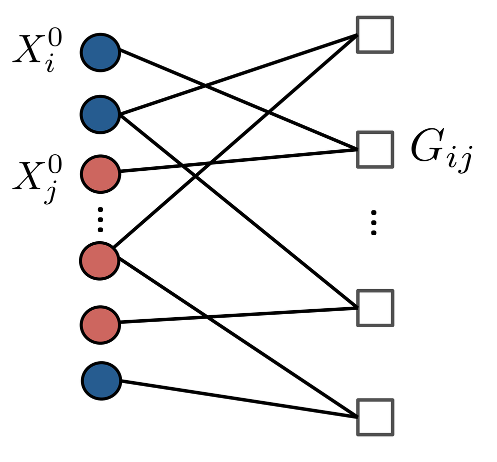
Adaptive interpolation scheme for inference problems with sparse underlying factor graph

Jean Barbier, Chun Lam Chan and Nicolas Macris

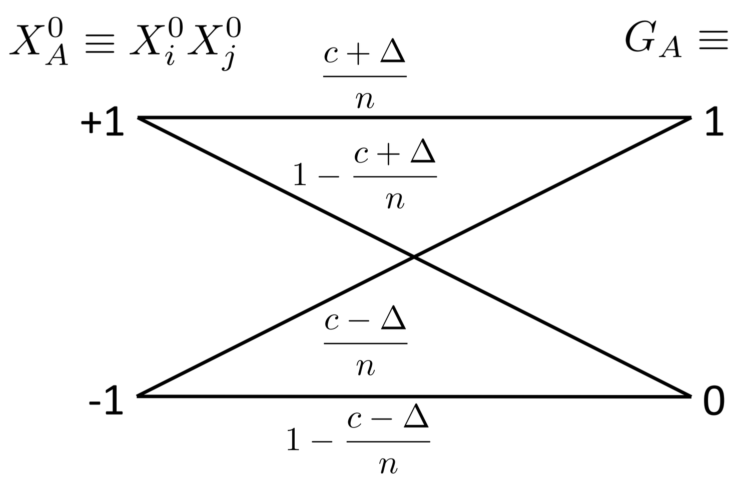
Laboratoire de Théorie des Communications (LTHC) - EPFL

EPFL Information Processing Group

Inference problem: stochastic block model



factor graph



(asymmetric) observation channel

- Ground truth: $\boldsymbol{X}^0 \in \{-1,+1\}^n$ uniformly distributed
- #observations m drawn from Poisson distribution $\sim \mathrm{Poi}(n(n-1)/2)$
- Each observation $G_A \in \{0,1\}$ follows the law

$$\mathbb{P}(G_A = 1 | X_A^0) = \frac{c}{n} + \frac{\Delta}{n} X_A^0$$

- Partition function:

$$\mathcal{Z} = \sum_{\boldsymbol{X} \in \{\pm 1\}^n} \prod_{A=1}^{m} (c+\Delta X_A)^{G_A} \left(1 - \frac{c}{n} - \frac{\Delta}{n} X_A\right)^{1-G_A}$$

- Posterior:

$$P(\boldsymbol{X}|\boldsymbol{G}) = \frac{1}{\mathcal{Z}} \prod_{A=1}^{m} (c+\Delta X_A)^{G_A} \left(1 - \frac{c}{n} - \frac{\Delta}{n} X_A\right)^{1-G_A}$$

Free entropy and the replica formula

- Free entropy $f \equiv \frac{1}{n}\mathbb{E}\ln \mathcal{Z}$: useful and link to mutual information, but the log-sum is intractable
- Conjecture from physics: $\lim_{n\to\infty} f$ is given by the variational problem of the potential function $f_{\rm RS}(M, \mathsf{x}, \tilde{\mathsf{x}})$ (M = scalar, $\mathsf{x}, \tilde{\mathsf{x}}$ = distributions)
- Story of this poster: **adaptive interpolation** as an optimistic strategy to prove this conjecture

Adaptive interpolation [1]

The interpolating free entropy

$$f_t := \frac{1}{n}\mathbb{E}\left[\ln \sum_{\boldsymbol{X}} e^{-\mathcal{H}_t(\boldsymbol{X})}\right]$$

provides a sum rule between f and the generalized potential function $\tilde{f}(\boldsymbol{M}, \mathbf{x}, \tilde{\mathbf{x}})$, where $\boldsymbol{M} = (M^{(1)}, M^{(2)}, \dots, M^{(T)})$, $\mathbf{x} = (\mathsf{x}^{(1)}, \mathsf{x}^{(2)}, \dots, \mathsf{x}^{(T)})$, $\tilde{\mathbf{x}} = (\tilde{\mathsf{x}}^{(1)}, \tilde{\mathsf{x}}^{(2)}, \dots, \tilde{\mathsf{x}}^{(T)})$.

The interpolating Hamiltonian $\mathcal{H}_t(\boldsymbol{X})$ is

$$\mathcal{H}_t(\boldsymbol{X}) = -\sum_{A=1}^{m^{(t)}} \Big(G_A \ln(c+\Delta X_A) + (1-G_A)\ln(1 - \frac{c+\Delta X_A}{n})\Big) - \sum_{t'=0}^{t}\sum_{i=1}^{n}\sum_{k=1}^{e_i^{(t')}} \ln(1 + X_i \tanh \tilde{L}_k^{(t')}) + \frac{\Delta}{T}\sum_{t'=0}^{t} M^{(t')} \sum_{i=1}^{n} X_i$$

- 1st line accounts for observations in SBM
  #observations $m^{(t)}$ is a Poisson number decreasing with t
- 2nd line accounts for decoupled messages
  From time $t-1$ to t, a message $M^{(t)}$ and a Poisson number $e_i^{(t)}$ of messages $\tilde{L}_k^{(t)}$ i.i.d. drawn from distribution $\tilde{\mathsf{x}}^{(t)}$ are added to every variable node i.

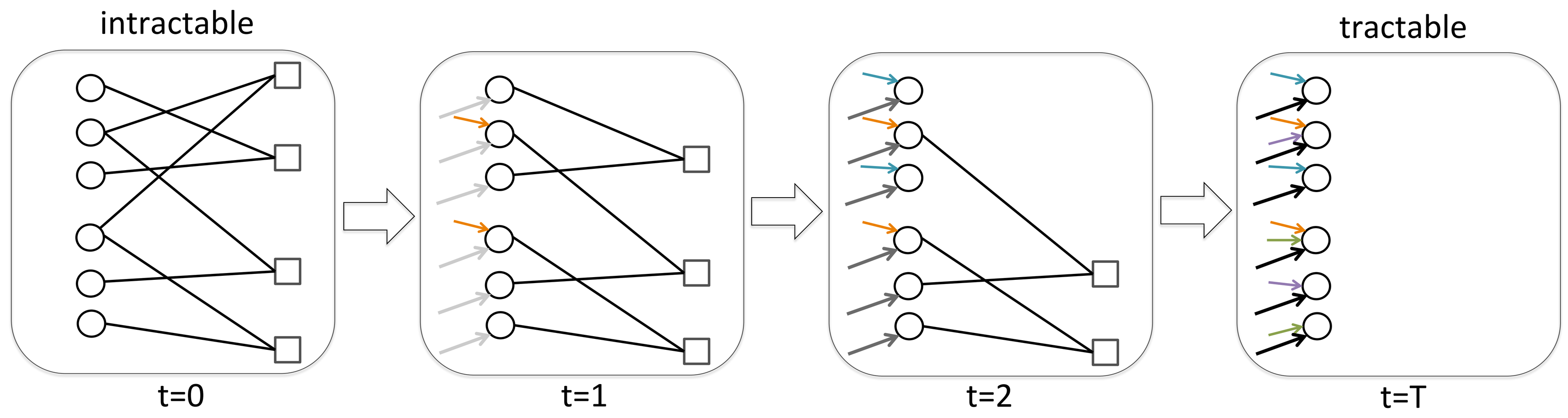


Figure: $t = 0$: a graph represents $\mathcal{H}_0$ and looks the same with the SBM; $t = T$: a decoupled graph represents $\mathcal{H}_T$

Trick of the interpolation:

$$f_0 = f_T - \sum_{t=1}^{T}(f_t - f_{t-1})$$

The interpolation scheme varies t from 0 to T

- $t = 0 \Rightarrow f_0$ recovers f
- $t = T \Rightarrow f_T$ recovers a term in $\tilde{f}(\boldsymbol{M}, \mathbf{x}, \tilde{\mathbf{x}})$
- $f_t - f_{t-1}$ can be converted to a nicer form using fundamental theorem of calculus

The sum rule yielded from the interpolation:

$$f = \tilde{f}(\boldsymbol{M}, \mathbf{x}, \tilde{\mathbf{x}}) + \frac{c}{2T}\sum_{t=1}^{T} \mathcal{R}_t^{\rm sparse} - \frac{\Delta}{2T}\sum_{t=1}^{T}(M^{(t)})^2$$

where

$$\mathcal{R}_t^{\rm sparse} = \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{p}\Big\{\Big(\frac{\Delta}{c}\Big)^p \Big(\mathbb{E}\langle Q_p^2\rangle_{t-1} - (q_p^{(t)})^2\Big) - 2\tilde{q}_p^{t}\big(\mathbb{E}\langle Q_p\rangle_{t-1} - q_p^{(t)}\big)\Big\}$$

- $q_p^{(t)} = \mathbb{E}[(\tanh L^{(t)})^p]$, $\tilde{q}_p^{(t)} = \mathbb{E}[(\tanh \tilde{L}^{(t)})^p]$ where $L^{(t)} \sim \mathsf{x}^{(t)}$, $\tilde{L}^{(t)} \sim \tilde{\mathsf{x}}^{(t)}$
- $Q_p = \frac{1}{n}\sum_{i=1}^{n} X_i^{(1)} \cdots X_i^{(p)}$
- $\langle - \rangle_{t-1}$ is the Gibbs measure associated with $\mathcal{H}_{t-1}$ s.t. **it depends on** $\tilde{\mathsf{x}}^{(1)}, \dots, \tilde{\mathsf{x}}^{(t-1)}$ **but not** $\tilde{\mathsf{x}}^{(t)}, \tilde{\mathsf{x}}^{(t+1)}, \dots$.

Canceling the remainder: We choose $M^{(t)} = 0$. Moreover, suppose $\mathbb{E}[\langle Q_p^2\rangle_t] = \mathbb{E}[\langle Q_p\rangle_t]^2$, then we can choose suitable $\tilde{\mathsf{x}}^{(t)}$ s.t. $\tilde{q}_p^{(t)} = \frac{1}{2}(\mathbb{E}\langle Q_p\rangle_{t-1} + q_p^{(t)})$ to cancel $\mathcal{R}_t^{\rm sparse}$.

Challenge and our related work

Conjecture 1: With some proper perturbation to $\mathcal{H}_t$ we conjecture full concentration of overlap

$$\lim_{n\to\infty} \mathbb{E}[\langle Q_p - \mathbb{E}\langle Q_p\rangle_t\rangle_t^2] = 0.$$

Conjecture 2: The generalized potential function $\tilde{f}$ shares the same extremum with $f_{\rm RS}$.

- For the SBM we can show the thermal concentration

$$\lim_{n\to\infty} \mathbb{E}[\langle Q_p - \langle Q_p\rangle_t\rangle_t^2] = 0.$$

- For any ferromagnetic models we can treat Conjecture 1 but not always Conjecture 2.
- For a special case of censored block model [2] we can treat Conjecture 1 & 2 and therefore give a full proof to the replica formula.

References

[1] J. Barbier and N. Macris, "The adaptive interpolation method: A simple scheme to prove replica formulas in bayesian inference," *CoRR*, vol. abs/1705.02780, 2017.

[2] J. Barbier, C. L. Chan, and N. Macris, "Adaptive path interpolation for sparse systems: Application to a simple censored block model," *CoRR*, vol. abs/1806.05121, 2018.