Approximate inference on the Ising grid

Brahim Khalil Abid, Ilyes Khemakhem, Rémi le Priol

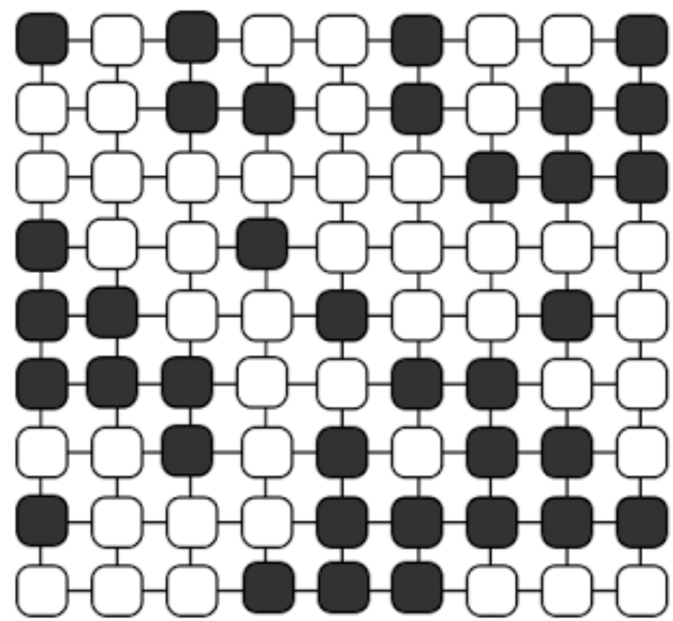
The Ising Model

Model: We consider a 2D grid $G = (V, E)$ where each nodes can take values -1 and +1. The probability for a certain configuration x to occur is proportional to $\exp\left(-E(x)\right)$ where the energy E(x) is written :

$$E(x) = -\frac{1}{2}\sum_{i} b_i x_i \ - \frac{1}{2}\sum_{(i,j)\in E} a_{i,j} x_i x_j$$

$a_{i,j}$ encodes the correlation between adjacent nodes. A large $a_{i,j}$ means that x_i and x_j will tend to have the same sign.
b_i encodes the potential at each node. $b_i > 0$ means that x_i will tend to be 1 and $b_i < 0$, -1.
In the following σ will stand for the logistic function $\sigma(z) = \frac{1}{1+e^{-z}}$

Problem: the canonical parameters θ=(a,b) are known. We would like to sample according to this distribution, infer the marginal probabilities, or equivalently the mean parameters $\mu = E[x]$.

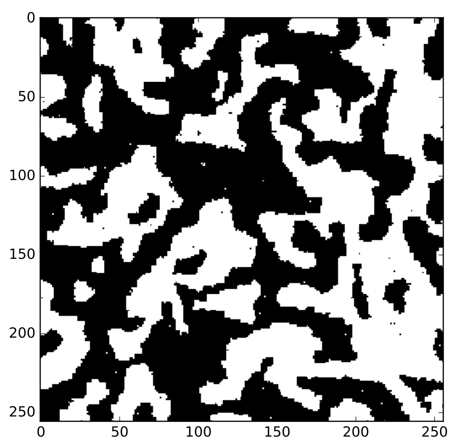


Block Gibbs Sampling

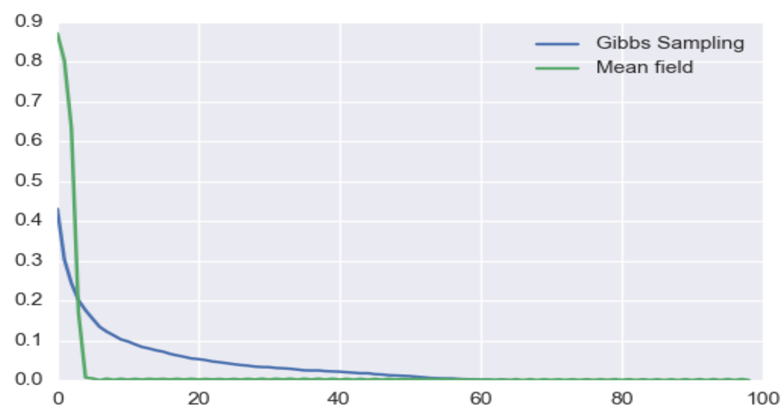
We start from an initial observation of the grid. We split the grid in two halves corresponding to the black and white boxes of a checkerboard. We then sample alternatively each half conditionally to the observation we have on the second half. Each white box being independent from the other white boxes conditionally to its black neighbors, we get the update probability:

$$\mathrm{P}(x_i = 1) = \sigma\left(b_i + \sum_{j \in N(i)} a_{i,j} x_j\right)$$

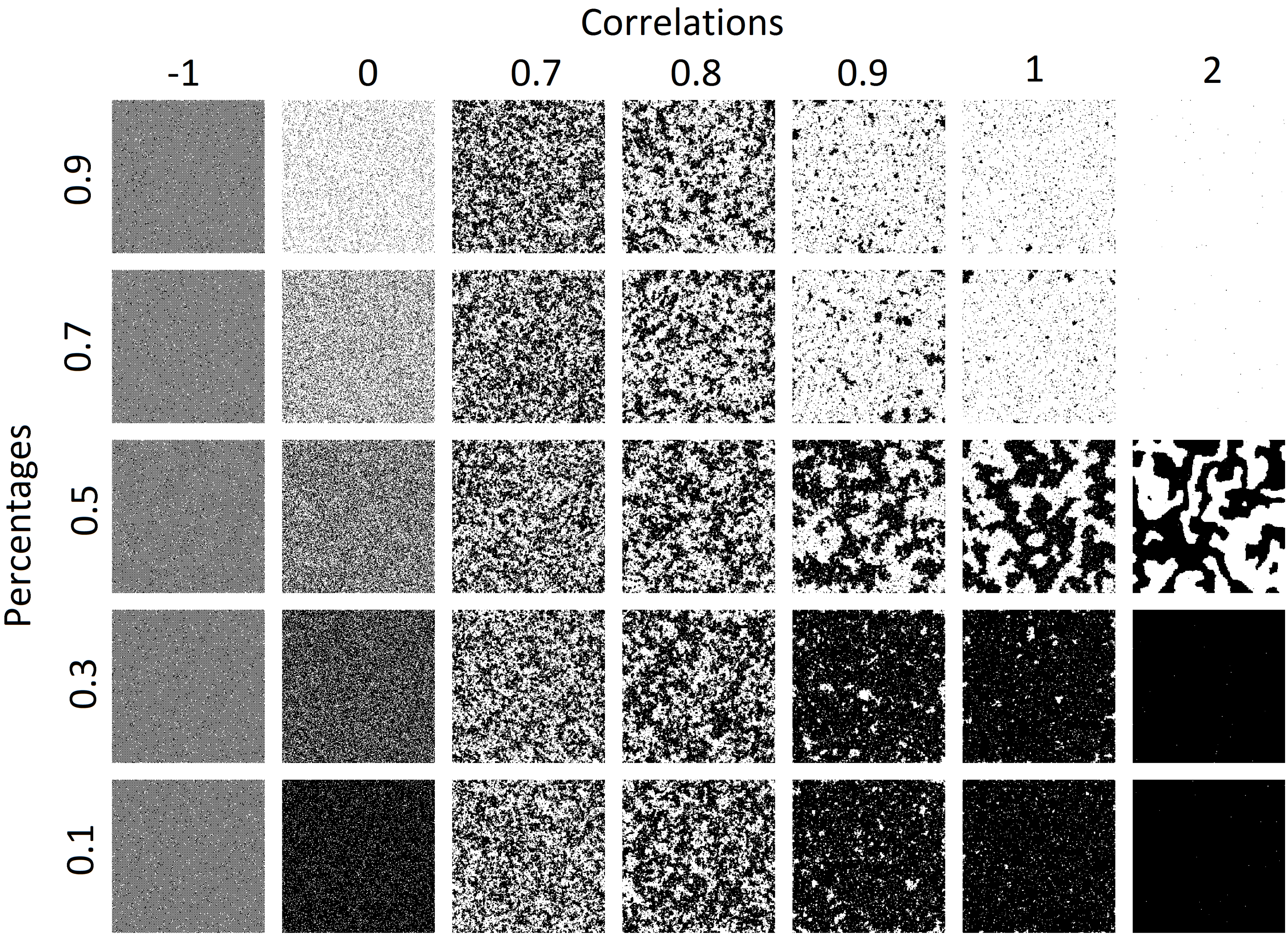
Example with a=1.78 and b=0



Gibbs sampling is consistently slower than mean field.



The plots against were drawn from a model with various uniform correlations $a_{i,j}$ and b_i=0. We also varied the initial percentage of +1.



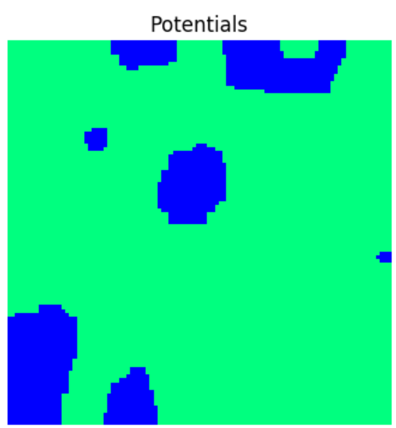
Correlations
-1
0
0.7
0.8
0.9
1
2
Percentages
0.9
0.7
0.5
0.3
0.1

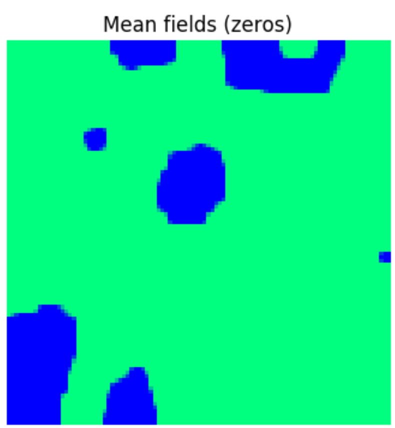
Mean Field

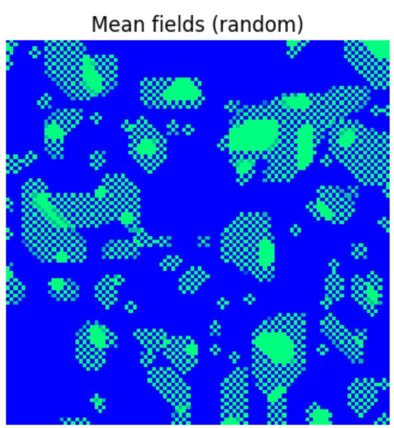
The inference problem on the Ising Grid is intractable. Instead of solving it, we look for the product model (empty graph) best approximating our model with regard to the Kullback Leibler divergence. We will take the mean parameters of this model as an approximation for the real mean parameters. With μ_i denoting the mean value at point i, minimizing the divergence gives the update :

$$\mu_i = 2 * \sigma\left(b_i + \sum_{j \in N(i)} a_{i,j}\mu_j\right) - 1$$

Interestingly, the mean field problem is non convex. Different initialization can yield different fixed points, as illustrated below with a=10 and the potentials b as plotted. The starting point 0 tends to give the expected result.

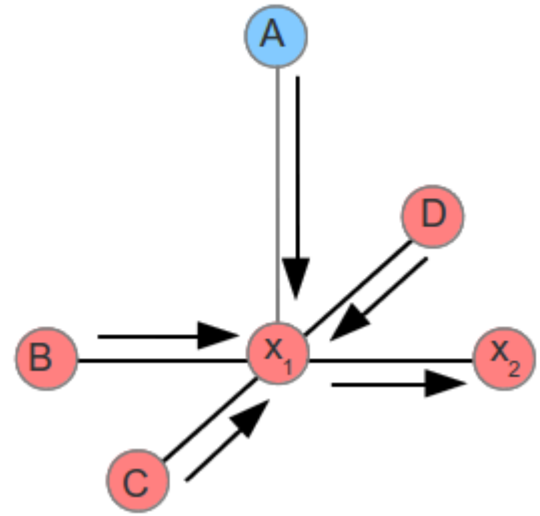






Loopy belief propagation

The belief propagation is a message passing algorithm that computes exact inference on trees. It can be used in graphs with cycles as well, however its convergence is not guaranteed and the result is not exact. It is then called loopy belief propagation (LBP).
Each node send a message to its neighbors telling them how likely he believes they are in a certain state. Messages are exchanged between nodes until the beliefs are stable (if ever).



Message passing from x_1 to x_2 with potential A and neighbors B, C, D

$$m_{j\to i}(x_i) = \sum_{x_j} \psi_j(x_j)\psi_{j,i}(x_i,x_j) \prod_{k\in N(j)\backslash\{i\}} m_{k\to j}(x_j)$$

$$\psi_i(x_i) = \exp\left(\frac{1}{2} b_i . x_i\right)$$

$$p(x_i) = \frac{1}{Z}\psi_i(x_i) \prod_{k\in N(i)} m_{k\to i}(x_i)$$

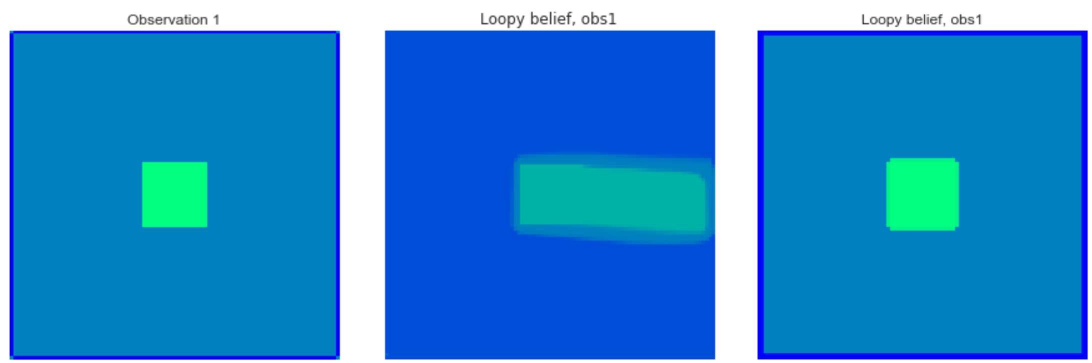
$$\psi_{i,j}(x_i,x_j) = \exp\left(\frac{1}{2} a_{i,j} . x_i . x_j\right)$$

Convergence of LBP

Messages tend to explode. The log-sum-exp trick proved to be inefficient. We had to normalize the messages at each iteration so that messages sent between two nodes sum to 1 on the different states.

$$m_{i\to j}(x_j) = \frac{m_{i\to j}(x_j)}{m_{i\to j}(1) + m_{i\to j}(-1)}$$

The choice of the message passing scheme also impacts the results. Different sequences will generally produce different results as seen against.



Even so, our algorithm does not produce very convincing results.

Results with observations

Practically, we want to start from observed nodes and infer the rest of the distribution. This is what we implemented for mean field and loopy belief propagation. On the charts below, the first line is the observation, the second is obtained by mean field, and the third by loopy belief propagation. In every cases, a=3. On the left, b=0, and on the right b is random.