RBM with continuous hidden layer variables

July 14, 2022

1 Relevant literature

1. This paper from high-energy physics derives a similar auxiliary field x and free energy $F(x)$ that could be used for sampling an Ising model. Apparently they were not aware of the microscopic Hamiltonian $H_0(\sigma, x)$ that can be used to sample Ising spins σ. They also missed some significant benefits of moving to continuous variables. A continuous HMC or Langevin sampling method could get huge benefits from FFT for long-range interactions. Also, perhaps "dynamical Fourier acceleration" could give a similar benefit to cluster moves. `https://arxiv.org/abs/1912.03278`. *Ostmeyer et al., "The Ising Model with Hybrid Monte Carlo," CPC 2019.*

2. This earlier paper from NeurIPS got a ton more visibility (50 citations) and seems to be the first use of HS transformation to map from Ising to continuous fields. Our mathematical results mostly reproduce theirs. *Yichuan Zhang et al, "Continuous Relaxations for Discrete Hamiltonian Monte Carlo" NeurIPS 2012.* `https://proceedings.neurips.cc/paper/2012/hash/c913303f392ffc643f7240b180602652-Abstract.html`

3. Some similar ideas appear here: *Pakman and Paninski "Auxiliary-variable Exact Hamiltonian Monte Carlo Samplers for Binary Distributions," NeurIPS 2013* `https://proceedings.neurips.cc/paper/2013/hash/a7d8ae4569120b5bec12e7b6e9648b86-Ab` `html`

4. Another relevant concept to us seems to be "slice sampling" which introduces an auxiliary field (i.e. "hidden layer"), and then does alternating Gibbs sampling on physical and auxiliary fields. This paper connects HMC with slice sampling: *Yizhe Zhang et al., "Towards Unifying Hamiltonian Monte Carlo and Slice Sampling," NeurIPS 2016.* `https://proceedings.neurips.cc/paper/2016/hash/3cef96dcc9b8035d23f69e30bb19218a-Abstract.html`.

2 Mapping of the Ising model

Consider an Ising model Hamiltonian,

$$H = \sum_{ij} J_{ij}\sigma_i\sigma_j + \sum_i b_i\sigma_i. \tag{1}$$

In matrix notation,

$$H = \sigma^T J\sigma + b^T\sigma, \tag{2}$$

where J is a symmetric matrix. Our purpose is to sample σ from the distribution $P(\sigma) \propto \exp(-\beta H)$.

Extend the configuration space to include a vector of auxiliary ("hidden") variables x in addition to the usual ("visible") spins σ. Define a "microscopic" Hamiltonian that linearly couples σ and x,

$$H_0 = 2x^T W\sigma + b^T\sigma + x^T x. \tag{3}$$

The joint-probability for the full configuration (σ, x) is,

$$P(\sigma, x) \propto e^{-\beta H_0}. \tag{4}$$

Upon integrating out x, we should find the correct Boltzmann distribution

$$e^{-\beta H} \propto \int e^{-\beta H_0} dx, \tag{5}$$

where $dx = (dx_1 dx_2 \ldots)$. It remains to determine the weight matrix W such that Eq. (5) is satisfied. We expand the right-hand side of Eq. (5),

$$e^{-\beta H} \propto \int e^{-\beta(2x^T W\sigma + b^T\sigma + x^T x)} dx. \tag{6}$$

Since integration runs over all x, we are free to shift

$$x \to x - W\sigma. \tag{7}$$

The constraint becomes

$$e^{-\beta\sigma^T J\sigma} \propto \int e^{-\beta A} dx, \tag{8}$$

where

$$A = 2(x - W\sigma)^T W\sigma + (x - W\sigma)^T(x - W\sigma). \tag{9}$$

Expanding in powers of σ, we find

$$\begin{aligned} A = {} & \sigma^T\left(-2W^T W + W^T W\right)\sigma \\ & + 2x^T W\sigma - x^T W\sigma - \sigma^T W^T x \\ & + x^T x. \end{aligned} \tag{10}$$

The terms linear in x cancel because scalars are invariant under transpose, $x^T W\sigma = \sigma^T W^T x$, leaving

$$A = -\sigma^T W^T W\sigma + x^T x. \tag{11}$$

The constraint becomes,

$$e^{-\beta\sigma^T J\sigma} \propto e^{+\beta\sigma^T W^T W\sigma} \int e^{-\beta x^T x} dx. \tag{12}$$

The final integral is an irrelevant constant, so we are left with,

$$\sigma^T W^T W\sigma = -\sigma^T J\sigma + \text{const.} \tag{13}$$

This equation can be satisfied by selecting the symmetric weight matrix,

$$W = W^T = \sqrt{cI - J}, \tag{14}$$

where the constant c should be sufficiently large to ensure that $cI - J$ is semi-positive definite. For example, $c = \text{eigmax}(J)$. Note that $\sigma^T\sigma$ is a constant because each Ising spin is $\sigma_i = \pm 1$.

Typically the couplings J_{ij} will be translation invariant. I.e., the matrix elements only depend on the displacement between sites i and j. Then J and W diagonalize in Fourier space. We can efficiently apply W to a vector as follows: (1) Use an FFT to transform the vector into Fourier space, (2) multiply the vector components by a diagonal matrix containing the eigenvalues of W, and (3) transform back to real-space using an inverse FFT.

3 Direct sampling of auxiliary field x

We can calculate an effective free energy $F(x)$ by summing over all possible σ,

$$P(x) \propto e^{-\beta F} \propto \sum_{\{\sigma\}} e^{-\beta H_0}.$$

This may be written,

$$e^{-\beta F} = e^{-\beta x^T x} \sum_{\{\sigma\}} e^{-\beta B^T \sigma},$$

where

$$B = 2W^T x + b,$$

can be viewed as an effective field. We may calculate,

$$\begin{aligned}\sum_{\{\sigma\}} e^{-\beta B^T\sigma} &= (\sum_{\sigma_1} \cdots \sum_{\sigma_N}) \prod_i e^{-\beta B_i \sigma_i} \\ &= \prod_i \sum_{\sigma_i = \pm 1} e^{-\beta B_i \sigma_i} \\ &\propto \prod_i \cosh(\beta B_i).\end{aligned}$$

Then

$$F = x^T x - \beta^{-1} \sum_i \log\left(\cosh(\beta B_i)\right).$$

We can use this free energy to sample auxiliary fields x without the presence of σ. Once x has been decorrelated, we can immediately sample σ from the joint distribution $P(\sigma|x) \propto \exp(-\beta H_0)$, where x is held fixed.

To perform Langevin or HMC type sampling on x, we will need to calculate the gradient,

$$\frac{\partial F}{\partial x_j} = 2x_j - \beta^{-1} \sum_i \frac{\partial}{\partial x_j} \log\left(\cosh(\beta B_i)\right).$$

Note that

$$\frac{\partial}{\partial x_j} \log\left(\cosh(\beta B_i)\right) = \beta \tanh(\beta B_i) \frac{\partial B_i}{\partial x_j}.$$

Let e_i denote the unit vector (a single nonzero element at index i) such that $B_i = e_i^T B$. Then

$$\frac{\partial B_i}{\partial x_j} = 2e_i^T W^T e_j = 2W_{ji}.$$

The force becomes

$$\frac{\partial F}{\partial x_j} = 2x_j - \sum_i W_{ji} \tanh(\beta B_i).$$

Our final expression simplifies in matrix notation,

$$\frac{\partial F}{\partial x} = 2x - Wv,$$

where the vector v has elements

$$v_i = \tanh(\beta B_i).$$