ETH
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
February 8, 2021

Problem 1

(a) It follows directly from the definition of $\theta_{s,t}$ that $\theta_{s,t}$ is the smallest number $\tilde{\theta} \geq 0$ such that

$$\left| \frac{|\langle Au, Av\rangle|}{\|u\|_2 \|v\|_2} \right| \leq \tilde{\theta},$$

for all disjointly supported s-sparse and t-sparse vectors $u \in \mathbb{C}^N \backslash \{0\}$ and $v \in \mathbb{C}^N \backslash \{0\}$, respectively. Therefore,

$$\begin{aligned}
\theta_{s,t} &= \max_{\substack{u,v\in\mathbb{C}^N \text{ disjointly } s,t\text{-sparse,} \\ \|u\|_2=\|v\|_2=1}} |\langle Au, Av\rangle| \\
&= \max_{\substack{S,T\subset\{1,\dots,N\}, S\cap T=\emptyset, \\ |S|\leq s, |T|\leq t}} \max_{\substack{u\in\mathbb{C}^{|S|}, v\in\mathbb{C}^{|T|}, \\ \|u\|_2=\|v\|_2=1}} |\langle Au, Av\rangle| \\
&= \max_{\substack{S,T\subset\{1,\dots,N\}, S\cap T=\emptyset, \\ |S|\leq s, |T|\leq t}} \max_{\substack{u\in\mathbb{C}^{|S|}, v\in\mathbb{C}^{|T|}, \\ \|u\|_2=\|v\|_2=1}} |\langle A_S u, A_T v\rangle| \\
&= \max_{\substack{S,T\subset\{1,\dots,N\}, S\cap T=\emptyset, \\ |S|\leq s, |T|\leq t}} \max_{\substack{u\in\mathbb{C}^{|S|}, v\in\mathbb{C}^{|T|}, \\ \|u\|_2=\|v\|_2=1}} |\langle (A_T^H A_S)u, v\rangle|.
\end{aligned} \tag{1}$$

Note that for every $S, T \subset \{1, \dots, N\}$ with $S \cap T = \emptyset, |S| \leq s, |T| \leq t$ and $u \in \mathbb{C}^{|S|}, v \in \mathbb{C}^{|T|}$ with $\|u\|_2 = \|v\|_2 = 1$, the Cauchy–Schwarz inequality yields

$$|\langle (A_T^H A_S)u, v\rangle| \leq \|(A_T^H A_S)u\|_2 \|v\|_2 = \|(A_T^H A_S)u\|_2.$$

On the other hand, $A_T^H A_S \in \mathbb{C}^{|T|\times|S|}$ and hence $(A_T^H A_S)u \in \mathbb{C}^{|T|}$. Therefore, if $(A_T^H A_S)u \neq 0$, $v = (A_T^H A_S)u/\|(A_T^H A_S)u\|_2$ satisfies $v \in \mathbb{C}^{|T|}$ with $\|v\|_2 = 1$, and we get

$$\begin{aligned}
|\langle (A_T^H A_S)u, v\rangle| &= |\langle (A_T^H A_S)u, (A_T^H A_S)u/\|(A_T^H A_S)u\|_2\rangle| \\
&= \frac{\|(A_T^H A_S)u\|_2^2}{\|(A_T^H A_S)u\|_2} = \|(A_T^H A_S)u\|_2.
\end{aligned}$$

Hence,

$$\max_{\substack{u\in\mathbb{C}^{|S|}, v\in\mathbb{C}^{|T|}, \\ \|u\|_2=\|v\|_2=1}} |\langle (A_T^H A_S)u, v\rangle| = \max_{\substack{u\in\mathbb{C}^{|S|}, \\ \|u\|_2=1}} \|(A_T^H A_S)u\|_2 = \max_{\substack{u\in\mathbb{C}^{|S|}, \\ \|u\|_2\leq 1}} \|(A_T^H A_S)u\|_2.$$

Combining this with (1), we obtain

$$\begin{aligned}\theta_{s,t} &= \max_{\substack{S,T\subset\{1,\dots,N\},S\cap T=\emptyset,\\ |S|\le s,|T|\le t}} \max_{\substack{u\in\mathbb{C}^{|S|},\\ \|u\|_2\le 1}} \|(A_T^H A_S)u\|_2 \\ &= \max_{\substack{S,T\subset\{1,\dots,N\},S\cap T=\emptyset,\\ |S|\le s,|T|\le t}} \|A_T^H A_S\|_2.\end{aligned}$$

(b) Let $u, v \in \mathbb{C}^N$ be disjointly supported s-sparse and t-sparse vectors, respectively, let $S := \operatorname{supp}(u) \cup \operatorname{supp}(v)$, and let $u_S, v_S \in \mathbb{C}^{|S|}$ be the restrictions of $u, v \in \mathbb{C}^N$ to S. Since u and v have disjoint supports, we have $\langle u_S, v_S\rangle = 0$ and hence

$$|\langle Au, Av\rangle| = |\langle A_S u_S, A_S v_S\rangle - \langle u_S, v_S\rangle| = |\langle (A_S^H A_S - \mathrm{I}_{|S|})u_S, v_S\rangle|.$$

Applying the Cauchy–Schwarz inequality and the relation

$$\|u_S\|_2 \|A_S^H A_S - \mathrm{I}_{|S|}\|_2 = \|u_S\|_2 \max_{\substack{x\in\mathbb{C}^{|S|},\\ \|x\|_2\le 1}} \|(A_S^H A_S - \mathrm{I}_{|S|})x\|_2 \ge \|(A_S^H A_S - \mathrm{I}_{|S|})u_S\|_2,$$

we get

$$|\langle Au, Av\rangle| \le \|(A_S^H A_S - \mathrm{I}_{|S|})u_S\|_2 \|v_S\|_2 \le \|A_S^H A_S - \mathrm{I}_{|S|}\|_2 \|u_S\|_2 \|v_S\|_2.$$

Based on the lemma in the problem statement and using $\|u_S\|_2 = \|u\|_2$, $\|v_S\|_2 = \|v\|_2$, this allows us to conclude that

$$|\langle Au, Av\rangle| \le \delta_{s+t} \|u\|_2 \|v\|_2,$$

which, in turn, proves

$$\theta_{s,t} \le \delta_{s+t}.$$

(c) (i) Note that each $j \in T$ belongs to exactly s sets S_i, so that

$$u = \frac{1}{s}\sum_{i=1}^{t} u_{S_i} \quad \text{and} \quad \|u\|_2^2 = \frac{1}{s}\sum_{i=1}^{t} \|u_{S_i}\|_2^2.$$

(ii) Using (c)(i) and the triangle inequality, we get

$$\begin{aligned}|\langle Au, Av\rangle| &\le \frac{1}{s}\sum_{i=1}^{t} |\langle Au_{S_i}, Av\rangle| \le \frac{1}{s}\sum_{i=1}^{t} \theta_{s,r}\|u_{S_i}\|_2\|v\|_2 \\ &= \theta_{s,r}\,\frac{1}{s}\left(\sum_{i=1}^{t}\|u_{S_i}\|_2\right)\|v\|_2,\end{aligned} \tag{2}$$

where in the second inequality we used that u_{S_i} and v are disjointly supported s-sparse and r-sparse vectors, respectively. Moreover, note that the Cauchy-

Schwarz inequality yields

$$\left(\sum_{i=1}^{t}\|u_{S_i}\|_2\right)^2 \le \left(\sum_{i=1}^{t}\|u_{S_i}\|_2^2\right)\left(\sum_{i=1}^{t}1\right) = t\left(\sum_{i=1}^{t}\|u_{S_i}\|_2^2\right).$$

This together with (2) and subproblem (c)(i) yields

$$\begin{aligned}|\langle Au, Av\rangle| &\le \theta_{s,r}\,\frac{\sqrt{t}}{s}\left(\sum_{i=1}^{t}\|u_{S_i}\|_2^2\right)^{1/2}\|v\|_2\\ &= \theta_{s,r}\sqrt{\frac{t}{s}}\,\|u\|_2\|v\|_2.\end{aligned}$$

Problem 2

(a) (i) Let us assume that $\|x\|_{2,\infty} = 0$. This implies, that for every $M > 0$,

$$\operatorname{card}\big(\{j \in \{1, \ldots, N\} \colon |x_j| \geq t\}\big) \leq \frac{M^2}{t^2}, \text{ for all } t > 0.$$

Hence, for every $t > 0$,

$$\operatorname{card}\big(\{j \in \{1, \ldots, N\} \colon |x_j| \geq t\}\big) \leq \frac{M^2}{t^2}, \text{ for all } M > 0.$$

Consequently, by choosing $M > 0$ but arbitrarily small, we get

$$\operatorname{card}\big(\{j \in \{1, \ldots, N\} \colon |x_j| \geq t\}\big) = 0, \tag{3}$$

for every $t > 0$. Now, taking t in (3) arbitrarily small, we can conclude that $x = 0$.

(ii) The statement is obvious for $\lambda = 0$. Hence, we can assume that $\lambda \neq 0$. Now, observe that

$$\{j \in \{1, \ldots, N\} \colon |\lambda x_j| \geq t\} = \{j \in \{1, \ldots, N\} \colon |x_j| \geq t/|\lambda|\}.$$

Therefore,

$$\begin{aligned}
\|\lambda x\|_{2,\infty} &= \inf\left\{ M \geq 0 \colon \operatorname{card}\big(\{j \in \{1, \ldots, N\} \colon |\lambda x_j| \geq t\}\big) \leq \frac{M^2}{t^2}, \text{ for all } t > 0 \right\} \\
&= \inf\left\{ M \geq 0 \colon \operatorname{card}\big(\{j \in \{1, \ldots, N\} \colon |x_j| \geq t/|\lambda|\}\big) \leq \frac{M^2}{t^2}, \text{ for all } t > 0 \right\} \\
&= \inf\left\{ M \geq 0 \colon \operatorname{card}\big(\{j \in \{1, \ldots, N\} \colon |x_j| \geq t\}\big) \leq \frac{M^2}{|\lambda|^2 t^2}, \text{ for all } t > 0 \right\} \\
&= |\lambda| \inf\left\{ M \geq 0 \colon \operatorname{card}\big(\{j \in \{1, \ldots, N\} \colon |x_j| \geq t\}\big) \leq \frac{M^2}{t^2}, \text{ for all } t > 0 \right\} \\
&= |\lambda| \|x\|_{2,\infty}.
\end{aligned}$$

(b) (i) Note that

$$\operatorname{card}\big(\{j \in \{1, 2\} \colon |x_j| \geq t\}\big) = \begin{cases} 2, & \text{if} \quad t \leq 2^{-1/2}, \\ 1, & \text{if} \quad 2^{-1/2} < t \leq 1, \\ 0, & \text{if} \quad t > 1. \end{cases}$$

Therefore, for all $t > 0$,

$$\operatorname{card}\big(\{j \in \{1, 2\} \colon |x_j| \geq t\}\big) \leq \frac{1}{t^2}.$$

On the other hand,

$$\operatorname{card}\big(\{j \in \{1, 2\} \colon |x_j| \geq 1\}\big) = 1,$$

and hence, we get

$$\|x\|_{2,\infty} = \inf\left\{ M \geq 0\colon\ \operatorname{card}\bigl(\{j \in \{1,2\}\colon |x_j| \geq t\}\bigr) \leq \frac{M^2}{t^2},\ \text{ for all } t > 0 \right\} = 1.$$

Next, using

$$\operatorname{card}\bigl(\{j \in \{1,2\}\colon |x_j| \geq t\}\bigr) = \operatorname{card}\bigl(\{j \in \{1,2\}\colon |y_j| \geq t\}\bigr),\ \text{ for all } t > 0,$$

yields

$$\|y\|_{2,\infty} = \|x\|_{2,\infty} = 1.$$

(ii) Thanks to (a)(ii), we have

$$\|x+y\|_{2,\infty} = \|(1+2^{-1/2}, 1+2^{-1/2})\|_{2,\infty} = (1+2^{-1/2})\|(1,1)\|_{2,\infty}.$$

We are therefore left with having to calculate $\|(1,1)\|_{2,\infty}$. Let us fix $z := (1,1)$. Then

$$\operatorname{card}\bigl(\{j \in \{1,2\}\colon |z_j| \geq t\}\bigr) = \begin{cases} 2, & \text{if } \quad t \leq 1, \\ 0, & \text{if } \quad t > 1, \end{cases}$$

and hence

$$\|z\|_{2,\infty} = \inf\left\{ M \geq 0\colon\ \operatorname{card}\bigl(\{j \in \{1,2\}\colon |z_j| \geq t\}\bigr) \leq \frac{M^2}{t^2},\ \text{ for all } t > 0 \right\} = 2^{1/2}.$$

Consequently, we get

$$\|x+y\|_{2,\infty} = 2^{1/2}(1+2^{-1/2}) = 2^{1/2}+1 > 2 = \|x\|_{2,\infty} + \|y\|_{2,\infty}.$$

(c) (i) If $|x_j^1 + \cdots + x_j^k| \geq t$ for some $j \in \{1,\ldots,N\}$, then we have that $|x_j^i| \geq t/k$ for this j and some $i \in \{1,\ldots,k\}$. This allows us to conclude that

$$\bigl\{j \in \{1,\ldots,N\}\colon |x_j^1 + \cdots + x_j^k| \geq t\bigr\} \subset \bigcup_{i \in \{1,\ldots,k\}} \bigl\{j \in \{1,\ldots,N\}\colon |x_j^i| \geq t/k\bigr\}.$$

(ii) From (c)(i) we get that

$$\begin{aligned} &\operatorname{card}\Bigl(\bigl\{j \in \{1,\ldots,N\}\colon |x_j^1 + \cdots + x_j^k| \geq t\bigr\}\Bigr) \\ &\leq \sum_{i \in \{1,\ldots,k\}} \operatorname{card}\Bigl(\bigl\{j \in \{1,\ldots,N\}\colon |x_j^i| \geq t/k\bigr\}\Bigr) \\ &\leq \sum_{i \in \{1,\ldots,k\}} \frac{\|x^i\|_{2,\infty}^2}{(t/k)^2} \\ &= \frac{k^2\bigl(\|x^1\|_{2,\infty}^2 + \cdots + \|x^k\|_{2,\infty}^2\bigr)}{t^2}. \end{aligned}$$

We therefore obtain

$$\|x^1 + \cdots + x^k\|_{2,\infty} \leq k\big(\|x^1\|_{2,\infty}^2 + \cdots + \|x^k\|_{2,\infty}^2\big)^{1/2}.$$

(iii) We have

$$\begin{aligned}
&\frac{\big(\|x^1\|_{2,\infty}^2 + \cdots + \|x^k\|_{2,\infty}^2\big)}{\big(\|x^1\|_{2,\infty} + \cdots + \|x^k\|_{2,\infty}\big)^2} \\
&= \left(\frac{\|x^1\|_{2,\infty}}{\big(\|x^1\|_{2,\infty} + \cdots + \|x^k\|_{2,\infty}\big)}\right)^2 + \cdots + \left(\frac{\|x^k\|_{2,\infty}}{\big(\|x^1\|_{2,\infty} + \cdots + \|x^k\|_{2,\infty}\big)}\right)^2 \\
&\leq \frac{\|x^1\|_{2,\infty}}{\big(\|x^1\|_{2,\infty} + \cdots + \|x^k\|_{2,\infty}\big)} + \cdots + \frac{\|x^k\|_{2,\infty}}{\big(\|x^1\|_{2,\infty} + \cdots + \|x^k\|_{2,\infty}\big)} = 1,
\end{aligned}$$

and hence

$$\big(\|x^1\|_{2,\infty}^2 + \cdots + \|x^k\|_{2,\infty}^2\big)^{1/2} \leq \|x^1\|_{2,\infty} + \cdots + \|x^k\|_{2,\infty}. \tag{4}$$

Combining this with (c)(ii), we obtain

$$\|x^1 + \cdots + x^k\|_{2,\infty} \leq k\big(\|x^1\|_{2,\infty} + \cdots + \|x^k\|_{2,\infty}\big).$$

(d) For every $k \in \{1, \ldots, N\}$, we can write

$$\|x\|_2^2 = \sum_{j=1}^{N} (x_j^*)^2 \geq \sum_{j=1}^{k} (x_j^*)^2 \geq k(x_k^*)^2.$$

Raising to the power $1/2$ and taking the maximum over k yields the desired result.

Problem 3

(a) (i) We take two indices $i \neq j$ and assume that there exists at least one element $\theta \in \mathbb{H}^d$ in the intersection $\mathbb{H}_i \cap \mathbb{H}_j$. By the triangle inequality and the definition of $\mathbb{H}_i$ and $\mathbb{H}_j$, we have

$$d_H(\theta^i, \theta^j) \leq d_H(\theta^i, \theta) + d_H(\theta^j, \theta) \leq \varepsilon/2 + \varepsilon/2 \leq \varepsilon.$$

However, by definition of an ε-packing, we must have $d_H(\theta^i, \theta^j) > \varepsilon$, which results in a contradiction and thereby concludes the proof of the sets $\{\mathbb{H}_i\}_{i=1}^M$ being disjoint.

(ii) For a given integer $k \in [0, d]$, the points in $\mathbb{H}^d$ that are at distance k/d of θ^i are exactly the θ obtained by flipping k different coordinates of θ^i. We can therefore observe that there are exactly $\binom{d}{k}$ points at distance k/d from θ^i. Summing over all the integers k such that $k/d \leq \varepsilon/2$ or, equivalently, $k \leq d\varepsilon/2$, one obtains

$$|\mathbb{H}_i| = \sum_{k=0}^{n} \binom{d}{k},$$

with $n = \lfloor d\varepsilon/2 \rfloor$.

(iii) We argue as follows:

$$\begin{aligned}
\log M(\varepsilon; \mathbb{H}^d, d_H) + \log \binom{d}{n} &= \log \left[M(\varepsilon; \mathbb{H}^d, d_H) \binom{d}{n} \right] \\
&\leq \log \left[M(\varepsilon; \mathbb{H}^d, d_H) \sum_{k=0}^{n} \binom{d}{k} \right] \\
&= \log \left[\sum_{i=1}^{M} \sum_{k=0}^{n} \binom{d}{k} \right] \\
&\overset{(a)(ii)}{=} \log \left[\sum_{i=1}^{M} |\mathbb{H}_i| \right] \\
&\overset{(b)(i)}{\leq} \log |\mathbb{H}^d| = \log 2^d = d \log 2,
\end{aligned}$$

where we used the shorthand M for $M(\varepsilon; \mathbb{H}^d, d_H)$. The last inequality comes from the fact that the $\mathbb{H}_i$ are disjoint subsets of $\mathbb{H}^d$, so that the sum of their cardinalities is bounded by the cardinality of $\mathbb{H}^d$. Rearranging terms yields the desired result according to

$$\log M(\varepsilon; \mathbb{H}^d, d_H) \leq d \log 2 - \log \binom{d}{n}. \tag{5}$$

(b) (i) Following the hint, we note that

$$\begin{aligned}\frac{\mathbb{P}[Y=\ell]}{\mathbb{P}[Y=\ell-1]} &= \frac{\binom{d}{\ell}(n/d)^{\ell}(1-n/d)^{d-\ell}}{\binom{d}{\ell-1}(n/d)^{\ell-1}(1-n/d)^{d-\ell+1}} \\ &= \frac{d-\ell+1}{\ell}\frac{n/d}{1-n/d} \\ &= \frac{d-\ell+1}{d-n}\ \frac{n}{\ell}. \end{aligned} \tag{6}$$

Both of the fractions in (6) are larger than 1 for $\ell \le n$ and smaller than 1 for $\ell > n$. This means that $\mathbb{P}[Y=\ell]$ is maximized at $\ell = n$, which is the desired result.

(ii) Following the hint, we have

$$\begin{aligned} 1 &= \sum_{\ell=0}^{d} \mathbb{P}[Y=\ell] \\ &\overset{(b)(i)}{\le} (d+1)\mathbb{P}[Y=n] \\ &= (d+1)\binom{d}{n}(n/d)^n(1-n/d)^{d-n}. \end{aligned}$$

Taking the logarithm, we obtain

$$\begin{aligned} \log\binom{d}{n} &\ge -\log\left\{(d+1)(n/d)^n(1-n/d)^{d-n}\right\} \\ &= -n\log(n/d)-(d-n)\log(1-n/d)-\log(d+1) \\ &= d\left\{-(n/d)\log(n/d)-(1-n/d)\log(1-n/d)\right\}-\log(d+1) \\ &= d\,\phi(n/d)-\log(d+1), \end{aligned}$$

which is the desired result.

(c) We argue as follows:

$$\begin{aligned} \frac{\log M(\varepsilon;\mathbb{H}^d,d_H)}{d} &\overset{(5)}{\le} \log 2 - \frac{\log\binom{d}{n}}{d} \\ &\overset{(b)(ii)}{\le} \log 2 - \phi(n/d) + \frac{\log(d+1)}{d} \\ &= (n/d)\log 2 + (1-n/d)\log 2 + (n/d)\log(n/d) \\ &\quad + (1-n/d)\log(1-n/d) + \frac{\log(d+1)}{d} \\ &= (n/d)\log\left(\frac{n/d}{1/2}\right) + (1-n/d)\log\left(\frac{1-n/d}{1/2}\right) + \frac{\log(d+1)}{d} \\ &= D\big((n/d)\|(1/2)\big) + \frac{\log(d+1)}{d}. \end{aligned}$$

Problem 4

(a) Consider the points $x_1 = 0$ and $x_2 = 1$ in $\mathbb{R}$. The four possible labelings

$$\begin{cases} h_{(-1,-1)}(x_1) = 0 & \text{and} \quad h_{(-1,-1)}(x_2) = 0, \\ h_{(0,0)}(x_1) = 1 & \text{and} \quad h_{(0,0)}(x_2) = 0, \\ h_{(1,1)}(x_1) = 0 & \text{and} \quad h_{(1,1)}(x_2) = 1, \\ h_{(0,1)}(x_1) = 1 & \text{and} \quad h_{(0,1)}(x_2) = 1, \end{cases}$$

are produced by $\mathcal{H}_1$. Therefore, there is a set of 2 points shattered by $\mathcal{H}_1$, which implies $\dim_{VC}(\mathcal{H}_1) \geq 2$.

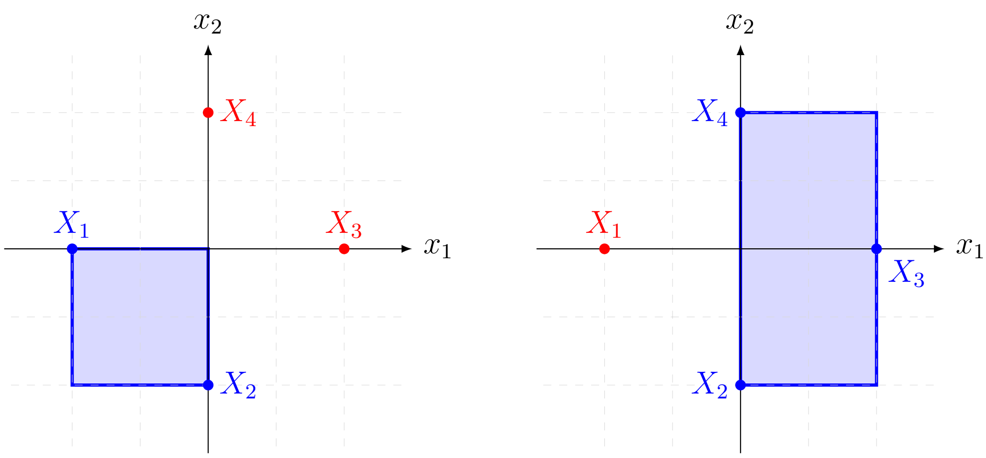
On the other hand, for any set of three distinct points x_1, x_2, and x_3 that we choose without loss of generality such that $x_1 < x_2 < x_3$, there is no closed interval containing x_1 and x_3 but not x_2. Therefore, there is no set of 3 points shattered by $\mathcal{H}_1$, which implies $\dim_{VC}(\mathcal{H}_1) < 3$.

We have therefore proven that $\dim_{VC}(\mathcal{H}_1) = 2$.

(b) (i) The 4 points $X_1 = (-1, 0)$, $X_2 = (0, -1)$, $X_3 = (1, 0)$, and $X_4 = (0, 1)$ are shattered by $\mathcal{H}_2$. To see this, we fix a labeling $(y_1, y_2, y_3, y_4) \in \{0, 1\}^4$. The rectangle $h_{(-y_1,-y_2,y_3,y_4)} \in \mathcal{H}_2$ labels correctly all four points:

$$\begin{cases} h_{(-y_1,-y_2,y_3,y_4)}(X_1) = y_1, \\ h_{(-y_1,-y_2,y_3,y_4)}(X_2) = y_2, \\ h_{(-y_1,-y_2,y_3,y_4)}(X_3) = y_3, \\ h_{(-y_1,-y_2,y_3,y_4)}(X_4) = y_4. \end{cases}$$

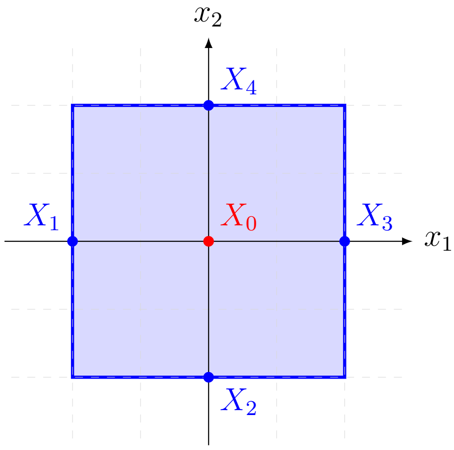
The figures below picture the rectangle $h_{(-1,-1,0,0)}$ (left) and the rectangle $h_{(0,-1,1,1)}$ (right).



(ii) Assume that we can find a_1, b_1, a_2, and b_2 such that

$$\begin{cases} h_{(a_1,a_2,b_1,b_2)}(0,0) = 0, \\ h_{(a_1,a_2,b_1,b_2)}(-1,0) = h_{(a_1,a_2,b_1,b_2)}(1,0) = 1, \\ h_{(a_1,a_2,b_1,b_2)}(0,-1) = h_{(a_1,a_2,b_1,b_2)}(0,1) = 1. \end{cases}$$

Since $h_{(a_1,a_2,b_1,b_2)}(-1,0) = h_{(a_1,a_2,b_1,b_2)}(1,0) = 1$, we must have $a_1 \leq -1$ and $1 \leq b_1$, which implies $a_1 \leq 0 \leq b_1$. Likewise, since $h_{(a_1,a_2,b_1,b_2)}(0,-1) = h_{(a_1,a_2,b_1,b_2)}(0,1) = 1$, we must have $a_2 \leq -1$ and $1 \leq b_2$, which implies $a_2 \leq 0 \leq b_2$. The inequalities $a_1 \leq 0 \leq b_1$ and $a_2 \leq 0 \leq b_2$ together imply $h_{(a_1,a_2,b_1,b_2)}(0,0) = 1$, which yields a contradiction, thereby concluding the proof.



(iii) Take any set of five distinct points in $\mathbb{R}^2$. We call X_1 the "leftmost" point (the point of smallest first coordinate x_1^-), X_2 the "lowest" point (the point of smallest second coordinate x_2^-), X_3 the "rightmost" point (the point of largest first coordinate x_1^+) and X_4 the "highest" point (the point of largest second coordinate x_2^+). Note that these extremal points X_1, X_2, X_3, and X_4 are not necessarily distinct (e.g., there could be a point of both largest first and largest second coordinates, i.e., $X_3 = X_4$). We consider the labeling y that assigns 1 to X_1, X_2, X_3 and X_4 and 0 to a point X_0 distinct from X_1, X_2, X_3, and X_4 (which exists since we consider 5 points in total). If it exists, a rectangle $h_{(a_1,a_2,b_1,b_2)}$ realizing the desired labeling has to be such that $a_1 \leq x_1^-$, $x_1^+ \leq b_1$, $a_2 \leq x_2^-$, and $x_2^+ \leq b_2$. Since by construction of X_1, X_2, X_3 and X_4, we must have $x_1^- \leq x_1^0 \leq x_1^+$ and $x_2^- \leq x_2^0 \leq x_2^+$, we necessarily get $a_1 \leq x_1^0 \leq b_1$ and $a_2 \leq x_2^0 \leq b_2$, which does not produce the correct labeling for the point X_0. This contradiction proves that there is no rectangle $h_{(a_1,a_2,b_1,b_2)}$ yielding the desired labelling y. Therefore, no set of 5 points can be shattered by $\mathcal{H}_2$.

(iv) The class $\mathcal{H}_2$ shatters *a* set of 4 points but does not shatter *any* set of 5 points. By definition of VC dimension, we therefore have $\dim_{VC}(\mathcal{H}_2) = 4$.