1 Ising model and Glauber dynamics

MCMC are a class of algorithms that allow to sample from distributions that typically have large state spaces or are analytically intractable. We have already seen in previous classes an application of MCMC to the problem of coloring graphs. This is part of a larger set of ideas that can be conveniently applied when the state space has product form $S = \mathcal{X}^N$ where $\mathcal{X}$ is finite alphabet (say the colors $\{1, \cdots, q\}$) and N is very large (say the number of vertices in the graph that is colored). These algorithms go under the name of Glauber dynamics or heat bath dynamics or Gibbs sampler.[1] The goal of this chapter is to provide a short and elementary introduction with the particular example of the Ising model in mind. We will also use this particular dynamics as a concrete example to illustrate the coupling from the past method in the last two lectures.

1.1 Introduction to the Ising model

Consider a graph $G = (V, E)$ with $V = \{1, 2, ..., N\}$, and a binary alphabet $\mathcal{X} = \{1, -1\}$. Variables $\sigma_v \in \mathcal{X}$ are "attached" to the vertices $v \in V$. These variables are called "spins" for reasons explained briefly later on. The state space is made of spin assignments $\underline{\sigma} = (\sigma_1, \ldots, \sigma_N) \in \mathcal{X}^N$. If you wish you can think of the spin assignments as functions from V to the state space S. The probability distribution defining the "Ising model" is:

$$\mu(\underline{\sigma}) = \frac{1}{Z} \exp\Big(\sum_{(v,w)\in E} \beta J_{vw} \sigma_v \sigma_w + \sum_{v \in V} \beta h_v \sigma_v \Big), \tag{1}$$

where $\beta > 0$, J_{vw} and $h_v \in \mathbb{R}$ and Z is a normalization constant - called the partition function - equal to

$$Z = \sum_{\underline{\sigma} \in \{-1,+1\}^N} \exp\Big(\sum_{(v,w)\in E} \beta J_{vw} \sigma_v \sigma_w + \sum_{v \in V} \beta h_v \sigma_v \Big) \tag{2}$$

Such a distribution is of the form

$$\mu(\underline{\sigma}) = \frac{1}{Z} \exp(-\beta \mathcal{H}(\underline{\sigma})), \tag{3}$$

where

$$\mathcal{H}(\underline{\sigma}) = - \sum_{(v,w)\in E} J_{vw} \sigma_v \sigma_w - \sum_{v \in V} h_v \sigma_v \tag{4}$$

is a "cost" function called the Hamiltonian of the model.

Let us briefly discuss the physical origin (and interpretation) of this probability distribution. The Ising model (1920) was originally invented as a toy model for magnetic materials. Such models are of utmost importance in physics and statistical mechanics. One imagines that in these material there are magnetic degrees of freedom - the magnetic moments of atoms - attached to the sites of a cubic grid which represent the atoms of a crystal. For the simplest crystalline arrangement of atoms the graph would be a cubic grid of size $L \times L \times L = N$ where L is a linear dimension of the sample. The spins model the magnetic moments of the atoms. Each magnetic moment behaves like a little magnet oriented south→north $= +1$ or north→south$= -1$. These magnetic moments interact and each assignment has an energy cost equal to the Hamiltonian $\mathcal{H}(\underline{\sigma})$. The real numbers J_{vw} are related to the mutual interaction between neighboring magnetic moments, and h_v is a bias related to their interaction with external applied magnetic fields. At zero temperature the system tries to find the assignment which minimises this energy function. Note that this is for us an optimisation problem (that nature essentially solves). At finite temperature β^{-1}

[1]Glauber dynamics was initially introduced by Roy Glauber in 1963 in "Time dependent statistics of the Ising model", J. Math. Phys. 4 (1963) 294.

(and when the system is a thermal equilibrium) the magnetic moments fluctuate randomly and the basic laws of statistical mechanics state that they behave like random variables distributed according to the Gibbs distribution $\mu(\underline{\sigma})$.

The most important quantity that one would like to compute is perhaps the magnetization defined as follows

$$m = \frac{1}{n}\sum_{v=1}^{N}\langle s_v\rangle \tag{5}$$

where $\langle s_v\rangle$ is the expectation[2] of s_v with respect to $\mu(\underline{\sigma})$,

$$\langle s_v\rangle = \sum_{\underline{\sigma}\in\{-1,+1\}^N} \sigma_v\, \mu(\underline{\sigma}) \tag{6}$$

This quantity represents the total magnetic moment of the system. If it is non zero you have a magnet (e.g. that sticks to your fridge).

More generally one would like to compute the average value of "local observables", i.e. functions $A(\{\sigma_v\}_{v\in A})$ of some finite subset of spins $A \subset V$,

$$\langle A(\{\sigma_v\}_{v\in A})\rangle = \sum_{\underline{\sigma}\in\{-1,+1\}^N} A(\{\sigma_v\}_{v\in A})\, \mu(\underline{\sigma}) \tag{7}$$

If we knew how to calculate Z then we could easily compute such averages. Indeed the reader can check that

$$\langle s_v\rangle = \beta^{-1}\frac{\partial}{\partial h_v}\ln Z \tag{8}$$

This is one of the reasons why the quantity $\ln Z$ plays a very important role, and one of the major aims in statistical mechanics is to compute the "free energy" $f = -\beta^{-1}\lim_{n\to+\infty}\frac{1}{n}\ln Z$. This however is very hard.

MCMC methods allow to (numerically) compute averages of observables $\langle A(\{\sigma_v\}_{v\in A})\rangle$ by sampling from the Gibbs distribution without computing the normalisation factor Z.

Before dwelling into the MCMC methods for this model let us give a few important examples of special cases of the model and discuss other interpretations beyond physics.

Standard Ising model. See Figure 1. The graph G is a cubic d-dimensional grid. The vertex set is $V = \{v = (i_1, \cdots, i_d) \mid -L \le i_\ell \le L, \ell = 1, \cdots, d\}$ where L is an integer, and the edge set is $E = \{(v, w) \in V \times V \mid |v - w| = 1\}$. The total number of vertices is $(2L+1)^d$. One generally distinguishes two important physical models. The "ferromagnetic" model where $J_{vw} = J > 0$ for nearest neighbords; and the "antiferromagnetic" model where $J_{vw} = -J < 0$ for nearest neighbords. In the former case the spins have a tendency to "align" while in the second they have a tendency to "anti-align". Generally one expects that at high temperatures $\beta \to 0$ the spin assignments are pretty much random since the Gibs distribution is more or less uniform. However in the ferromagnetic case at low temperature and no external magnetic field one expects that there are two types of typical spin assignments: those with positive magnetization (most spins are $+1$) and those with negative magnetization (most spins are -1). The state space is in this case not very well connected and one expects that it will be difficult to sample at low temperatures.

Ising model on complete graph. This is an important special case because it leads to an exactly solvable model. It is also an important case where MCMC methods can be studied in some detail. The graph G is complete (so all $N(N-1)/2$ edges are present) and one sets $J_{vw} = J/N$ where $J > 0$ and also $h_v = h$ a constant. The scaling with N is important in order to have a well defined distribution. This

[2]For historical reasons we use the symbol $\langle - \rangle$ for the expectation $\mathbb{E}$ with respect to Gibbs distributions.

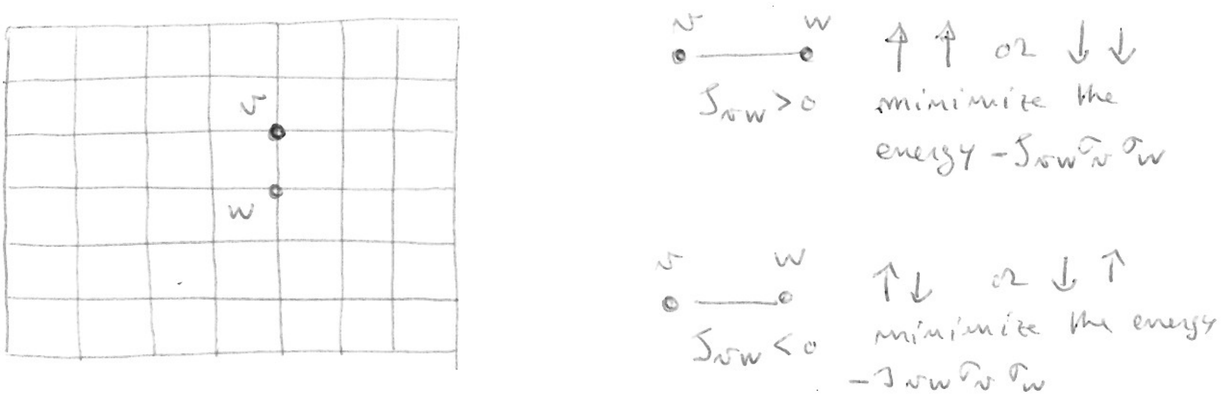

Figure 1: Square grid for the two-dimensional standard Ising model on $\{-L, \cdots, L\}^2$. Usually $J_{vw} = J$ and $h_v = h$ are constant for the standard ising model. When $J_{vw} = J > 0$ we speak of a ferromagnetic model because spin have a tendency to align in order to minimise their energy. When $J_{vw} = J < 0$ we speak of an anti-ferromagnetic model because spin have a tendency to anti-align in order to minimise their energy.

is a ferromagnetic case and again (as discussed later on) it is difficult to sample from the Gibbs measure at low temperatures. This model admits an easy solution that displays a phase transition. It is out of the scope to discuss this solution in detail here, but to better understand the behaviour of the MCMC it is useful to have the general picture in mind. Figure 2 shows the total average magnetization defined as follows:

$$m_{\pm} = \lim_{h \to 0_{\pm}} \langle \frac{1}{N} \sum_{v=1}^{N} \sigma_v \rangle \tag{9}$$

Other interpretations of the model. If you thought that this topic is specific to physics then be aware that: Ising type models and generalizations have found interpretations and applications independent from physics, e.g. in image processing, social networks, epidemiology, voter models, community detection, machine learning, coding theory etc (a non-exhaustive list). They are much studied by mathematicians, specially in probability theory, and in theoretical computer science. Let us briefly give a simplistic voter model interpretation. Each vertex v is a person that votes $\sigma_v = \pm 1$ (think of your favorite yes/no societal issue). Persons v and w related by an edge are friends ($J_{vw} > 0$) or enemies ($J_{vw} < 0$). Persons not related by an edge don't know each other. In our model the society is a bit simple minded: friends with $J_{vw} > 0$ tend to vote similarly, while enemies with $J_{vw} < 0$ tend to vote in opposite ways. The biases h_v (positive or negative) may model a prior opinion that each person has a priori and influences his decision. Sampling from $\mu(\underline{\sigma})$ would correspond to evaluate the voting pattern of the population.

1.2 Metropolis dynamics

The simplest Metropolis dynamics introduced in the previous lectures can be formulated as follows. Suppose at some step of the algorithm the spin assignment is $\underline{\sigma}$. One chose a vertex v uniformly at random (i.e. with probability $1/N$) and considers the spin assignment $\underline{\sigma}^{(v)}$ where the initial spin σ_v is flipped i.e. $\sigma_v^{(v)} = -\sigma_v$. Let

$$\begin{aligned} \Delta E(\underline{\sigma} \to \underline{\sigma}^{(v)}) &= \mathcal{H}(\underline{\sigma}^{(v)}) - \mathcal{H}(\underline{\sigma}) \\ &= 2\sigma_v \Big(\sum_w J_{vw} \sigma_w + 2h_v \Big) \end{aligned} \tag{10}$$

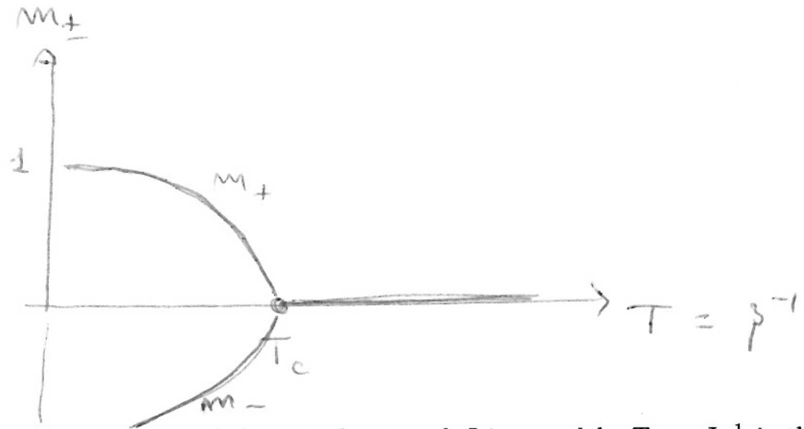

Figure 2: Magnetization of the complete graph Ising model. $T_c = J^{-1}$ is the phase transition point. Above this temperature we have $m_\pm = 0$. Below this temperature there are two possible and opposite non-zero magnetisations. They are selected by a small "symmetry breaking field" $h \to 0_\pm$.

the energy change. If the energy change is negative or zero one performs the move with probability one; while if the energy change is strictly positive one performs the move with probability $\exp(-\beta\Delta E(\underline{\sigma} \to \underline{\sigma}^{(v)})$.

This can be summarised as follows. Set

$$A(\underline{\sigma} \to \underline{\sigma}^{(v)}) = \min\bigl(1, e^{-\beta\Delta E(\underline{\sigma}\to\underline{\sigma}^{(v)})}\bigr) \tag{11}$$

The Markov chain has transition probabilities

$$p_{\underline{\sigma}\to\underline{\tau}} = \frac{1}{N}\sum_{v=1}^{N} A(\underline{\sigma} \to \underline{\sigma}^{(v)}) 1(\underline{\tau} = \underline{\sigma}^{(v)}) + \Bigl(1 - \frac{1}{N}\sum_{v=1}^{N} A(\underline{\sigma} \to \underline{\sigma}^{(v)})\Bigr) 1(\underline{\tau} = \underline{\sigma}) \tag{12}$$

Exercise. Translate this chain in the notations of left week 9. Identify the base chain and the acceptance probabilities. Check the detailed balance condition.

1.3 Glauber dynamics

This chain presents some advantages with respect to the metropolis dynamics because it is easier to analyse. For example we will see in the last chapter that it has nice monotonicity properties that allow to use it in the coupling from the past method.

A word about terminology: it is also often called heat bath dynamics or Gibbs sampling.

As before at a given time step one considers the move $\underline{\sigma} \to \underline{\sigma}^{(v)})$ where the vertex v has been chosen at random and the spin σ_v is flipped i.e. $\sigma_v^{(v)} = -\sigma_v$. Again one computes the energy change $\Delta E(\underline{\sigma} \to \underline{\sigma}^{(v)})$. One performs the move with probability

$$\frac{1}{2}\Bigl(1 - \tanh(\frac{\beta\Delta E(\underline{\sigma} \to \underline{\sigma}^{(v)})}{2})\Bigr)$$

and does not perform the move with probability

$$\frac{1}{2}\Bigl(1 + \tanh(\frac{\beta\Delta E(\underline{\sigma} \to \underline{\sigma}^{(v)})}{2})\Bigr)$$

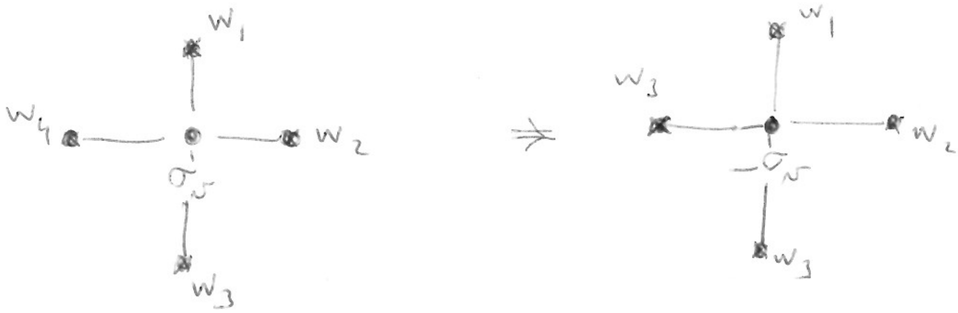

Figure 3: The energy change of a move depends only on the local neighbourhood of the selected vertex

Introduce the quantity $h_v^{loc} = h_v + \sum_w J_{vw}\sigma_w$ which is interpreted as the total "local magnetic field" at vertex v (see figure 3). The two probabilities above can be written as

$$\frac{1}{2}(1 + \sigma_v^{(v)} \tanh(\beta h_v^{(loc)})) \qquad \text{and} \qquad \frac{1}{2}(1 - \sigma_v^{(v)} \tanh(\beta h_v^{(loc)}))$$

The detailed balance condition can be checked directly from these formulas. But it can also be checked more quickly on the more general formulation explained in the next paragraph.

Simulations on a complete graph. Here we want to discuss briefly simulation results for one of the simplest situations: that of a complete graph with ferromagnetic constant interactions $J_{vw} = J/N > 0$ and zero external magnetic field $h_v = 0$. Furthermore we normalise without loss of generality $J = 1$ (this amounts to redefine the temperature scale). In this case the total average magnetisation is simply

$$\bar{m} = \langle \frac{1}{N}\sum_{v=1}^{N} \sigma_v \rangle = \frac{1}{N}\sum_{v=1}^{N} \langle \sigma_v \rangle = 0 \tag{13}$$

The first equality is a definition (of average magnetisation per variable), the second equality comes from the linearity of the expectation (with respect to the Gibbs distribution), and the third equality comes from the spin-flip symmetry of the Ising model Hamiltonian when $h_v = 0$.

We now run the Glauber dynamics starting fro a random initial condition $\underline{\sigma}(0)$ and produce a sequence of spin assignments $\underline{\sigma}(t)$, $t = 0, 1, \cdots, T$. For finite N and $T \to +\infty$ the chain will mix (since detailed balance condition is satisfied, and the chain is irreducible and aperiodic). Therefore we expect that when we run experiments, the average empirical magnetization tends to $\bar{m}$,

$$\lim_{T\to+\infty} \lim_{t\to\infty} \frac{1}{T}\int_t^{t+T} dt\, \frac{1}{N}\sum_{v=1}^{N} \sigma_v(t) = m = 0 \tag{14}$$

when $1 << t << T$. If we run the Markov chain for long enough - for finite values of N - this is what we observe. However a more detailed view of the instantaneous magnetization defined as

$$m(t) = \frac{1}{N}\sum_{v=1}^{N} \sigma_v(t) \tag{15}$$

is very instructive. Figure 4 shows three plots of the fluctuations of this object as a function of time for three inverse temperatures $\beta = 0.8$ (above the critical temperature), $\beta = 1$ (at the critical temperature), $\beta = 1.2$ (below the critical temperature).

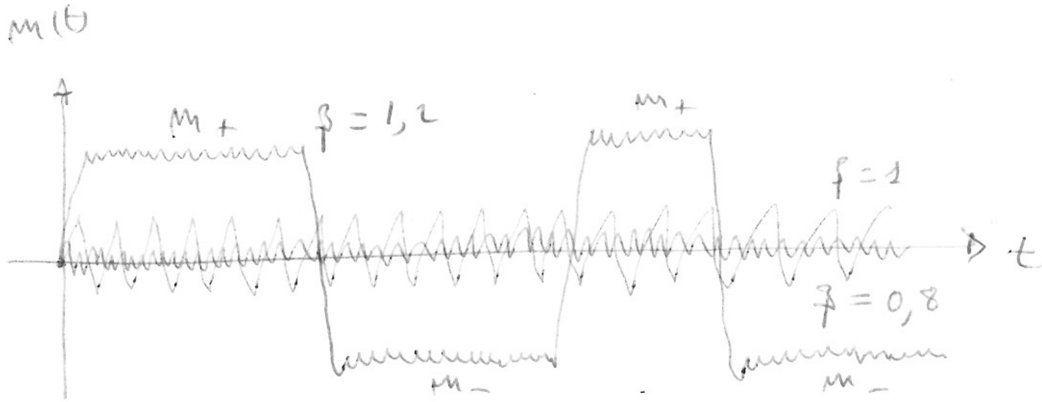

Figure 4: Empirical magnetization $m(t)$ as a function of time for three different temperatures when computed by Glauber dynamics (initialisation is random).

We observe that for $\beta = 0.8$ the curve has stationary fluctuations around zero. At the critical temperature $\beta = 1$ the curve still fluctuates around zero, however the fluctuations are markedly higher. This is a signal of the phase transition. For $\beta = 1.2$ we observe that the chain makes transition between two typical states where the magnetization is non zero (positive or negative) for a very long time. On average the total magnetization is zero, but for long times the chain remains stuck in states of positive and negative magnetisations.

These numerical observations show that the mixing time behaves very differently above and below the critical temperature. It is known that above the critical temperature (so $\beta < 1$) the mixing time behaves as $O(N^2 \log N)$; but below the critical temperature (so $\beta > 1$) it behaves as $O(e^N)$. Therefore although the ergodic theorem holds for finite N, it is not very relevant for the "practical" situation of a very large system that we can simulate or observe for a large but finite time. This limit is captured by $N \to +\infty$ first and $T \to \infty$ after. In this limit, at low temperatures the Markov chain remains stuck in part of the state space (the state space effectively becomes disconnected) and we say that ergodicity is broken. Mathematically speaking in (14) we cannot exchange limits of large time and large system ($N \to +\infty$) do not commute.

1.4 General Gibbs sampling (or generalisation of Glauber)

Here we give a more conceptual from of the Glauber dynamics. The application to the Ising model yields the rules of the previous paragraph.

Let $\pi(\underline{x})$ a probability distribution over states $\underline{x} = (x_1, \cdots, x_N)$ where $x_j \in \mathcal{X}$. Thus the state space is $\mathcal{X}^N$. The sampler is constructed as follows:

- Take $\underline{x} \in \mathcal{X}^N$.
- Select vertex $v \in V$ uniformly at random.
- Compute the probability of $\underline{y}$ conditional on $\{y_w = x_w, w \neq v\}$,

$$\mathbb{P}(\underline{y} \mid \{y_w = x_w, w \neq v\}) = \frac{\pi(x_1, \cdots x_{v-1}, y_v, x_{v+1}, \cdots, x_N)}{\sum_{y_v \in \mathcal{X}} \pi(x_1, \cdots x_{v-1}, y_v, x_{v+1}, \cdots, x_N)} \tag{16}$$

- Make the move $\underline{x} \to \underline{y}$ with probability

$$\mathbb{P}(\underline{y} \mid \{y_w = x_w, w \neq v\}) \tag{17}$$

- Go to first item and iterate.

Summarizing, the transition probability of this chain is given by:

$$p_{\underline{x}\to\underline{y}} = \begin{cases} \frac{1}{N}\frac{\pi(x_1,\cdots x_{v-1},y_v,x_{v+1},\cdots,x_N)}{\sum_{y_v\in\mathcal{X}}\pi(x_1,\cdots x_{v-1},y_v,x_{v+1},\cdots,x_N)}, & \text{if } \underline{y} = (x_1,\cdots,x_{v-1},x_v,x_{v-1},\cdots,x_N) \text{ for } v\in\{1,\cdots,N\},\\ 0 & \text{otherwise.}\end{cases}$$

Exercise: From the formula for the transition probability check the detailed balance condition $\pi(\underline{x})p_{\underline{x}\to\underline{y}} = \pi(\underline{y})p_{\underline{y}\to\underline{x}}$ for all pairs of states $\underline{x}$ and $\underline{y}$. (see solution of homework set 10).

Here we verify that this chain reduces to the usual Glauber dynamics in the case of the Ising model. Taking $\mu(\underline{\sigma})$ for $\pi(\underline{x})$ and start at a configuration $\underline{\sigma}$. We consider a move $\underline{\sigma}\to\underline{\tau}$ where $\tau_w = \sigma_w$ for $w\neq v$ (and $\tau_v = \pm\sigma_v$). Inspection of the Ising Hamiltonian yields:

$$\mu(\underline{\tau}\mid\{\tau_w=\sigma_v, w\neq v\}) = \frac{e^{\beta\tau_v(\sum_w J_{vw}\sigma_w+h_v)}e^{\text{terms indep of }\sigma_v}}{\sum_{\tau_v=\pm1}e^{\beta\tau_v(\sum_w J_{vw}\sigma_w+h_v)}e^{\text{terms indep of }\sigma_v}} \tag{18}$$

The terms independent of σ_v in the numerator and denominator are the same and can be simplified. Performing the sum over τ_v in the denominator we find

$$\begin{aligned}\mu(\underline{\tau}\mid\{\tau_w=\sigma_v, w\neq v\}) &= \frac{e^{\tau_v(\sum_w \beta J_{vw}\sigma_w+\beta h_v)}}{2\cosh(\sum_w \beta J_{vw}\sigma_w+\beta h_v)}\\ &= \frac{1}{2}(1+\tau_v\tanh(\beta h_v+\sum_w \beta J_{vw}\sigma_w)\\ &= \frac{1}{2}(1+\tau_v\tanh(\beta h_v^{loc})\end{aligned} \tag{19}$$