iTrust: Ising Machines for Trust-Region Optimisation

Sayantan Pramanik[1,2], Kaumudibikash Goswami[3], Sourav Chatterjee[1], and M Girish Chandra[4]

[1]Corporate Incubation, TATA Consultancy Services, India
[2]Robert Bosch Centre for Cyber-Physical Systems, Indian Institute of Science
[3]QICI Quantum Information and Computation Initiative, The University of Hong Kong
[4]TCS Research, TATA Consultancy Services, India

In this work, we present a heretofore unseen application of Ising machines to perform trust region-based optimisation with box constraints. This is done by considering a specific form of opto-electronic oscillator-based coherent Ising machines with clipped transfer functions, and proposing appropriate modifications to facilitate trust-region optimisation. The enhancements include the inclusion of non-symmetric coupling and linear terms, modulation of noise, and compatibility with convex-projections to improve its convergence. The convergence of the modified Ising machine has been shown under the reasonable assumptions of convexity or invexity. The mathematical structures of the modified Ising machine and trust-region methods have been exploited to design a new trust-region method to effectively solve unconstrained optimisation problems in many scenarios, such as machine learning and optimisation of parameters in variational quantum algorithms. Hence, the proposition is useful for both classical and quantum-classical hybrid scenarios. Finally, the convergence of the Ising machine-based trust-region method, has also been proven analytically, establishing the feasibility of the technique.

1 Introduction

Ising models have traditionally been used to solve NP-hard combinatorial optimisation problems [1, 2] by exploiting the adiabatic evolution of a physical system. Specifically, such problems are solved by mapping them to the ground-state search problem of the Ising model [3], where the ground state encodes its optimal solution.

Sayantan Pramanik: sayantan.pramanik@tcs.com

Among various methods of realizing an Ising model of coupled artificial spins [4, 5, 6], an important approach is to utilise opto-electronic-oscillators (OEOs) for building a coherent Ising machine (CIM) [7, 8, 9]. A CIM implements a network of artificial spins with bistable coherent optical states for mapping the optimisation problems to the ground state of the Ising model [3, 7]. The OEO-based CIM approach particularly stands out for its cost-effectiveness, ambient operation, and scope for miniaturization [9]. Being inherently gain dissipative, it naturally approaches the optimal solution [7, 10].

In this work, we present a new application of OEO-CIMs to unconstrained optimisation. This is the first time we have analytically proven the viability of Ising machines to perform trust-region-based optimization [11, 12, 13] and refer to the technique as iTrust. We refer the reader to [11] for a comprehensive overview of trust-region methods. The main advantage of iTrust stems from the avoidance of matrix-inversion, along with the other aforementioned benefits of OEO-CIMs. This opens up a new avenue of applications where the Ising machines may be used to optimise the parameters in arbitrary objective functions, with an important example being the objective (loss/reward/penalty) functions of machine learning (ML) [14, 15, 16], quantum ML (QML) [17, 18, 19], and quantum-inspired ML (QiML) [20] models and variational quantum algorithms (VQAs) [21]. Hence, iTrust finds applicability in both classical and quantum-classical hybrid computing. More generally, the optimisation of any parametrised, unconstrained objective function $f : \mathbb{R}^n \to \mathbb{R}$ is within the purview of iTrust. We denote the parameters of the objective function $f(\cdot)$ with the vector $\boldsymbol{\theta} \in \mathbb{R}^n$. For completeness, the overarching problem that we attempt to solve using iTrust is:

Problem 1.

$$\min_{\boldsymbol{\theta}\in\mathbb{R}^n} f(\boldsymbol{\theta}), \tag{1}$$

with the aim of finding a point $\boldsymbol{\theta}^*$ which satisfies second-order optimality conditions [11], under the following generic assumption [11]:

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

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 invex [22] functions in Section 2.1. Finally, we describe the proposed algorithm iTrust in Section 3, before showing its convergence to second-order optimal solutions of Problem 1 in Theorem 3. Conclusions and future outlook are in Section 4.

2 Economical Coherent Ising Machine

For iTrust, we consider the poor man's CIM introduced in [7] with clipped nonlinearity [8], and refer to it as the Economical CIM (ECIM). It is then modified to find ε-suboptimal solutions of the following problem with $\boldsymbol{J}$ as the coupling-matrix, and $\boldsymbol{h}$ as the external field:

Problem 2.

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

Inspired by an earlier work [23], our modifications include setting $\alpha = 1$ and viewing β as the step-size in equation 8 of [8]. 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 [24, 9]. The clipping voltage is set to $\pm\Delta$, and finally, the ECIM is made compatible with the definition of projection to the convex box $\mathscr{C} = [-\Delta, \Delta]^n$. As a result, the iterative update equation of the modified ECIM is given by:

$$\boldsymbol{s}^{(k+1)} = \Pi_{\mathscr{C}}\left(\boldsymbol{s}^{(k)} - \beta_k\left(\nabla E(\boldsymbol{s}^{(k)}) - \boldsymbol{\zeta}^{(k)}\right)\right), \tag{3}$$

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

2.1 Convergence of ECIM

In this section, we present the convergence-results of the modified ECIM through the following informal Theorems. Their formal statements and proofs have not been included for 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 $\mathscr{C}$ 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 $\sum_{k=0}^{\infty}\beta_k^2 < \infty$, $\lim_{k\to\infty}(E(\boldsymbol{s}^{(k)}) - E^*) = 0$, where $E^* = \min_{\boldsymbol{s}\in\mathscr{C}} E(\boldsymbol{s})$.*

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

If $\boldsymbol{\mathcal{s}}$ 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 [13]:

$$-E(\boldsymbol{\mathcal{s}}) \geq c|E(\boldsymbol{s}^*)|. \tag{4}$$

3 iTrust

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

Problem 3.

$$\min_{||\boldsymbol{p}||_2\leq\delta_t} \left(m_t(\boldsymbol{p}) \triangleq \langle \nabla f(\boldsymbol{\theta}^{(t)}), \boldsymbol{p}\rangle + \frac{1}{2}\langle \boldsymbol{p}, \boldsymbol{H}(\boldsymbol{\theta}^{(t)})\boldsymbol{p}\rangle \right), \tag{5}$$

where $\nabla f(\boldsymbol{\theta}^{(t)})$ and $\boldsymbol{H}(\boldsymbol{\theta}^{(t)})$ are the gradient and Hessian of f at $\boldsymbol{\theta}^{(t)}$, respectively. A major disadvantage of using the method proposed in Algorithm 3.14 proposed in [13] to find $\boldsymbol{p}^*_{(t)}$ 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 $\boldsymbol{p}^*_{(t)}$. We achieve this by exploiting the structural similarity Problems 2 and 3. Specifically, at each iteration t, $\boldsymbol{J}$ is set to $\boldsymbol{H}(\boldsymbol{\theta}^{(t)})$, $\boldsymbol{h}$ 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 $\nabla E(\boldsymbol{s}^{(k)})$ could not have been provided to the ECIM without additional overheads in the form of ancillary spins [23, 24]. Also, by design, the constraint set of Problem 3 is contained in $[-\Delta_t, \Delta_t]^n$. As a result, it is guaranteed that $E_t(\boldsymbol{s}^*_{(t)}) \leq m_t(\boldsymbol{p}^*_{(t)})$. We name this technique of using the ECIM for trust-region optimisation as *i*Trust. The workflow for *i*Trust has been portrayed in Algorithm 1, which draws inspiration from, and is an amalgamation of, Algorithms 4.1 and 4.2 of [11, 13], respectively.

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 $\mathscr{S}$. This claim is formalised in the form of the following theorem [11, 13], the proof of which has been omitted for brevity:

Theorem 3 (Convergence of *i*Trust)**.** *Let assumption 1 be true, and let* $(\boldsymbol{\theta}^{(t)})$ *be the sequence of iterates generated by Algorithm 1 such that equation (4) is satisfied at each iteration. Then we have that:*

$$\lim_{t\to\infty} ||\nabla f(\boldsymbol{\theta}^{(t)})||_2 = 0. \tag{6}$$

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

4 Conclusions and Outlook

In this paper, we introduced *i*Trust, 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 *i*Trust. We look forward to validate our theoretical results by experimenting extensively with the proposed algorithm. Possible future directions may include the investigation of the performance of the ECIM for other classes of objective functions besides convex and invex ones. Variants of *i*Trust can also be constructed that are compatible with natural gradient descent [25, 26], by replacing the Hessian with the Fisher Information Matrix. *i*Trust may be further augmented by zeroth order methods like SPSA [27] in scenarios where evaluation of the gradients, Hessian, and Fisher information matrix is computationally expensive [28]. Lastly, the advantages of the

Algorithm 1: *i*Trust

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

begin
- **for** $t \in [T]$ **do**
  - evaluate $\nabla f(\boldsymbol{\theta}^{(t)})$ and $\boldsymbol{H}(\boldsymbol{\theta}^{(t)})$;
  - $\boldsymbol{J}^{(t)} \leftarrow \boldsymbol{H}(\boldsymbol{\theta}^{(t)})$;
  - $\boldsymbol{h}^{(t)} \leftarrow \nabla f(\boldsymbol{\theta}^{(t)})$;
  - $\Delta_t \leftarrow \delta_t$;
  - initialise $\boldsymbol{s}^{(0)}$ randomly in $\mathscr{C}_t = [-\Delta_t, \Delta_t]^n$;
  - **for** $k \in [K]$ **do**
    - sample $\boldsymbol{\zeta}^{(k)} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$;
    - $\boldsymbol{s}^{(k+1)} = \Pi_{\mathscr{C}_t}\left(\boldsymbol{s}^{(k)} - \beta_k\left(\nabla E_t(\boldsymbol{s}^{(k)}) - \boldsymbol{\zeta}^{(k)}\right)\right)$;
  - **end**
  - calculate $\rho_t = \frac{f(\boldsymbol{\theta}^{(t)} + \boldsymbol{s}^{(K)}) - f(\boldsymbol{\theta}^{(t)})}{E_t(\boldsymbol{s}^{(K)})}$;
  - **if** $\rho_t < \mu$ **then**
    - $\delta_{t+1} = \gamma_1 \delta_t$;
    - **continue**;
  - else
    - **if** $\rho_t > (1-\mu)$ *and* $||\boldsymbol{s}^{(K)}||_\infty = \delta_t$ **then**
      - $\delta_{t+1} = \min(\gamma_2 \delta_t, \delta_{\max})$;
    - **else**
      - $\delta_{t+1} = \delta_t$;
    - **end**
  - **end**
  - **if** $\rho_t > \eta$ **then**
    - $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} + \boldsymbol{s}^{(K)}$;
  - **else**
    - $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$;
  - **end**
- **end**
- **return** $\boldsymbol{\theta}^{(T)}$

end

ECIM over noisy projected gradient descent for the subproblem-minimisation 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] Andrew Lucas. Ising formulations of many np problems. *Frontiers in Physics*, 2, 2014.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] Timothée Leleu, Yoshihisa Yamamoto, Shoko Utsunomiya, and Kazuyuki Aihara. Combinatorial optimization using dynamical phase transitions in driven-dissipative systems. *Physical Review E*, 95(2), February 2017.

[11] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. Springer, New York, NY, USA, 2e edition, 2006.

[12] D. C. Sorensen. Newton's method with a model trust region modification. *SIAM Journal on Numerical Analysis*, 19(2):409–426, 1982.

[13] Jorge J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, 4(3):553–572, 1983.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] Larry Huynh, Jin Hong, Ajmal Mian, Hajime Suzuki, Yanqiu Wu, and Seyit Camtepe. Quantum-inspired machine learning: a survey, 2023.

[21] 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.

[22] Hamed Karimi, Julie Nutini, and Mark Schmidt. Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition, 2020.

[23] Sayantan Pramanik, Sourav Chatterjee, and Harshkumar Oza. Convergence analysis of opto-electronic oscillator based coherent ising machines. In *2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS)*, pages 1076–1081, 2024.

[24] Abhishek Kumar Singh, Kyle Jamieson, Davide Venturelli, and Peter McMahon. Ising machines' dynamics and regularization for near-optimal large and massive mimo detection, 2021.

[25] Shun-ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, February 1998.

[26] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum natural gradient. *Quantum*, 4:269, may 2020.

[27] James C. Spall. An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins Apl Technical Digest*, 19:482–492, 1998.

[28] Julien Gacon, Christa Zoufal, Giuseppe Carleo, and Stefan Woerner. Simultaneous perturbation stochastic approximation of the quantum fisher information. *Quantum*, 5:567, October 2021.