
Training Machine Learning Models with Ising Machines

Sayantana Pramanik^{1,2*} Kaumudibikash Goswami³ Sourav Chatterjee¹
M Girish Chandra¹

¹TATA Consultancy Services, India ²Indian Institute of Science

³QICI, The University of Hong Kong

Abstract

In this study, we use Ising machines to help train machine learning models by employing a suitably tailored version of opto-electronic oscillator-based coherent Ising machines with clipped transfer functions to perform trust region-based optimisation with box constraints. To achieve this, we modify such Ising machines by including non-symmetric coupling and linear terms, modulating the noise, and introducing compatibility with convex-projections. The convergence of this method, dubbed *iTrust* has also been established analytically. We validate our theoretical result by using *iTrust* to optimise the parameters in a quantum machine learning model in a binary classification task. The proposed approach achieves similar performance to other second-order trust-region based methods while having a lower computational complexity. Our work serves as a novel application of Ising machines and allows for unconstrained optimisation problems to be performed on energy-efficient computers with non von Neumann architectures.

1 Introduction

Traditionally, the utility of Ising machines has been limited to solving combinatorial optimisation problems [1, 2, 3] with polynomial resources by mapping them onto ground-state search problems of the Ising model [4, 5], using them as machine learning models [6, 7, 8, 9], or modelling optical neural networks [10]. While a variety of approaches for realizing this model of artificial spins network has been demonstrated in literature [11, 12, 13], the approach of employing opto-electronic-oscillators (OEOs) for building a coherent Ising machine (CIM) [14] has been lately gaining a lot of attention because of its cost-effective implementation, ambient operation, and scope for miniaturization [15].

In this work, we present a novel application of OEO-CIMs to unconstrained optimisation, and provide analytical proof of convergence of Ising machines to perform trust region-based optimization [16, 17, 18]. We refer to this technique as *iTrust* (Ising machines for trust-region optimisation). Along with the other aforementioned benefits of OEO-CIMs, the main advantage of *iTrust* stems from avoiding matrix-inversion and Cholesky decomposition of the Hessian. This opens up a new avenue of applications where the Ising machines may be used to optimise any parameterised, unconstrained objective function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. We denote the parameters of the objective function $f(\cdot)$ with the vector $\theta \in \mathbb{R}^n$. Particularly, using *iTrust*, we aim to find the optimal point θ^* that minimises the objective function:

Problem 1.

$$f(\theta^*) := \min_{\theta \in \mathbb{R}^n} f(\theta), \quad (1)$$

where θ^* satisfies second-order optimality conditions [16], under the following assumption [16]:

*Correspondence to <sayantan.pramanik@tcs.com, sayantanp@iisc.ac.in>

Assumption 1. If $\theta^{(0)}$ is the starting point of an iterative algorithm, then the function $f(\cdot)$ is bounded below on the level set $\mathcal{S} = \{\theta \mid f(\theta) \leq f(\theta^{(0)})\}$ by some value f^* , such that $f^* \leq f(\theta) \forall \theta \in \mathcal{S}$. Further, f is twice continuously differentiable on \mathcal{S} .

This allows us to use *i*Trust as an optimisation procedure for training models in traditional machine learning (ML) [19, 20, 21], quantum ML (QML) [22, 23, 24], quantum-inspired ML (QiML) [25], and variational quantum algorithmic (VQA) [26] models. Such optimisation problems are conventionally tackled by digital computers based on von Neumann architecture, leading to substantial memory and energy consumption, also known as ‘von Neumann bottleneck’ [27]. In contrast, since *i*Trust is based on Ising machines, it may potentially lead to more energy-efficient protocols [28, 29, 1] with an increased clock-speed [30]. This paper (along with a contemporary studies in [31] and [32] - which use analog thermodynamic computers to perform natural gradient descent, and quantum linear solvers [33, 34, 35] to calculate the Newton-update, respectively) hopes to open up new avenues of research where benefits of new-compute paradigms are reaped not only by using them as ML models, but also by employing them to aid in the training of models.

The remainder of this extended abstract is organised as follows: we propose essential modifications to a specific type of CIMs to make them compatible for trust-region optimisation in Section 2, and analytically examine its performance on convex objective functions with bounded gradients, and on smooth, locally-Polyak-Łojasiewicz (PŁ) [36] functions in Section 2.1. We describe the proposed algorithm *i*Trust in Section 3, before showing its convergence to second-order optimal solutions of Problem 1 in Theorem 3. We then proceed to demonstrate its efficacy through numerical experiments in Section 4 Conclusions and future outlook are in Section 5.

2 Economical Coherent Ising Machine

For *i*Trust, we consider the poor man’s CIM introduced in [14] with clipped nonlinearity [37], and refer to it as the Economical CIM (ECIM). It is then modified to find ε -suboptimal solutions of the following problem with \mathbf{J} as the coupling-matrix, and \mathbf{h} as the external field:

Problem 2.

$$\min_{\mathbf{s} \in [-\Delta, \Delta]^n} \left(E(\mathbf{s}) \triangleq \frac{1}{2} \langle \mathbf{s}, \mathbf{J}\mathbf{s} \rangle + \langle \mathbf{h}, \mathbf{s} \rangle \right) \quad (2)$$

Inspired by an earlier work [38], our modifications include setting $\alpha = 1$ and viewing β as the step-size in equation 8 of [37]. The variance of the injected noise is modulated, and varying step-sizes β_k are considered to facilitate better convergence. Provisions for accommodating non-symmetric coupling and linear terms are also made without relying on ancillary spins [39, 15]. The clipping voltage is set to $\pm\Delta$, and finally, the ECIM is made compatible with the definition of projection to the convex box $\mathcal{C} = [-\Delta, \Delta]^n$. As a result, the iterative update equation of the modified ECIM is given by:

$$\mathbf{s}^{(k+1)} = \Pi_{\mathcal{C}} \left(\mathbf{s}^{(k)} - \beta_k \left(\nabla E(\mathbf{s}^{(k)}) - \zeta^{(k)} \right) \right), \quad (3)$$

where $\zeta^{(k)} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, and $\Pi_{\mathcal{C}}(\cdot)$ is the projection operator to \mathcal{C} .

2.1 Convergence of ECIM

In this section, we present the convergence-results of the modified ECIM for convex or locally-Polyak-Łojasiewicz (PŁ) $E(\cdot)$. While PŁ may not be as popular a condition as convexity, it is definitely more general. For instance, PŁ (or PŁ*) functions have been shown to include neural networks with ReLU activations and quadratic losses where convexity cannot be assumed [40, 41, 42]. Further, [36] argues and proves that among Lipschitz-smooth functions such as *strongly convex*, *essentially strongly convex*, *weakly strongly convex*, and functions obeying the *restricted secant inequality*, PŁ functions entail the weakest assumptions. A more detailed exposition on the relations and implications between function-classes may be found in Theorem 2 of [36]. Furthermore, it is known that PŁ functions obey the Polyak-Łojasiewicz inequality. We, however, require the objective function to be PŁ *locally* on the constraint set \mathcal{C} , i.e., for some $\mu > 0$ and for all $\mathbf{s} \in \mathcal{C}$,

$$\|\nabla E(\mathbf{s})\|_2^2 \geq 2\mu(E(\mathbf{s}) - E^*). \quad (4)$$

We now state the convergence results through the following informal Theorems. Their formal statements and proofs have not been included in adherence to the page-limits.

Theorem 1 (Informal). *For convex $E(\cdot)$ with bounded gradients, the ECIM in equation (3) finds an ε -suboptimal solution to Problem 2 in \mathcal{C} with fixed step-sizes in $\mathcal{O}(1/\varepsilon^2)$ iterations. With diminishing step-sizes such that $\sum_{k=0}^{\infty} \beta_k = \infty$ and $\sum_{k=0}^{\infty} \beta_k^2 < \infty$, $\lim_{k \rightarrow \infty} (E(\mathbf{s}^{(k)}) - E^*) = 0$, where $E^* = \min_{\mathbf{s} \in \mathcal{C}} E(\mathbf{s})$.*

Theorem 2 (Informal). *For smooth $E(\cdot)$ that obeys the PL inequality locally, the ECIM in equation (3) finds an ε -suboptimal solution to Problem 2 in \mathcal{C} with fixed step-sizes in $\mathcal{O}(\ln(1/\varepsilon))$ iterations.*

If \mathcal{J} is the output of the ECIM, then the above results may be unified into the following equation for some constant $c \in (0, 1]$, as suggested in [18]:

$$-E(\mathcal{J}) \geq c|E(\mathbf{s}^*)|. \quad (5)$$

3 iTrust

Very briefly, the update $\mathbf{p}_{(t)}^*$ to $\boldsymbol{\theta}^{(t)}$ at the iteration t of a Newton-like trust-region method is found from the minimiser of:

Problem 3.

$$\min_{\|\mathbf{p}\|_2 \leq \delta_t} \left(m_t(\mathbf{p}) \triangleq \langle \nabla f(\boldsymbol{\theta}^{(t)}), \mathbf{p} \rangle + \frac{1}{2} \langle \mathbf{p}, \mathbf{H}(\boldsymbol{\theta}^{(t)}) \mathbf{p} \rangle \right), \quad (6)$$

where $\nabla f(\boldsymbol{\theta}^{(t)})$ and $\mathbf{H}(\boldsymbol{\theta}^{(t)})$ are the gradient and Hessian of f at $\boldsymbol{\theta}^{(t)}$, respectively. If the radius of the trust-region at iteration t is δ_t , then the feasible set, which is a ball¹, is represented with $\mathcal{B}_t = \{\mathbf{z} \in \mathbb{R}^n \mid \|\mathbf{z} - \boldsymbol{\theta}^{(t)}\|_2 \leq \delta_t\}$.

A major disadvantage of using the method proposed in Algorithm 3.2 stated in [18] to find $\mathbf{p}_{(t)}^*$ is the repeated requirement for Cholesky decomposition and inversion of the Hessian, both of which are in $\mathcal{O}(n^3)$. This becomes prohibitively expensive for problems where n is large, for instance machine learning models with millions of parameters. We aim to alleviate this problem by using the enhanced ECIM to find $\mathbf{p}_{(t)}^*$. We achieve this by exploiting the structural similarity Problems 2 and 3. Specifically, at each iteration t , \mathbf{J} is set to $\mathbf{H}(\boldsymbol{\theta}^{(t)})$, \mathbf{h} to $\nabla f(\boldsymbol{\theta}^{(t)})$, and Δ to δ_t . Here, the importance of the inclusion of linear terms in the Ising machine becomes clear, without which the gradient $\nabla E(\mathbf{s}^{(k)})$ could not have been provided to the ECIM without additional overheads in the form of ancillary spins [38, 39].

Remark 1. *It is interesting to note that if the coupling matrix $\mathbf{J}^{(t)}$ is positive semidefinite at the iteration t , then as per the definition of convexity, the objective function of the trust-region subproblem is convex. Additionally, since the coupling matrix is equal to the Hessian $\mathbf{H}(\boldsymbol{\theta}^{(t)})$, this also implies that the objective function f is convex in the region around $\boldsymbol{\theta}^{(t)}$. Thus, in a convex region of the original problem, the result in Theorem 1 becomes applicable for the ECIM.*

Further, we distinguish between the minimisers of $E_t(\mathbf{s})$ and $m_t(\mathbf{p})$ on the sets \mathcal{C}_t and \mathcal{B}_t by denoting them with $\mathbf{s}_{(t)}^*$ and $\mathbf{p}_{(t)}^*$, respectively.

Remark 2. *We would like to emphasize that the box \mathcal{C}_t and the ball \mathcal{B}_t share a common centre $\boldsymbol{\theta}^{(t)}$, and by design, the side-length of the box is set equal to the diameter of the ball at each iteration. Thus, the ball is contained completely within the box: $\mathcal{B}_t \subset \mathcal{C}_t^2$. Now, since the objective function*

¹To avoid situations where the optimisation Problem 1 has a *poor scaling* with respect to the decision variables $\boldsymbol{\theta}$, *elliptical* trust regions may be employed by replacing the constraint of Problem 3 with:

$$\|\mathbf{D}\mathbf{p}\|_2 \leq \delta, \quad (7)$$

where $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$ with $d_i \geq 0$. The elements d_i are adjusted according to the *sensitivity* of $f(\cdot)$ to $\boldsymbol{\theta}_i$: if $f(\cdot)$ varies highly with a small change in $\boldsymbol{\theta}_i$, then a large value of d_i is used; and vice versa [16].

²Differential scaling with respect to different components of the decision variables may be handled by setting individual Δ_i for each coordinate $\boldsymbol{\theta}_i$ such that the elliptical trust-region from equation (7) lies within the box defined by the Δ_i s.

of the Problems 3 and 2 are identical, and the constraint set of the former is contained in that of the latter, we have:

$$E_t(\mathbf{s}^*_{(t)}) \leq m_t(\mathbf{p}^*_{(t)}). \quad (8)$$

This means that if the ECIM and the Algorithm 3.14 in [18] can both reach near-optimal solutions of their respective optimisation problems, then the objective value obtained by the ECIM is guaranteed to be better. This results in a higher reduction in the value of $f(\boldsymbol{\theta})$ at each iteration.

We name this technique of using the ECIM for trust-region optimisation as *iTrust*. The workflow for *iTrust* has been portrayed in Algorithm 1, which draws inspiration from, and is an amalgamation of, Algorithms 4.1 and 4.2 of [16] and [18], respectively.

Algorithm 1: *iTrust*

input: initial point $\boldsymbol{\theta}^{(0)} \in \mathbb{R}^n$; maximum trust-region radius $\delta_{\max} > 0$; initial radius $\delta_0 \in (0, \delta_{\max}]$; thresholds on ρ_t : $0 < \mu < \eta < 1$; radius-updation parameters $\gamma_1 < 1$ and $\gamma_2 > 1$; noise variance σ^2 ; sequence of step-sizes (β_k) ; and number of iterations T and K

```

1 begin
2   for  $t \in [T]$  do
3     evaluate  $\nabla f(\boldsymbol{\theta}^{(t)})$  and  $\mathbf{H}(\boldsymbol{\theta}^{(t)})$ ;
4      $\mathbf{J}^{(t)} \leftarrow \mathbf{H}(\boldsymbol{\theta}^{(t)})$ ;
5      $\mathbf{h}^{(t)} \leftarrow \nabla f(\boldsymbol{\theta}^{(t)})$ ;
6      $\Delta_t \leftarrow \delta_t$ ;
7     initialise  $\mathbf{s}^{(0)}$  randomly in  $\mathcal{C}_t = [-\Delta_t, \Delta_t]^n$ ;
8     for  $k \in [K]$  do
9       sample  $\boldsymbol{\zeta}^{(k)} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ ;
10       $\mathbf{s}^{(k+1)} = \Pi_{\mathcal{C}_t} \left( \mathbf{s}^{(k)} - \beta_k \left( \nabla E_t(\mathbf{s}^{(k)}) - \boldsymbol{\zeta}^{(k)} \right) \right)$ ;
11    end
12    calculate  $\rho_t = \frac{f(\boldsymbol{\theta}^{(t)} + \mathbf{s}^{(K)}) - f(\boldsymbol{\theta}^{(t)})}{E_t(\mathbf{s}^{(K)})}$ ;
13    if  $\rho_t < \mu$  then
14       $\delta_{t+1} = \gamma_1 \delta_t$ ;
15      continue;
16    else
17      if  $\rho_t > (1 - \mu)$  and  $\|\mathbf{s}^{(K)}\|_{\infty} = \delta_t$  then
18         $\delta_{t+1} = \min(\gamma_2 \delta_t, \delta_{\max})$ ;
19      else
20         $\delta_{t+1} = \delta_t$ ;
21      end
22    end
23    if  $\rho_t > \eta$  then
24       $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \mathbf{s}^{(K)}$ ;
25    else
26       $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$ ;
27    end
28  end
29  return  $\boldsymbol{\theta}^{(T)}$ 
30 end

```

We claim that this technique of employing ECIMs to solve the subproblem of trust-region methods converges (or tends to converge to) second-order optimal solutions of Problem 1 in \mathcal{S} . This claim is formalised in the form of the following theorem [16, 18], the proof of which has been omitted for brevity:

Theorem 3 (Convergence of *iTrust*). *Let assumption 1 be true, and let $(\boldsymbol{\theta}^{(t)})$ be the sequence of iterates generated by Algorithm 1 such that equation (5) is satisfied at each iteration. Then we have*

that:

$$\lim_{t \rightarrow \infty} \|\nabla f(\boldsymbol{\theta}^{(t)})\|_2 = 0. \quad (9)$$

Moreover, if \mathcal{S} is compact, the either Algorithm 1 terminates at a point $\boldsymbol{\theta}^{(T)} \in \mathcal{S}$ where $\nabla f(\boldsymbol{\theta}^{(T)}) = 0$ and $\mathbf{H}(\boldsymbol{\theta}^{(T)}) \succcurlyeq 0$; or $(\boldsymbol{\theta}^{(t)})$ has a limit point $\boldsymbol{\theta}^* \in \mathcal{S}$ such that $\nabla f(\boldsymbol{\theta}^*) = 0$ and $\mathbf{H}(\boldsymbol{\theta}^*) \succcurlyeq 0$.

4 Empirical Evaluation

In this section, we will demonstrate the efficacy for the proposed method through numerical experiments. Specifically, *iTrust* will be applied to optimise the parameters in a quantum machine learning (QML) model that performs binary classification. QML is another instance where an *alternate-compute* paradigm is used to enhance machine learning through the introduction of quantum models as hypothesis functions that are non-trivial to simulate classically. QML models have been found to provide advantages in laboratory setting in terms of the number of parameters [43], the volume of training data required [44], and the number of iterations/epochs the models are trained for [45]. Nevertheless, in recent times, the use of variational models [26] for QML has garnered some criticism that questions their advantage [46], especially due to the presence of barren plateaus [47, 48] and classical simulability [49, 50]. However, proving a quantum-advantage in ML is far from the scope of this study, and neither does the *iTrust* algorithm alleviate the issues of barren plateaus or classical simulability in QML.

Here, we aim to enhance the training of a small QML model to perform classification of the Iris dataset [51]. Since our focus is only on facilitating the training of models and not on their generalisability, only the training error at each iteration will be observed and reported as a measure of the performance of *iTrust*. The Iris dataset consists of four floral-features that may be used to categorise the flowers into three distinct classes, only the first two of which have been used here.

The details of the quantum classification model are as follows: the four features were encoded into the states of four qubits using `AngleEmbedding` with a combination of Hadamard and R_Z gates on each of the qubits. The features were first scaled to lie in the range of $[0, \pi]$ before being passed as parameters into the R_Z gates. Subsequently, three layers of the `BasicEntanglingLayers` ansatz from PennyLane [52] were appended to the circuit. The gates in the ansatz contain learnable parameters that were optimised. Finally, the expectation value of the Pauli- Z operator of the first qubit was measured. To calculate the *empirical risk*, the label for each datapoint was expressed as $\{\pm 1\}$ and the mean squared error was evaluated over the entire training set.

The performance of *iTrust* was benchmarked against those of two other algorithms: gradient-descent (GD) which only uses the first-order derivatives; and Algorithm 3.2 from [18] - henceforth referred to as More & Sorensen (MnS) - which like *iTrust* additionally requires the Hessian. Hence, GD requires only n evaluations of the quantum circuit at each iteration to estimate the gradient using the Parameter-Shift Rule [53], while the second-order methods need $2n^2 + n$ circuit executions (2 for the gradients of each coordinate, 4 for each of the off-diagonal terms of the Hessian, and an additional 1 for the diagonal terms [54]). Each of the algorithms was run for 100 iterations with the same initial point; and the experiment was repeated 10 times with different starting points sampled from a uniform distribution. A learning rate of 0.1 was used for GD. The hyperparameters for the second-order methods were: $\delta_{\max} = \delta_0 = 0.5$, $\eta = 0.1$, $\mu = \eta - 0.001$, $\gamma_1 = 0.75$, and $\gamma_2 = 1.25$. In addition, for *iTrust*, β was set to 0.5, with $K = 10$.

4.1 Numerical Results

The results of the aforementioned experiments have been reported in Figure 1, where 1a shows the training loss at each iteration of training. The bold lines denote the mean, while the shaded regions indicate the standard deviation around the mean across the 10 experiments. It may be noted that the second-order methods outperformed the first-order one, as expected. Between MnS and *iTrust*, the former obtained a quicker reduction in loss at the initial iterations, with the latter catching up soon. As the training progressed, *iTrust* converged to a marginally lower value of loss compared to MnS and was found to be more stable, owing to its lower standard deviation.

However, as detailed earlier, the second-order methods come at an increased overhead of calculation the Hessian for QML models. To check if this overhead eclipses the benefit of a reduced number

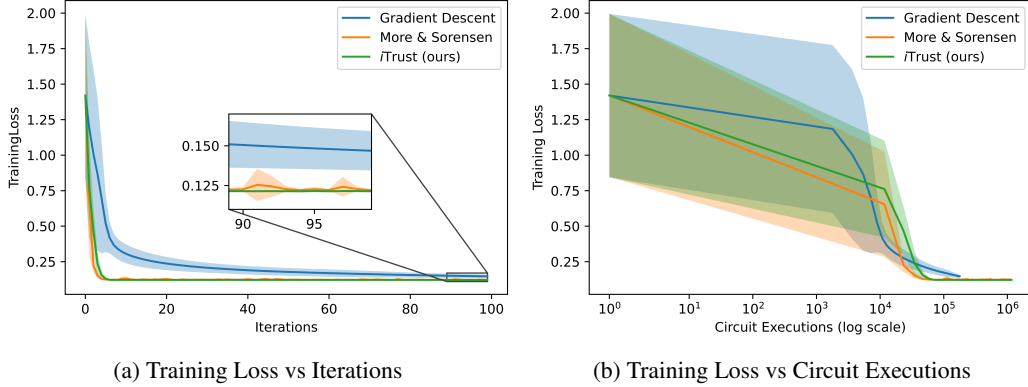


Figure 1: Fig. (a) shows the training loss at each iteration of training; while Fig. (b) reports the training loss against the number of circuits executed on a logarithmic scale.

of iterations, Figure 1b demonstrates the training loss against the total number of circuit-executions on a logarithmic scale. It is apparent from the plot that the second-order methods perform better towards the initial epochs and that the performance of MnS is slightly better than that of *iTrust*. But, one may recall that MnS requires Cholesky decomposition of the Hessian, whose complexity scales cubically with n . In comparison, *iTrust* forgoes this extra complexity while still retaining comparable performance.

With the above results in mind, we propose a training schedule where *iTrust* is used in the initial phase to get a quick reduction in the training loss, followed by the utilisation of GD until convergence. This method is markedly distinct from the existing convention of starting with GD to reach a *Newton's region*, followed by the use of second-order Newton's method. It must be noted at this point that evaluating the performance against the number of circuit/function executions would be unnecessary for models where the gradients and Hessians may be calculated (or estimated) with similar complexity as the function execution. In such cases, the advantages of *iTrust* (and MnS) become more pronounced.

5 Conclusions and Outlook

In this paper, we introduced *iTrust*, an algorithm that leverages Ising machines for trust-region based optimisation. In doing so, we proposed necessary modifications to the Ising machine, and proved the feasibility and convergence of *iTrust*. The use of Ising machines provides the potential for higher clock-speeds and reduced energy-consumption compared to conventional approaches. We validated our theoretical results by introducing the *iTrust* as an optimiser in a simple quantum machine learning model to perform binary classification, and compared its performance against other first and second-order methods. We find that *iTrust* delivers similar performance to the other trust-region based method, but has the advantage of avoiding Hessian inversion and Cholesky decomposition. In this way, we extend the previously allowed class of optimisation problems using the Ising machines and open up the possibility of training machine learning models with new compute paradigms.

Possible future directions may include generalising the ECIM for non-convex and non-PL objective functions. Variants of *iTrust* can also be constructed that are compatible with natural gradient descent [55, 56], by replacing the Hessian with the Fisher Information Matrix. *iTrust* may be further augmented by zeroth order methods like SPSA [57] in scenarios where evaluation of the gradients, Hessian, and Fisher information matrix is computationally expensive [58]. Lastly, the advantages of the ECIM over noisy projected gradient descent for the subproblem minimisation in terms of the clock-speed and energy-consumption can also be examined. We hope that this paper opens up new avenues of research in the analytical and empirical exploration of new applications of Ising machines.

References

- [1] Kirill P. Kalinin, George Mourgias-Alexandris, Hitesh Ballani, Natalia G. Berloff, James H. Clegg, Daniel Cletheroe, Christos Gkantsidis, Istvan Haller, Vassily Lyutsarev, Francesca Parmigiani, Lucinda Pickup, and Antony Rowstron. Analog iterative machine (aim): using light to solve quadratic optimization problems with mixed variables, 2023.
- [2] Peter L. McMahon, Alireza Marandi, Yoshitaka Haribara, Ryan Hamerly, Carsten Langrock, Shuhei Tamate, Takahiro Inagaki, Hiroki Takesue, Shoko Utsunomiya, Kazuyuki Aihara, Robert L. Byer, M. M. Fejer, Hideo Mabuchi, and Yoshihisa Yamamoto. A fully programmable 100-spin coherent ising machine with all-to-all connections. *Science*, 354(6312):614–617, November 2016.
- [3] Ising machines as hardware solvers of combinatorial optimization problems - Nature Reviews Physics — nature.com. <https://www.nature.com/articles/s42254-022-00440-8>. [Accessed 11-09-2024].
- [4] Takahiro Inagaki, Yoshitaka Haribara, Koji Igarashi, Tomohiro Sonobe, Shuhei Tamate, Toshimori Honjo, Alireza Marandi, Peter L McMahon, Takeshi Umeki, Koji Enbutsu, et al. A coherent ising machine for 2000-node optimization problems. *Science*, 354(6312):603–606, 2016.
- [5] Andrew Lucas. Ising formulations of many np problems. *Frontiers in Physics*, 2, 2014.
- [6] Fabian Böhm, Diego Alonso-Urquijo, Guy Verschaffelt, and Guy Van der Sande. Noise-injected analog ising machines enable ultrafast statistical sampling and machine learning. *Nature Communications*, 13(1), October 2022.
- [7] Xiao-Yang Liu and Ming Zhu. K-spin ising model for combinatorial optimizations over graphs: A reinforcement learning approach. In *OPT 2023: Optimization for Machine Learning*, 2023.
- [8] Jérémie Laydevant, Danijela Marković, and Julie Grollier. Training an ising machine with equilibrium propagation. *Nature Communications*, 15(1), April 2024.
- [9] Shaila Niazi, Shuvro Chowdhury, Navid Anjum Aadit, Masoud Mohseni, Yao Qin, and Kerem Y. Camsari. Training deep boltzmann networks with sparse ising machines. *Nature Electronics*, 7(7):610–619, June 2024.
- [10] Yoshihisa Yamamoto, Kazuyuki Aihara, Timothee Leleu, Ken-ichi Kawarabayashi, Satoshi Kako, Martin Fejer, Kyo Inoue, and Hiroki Takesue. Coherent ising machines—optical neural networks operating at the quantum limit. *npj Quantum Information*, 3(1):49, 2017.
- [11] K. Kim, M.-S. Chang, S. Korenblit, R. Islam, E. E. Edwards, J. K. Freericks, G.-D. Lin, L.-M. Duan, and C. Monroe. Quantum simulation of frustrated ising spins with trapped ions. *Nature*, 465(7298):590–593, June 2010.
- [12] Zhe Wang, Alireza Marandi, Kai Wen, Robert L. Byer, and Yoshihisa Yamamoto. Coherent ising machine based on degenerate optical parametric oscillators. *Physical Review A*, 88(6), December 2013.
- [13] R. Barends, A. Shabani, L. Lamata, J. Kelly, A. Mezzacapo, U. Las Heras, R. Babbush, A. G. Fowler, B. Campbell, Yu Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, P. J. J. O’Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, E. Solano, H. Neven, and John M. Martinis. Digitized adiabatic quantum computing with a superconducting circuit. *Nature*, 534(7606):222–226, June 2016.
- [14] Fabian Böhm, Guy Verschaffelt, and Guy Van der Sande. A poor man’s coherent ising machine based on opto-electronic feedback systems for solving optimization problems. *Nature Communications*, 10(1), August 2019.
- [15] A. Prabhakar, P. Shah, U. Gautham, V. Natarajan, V. Ramesh, N. Chandrachoodan, and S. Tayur. Optimization with photonic wave-based annealers. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 381(2241), December 2022.

- [16] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, 2e edition, 2006.
- [17] D. C. Sorensen. Newton’s method with a model trust region modification. *SIAM Journal on Numerical Analysis*, 19(2):409–426, 1982.
- [18] Jorge J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, 4(3):553–572, 1983.
- [19] Amanpreet Singh, Narina Thakur, and Aakanksha Sharma. A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pages 1310–1315, 2016.
- [20] Mohamed Alloghani, Dhiya Al-Jumeily, Jamila Mustafina, Abir Hussain, and Ahmed J. Aljaaf. *A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science*, page 3–21. Springer International Publishing, September 2019.
- [21] Rui Nian, Jinfeng Liu, and Biao Huang. A review on reinforcement learning: Introduction and applications in industrial process control. *Computers & Chemical Engineering*, 139:106886, 2020.
- [22] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, November 2019.
- [23] Alejandro Perdomo-Ortiz, Marcello Benedetti, John Realpe-Gómez, and Rupak Biswas. Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers. *Quantum Science and Technology*, 3(3):030502, June 2018.
- [24] M. Cerezo, Guillaume Verdon, Hsin-Yuan Huang, Lukasz Cincio, and Patrick J. Coles. Challenges and opportunities in quantum machine learning. *Nature Computational Science*, 2(9):567–576, September 2022.
- [25] Larry Huynh, Jin Hong, Ajmal Mian, Hajime Suzuki, Yanqiu Wu, and Seyit Camtepe. Quantum-inspired machine learning: a survey, 2023.
- [26] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, August 2021.
- [27] Levente Hajdu, Jérôme Lauret, and Radomir A. Mihajlović. *Grids, Clouds, and Massive Simulations*, page 308–340. IGI Global, 2014.
- [28] Jia Si, Shuhan Yang, Yunuo Cen, Jiaer Chen, Yingna Huang, Zhaoyang Yao, Dong-Jun Kim, Kaiming Cai, Jerald Yoo, Xuanyao Fong, and Hyunsoo Yang. Energy-efficient superparamagnetic ising machine and its application to traveling salesman problems. *Nature Communications*, 15(1), April 2024.
- [29] Yimin Wang and Xuanyao Fong. Energy-efficient ising machines using capacitance-coupled latches for maxcut solving. In *2024 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, May 2024.
- [30] Navid Anjum Aadit, Andrea Grimaldi, Mario Carpentieri, Luke Theogarajan, John M. Martinis, Giovanni Finocchio, and Kerem Y. Camsari. Massively parallel probabilistic computing with sparse ising machines. *Nature Electronics*, 5(7):460–468, June 2022.
- [31] Kaelan Donatella, Samuel Duffield, Maxwell Aifer, Denis Melanson, Gavin Crooks, and Patrick J. Coles. Thermodynamic natural gradient descent, 2024.
- [32] Pingzhi Li, Junyu Liu, Hanrui Wang, and Tianlong Chen. Hybrid quantum-classical scheduling for accelerating neural network training with newton’s gradient descent, 2024.
- [33] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.*, 103:150502, Oct 2009.

- [34] Andrew M. Childs, Robin Kothari, and Rolando D. Somma. Quantum algorithm for systems of linear equations with exponentially improved dependence on precision. *SIAM Journal on Computing*, 46(6):1920–1950, January 2017.
- [35] David Jennings, Matteo Lostaglio, Sam Pallister, Andrew T Sornborger, and Yiğit Subaşı. Efficient quantum linear solver algorithm with detailed running costs, 2023.
- [36] Hamed Karimi, Julie Nutini, and Mark Schmidt. Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition, 2020.
- [37] Fabian Böhm, Thomas Van Vaerenbergh, Guy Verschaffelt, and Guy Van der Sande. Order-of-magnitude differences in computational performance of analog ising machines induced by the choice of nonlinearity. *Communications Physics*, 4(1), July 2021.
- [38] Sayantan Pramanik, Sourav Chatterjee, and Harshkumar Oza. Convergence analysis of optoelectronic oscillator based coherent ising machines. In *2024 16th International Conference on COMMunication Systems & NETWORKS (COMSNETS)*, pages 1076–1081, 2024.
- [39] Abhishek Kumar Singh, Kyle Jamieson, Davide Venturelli, and Peter McMahon. Ising machines’ dynamics and regularization for near-optimal large and massive mimo detection, 2021.
- [40] Guillaume Garrigos and Robert M. Gower. Handbook of convergence theorems for (stochastic) gradient methods, 2024.
- [41] J. Bolte, A. Daniilidis, A. S. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions, 2006.
- [42] Chaoyue Liu, Libin Zhu, and Mikhail Belkin. Loss landscapes and optimization in over-parameterized non-linear systems and neural networks, 2021.
- [43] The power of quantum neural networks - Nature Computational Science — nature.com. <https://www.nature.com/articles/s43588-021-00084-1>. [Accessed 11-09-2024].
- [44] Event Classification with Quantum Machine Learning in High-Energy Physics - Computing and Software for Big Science — link.springer.com. <https://link.springer.com/article/10.1007/s41781-020-00047-7>. [Accessed 11-09-2024].
- [45] Quantum convolutional neural networks for high energy physics data analysis — journals.aps.org. <https://journals.aps.org/prresearch/abstract/10.1103/PhysRevResearch.4.013231>. [Accessed 11-09-2024].
- [46] Joseph Bowles, Shah Nawaz Ahmed, and Maria Schuld. Better than classical? the subtle art of benchmarking quantum machine learning models, 2024.
- [47] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1), nov 2018.
- [48] Martin Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J. Coles, Lukasz Cincio, Jarrod R. McClean, Zoë Holmes, and M. Cerezo. A review of barren plateaus in variational quantum computing, 2024.
- [49] M. Cerezo, Martin Larocca, Diego García-Martín, N. L. Diaz, Paolo Braccia, Enrico Fontana, Manuel S. Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, Eric R. Anschuetz, and Zoë Holmes. Does provable absence of barren plateaus imply classical simulability? or, why we need to rethink variational quantum computing, 2024.
- [50] Pablo Bermejo, Paolo Braccia, Manuel S. Rudolph, Zoë Holmes, Lukasz Cincio, and M. Cerezo. Quantum convolutional neural networks are (effectively) classically simulable, 2024.
- [51] R. A. Fisher. Iris. UCI Machine Learning Repository, 1988. DOI: <https://doi.org/10.24432/C56C76>.

- [52] Ville Bergholm, et al. Pennylane: Automatic differentiation of hybrid quantum-classical computations, 2022.
- [53] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Physical Review A*, 99(3), mar 2019.
- [54] Andrea Mari, Thomas R. Bromley, and Nathan Killoran. Estimating the gradient and higher-order derivatives on quantum hardware. *Physical Review A*, 103(1), January 2021.
- [55] Shun-ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, February 1998.
- [56] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum natural gradient. *Quantum*, 4:269, may 2020.
- [57] James C. Spall. An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins Apl Technical Digest*, 19:482–492, 1998.
- [58] Julien Gacon, Christa Zoufal, Giuseppe Carleo, and Stefan Woerner. Simultaneous perturbation stochastic approximation of the quantum fisher information. *Quantum*, 5:567, October 2021.