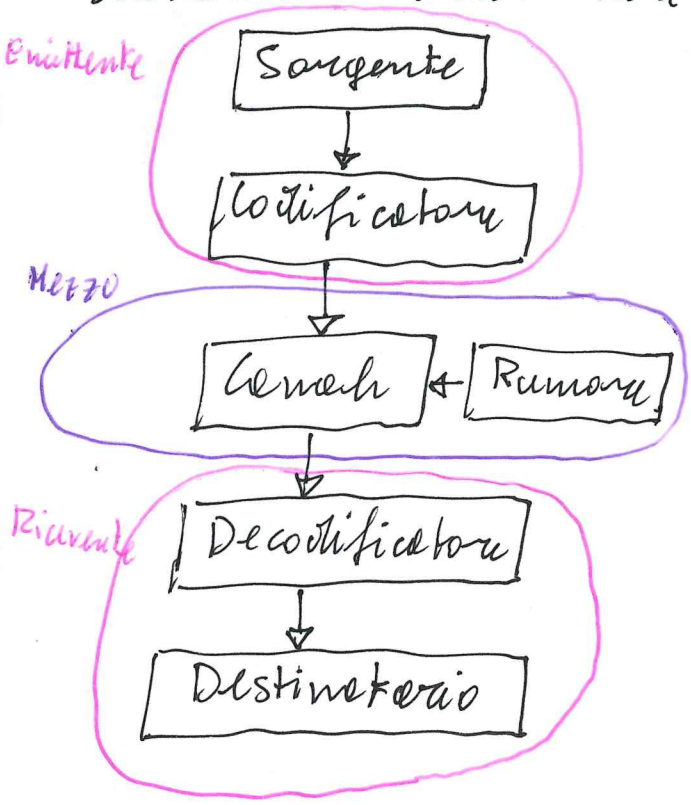
PLS 2018

NICOLA.ARCOZZI@UNIBO.IT

IL CUBO DI HAMMING

Sistema di comunicazione



Terminologia di base.

Alfabeto del codice ("caratteri"):

$A = \{a_1, \dots, a_D\}$. **Binario**: $\{0, 1\}$

Parole di lunghezza n:

$x = (\xi_1, \dots, \xi_n)$ con $\xi_j \in A$

$A^n = A \times \dots \times A$

Codice (con parole di lunghezza n)

$C \subseteq A^n$

Se $x \in C$, x è una **parola** del codice.

Oss. vi sono codici con parole di lunghezza variabile

Es. codice Morse $A = \{\cdot, -\}$

Messaggio: $x_1 x_2 \dots x_n$

con $x_j \in C$.

"Rumore": errori di trasmissione.

Problema 1: rilevazione degli errori

Problema 2: correzione degli errori.

Alfabeto binario: $A = \{0, 1\}$

Stringhe: Parole di lunghezza n:

n	A^n	$\# A^n$
1	0, 1	2
2	00, 10, 01, 11	2^2
3	000, 100, 010, 110, 100, 110, 101, 111	2^3
4	...	2^4
⋮		⋮
n	...	2^n
⋮		⋮

Esempio di codice binario con $n = 2$.

Lettere: a, b

$a \longmapsto 00$

$b \longmapsto 11$

$C = \{00, 11\}$

Sorgente abba

Codificatore 00111100

Canale ↓

Decodificatore 00 11 11 00

Destinatario abba

Supponiamo di aver ricevuto 01 11 00

Cosa ne deduciamo?

01 11 00 $\longmapsto$? b a

C'è stato 1 errore di trasmissione nella prima parola.

$C = \{00, 11\}$:

- rivela 1 errore per parola
- non rivela 2 errori per parola
- non permette la correzione di 1 errore per parola (perché? esercizio)

Esercizio.

Trovare un codice binario (alfabeto $A = \{0,1\}$) con due parole codice (a, b) che permette:

- la correzione di 1 errore per parola
- la rilevazione di 2 errori per parola.

Controllo di parità.

Obiettivo: codificare le lettere $\mathcal{L} = A^n$ $(A = \{0,1\})$ mediante un codice che permette la rivelazione di 1 errore / parola.

Soluzione: Alle lettere

$x = \varepsilon_1 \cdots \varepsilon_n \quad \varepsilon_j \in \{0,1\}$

associamo le parole-codice

$x' = \varepsilon_1 \cdots \varepsilon_n \varepsilon_{n+1}$

dove $\varepsilon_{n+1} = \begin{cases} 0 \text{ se } \varepsilon_1 + \cdots + \varepsilon_n \text{ è pari} \\ 1 \text{ se } \varepsilon_1 + \cdots + \varepsilon_n \text{ è dispari} \end{cases}$

Le parole-codice stanno in A^{n+1}:

$C \subseteq A^{n+1}$

Esercizio: Dimostrare che c'è rilevazione di 1 errore/parola. Mostrare che non c'è correzione di 1 errore/parola.

La distanza di Hamming (1950)

ci permette di inquadrare il problema di rilevare e correggere errori.

Cubo di Hamming di dimensione n:

$H_n = A^n$, $A = \{0,1\}$

(parole binarie di lunghezza n)

Distanza di Hamming:

se $x = x_1 \cdots x_n$, $y = y_1 \cdots y_n \in H_n$

allora $d(x,y) = \#\{j : x_j \neq y_j ; 1 \leq j \leq n\}$

$0 \leq d(x,y) \leq n$

Proprietà elementari:

(1) $d(x,y) \geq 0$ e $d(x,y) = 0 \iff x = y$

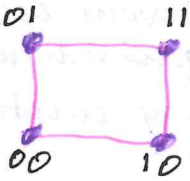
(2) $d(x,y) = d(y,x)$

(3) $d(x,y) \leq d(x,z) + d(z,y)$: disuguaglianza triangolare

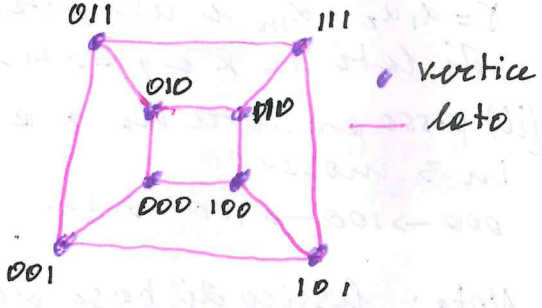
(1)-(2)-(3) sono gli assiomi delle distanze, in generale.

Es. $n=1$

$n=2$



$n=3$



— unisce x, y se $d(x,y) = 1$

Esercizio. Dimostrare che se $x, y \in H_n$, allora $d(x,y) = k$ $\iff$ il numero di "lati" più piccolo per andare da x a y è k.

soluzioni nel caso particolare

$000, 111 \qquad d(x,y) = 3$

$x \quad y$

(i) ogni lato indica il cambiamento di una cifra; ce ne devono essere almeno tre, quindi un percorso da x a y contiene almeno 3 lati:

$\gamma = \alpha_1 \alpha_2 \cdots \alpha_m$ è un percorso di lati da x a $y \Rightarrow m \geq 3$.

(ii) posso andare da x a y in 3 mosse:

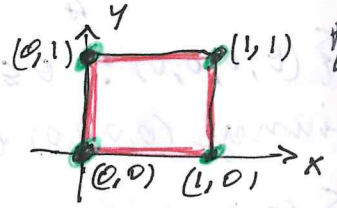
$000 \to 100 \to 110 \to 111$.

Nota: lessico di base dei grafi: vertici, lati, cammini.

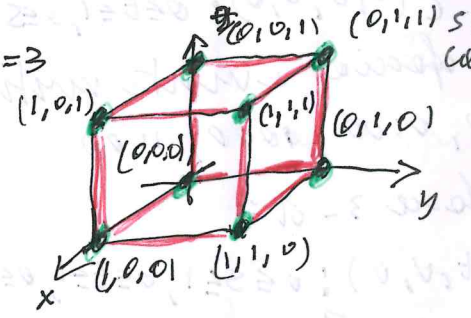
Rappresentare il cubo di Hamming come vertici di un cubo in n dimensioni

$n = 1$ retta cartesiana

$n = 2$ piano cartesiano



$n = 3$ spazio cartesiano



$n = 4$? (x, y, z, w): $x \in \mathbb{R}, y \in \mathbb{R}, z \in \mathbb{R}, z \in \mathbb{R}$

$\mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} = \mathbb{R}^4$

Possiamo immaginare:
vertici, lati, facce, iperfacce
nel cubo Q_4 a 4-dim.

Vertici: $(0,0,0,0)$, $(1,0,0,0)$, ...
elementi in H_4

Spigoli
~~Lati~~: $\{(t,0,0,0) : 0 \le t \le 1\}$
congiunge $(0,0,0,0)$ e $(1,0,0,0)$

Facce: $\{(s,t,0,0) : 0 \le t \le 1, 0 \le s \le 1\}$
è le facce avente vertici
$0000, 0100, 1000, 1100$

Iperfacce 3-dim:
$\{(s,t,v,0) : 0 \le s \le 1, 0 \le t \le 1, 0 \le v \le 1\}$
corrisponde a ...

Il cubo in sé:
$Q_4 = \{(s,t,v,w) : 0 \le s \le 1, \dots, 0 \le v \le 1\}$

Possiamo ~~questo~~ In generale:

$Q_n \subseteq \mathbb{R}^n$: cubo in n-dimensioni
$H_n \subseteq Q_n$: insieme dei vertici
Se $x, y \in H_n$: c'è un lato
tra x e y $\Leftrightarrow$ $d(x,y) = 1 \Leftrightarrow$
x e y sono estremi di
uno spigolo di Q_n.

Le proprietà di Q_n sono
legate a quelle di H_n.

Problema. Quante iperfacce
3-dim ha Q_4?
Quante facce 2-dim?
Quanti spigoli?
Vertici?

Geometria e combinatoria in H_n (e Q_n).

$\#(H_n) = 2^n$.

Quanti sono i lati?

$n=1$ 1 lato

$n=3$ 12 lati

$n=2$ 4 lati

Formula generale.

$$\frac{2^n \times n}{2} = 2^{n-1} \times n = l_n$$

n	l_n
1	1
2	4
3	12
4	32
5	80
6	192
7	448

Perché?

$\forall x \in H_n$ ($\#(H_n) = 2^n$)

x è estremo di 2^n lati:

$2^n \times n$

ma ogni lato è stato contato 2 volte:

$\frac{2^n \times n}{2}$

Quante facce 2-d?

n	f_n
1	0
2	1
3	6
4	24
.	

1. Scelgo un vertice x (delle faccie): 2^n

2. Scelgo un lato α (delle faccie) avente estremo x: n

3. Scelgo un secondo lato β (delle faccie) avente estremo x: $(n-1)$

In tutto: $2^n \times n \times (n-1)$ scelte.

Quante volte è stata scelta F?

2^2 (numero vertici di F) $\times$ 2 (scelte (α,β) o (β,α)).

$$f_n = \frac{2^n \times n \times (n-1)}{2^2 \times 2} = 2^{n-2} \times \frac{n \times (n-1)}{2}$$

$f_4 = 24$

Numero delle facce k-dim in H_n: $2^{n-k} \cdot \binom{n}{k} =$

$= 2^{n-k} \cdot \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{1 \cdot 2 \cdot \ldots \cdot k}$

$k = 0, 1, 2$: o.k.

Come dimostrarlo in generale?

Ragionamento diretto

oppure

Induzione (su k).

~~$A_{n,k} = A_{n,k-1} \cdot \frac{(n-(k-1))}{2}$~~

$= A_{n,k}$: numero delle k-dim. facce in H_n.

$$A_{n,k} = \frac{A_{n,k-1} \cdot (n-(k-1))}{\#\{(k-1) \text{ facce in } H_k\}}$$

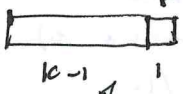
$$= \frac{A_{n,k-1} \cdot (n-k+1)}{A_{k,k-1}} =$$

$$= \frac{A_{n,k-1} \cdot (n-k+1)}{2 \cdot k}$$

k	$A_{k,k-1}$
1	2
2	4
3	6
4	8
⋮	⋮
k	2k

$$= \frac{2^{n-(k-1)}}{2} \cdot \frac{n \cdot \ldots \cdot (n-k+2)(n-k+1)}{1 \cdot \ldots \cdot (k-1) \cdot k}$$

o.k.



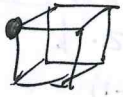
fisso un vertice: 2 scelte

Fisso una posizione: k scelte

$B_{n,k}$: punti a distanza k da un punto di H_n.

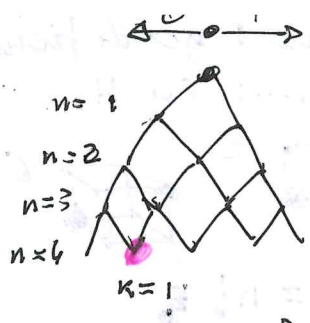
$n=1$ $B_{1,0}=1$ $B_{1,1}=1$

$n=2$  $B_{2,0}=1$ $B_{2,1}=2$ $B_{2,2}=1$

$n=3$  $B_{3,0}=1$ $B_{3,1}=3$ $B_{3,2}=3$ $B_{3,3}=1$

$B_{n,k} = \binom{n}{k}$: triangolo di Tartaglia.

Dim. $x = 00\cdots0 \in H_n$.
I punti a distanza k da x sono $y = \varepsilon_1 \cdots \varepsilon_n$ dove k bit sono 1 (e $n-k$ sono 0).
Questo numero è $\binom{n}{k}$.
Se non lo si sa, lo si può dimostrare

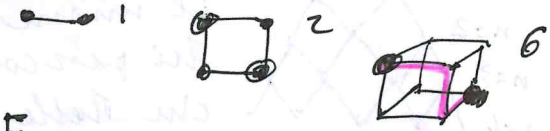


Calcolare il numero di percorsi che dalla cima portano nel punto della riga n al posto k: $\binom{n}{k}$.

Contare geodetiche.
Siano $x, y \in H_n$; $d(x,y) = k \leq n$.
Un cammino da x a y è una geodetica se $\#(\gamma) = k$ (se la sua lunghezza è k).
Quante sono le geodetiche di lunghezza k in H_n:

$G_{n,k}$?

Sottoproblema: geodetiche tra $00\ldots0$ e $11\ldots1$ in H_n.



$F_n = n \cdot F_{n-1} = n!$

Quindi $G_{n,n} = 2^n \cdot n!$

(numero delle geodetiche orientate).

$G_{n,k} = 2^n \cdot n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$

↑ punto iniziale

Contare le isometrie di H_n (quindi di Q_n).

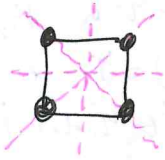
Una funzione $H_n \xrightarrow{F} H_n$ è una isometria $\iff$
$\forall x, y \in H_n : d(F(x), F(y)) = d(x,y)$.

$n=1$

$01 \to 0$
$11 \to 1$
0
$01 \to 1$
$11 \to 0$

$I_1 = 2$

$n=2$



4 simmetrie assiali
4 rotazioni:
$0, \pi/4, \pi/2, \frac{3}{4}\pi$

$I_2 = 8$

Sono tutte? C'è un modo meno analitico per contarle?

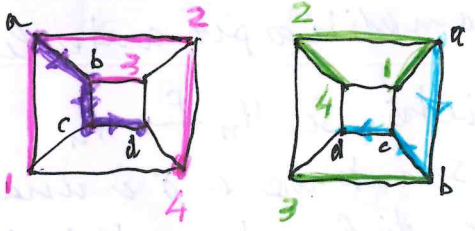
Proprietà. Se $H_n \xrightarrow{F} H_n$ è un'isometria e γ è una geodetica di lunghezza k in H_n, allora $T(\gamma)$ è una curva geodetica di lunghezza k in H_n.

Dim. esercizio.

Teorema. Sieno γ_0, γ_1 geodetiche di lunghezza n in H_n. Allora esiste un'unica isometria $H_n \xrightarrow{T} H_n$ tale che $T(\gamma_0) = \gamma_1$ e che conserva l'orientamento:

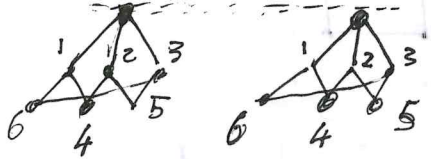
$\gamma_0 = x_0 x_1 x_2 \cdots x_n$
$\gamma_1 = y_0 y_1 y_2 \cdots y_n$ $\Longleftrightarrow T(x_j) = y_j$

Vediamolo per H_3



Dato il Teorema:
$I_n = G_{n,n} = 2^n \cdot n!$

Oppure posso mostrare che $I_n = 2^n \cdot n!$ e usare le proposizioni per dimostrare il Teorema



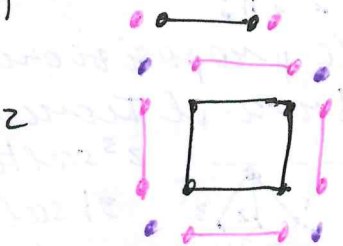
2^3 scelte
$3!$ scelte

oss. In $\mathbb{R}^n$ è ovvio: la scelta della nuova origine degli assi e delle nuove direzioni coordinate determina l'isometria.

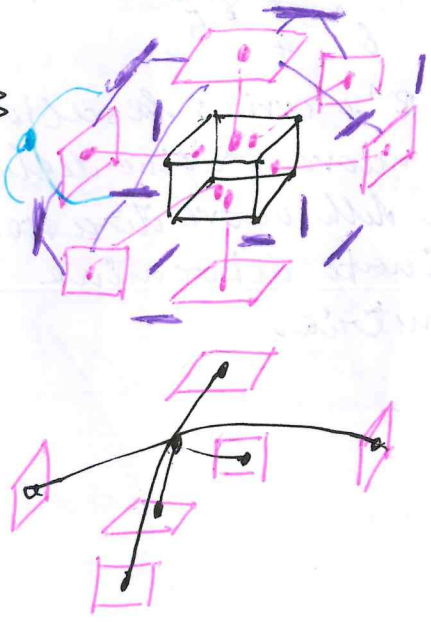
Geometria di H_n e Q_n $(n \leq 4)$

$n = 1$

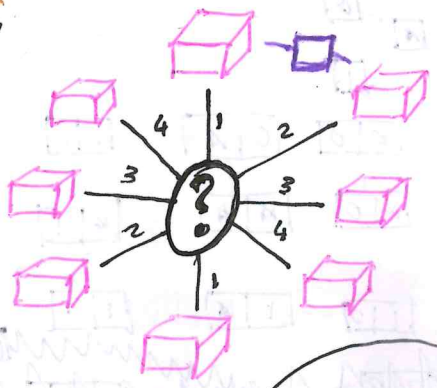
$n = 2$

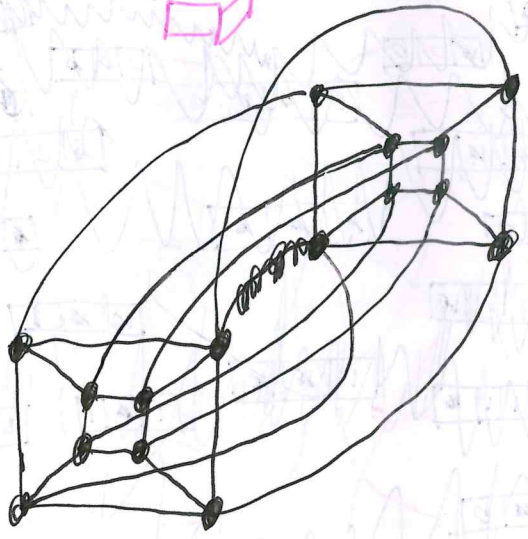


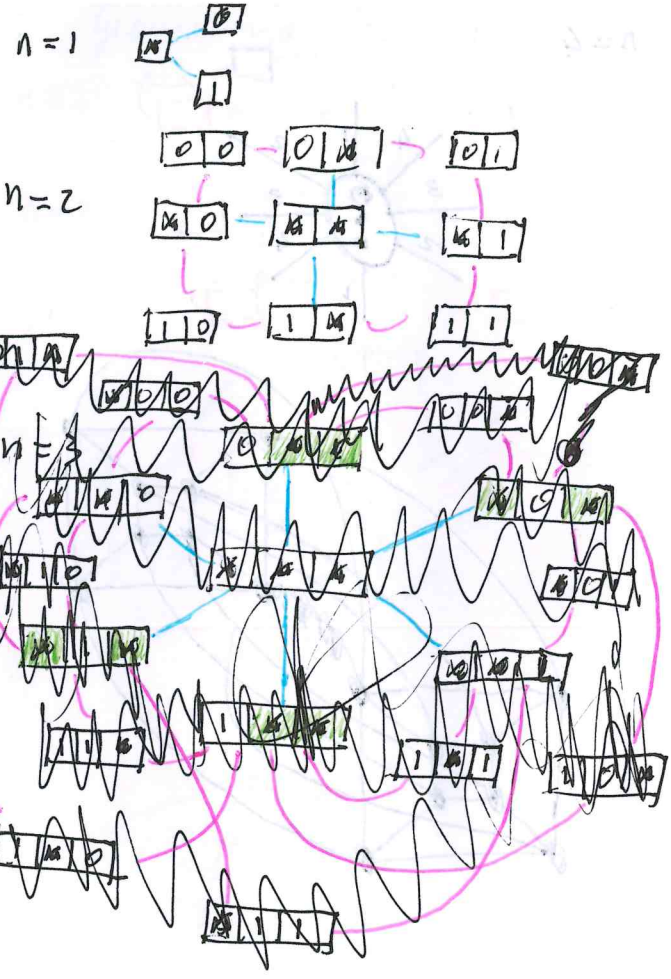
$n = 3$

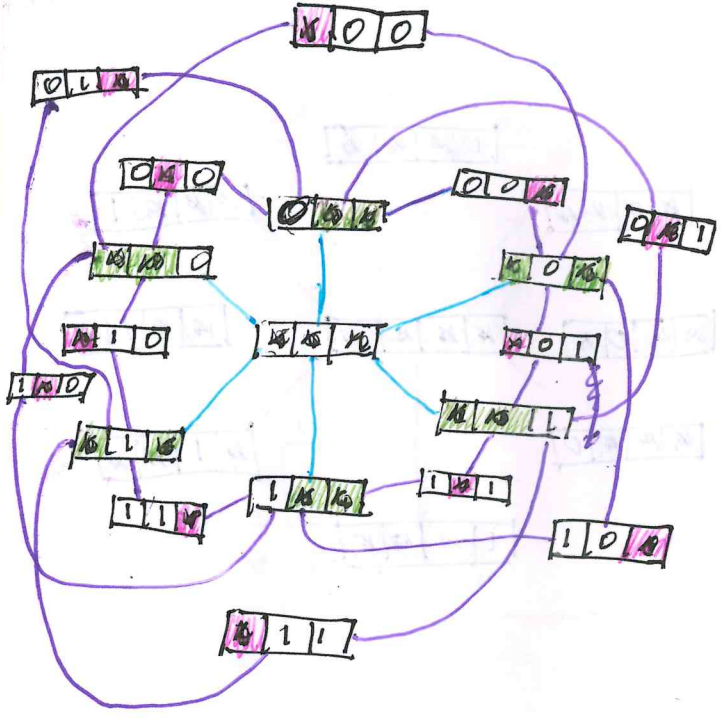


$n = 4$

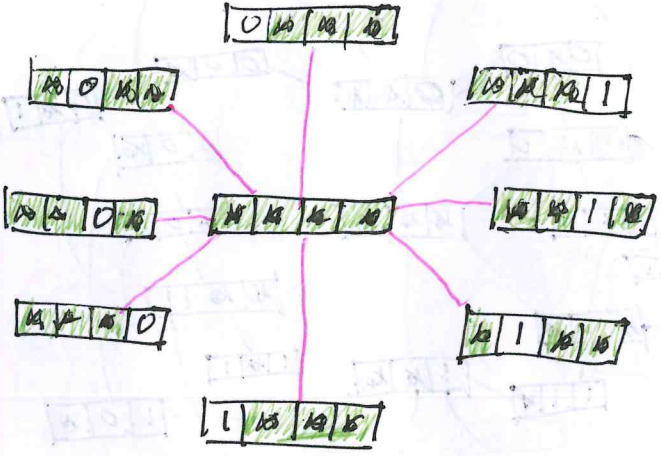




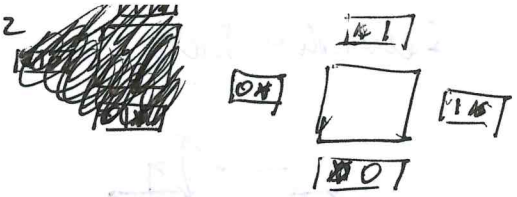

n = 1
n = 2
n = 3



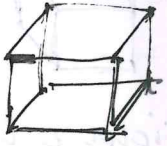
n = 4

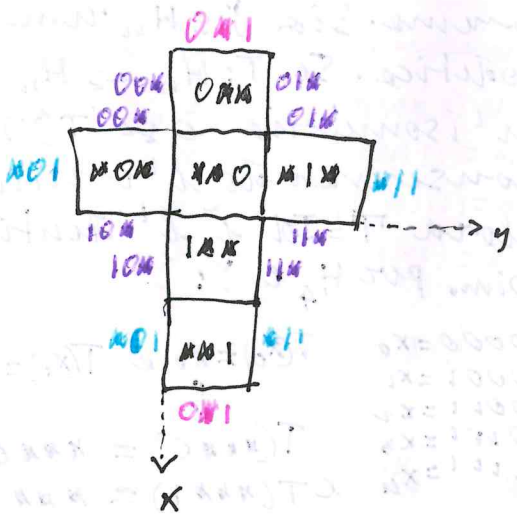


n : 1 → 2

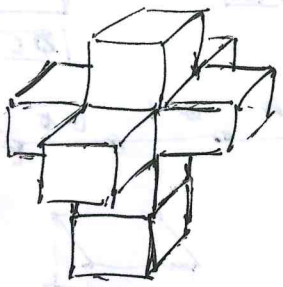


n : 2 → 3





Salvador Dalí



Geodetiche e isometrie.

Teorema. Sia $\gamma \subseteq H_n$ una geodetica. Se $T: H_n \to H_n$ è un'isometria e se $T(\gamma) = \gamma$ (conservando l'ordine) allora $T = Id$ è l'identità.

Dim. per H_4 e γ:

$0000 = x_0$
$0001 = x_1$
$0011 = x_2$
$0111 = x_3$
$1111 = x_4$

$T(x_0) = x_0$ e $T(x_1) = x_1$

$\Downarrow$

$T(****0) = ****0$
e $T(****1) = ****1$

~~Dimostrazione del teorema~~

Prendiamo $1100 = y$: $d(y, x_0) = 2$
$d(y, x_1) = 3$

Se $T(y) = ****1$, poiché $d(T(y), x_0) = 2$ deve essere del tipo:

$T(y) = 0001$ (con un 1 messo in una delle prime 3 posizioni).

Ma così $d(T(y), x_1) = 1 \neq 2$, assurdo.

Iterando il ragionamento:

$T(***0) = ***0$ $\quad T(***1) = ***1$
$T(**01) = **01$ $\quad T(**11) = **11$
$T(*011) = *011$ $\quad T(*111) = *111$
$T(0111) = 0111$ $\quad T(1111) = 1111$

A ritroso:

$T(x_2) = x_2 \Rightarrow T(1011) = 1011$;
$T|_{**11} = Id$

Iterando: $T|_{****} = Id$; $T|_{H_4} = Id$. ∎

Codice di Hamming.

Codifica $x = x_1 x_2 x_3 x_4 \in H_4$ aggiungendo 3 controlli di parità.

Versione $[8,4]$

$x_1 + x_2 + x_4 = y_1$

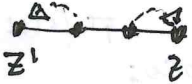
$x_1 + x_3 + x_4 = y_2$

$x_2 + x_3 + x_4 = y_3$

$x_1 + x_2 + x_3 = y_4$

Affinché il codice sia ad autocorrezione di un errore occorre che

$x' \neq x'' \Rightarrow d(x'y', x''y'') \geq 3$



Versione $[7,3]$

$x_1 + x_2 + x_4 = y_1$

$x_1 + x_3 + x_4 = y_2$

$x_2 + x_3 + x_4 = y_3$

Supp. $d(x', x'') = 1$

C'è una cifra diversa:

p.es. x_1, che cambia $y_1 y_2$

o x_2, " " $y_1 y_3$

o x_3, " " $y_2 y_3$

o x_4, " " $y_1 y_2 y_3$

Ci sono due cifre diverse:

$x_1 x_2 \longrightarrow y_2 y_3$

$x_1 x_3 \longrightarrow y_1 y_3$

$x_2 x_3 \longrightarrow y_1 y_2$

Il codice sta in $\mathbb{H}_7$
e ha 2^4 parole, ~~cioè~~
ciascuna delle quali è
al centro di una sfera
di volume $1+7=8$:

$$8 \times 2^4 = 2^7$$

il codice è ottimale.