

A Comparison of Optimization in Artificial Neural Networks and Ising Systems

Iliana Balaskas

April 25, 2026

Abstract

For many years, artificial neural networks (ANNs) have been described as computing systems with loosely inspired biological similarities. Despite these similarities, modern systems differ significantly from known neural processes in the brain. In contrast, energy-based computational models, such as Boltzmann and Ising machines, show promising similarities to these neural networks. Because both approaches solve optimization problems with drastically different processes, comparing gradient-based mechanisms to energy minimizing dynamics provides useful insight into how Ising computations differ from conventional neural network optimization frameworks. In this paper, we compare these two approaches by examining their energy landscapes, updating process, and convergence behavior!. Through this, we will show how stochastic sampling methods provide an alternative computational option! to gradient-based systems in neural networks.

1 Background

Optimization plays a key role in modern computational methods to successfully solving a variety of problems. Artificial neural networks are one of the most used frameworks for optimization. Ising machines and other similar energy based systems offer an alternative approach on minimizing physical energy functions defined over various interacting variables.

While both frameworks solve optimization problems, the two methods rely of fundamentally different dynamical mechanisms. ANNs use gradient based update to adjust itself, while Ising systems typically form updates through stochastic transitions that reduce the system's energy. Analyzing the differences between these approaches offers insight as to how energy minimization procedures offer an alternative computations strategy to deterministic gradient descent displayed in ANNs.

1.1 Artificial Neural Networks

Modern artificial neural networks are described as computing systems similar to and inspired by the brain. However, the mechanism with which ANNs preform has no known biological equivalent; ANNs solve optimization problems by minimizing a cost function. Successful optimization therefore relies on the systems ability to navigate the landscape of the network using gradient decent.

The learning of the system corresponds to moving through an n-dimensional parameter landscape toward a minimum satisfying $\nabla f(\theta) = 0$!. In this framework, parameters update in iterations following

$$\theta \leftarrow \theta - \alpha \nabla f(\theta)$$

where the gradient determines the direction of the steepest descent. Since this update relies only on local information, convergence at a nearby stationary point is significantly more common than acquiring a global minimum.

1.2 Ising Models

In contrast to ANNs, Ising systems solve optimization problems by minimizing an energy function defined over various interacting variables. Optimization in the Ising models is expressed as minimizing the Hamiltonian, such that,

$$H = - \sum_{i,j} J_{ij} \sigma_i \sigma_j - \sum_i h_i \sigma_i$$

Here, the variable $\sigma_i \in \{-1, 1\}$ represents the spin states, and the coefficient J_{ij} determines pairwise interactions between variables. Rather than adjusting parameters continuously, optimization occurs by searching over many discrete configurations of spin assignments.

A variety of optimization problems are expressed in this form. For example, the Max Cut problem can be written as an Ising energy minimization problem by choosing couplings $J_{ij} = -W_{ij}$. This allows the Hamiltonian to represent the cut value up to some constant offset. Such mapping demonstrates how Ising systems can provide a general framework for solving discrete optimization problems using energy minimization.

1.2.1 Gibbs Sampling and Markov-Chain

Unlike gradient descent, which updates using deterministic directions, Ising styled systems explore the space through stochastic transitions governed by probability distributions. Configurations evolve according to the probability distribution of the form

$$\pi(s) \propto e^{-H(s)}$$

Gibbs distributions describe equilibrium behavior in Markov-chain sampling algorithms and are the basis of the procedure for Boltzmann machines. Since stochastic transition allows for a temporary increase in energy, a system can sometimes escape local minima, thereby allowing it to continue exploring the configuration landscape. This is a key distinguishing factor of energy based optimization from deterministic descent methods. The systems ability to successfully escape a local minima can often allow it to reach a final result much closer to the global value.

1.2.2 Langevin Dynamics and Stochastic Relaxation

Another factor that plays a key role to the success of energy based computations is stochastic relaxation. Simply put, stochastic relaxation is a random change made in the immediate "neighborhood" where the system is currently sitting at. This occurs by sampling from a local conditional probability distribution in the given neighborhood and can be modeled such that,

$$m\dot{v}(t) = -\lambda v(t) + \eta(t)$$

where m is the effective mass of the system, $v(t)$ is the velocity of the variable that is changing over time, $\dot{v}(t)$ is the acceleration of the given variable, λ is the dampening coefficient, $\eta(t)$ is the term representing random noise acting on the system. Through this equation, we see that the dampening term drive the system toward some local equilibrium,

while the stochastic noise term allows for a temporary increase in energy that permits the system to explore different states in the energy landscape without "getting stuck".

2 Comparing Frameworks

Now with a clear understanding of the various optimization frameworks and their components, one can compare how artificial neural networks and Ising systems represent variables, explore solution spaces, and converge toward optimal configurations through fundamentally different dynamical mechanisms.

2.0.1 Continuous vs Discrete Optimization

A key distinction between neural network and Ising style optimization is highlights in the representation of the variables being optimized. Artificial neural networks update continuous value parameters that specifically define the structure of some model. Comparatively, Ising systems update discrete valued states that present direct, specific solutions for the system.

Since neural networks optimize parameters rather than providing solutions, converge to an acceptable solution is directly dependent on identifying what factors contribute to minimizing the the prediction error, as opposed to determining configurations that minimize the overall energy function. While conversely, Ising systems represent solutions explicitly as spin configurations. With such systems, optimization corresponds to selecting a minimum energy configuration of the system that yields a satisfactory result.

As a result, these differing methods change the overall structure of the optimization landscape. In ANNs, the system is usually smooth and differentiable, making gradient based updates possible. While with Ising systems, the landscape is much more discrete, therefore requiring stochastic sampling methods to efficiently explore and optimize.

2.0.2 Deterministic versus Stochastic Updates

Gradient descent produces deterministic updates that are defined entirely by the information in the a local neighborhood. Once a direction of the descent is determined, the next steps in the optimization process are fixed, leading to a local minima that cannot be escaped.

This can be compared to energy based systems, which evolve through stochastic updates derived from probability distributions in the neighborhood. In Gibbs sampling, transitions between configurations occur with probabilities determined by energy differences rather than gradient directions. This allows the system to explore multiple possible solution paths simultaneously rather than committing to a single predictable path for descent.

Because stochastic updates permit temporary increases in energy, they provide a mechanism for escaping local minima that differs from deterministic optimization strategies. This feature is particularly important in discrete optimization problems where the objective landscape contains many competing local solutions.

2.0.3 Role of Energy Landscape and Convergence Behavior

Both ANNs and Ising systems can be interpreted as navigating energy landscapes, but with different methods. The structure of these landscapes differs depending on the variables being optimized. The convergence behavior of gradient based networks depends primarily on the geometry of the parameter landscape defined by the cost function. In many cases, gradient descent converges to nearby stationary points that are not globally optimal but remain stable under local perturbations. As a result, performance depends strongly on initialization and parametrization.

In comparison, Ising systems converge toward equilibrium probability distributions determined by the Hamiltonian rather than toward a single deterministic solution. Instead of identifying one minimum directly, the system samples configurations according to their relative energies and gradually concentrates a probability near optimal solutions. This sampling based convergence allows the system to explore multiple candidate configurations before settling into low energy states, thereby improving the final result. Since stochastic relaxation supports exploration across the energy landscape, Ising style computation is a very successful alternative for navigating optimization problems with complex combinatorial structure. These differences highlight a key distinction between parameter learning systems and configuration search systems; neural networks optimize representations of solutions, while Ising systems optimize solutions directly.

3 Implications

The differences between gradient based learning and stochastic energy minimization suggest that Ising style computation represents a distinct optimization model rather than a variation of neural network learning. Instead of computing gradients explicitly, energy based systems rely on local interactions between variables to guide global system behavior. As a result, optimization emerges from the collective dynamics of all the interacting components, rather than from centralized parameter updates.

This computation style allows many combinatorial optimization problems to be expressed directly as Hamiltonian minimization tasks. Because the structure of the optimization problem is embedded into the coupling matrix of the system, solution search becomes equivalent to identifying low-energy configurations. This approach differs fundamentally from gradient-based optimization, which requires explicit evaluation of derivatives at each update step.

Stochastic sampling methods such as Gibbs and Langevin dynamics therefore provide a computational mechanism for solving optimization problems that does not rely on differentiability nor continuous parameter updates. Instead, these methods exploit probabilistic transitions to explore configuration space efficiently, thereby identifying candidate solutions through relaxation toward equilibrium states.

3.1 Problem Suitability

Artificial neural networks and Ising systems are designed for different classes of optimization problems, and neither framework can be considered universally superior. Gradient based neural networks are well suited for optimization over continuous parameter spaces where derivative information is available and efficient updates can be computed directly. In contrast, Ising systems naturally represent discrete configuration space optimization problems through Hamiltonian minimization and stochastic sampling. Because many combinatorial problems can be mapped directly onto Ising energy functions, energy based computation provides a structured approach for solving optimization tasks that are difficult to express using gradient based parameter updates. As a result, the effectiveness of each framework depends primarily on the structure of the optimization problem being considered rather than on the update mechanism alone.

4 Conclusion

Artificial neural networks and Ising systems both perform optimization over high-dimensional landscapes, but they rely on fundamentally different update mechanisms. Gradient based learning operates through deterministic parameter updates guided by local derivative information, while energy based computation proceeds through stochastic relaxation toward equilibrium configurations. Because these approaches explore optimization landscapes using different dynamical principles, comparing them clarifies how stochastic sampling methods provide an alternative framework for solving complex optimization problems. Understanding these differences highlights the role of energy minimization dynamics in Ising systems and illustrates how this style of optimization offers capabilities distinct from conventional gradient based learning methods.

References

- [1] Electronics (MDPI), *Ising Machine-Based Optimization Methods*, Journal Article, <https://www.mdpi.com/2079-9292/10/21/2689>.
- [2] Nature Communications, *ON-OFF Neuromorphic Ising Machines Using Fowler–Nordheim Annealers*, Journal Article, <https://www.nature.com/articles/s41467-025-58532-9>.
- [3] A. Zador, *Modern Artificial Neural Networks: Is Evolution Required?*, Neural Computation, <https://direct.mit.edu/neco/article/35/5/763/115254/Modern-Artificial-Neural-Networks-Is-Evolution>.
- [4] arXiv, *Recent Advances in Ising Machine Architectures*, Preprint, <https://arxiv.org/html/2511.00746>.
- [5] National Institutes of Health (PMC), *Energy-Based Optimization and Ising Machine Hardware Review*, Journal Article, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10533504/>.
- [6] arXiv, *Annealing-Based Optimization with Ising Systems*, Preprint, <https://arxiv.org/html/2403.01320v2>.
- [7] M. Huang, *ECE 205 Lecture Slides*, University of Rochester, Lecture Slides.