\mathsf{H}}x\|_2 \\ &\overset{\text{(b)}}{=} \|AD_{\mathcal{T}^c}A^{\mathsf{H}}AB^{\mathsf{H}}x\|_2,\end{aligned}$$

where (a) follows by unitarity of B and (b) is by unitarity of A.

(b) Note that if $\operatorname{supp}(A^{\mathsf{H}}x) \subseteq \mathcal{S}$, then $P_{\mathcal{S}}(A)x = x$. Thus,

$$\begin{aligned}\|P_{\mathcal{T}}(A)Ux\|_2 &= \|P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A)x\|_2 \\ &\overset{\text{(a)}}{=} \|P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A)P_{\mathcal{S}}(A)x\|_2 \\ &\overset{\text{(b)}}{\leq} |||P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A)|||_2\|P_{\mathcal{S}}(A)x\|_2, \end{aligned} \tag{18}$$

where in (a) we used that $P_{\mathcal{S}}(A)$ is an orthogonal projection (specifically, $P_{\mathcal{S}}(A)^2 = P_{\mathcal{S}}(A)$) and (b) follows from Theorem H7. Using the result in subproblem (a), we obtain

$$\begin{aligned}\|P_{\mathcal{T}^c}(B)x\|_2 &= \|P_{\mathcal{T}^c}(A)Ux\|_2 \\ &= \|(I - P_{\mathcal{T}}(A))Ux\|_2 \\ &\overset{\text{(a)}}{\geq} \|Ux\|_2 - \|P_{\mathcal{T}}(A)Ux\|_2 \\ &\overset{\text{(b)}}{\geq} \|Ux\|_2 - |||P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A)|||_2\|P_{\mathcal{S}}(A)x\|_2 \\ &\overset{\text{(c)}}{=} (1 - |||P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A)|||_2)\|P_{\mathcal{S}}(A)x\|_2,\end{aligned}$$

where I denotes the $N \times N$ identity matrix, (a) is by the reverse triangle inequality, (b) follows from (18), and in (c) we used that U is unitary as well as the fact that $P_{\mathcal{S}}(A)x = x$.

(c) For $x \in \mathbb{C}^N$, we have

$$\begin{aligned}\|x\|_2 &= \|P_{\mathcal{S}}(A)x + P_{\mathcal{S}^c}(A)x\|_2 \\ &\overset{\text{(a)}}{\leq} \|P_{\mathcal{S}}(A)x\|_2 + \|P_{\mathcal{S}^c}(A)x\|_2 \\ &\overset{\text{(b)}}{\leq} \frac{\|P_{\mathcal{T}^c}(B)P_{\mathcal{S}}(A)x\|_2}{(1 - |||P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A)|||_2)} + \|P_{\mathcal{S}^c}(A)x\|_2 \\ &\overset{\text{(c)}}{\leq} \frac{\|P_{\mathcal{T}^c}(B)(x - P_{\mathcal{S}^c}(A)x)\|_2}{(1 - |||P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A)|||_2)} + \|P_{\mathcal{S}^c}(A)x\|_2\end{aligned}$$

$$\overset{\text{(d)}}{\leq} \frac{\|P_{\mathcal{T}^c}(B)x\|_2 + \|P_{\mathcal{T}^c}(B)P_{\mathcal{S}^c}(A)x\|_2}{(1 - \|\|P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A)\|\|_2)} + \|P_{\mathcal{S}^c}(A)x\|_2$$
$$\overset{\text{(e)}}{\leq} \frac{\|P_{\mathcal{T}^c}(B)x\|_2 + \|P_{\mathcal{S}^c}(A)x\|_2}{(1 - \|\|P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A)\|\|_2)} + \|P_{\mathcal{S}^c}(A)x\|_2,$$

where (a) is by the triangle inequality, (b) follows from the result in subproblem (b) together with $P_{\mathcal{S}}(A)^2 = P_{\mathcal{S}}(A)$, in (c) we used $P_{\mathcal{S}}(A) = I - P_{\mathcal{S}^c}(A)$, (d) holds by the triangle inequality, and in (e) we used the fact that $P_{\mathcal{T}^c}(B)$ is an orthogonal projection (specifically, $\|P_{\mathcal{T}^c}(B)y\|_2 \leq \|y\|_2$, for all $y \in \mathbb{C}^N$).

(d) Consider the SVD $C = U_C \Sigma V_C^{\mathsf{H}}$, where $U_C, V_C \in \mathbb{C}^{N \times N}$ are unitary matrices and $\Sigma \in \mathbb{C}^{N \times N}$ is a diagonal matrix containing the singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$ of C, where $r \in \{1, \ldots, N\}$ denotes the rank of C. As U_C and V_C are unitary, we have

$$\|\|C\|\|_2 = \max_{\|x\|_2=1} \|Cx\|_2 = \max_{\|x\|_2=1} \|U_C \Sigma V_C^{\mathsf{H}} x\|_2 = \max_{\|y\|_2=1} \|\Sigma y\|_2 = \sigma_1$$

and

$$\mathrm{Tr}(CC^{\mathsf{H}}) = \mathrm{Tr}(U_C \Sigma V_C^{\mathsf{H}} V_C \Sigma^{\mathsf{H}} U_C^{\mathsf{H}}) = \mathrm{Tr}(\Sigma\Sigma^{\mathsf{H}}) = \sum_{j=1}^{r} \sigma_j^2.$$

Thus, $\|\|C\|\|_2 \leq \sqrt{\mathrm{Tr}(CC^{\mathsf{H}})}$. As $A = (a_1, \ldots, a_N)$ is a unitary matrix, we have

$$\begin{aligned}
\mathrm{Tr}(CC^{\mathsf{H}}) &= \mathrm{Tr}(A^{\mathsf{H}} CC^{\mathsf{H}} A) \\
&= \sum_{i=1}^{N} \langle CC^{\mathsf{H}} a_i, a_i \rangle \\
&= \sum_{i=1}^{N} \|C^{\mathsf{H}} a_i\|_2^2 \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \left|\langle C^{\mathsf{H}} a_i, a_j \rangle\right|^2 \\
&= \sum_{i,j=1}^{N} |\langle a_i, Ca_j \rangle|^2,
\end{aligned}$$

and the desired inequality $\|\|C\|\|_2 \leq \sqrt{\sum_{i,j=1}^{N} |\langle a_i, Ca_j \rangle|^2}$ follows.

(e) Application of the result in subproblem (d) results in

$$\|\|P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A)\|\|_2 \leq \sqrt{\sum_{i,j=1}^{N} |\langle a_i, P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A)a_j \rangle|^2}$$

$$
\begin{aligned}
&\stackrel{\text{(a)}}{=} \sqrt{\sum_{i,j=1}^{N} |\langle P_{\mathcal{T}}(A)a_i, UP_{\mathcal{S}}(A)a_j\rangle|^2} \\
&= \sqrt{\sum_{i,j=1}^{N} |\langle AD_{\mathcal{T}}A^{\mathsf{H}}a_i, UAD_{\mathcal{S}}A^{\mathsf{H}}a_j\rangle|^2} \\
&= \sqrt{\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{S}} |\langle a_i, Ua_j\rangle|^2} \\
&= \sqrt{\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{S}} |\langle U^{\mathsf{H}}a_i, a_j\rangle|^2} \\
&= \sqrt{\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{S}} |\langle BA^{\mathsf{H}}a_i, a_j\rangle|^2} \\
&= \sqrt{\sum_{i\in\mathcal{T}}\sum_{j\in\mathcal{S}} |\langle b_i, a_j\rangle|^2} \\
&\le \mu([A\ B])\sqrt{|\mathcal{S}||\mathcal{T}|},
\end{aligned}
$$

where in (a) we used that $P_{\mathcal{T}}(A)^{\mathsf{H}} = P_{\mathcal{T}}(A)$, which follows from the fact that orthogonal projections are self-adjoint.

(f) Thanks to the results in subproblems (c) and (e), we have

$$
\begin{aligned}
\|x\|_2 &\le \frac{\|P_{\mathcal{S}^c}(A)x\|_2 + \|P_{\mathcal{T}^c}(B)x\|_2}{1 - |\!|\!| P_{\mathcal{T}}(A)UP_{\mathcal{S}}(A) |\!|\!|_2} + \|P_{\mathcal{S}^c}(A)x\|_2 \\
&\le \frac{\|P_{\mathcal{S}^c}(A)x\|_2 + \|P_{\mathcal{T}^c}(B)x\|_2}{1 - \mu([A\ B])\sqrt{|\mathcal{S}||\mathcal{T}|}} + \|P_{\mathcal{S}^c}(A)x\|_2 \\
&\le \left(1 + \frac{1}{1 - \mu([A\ B])\sqrt{|\mathcal{S}||\mathcal{T}|}}\right)(\|P_{\mathcal{S}^c}(A)x\|_2 + \|P_{\mathcal{T}^c}(B)x\|_2).
\end{aligned}
$$

Problem 3

(a) Fix $\epsilon > 0$ arbitrarily. Consider a minimal (2ϵ)-bracket-covering $\{[\ell_j, u_j]\}_{j=1}^m$ of $\mathcal{F}$ with respect to $\|\cdot\|_{\mathcal{P}}$, where $m := N_{[\,]}(2\epsilon, \mathcal{F}, \|\cdot\|_{\mathcal{P}})$. Define $g_j(x) := \frac{u_j(x)+\ell_j(x)}{2}$, for $j \in \{1, \dots, m\}$. We claim that the set $\{g_j\}_{j=1}^m$ is an ϵ-covering of $\mathcal{F}$ with respect to $\|\cdot\|_{\mathcal{P}}$. As the brackets $\{[\ell_j, u_j]\}_{j=1}^m$ are a (2ϵ)-bracket-covering of $\mathcal{F}$, for every $f \in \mathcal{F}$, there is a $j \in \{1, \dots, m\}$ such that $\ell_j(x) \leq f(x) \leq u_j(x)$, for all $x \in \mathcal{X}$. Hence, for all $x \in \mathcal{X}$ we have

$$\begin{aligned} & \ell_j(x) \leq f(x) \leq u_j(x) \\ \Rightarrow \qquad & -\frac{1}{2}(u_j(x) - \ell_j(x)) \leq f(x) - g_j(x) \leq \frac{1}{2}(u_j(x) - \ell_j(x)) \\ \Rightarrow \qquad & |f(x) - g_j(x)| \leq \frac{1}{2}(u_j(x) - \ell_j(x)). \end{aligned}$$

Taking expectations on both sides and using the hint yields $\|f - g_j\|_{\mathcal{P}} \leq \frac{1}{2}\|u_j - \ell_j\|_{\mathcal{P}} \leq \epsilon$. We have therefore found an ϵ-covering $\{g_j\}_{j=1}^m$ with m elements which implies that the cardinality of a minimal covering must satisfy $N(\epsilon, \mathcal{F}, \|\cdot\|_{\mathcal{P}}) \leq m = N_{[\,]}(2\epsilon, \mathcal{F}, \|\cdot\|_{\mathcal{P}})$, as desired.

(b) i. Fix $f \in \mathcal{F}$ arbitrarily and let $j \in \{1, \dots, m\}$ be such that $\ell_j \leq f \leq u_j$. Now we estimate

$$\begin{aligned} \left(\frac{1}{n}\sum_{i=1}^n f(X_i)\right) - \mathbb{E}[f(X)] &= \left(\frac{1}{n}\sum_{i=1}^n f(X_i)\right) - \mathbb{E}[u_j(X)] + \mathbb{E}[u_j(X)] - \mathbb{E}[f(X)] \\ &\overset{(a)}{\leq} \left(\frac{1}{n}\sum_{i=1}^n u_j(X_i)\right) - \mathbb{E}[u_j(X)] + \mathbb{E}[u_j(X) - \ell_j(X)] \\ &\overset{(b)}{\leq} \left(\frac{1}{n}\sum_{i=1}^n u_j(X_i)\right) - \mathbb{E}[u_j(X)] + \epsilon, \end{aligned}$$

where in (a) we used $\ell_j \leq f \leq u_j$ in combination with the hint, and in (b) we employed

$$\mathbb{E}[u_j(X) - \ell_j(X)] = \mathbb{E}[|u_j(X) - \ell_j(X)|] = \|u_j - \ell_j\|_{\mathcal{P}} \leq \epsilon.$$

ii. From the result in subproblem (b).i. it follows that

$$\sup_{f\in\mathcal{F}} \left(\left(\frac{1}{n}\sum_{i=1}^n f(X_i)\right) - \mathbb{E}[f(X)]\right) \leq \epsilon + \max_{j\in\{1,\dots,m\}} \left(\left(\frac{1}{n}\sum_{i=1}^n u_j(X_i)\right) - \mathbb{E}[u_j(X)]\right).$$

Hence, the event

$$\sup_{f\in\mathcal{F}}\left(\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right)-\mathbb{E}[f(X)]\right)>2\epsilon$$

implies the event

$$\max_{j\in\{1,\ldots,m\}}\left(\left(\frac{1}{n}\sum_{i=1}^{n} u_j(X_i)\right)-\mathbb{E}[u_j(X)]\right)>\epsilon.$$

We can therefore upper-bound as follows

$$\begin{aligned}
&\mathbb{P}\left(\sup_{f\in\mathcal{F}}\left(\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right)-\mathbb{E}[f(X)]\right)>2\epsilon\right)\\
&\leq \mathbb{P}\left(\max_{j\in\{1,\ldots,m\}}\left(\left(\frac{1}{n}\sum_{i=1}^{n} u_j(X_i)\right)-\mathbb{E}[u_j(X)]\right)>\epsilon\right)\\
&= \mathbb{P}\left(\exists j\in\{1,\ldots,m\}:\left(\frac{1}{n}\sum_{i=1}^{n} u_j(X_i)\right)-\mathbb{E}[u_j(X)]>\epsilon\right)\\
&\overset{(a)}{\leq} \sum_{j=1}^{m}\mathbb{P}\left(\left(\frac{1}{n}\sum_{i=1}^{n} u_j(X_i)\right)-\mathbb{E}[u_j(X)]>\epsilon\right)\\
&\leq \sum_{j=1}^{m}\mathbb{P}\left(\left|\left(\frac{1}{n}\sum_{i=1}^{n} u_j(X_i)\right)-\mathbb{E}[u_j(X)]\right|>\epsilon\right)\xrightarrow[n\to\infty]{(b)}0,
\end{aligned}$$

where in (a) we used the union bound and in (b) the weak law of large numbers.

iii. We have

$$\begin{aligned}
&\sup_{f\in\mathcal{F}}\left|\mathbb{E}[f(X)]-\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right)\right|\\
&=\sup_{f\in\mathcal{F}}\max\left\{\mathbb{E}[f(X)]-\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right),\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right)-\mathbb{E}[f(X)]\right\}\\
&=\max\left\{\sup_{f\in\mathcal{F}}\left(\mathbb{E}[f(X)]-\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right)\right),\sup_{f\in\mathcal{F}}\left(\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\right)-\mathbb{E}[f(X)]\right)\right\}
\end{aligned}$$

and hence

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{f\in\mathcal{F}}\left|\mathbb{E}[f(X)]-\left(\frac{1}{n}\sum_{i=1}^{n}f(X_i)\right)\right|>2\epsilon\right)\\
\leq\ &\mathbb{P}\left(\sup_{f\in\mathcal{F}}\left(\left(\frac{1}{n}\sum_{i=1}^{n}f(X_i)\right)-\mathbb{E}[f(X)]\right)>2\epsilon\right)\\
&+\mathbb{P}\left(\sup_{f\in\mathcal{F}}\left(\mathbb{E}[f(X)]-\left(\frac{1}{n}\sum_{i=1}^{n}f(X_i)\right)\right)>2\epsilon\right)\\
&\xrightarrow[n\to\infty]{(a)} 0,
\end{aligned}
$$

where in (a) we used the result from subproblem (b).ii. as well as (9) from the problem statement.

(c) Since, by the assumption in the problem statement, $N_{[\,]}(\epsilon,\mathcal{F},\|\cdot\|_{\mathcal{P}})<\infty$ holds for all $\epsilon>0$, we have in particular $N_{[\,]}(\frac{\delta}{2},\mathcal{F},\|\cdot\|_{\mathcal{P}})<\infty$. Thus the result from subproblem (b).iii. with $\epsilon=\frac{\delta}{2}$ implies

$$
\lim_{n\to\infty}\mathbb{P}\left(\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}f(X_i)-\mathbb{E}[f(X)]\right|>\delta\right)=0
$$

as desired.

Problem 4

We start by proving the claim in the hint, namely that, for $u \in \mathbb{C}^s$ and $v \in \mathbb{C}^s$,

$$\text{if } \max_{i=1,\dots,s} |u_i| \leq \min_{i=1,\dots,s} |v_i|, \quad \text{then } \|u\|_2 \leq \frac{\|v\|_1}{\sqrt{s}}. \tag{19}$$

This is accomplished by first observing that

$$\|u\|_2^2 = \sum_{i=1}^{s} |u_i|^2 \leq \sum_{i=1}^{s} \max_{i=1,\dots,s} |u_i|^2 = s \left(\max_{i=1,\dots,s} |u_i| \right)^2 \leq s \left(\min_{i=1,\dots,s} |v_i| \right)^2, \tag{20}$$

where the last inequality is a consequence of the assumption on u and v in (19). One further has

$$s \left(\min_{i=1,\dots,s} |v_i| \right)^2 = s \left(\frac{1}{s} \sum_{i=1}^{s} \min_{i=1,\dots,s} |v_i| \right)^2 \leq s \left(\frac{1}{s} \sum_{i=1}^{s} |v_i| \right)^2 = \frac{\|v\|_1^2}{s}, \tag{21}$$

so that taking square roots in (20) and (21) yields the claim.

Let us now fix a vector $x \in \mathbb{C}^N$ and decompose it into a sum of vectors each of sparsity s. More concretely, we define the index set $S_1 \subseteq \{1, \dots, N\}$ to be given by the locations of the s largest (in terms of absolute value) entries of x, $S_2 \subseteq \{1, \dots, N\} \setminus S_1$ to be the locations of the s largest entries of x that are not in S_1, $S_3 \subseteq \{1, \dots, N\} \setminus S_1 \cup S_2$ to be the locations of the s largest entries of x that are not in $S_1 \cup S_2$. We continue this procedure to construct the index sets $S_1, \dots, S_k$, for some integer $k \geq 1$, such that $S_1 \cup S_2 \cup \dots \cup S_k = \{1, \dots, N\}$, where the last index set S_k might contain fewer than s elements. Given an index set S, we use x_S to denote the vector obtained from x by setting to zero all the components that are not indexed by S. We then have the decomposition into disjoint s-sparse vectors according to

$$x = \sum_{i=1}^{k} x_{S_i},$$

and, by the triangle inequality, we further get

$$\|\Phi x\|_2 = \left\| \sum_{i=1}^{k} \Phi x_{S_i} \right\|_2 \leq \sum_{i=1}^{k} \|\Phi x_{S_i}\|_2. \tag{22}$$

As the vectors x_{S_i}, for $i = 1, \dots, k$, are s-sparse, by definition of the restricted isometry constant (Definition H4 in the Handout), we have

$$\left| \|\Phi x_{S_i}\|_2^2 - \|x_{S_i}\|_2^2 \right| \leq \delta_s \|x_{S_i}\|_2^2,$$

which directly implies

$$\|\Phi x_{S_i}\|_2 \leq \sqrt{1+\delta_s}\,\|x_{S_i}\|_2 \,. \tag{23}$$

By construction, for $i = 2, \ldots, k$, all the non-zero components of x_{S_i} are smaller than or equal to (in absolute value) the non-zero components of $x_{S_{i-1}}$. Therefore, upon application of (19) with u the vector containing the non-zero components of x_{S_i} and v the vector containing the non-zero components of $x_{S_{i-1}}$, we obtain

$$\|x_{S_i}\|_2 \leq \frac{\left\|x_{S_{i-1}}\right\|_1}{\sqrt{s}}, \quad \text{for all } i = 2, \ldots, k. \tag{24}$$

Combining the previous results according to

$$\begin{aligned}
\|\Phi x\|_2 \overset{(22)}{\leq} \sum_{i=1}^{k} \|\Phi x_{S_i}\|_2 \overset{(23)}{\leq} \sqrt{1+\delta_s}\sum_{i=1}^{k} \|x_{S_i}\|_2 &= \sqrt{1+\delta_s}\left(\|x_{S_1}\|_2 + \sum_{i=2}^{k} \|x_{S_i}\|_2\right) \\
&\overset{(24)}{\leq} \sqrt{1+\delta_s}\left(\|x_{S_1}\|_2 + \frac{\sum_{i=2}^{k}\left\|x_{S_{i-1}}\right\|_1}{\sqrt{s}}\right) \\
&\leq \sqrt{1+\delta_s}\left(\|x\|_2 + \frac{\|x\|_1}{\sqrt{s}}\right)
\end{aligned}$$

yields the desired result.