

Towards Evolutionary Optimization Using the Ising Model

Simon Klüttermann
 simon.kluettermann@cs.tu-dortmund.de

Abstract—In this paper, we study the problem of finding the global minima of a given function. Specifically, we consider complicated functions with numerous local minima, as is often the case for real-world data mining losses.

We do so by applying a model from theoretical physics to create an Ising model-based evolutionary optimization algorithm.

Our algorithm creates stable regions of local optima and a high potential for improvement between these regions. This enables the accurate identification of global minima, surpassing comparable methods, and has promising applications to ensembles.

Index Terms—Evolutionary Optimization, Evolutionary Data Mining, Optimization Algorithms, Physics-inspired Algorithms, Ensembles

I. INTRODUCTION

The development of gradient descent [1] made it possible to create and improve many data mining algorithms, for example, by allowing methods based on neural networks to be trained efficiently. Still, gradient descent methods have their limits. Most importantly, gradient descent requires a continuous loss function and thus cannot be applied to every type of model. While we have historically preferred to describe complicated relations using intricate mathematical formulas, with gradient descent, we must hide these precise mathematical objects within huge neural network matrices. This is not always possible [2] and made it almost impossible to understand the underlying relation [3].

A potential solution to improve this would be to employ more specialized models and optimize them using a different paradigm, such as evolutionary optimization [4], [5]. While evolutionary optimization is often much more time-consuming, variants are applicable to almost every minimization task. Neither the loss function nor its inputs need to be continuous, and it can even allow for optimizing more complicated structures, such as lists of variable length or a tree encoding a mathematical function.

However, due to the increased time cost, evolutionary optimization is typically specialized only to find a good enough optimum. While evolutionary approaches are less likely to get stuck in a local optimum than gradient descent methods, we will see later that they are usually not able to even come close to the global optimum. While this is a good approach, for example, in hyperparameter optimization [6], where function evaluations are very costly, and the performance increase might often be tiny, the situation is arguably different in data mining. When searching for an explainable model, only locally optimal solutions also represent a flawed explanation of the underlying process, as there is an explanation that describes

it more effectively. This also implies that a locally optimal solution generalizes less well to a new dataset.

To test algorithms in this problem situation, we propose a test function in this paper, which features many local minima of varying qualities. These are also distributed in a way that makes it easy to evaluate the number of local minima found.

Using this test situation, we propose a new evolutionary optimization paradigm that is particularly effective in finding global minima. This algorithm is inspired by the Ising model [7] from statistical physics, which is often studied because it exhibits the emergence of regions with slightly different behavior. We observe that when extending this algorithm to evolutionary optimization, the emergence of regions remains. This enables our optimization algorithm to consider a much wider range of possible solutions, thereby increasing the likelihood of finding the global minimum.

Additionally, we notice that our optimization can be very helpful for ensemble models [8]. While finding the global optimum of a loss function usually implies a valuable data mining model, combining different models into an ensemble can be even more effective. This is because while a single model might make mistakes, these mistakes are often canceled by averaging multiple ones [9].

However, to do so, we must find multiple significantly different solutions to a given data mining task, as repeating the same errors does not allow them to cancel each other out. This is not necessarily an easy task, as it often relies on different initializations finding different local minima. And even if it works, it requires running a potentially very slow evolutionary optimization algorithm many times.

The diversity of solutions our algorithm provides can solve both problems. Instead of requiring multiple optimizations, we can benefit from different solutions in different regions.

Concretely, our contributions in this paper are threefold:

- Definition of an easy-to-study function for the problem of evolutionary global minima optimization.
- Suggestion of a new physics-based optimization algorithm.
- Evaluation of its properties and benefits.

To help with further research, our code is available at <https://github.com/psorus/IsingEvo>.

II. RELATED WORK

A. Evolutionary Optimization

The problem of optimization [10] is central to machine learning. As a result, there are several ways to address this

issue. A common method is Gradient Descent [1], which efficiently trains complicated neural networks. However, Gradient Descent requires a continuous loss function and struggles to handle randomness effectively. Both are conditions that restrict the type of problems that can be optimized. Additionally, Gradient Descent can often not find the global minimum of a loss function, but only a local minimum.

Evolutionary optimization solves some of these problems. Instead of requiring a continuous loss function differential, it alters the current input randomly and either accepts or rejects the changed value depending on whether it decreases the loss function. This allows applying it to both functions of non-numeric values, as well as to loss functions that are not differentiable [4].

Additionally, there are numerous variants of evolutionary optimization algorithms. Notable are Genetic Algorithms [11], Particle Swarm Optimization [12], and Differential Evolution [13]. A few specific algorithms are described in Sec. III-A.

When searching for the global optimum, the optimization algorithm needs to be able to make a locally suboptimal choice. While this is easy to implement for an evolutionary algorithm, by adding a probability to create a suboptimal choice, it is not optimal. Theoretically, it might be possible to go from a local optimum to a different one, but this is still very unlikely, as it often requires choosing the suboptimal choice many times in a row. Instead, we suggest a method to increase the variance of inputs tried, which makes it considerably more likely to find the global minimum.

B. Ising Model

We achieve this by utilizing an Ising model-inspired algorithm. This model, studied in 1924 by Ernest Ising in his dissertation [7], was initially used to model spin interactions: Two neighboring spins are either aligned ($s_i \cdot s_j = 1$) or opposite to each other ($s_i \cdot s_j = -1$), and the energy of the system (similar to a loss function) depends on this interaction $E = -\sum_{i \text{ neighbour of } j} s_i \cdot s_j$.

It is one of the most commonly studied models in statistical physics, because it is one of the simplest models to show a phase transition [14]. This means its behavior is drastically different based on one control parameter (the temperature T or the more commonly used inverse temperature $\beta = \frac{1}{T}$): For low temperatures, spins are aligned and stable regions form. But for high temperatures, the spins become chaotic.

Notably, this effect also depends on the dimension in which it is studied. In one dimension, every spin has only two neighbors, and the system behaves unremarkably. However, it exhibits the aforementioned phase transitions in two or more dimensions (i.e., four or more neighbors).

While the Ising model has originally been used to model magnetism, by now there are many more applications: From protein folding [15] to social sciences [16] and modeling the stock market [17], by now there are many more problems that can be modeled by an Ising model.

C. Science-Inspired Machine learning

There is a deep connection between machine learning and natural sciences, such as physics. Not only are machine learning methods often used to conduct scientific research [18]–[20], but knowledge from science is also frequently utilized to enhance machine learning methods, invent new ones, or elucidate existing ones.

Statistical physics can be used to explain the behavior of neural networks [21] and has directly inspired the development of Boltzmann machines [22]. Particle swarm optimization [12] and Hopfield networks [23] are inspired by the behavior of animals and neurons, respectively.

Notably, Simulated Annealing [24], a commonly used evolutionary optimization strategy inspired by metallurgical research, utilizes a temperature variable to control the randomness of a system and guide it toward an optimal state. We will use this algorithm as a competitor.

Additionally, Ising models have been utilized to comprehend the training of neural networks [25] and have inspired Markov random fields [26], which are employed to model joint probability distributions. Still, to the best of our knowledge, they have not been used as a basis for an evolutionary algorithm.

D. Ensemble methods

Replacing a complex model with an ensemble of models can often be an effective way to enhance the model's performance and increase its reliability. While there are many approaches to ensembles (like stacking or boosting [27], [28]), the most commonly studied task combines multiple independent models using a combination function [29]. For this to help, we require the errors of ensemble submodels to differ from each other [9].

This difference can be achieved in one of two ways. Either by changing the setup of each submodel, or by, for example, training different types of models. However, this makes the resulting combination task harder and can often (especially for unsupervised tasks) limit the resulting performance [30], [31].

So instead, submodels are often identical, except for a random seed [32]. However, combining this with an optimization algorithm like gradient descent [1] or evolutionary optimization [4] means that we can only achieve the variance needed for the ensemble to work when there are many different, equally easily reachable local minima. Since this assumption is often not fulfilled and requires retraining many submodels multiple times, we instead suggest a new optimization method that can directly find a set of local minima and thus an entire ensemble during training.

III. ISING BASED EVOLUTION

We describe our approach in Alg. 1. We present here the two-dimensional version, but this can be easily adapted by modifying what it means for two indices to be neighboring. While Ising-based evolution is similar to cellular evolution

[33] as updates happen locally, the likelihood of changing a value is given by Eq. (1):

$$p(\text{old}, \text{new}) = \begin{cases} 1 & \text{if } \text{old} > \text{new} \\ \exp(\beta \cdot (\text{old} - \text{new})) & \text{else} \end{cases} \quad (1)$$

When minimizing the existing value, the probability is $p = 1$, and such the update step happens. However, if it is not improved, there is still a chance that the value will get updated. This is necessary for the method to avoid getting stuck in local minima. Here, the probability is bigger when β is smaller. This is the same probability used in the usual Ising model to go to a state of different energy.

We chose here a mutation of our value that is either sexual or asexual: Either two values are averaged, or one value is changed according to a normal distribution. Still, this mutation is not part of our algorithm and can be changed at will. We will also use an equivalent mutation step in our competitor algorithms.

Algorithm 1 Ising based evolution

```

Require:  $f(x)$ 
Require:  $\text{Region}$ ,  $\text{wid}$ ,  $\text{hei}$ ,  $\beta$ ,  $n_{\text{steps}} > \text{wid} \cdot \text{hei}$ 
Require:  $\text{uniform}(\text{Region})$ ,  $\text{normal}(\text{mean}, \text{std})$ 
for  $i < \text{wid}$  do
    for  $j < \text{hei}$  do
         $\text{population}_{i,j} \leftarrow \text{uniform}(\text{Region})$ 
         $\text{qual}_{i,j} \leftarrow f(\text{population}_{i,j})$ 
         $n_{\text{steps}} \leftarrow n_{\text{steps}} - 1$ 
    end for
end for
 $\text{solution} \leftarrow \text{population}_{\text{argmin}(\text{qual})}$ 
 $\text{qual}_{\text{min}} \leftarrow \min(\text{qual})$ 
while  $n_{\text{steps}} > 0$  do
     $i \leftarrow \text{uniform}(0, \text{wid})$ 
     $j \leftarrow \text{uniform}(0, \text{hei})$ 
     $\bar{i}, \bar{j} \leftarrow \text{neighbour}(i, j)$ 
     $\text{test} \leftarrow \frac{\text{population}_{i,j} + \text{population}_{\bar{i}, \bar{j}}}{2}$ 
    if  $\text{uniform}(0, 1) < 0.5$  then
         $\text{test} \leftarrow \text{normal}(\text{mean} = \text{population}_{i,j}, \text{std} = 100)$ 
    end if
     $\text{quality}_{\text{test}} \leftarrow f(\text{test})$ 
    if  $\text{uniform}(0, 1) < \exp(\beta \cdot (\text{qual}_{i,j} - \text{quality}_{\text{test}}))$  then
         $\text{population}_{i,j} \leftarrow \text{test}$ 
         $\text{qual}_{i,j} \leftarrow \text{quality}_{\text{test}}$ 
        if  $\text{qual}_{\text{test}} < \text{qual}_{\text{min}}$  then
             $\text{solution} \leftarrow \text{test}$ 
             $\text{qual}_{\text{min}} \leftarrow \text{quality}_{\text{test}}$ 
        end if
    end if
     $n_{\text{steps}} \leftarrow n_{\text{steps}} - 1$ 
end while
return  $\text{solution}$ 

```

A. Comparison algorithms

We chose five different evolutionary optimization strategies to compare our algorithm to.

Cellular evolution [33] is a similar approach to Alg. 1. The main difference to our approach is that the probability of choosing a suboptimal solution is fixed at 10%. While this change may seem small, it would already be sufficient to prevent the emergence of a phase transition in the physical model.

Simulated Annealing [24] is also a physics-inspired algorithm in which one solution is updated based on mutations, with a probability given by Eq. (1). As it is inspired by the cooling of a metal, the value of β increases with the number of update steps i . A common way to parameterize this is $\beta = \beta_0 \cdot i$. However, to allow for more variance with a high number of update steps and have a stronger competitor, we choose $\beta = \beta_0 \cdot \sqrt{i}$.

Mutation evolution is a much simpler algorithm, which generates a new value through a standard distribution $\text{newvalue} = \text{normal}(\text{mean} = \text{current Value}, \text{std} = 100)$. The value is updated if the new value has a lower score than the old one. We employ this algorithm here, similar to an ablation study, as it utilizes a similar update step to our algorithm. Still, Mutation evolution is a case of differential evolution [13].

Mixture evolution is a further case of differential evolution [13], which we use like an ablation study. Here we are given a population of 100 random values, and we randomly average the value of two instances to generate a new one. If this generated value outperforms one of its parents, or in at least 10% of cases, we keep it. While Mutation evolution could be seen as inspired by asexual reproduction, Mixture evolution is more similar to sexual reproduction.

Random Search is our final and simplest algorithm. Instead of using a procedure to update the current value, we randomly guess it. While this algorithm cannot get stuck in local minima, it can also not benefit from any local feedback. The likelihood of reaching the optimal state is the same, whether in the second-best state or in the worst one.

IV. TEST DESCRIPTION

To test how well a particular algorithm performs in finding the global minimum of a function, we need to evaluate it on a function with many local optima. To achieve this, we suggest minimizing Equation (2).

$$f(x) = \left| \sin \left(\frac{x+1}{100} \right) \right|, \quad x \in \mathbb{Z} \quad (2)$$

We visualize this function in Fig. 1.

While Eq. (2) is almost a simple sine function, the evaluation over the range of integers means that most minima are only local minima (See Theorem 1). Still, the periodicity and improving approximation of a real number impart the function with certain general properties. For example $f(x) \approx 0 \rightarrow f(n \cdot x) \approx 0 \forall n \in \mathbb{N}$.

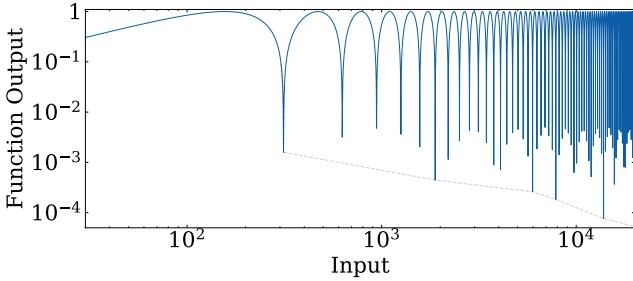


Fig. 1. Visualization of Eq. (2). Because of the double logarithmic axes, the sinus looks unusual. But this highlights that the local minima are not randomly distributed.

Theorem 1. *The only global optimum to function (2) is $x = -1$.*

Proof.

$$f(-1) = |\sin(0)| = 0; \quad f(x) \geq 0 \quad \forall x$$

$$f(x) = 0, x \neq -1 \implies \frac{x+1}{100} = k \cdot \pi, k \in \mathbb{Z} \setminus \{0\}$$

$$\implies \pi = \frac{x+1}{100 \cdot k} \in \mathbb{Q} \quad \text{which contradicts that } \pi \text{ is irrational.}$$

□

Thus, we can exclude the global minimum by only considering $x \in \mathbb{N}$. We also only evaluate $x < 10^5$ to ensure an optimal result and to put evaluation costs into perspective. In total, this means there are 318 local minima of our function.

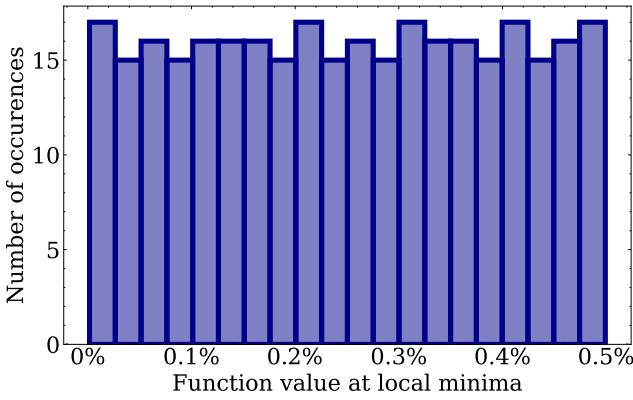


Fig. 2. Distribution of locally minimal function values.

Further, as shown in Fig. 2, the quality of minima is equally distributed. This means, we can use the average found function value, to evaluate an optimization algorithm: When an algorithm only finds a local minimum, its average performance will be about $0.25\% = 2.5 \cdot 10^{-3}$. A more efficient algorithm would instead find the lowest point of our range, $f(84822) \approx 1.64 \cdot 10^{-5}$.

V. RESULTS

Using this setup, we randomly initialize and run each algorithm 100 times, measuring the average result found. Each algorithm can evaluate the function up to 10^5 times to find the lowest function value. In theory, every function value in our range could be evaluated, guaranteeing that the lowest value is found. Still, no algorithm reliably finds the lowest performance all the time since we allow our algorithm to make the same function call multiple times. While repetitions should usually be cached in an actual application, this implicitly allows us to punish algorithms that do not efficiently explore the entire search space and guarantees that algorithms run in finite time.

The resulting average function values are shown in Fig. 3.

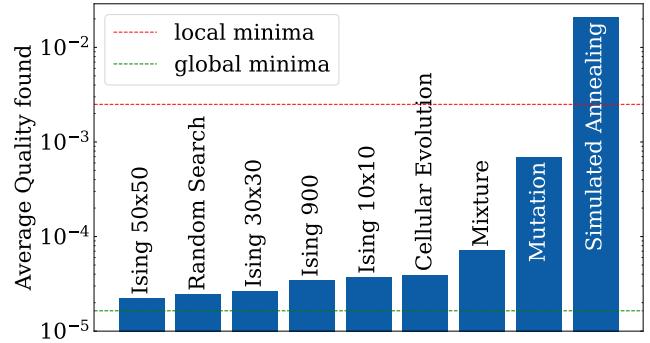


Fig. 3. Average minimal value found for different evolutionary optimization algorithms.

The Mixture algorithm does not improve much upon finding a local minimum. Simulated Annealing also freezes quickly. Interestingly, the Mixture algorithm does much better than both. This is likely because if $f(x) \approx 0 \approx f(y)$ this implies that $f(\frac{x+y}{2}) \approx 0$ or $f(\frac{x+y}{2}) \approx 1$, which is a property we can not expect to generalize to different tasks.

Each of our Ising-based models performs much better, with the best reaching an average performance of $\approx 2.2 \cdot 10^{-5}$, close to the theoretical optimum. We will spend the remainder of this section to further understand this concept. Importantly, our Ising-based optimization outperforms the related Cellular Evolution. We used $\beta = 100$ here, but choosing a different value does not significantly affect the resulting quality, up to the limit studied in Section V-A.

The Random Search model seems to be the most robust competitor algorithm. This is a result of our construction: The likelihood that it does not find the optimal value is $(1 - \frac{1}{10^5})^{10^5} \approx 0.3679 \approx \frac{1}{e}$, and we could improve it further by removing duplicate function calls. Still, while it can find absolute minima relatively commonly, it cannot benefit from local feedback. This means its practical applications are limited, as most practical functions do not contain as many local minima, and a higher number of optimizable parameters exponentially increases the number of samples to be tried.

We study this further by comparing the performance of our Ising-based approach with that of Random Search as a function of the number of function evaluations allowed. This is shown in Fig. 4.

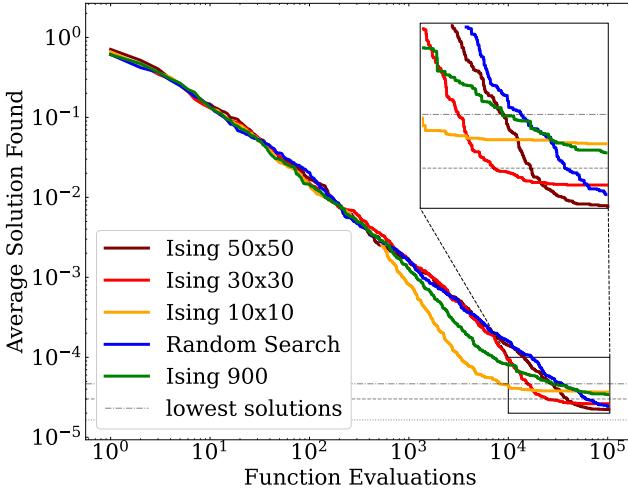


Fig. 4. Average lowest value found as a function of the number of function evaluations used. We zoom in on this region because the most interesting changes happen in the lower right corner. We also add the lowest three possible solutions of Eq. (2) to be found as horizontal lines.

It seems that for the first function evaluations, each algorithm shown performs very similarly. This is because the population is first initialized randomly, making the Ising-based algorithm equivalent to the Random Search for 10^2 to $25^2 \approx 10^{2.8}$ function evaluations. Afterward, each Ising-based algorithm decreases more quickly than the Random Search algorithm, which means that they are faster in reaching a better solution than the Random Search and demonstrate the benefit of local feedback in our algorithms. The change might look small, but due to the double logarithmic axes, even a small change can result in a tenfold speed increase. But the model performance converges at some point and does not improve further. Both the resulting quality and the cost of evaluations seem to be a function of the size of the population of our algorithm: the bigger the board, the longer we follow the Random Search performance, but the lower the resulting quality is also. And ultimately, there is a point where a 50x50 population converges at a quality below Random Search in $\approx 50\%$ fewer function evaluations.

As an ablation study, we also show a one-dimensional Ising-based model with the same population size of the 30x30 two-dimensional model. However, it appears to converge more slowly and exhibit poorer performance than its two-dimensional alternative. This is expected from the behaviour of the physical Ising model and will be further understood in the following subsections.

A. Phase transitions

The Ising model is often studied in physics because it is one of the simplest models that can show phase transitions. While for high temperatures (low β), the system is chaotic and almost random, for low temperatures (high β), relatively stable regions of aligned spins are formed. Interestingly, this effect does not appear for the one-dimensional Ising model but

only in at least two dimensions. We aim to investigate whether a similar transition also occurs here.

To illustrate this, we present the current state of our population for various temperatures after different numbers of update steps for the 30x30-sized algorithm in Fig. 5.

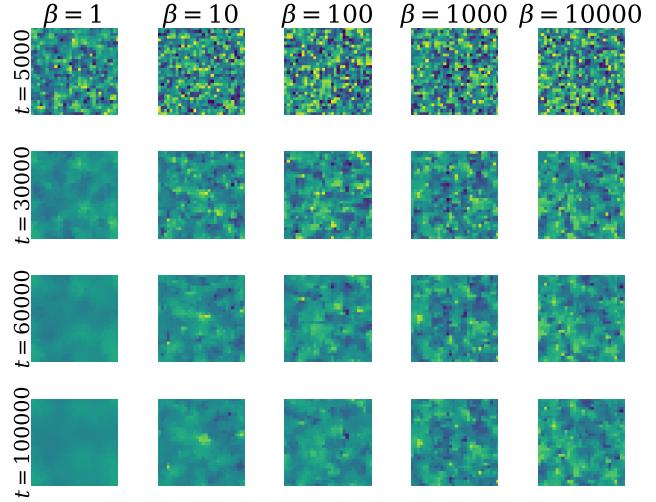


Fig. 5. Current state (color) over the population, as a temperature and update step function.

Similar to the physical effect, we see a relatively stable distribution. After the almost random initialization, each model exhibits unique structures, which persist even after further update steps, albeit with slight changes (for example, the yellow spot in the center of $\beta = 10$).

Similarly, this distribution also depends on the temperature value, just like the physical effect. For a high temperature (low β), the resulting structures seem to disappear the most with increasing update steps. However, for low temperatures (high β), the results become more chaotic, and stable regions emerge.

But after 10^5 update steps, even the almost convergent distribution of $\beta = 1$ still differs by about $\approx 15\%$ of the available range in its solutions.

To further characterize how the distribution depends on temperature and the update step, we visualize the standard deviation of these images as a function of both. We show this for a two-dimensional model in Fig. 6 and a one-dimensional one in Fig. 7.

As we have seen before, in the two-dimensional case, the system becomes more chaotic with a higher value of β , while also needing a longer time to converge. We also almost see phases: For a low number of evaluations and a high value of β , the variance is high, while for a high number of evaluations and a low β it is significantly smaller. However, importantly, the population variance does not drop to zero, unlike the other evolution algorithms. This is not only an effect of continuing random mutation steps, as these should at most create a relative standard deviation of $\frac{100}{10^5} = 0.001$ (See Alg. 1).

Interestingly, and similarly to the physical effect, the one-dimensional case appears to behave quite differently from the

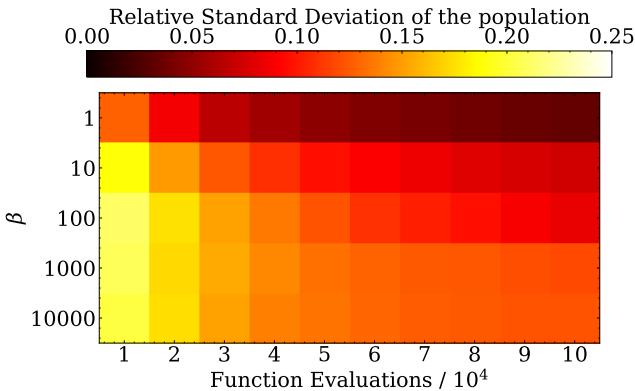


Fig. 6. Relative standard deviation (standard deviation divided by 10^5) of models with different β value during the optimization process for a two-dimensional ising model of population size 30x30.

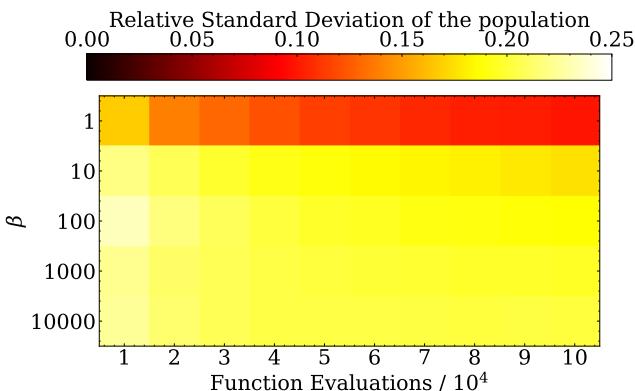


Fig. 7. Relative standard deviation (standard deviation divided by 10^5) of models with different β value during the optimization process for a one-dimensional ising model of population size 900.

two-dimensional one. While the two-dimensional model shows a gradual transition from the low to the high variance phase, the one-dimensional one freezes out almost completely for every $\beta > 1$.

B. Ensembles

Finally, we aim to investigate how effectively our Ising-based approach can be utilized to identify not only the global minimum of a function but also multiple distinct ones that can be combined into an ensemble. To do this, we test not how often the global minimum is found but how many of the lowest ten solutions are found. This is shown in Fig. 8.

Our Ising-based approach identifies most of these minima, with an average of 7.44 out of 10 valuable minima being found. Similar to our analysis in Fig. 3, a larger size of the Ising population appears to be beneficial, and here the improvement over the Random Search algorithm increases, as even a 30x30 population is sufficient to outperform it.

This demonstrates that utilizing submodels from various regions can yield ensembles that outperform those randomly initialized. We can create an ensemble from our results by requiring only one optimization loop, rather than multiple

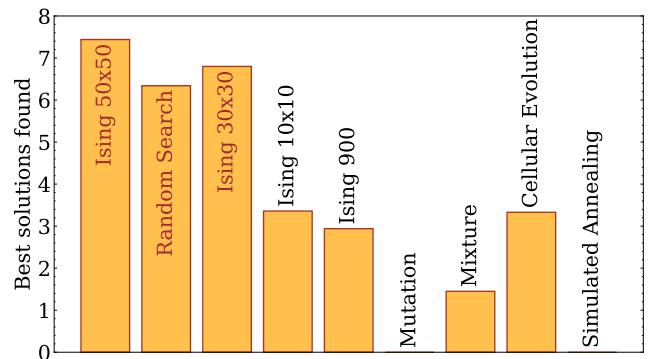


Fig. 8. When not only searching for the best solution but also for the 9 best local optima, how many of these are found on average by a given method.

ones (one per submodel). Additionally, we could link the number of different solution regions to the number of ensemble submodels needed. But this still requires further study.

VI. CONCLUSION

This paper introduced an evolutionary optimization algorithm based on Ising models. Our method is capable of finding global minima even in highly complex loss functions. It achieves this by leveraging knowledge from theoretical physics, adapting a model that demonstrates the emergence of semi-stable regions, and then adjusting these regions to evolutionary optimization. We have demonstrated that our algorithm exhibits properties similar to those that make Ising models so interesting in physics.

We can utilize these regions to perform what we call high-variance optimization. This kind of optimization allows us not only to find better optima but also has interesting applications for ensembles.

In further studies, we aim to characterize other shapes of our population (higher dimensionality, non-square shapes) and investigate whether we can improve the trade-off between quality and time cost by using a region-based initialization. Applying this method to various other loss functions and mutation procedures would also be interesting.

Further, we are interested in the possibility of learning complicated and inherently explainable models (like a function tree) using an algorithm like ours.

REFERENCES

- [1] S. Ruder, “An overview of gradient descent optimization algorithms,” *arXiv preprint arXiv:1609.04747*, 2016.
- [2] K. Scaman and A. Virieux, “Lipschitz regularity of deep neural networks: Analysis and efficient estimation,” in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, ser. NIPS’18. Red Hook, NY, USA: Curran Associates Inc., 2018, p. 3839–3848.
- [3] N. Burkart and M. F. Huber, “A survey on the explainability of supervised machine learning,” *ArXiv*, vol. abs/2011.07876, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:226964842>
- [4] C. Blum, R. Chiong, M. Clerc, K. De Jong, Z. Michalewicz, F. Neri, and T. Weise, *Evolutionary Optimization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 1–29. [Online]. Available: https://doi.org/10.1007/978-3-642-23424-8_1

- [5] J. A. Vrugt and B. A. Robinson, "Improved evolutionary optimization from genetically adaptive multimethod search," *Proceedings of the National Academy of Sciences*, vol. 104, no. 3, pp. 708–711, 2007. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.0610471104>
- [6] C. Wang, Q. Wu, M. Weimer, and E. E. Zhu, "Flaml: A fast and lightweight automl library," in *Fourth Conference on Machine Learning and Systems (MLSys 2021)*, April 2021. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/flaml-a-fast-and-lightweight-automl-library/>
- [7] E. Ising, "Beitrag zur Theorie des Ferromagnetismus," *Zeitschrift für Physik*, vol. 31, no. 1, pp. 253–258, Feb. 1925.
- [8] A. Zimek, R. J. Campello, and J. Sander, "Ensembles for unsupervised outlier detection: Challenges and research questions a position paper," *SIGKDD Explor. Newsl.*, vol. 15, no. 1, p. 11–22, mar 2014. [Online]. Available: <https://doi.org/10.1145/2594473.2594476>
- [9] K. Ali and M. Pazzani, "Error reduction through learning multiple descriptions," *Machine Learning*, vol. 24, 11 1997.
- [10] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012.
- [11] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091–8126, Feb 2021. [Online]. Available: <https://doi.org/10.1007/s11042-020-10139-6>
- [12] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, 1995, pp. 1942–1948 vol.4.
- [13] R. Storn and K. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, Dec 1997. [Online]. Available: <https://doi.org/10.1023/A:1008202821328>
- [14] A. Taroni, "90 years of the ising model," *Nature Physics*, vol. 11, no. 12, pp. 997–997, Dec 2015. [Online]. Available: <https://doi.org/10.1038/nphys3595>
- [15] A. Bakke and J. S. Høye, "One-dimensional ising model applied to protein folding," *Physica A: Statistical Mechanics and its Applications*, vol. 323, pp. 504–518, 2003. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378437103000189>
- [16] C. Li, F. Liu, and P. Li, "Ising model of user behavior decision in network rumor propagation," *Discrete Dynamics in Nature and Society*, vol. 2018, p. 5207475, Aug 2018. [Online]. Available: <https://doi.org/10.1155/2018/5207475>
- [17] L. Zhao, W. Bao, and W. Li, "The stock market learned as ising model," *Journal of Physics: Conference Series*, vol. 1113, no. 1, p. 012009, nov 2018. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/1113/1/012009>
- [18] C. Xu and S. A. Jackson, "Machine learning and complex biological data," *Genome Biology*, vol. 20, no. 1, p. 76, Apr 2019. [Online]. Available: <https://doi.org/10.1186/s13059-019-1689-0>
- [19] Y.-H. He, E. Heyes, and E. Hirst, "Machine learning in physics and geometry," 03 2023.
- [20] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, and A. Tkatchenko, "Combining machine learning and computational chemistry for predictive insights into chemical systems," *Chemical Reviews*, vol. 121, no. 16, pp. 9816–9872, Aug 2021. [Online]. Available: <https://doi.org/10.1021/acs.chemrev.1c00107>
- [21] K. Weng, A. Cheng, Z. Zhang, P. Sun, and Y. Tian, "Statistical physics of deep neural networks: Initialization toward optimal channels," *Phys. Rev. Res.*, vol. 5, p. 023023, Apr 2023. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevResearch.5.023023>
- [22] A. Patel and R. K. Rama, "An overview of boltzmann machine and its special class," 09 2020.
- [23] H. Ramsauer, B. Schäfl, J. Lehner, P. Seidl, M. Widrich, L. Gruber, M. Holzleitner, M. Pavlović, G. Sandve, V. Greiff, D. Kreil, M. Kopp, G. Klambauer, J. Brandstetter, and S. Hochreiter, "Hopfield networks is all you need," 07 2020.
- [24] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science (New York, N.Y.)*, vol. 220, pp. 671–80, 06 1983.
- [25] G. Tkacik, E. Schneidman, M. Berry II, and W. Bialek, "Ising models for networks of real neurons," 12 2006.
- [26] R. Kindermann and L. Snell, *Markov random fields and their applications*. American Mathematical Society, 1980, vol. 1.
- [27] M. O. Sandim, "Using stacked generalization for anomaly detection," Ph.D. dissertation, 2017.
- [28] Y. Freund and R. E. Schapire, "A short introduction to boosting," in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*. Morgan Kaufmann, 1999, pp. 1401–1406.
- [29] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, Aug 1996.
- [30] S. Klütermann and E. Müller, "Evaluating and comparing heterogeneous ensemble methods for unsupervised anomaly detection," 06 2023, pp. 1–8.
- [31] A. Lazarevic and V. Kumar, "Feature bagging for outlier detection," vol. 21, 01 2005, pp. 157–166.
- [32] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. [Online]. Available: <https://doi.org/10.1023/a:1010933404324>
- [33] N. Noman and H. Iba, "Cellular differential evolution algorithm," in *AI 2010: Advances in Artificial Intelligence*, J. Li, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 293–302.