Simulated Quantum Annealing of Feed-Forward Neural Network Synaptic Weights

Justin R. Fletcher, *Student Member, IEEE,* Michael J. Mendenhall, *Member, IEEE*

Abstract—In this paper we present a methodology for feed-forward neural network synaptic weight selection which employs a quantum-inspired variant of simulated annealing.
The algorithm presented in this paper does not require any explicit temperature schedule tuning, and runs continuously.

Index Terms—simulated quantum annealing, neural networks. stochastic gradient descent.

◆

1 INTRODUCTION

.. the most common method of weight selection is back-propagation. Back propagation is a [direct] gradient descent method, which traverses the error surface along its maximum gradient path. The algorithm converges once a minimum located, because there is no longer a negative gradient for the algorithm to follow. While this method can be very successful, it often suffers from premature convergence. That is, the algorithm finds a local, rather than global, error surface minima.

2 LITERATURE REVIEW

Simulated annealing was first applied to neural network weight selection

Though simulated annealing is effective at finding synaptic weight configurations which perform comparatively well [], the method has several significant shortcomings. Neural networks trained using simulated annealing are often converge slowly relative to other conventional methods of weight selection such as back-propagation []. Simulated annealing also requires the a priori specification of a temperature schedule, which must cool sufficiently slowly to ensure convergence to a global, rather than local, minimum. The amount of cooling time required is highly-problem dependent. These limitations motivated the study of alternative methods, which led to the work presented in this paper.

Simulated annealing is a stochastic optimization algorithm which can be used to find the global minimum of a cost function mapped from the configurations of a combinatorial optimization problem. The concept of simulated annealing was introduced in by Kirkpatrick et. al. in [?] as an application of the methods of statistical mechanics to the problem of discrete combinatorial

optimization. Specifically, simulated annealing is an extension of the Metropolis-Hastings [?] algorithm which can be used to estimate the ground energy state of a many-body systems at thermal equilibrium. Kirkpatrick et. al. applied the Metropolis-Hastings algorithm sequentially, with decreasing temperature values in order to approximate a solid slowly cooling to low temperatures. Later work by Goffe [?], Corana et al. [?], and Lecchini-Visintini et. al. [?] extended simulated annealing to the continuous domain. When considering only the influence of classical thermal fluctuations in particle energy levels, the probability of a particle traversing a barrier of height ΔV at a temperature T is on the order of:

$$\mathcal{P}_t = e^{-\frac{\Delta V}{T}}. \tag{1}$$

This probability forms the basis for the Metropolis acceptance criterion

2.1 Paper Organization

Section contains a discussion of the quantum-inspired simulated annealing algorithm that will be used to select

3 PROBLEM FORMULATION AND OVERVIEW

In this paper we present a methodology for feed-forward neural network weight selection which employs a quantum-inspired variant of simulated annealing. In the following sections CSA and QSA are introduced and compared. CSA and QSA will then be applied to the pro

As discussed in Section 2, simulated annealing often converges to a local, rather than global minimum. One way to solve this problem is to create a mechanism by which the system can tunnel out of local minima to nearby, lower-cost configurations. This is analogous to the quantum mechanical phenomenon of tunneling. A particle with energy E incident upon a potential energy barrier of height $\Delta V > E$ has a non-zero probability of being found in, or past, the barrier. Classically, this behavior is forbidden. The probability of tunneling, $\mathcal{P}_t$, through a step barrier of height ΔV is described by:

$$\mathcal{P}_t = e^{-\frac{w\sqrt{\Delta V}}{\Gamma}} \tag{2}$$

- *J.F. and M.M. are with the Department of Electrical and Computer Engineering, Air Force Institute of Technology.*

where Γ is the tunneling field strength [**?**]. Figure **??** depicts a one-dimensional example of quantum tunneling.

It is instructive to contrast Eq. 1 and Eq. 2. Both describe the same value, but the importance of the width and height of the traversed barrier in the two equations is considerably different. For systems in which quantum tunneling is possible, the probability of penetrating a barrier of height ΔV is increased by a factor of approximately $e^{\Delta V}$, for large values of ΔV. This relationship is depicted graphically in Fig. **??** which shows the probability of barrier traversal for a system which allows quantum fluctuations, divided by the same probability for a system which only considers thermal fluctuations. Therefore, physical models which considers quantum effects are much more likely predict penetration of tall, thin energy barriers than those which only include classical thermal effects.

3.1 Classical and Quantum Simulated Annealing

The formulation of QSA described in this paper arises from the construction of a neighborhood function which approximates the transition probabilities of a quantum system traversing a potential energy surface (PES). [1]

3.2 Annealing Synaptic Weights

In order to apply the techniques of QSA and CSA, we most first formulate the problem of feed-forward neural network synaptic weight selection as a combinatorial optimization problem. Each synaptic weight in a feed-forward neural network may be encoded as a real-valued element in a 3-dimensional relation matrix, denoted as W_{ijk}. In this encoding scheme, for a given [layer] of the matrix, the row and column indexes indicate the pre-synaptic and post-synaptic neurons, respectively. The absence of a synaptic connection is indicated by a 0. A nonexistent synapse can be caused by the absence of either the presynaptic or post-synaptic neuron, or by the absence of a connection between the neurons. This weight encoding scheme is depicted graphically in Fig. **??**. The weight matrix, $\mathcal{W}$, therefore encodes a configuration in the solution space of the problem, which corresponds to some cost value $\mathcal{C}(\mathcal{W})$. We may now attempt to minimize $\mathcal{C}(\mathcal{W})$ by perturbing $\mathcal{W}$. With this framework in place, we can apply the techniques of CSA and QSA to the problem of weight selection.

3.3 Neighborhood Functions

All variations of simulated annealing require the specification of a neighborhood function, which determines the way in which new states may be generated from the current state. Let us define a generic neighborhood function, $\mathcal{N}$, which operates on $\mathcal{W}$, thereby changing its value. In the following section we detail several possible realizations of $\mathcal{N}$, some of which constitute an approximation to system performing quantum tunneling.

3.3.1 Classical Neighborhood Functions

We define $\mathcal{N}_c$ to be the classical neighborhood function, which is defined as

$$\mathcal{N}_c(\mathcal{W}) = \mathcal{W} + \alpha \mathcal{R}_n \qquad (3)$$

where $\mathcal{R}_n$ is a random matrix with dimensionality equal to that of $\mathcal{W}$ which has been normalized such that the elements of the matrix sum to 1, and α is the learning rate. This neighborhood function will produce a weight matrix, $\mathcal{W}'$, which will have an L^2 distance of exactly α from $\mathcal{W}$ in the weight space.

3.3.2 Quantum Neighborhood Function

... Consider the hypothetical cost surface depicted in Fig. **??**.

3.3.3 Anisotropic Quantum Neighborhood Function

In the course of developing this framework, it was noted that it is sometimes advantageous to move along an error surface in exactly one dimension, as show in Fig. **??** This corresponds to modifying a single synaptic weight a single epoch. Though it is possible that this could occur by chance, the likelihood of a single-weight modification is inversely proportional to the number of weights in the network. In physical science a force acting preferentially in a subset of the total dimensionality of the degrees of freedom in which it could act is referred to as an anisotropy. This concept has been extended to the weight selection scheme presented in this paper in the form of a value, a, specifying the anisotropicity of a weight matrix perturbation. This value serves as a nonlinear transform on the perturbation matrix, thus concentrating the perturbation in a subset of the available degrees of freedom. Given as

By inspection, when a is 0, the perturbation will be unchanged and will be distributed evenly across all degrees of freedom, on average. Conversely, a is 1, all the perturbation will occur in a single dimension, thus changing only one synaptic weight.

3.4 Cost Functions

4 IMPLEMENTATION DETAILS

5 EXPERIMENTS AND RESULTS

---

1. This concept is nearly mathematically identical to the fast simulated annealing model described in []. However, the QSA algorithm presented in this paper was arrived at independently, though an attempt to incorporate quantum mechanical phenomena into neural network weight selection.

PLACE PHOTO HERE

Justin R. Fletcher received the B.S degree in Computer Engineering from Embry Riddle Aeronautical University in 2012 and is currently pursuing the M.S. degree in Computer Science from the Air Force Institute of Technology (AFIT). His research interests include neural networks, quantum mechanics, and stuff. [do this later]

PLACE PHOTO HERE

Michael J. Mendenhall received the B.S. degree in Computer Engineering from Oregon State University, Corvallis, OR in 1996, the M.S. degree in Computer Engineering from the Air Force Institute of Technology (AFIT), Wright-Patterson AFB, OH, in 2001, and the Ph.D. degree in Electrical Engineering from Rice University, Houston, TX, in 2006. Currently, he is an Assistant Professor of Electrical Engineering at AFIT. He recieved the Dr. Leslie M. Norton teaching award by the AFIT student association and an honorable mention for the John L. McLucas basic research award at the Air Force level, both in 2010. His research interests are in hyperspectral signal/image processing, hyperspectral signature modeling, and computational intelligence.