**ETH**
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair for Mathematical Information Science

Prof. Dr. H. Bölcskei
Sternwartstrasse 7
CH-8092 Zürich

# Mathematics of Information

Spring semester 2022

## Solutions to Problem Set 9

## Problem 1 Metric entropy and minimax code length ☕

We first show $L(\epsilon,\mathcal{C}) \geq \lceil \log_2 N(\epsilon;\mathcal{C},\rho)\rceil$. Let $E, D$ be an encoder decoder pair that achieve error $\epsilon$ with code length $L(\epsilon,\mathcal{C})$. Next we define the sets

$$\mathcal{S} = \{E(f) \mid f \in \mathcal{C}\} \subset \{0,1\}^{L(\epsilon,\mathcal{C})} \quad \text{and} \quad \mathcal{X} = \{D(s) \mid s \in \mathcal{S}\} \subset \mathcal{C}$$

The cardinality of these sets fulfills $\#\mathcal{X} \leq \#S \leq \#\{0,1\}^{L(\epsilon,\mathcal{C})} = 2^{L(\epsilon,\mathcal{C})}$. It remains to show that $\mathcal{X}$ is a cover for $\mathcal{C}$. Indeed, $\forall f \in \mathcal{C}$ we have by definition $E(f) \in \mathcal{S}$ and therefore $D(E(f)) \in \mathcal{X}$. Furthermore, $\|D(E(f)) - f\|_{L^2(\Omega)} \leq \epsilon$ since the pair $E, D$ achieve this error. Hence, $\mathcal{X}$ is an $\epsilon$ covering with at most $2^{L(\epsilon,\mathcal{C})}$ elements. Thus, the cardinality of the minimal covering must fulfill $N(\epsilon;\mathcal{C},\rho) \leq 2^{L(\epsilon,\mathcal{C})}$ and therefore $\log_2 N(\epsilon;\mathcal{C},\rho) \leq L(\epsilon,\mathcal{C})$. Taking $\lceil\cdot\rceil$ on both sides, where the right-hand side does not change as it is already integer, yields $\lceil \log_2 N(\epsilon;\mathcal{C},\rho)\rceil \leq L(\epsilon,\mathcal{C})$.

We now show $L(\epsilon,\mathcal{C}) \leq \lceil \log_2 N(\epsilon;\mathcal{C},\rho)\rceil$. Let $\{f_1,\dots,f_{N(\epsilon)}\}$ be a minimal $\epsilon$-covering of $\mathcal{C}$. I.e., for any $f \in \mathcal{C}$, $\exists i \in \{1,\dots,N(\epsilon)\}$ with

$$\|f - f_i\|_{L^2(\Omega)} \leq \epsilon. \tag{1}$$

Now we define an encoder $E$ and decoder $D$ as follows. For $f \in \mathcal{C}$ it identifies the index $i \in \{1,\dots,N(\epsilon)\}$ such that (1) holds and encodes $i$ as a bit string of fixed length $\lceil \log_2 N(\epsilon;\mathcal{C},\rho)\rceil$. Given a bit string, the decoder $D$ decodes the $i$ and produces $f_i$.

## Problem 2 Covering of unit cube. ☕

Following the proof of the one-dimensional case presented in the lecture, we proceed in two steps.

First, we prove an upper-bound on the covering number. This amounts to constructing an $\varepsilon$-cover of $\mathcal{C}^d$. To this end, we divide $\mathcal{C}^d$ up into smaller cubes of side length $2\varepsilon$ by setting

$$x_{(i_1,\dots,i_d)} := (-1+\varepsilon) + 2(i_1-1)\varepsilon, \dots, (-1+\varepsilon) + 2(i_d-1)\varepsilon),$$

for $i_j \in [1, L]$, $j \in \{1,\dots,d\}$ and $L = 1 + \lfloor \varepsilon^{-1} \rfloor$. That guarantees that for every $x \in \mathcal{C}^d$, there exists an index $(i_1,\dots,i_d)$ such that $\|x - x_{(i_1,\dots,i_d)}\|_\infty \leq \varepsilon$, which, in turn, establishes

$$N(\varepsilon;\mathcal{C}^d,\|\cdot\|_\infty) \leq \left(1 + \lfloor \varepsilon^{-1} \rfloor\right)^d \leq \left(1 + \varepsilon^{-1}\right)^d,$$

and hence yields an upper bound on the metric entropy according to

$$\log_2 N(\varepsilon;\mathcal{C}^d,\|\cdot\|_\infty) \leq \log_2\left\{\left(1+\varepsilon^{-1}\right)^d\right\} \asymp d\log\left(\varepsilon^{-1}\right). \tag{2}$$

To obtain the lower bound, we notice that the points

$$x_{(i_1,\dots,i_d)} \coloneqq (-1+2(i_1-1)\varepsilon, \dots, -1+2(i_d-1)\varepsilon),$$

with $i_j \in [1, L]$, $j \in \{1, \dots, d\}$ and $L = 1 + \lfloor \varepsilon^{-1} \rfloor$, form an $\varepsilon$-packing of $\mathcal{C}^d$, which establishes

$$\left(1 + \left\lfloor \frac{1}{2\varepsilon} \right\rfloor\right)^d \leq M(2\varepsilon; \mathcal{C}^d, \|\cdot\|_\infty).$$

Using the relation between packing and covering numbers presented in the lecture, one obtains the desired lower bound

$$d \log\left(\varepsilon^{-1}\right) \asymp \log_2 M(2\varepsilon; \mathcal{C}^d, \|\cdot\|_\infty) \leq \log_2 N(\varepsilon; \mathcal{C}^d, \|\cdot\|_\infty). \tag{3}$$

Combining the upper bound (2) and the lower bound (3), one obtains the desired result

$$\log_2 N(\varepsilon; \mathcal{C}^d, \|\cdot\|_\infty) \asymp d \log\left(\varepsilon^{-1}\right).$$

## Problem 3 Metric entropy of parametric class of functions ☕☕

a)

$$N(2\epsilon; \mathcal{C}, \rho) \leq M(2\epsilon; \mathcal{C}, \rho) \leq N(\epsilon; \mathcal{C}, \rho) \leq M(\epsilon; \mathcal{C}, \rho). \tag{4}$$

b) i) For every $f_\theta \in \mathcal{F}$, we can find a $\theta_i$ in the set $\{\theta_0, \dots, \theta_T, \theta_{T+1}\}$, such that $|\theta_i - \theta| \leq \epsilon$. The Hint implies $|ln(1+x)| \leq |x|$ for $x \geq 0$, and we thus have

$$\begin{aligned}\|f_{\theta_i} - f_\theta\|_\infty &= \max_{x\in[0,1]} \left| \ln(1+\theta_i x) - \ln(1+\theta x) \right| = \max_{x\in[0,1]} \left| \ln\left(\frac{1+\theta_i x}{1+\theta x}\right) \right| \\ &= \max_{x\in[0,1]} \left| \ln\left(1 + \frac{(\theta_i - \theta)x}{1+\theta x}\right) \right| \leq \max_{x\in[0,1]} \left| \frac{(\theta_i-\theta)x}{1+\theta x} \right| \leq |\theta_i - \theta| \leq \epsilon,\end{aligned}$$

where we assumed $\theta_i \geq \theta$ without loss of generality. Otherwise, we repeat the same steps with the roles of $\theta$ and $\theta_i$ exchanged.

Therefore, we can conclude that the set $\{f_{\theta_0}, \dots, f_{\theta_T}, f_{\theta_{T+1}}\}$ constitutes an $\epsilon$-covering of $\mathcal{F}$. An upper bound on the covering number is hence given by $N(\epsilon; \mathcal{F}, \|\cdot\|_\infty) \leq T + 2 \leq \frac{1}{2\epsilon} + 2$.

ii) We construct an explicit packing as follows. Set $T = \lfloor \frac{1}{3\epsilon} \rfloor$, and for $i = 0, 1, \dots, T$, define $\theta_i = 3\epsilon i$. The Hint implies $\left|\frac{x}{1+x}\right| \leq |ln(1+x)|$, for $x \geq 0$. For all $i, j$ with $i \neq j$

and assuming $\theta_i \geq \theta_j$ without loss of generality, we thus have

$$\|f_{\theta_i} - f_{\theta_j}\|_\infty = \max_{x\in[0,1]} \left|\ln(1+\theta_i x) - \ln(1+\theta_j x)\right| = \max_{x\in[0,1]} \left|\ln\left(\frac{1+\theta_i x}{1+\theta_j x}\right)\right|$$
$$= \max_{x\in[0,1]} \left|\ln\left(1 + \frac{(\theta_i-\theta_j)x}{1+\theta_j x}\right)\right| \geq \max_{x\in[0,1]} \left|\left(\frac{(\theta_i-\theta_j)x}{1+\theta_j x}\right)\Big/\left(1+\frac{(\theta_i-\theta_j)x}{1+\theta_j x}\right)\right|$$
$$= \max_{x\in[0,1]} \left|\left(\frac{(\theta_i-\theta_j)x}{1+\theta_j x}\right)\Big/\left(\frac{1+\theta_j x+\theta_i x-\theta_j x}{1+\theta_j x}\right)\right| = \max_{x\in[0,1]} \left|\frac{(\theta_i-\theta_j)x}{1+\theta_i x}\right|$$
$$\geq \max_{x\in[0,1]} \left|\frac{(\theta_i-\theta_j)x}{2}\right| = \left|\frac{\theta_i-\theta_j}{2}\right| = \left|\frac{3\epsilon(i-j)}{2}\right| > \epsilon,$$

by definition of $\theta_i$. We can therefore conclude that $\{f_{\theta_0}, \ldots, f_{\theta_T}\}$ is an $\epsilon$-packing and the corresponding packing number satisfies $M(\epsilon; \mathcal{F}, \|\cdot\|_\infty) \geq T+1 \geq \frac{1}{3\epsilon}$.

iii) By subproblems (b.i), (b.ii), and (4) we obtain

$$\frac{1}{6\epsilon} \leq M(2\epsilon; \mathcal{F}, \|\cdot\|_\infty) \leq N(\epsilon; \mathcal{F}, \|\cdot\|_\infty) \leq \frac{1}{2\epsilon} + 2,$$

which allows us to conclude that $\log N(\epsilon; \mathcal{F}, \|\cdot\|_\infty) \asymp \log(1/\epsilon)$, as $\epsilon \to 0$.

## Problem 4 Optimality with Johnson-Lindenstrauss.☕

a) Fix $i \in \{1, \ldots, m\}$. Then,

$$\|f(x_i) - y\|_2 = \left\|\frac{1}{m}\sum_{j=1}^m f(x_i) - \frac{1}{m}\sum_{j=1}^m f(x_j)\right\|_2 = \left\|\frac{1}{m}\sum_{j=1}^m (f(x_i) - f(x_j))\right\|_2$$
$$\leq \frac{1}{m}\sum_{j=1}^m \|f(x_i) - f(x_j)\|_2 \leq \frac{1}{m}\sum_{j=1}^m \sqrt{1+\delta}\,\|x_i - x_j\|_2$$
$$< \frac{1}{m} m\sqrt{2}\cdot\sqrt{2} = 2,$$

where the first inequality is by the triangle inequality, the second inequality follows from (5) in the problem statement, and the last inequality is by the assumption $\delta < 1/2$ and the relation $\|x_i - x_j\|_2 = \sqrt{2}$, which follows from $\{x_j\}_{j=1}^m$ being an orthonormal basis. We have therefore established that $\|f(x_i) - y\|_2 < 2$, or equivalently $f(x_i) \in \mathcal{B}(y, 2)$, and this holds for all $i$ such that $1 \leq i \leq m$.

b) We have established in the previous subproblem that $\{f(x_j)\}_{j=1}^m \subset \mathcal{B}(y, 2)$, so that we are left with having to show that $\|f(x_i) - f(x_j)\|_2 > 1$, for $1 \leq i \neq j \leq m$. From (5) in the problem statement, we have

$$\|f(x_i) - f(x_j)\|_2^2 \geq (1-\delta)\|x_i - x_j\|_2^2 = 2(1-\delta) > 1,$$

where the last inequality is thanks to the assumption $\delta < 1/2$. Taking the square root yields the desired result.

c) Using the volume ratio estimate (Lemma 10.5), one gets

$$M(1;\mathcal{B}(y,2),\|\cdot\|_2) \le \frac{vol(2\mathcal{B}(y,2)+\mathcal{B}(y/2,1))}{vol(\mathcal{B}(y/2,1))},$$

where we chose $\mathcal{B}=\mathcal{B}(y,2)$ and $\mathcal{B}'=\mathcal{B}(y/2,1)$. Simplifying with the equalities $\mathcal{B}(y,2) = 2\mathcal{B}(y/2,1)$ and $vol(5\mathcal{B}(y/2,1)) = 5^k vol(\mathcal{B}(y/2,1))$, we obtain

$$M(1;\mathcal{B}(y,2),\|\cdot\|_2) \le 5^k.$$

Taking the logarithm on both sides, and setting $C \coloneqq (\log(5))^{-1}$, yields the desired result according to

$$C\log\left(M(1;\mathcal{B}(y,2),\|\cdot\|_2)\right) \le k.$$

d) Assume for the sake of contradiction that a function $f$ satisfying (5) in the problem statement exists. From the result of subproblem (b), there would hence exist a 1-packing of $(\mathcal{B}(y,2),\|\cdot\|_2)$ with $m$ elements, namely $\{f(x_j)\}_{j=1}^m$ with $y=(1/m)\sum_{j=1}^m f(x_j)$, which, in turn, would imply

$$C\log(m) \le C\log\left(M(1;\mathcal{B}(y,2),\|\cdot\|_2)\right) \overset{(*)}{\le} k.$$

Here, $(*)$ has been established in subproblem (c). Therefore, if $C\log(m) > k$, a function $f\colon \mathbb{R}^m \to \mathbb{R}^k$ satisfying (5) in the problem statement cannot exist.

## Problem 5 Volume ratio estimate for the Hamming cube ☕☕

a) We verify that $d$ satisfies the defining properties of a metric on $\mathbb{H}^n$:

- $d(x,y)\ge 0$ for $x,y\in\mathbb{H}^n$ is satisfied by definition;
- $d(x,y)=0$ implies that $x_i=y_i$ for all $i\in[n]$ which, in turn, implies $x=y$;
- $d(x,y)=\#\{i\in[n]\,|\,x_i\ne y_i\}=\#\{i\in[n]\,|\,y_i\ne x_i\}=d(y,x)$;
- Fix $x,y,z\in\mathbb{H}^n$ and note that, for all $i\in[n]$ such that $x_i\ne z_i$, one has either $x_i\ne y_i$ or $y_i\ne z_i$. Thus, we get

$$\begin{aligned} d(x,z) &= \#\{i\in[n]\,|\,x_i\ne z_i\}\\ &\le \#\left\{\{i\in[n]\,|\,x_i\ne y_i\}\cup\{i\in[n]\,|\,y_i\ne z_i\}\right\}\\ &\le \#\{i\in[n]\,|\,x_i\ne y_i\}+\#\{i\in[n]\,|\,y_i\ne z_i\}\\ &= d(x,y)+d(y,z). \end{aligned}$$

b) $\mathcal{B}(x,m)$ is the set of points at distance $d$ less than or equal to $m$ from $x$ and as such can be expressed as

$$\mathcal{B}(x,m)=\bigcup_{k=0}^m \{y\in\mathbb{H}^n\,|\,d(x,y)=k\}.$$

As this union is over disjoint sets, it follows that

$$\#\mathcal{B}(x,m)=\sum_{k=0}^m \#\{y\in\mathbb{H}^n\,|\,d(x,y)=k\}.$$

The set $\{y\in\mathbb{H}^n\,|\,d(x,y)=k\}$ is the set of points in $\mathbb{H}^n$ which differ from $x$ in exactly $k$

coordinates. There are $\binom{n}{k}$ possible choices for these coordinates, so that

$$\#\mathcal{B}(x,m) = \sum_{k=0}^{m} \binom{n}{k}. \tag{5}$$

c) By definition, for every $x \in \mathbb{H}^n$, one can find $x_j \in \{x_1, \dots, x_{N(m;\mathbb{H}^n,d)}\}$ such that $d(x,x_j) \leq m$, or equivalently, $x \in \mathcal{B}(x_j,m)$. Therefore,

$$\mathbb{H}^n \subseteq \bigcup_{j=1}^{N(m;\mathbb{H}^n,d)} \mathcal{B}(x_j,m),$$

which implies

$$\#\mathbb{H}^n \leq \sum_{j=1}^{N(m;\mathbb{H}^n,d)} \#\mathcal{B}(x_j,m).$$

Now using $\#\mathbb{H}^n = 2^n$ and, from (5), $\#\mathcal{B}(x_j,m) = \sum_{k=0}^{m} \binom{n}{k}$, we get

$$2^n \leq N(m;\mathbb{H}^n,d) \sum_{k=0}^{m} \binom{n}{k},$$

which, after rearranging terms, yields the desired result

$$\frac{2^n}{\sum_{k=0}^{m} \binom{n}{k}} \leq N(m;\mathbb{H}^n,d). \tag{6}$$

d) Take $i,j$ such that $1 \leq i \neq j \leq M(m;\mathbb{H}^n,d)$, so that $d(x_i,x_j) > m$, and fix $x \in \mathcal{B}(x_i,\lfloor m/2 \rfloor)$. We need to show that $x \notin \mathcal{B}(x_j,\lfloor m/2 \rfloor)$, or, equivalently, that $d(x,x_j) > \lfloor m/2 \rfloor$. By the triangle inequality, one has

$$d(x,x_j) \geq d(x_i,x_j) - d(x,x_i) > m - \lfloor m/2 \rfloor = \lceil m/2 \rceil \geq \lfloor m/2 \rfloor.$$

Therefore, the balls $\{\mathcal{B}(x_j,\lfloor m/2 \rfloor)\}_{j=1}^{M(m;\mathbb{H}^n,d)}$ are, indeed, disjoint.

e) Since $\mathcal{B}(x_j,\lfloor m/2 \rfloor) \subseteq \mathbb{H}^n$ by definition, one has $\bigcup_{j=1}^{M(m;\mathbb{H}^n,d)} \mathcal{B}(x_j,\lfloor m/2 \rfloor) \subseteq \mathbb{H}^n$. This implies the following estimate on the cardinalities:

$$\#\left\{ \bigcup_{j=1}^{M(m;\mathbb{H}^n,d)} \mathcal{B}(x_j,\lfloor m/2 \rfloor) \right\} \leq \#\mathbb{H}^n,$$

which, using that the balls are disjoint, yields

$$\sum_{j=1}^{M(m;\mathbb{H}^n,d)} \#\mathcal{B}(x_j,\lfloor m/2 \rfloor) \leq 2^n.$$

With (5), one therefore gets

$$M(m;\mathbb{H}^n,d) \sum_{k=0}^{\lfloor m/2 \rfloor} \binom{n}{k} \leq 2^n,$$

which, after rearranging terms, yields the desired result

$$M(m;\mathbb{H}^n,d) \leq \frac{2^n}{\sum_{k=0}^{\lfloor m/2 \rfloor} \binom{n}{k}}. \tag{7}$$

f) Following the hint, we have $N(m;\mathbb{H}^n,d) \leq M(m;\mathbb{H}^n,d)$. Combining this result with (6) and (7) yields the desired result

$$\frac{2^n}{\sum_{k=0}^{m} \binom{n}{k}} \leq N(m;\mathbb{H}^n,d) \leq M(m;\mathbb{H}^n,d) \leq \frac{2^n}{\sum_{k=0}^{\lfloor m/2 \rfloor} \binom{n}{k}}.$$