MCMC for Sampling from Markov Random Fields: Gibbs and Metropolis Algorithms for Image Denoising

Davide Patricelli
Ghent University

May 2024

1 Introduction

This project explores the use of Markov Chain Monte Carlo (MCMC) algorithms for sampling from Markov Random Fields (MRFs), with a focus on binary image models derived from statistical physics. The objective is to study how different MCMC schemes—Metropolis-Hastings and Gibbs sampling—behave when applied to inference and image denoising.

The work begins from a MATLAB implementation of a Metropolis sampler for two models:

- the classical Ising model;
- an anisotropic second-order model with distinct horizontal, vertical, and diagonal couplings.

I reimplemented both models in Python, using `numpy` for computation and `matplotlib` for visualization. The implementation was then extended to include:

- a Gibbs sampler for both isotropic and anisotropic models;
- an optional data-likelihood term for image denoising (posterior sampling);
- simulated annealing, i.e. gradual reduction of the temperature parameter to approach MAP solutions.

This framework allows direct comparison between Metropolis and Gibbs sampling in terms of convergence, cluster formation, and denoising quality. The parameters α, β, and T retain their standard physical interpretation: local bias, interaction strength, and thermal noise. Their combined effect controls how strongly the field enforces spatial coherence versus random fluctuation.

2 Task 1: Model Definition

2.1 MRF and Ising Model

A Markov Random Field (MRF) is an undirected graph $G = (\mathbf{X}, \mathbf{E})$ where $\mathbf{X} = \{X_1, \dots, X_N\}$ are random variables and $\mathbf{E}$ are edges representing conditional dependencies. Each variable X_i depends only on its neighborhood $N(X_i)$:

$$P(X_i|\mathbf{X}_{/X_i}) = P(X_i|N(X_i)).$$

The joint distribution has the Gibbs form

$$P(\mathbf{x}) = \frac{1}{Z} \exp\left(-\frac{H(\mathbf{x})}{T}\right), \tag{1}$$

where $H(\mathbf{x})$ is the energy and Z the partition function. For a binary MRF (the Ising model),

$$H(\mathbf{x}) = \alpha \sum_i x_i + \beta \sum_{\langle i,j \rangle} x_i x_j,$$

with β controlling neighbor coupling and α an external bias.

The model was also extended to an *anisotropic* version, introducing different couplings for horizontal, vertical, and diagonal neighbors $(\beta_h, \beta_v, \beta_{d1}, \beta_{d2})$, allowing direction-dependent smoothness.

2.2 MCMC Sampling

MCMC methods generate samples from complex distributions by constructing a Markov chain that converges to the desired stationary distribution. In this context, they produce sequences of configurations proportional to $\exp(-H/T)$. As temperature decreases, the system tends toward minimum-energy (maximum a posteriori) states.

2.3 Metropolis Algorithm

Starting from configuration $\mathbf{x}_0$, a new candidate $\mathbf{x}_c$ is proposed by flipping a random site. The move is accepted with probability

$$p_{\text{acc}} = \min\left(1, \exp\left[-\frac{\Delta H}{T}\right]\right),$$

where $\Delta H = H(\mathbf{x}_c) - H(\mathbf{x}_0)$. If $\Delta H < 0$, the change is always accepted. This scheme does not require evaluating Z.

For posterior sampling in image denoising, the energy includes an additional *likelihood term*:

$$H(\mathbf{x}|\mathbf{y}) = \alpha \sum_i x_i + \beta \sum_{\langle i,j \rangle} x_i x_j - \lambda \sum_i y_i x_i,$$

where $\lambda > 0$ enforces similarity between the latent x_i and the observed pixel y_i.

3 Task 2: Convergence Considerations

Empirically, around 50 sweeps per pixel often suffice for qualitative convergence to a low-energy configuration, though this depends on temperature T and coupling β. Large $|\beta|$ encourages strong local agreement, while low T suppresses noise and accelerates stabilization into coherent clusters. However, global optimality is not guaranteed; the chain may become trapped in local minima. Simulated annealing mitigates this by gradually reducing T during iterations.

4 Task 3: Python Implementation

The MATLAB code was ported to Python for transparency and modularity. Core differences:

- random site order is generated using `numpy.random.shuffle`;
- computation of neighbor sums uses direct indexing over flattened arrays;
- plotting uses `matplotlib.pyplot`;
- the code now supports both prior-only and posterior (data-driven) sampling.

The repository consists of `main.py` plus four samplers:

- `Gibbs.py` and `Metropolis.py` for isotropic (Ising) models;
- `AGibbs.py` and `AMetropolis.py` for anisotropic models.

Each sampler can optionally apply simulated annealing:

$$T_i = T_0 \left(\frac{T_f}{T_0}\right)^{i/(N_{\text{iters}}-1)}.$$

5 Task 4: Gibbs Sampling and Comparison

5.1 Gibbs Sampler

Gibbs sampling updates each site X_k by resampling it from its conditional distribution given its neighbors and the observed data:

$$P(X_k = +1|\text{rest}) = \frac{1}{1 + \exp(2h_k/T)},$$

where $h_k = \alpha + \beta \sum_{n \in N(k)} X_n - \lambda y_k$ is the local field. This local sampling replaces the accept/reject mechanism of Metropolis and directly draws from the conditional probabilities.

5.2 Comparison with Metropolis

Both methods converge toward similar equilibrium states. Gibbs is simpler conceptually but can mix more slowly since it always updates deterministically by sampling the local conditional. Metropolis can traverse states faster when temperature is well tuned, especially with annealing. In practice, both produce consistent denoised images when applied to the same posterior energy.

5.3 Results

Behavior depends mainly on parameters $(\alpha, \beta, T, \lambda, N_{\text{iters}})$:

- α biases toward one spin value;
- $\beta < 0$ enforces neighbor alignment and yields smooth regions;
- large T increases randomness;
- λ controls how strongly the reconstructed image follows the noisy input;
- higher iteration count improves stability but increases computation.

With $T = 0.1$, $\beta = -0.9$, and $\alpha = 0$, $\lambda = 0$. both samplers form clear spin clusters resembling coherent image regions. For anisotropic couplings, e.g. larger $|\beta_{d2}|$ than other directions, elongated oblique structures emerge, confirming direction-dependent smoothness.

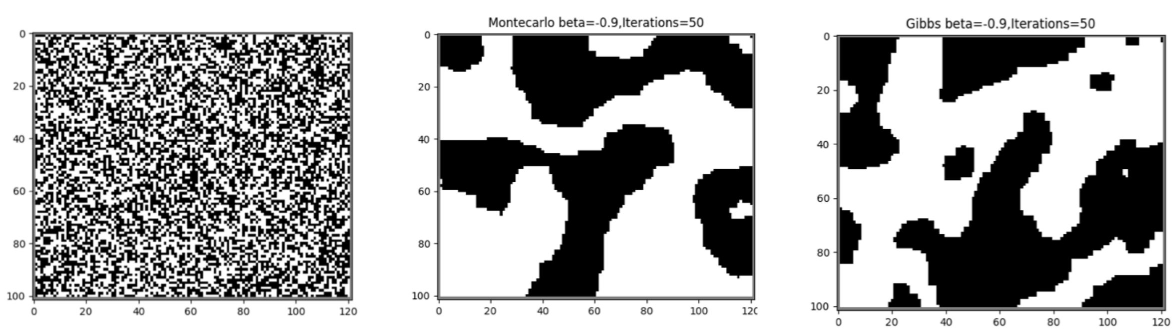

Figure 1: Formation of coherent clusters using Metropolis and Gibbs samplers.

6 Image Denoising via MAP Estimation

Binary image denoising can be expressed as MAP inference:

$$\hat{x}_{MAP} = \arg\max_x P(x|y) = \arg\min_x H(x|y).$$

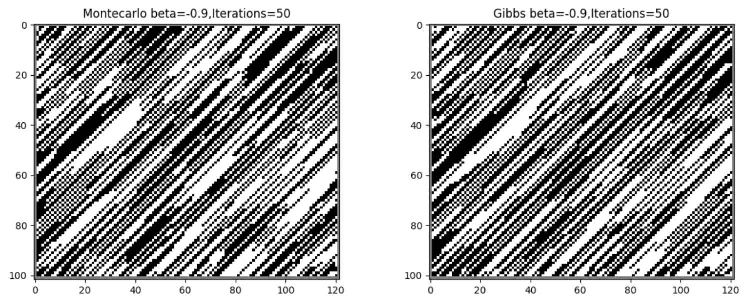

Figure 2: Anisotropic model result with directional clustering.

The posterior combines a smoothness prior (MRF) with a fidelity term $\lambda y_i x_i$. The Gibbs and Metropolis samplers approximate this MAP configuration through stochastic relaxation. Annealing lowers the temperature gradually, allowing the chain to explore at high T and settle into low-energy minima at low T. For $T_0 = 1.5$ and $T_f = 0.8$, the exponential cooling schedule used was:

$$T_i = T_0 \left(\frac{T_f}{T_0} \right)^{i/N_{\text{iters}}}.$$

7 Phase Transition and the Random Field Ising Model

The denoising setup can be understood as a realization of the **Random Field Ising Model (RFIM)**. Each pixel (spin) $x_i \in \{-1, +1\}$ interacts with its neighbors through a uniform coupling β and experiences a local external field proportional to the observed image:

$$H(x|y) = \beta \sum_{\langle i,j \rangle} x_i x_j - \lambda \sum_i y_i x_i + \alpha \sum_i x_i.$$

The local field $h_i = -\lambda y_i$ introduces spatial disorder: when y corresponds to a clean image, these fields are coherent; when noise is added, neighboring sites impose conflicting alignments. The resulting competition between the coupling term (favoring order) and the random field (inducing disorder) is the hallmark of the RFIM.

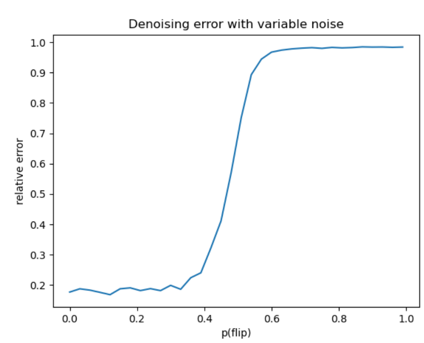

Figure 4: Measured transition in reconstruction quality: for small noise, ordered configurations persist; beyond a critical flip probability, global structure collapses.

In the classical Ising model, temperature drives a transition from an ordered to a disordered phase. In the RFIM, the key control parameter is the level of randomness in the external field. For weak disorder, the system maintains long-range order; for strong disorder, the influence of the random field destroys global alignment, yielding a disordered phase. This transition has been extensively studied in statistical physics and shares structural analogies with information-theoretic detectability transitions [3].

Tne results reproduce this qualitative behavior. As the flip probability of the observed image increases, the external field becomes inconsistent, and reconstruction error rises sharply. Below a critical noise level, the system remains ordered and the recovered image retains its structure. Beyond it, disorder dominates, the posterior becomes fragmented, and denoising fails. This empirical transition mirrors the RFIM order–disorder phenomenon, illustrating how inference tasks under noisy observations inherit the same mechanisms governing phase transitions in disordered spin systems.

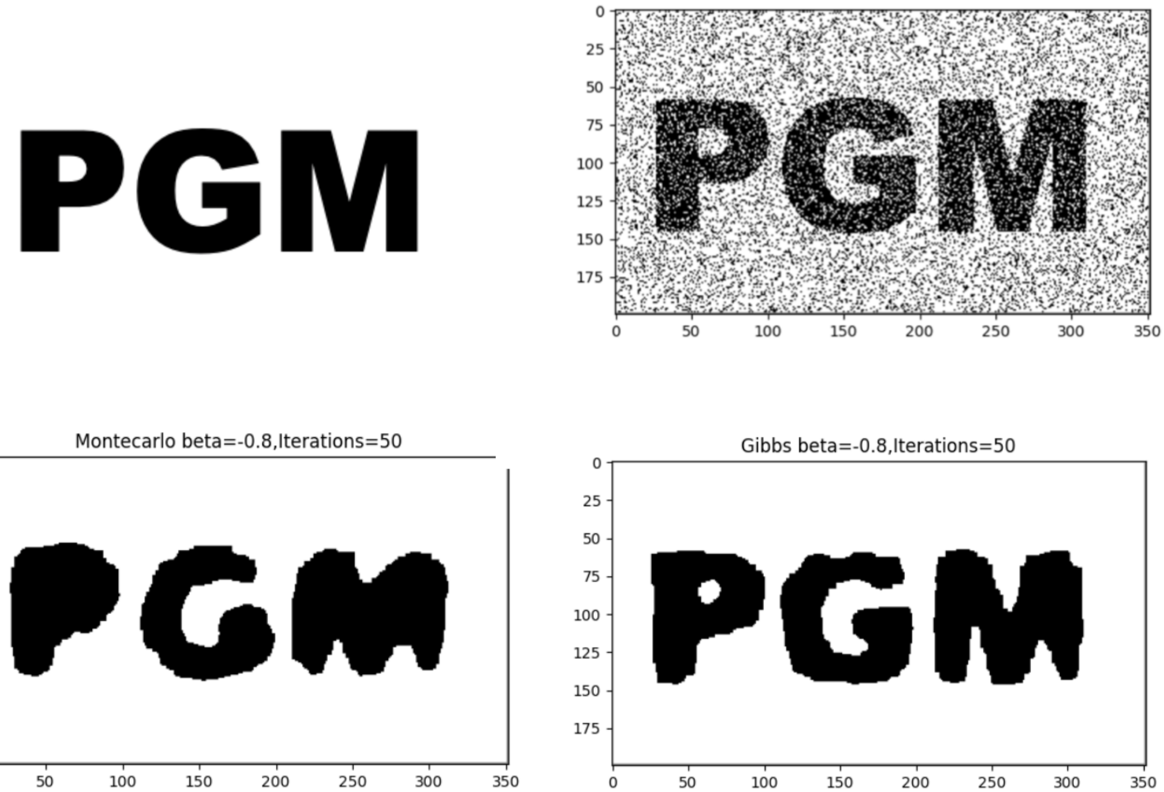

Figure 3: From noisy input (left) to denoised reconstruction (right). The Gibbs sampler produced smoother homogeneous regions.

8 Discussion and Outlook

The project demonstrates, in a compact setting, how MCMC techniques from statistical physics directly implement probabilistic inference on images.

- **Physical–probabilistic link:** Temperature controls stochasticity; strong coupling induces order. Image denoising thus parallels low-temperature Ising relaxation.
- **Posterior sampling:** Including a data-likelihood term $(-\lambda y_i x_i)$ converts the MRF prior into a proper posterior, bridging random field modeling and Bayesian inference.
- **Annealing:** Gradual temperature decay moves sampling toward MAP estimation while avoiding local traps.
- **Anisotropy:** Direction-dependent couplings replicate structural priors, such as oriented patterns or texture anisotropy.

The implementation remains qualitative—convergence and denoising quality are assessed visually. Quantitative extensions could include metrics (PSNR, SSIM), alternative noise models, and accelerated samplers (e.g. block-Gibbs or Metropolis-within-Gibbs).

References

[1] Theory slides: Probabilistic graphical models course at ghent university (ay 2022-2023).

[2] David Barber. *Bayesian Reasoning and Machine Learning.* 2007–2020.

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

[4] A. Pizurica. Probabilistic reasoning: When the environment is uncertain.

[5] Chang Yue. Markov random fields and gibbs sampling for image denoising. Project Report.