I Introduction

why these lectures rely on this school

- information everywhere, "big data" era

  - test, sounds, image... = computer files = 0110...

  - basic problems: compressing these files, how much information in it? (no sound compression)

- transmitting through noisy channels (interference of cell phones)

formalized by Shannon, 48, A mathematical theory of communication, fundamental paper

- other fields of information treatment

  - DNA = RNA = proteins = cells = organs = organisms = all of nature, regulations (mitigation)

  - photons = retina = optical nerve = brain perception

- language, possibility of error correction

- relationship with the school

  - enlightens fundamental features of stat mech (entropy)

  - one of the fruitful interdisciplinary applications of stat mech of disordered systems (noise = disorder)

  

  \[ \rightarrow \text{relations with C. Moore} \]

  \[ \rightarrow \text{G. Eliot} \]

  \[ \rightarrow \text{H. Cohn (packing in Hamming space)} \]

  \[ \rightarrow \text{F. Rieke} \]

  \[ \rightarrow \text{F. Ricci-Tersenghi} \]

deviations from original title of the lecture, but we'll see at the end a few of "easy computation", to be completed by Federico

bibliography:

outline of the lecture

exercises on:

  www, phys. ens.fr/~guillemin/bouche.html
II The meaning of entropy

II. A. Definition

In stat mech., \( S = k \ln N \) microcanonical

\[ S = -k \sum \frac{p(b) \ln p(b)}{\text{canonical}} \quad p(b) = \frac{e^{-\beta H(b)}}{Z} \]

If \( p(b) = \frac{1}{N} \) on all configurations reduces to microcanonical

Shannon’s definition and notation:

\[ p = \{ p(x), x \in X \} \quad \text{probability law} \]

\[ H(p) = -\sum_{x \in X} p(x) \log_2 p(x) \quad \text{choice of units of} \quad k = \frac{1}{\ln 2} \]

\[ 0 \ln 0 = 0 \]

"Entropy is a measure of (the lack of) information."

\[ p(x) = \delta_{x, x_0} \quad \Rightarrow \quad H(p) = 0 \]

\[ p(x) = \frac{1}{|X|} \quad \Rightarrow \quad H(p) = \log_2 |X| \quad \text{maximal (exercise, to be done with } D(p \| q)) \]

and in general \( H(p) \in [0, \log_2 |X|] \)

\( \Rightarrow \) grows when \( p \) spreads out, i.e., when randomness \( f \), but

why this precise form? many other functions could be used

we shall see that this is the “right” def

\[ \text{if } |X| > 2, \quad h(p) = -p \log_2 p - (1-p) \log_2 (1-p) \]

\[ \frac{1}{2} + p \]
II. B. A guessing game

2 players. Alice draws \( x \) from \( X \), \( x \in X \) (prob. \( p(x) = P(X=x) \)) - outcome \( x \).

Bob does not see \( x \), asks yes/no questions to Alice to determine \( x \).

Example: \( X = \{A, B, C, D\} \)

Tree of questions:

\[
\begin{array}{c}
A \\
\downarrow \\
A \\
\downarrow \\
A \\
\downarrow \\
A \\
\downarrow \\
A \\
\downarrow \\
\downarrow \\
B \\
\downarrow \\
C \\
\downarrow \\
D
\end{array}
\]

\( p_A = p_B = p_C = p_D \), \( p_A = p_B = p_C = p_D = 1 \), \( p_A = 2 \), \( p_C = p_D = 3 \)

But Bob can also play with:

\[
\begin{array}{c}
A \\
\downarrow \\
A \\
\downarrow \\
A \\
\downarrow \\
A \\
\downarrow \\
A \\
\downarrow \\
B \\
\downarrow \\
C \\
\downarrow \\
D
\end{array}
\]

What's the best strategy? It for Bob to conclude as fast as possible, it depends on \( p(x) \), if \( A \) is very probable, \( 2 \) is better.

Call \( T \) the tree of questions, \( \bar{c}(T) \) the nb of questions to conclude \( x \) in \( T \).

\[
\bar{c}(T) = \sum_{x \in X} p(x) \bar{c}_x(T)
\]

average nb of questions asked.

\( T^* \) the choice that minimizes \( \bar{c}(T) \).

Claim (Shannon's source coding th):

\[
H(p) \leq \bar{c}(T^*) < H(p) + 1
\]

\( \Rightarrow H(p) \) really quantifies the lack of information on the output of \( p \), as the nb of binary questions to ask to determine the outcome.
Elements of the proof:

- **Kraft Inequality:** \( \forall T, \sum_{x \in X} 2^{-p_x(T)} \leq 1 \)

  - If \( |X| \) grows, some \( p_x(T) \) must be big to compensate

  **Proof:** \( \epsilon_{\text{max}} = \max_x p_x \)

  - \( \epsilon_{\text{max}} \) nodes at depth \( \epsilon_{\text{max}} \)
  - Each \( p_x \) projects a "shadow" on \( 2^{\epsilon_{\text{max}} - p_x} \)
  - Mode shadows do not intersect \( \Rightarrow \sum_{x \in X} 2^{\epsilon_{\text{max}} - p_x} \leq 2^{\epsilon_{\text{max}}} \)
  - By definition \( \Rightarrow \) answer found \( \Rightarrow \) no other \( \epsilon < \epsilon_{\text{max}} \)

Kullback–Leibler Divergence

- \( p, q \) two prob. laws on \( X \), def \( D(p \| q) = \sum_{x \in X} p(x) \log_2 \frac{p(x)}{q(x)} \)

- Properties:
  - \( D(p \| q) \geq 0 \)
  - \( D(p \| q) = -\sum_{x \in X} p(x) \log_2 \frac{q(x)}{p(x)} \)
  - \( \log_2 \left( \sum_{x \in X} p(x) \frac{q(x)}{p(x)} \right) = \log_2 (1) = 0 \)

\[ D(p \| q) = 0 \quad \Rightarrow \quad p(x) = q(x) \quad \forall x \in X \]

  - Because log strictly concave, Jensen, saturated \( \Rightarrow \theta^* = \text{int} \)

- \( \theta^* \) with \( q^* = \frac{p(x)}{|X|} \)

\[ D(p \| q^*) = -H(p) + \log_2 (|X|) \]

- \( H(p) \leq \log_2 |X| \), equally, \( -\theta \leq q \cdot p \cdot q \)

**Proof of Jensen:** \( \mathbb{E} \left[ f(X) \right] \leq f \left( \mathbb{E}[X] \right) \) for \( f \) concave

\[ X = \mathbb{E}[X] + (X - \mathbb{E}[X]) \]

\[ f(a + b) \leq f(a) + (b-a) \lambda \]

\[ f(X) \leq f(\mathbb{E}[X]) + (X - \mathbb{E}[X]) \lambda \]

\[ \mathbb{E}[f(X)] \leq f(\mathbb{E}[X]) + \lambda \]
Putting together the two

\[ q(x) = \sum \frac{1}{2} 2^{-x(t)} \]  

for a valid tree, \( \gamma = \sum_{x \in X} 2^{-x(t)} \)

\[ D(p||q) = \sum_{x} p(x) \log_{2} \left( \frac{\frac{p(x)}{\gamma}}{\frac{1}{2} 2^{-x(t)}} \right) = -H(p) - \log_{2}(\gamma) + \varepsilon(t) \]

Kraft \( \Rightarrow \gamma \leq 1 \) \( \log_{2}(\gamma) \leq 0 \)

\[ \varepsilon(t) - H(p) = D(p||q) - \log_{2}(\gamma) \geq 0 \]

\( \Rightarrow H(p) \leq \varepsilon(t) \quad \forall \text{valid } t \)

This proves the lower bound of Shannon's ch

Upper bound: one can exhibit a tree which achieves it:

1. If \( \{ x \} \) is a set of integers satisfying Kraft, then there exists a tree with \( x(t) = x \) such that the \( x_k \) in sequence, take the first node in lexicographic order of depth \( t \), then remove what is below and continue.

2. Let \( \bar{e}_x = \left( \log_{2} \frac{1}{p(x)} \right) \geq -\log_{2} p(x) \), \( \sum_{x} 2^{-e_x} \leq \sum_{x} p(x) = 1 \) \( \Rightarrow \) Kraft,

\[ \sum_{x} p(x) e_x \leq \sum_{x} p(x) \left( \log_{2} \frac{1}{p(x)} + 1 \right) = H(p) + 1 \]

Huffman coding to find the optimal choice of \( \{ e_x \} \) given \( p(x) \): group the two least likely symbols to make a single one, and iterate.
II.C. Data compression

Back to original problem: \( x_1 x_2 \ldots x_n \rightarrow 01001\ldots 01 \)

We have actually solved this problem.

Tree:

- Left branch \( \rightarrow 0 \)
- Right branch \( \rightarrow 1 \)

Each symbol \( x \) associated to a string \( w_x(T) \) of 0, 1, of length \( l_x(T) \)

\( x_1 \ldots x_n \rightarrow w_{x_1}(T) w_{x_2}(T) \ldots w_{x_n}(T) \)

The string is uniquely decodable (prefix-free): just follow the game.

\[
\begin{array}{cccc}
0 & 1 & 0 & 110 \\
A & B & A & C \\
\end{array}
\]

If the \( x_i \) are iid with law \( p(x) \), the length of the total sequence will be \( n \bar{c}(T) \)

\[ H(p) \text{ is the average number of bits (within 1) per symbol necessary to compress a sequence of symbols generated by a source of proba } p \]

\( H(p) \): we gzip to measure the entropy of English.

- not iid but short correlation
- this is for loss-less compression
- in image/sound compression some mistakes are tolerable, compromise between accuracy of reconstruction and rate of compression: rate-distortion theory
II. D. Mutual information

useful for the following and for other lectures (Fleet)

\((x, y)\) a pair of (a priori correlated) r.v. on \((X \times X') \) not necessarily the same

\(p_{x,y}(x,y) = P[x=x \text{ and } y=y] \) joint law

\[ p_x(x) = \sum_{y \in X} p_{x,y}(x,y) \] marginal law

\[ p_y(y) = \sum_{x \in X} p_{x,y}(x,y) \]

\[ p_{x|y}(x|y) = \frac{p_{x,y}(x,y)}{p_y(y)} \] conditional law, \(\sum_x p_{x|y}(x|y) = 1\)

\(H(x) = -\sum_x p_x(x) \log_2 p_x(x)\)

\(H(y) = -\sum_y p_y(y) \log_2 p_y(y)\)

\[H(x|y) = \sum_y \frac{p_{x|y}(x|y)}{p_y(y)} \log_2 \frac{p_{x|y}(x|y)}{p_x(x)}\]

\[H(x;y) = D(p_{x,y} \parallel p_x p_y) = \sum_{x,y} p_{x,y}(x,y) \log_2 \frac{p_{x,y}(x,y)}{p_x(x) p_y(y)}\]

**Properties:**

- \(I \geq 0\) (we have seen this A D)
- \(I = 0 \iff p_{x,y} = p_x p_y \iff X \text{ and } Y \text{ independent}\)
- \(I(x; y) = \sum_{x,y} p_{x,y}(x,y) \log_2 \left( \frac{p_{x,y}(x,y)}{p_x(x)} \right) = H(x) - H(x|y)\)
- \(= H(y) - H(y|x)\)
- \(= H(x) + H(y) - H(x,y)\)
- Hence, \(H(x|y) \leq H(x)\) conditioning reduces entropy
- \(I(x; y)\) measures how much you know (in bits) about one of the two r.v. if the other is revealed to you
- also, \(H(x,y) = H(x) + H(y)\)
additional remarks/ exercises:

- proof that canonical Gibbs-Boltzmann is the one maximizing entropy under constraint on average energy

- proof that uniform law is the one with maximal entropy

- in Shannon’s paper, proof that if one assumes

  1. $H(p_1, \ldots, p_n)$ continuous

  2. $H\left( \frac{p_1}{p_n}, \ldots, 1/p_n \right)$ $\propto$ with $H$

  3. $H$ "additive" under decomposition

  then only possibility is $-\sum_x p_i \ln p_i$, within a multiplicative constant
III. Communication over noisy channels

III.A. Definitions

- Binary Erasure Channel (BEC)

```
input   output
x        y
```

Examples:

- **Binary Erasure Channel (BEC)**
  
  \[ P(x) = \begin{cases} 1 - \varepsilon & \text{if } x = 0 \\ \varepsilon & \text{if } x = 1 \end{cases} \]

- **Binary Symmetric Channel (BSC)**

\[ P(x) = \begin{cases} 1 - p & \text{if } x = 0 \\ p & \text{if } x = 1 \end{cases} \]

Capacity of a channel:

\[ C = \max_{P_X} I(X;Y) \]

Exercises:

- Compute \( C \) for BEC and BSC, show that:
  
  \[ C = 1 - \varepsilon \] for the BEC

  \[ C = 1 - h(p) \] for the BSC, \( h(p) = -p \log_2 p - (1-p) \log_2 (1-p) \)

...
proof of the capacity

- $\text{BEC}$
  \[ p_X \text{ uniformly distributed over } \{1, \ldots, n\} \]
  \[ I(X; Y) = H(Y) - H(Y|X) \]
  \[ \overset{\text{indep. } Y\text{ on } X}{\sim} \]
  \[ H(p) \]
  \[ Y = 0 \quad \text{with proba } \ a \ (1-p) + (1-a)p \]
  \[ \implies H(Y) = \mathbb{1}(a (1-p) + (1-a)p) \]
  \[ \text{maximized with } a = \frac{1}{2} \quad \text{if } H(Y) = \mathbb{1}(\frac{1}{2}) = 1 \]
  \[ C = 1 - h(p) \]

- $\text{BEC}$
  \[ H(Y|X) = \mathbb{1}(E) \]
  \[ \text{indep. on } a \quad \text{within } \{0, 1\} \quad \text{in } \{1, 1\} \]

  \[ H(Y) = \begin{cases} 0 & \text{with proba } 1(1-E) \\ 1 & (1-a)(1-E) \\ E & \text{rest} \end{cases} \]

  \[ H(Y) = -E \ log E - a(1-E) \ log_a(a(1-E)) - (1-a)(1-E) \ log((1-a)(1-E)) \]

  \[ = -E \ log E - (1-E) \ log(1-E) - (1-a) \ log(a) - (1-a) \ log((1-a)) \]

  \[ \implies I(X; Y) = (1-E) \ \mathbb{1}(E), \quad \text{max in } a = \frac{1}{2} \quad \text{(again obvious)} \]
  \[ \implies C = 1 - E \]
Encoding and decoding

- Noise of the channel degrades information → fight it by transmitting more information, i.e., add redundancy

\[ \text{message} \rightarrow \text{encoded message} \rightarrow \text{channel} \rightarrow \text{corrupted encoded message} \rightarrow \text{decoder} \rightarrow \text{decoded message} \]

- Senders and receivers agree on the encoding and decoding procedure beforehand

\[
\text{rate of a code} = \frac{\text{no. of bits of message}}{\text{no. of bits of encoded message}} < 1, \quad \text{measure of redundancy should be as big as possible}
\]

### III.8 Naïve coding

- Simplest way to be redundant: repeat oneself

**Encoding:**
- 0 → 000
- 1 → 111

**Decoding:**
- 000 → 0
- 001, 100, 010 → 0
- 011, 101, 110 → 1
- 111 → 1

- On the BSC, with \( p < \frac{1}{2} \)

**Decoding:**

- \( 000 \rightarrow 0 \)
- \( 001, 010, 100 \rightarrow 0 \)
- \( 011, 101, 110 \rightarrow 1 \)
- \( 111 \rightarrow 1 \)

- Majority rule (better take an odd number of repetitions to avoid it)

**We want to transmit 0 or 1 with the same prob.**

- Without repetition, \( P(w) = p \) → just take output as a guess
- With repetition, \( P(w) = 3p^2(1-p) + p^3 \) → two or three flips induce a mistake and can be corrected

\[
\begin{align*}
P(w) & \sim \begin{cases} 1, & p = \frac{1}{2} \\ 0, & p \neq \frac{1}{2} \end{cases} \\
\text{better, but: } & R = \frac{1}{3} \\
\text{to have } & P(w) \rightarrow 0 \text{ repeat } 5,7, \ldots \text{ times, but then } R \rightarrow 0
\end{align*}
\]

- Can one do better?
III, C. Shannon channel coding theorem

quite surprisingly, answer of Shannon then is yes:

\[ \exists \text{ codes (of growing size)} \text{ with } \text{Pen} \to 0 \text{ for any } R < C \]

* C is the ultimate limit for the rate
* yet it can be approached arbitrarily close
* statement true in the thermodynamic limit

more formal definitions

encoding: \( x = \text{encoding} (z) \) injective

i.e. we choose \( 2^L \) "codewords" \( \mathcal{C} \subseteq \{0,1\}^N \) \( \alpha = 1, 2, \ldots, 2^L \) to represent the possible messages

intuitively, if \( z \) is transmitted, \( x = z \) has some because of the noise,

\( \hat{x} \): estimation of the sent \( z \) given the received \( y \)

then \( \hat{x} = \text{decoding} (\hat{z}) \)

assuming the \( 2^L \) \( z \) are equiprobable, then the \( 2^L \) \( \hat{z} \) are equiprobable,

we can forget about \( z \) and \( \hat{z} \): the problem is thus:

\( x \in \mathcal{C} \) a.a.r. \( \Rightarrow \) computed \( \Rightarrow \) reconstruct \( z \)
decoding as an inference problem (for one given codeword)

X random codeword, \( p_X(x) = \frac{1}{2^L} 1(x \in \mathcal{C}) \) indicator function

Y random output \( p_{Y|X}(y|x) = \prod_{i=1}^{N} q(y_i | x_i) \)

with \( q \) describing the channel for BEC

\[ q(0|0) = 1 - E \]
\[ q(1|0) = E \]
\[ q(1|1) = 1 - E \]
\[ q(0|1) = E \]
\[ q(1|1) = 0 \]

Given \( y \) we have to guess \( x \)

Bayes theorem:

\[ p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x) p_X(x)}{p_Y(y)} = \frac{1}{Z(y)} \prod_{i=1}^{N} q(y_i | x_i) \]

as there is noise we cannot a priori do better than assigning probabilities to the various possible input code words

\( \hat{x}(y) \) depends on the measure of error one wants to minimize

\[ \hat{x}(y) = \arg \max_{x} p_{X|Y}(x|y) \] minimizes \( P[\hat{x} \neq x] \)

Block Maximal A Posteriori (MAP) decoding

\( \hat{x} \) Maximum Likelihood (ML) decoding

\[ \hat{x}_L(y) = \arg \max_{x_L} p_{X|Y}(x_L|y) \] minimizes the number of decoded bits in marginal law of

\[ p_{X|Y} \]

more on that in Kazahara's lectures

for the BEC:

\[ y = (0, 1, *, 1, *, 0, *, 0) \]

those are free

\[ p_{X|Y}(x|y) \] uniform over those code words which agree with \( y \) on the uncoded place

\[ B(y) = \{ x \in \mathcal{C} : \min_j \{ 0, 1, 0 \} \} \]

\[ = \frac{1}{|\mathcal{C} \cap B(y)|} \]

\[ \delta(z \in \mathcal{C} \cap B(y)) \]
Shannon Random Code Ensemble

Definition of REM (Birkho's lectures), also in the treatment
\[ Y = \left\{ \frac{Z^{(x)}}{Z} \right\} \text{ for } x = 1, \ldots, L \text{ obtained by taking } Z^{(x)} = 1 \text{ probe half independently } \forall x, \alpha \]

very simple, seems crazy, does not emerge that the \( z^{(x)} \) are all different but
in the limit \( N, L \to \infty \) very few collisions

on the BEC: suppose w.h.p. that \( z^{(1)} \) has been transmitted
\[ N_x(y) \text{ : no of erased bits in } y, \text{ i.i.d. } \text{ Bin}(N_E) \text{ or } \text{ NE in the thermodynamic limit} \]

How many codewords \( \neq z^{(1)} \) in \( B(y) \)?
\[ \text{w.p. } X^o = \text{ Bin} \left( 2^{L-1}, \left( \frac{1}{2} \right)^{N-N_x(y)} \right) \]
\[ \text{prob that } z^{(1)} \text{ agrees with } z^{(1)} \text{ on the non erased bits} \]
\[ \mathbb{E} \left[ X^o \right] \approx 2^{L-N+EN} = 2^{N(R-1-E)} \quad \text{if } R < 1 - E = C_{N \to \infty} \]

=> as in the REM, a first moment method \( X^o = 0 \) w.h.p.
\[ \Rightarrow \text{ one can recover } z^{(1)} \text{ as the only codeword compatible with the received bits in } y \]

=> proves Shannon's channel coding theorem for the BEC,

one can decode with vanishing error probability with rates up to capacity

proof for the BSC no homework, text and solution (in french) on the webpage

\[ \Delta \text{ notation } \triangleq \text{ invested} \]

very simple random code achieves capacity

BUT: needs exponentially large (in \( N \)) memory to store the codewords

needs exponential time to decode

\[ \Rightarrow \text{ needs to add structure to the codebook, can this be done and still achieve capacity?} \]

\[ \text{to be seen in the following} \]

With respect to erasures and to the code

where we have average error probability \( \to 0 \), can be treated to maximal (over the codewords) error probability \( \to 0 \), hence \( \exists \) one code with max error prob \( \to 0 \)

by approximating the worst half codewords, does not change the rate

\[ \text{converse with Fano, } P_u > 0 \text{ if } R > C \text{, actually } P_u \to 0 \text{ in } N \]
IV. A linear codes

need to add some structure to the codebook. \( G = \{ x^{(k)} \} \subset \{ 0,1 \}^N \) is made decoding Easy

\( \{ 0,1 \}^N \) is a linear space over \( \{ 0,1 \} = \mathbb{Z}_2 \), with:

- \( 0 + 0 = 0 \)
- \( 0 + 1 = 1 \)
- \( 1 + 1 = 0 \)

addition mod 2

\( G \) is a linear code if it is a linear subspace of \( \{ 0,1 \}^N \)

\[ \text{for } \{ x,y \} \in G \implies x+y = (x_1+y_1, \ldots, x_N+y_N) \in G \]

linear algebra on \( \{ 0,1 \} \) instead of \( \mathbb{R} \), works pretty the same

- \( \dim G = \text{nb of linearly independent code words} = |G| = 2^d \)

- \( G \) can be specified as \( \ker H = \{ x \in \{ 0,1 \}^N \mid Hx = 0 \} \)

\[ H = \begin{pmatrix} \mathbb{M} \end{pmatrix} \]

or as \( \text{Im } G = \{ x : Gx \in \{ 0,1 \}^M \} \), \( G = \{ Gx : x \in \{ 0,1 \}^N \} \)

- \( H \) parity check matrix, one line of \( H \) of \( x \) of the form \( x_1^2 + x_2 + x_3 + x_4 = 0 \)

- parity of the number of 1's must be even

if these equations are linearly independent, \( |G| = 2^d \)

\[ \dim \ker H = N - \text{rank H} \]

- always, independence \( \iff \) full rank

- codes of the code \( R = \frac{N-M}{N} = 1 - \frac{M}{N} \)

(assuming equations independent, in particular \( M \leq N \))

\( G \) generator matrix, allows to encode the codewords very easily

---

3) Tanner graph representation of parity checks:

- n: \( n \) nodes
- m: \( m \) edges

\[ \mathbb{P} \times (a) = \Delta \left( \sum_{i=1}^{n} H_{ai} x_i = 0 \right) \]

edge between \( a \) and \( i \) \( \iff \) \( H_{ai} = 1 \)

\[ \mathbb{D}_a = \{ i : H_{ai} = 1 \} \quad \mathbb{D}_i = \{ a : H_{ai} = 1 \} \]

neighbourhood \( \mathbb{D}_a \) of the nodes

all codewords are equivalent in a linear code, \( (0 \ldots 0) \) is always a codeword

\[ R = (a) \]

\[ \sum_{i=1}^{n} x_i = 0 \mod 2 \]

\[ \prod_{i=1}^{n} x_i = 1 \]

---

x: exclusive OR of \{True, False\} \( \iff \text{known as a CSP as XOR gate} \)

mention \( z \); for \( n \) variables

\[ z: \text{exclusive OR of } z \text{ of all variables} \]

\[ \sum_{i=1}^{n} x_i = 0 \mod 2 \]

\[ \prod_{i=1}^{n} x_i = 1 \]

also p-spin interactions

---

= lecture of Bode, Ribeiro, Terazaki, Maseu...
IV.2. Definition of the simple ensemble

given the success of Shannon's random code, try a random $H$
but "few density" : a few $1's$ by row and column of $H \to$ few operations, faster

take $H$ such that: $|\partial a| = k \forall a$ i.e. $k$ $1's$ in each column
with $\ell, k$ finite in the thermodynamic limit

$\Rightarrow N = M \ell k$, total number of $1's$ in the matrix $R = 1 - \ell k$

how to do this in practice:

\[
\begin{array}{c}
\text{random matching (permutation)} \\
\text{of those } N \ell = M \ell k
\end{array}
\]

half edges

there can be pb if:

\[
\sum a_i \text{ number of parallel edges } \equiv 0 \mod 2 \forall a_i
\]

defect really matters, will be rare in the thermodynamic limit if $M \ll N$
the equations are linearly independent (i.e. successive $a_i$ are distinct)

\[\text{crucial property: locally tree-like in the thermodynamic limit}\]

no short loops (with high probability)

\[\text{intuitive explanation: exploring from one vertex, } \]
choose finite $N$ of neighbors out of an extensive one, proven to choose twice in the same finite set is $O(1/N)$

△ not the tree of the guessing game
Assume codeword \((0, \ldots, 0)\) has been transmitted over BEC(\(E\))

\(\rightarrow\) because all codewords are equivalent, but we must pretend not to know that it is \((0, \ldots, 0)\)

received \(y = (0, *, 0, 0, \ldots, *)\)

sender and receiver have agreed on an LDPC

what can the receiver do?

the first * can be set to 0 with the first parity check

then the second * as well

in all checks where there is a single * \(\rightarrow\) set it to 0

continue iteratively

when it stops, either no * remains \(\Rightarrow\) perfect decoding

or at least two * in the checks that remain

polynomial time algorithm

up to which value of \(E\) will the first situation occur?

Formulation of the algorithm as message passing (if receiver's estimate \(u \in \{0, *, O\}\)

\(u_{a \rightarrow i} = 0 \text{ if } b_{j \rightarrow a} = 0 \quad \forall j \in \partial a \setminus i\)

\(= * \text{ otherwise}\)

\(= 0 \text{ otherwise}\)

\(b_i = \{0\} \text{ if all } i \in v \quad \forall v \in \partial i\)

\(= 0 \text{ otherwise}\)

in discrete time \(u[t] = * \text{ for all } a \rightarrow i\)

\(\hat{b}[t] = \{0\} \quad \forall v \in \partial i\)

\(\hat{b}[t] = \{0\} \quad \forall v \in \partial i\)

\(\hat{b}[t] = \{0\} \quad \forall v \in \partial i\)
monotonicity messages switch from $\ast$ to $0$, at the end some final state for the $k_i$ thus the original decoder, for any update schedule

Probabilistic analysis

$\mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0)$ with respect to the choice of the $H$

$\mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0)$ with respect to the choice of the channel

$\mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0)$ with respect to the choice of an edge uniformly at random

$\mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0) = 0$

$\mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0) = (\mathcal{Z}_{\mathcal{X}^{(t-1)}})^{L_{\mathcal{X}^{(t)}}}$

$\mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0) = 1 - \varepsilon (1 - \mathcal{Z}_{\mathcal{X}^{(t-1)}})^{L_{\mathcal{X}^{(t)}}}$

properties of $f$

$f(0) = 1 - \varepsilon$

$f(1) = 1$

E small $\mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0) = \mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0)$

E big $\mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0) = \mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0)$

Plot of the iteration plateau in $\mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0)$ exponent $1/2$, similarly with $C(t)$

$P \left[ L_{\mathcal{X}^{(t-1)}} = \ast \right] = \varepsilon \left( 1 - \mathcal{Z}_{\mathcal{X}^{(t)}}(L_{\mathcal{X}^{(t)}} = 0) \right)^{L_{\mathcal{X}^{(t)}}}$
for $E < E_{BP}$, perfect decoding, in linear time, for a positive rate $\Rightarrow$ non trivial results

numerical values

<table>
<thead>
<tr>
<th>$t$</th>
<th>$k$</th>
<th>$E_{BP}$</th>
<th>$E_{MAP}$</th>
<th>$R$</th>
<th>$E_{Sk}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>0,647</td>
<td>0,946</td>
<td>1/4</td>
<td>0,25</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>0,518</td>
<td>0,591</td>
<td>2/5</td>
<td>0,4</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>0,423</td>
<td>0,488</td>
<td>1/2</td>
<td>0,5</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0,506</td>
<td>0,665</td>
<td>1/3</td>
<td>0,333</td>
</tr>
</tbody>
</table>

$E_{Sk} = 1 - R$ the maximal level of noise a code of this rate could correct

$E_{BP} < E_{Sk}$ in this because of the code?

or because of the algorithm to decode?

in other words, for $E > E_{BP}$, since $E_{BP}$ has inferred all possible implications, one has a linear system of equations on $N'$ variables, $M'$ equations, has this system, a single solution? then perfect decoding would still be possible by exhaustive algorithm (or some elimination here)

... or more than one solution, then perfect decoding is not possible, whatever the algorithm...
\[ N' = N \left( 1 - \frac{k}{k+1} \right)^e \]

\[ M' = M \left( 1 - \frac{k}{k+1} \right)^{(1-e)} \]

with \( k = 3^\infty \), the number of undecorated variables.

At least two
in the invariant
magnets
are not to be found
independently

assuming the equations are independent,

\[ m_b \text{ of solution} = \frac{N' - M'}{N} = 2 \]

with \( P(\varepsilon, q) = \varepsilon \left( 1 - \frac{k}{k+1} \right)^e - \frac{\varepsilon}{k} \left( 1 - \frac{k}{k+1} \frac{k}{k+1} \right) \)

\[ \frac{d}{dq} P = \varepsilon k' \left( 1 - \frac{k}{k+1} \right)^e + \varepsilon k' \left( 1 - \frac{k}{k+1} \right) \left( 1 - \frac{k}{k+1} \right) \]

\[ = \varepsilon k' \left[ - \left( 1 - \frac{k}{k+1} \right)^e + 1 - \frac{k}{k+1} \right] \]

= 0 when \( \varepsilon \) is a fixed point of the iterations

\[ \varepsilon < \varepsilon_{bp} \]

\[ \varepsilon = \varepsilon_{bp} \]

\[ \varepsilon \in [\varepsilon_{bp}, \varepsilon_{map}] \]

\[ \varepsilon = \varepsilon_{map} \]

Use Frisch-Bose potential (with - sign, \( 1 - \frac{k}{k+1} \))

plats shown
by Florent
are not exactly the same places,
\( m_c \neq m_b \)
in terms of magnetization

\[ \varepsilon > \varepsilon_{map} \]

computational gap, as in many influence pb (Flourent)

\( \varepsilon \) = \( \frac{T}{T_c} \), but not completely
IV. D. More generic LDPC ensembles

Instead of fixed degrees \((e, k)\) distribution of degrees

\(\lambda_e\) function of variables that have degree \(e\)

\(p_k\) checks

\(H\) uniform under this constraint

\(<e> = \frac{\sum e \lambda_e}{\sum \lambda_e}, \quad <k> = \frac{\sum k \lambda_k}{\sum \lambda_k}\)

\(N < e > = M < k >\) to match

Still locally tree like

if one draws at random an edge, ends up on \(e\) with degree \(k+1\)

with prob \(\lambda_e = \frac{(e+1)\lambda_{e+1}}{<e>}\)

with prob \(\lambda_k = \frac{k\lambda_{k+1}}{<k>}\)

\(\tilde{\lambda}(z) = \sum e \lambda_e z^e, \quad \tilde{\phi}(z) = \sum k \lambda_k z^k\)

\(\tilde{\lambda}(z) = \frac{\tilde{\phi}(z)}{\lambda(<e>)}\)

\(\varphi(e, z) = E \lambda \left( 1 - \tilde{\phi}(z) \right) - \frac{<e>}{<k>} \left( 1 - \frac{\tilde{\phi}(z)}{\lambda(<e>)} \right)\)

Then, exercise:

\(\varphi(0, z) = \frac{\tilde{\phi}(z)}{1 - \tilde{\phi}(z)}\)

\(\varphi(0, z) = 1 - \frac{\tilde{\phi}(z)}{1 - \tilde{\phi}(z)}\)

Some reasoning as before, just add the proba of finding a vertex of a certain degree

generalizes the case with fixed \(k, e\)

one can find choices of \(\{\lambda_e, \lambda_k\}\) such that \(\epsilon_{BP}\) is arbitrarily close to \(\epsilon_{BH}\)

seems great but

not obvious that they are the best in

practice, depends on the finite \(N\) behavior,

and prefactors in the linear complexity
diverges when \(\epsilon_{BP} < \epsilon_{BH}\) (max. degree \(k\))
what we have done is a very specific (and simple) exemple of the reasoning of the cavity method: Let's abstract the main points, besides the IT context:

- variables \( x = (x_1, ..., x_N) \in \mathbb{R}^N \) (here the encoded message to be recovered)
- quenched randomness \( \mathbb{F} \) (here the graph/parity check matrix + realization of the noise)
- defining a factor graph with \( M \) interactions

\[
\mu(x; \mathbb{F}) = \frac{1}{Z(\mathbb{F})} \prod_{i=1}^{N} \pi_i(x_i; \mathbb{F}) \prod_{a=1}^{M} \pi_a(x_{2a-1}, x_{2a})
\]

- probability law \( \mu(x; \mathbb{F}) \)
- constraints on the channel codewords
- goal of the cavity method:
  - compute \( \Phi = \lim_{N \to \infty} -\frac{1}{N} \mathbb{E}\left[\ln Z(\mathbb{F})\right] \) quenched free entropy density
  - marginal laws \( \mu(x_i; \mathbb{F}) \) "magnetization"

strategy: tree like factor graph
- trees exactly solved by message passing (here the messages were \( \frac{1}{N} \) in general non-trivial function of \( x \))
- long loops \( \rightarrow \) "inherent" \( \rightarrow \) RS cavity method
- induce long range correlations \( \rightarrow \) RSB

\( \Delta \) LDPC - planted models, \( \rightarrow \) the replica line, don't take as completely generic the phenomenon / transitions found here

\( \Rightarrow \) of Fortuin's lectures

- Screw & G
- Richardson & Urbanke
- Malyshkin & Skhulkiki - Spectra
- Kschischang - Scale \& 2000
- Montanari = 2000