A dynamic minimum probability flow approach to infer parallel Glauber updates

Luca Raffo

EPFL, Institute of Mathematics - `luca.raffo@epfl.ch`

Capital Fund Management - `luca.raffo@cfm.com`

October 2025

Abstract

There exist many different algorithms to infer the parameters of an Ising model given some i.i.d. samples, the most widely used being maximum pseudolikelihood and interaction screening. Firstly, we show that the latter is a specific case of a more general inference method, called minimum probability flow. Then, we extend this inference method to the non-static samples regime, in which we assume that our data have been generated by an underlying parallel Glauber dynamics. The most widely used inference method for this setting is maximum likelihood, and we show that our dynamic formulation of minimum probability flow provides better inferences in generic regimes.

Contents

1 Introduction

The Ising model is one of the most fundamental models in statistical physics, originally introduced to describe collective phenomena such as ferromagnetism [2, 1].

A sample from this model consists of a system of `n` binary variables, denoted as spins, $s_i \in \{-1, 1\}$, where i ranges from 1 to `n`. The probability of a configuration, for given `inverse temperature` β, `magnetic field` h and symmetric `coupling matrix` J is

$$\mathbb{P}_{J,h,\beta}(s) = \frac{1}{Z_{J,h,\beta}} \exp\left[\beta\left(\sum_{i<j} J_{ij} s_i s_j + \sum_i h_i s_i\right)\right], \tag{1}$$

where $Z_{J,h,\beta}$ is the constant that ensures normalization. To ease notation, throughout this manuscript we will assume that $\beta = 1$ and h is the zero vector, so that our probability mass function will be

$$\mathbb{P}_J(s) = \frac{1}{Z_J} \exp\left[\left(\sum_{i<j} J_{ij} s_i s_j\right)\right]. \tag{2}$$

Nonetheless our arguments hold the most general settings, at the cost of a more cumbersome notation.

Despite its apparent simplicity, the Ising model captures rich behaviors such as phase transitions and collective alignment, and it serves as a canonical framework for studying high-dimensional inference, energy-based models, and network interactions [1, 8, 4, 6].

Notice that, in order to generate samples from (2), we cannot rely on a brute-force approach, that is, explicitly evaluating all the $2^{\mathtt{n}}$ possible configurations, because already in systems of size $\mathtt{n} \approx 50$ this requires an inaccessible amount of time to be computed.

In practice, samples from (2) are obtained by running a Markov Chain Monte Carlo (MCMC) procedure whose stationary distribution coincides with the Ising measure. Starting from an initial random configuration $s^{(0)} \in \{-1, +1\}^n$, the system is evolved according to a transition rule that preserves the target distribution. After a sufficient number of iterations, the configuration $s^{(T)}$ can be regarded as an approximate sample from the Ising model.

A common choice for this transition rule is the parallel Glauber dynamics. In this scheme, all spins are updated simultaneously according to their conditional distributions given the current configuration. Given $s^{(t)} = (s_1^{(t)}, \dots, s_n^{(t)})$, each spin s_i is resampled independently from

$$\mathbb{P}(s_i^{(t+1)} = +1 \mid s^{(t)}) = \frac{1}{1 + \exp\left(-2\,\theta_i^{(t)}\right)}, \tag{3}$$

where the `effective local field` acting on spin i is

$$\theta_i^{(t)} = \sum_{j \neq i} J_{ij}\, s_j^{(t)}. \tag{4}$$

Iterating this update rule defines a Markov chain $s^{(t+1)} = F(s^{(t)})$ whose invariant distribution is the Ising measure (2).

Remarkably, in case the `coupling matrix` we use during the updates is not symmetric, the stationary distribution does not have the form of (2), and we call it a non equilibrium steady-state (or NESS). Unlike the sequential Glauber dynamics, which updates one spin at a time, the parallel version performs all updates simultaneously, enabling efficient vectorized implementations on modern hardware.

In static inference problems, we assume to have access to some i.i.d. samples from (2), and our objective is to reconstruct the `coupling matrix` J.

In dynamic inference problems, we assume to have access to a trajectory given by a parallel Glauber dynamics, and we want to infer the matrix J ruling these updates.

2 Static regime

The natural approach for any inference problem is maximum likelihood. In short, we look for the parameters J that make the observed data as probable as possible under the model distribution. Formally, given a set of independent samples $\mathcal{D} = \{s^{(1)}, ..., s^{(M)}\}$, drawn from an unknown Ising measure $\mathbb{P}_{J^\star}$, we define the negative log-likelihood function

$$\ell_{NLL}(J) = -\frac{1}{M}\sum_{m=1}^{M} \log \mathbb{P}_J(s^{(m)}), \tag{5}$$

where $\mathbb{P}_J(s)$ is given by (2). Expanding the expression, we obtain

$$\ell_{NLL}(J) = -\frac{1}{M}\sum_{m=1}^{M}\left[\sum_{i<j} J_{ij} s_i^{(m)} s_j^{(m)} - \log Z_J\right]. \tag{6}$$

The optimal couplings are thus obtained by minimizing (6), or equivalently by minimizing the Kullback–Leibler divergence between the empirical distribution of the data and the model:

$$\arg\min_J \mathrm{KL}(\hat{p} \,\|\, \mathbb{P}_J)\,, \qquad \hat{p}(s) = \frac{1}{M}\sum_{m=1}^{M} \delta_{s^{(m)}}.$$

Taking derivatives of (6) with respect to J_{ij} yields the classical *moment matching condition*

$$\frac{\partial \ell_{NLL}(J)}{\partial J_{ij}} = -\frac{1}{M}\sum_{m=1}^{M} s_i^{(m)} s_j^{(m)} + \mathbb{E}_{\mathbb{P}_J}[s_i s_j] = 0. \tag{7}$$

This expression shows that maximum likelihood estimation enforces the model correlations $\mathbb{E}_{\mathbb{P}_J}[s_i s_j]$ to match the empirical ones computed from the data.

However, the second term in (7) is generally intractable. Computing

$$\mathbb{E}_{\mathbb{P}_J}[s_i s_j] = \frac{1}{Z_J}\sum_{s\in\{-1,+1\}^n} s_i s_j \exp\left(\sum_{k<\ell} J_{k\ell} s_k s_\ell\right)$$

requires summing over all 2^n possible configurations. Even for moderate system sizes, this enumeration becomes computationally impossible.

In principle, one could approximate these expectations using Monte Carlo methods—sampling configurations from $\mathbb{P}_J$ through a Markov Chain Monte Carlo algorithm such as Glauber dynamics or Metropolis–Hastings (this is called Boltzmann machine learning). Yet, this introduces a nested sampling problem: at every gradient step, the chain must be run until approximate equilibrium, dramatically increasing computational cost.

Consequently, direct maximum likelihood inference is practically infeasible for all but the smallest systems. This motivates the use of surrogate estimators that avoid the computation of the partition function Z_J, such as the *maximum pseudolikelihood* and the *minimum probability flow* methods, which we discuss next.

2.1 Maximum pseudolikelihood

The maximum pseudolikelihood estimator (MPL) was introduced as a tractable alternative to the full likelihood, replacing the joint probability $\mathbb{P}_J(s)$ by the product of its conditional distributions. Formally, we define the *negative pseudolikelihood* as

$$-\prod_{m=1}^{M}\prod_{i=1}^{\mathtt{n}} \mathbb{P}_J\left(s_i^{(m)} \,\middle|\, s_{\setminus i}^{(m)}\right), \tag{8}$$

where $s_{\setminus i}$ denotes all spins except s_i. Minimizing (8) yields the *maximum pseudolikelihood estimator*. For the Ising model (2), the conditional probability of a single spin s_i given all others is available in closed form:

$$\mathbb{P}_J(s_i = +1 \mid s_{\setminus i}) = \frac{1}{1+\exp(-2\,\theta_i)}, \qquad \text{where} \quad \theta_i = \sum_{j\neq i} J_{ij}s_j. \tag{9}$$

Hence,

$$\mathbb{P}_J(s_i \mid s_{\setminus i}) = \frac{\exp(s_i\theta_i)}{2\cosh(\theta_i)}.$$

Substituting this expression into (8) and taking the logarithm gives the negative pseudolog-likelihood:

$$\ell_{\mathrm{NPL}}(J) = -\frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{\mathtt{n}}\left[s_i^{(m)}\,\theta_i^{(m)}(J) - \log\Big(2\cosh\Big(\theta_i^{(m)}(J)\Big)\Big)\right], \tag{10}$$

which can be reformulated as

$$\ell_{\text{NPL}}(J) = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{\mathtt{n}} \log\left(1 + e^{-2s_i^{(m)} \theta_i^{(m)}(J)}\right). \tag{11}$$

The derivative of ℓ_{NPL} with respect to J_{ij} can be computed explicitly and it is a sum of local, tractable terms involving only observed quantities and the current parameters J.

2.2 Interaction screening

The *Interaction Screening Estimator* (ISE), introduced in [3], provides an efficient and statistically optimal approach to infer the couplings J of the Ising model without evaluating the partition function Z_J. The key idea is to construct a convex objective that vanishes at the true parameters by enforcing a local stationarity condition for each spin, effectively "screening out" indirect interactions mediated through other nodes.
Starting from the conditional probability of a single spin given all others,

$$\mathbb{P}_J(s_i \mid s_{\setminus i}) = \frac{\exp\left(s_i \sum_{j \neq i} J_{ij} s_j\right)}{2 \cosh\left(\sum_{j \neq i} J_{ij} s_j\right)},$$

its conditional mean is

$$\mathbb{E}_{\mathbb{P}_J}[s_i \mid s_{\setminus i}] = \tanh\left(\sum_{j \neq i} J_{ij} s_j\right).$$

At equilibrium, the true parameters $J^\star$ satisfy the self-consistency relation

$$\mathbb{E}_{\mathbb{P}_{J^\star}}\left[s_i \exp\left(-s_i \sum_{j \neq i} J^\star_{ij} s_j\right)\right] = 0.$$

Replacing the expectation with its empirical counterpart yields the *interaction screening loss* for node i:

$$\ell_{IS}^{(i)}(J_i) = \frac{1}{M} \sum_{m=1}^{M} \exp\left(-s_i^{(m)} \sum_{j \neq i} J_{ij} s_j^{(m)}\right), \tag{12}$$

where $J_i = (J_{i1}, \ldots, J_{i,i-1}, J_{i,i+1}, \ldots, J_{i\mathtt{n}})$ are the couplings incident to spin i. The Interaction Screening Estimator is then defined as the minimizer of this convex objective:

$$\arg\min_{J_i} \ell_{IS}^{(i)}(J_i),$$

where the ℓ_1-regularization promotes sparsity in the inferred interactions. Each optimization problem is independent across spins and can be efficiently solved with standard convex solvers such as coordinate descent or gradient-based methods.

2.3 Static minimum probability flow

The *Minimum Probability Flow* (MPF) estimator [7] provides an elegant alternative to likelihood-based inference by entirely bypassing the computation of the partition function Z_J. The main idea is to define a continuous-time dynamics over the configuration space whose stationary distribution coincides with the Ising measure $\mathbb{P}_J$, and to adjust the parameters J such that the empirical distribution of the data remains approximately invariant under this dynamics at initial time.

Formally, let $\mathbb{P}_J(s)$ be the target Ising distribution (2) and let $\hat{p}(s)$ denote the empirical distribution associated with the dataset $\mathcal{D} = \{s^{(1)}, \ldots, s^{(M)}\}$:

$$\hat{p}(s) = \frac{1}{M} \sum_{m=1}^{M} \delta_{s^{(m)}}.$$

We consider a continuous-time Markov process with transition rates $W_{s \to s'}(J)$ that preserve detailed balance with respect to $\mathbb{P}_J$:

$$\frac{W_{s\to s'}(J)}{W_{s'\to s}(J)} = \frac{\mathbb{P}_J(s')}{\mathbb{P}_J(s)} = \exp\Big(E_J(s) - E_J(s')\Big), \qquad E_J(s) = -\sum_{i<j} J_{ij} s_i s_j.$$

The time evolution of a probability distribution $p_t(s)$ under this dynamics is governed by the master equation

$$\frac{dp_t(s)}{dt} = \sum_{s'} \big[W_{s'\to s}(J) p_t(s') - W_{s\to s'}(J) p_t(s)\big].$$

The MPF approach penalizes the initial *probability flow* from the empirical distribution $\hat{p}$ toward configurations not observed in the data. The objective function is thus defined as the instantaneous rate of change of the Kullback–Leibler divergence between p_t and $\hat{p}$, evaluated at $t = 0$:

$$\ell_{MPF}(J) = \frac{d}{dt}\mathrm{KL}(\hat{p} \,\|\, p_t)\bigg|_{t=0} = \sum_{s\in\mathcal{D}} \sum_{s'\notin\mathcal{D}} \hat{p}(s)\, W_{s\to s'}(J). \tag{13}$$

Minimizing $\ell_{MPF}(J)$ forces the probability flow out of the data manifold to vanish, ensuring that $\hat{p}$ is approximately stationary under the dynamics induced by J. Hence, the MPF estimator is defined as

$$\arg\min_J \ell_{MPF}(J).$$

In the Ising model, a natural and efficient choice of connectivity restricts transitions to single-spin flips:

$$s' = F_i(s) \quad \text{such that} \quad s'_i = -s_i, \quad s'_j = s_j \text{ for } j \neq i.$$

Using this structure, the transition rate is taken as

$$W_{s\to F_i(s)}(J) = \exp\left(-\frac{1}{2}\big[E_J(F_i(s)) - E_J(s)\big]\right) = \exp\left(-s_i \sum_{j\neq i} J_{ij} s_j\right). \tag{14}$$

Substituting (14) into the definition of $\ell_{MPF}(J)$ yields a fully tractable objective:

$$\ell_{MPF}(J) = \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{\mathtt{n}} \exp\Big(-s_i^{(m)}\theta_i^{(m)}\Big). \tag{15}$$

This loss depends only on observed spins and current parameters J, without any normalization constant or sampling step. Taking derivatives of (15) provides an explicit and efficiently computable gradient, which can be used directly in gradient-based optimization schemes.

Interestingly, the MPF loss (15) can be written as a sum of node-wise terms,

$$\ell_{MPF}(J) = \sum_{i=1}^{\mathtt{n}} \ell_{IS}^{(i)}(J_i),$$

where each component $\ell_{IS}^{(i)}(J_i)$ coincides with the interaction screening objective introduced in (12). Consequently, performing gradient descent on $\ell_{MPF}(J)$ in a *spin-by-spin* (or node-wise) fashion is equivalent to minimizing the corresponding Interaction Screening loss for each spin independently.

3 Dynamic regime

In the *dynamic regime*, we are no longer given independent equilibrium samples from the Ising measure, but rather an entire trajectory generated by a stochastic dynamics whose stationary distribution is (or is expected to be) of the form (2). Specifically, we consider a sequence of spin configurations

$$s^{(0)}, s^{(1)}, \dots, s^{(T)},$$

obtained by evolving an initial state $s^{(0)}$ under the *parallel Glauber dynamics* introduced earlier. At each time step, all spins are simultaneously resampled according to their conditional probabilities given the current configuration:

$$\mathbb{P}_J\big(s_i^{(t+1)} = +1 \,\big|\, s^{(t)}\big) = \frac{1}{1 + \exp\big(-2\,\theta_i^{(t)}(J)\big)}, \qquad \theta_i^{(t)}(J) = \sum_{j \neq i} J_{ij} s_j^{(t)}.$$

Our goal in the dynamic setting is to infer the coupling matrix J that best explains the observed sequence of transitions under this update rule.

3.1 Maximum likelihood

Since spins are updated independently given $s^{(t)}$, the transition probability factorizes as

$$\mathbb{P}_J\big(s^{(t+1)} \mid s^{(t)}\big) = \prod_{i=1}^{\mathtt{n}} \mathbb{P}_J\big(s_i^{(t+1)} \mid s^{(t)}\big) = \prod_{i=1}^{\mathtt{n}} \frac{\exp\big(s_i^{(t+1)} \theta_i^{(t)}(J)\big)}{2\cosh\big(\theta_i^{(t)}(J)\big)}.$$

The likelihood of an observed trajectory is then

$$\prod_{t=0}^{T-1} \mathbb{P}_J\big(s^{(t+1)} \mid s^{(t)}\big),$$

and its negative log-likelihood reads

$$\ell_{\text{D-NLL}}(J) = -\frac{1}{T} \sum_{t=0}^{T-1} \sum_{i=1}^{\mathtt{n}} \left[s_i^{(t+1)}\, \theta_i^{(t)}(J) - \log\Big(2\cosh\big(\theta_i^{(t)}(J)\big)\Big) \right], \tag{16}$$

and it can be reformulated with

$$\ell_{\text{D-NLL}}(J) = \frac{1}{T} \sum_{t=0}^{T-1} \sum_{i=1}^{\mathtt{n}} \log\Big(1 + e^{-2\, s_i^{(t+1)}\, \theta_i^{(t)}(J)}\Big). \tag{17}$$

Minimizing (16) yields the *dynamic maximum likelihood estimator*, which identifies the coupling matrix J that makes the observed sequence of Glauber updates most probable under the model. Notably, the loss (16) has the same local structure as the static pseudolikelihood (11), but now the "responses" $s_i^{(t+1)}$ correspond to the spins at the next time step rather than the current configuration.

3.2 Dynamic minimum probability flow

The Minimum Probability Flow (MPF) principle can be extended to the dynamic setting, yielding what we call the conditional or dynamic MPF estimator.

The idea is to enforce that the empirical conditional distribution of spin transitions remains approximately invariant under an infinitesimal dynamics parameterized by J. Given an observed trajectory $\{s^{(t)}\}_{t=0}^{T}$, for each spin i we define the empirical conditional law

$$\hat{p}_i(y_i \mid x) = \frac{1}{T}\sum_{t=0}^{T-1} \delta_{s^{(t)},x}\,\delta_{s_i^{(t+1)},y_i}, \qquad x = s^{(t)} \in \{\pm 1\}^n, \quad y_i \in \{\pm 1\}.$$

We assume that the true update rule follows the logistic form

$$p_{J,i}(y_i \mid x) = \frac{\exp(y_i\,\theta_{i,J}(x))}{2\cosh(\theta_{i,J}(x))}, \qquad \theta_{i,J}(x) = \sum_{j\neq i} J_{ij}\,x_j.$$

For each fixed configuration x, consider a continuous-time Markov process on the binary variable $y_i \in \{-1,+1\}$ with transition rates $g_{J,i}(y_i' \mid y_i; x)$ satisfying detailed balance with respect to $p_{J,i}(y_i \mid x)$:

$$g_{J,i}(y_i' \mid y_i; x)\,p_{J,i}(y_i \mid x) = g_{J,i}(y_i \mid y_i'; x)\,p_{J,i}(y_i' \mid x). \tag{18}$$

Following [7], we choose the symmetric generator

$$g_{J,i}(y_i' \mid y_i; x) = \exp\!\left[-\tfrac{1}{2}\big(E_{J,i}(y_i' \mid x) - E_{J,i}(y_i \mid x)\big)\right], \qquad E_{J,i}(y_i \mid x) = -\,y_i\,\theta_{i,J}(x),$$

which, for binary y_i, yields the single transition rate

$$g_{J,i}(-y_i \mid y_i; x) = \exp[-\,y_i\,\theta_{i,J}(x)]. \tag{19}$$

Let $p_{t,i}(y_i \mid x)$ denote the conditional distribution at time t evolving according to this generator, with initial condition $p_{0,i}(y_i \mid x) = \hat{p}_i(y_i \mid x)$. The dynamic MPF objective is defined as the initial time derivative of the conditional Kullback–Leibler divergence:

$$\ell_{\text{D-MPF}}(J) = \sum_{i=1}^{\mathtt{n}} \frac{d}{dt} \mathbb{E}_{x\sim\hat{p}(x)} \mathrm{KL}\big(\hat{p}_i(\cdot \mid x) \,\|\, p_{t,i}(\cdot \mid x)\big)\Big|_{t=0}. \tag{20}$$

We now compute this derivative explicitly. For fixed i and x,

$$\mathrm{KL}\big(\hat{p}_i(\cdot \mid x)\|p_{t,i}(\cdot \mid x)\big) = \sum_{y_i=\pm 1} \hat{p}_i(y_i \mid x) \log \frac{\hat{p}_i(y_i \mid x)}{p_{t,i}(y_i \mid x)}.$$

Differentiating with respect to t and evaluating at $t = 0$,

$$\frac{d}{dt}\mathrm{KL}\Big|_{t=0} = -\sum_{y_i} \frac{\dot{p}_{t,i}(y_i \mid x)|_{t=0}}{p_{0,i}(y_i \mid x)} \hat{p}_i(y_i \mid x) = -\sum_{y_i} \dot{p}_{t,i}(y_i \mid x)|_{t=0},$$

where the last equality uses $p_{0,i} = \hat{p}_i$. The master equation for the conditional process reads

$$\dot{p}_{t,i}(y_i \mid x) = \sum_{y_i'} \big[g_{J,i}(y_i \mid y_i'; x) p_{t,i}(y_i' \mid x) - g_{J,i}(y_i' \mid y_i; x) p_{t,i}(y_i \mid x)\big].$$

Substituting $p_{0,i} = \hat{p}_i$ and summing over y_i, we obtain

$$\frac{d}{dt}\mathrm{KL}\Big|_{t=0} = \sum_{y_i} \hat{p}_i(y_i \mid x) \sum_{y_i' \neq y_i} g_{J,i}(y_i' \mid y_i; x).$$

Averaging over $x \sim \hat{p}(x)$ and summing over all spins i yields

$$\ell_{\text{D-MPF}}(J) = \sum_{i=1}^{\mathtt{n}} \mathbb{E}_{(x,y_i)\sim\hat{p}_i}\, g_{J,i}(-y_i \mid y_i; x) = \sum_{i=1}^{\mathtt{n}} \mathbb{E}_{(x,y_i)\sim\hat{p}_i}\, \exp[-\, y_i\, \theta_{i,J}(x)]. \tag{21}$$

Replacing the expectation by the empirical average over the observed trajectory gives

$$\ell_{\text{D-MPF}}(J) = \frac{1}{T} \sum_{t=0}^{T-1} \sum_{i=1}^{\mathtt{n}} \exp[-\, s_i^{(t+1)}\, \theta_i^{(t)}(J)]. \tag{22}$$

Each row $J_{i\cdot}$ can thus be optimized independently, leading to a set of convex node-wise estimation problems analogous to those of the static MPF or Interaction Screening estimators.

The four loss functions we derived show strong simmetries between them.

Static objectives	
Logistic (NPL)	$\ell_{\mathrm{NPL}}(J) = \dfrac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{\mathtt{n}} \log\bigl(1 + e^{-2\, s_i^{(m)}\, \theta_i^{(m)}(J)}\bigr)$
Exponential (MPF)	$\ell_{\mathrm{MPF}}(J) = \dfrac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{\mathtt{n}} \exp\bigl(-\ s_i^{(m)}\, \theta_i^{(m)}(J)\bigr)$

Table 1: Static: logistic (pseudolikelihood) vs. exponential (MPF).

Dynamic objectives	
Logistic (D-NLL)	$\ell_{\text{D-NLL}}(J) = \dfrac{1}{T}\sum_{t=0}^{T-1}\sum_{i=1}^{\mathtt{n}} \log\bigl(1 + e^{-2\, s_i^{(t+1)}\, \theta_i^{(t)}(J)}\bigr)$
Exponential (D-MPF)	$\ell_{\text{D-MPF}}(J) = \dfrac{1}{T}\sum_{t=0}^{T-1}\sum_{i=1}^{\mathtt{n}} \exp\bigl(-\ s_i^{(t+1)}\, \theta_i^{(t)}(J)\bigr)$

Table 2: Dynamic: logistic (MLE) vs. exponential (conditional MPF).

4 Numerical experiments

We compare the performance of our dynamic algorithms in the following settings.

1. Firstly, we generate an asymmetric matrix J whose entries are uniformly ditributed in $[-\beta, \beta]$, with $\beta = 3.0$.
2. Then we generate a trajectory whose length is 500.
3. We plot the evolution of the reconstruction's error as we increase the number of data used in the inferences. The reconstruction's error is

defined as $\frac{\|J-\hat{J}\|_2}{\|J\|_2} * \frac{1}{\mathtt{n}}$ (accordingly with [5]). Both MLE and MPF are implemented with gradient descent, whose parameters are fixed to $\mathtt{lr} = 0.1$, $\mathtt{mle\ steps} = 30000$, $\mathtt{mpf\ steps} = 10000$.

We repeated this for 50 runs and we plot below the average reconstruction's error and a 95 percent confidence interval.

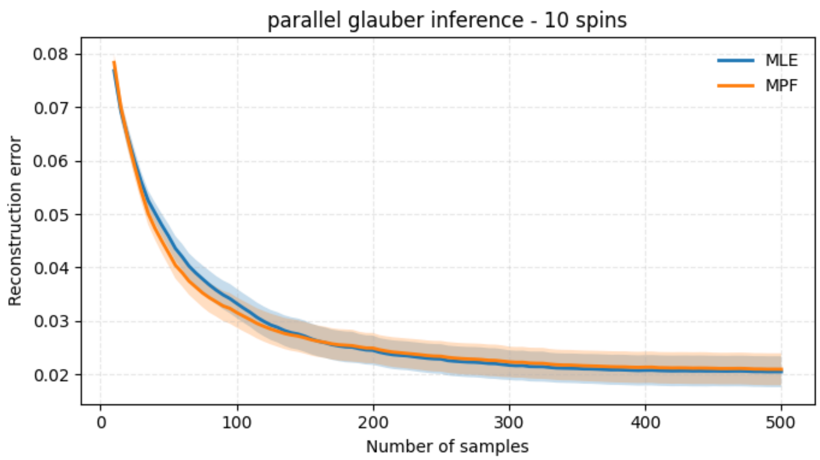

Figure 1: Times series of the reconstruction's error.

References

[1] Stephen G. Brush. History of the lenz-ising model. *Reviews of Modern Physics*, 39(4):883–893, 1967.

[2] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

[3] Andrey Y. Lokhov, Marc D. Vuffray, Sidhant Misra, and Michael Chertkov. Optimal structure and parameter learning of ising models. *Science Advances*, 4(3):e1700791, 2018.

[4] Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009.

[5] H. Chau Nguyen, Riccardo Zecchina, and Johannes Berg. Inverse statistical problems: from the inverse ising problem to data science. *Advances in Physics*, 66(3):197–261, June 2017.

[6] Pradeep Ravikumar, Martin J. Wainwright, and John D. Lafferty. High-dimensional ising model selection using ℓ_1-regularized logistic regression. *Annals of Statistics*, 38(3):1287–1319, 2010.

[7] Jascha Sohl-Dickstein, Peter Battaglino, and Michael Robert DeWeese. A new method for parameter estimation in probabilistic models: Minimum probability flow. *CoRR*, abs/2007.09240, 2020.

[8] Julia M. Yeomans. *Statistical Mechanics of Phase Transitions.* Oxford University Press, 1992.