Improving photonic Ising machines by incorporating the Adam-optimizer

Stijn Van Vooren,[1] Guy Van der Sande,[1] and Guy Verschaffelt[1]

[1] *Applied Physics Research Group, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium*

Optimization problems are crucial across various fields, including logistics, network design, and machine learning. However, growing ecological challenges and the end of Moore's law highlight the need for energy-efficient, alternative computing methods. Ising machines have emerged as a promising solution, with several physical implementations already in existence. Optimization problems can be mapped onto Ising machines, where the ground state of the Ising machine then corresponds to the optimal solution, which it naturally tends to. Photonic implementations in particular are promising due to their expected high operational speed and energy efficiency. Traditionally, photonic Ising machines have used gradient descent for optimization. In this contribution, we discuss how momentum and Adam-optimizer – well-known from neural network training - can be added to these Ising machines' temporal evolution. Using numerical simulations of such Ising machines, we show that Adam's adaptive learning rate and use of gradient moments enable faster convergence to the optimum solution as compared to classical methods. Simulations on Max-Cut benchmarks demonstrate that Adam- based Ising machines reduce time to solution by orders of magnitude over gradient descent. These findings suggest that incorporating Adam consistently enhances performance of Ising machines.

Introduction

Computing technology has long relied on making transistors smaller to improve performance. However, we are now reaching physical limits where transistors cannot be made much smaller without running into fundamental problems, marking the end of Moore's law [1]. This challenge means we need to find new ways to advance computing power beyond just making components smaller. To keep advancing computing power, researchers are now looking beyond traditional hardware, exploring natural computing systems that harness physical phenomena, alongside new approaches in software and system design [2].

One particularly promising new direction is the development of Ising machines, which represent a fundamentally different approach to computing compared to traditional computers. These machines are specifically designed to solve complex problems that regular computers often struggle with, such as optimization tasks, statistical sampling, and certain types of machine learning problems [3], [4], [5]. Nature shows us powerful ways to solve complex problems. When Japanese researchers placed food sources on a map representing Tokyo's stations, simple slime mold grew to connect them in almost the same pattern as Tokyo's optimized rail system. Like these natural systems, Ising machines harness nature's inherent ability to find the most efficient, lowest-energy solutions. The basic idea behind an Ising machine is that: it transforms difficult computational problems into a physical system where the solution naturally emerges when the system finds its lowest energy state, which it naturally tends to.

Currently, researchers are exploring several different ways to implement and build these machines. One approach uses networks of synchronized oscillators. These machines are practical because they can work at normal room temperature and don't need much power to operate [6]. Another method uses quantum effects at very low temperatures, called quantum annealing. While this approach shows great promise for certain types of problems, scientists are still working on making these machines more reliable and capable of handling larger problems [7]. A third approach leverages photonics for information processing, offering rapid computational speeds and the ability to execute numerous calculations concurrently through the encoding of information in light waves. Notably, these systems are capable of managing large-scale problems while also being scalable in terms of system size, allowing for adaptation to smaller configurations without compromising performance [8].

In this paper, we will focus on Ising machines as a versatile computational paradigm, irrespective of their physical implementation. Many of the Ising machine implementations make use of the temporal evolution of continuous "spin" variables to find the ground state of a binary spin problem, enabling the application of gradient descent techniques that are not feasible for binary problems.
In our exploration, we propose replacing traditional gradient descent methods with more advanced optimization solvers, specifically momentum and the Adam optimizer. This work is inspired by [9], which suggests the potential application of the Adam optimizer in the context of Ising machines. These modern approaches, well-known for their efficiency in training neural networks, offer significant advantages in terms of convergence speed and stability. The Adam optimizer, in particular, has achieved considerable success across various machine learning and signal processing tasks, often outperforming traditional gradient descent by providing faster convergence and greater robustness against local minima. Despite its widespread adoption in these fields, Adam has not been applied directly to the temporal dynamics of Ising machines. Previous research has hinted at the potential benefits of integrating Adam with Ising machines [paperreference], but this combination has yet to be fully explored. We argue that Adam's adaptive learning rates and built-in momentum can effectively enhance the dynamic response of Ising machines, allowing for more efficient and stable convergence to optimal states. This improvement could lead to substantial gains in computation speed and energy efficiency, offering a promising new direction for optimization techniques.

Results

We investigate binary optimization problems, where we want to minimize an objective function f

$$\min\{ f(\sigma) \mid \sigma \in \{-1,+1\}^{n}\},$$

that can be written in the form of the Ising Hamiltonian, i.e.

$$f(\sigma) = H_{\text{Ising}} = -\frac{1}{2}\sum_{i,j} J_{ij}\sigma_i\sigma_j - \sum_i b_i\,\sigma_i \quad ,$$

where $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_n\}\ is\ the\ state\ of\ the\ system,$ $\sigma_i \in \{-1,+1\}$ are binary spin variables, $J_{ij} \in \mathrm{R}$ represents the interaction strengths between variables, commonly referred to as the interaction matrix. In this study, we assume zero bias, i.e., $b_i = 0$.
The solution to our problem is called the ground state.
To leverage continuous optimization techniques, we replace our binary spin variables $\sigma_i \in \{-1,+1\}$ with continuous spin variables $x_i \in R$, effectively also replacing our original binary objective function $f(\sigma)$ with a continuous objective function

$$f(x) = H_{\text{cont}} = -\frac{1}{2}\sum_{i,j} J_{ij} x_i x_j \; .$$

In order to read out the binary solution corresponding to the obtained continuous solution, we take the sign of our state $\sigma_i = sign(x_i)$. By modifying the optimization process in this way, we are essentially solving a different, continuous problem. To ensure that the solutions tend toward binary outcomes, which are the desired solutions for the original problem, a non-linearity is introduced to encourage binary behavior in the spin variables. A common approach is to incorporate a double-well potential into the Ising Hamiltonian. This modification adjusts the objective function by creating deep minima at the binary values ($x_i = \pm 1$), while increasing the function value elsewhere ($x_i \neq \pm 1$). As a result, this promotes convergence toward binary solutions, ensuring that the system's behavior remains aligned with the original binary optimization problem. The resulting objective function after incorporating this non-linearity into the Ising Hamiltonian can be expressed as follows:

$$f(x) = \sum_i \left(\frac{\alpha - 1}{2} x_i^2 - \frac{x_i^4}{4} \right) - \frac{1}{2} \sum_{i,j} J_{ij} x_i x_j .$$

This results in an expression that is not the objective function itself, but rather the negative gradient of the modified objective function.

Classical Ising machines use gradient descent to optimize this altered objective function. *Gradient descent*, applied to a continuous objective function f(x), follows the update rule:

$$\frac{dx}{dt} = -\nabla f(x).$$

To simulate this, we Euler expand it, and obtain the update rule

$$x_t = x_{t-1} - \Delta t \nabla f(x_{t-1}),$$

where Δt denotes the Euler timestep, which is generally chosen to be very small when the original differential equation models a physical system, and $x_t = x(t)$ represents the continuous spin variable at time t, while $x_{t-1} = x(t - \Delta t)$, and finally the gradient is given by $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$.

In this paper, instead of using gradient descent, like for classical Ising machines, we used Momentum and Adam optimizer to optimize this altered objective function. The update rules for *Momentum optimizer* are given by the discrete update rules

$$m_i = \beta_1 m_{i-1} + (1 - \beta_1) \nabla f(x_{i-1}) \;,$$
$$x_i = x_{i-1} - \eta m_i \;,$$
$$i \leftarrow i + 1.$$

The *Adam optimizer* update equations are given by the discrete update rules

$$m_i \leftarrow \beta_1 m_{i-1} + (1 - \beta_1) \nabla f(x_{i-1})$$
$$v_i \leftarrow \beta_2 v_{i-1} + (1 - \beta_2) \left(\nabla f(x_{i-1}) \right)^2$$
$$\widehat{m_i} \leftarrow \frac{m_i}{1 - \beta_1^i}$$

$$\widehat{v_i} \leftarrow \frac{v_i}{1-\beta_2^i}$$
$$x_i \leftarrow x_{i-1} - \eta \frac{\widehat{m_i}}{\sqrt{\widehat{v_i}} + \epsilon}$$
$$\mathrm{i} \leftarrow \mathrm{i} + 1$$

where for both Momentum and Adam optimizer η is the learning rate, $\beta_1, \beta_2 \in [0,1[$ and ϵ is a small number for numerical stability. We reformulated the update rules as Euler expansions of continuous coupled differential equations, enabling a transition from a discrete update rule to a continuous time-dependent system. This approach provides a foundation for potential physical implementations by establishing a framework that captures the system's temporal evolution.

We compared the performance of Ising machines utilizing gradient descent, momentum, and the Adam optimizer by simulation on two Max-Cut benchmark sets: Bigmaq g05 and Gset. Our results show that while momentum performs similarly to gradient descent, the Adam-based Ising machine achieves solutions orders of magnitude faster than both gradient descent and momentum. These findings suggest a promising direction for potential physical implementations and highlight that incorporating the Adam optimizer could consistently offer performance improvements. This shift in methodology not only reinforces the capabilities of Ising machines but also highlights the potential for cross-pollination between advances in machine learning and the evolving field of natural computing.

References

[1] T. N. Theis and H.-S. P. Wong, "The End of Moore's Law: A New Beginning for Information Technology," Computing in Science & Engineering, vol. 19, no. 2, 41–50. 41-50, 2017.

[2] C. E. Leiserson, N. C. Thompson, J. S. Emer, B. C. Kuszmaul, B. W. Lampson, D. Sanchez, and T. B. Schardl. "There's plenty of room at the Top: What will drive computer performance after Moore's law?," Science, vol. 368, 2020.

[3] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S. Uchaikin, J. Wang, B. Wilson, and G. Rose, "Quantum annealing with manufactured spins," Nature, vol. 473, no. 7346, 194–198, 2011.

[4] Y. Yamamoto, K. Aihara, T. Leleu, K.-i. Kawarabayashi, S. Kako, M. Fejer, K. Inoue, and H. Takesue, "Coherent Ising machines—optical neural networks operating at the quantum limit," Quantum Inf, vol. 3, no. 1, 1–15, 2017.

[5] F. Böhm, D. Alonso-Urquijo, G. Verschaffelt, and G. Van der Sande, "Noise-injected analog Ising machines enable ultrafast statistical sampling and machine learning," Nat Commun, vol. 13, no. 1, 584, 2022.

[6] Y. Zhang, Y. Deng, Y. Lin, Y. Jiang, Y. Dong, X. Chen, G. Wang, D. Shang, Q. Wang, H. Yu, and Z. Wang, "Oscillator-Network-Based Ising Machine," Micromachines, vol. 13, no. 7, 1016, 2022.

[7] L. P. Yulianti and K. Surendro, "Implementation of Quantum Annealing: A Systematic Review," IEEE Access, vol. 10, 73156–73177, 2022.

[8] D. Pierangeli, G. Marcucci, and C. Conti, "Large-Scale Photonic Ising Machine by Spatial Light Modulation," Phys. Rev. Lett., vol. 122, no. 21, 213902, 2019.

[9] R. A. Brown, D. Venturelli, M. Pavone, and D. E. Bernal Neira, "Accelerating Continuous Variable Coherent Ising Machines via Momentum," Integration of Constraint Programming, Artificial Intelligence, and Operations Research, B. Dilkina, Ed. Springer Nature Switzerland, vol. 14742,109–126, 2024.