MaxEnt 22 Institut Henri Poincaré

Local Max-Entropy and Free Energy Principles Solved by Belief Propagation

UNIVERSITÉ D'ARTOIS

Olivier Peltre

Univ. Artois, Laboratoire de Mathématiques de Lens (LML)

Introduction

UNIVERSITÉ D'ARTOIS

Belief Propagation (BP) is an old algorithm with many applications [Gallager 64, Pearl 82]
(low-density decoding, bayesian inference, AI, stereovision, condesed matter simulation ...)

[Yedidia 05]
Generalized Belief Propagation (GBP) works with hypergraphs (i.e. coverings) instead of graphs

This talk is not about GBP, [Kikuchi 51, Morita 57, Yedidia 05]
It is about the geometric and statistical problems solved by GBP stationary states

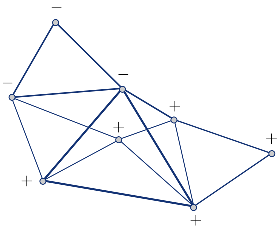
GBP can be viewed a soft constraint satisfaction solver trying to reach consistent
local probability distributions under energy conservation constraints.

Continuous-time diffusion equations generalise GBP with improved stability [P 20, 21]

Introduction

UNIVERSITÉ D'ARTOIS

Example: Spin Glasses



K random graph on a set of vertices $\Omega = \{1, \dots, N\}$

$(x_i) \in \{\pm 1\}^N$ random microstate of the system

$(w_{ij}) \in \mathbb{R}^{K_1}$ random couplings on the edges of K

$(b_i) \in \mathbb{R}^N$ random magnetic fields (or $b_i = B$)

Hamiltonian = Total Energy

$$H(x) = \sum_{ij \in K_1} w_{ij} x_i x_j + \sum_{i \in K_0} b_i x_i$$

Gibbs state = Graphical Model

$$p(x) = \frac{\mathrm{e}^{-\beta H(x)}}{\sum_x \mathrm{e}^{-\beta H(x)}}$$

Introduction

In thermodynamics, it is common to work with the following functionals:

- $S(\mathcal{U})$ *Boltzmann entropy* : defined as log-number of microstates having energy $\mathcal{U}$
- $\mathcal{F}(\beta,\mathcal{U})$ *variational free energy* : defined by $\mathcal{U} - \frac{1}{\beta} S(\mathcal{U})$
- $F(\beta)$ *free energy* : defined as $\min_{\mathcal{U}} \mathcal{F}(\beta,\mathcal{U})$, or from the partition function by $-\frac{1}{\beta} \ln Z(\beta)$

The Legendre transform relates these functionals, defining an energy-temperature relationship $\mathcal{U}(\beta)$

Thermodynamic equilibrium with different constraints: on $\mathcal{U}$ and β respectively

not tractable in practice

Short range interactions $\Rightarrow$ local approximations efficiently computed by GBP

Bethe-Kikuchi approximations

Cluster Variation Method

Outline

I.1. Microstates and Total Energy

UNIVERSITÉ D'ARTOIS

We consider a free sheaf $E : \mathcal{P}^{op}(\Omega) \to \mathbf{Set}$ of microstates, given by:

$$E_{\mathrm{a}} = \prod_{i \in \mathrm{a}} E_i \qquad \text{for all} \quad \mathrm{a} \subseteq \Omega$$

The restriction maps $\pi^{\mathrm{a} \to \mathrm{b}} : E_{\mathrm{a}} \to E_{\mathrm{b}}$ for every $\mathrm{b} \subseteq \mathrm{a}$ simply forget the state of atoms outside of b
We will write:

$$x_{\mathrm{a}|\mathrm{b}} = \pi^{\mathrm{a} \to \mathrm{b}}(x_{\mathrm{a}})$$

A statistical system on E_Ω is classically defined by a hamiltonian or energy function $H_\Omega : E_\Omega \to \mathbb{R}$
One can always define a finest covering $K \subseteq \mathcal{P}(\Omega)$ such that H_Ω decomposes as a sum of local interactions:

$$H_\Omega(x_\Omega) = \sum_{\mathrm{a} \in K} h_{\mathrm{a}}(x_{\Omega|\mathrm{a}})$$

I.2. Observables

Local observables form a cosheaf $\mathbb{R}^E : \mathcal{P}(\Omega) \to \mathbf{Alg}_c$ of commutative algebras

The extension maps $\pi^{\mathrm{a}\to\mathrm{b}\,*} : \mathbb{R}^{E_\mathrm{b}} \to \mathbb{R}^{E_\mathrm{a}}$ for all $\mathrm{b} \subseteq \mathrm{a}$ can be viewed as inclusions $\mathbb{R}^{E_\mathrm{b}} \subseteq \mathbb{R}^{E_\mathrm{a}}$

Def.

Observable fields on $K \subseteq \mathcal{P}(\Omega)$ form a chain complex $\big(C_\bullet(K, \mathbb{R}^E), \delta\big)$ where:

- $C_0 = \oplus_\mathrm{a} \mathbb{R}^{E_\mathrm{a}}$ consists of potentials $[h_\mathrm{a} : E_\mathrm{a} \to \mathbb{R} \mid \mathrm{a} \in K]$
- $C_1 = \oplus_{\mathrm{b}\subset\mathrm{a}} \mathbb{R}^{E_\mathrm{b}}$ consists of heat fluxes $[\varphi_{\mathrm{a}\to\mathrm{b}} : E_\mathrm{b} \to \mathbb{R} \mid \mathrm{a}, \mathrm{b} \in K \text{ and } \mathrm{b} \subseteq \mathrm{a}]$

and the codifferential $\delta : C_1 \to C_0$ acts by:

$$\delta\varphi_\mathrm{b}(x_\mathrm{b}) = \sum_{\mathrm{b}\subseteq\mathrm{a}} \varphi_{\mathrm{a}\to\mathrm{b}}(x_\mathrm{b}) - \sum_{\mathrm{c}\subseteq\mathrm{b}} \varphi_{\mathrm{b}\to\mathrm{c}}(x_{\mathrm{b}|\mathrm{c}})$$

Gauss thm. = Total energy conservation P 20

Two potentials h and h' are homologous, i.e. $h' = h + \delta\varphi$ if and only if:

$$\sum_{\mathrm{a}\in K} h_\mathrm{a}(x_{\Omega|\mathrm{a}}) = \sum_{\mathrm{a}\in K} h'_\mathrm{a}(x_{\Omega|\mathrm{a}})$$

I.3. Densities and Beliefs

UNIVERSITÉ D'ARTOIS

Local densities define a functor $\mathbb{R}^{E\,*} : \mathcal{P}(\Omega)^{op} \to \mathbf{Vect}$ of vector spaces

The pushforwards $\pi_*^{\mathrm{a}\to\mathrm{b}} : \mathbb{R}^{E_\mathrm{a}*} \to \mathbb{R}^{E_\mathrm{b}*}$ for all $\mathrm{b} \subseteq \mathrm{a}$ denote partial integration on fibers $\{x_{\mathrm{a}|\mathrm{b}} = x_\mathrm{b}\}$

Def. (Prop.)

Density fields on $K \subseteq \mathcal{P}(\Omega)$ form a cochain complex $\big(C_\bullet(K, \mathbb{R}^E)^*, d\big)$ where the differential $d : C_0^* \to C_1^*$ acts by:

$$dp_{\mathrm{a}\to\mathrm{b}}(x_\mathrm{b}) \;=\; p_\mathrm{b}(x_\mathrm{b}) \;-\; \sum_{x_{\mathrm{a}|\mathrm{b}} = x_\mathrm{b}} p_\mathrm{a}(x_\mathrm{a})$$

A density field $p \in C_0^*$ is said consistent if $dp = 0$

Def.

Beliefs lie in the convex subspace $\Delta_0 \subseteq C_0^*$ of positive local probability measures

Consistent beliefs lie in the convex polytope $\Gamma_0 \subseteq \Delta_0$ intersecting $\mathrm{Ker}(d)$

Consistent beliefs $p \in \Gamma_0$ act as local substitutes for global probabilities $p_\Omega \in \mathrm{Prob}(E_\Omega)$

(although positivity of the linear extension $p_\Omega \in \mathbb{R}^{E_\Omega *}$ is generally not guaranteed)

I.4. Local Gibbs Correspondence

UNIVERSITÉ D'ARTOIS

GBP tries to solve 2 different kinds of constraints simultaneously:

- Energy conservation on $h \in C_0$: potentials lie in $[h] = h + \delta C_1$ during evolution
- Marginal consistency on $p \in \Delta_0$: beliefs lie in $\Gamma_0 \subseteq \mathrm{Ker}(d)$ at equilibrium

A local correspondence $C_0 \to \Delta_0$ is therefore essential to the dynamic of GBP:

$$C_0 \xrightarrow[\text{linear}]{\zeta} C_0 \xrightarrow[\text{softmin}]{\rho^\beta} \Delta_0$$

Def.

The Gibbs correspondence for $\beta > 0$ is written $p = \rho^\beta(H)$ with $H = \zeta h$ as:

$$p_{\mathrm{a}}(x_{\mathrm{a}}) = \frac{1}{Z^\beta_{\mathrm{a}}} \mathrm{e}^{-\beta H_{\mathrm{a}}(x_{\mathrm{a}})} \qquad \text{with} \qquad H_{\mathrm{a}}(x_{\mathrm{a}}) = \sum_{\mathrm{b} \subseteq \mathrm{a}} h_{\mathrm{b}}(x_{\mathrm{a|b}})$$

- $\zeta : C_0 \to C_0$ is the zeta transform, combinatorial automorphism of C_0
- $\rho^\beta : C_0 \to \Delta_0$ is a softmin yielding local Gibbs states at inverse temperature $\beta = \frac{1}{T}$

I.5. Möbius inversion and Bethe numbers

UNIVERSITÉ D'ARTOIS

Thm. (Möbius inversion) (Dirichlet, Rota 64, P 20)

The Möbius transform $\mu : C_0 \to C_0$ inverting ζ acts by Dirichlet convolution as:

$$\mu(H)_{\mathrm{a}}(x_{\mathrm{a}}) = \sum_{\mathrm{b} \subseteq \mathrm{a}} \mu_{\mathrm{a} \to \mathrm{b}} \cdot H_{\mathrm{b}}(x_{\mathrm{a}|\mathrm{b}}) \qquad = \quad h_{\mathrm{a}}(x_{\mathrm{a}})$$

Coefficients $\mu_{\mathrm{a} \to \mathrm{b}} \in \mathbb{Z}$ can be recursively computed by N sparse matrix products,
denoting by N the maximal degree of $C_\bullet$ (= maximal length of a strict chain in K)

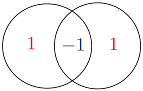
The reciprocal pair (ζ, μ) can moreover be extended to full automorphisms $C_\bullet \to C_\bullet$

Def. (Prop.)

Bethe-Kikuchi numbers $(c_{\mathrm{a}})_{\mathrm{a} \in K}$ are defined by the following equivalent formulas:

(i) $\forall \mathrm{b} \in K \qquad \sum_{\mathrm{a} \supseteq \mathrm{b}} c_{\mathrm{a}} = 1$

(ii) $\forall \mathrm{b} \in K \qquad c_{\mathrm{b}} = \sum_{\mathrm{a} \supseteq \mathrm{b}} \mu_{\mathrm{a} \to \mathrm{b}}$



Outline

UNIVERSITÉ D'ARTOIS

I. Local statistics : topology and combinatorics

II. Cluster variational principles solved by GBP

III. Example of diffusion with singularities

II.2 Regionwise Information Functionals

Def.

For all $\mathrm{a} \subseteq \Omega$ and $\beta > 0$ consider the following functionals:

- Shannon entropy given by: $S_\mathrm{a} : \mathrm{Prob}(E_\mathrm{a}) \to \mathbb{R}$

$$S_\mathrm{a}(p_\mathrm{a}) = -\sum_{E_\mathrm{a}} p_\mathrm{a} \ln(p_\mathrm{a})$$

- Variational free energy given by: $\mathcal{F}^\beta_\mathrm{a} : \mathrm{Prob}(E_\mathrm{a}) \times \mathbb{R}^{E_\mathrm{a}} \to \mathbb{R}$

$$\mathcal{F}^\beta_\mathrm{a}(p_\mathrm{a}, H_\mathrm{a}) = \mathbb{E}_{p_\mathrm{a}}[H_\mathrm{a}] - \frac{1}{\beta} S_\mathrm{a}(p_\mathrm{a})$$

- Free energy given by: $F^\beta_\mathrm{a} : \mathbb{R}^{E_\mathrm{a}} \to \mathbb{R}$

$$F^\beta_\mathrm{a}(H_\mathrm{a}) = -\frac{1}{\beta} \ln \sum_{E_\mathrm{a}} \mathrm{e}^{-H_\mathrm{a}}$$

Then:

(i) $F^\beta_\mathrm{a}(H_\mathrm{a}) = \min \mathcal{F}^\beta_\mathrm{a}(-, H_\mathrm{a})$ for all $H_\mathrm{a} \in \mathbb{R}^{E_\mathrm{a}}$

(ii) the differential $F^\beta_{\mathrm{a}*} : H_\mathrm{a} \mapsto \rho^\beta_\mathrm{a}(H_\mathrm{a})$ of F^β_a coïncides with the Gibbs state map

(iii) $\rho^\beta_\mathrm{a}(H_\mathrm{a})$ maximises S_a under the constraint $\mathbb{E}_{p_\mathrm{a}}[H_\mathrm{a}] = \mathcal{U}_\mathrm{a}$ for some $\mathcal{U}_\mathrm{a} \in \mathbb{R}$

II.1 Faithful Diffusions

UNIVERSITÉ D'ARTOIS

We consider continuous-time diffusion equations in C_0 defined by a flux functional $\Phi : C_0 \to C_1$

$$\frac{dv}{dt} = \delta\Phi(v) \qquad \text{with} \qquad v_{|t=0} = h$$

Energy conservation $\boxed{v \in [h] = h + \delta C_1}$ is therefore naturally enforced by any discrete integrator.

homology class

(GBP equivalent but with $\Delta t = 1$!)

Def.

Given $\beta > 0$, we call consistent manifold the subspace $\mathrm{Fix}^\beta \subseteq C_0$ defined by ;

$$\mathrm{Fix}^\beta = \{v \in C_0 \mid \; p = \rho^\beta(\zeta v) \in \Gamma_0\}$$

Given a flux functional $\Phi : C_0 \to C_1$, we say that:

- Φ is consistent if $\Phi_{|\mathrm{Fix}^\beta} = 0$
- Φ is faithful if $\delta\Phi(v) = 0 \;\Rightarrow\; v \in \mathrm{Fix}^\beta$ and Φ is consistent.

Consistent diffusions are stationary for $v \in [h] \cap \mathrm{Fix}^\beta$

Faithful diffusions are stationary only for $v \in [h] \cap \mathrm{Fix}^\beta$

II.3 Variational Free Energy Principle

UNIVERSITÉ D'ARTOIS

Problem 1.

Chose an inverse temperature $\beta > 0$ and local energies $H = \zeta h \in C_0$

Find $p \in \Gamma_0$ critical for the Bethe-Kikuchi variational free energy $\check{\mathcal{F}} : \Delta_0 \times C_0 \to \mathbb{R}$:

$$\check{\mathcal{F}}(p, H) = \sum_{\mathrm{a} \in K} c_{\mathrm{a}} \mathcal{F}_{\mathrm{a}}(p_{\mathrm{a}}, H_{\mathrm{a}})$$

Thm 1. [Yedidia 05, P 19]

Let $\beta > 0$ and $H = \zeta h \in C_0$

Solutions of problem 1 are given by $p = \rho^\beta(\zeta v)$ for all $v \in C_0$ such that:

$$v \in [h] \cap \mathrm{Fix}^\beta$$

II.4 Max-Entropy Principle

Université d'Artois

Problem 2.

Chose a mean energy $\mathcal{U} \in \mathbb{R}$ and local energies $H = \zeta h \in C_0$

Find $p \in \Gamma_0$ constrained to $\langle p, h \rangle = \mathcal{U}$ for the Bethe-Kikuchi entropy $\check{S} : \Delta_0 \to \mathbb{R}$:

$$\check{S}(p) = \sum_{\mathrm{a} \in K} c_{\mathrm{a}} S_{\mathrm{a}}(p_{\mathrm{a}})$$

Where $\langle p, h \rangle = \sum_{\mathrm{a} \in K} \mathbb{E}_{p_{\mathrm{a}}}[h_{\mathrm{a}}] = \sum_{\mathrm{a} \in K} c_{\mathrm{a}} \, \mathbb{E}_{p_{\mathrm{a}}}[H_{\mathrm{a}}] = \mathcal{U}$ $\quad \Leftarrow \quad$ Bethe mean energy is exact by $\sum_{\mathrm{a}} h_{\mathrm{a}} = \sum_{\mathrm{a}} c_{\mathrm{a}} H_{\mathrm{a}}$ and linearity of expectations

Thm 2.

Let $\mathcal{U} \in \mathbb{R}$ and $H = \zeta h \in C_0$

Solutions of problem 2 are given by $p = \rho^1(\zeta \bar{v})$ for all $\bar{v} \in C_0$ such that:

$$\exists \beta > 0 \qquad \bar{v} \in [\beta h] \cap \mathrm{Fix}^1$$

N.B. Letting $v = \beta \bar{v}$ one then recovers $v \in [h] \cap \mathrm{Fix}^\beta$

II.5 Free Energy Principle

Problem 3.

Chose an inverse temperature $\beta > 0$ and local energies $H = \zeta h \in C_0$

Find $V = \zeta v \in C_0$ constrained to $v \in [h]$ critical for the Bethe-Kikuchi free energy $\check{F} : C_0 \to \mathbb{R}$:

$$\check{F}(V) = \sum_{\mathrm{a} \in K} c_{\mathrm{a}} F_{\mathrm{a}}(V_{\mathrm{a}})$$

Let us call weakly consistent manifold $\mathrm{Fix}^{\beta}_{+} = \{v + \epsilon \,|\, v \in \mathrm{Fix}^{\beta}, \epsilon \in \mathrm{Ker}(c\zeta)\}$

(retracts linearly on Fix^{β}, coïncides with Fix^{β} when $c_{\mathrm{a}} \neq 0$ for all $\mathrm{a} \in K$)

Thm 3.

Let $\beta > 0$ and $H = \zeta h \in C_0$

Solutions of problem 1 are given by $V = \zeta v$ for all $v \in C_0$ such that:

$$v \in [h] \cap \mathrm{Fix}^{\beta}_{+}$$

Outline

III.1 Energy-Temperature Relationship(s)

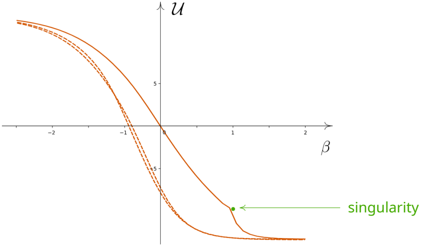
Problems 1, 2 and 3 all imply finding some $v \in [h] \cap \mathrm{Fix}^\beta$ (i.e. conservative and consistent), however with different constraints:

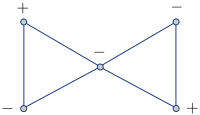
- Free energy principles (1,3) with temperature constraints on $\beta = \frac{1}{T}$
- The max-entropy principle (2) with mean energy constraints on $\mathcal{U} = \langle p, h \rangle$

Therefore solutions of (2) might not coïncide with solutions of (1) and (3).

There may exist multiple stationnary energies fore one given temperature constraint, and reciprocally.

Energy-Temperature Diagrams:





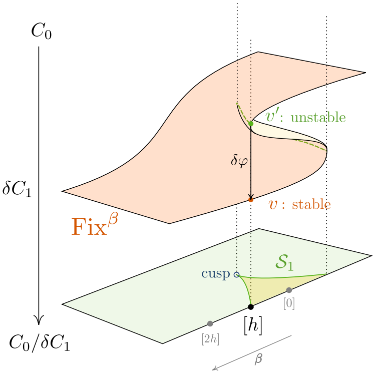
K graph with two loops

$$E_i = \{\pm\}$$
$$E_{ij} = \{\pm\}^2$$

III.2 Singularities on cyclic graphs

Singularities of the projection $\Pi : C_0 \to C_0/\delta C_1$ restricted to Fix^β are ranked by dimension

$$\mathcal{S}_k = \{ v \in \mathrm{Fix}^\beta \mid \mathrm{Ker}(\Pi_*) \cap \delta C_1 \text{ of dimension } k \}$$



This makes $\mathrm{Fix}^\beta = \sqcup_{k \geq 0} \mathcal{S}_k$ a stratified space

On simple graphs with 2+ loops one can parameterise the singular surface $\mathcal{S}_1$

(smooth in C_0)

Linearised diffusion does not restrict to an isomorphism of δC_1 on singularities

Continuous-time diffusion improves over GBP by yielding smooth trajectories, even nearby singularities

III.3 GBP and Bethe Diffusions

Let us define the conditional free energy functional $F_{\mathrm{a}\to\mathrm{b}} : \mathbb{R}^{E_\mathrm{a}} \to \mathbb{R}^{E_\mathrm{b}}$ by:

$$F_{\mathrm{a}\to\mathrm{b}}[V_\mathrm{a}|x_\mathrm{b}] = -\ln \sum_{x_{\mathrm{a}|\mathrm{b}}=x_\mathrm{b}} \mathrm{e}^{-V_\mathrm{a}(x_\mathrm{a})} \qquad \Rightarrow F_{\mathrm{a}\to\mathrm{b}}[V_\mathrm{b}|x_\mathrm{b}] = V_\mathrm{b}(x_\mathrm{b})$$

And the free energy gradient $\mathcal{D} : C_0 \to C_1$ by:

$$\mathcal{D}(V)_{\mathrm{a}\to\mathrm{b}}(x_\mathrm{b}) = V_\mathrm{b}(x_\mathrm{b}) - F_{\mathrm{a}\to\mathrm{b}}[V_\mathrm{a}|x_\mathrm{b}] \qquad = F_{\mathrm{a}\to\mathrm{b}}[V_\mathrm{b} - V_\mathrm{a} \,|\, x_\mathrm{b}]$$

We may then define the diffusion fluxes $\Phi_{\mathrm{GBP}} = -\delta\mathcal{D}\zeta$ and $\Phi_{\mathrm{Bethe}} = -\delta\mu\mathcal{D}\zeta$:

$$\begin{tikzcd} C_0 \arrow[r, "\zeta"] & C_0 \arrow[d, "\mathcal{D}"] \\ & C_1 \arrow[ul, "\delta"] \end{tikzcd}$$

$$\frac{dV_\mathrm{b}}{dt} = \sum_{\mathrm{a}\not\subseteq\mathrm{b}} \sum_{\mathrm{c}\subseteq\mathrm{b}} F_{\mathrm{a}\to\mathrm{c}}\big(V_\mathrm{a} - V_\mathrm{c}\big)$$

FAITHFUL

$$\begin{tikzcd} C_0 \arrow[r, "\zeta"] & C_0 \arrow[d, "\mathcal{D}"] \\ C_1 \arrow[u, "\delta"] & C_1 \arrow[l, "\mu"] \end{tikzcd}$$

$$\frac{dV_\mathrm{b}}{dt} = \sum_{\mathrm{a}\not\subseteq\mathrm{b}} c_\mathrm{a}\, F_{\mathrm{a}\to\mathrm{a}\cap\mathrm{b}}\big(V_\mathrm{a} - V_{\mathrm{a}\cap\mathrm{b}}\big)$$

LOCALLY FAITHFUL

Thm
(P 20)

Thank you!

UNIVERSITÉ D'ARTOIS

Repository:
github.com/opeltre/topos

References:

Belief Propagation

Gallager 1963 : *Low-Density Parity-Check Codes*, MIT Press

Pearl 1982 : *Reverend Bayes on Inference Engines : A Distribute Hierarchichal Approach*, AAAI-82

Yedidia Freeman Weiss 2005 : *Constructing Free Energy Approximations and GBP algorithms*, IEEE (51)

Peltre 2021 : *Belief Propagation as Diffusion*, GSI 21 Proceedings

Bethe-Kikuchi Free Energies

Bethe Bragg 1935 : *Statistical Theory of Superlattices,* Proceedings of the Royal Society of London

Kikuchi 1951: *A Theory of Cooperative Phenomena,* Phys. Rev. (81)

Morita 1957 : *Cluster Variation Method of Cooperative Phenomena and its Generalisation I*, J. Phys. Soc. Japan

Yedidia Freeman Weiss 2005 : *Constructing Free Energy Approximations and GBP algorithms*, IEEE (51)

Peltre 2022 : *Local Max-Entropy and Free Energy Principles solved by GBP*, MaxEnt 22 - arxiv: 2207.00841

PhD thesis

Message Passing Algorithms and Homology, 2020, Université Paris Cité - arxiv: 2009.11631