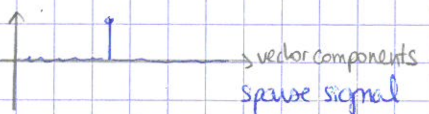
goo.gl/VdGVYf old lectures notes
goo.gl/PLJNoR exercises -

LECTURE I : DENOISING & STATISTICS

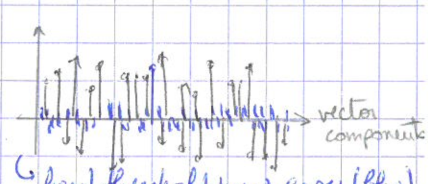
I a a simple example.

A denoising toy problem:



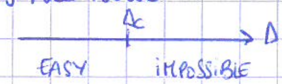
sparse signal

but observe



↳ hard thresholding → easy (blue)
↳ signal lost in the noise (black)

It seems that decreasing slowly the noise we go through a threshold value of the noise below which the denoising becomes easy after having been impossible



Formalisation : signal of length $\mathcal{N}$

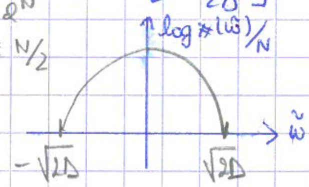
How components will have values $\in [\omega, \omega + \Delta\omega]$ → on avg. $\mathcal{N} \times \dfrac{e^{-\omega^2/2\Delta}}{\sqrt{2\pi\Delta}} \Delta\omega$

$\propto \exp\left[\ln N - \dfrac{\omega^2}{2\Delta}\right]$

$\tilde{\omega} = \dfrac{\omega}{\sqrt{\ln \mathcal{N}}}$ $\exp \ln N \left[1 - \dfrac{\tilde{\omega}^2}{2\Delta}\right]$

if $|\tilde{\omega}| > \sqrt{2\Delta}$: # of such components $\xrightarrow{\mathcal{N} \to +\infty} 0$ $\equiv$ REM $\begin{cases} \mathcal{N} = 2^N \\ \Delta = N/2 \end{cases}$

↳ in the thermodynamic limit, we will typically observe no component generated randomly with magnitude larger than $\sqrt{\ln \mathcal{N}\, 2\Delta}$



↳ $\boxed{\Delta_c = \dfrac{1}{2\ln \mathcal{N}}}$ → seems to be matching the little experiments limit on detectability!

(Donoho Johnston 1998 - universal denoiser threshold).

Ib denoising of a single variable

Consider a random variable X distributed according to $P_X(x)$. We want to transmit it through a noisy channel. Send ground truth $X^* \sim P_X(x^*)$

- Receive $Y = \sqrt{\lambda} X^* + Z$ $\begin{cases} \lambda = \text{signal to noise ration (SNR)} \\ Z \sim \mathcal{N}(0, I) \end{cases}$

AWGN: Additive white Gaussian noise -

first case: receiving many measurements: $\begin{cases} y_1 = \sqrt{\lambda} x^* + z_1 \\ y_2 = \sqrt{\lambda} x^* + z_2 \\ \vdots \\ y_N = \sqrt{\lambda} x^* + z_N \end{cases} \Rightarrow \vec{y}$

we implicitely defined the likelihood ≡ conditional probability: $P_{Y|X}(y|x) = \dfrac{e^{-\frac{(\sqrt{\lambda}x - y)^2}{2}}}{\sqrt{2\pi}}$

$\Rightarrow P_{Y|X}(\vec{y}|x) = \prod_{i=1}^{N} P_{Y|X}(y_i|x) = \dfrac{1}{(2\pi)^{N/2}} e^{-\sum_i (\sqrt{\lambda}x - y_i)/2}$ → proba on Y and not on X!

↳ maximum likelihood estimator: $\hat{x}(\vec{y}) = \underset{x}{\operatorname{argmax}}\ P_{Y|X}(\vec{y}|x) = \underset{x}{\operatorname{argmax}} \log P_{Y|X}(\vec{y}|x)$

↳ in our case: $= \underset{x}{\operatorname{argmax}} \sum_i - \dfrac{(\sqrt{\lambda}x - y_i)^2}{2} + \text{cste}$

$= \dfrac{1}{N\sqrt{\lambda}} \sum_i y_i \xrightarrow[N \to +\infty]{} x^*$ (exercise).

The maximum likelihood scheme always works when $N \to +\infty$. (Laplace 1810, Fisher 1912 ..)

second case: receiving only one measurement:

we'll need Bayes theorem: $P(B|A) = \dfrac{P(A|B)P(B)}{P(A)}$ | Laplace 1774, Bayes, price 1763 -

in our problem:

posterior distribution: $P(X|Y) = \dfrac{\overbrace{P(Y|X)}^{\text{likelihood}}\, P(X)}{P(Y)}$ ← prior distribution — The tricky part: most of the time you have no idea how data was generated...

$P(Y)$ ← normalization, "evidence"

$$\hookrightarrow P_{X|Y}(x|y) = \frac{1}{Z(y)} e^{-\frac{(\sqrt{\lambda}x - y)^2}{2\Delta}} P(x) = \frac{1}{Z(y)} \exp\left(-\log p_{Y|X}(y|x) - \log P_X(x)\right) \equiv \frac{e^{-\mathcal{H}(x,y)}}{Z(y)}$$

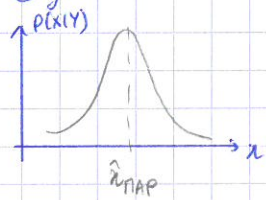
with $\begin{cases} \mathcal{H}(x,y) = \log p_{Y|X}(y|x) + \log P_X(x) \text{ -} \\ Z(y) = \int dx\, e^{-\mathcal{H}(x,y)} \text{ -} \end{cases}$ → and we made the stat mech connection - most probable ≡ ground state.

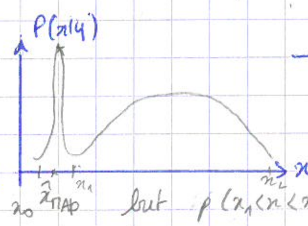
Now from the single measurement the most probable value of $\boxed{\hat{x}_{map} = \underset{x}{\arg\max}\, P_{X|Y}(x|y)}$

↳ MCMC, with decreasing temperature to find the ground state :)

Nevertheless, there is a subtlety here:

imagine the two scenari:





→ we need a principled way to take probability mass into account.

but $p(x_1 < x < x_2 | y)$ might $> p(x_0 < x < x_1 | y)$

<u>minimum mean squared error</u>: $\hat{x}_{MMSE}(y)$ such that $(\hat{x}_{MMSE} - x^*)^2$ be minimum -

Consider the <u>posterior risk</u>: $R = \int P(X|Y)(\hat{x} - x)^2 dx$ <u>average value</u> of squared error.

Minimize it: $\left.\dfrac{\partial R}{\partial \hat{x}}\right|_{\hat{x}_{MMSE}} = 0 \Rightarrow \boxed{\hat{x}_{MMSE} = \int dx\, x\, P_{X|Y}(x|y)}$

↳ think x = spin under Gibbs distribution - Good estimator = magnetization! $\hat{x}_{MMSE} = \langle x \rangle$

<u>Rk</u>: we could consider other risks, other moments of the gap.
The best strategy is problem dependent ($|x - \hat{x}| \Rightarrow \hat{x}$ = median).

Yet again, you might prefer to minimize the number of errors: $1 - \delta_{\hat{x}, x^*}$.
The associated risk is: $R = \int P(X|Y)(1 - \delta_{\hat{x}, x^*})\, dx \Rightarrow \boxed{\hat{x} = \underset{x}{\arg\max}\, P(x|y)}$

<u>Bayes optimal error</u>

difference with MAP for multi components: $\hat{x}_i = \underset{x_i}{\arg\max}\, P(x_i|y)$ BAYES OPTIMAL ERROR $\neq$ MAP $\vec{x}_{MAP} = \underset{\vec{x}}{\arg\max}\, P(\vec{x}|y)$

back to our specific problem: Redefining conviniently the partition sum $Z(y)$:

$$P(x|y) = \frac{1}{Z(y)} \exp\left[-\frac{\lambda x^2}{2} + y\sqrt{\lambda}x\right] P_X(x) \text{ -}$$

i/ the minimal mean squared error $MMSE = \underbrace{\mathbb{E}_{Y,X^*}}_{\text{"disorder average"}} \left[(\underbrace{\hat{x}_{MMSE}}_{\equiv\langle x\rangle} - x^*)^2\right]$

$$MMSE = q + q_0 - 2m \quad \text{with} \begin{cases} q = \mathbb{E}_{Y,X^*}(\langle x \rangle^2) \\ m = \mathbb{E}_{Y,X^*}(\langle x \rangle x^*) \\ q_0 = \mathbb{E}_{Y,X^*}(x^{*2}) \end{cases}$$

↳ interpretation in terms of replicas: OVERLAP OF 2 replicas $q = \mathbb{E}_{Y,X^*}\left(\langle x^{(1)} \rangle \langle x^{(2)} \rangle\right)$

ii/ <u>NISHIMORI RELATIONS</u>: $\mathbb{E}_{Y,X^*}\left[\langle f(x, x^*) \rangle\right] = \mathbb{E}_Y\left[\langle f(x^{(1)}, x^{(2)}) \rangle\right]$

<u>csq</u>: $m = q \Rightarrow MMSE = q_0 - m$.

<u>proof</u>: $\int dy\, dx^*\, P(y, x^*) \int dx\, P(x|y) f(x, x^*) = \int dy\, P(y) \int dx\, dx^*\, P(x^*|y) P(x|y) f(x, x^*)$

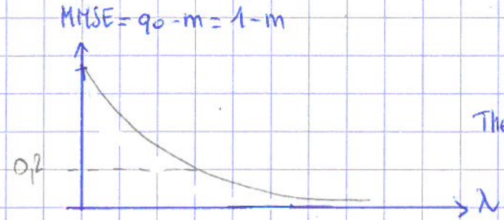
$= \int dy\, P(y) \int dx_1\, dx_2 \ldots$

11/07/2017

$$\mathcal{D}u = du\,\frac{e^{-u^2/2}}{\sqrt{2\pi}}$$

iii/ Assume we have a binary signal: $P_x(x) = \pm 1$ with proba $1/2$. $\Rightarrow \begin{cases} m = \int \mathcal{D}u \tanh(\lambda + \sqrt{\lambda} u) \\ q_0 = 1 \end{cases}$

MMSE $= q_0 - m = 1 - m$



The best one can do according to the SNR.

AVERAGE FREE ENERGY: $\boxed{F = -\mathbb{E}_{x^*,z}(\log Z) = -\int dx^* P_x(x^*) \int \mathcal{D}z \log \int dx\, e^{-\frac{\lambda}{2}x^2 + x\lambda x^* + x\sqrt{\lambda} z} P_x(x)}$

↳ we'll need $f_{\text{growing}}(\lambda^{(1)}, \lambda^{(2)}) = -\int dx^* P_x(x^*) \int \mathcal{D}z \log \int dx\, e^{-\frac{\lambda^{(1)}}{2}x^2 + x\lambda^{(2)} x^* + x\sqrt{\lambda^{(1)}} z} P_x(x)$.

↳ One can show that $\boxed{I(x;Y) = F + \lambda \frac{\mathbb{E}(x^2)}{2}}$ Ex: $\to \frac{\partial I}{\partial \lambda} = \frac{1}{2}\underbrace{\left(\langle x \rangle_{\text{MMSE}} - x^*\right)^2}_{=\text{MMSE}}$ I-MMSE theorem.

LECTURE II: SPIKE MATRIX MODEL

II a introduction

Our signal will have this time many component $\vec{x}^* = \begin{cases} x_1^* \\ \vdots \\ x_N^* \end{cases}$ with $x_i^* \sim P_x(x_i)$ i.i.d.

Our noisy measurement is a $N \times N$ matrix: $Y = \underbrace{\sqrt{\frac{\lambda}{N}}\, \vec{x}^* \vec{x}^{*T}}_{} + W$ with $w_{ij} = w_{ji} \sim \mathcal{N}(0,1)$.

perturbation to the random matrix W (wigner) corresponding to changing 1 e.v only → this is why thinking in terms of spectrum we talk about "spike".

QUESTION: Can we recover X^* from Y? Is there an efficient algorithm to do so? N^2 measures for N components

↳ EASY! if λ large enough (spike out of bulk)

↳ Rk: this definition can be extended to tensors: $y_{ijk} = \frac{\sqrt{\lambda}\, x_i^* x_j^* x_k^*}{N^{3-1/2}} + w_{ijk}$ ORDER 3, RANK 1 TENSOR

↳ much harder because there is no such thing as eigen decomposition for tensors

Application, binary variables + noise = random flips → SBM

Let's compute the posterior: $\begin{cases} P(\vec{X}|Y) \propto P_x(\vec{x}) \prod_{i \le j} e^{-\frac{1}{2}\left(y_{ij} - \sqrt{\frac{\lambda}{n}} x_i x_j\right)^2} \propto \prod P_x(x_i) \prod_{i \le j} e^{-\frac{\lambda}{2N} x_i^2 x_j^2 + y_{ij} x_i x_j \sqrt{\lambda/N}} \\ Z(Y) = \int d\vec{x} \prod_i P_x(x_i) \prod_{i \le j} e^{-\frac{\lambda}{2N} x_i^2 x_j^2 + y_{ij} x_i x_j \sqrt{\lambda/N}} \end{cases}$

Restricting ourselves to $x_i = \pm 1$: $(x_i^2 = 1) \Rightarrow P(\vec{X}|Y) \propto e^{\sqrt{\lambda} \sum_{i \le j} \frac{y_{ij}}{\sqrt{N}} x_i x_j}$ → Disordered Ising model with correlated interactions.

unlike the SK model, interactions are not i.i.d: $y_{ij} = w_{ij} + \sqrt{\frac{\lambda^*}{n}} x_i^* x_j^*$ ← let's assume there is a ground truth X^* that we might not know.

$$P(\vec{x}|Y) \propto e^{\sqrt{\lambda} \sum_{i \le j} \left(\frac{w_{ij}}{\sqrt{N}} + x_i^* x_j^* \sqrt{\frac{\lambda^*}{N}}\right) x_i x_j}$$

$(x_i^*)^2 = 1$

$$\propto \exp\left[\sqrt{\lambda} \sum_{i \le j} \left(\frac{1}{\sqrt{N}} w_{ij}\, x_i^* x_j^* + \sqrt{\frac{\lambda^*}{N}}\right) x_i x_i^* x_j x_j^*\right]$$

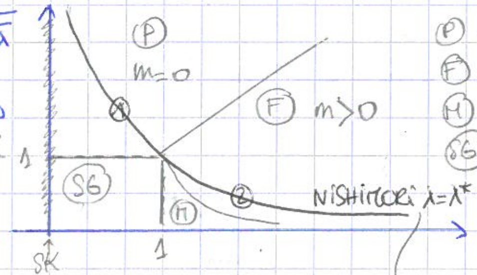
→ by the gauge transform we got rid of correlations at the price of a non zero mean

non-zero mean

$T = \frac{1}{\beta} = \frac{1}{\sqrt{\lambda}}$

diagram of model parameters

PHASE DIAGRAM

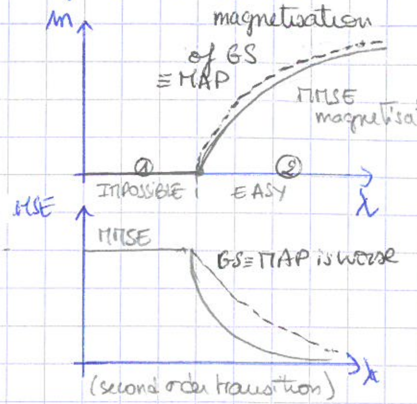


Ⓟ : paramagnetic ;
Ⓕ : ferromagnetic
Ⓜ : mixed : $m>0$, full RSB
ⓈⒼ : spin glass ; full RSB

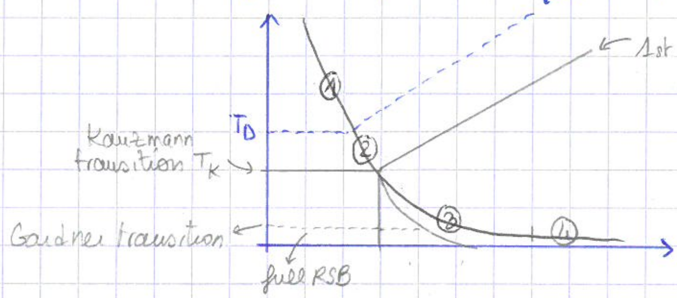
$T_c = \sqrt{\lambda^*}$

where the Nishimori relations hold.

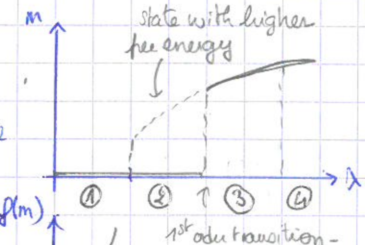
along the Nishimori line



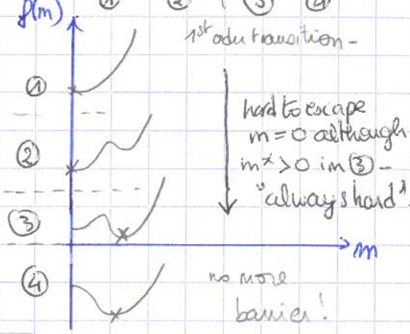
Now in the case of a tensor of order 3 :



now 3 phase along Nishimori line



COMPUTATIONAL COMPLEXITY :	
IMPOSSIBLE (indistinguishable)	②
POSSIBLE but HARD	③
EASY.	④



↳ This two cases are representative of the different scenario : ↳ 1st order with impossible → easy
↳ 2nd order with impossible → hard → easy.

We'll see how to make the necessary computations on the Nishimori line.

↳ REPLICA COMPUTATION ON THE EXERCISE SHEET.

$$\lim_{N\to\infty} \frac{F}{N} = \min_{m} \left[\mathcal{E}_{den}(\lambda m, \lambda m) + \frac{\lambda m^2}{4} \right]$$

↳ meq.

→ replicas lengthy (crazy as well...)
↳ rigorous proof