LA-UR-18-31544
Approved for public release; distribution is unlimited.

Title: Image classification using quantum inference on the D-Wave 2X

Author(s): Nguyen, Nga Thi Thuy
Kenyon, Garrett

Intended for: IEEE International Conference on Rebooting Computing, 2018-11-07/2018-11-09 (Washington, District Of Columbia, United States)

Issued: 2018-12-11

Disclaimer:
Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Image classification using quantum inference on the D-Wave 2X

Nga T. T. Nguyen
Information Sciences
Los Alamos National Laboratory
Los Alamos, New Mexico 87545, USA

Garrett T. Kenyon
Information Sciences
Los Alamos National Laboratory &
New Mexico Consortium
Los Alamos, New Mexico 87545, USA
gkenyon@lanl.gov

***Abstract*—We use a quantum annealing D-Wave 2X computer to obtain solutions to NP-hard sparse coding problems. To reduce the dimensionality of the sparse coding problem to fit on the quantum D-Wave 2X hardware, we passed downsampled MNIST images through a bottleneck autoencoder. To establish a benchmark for classification performance on this reduced dimensional data set, we built two deep convolutional neural networks (DCNNs). The first DCNN used an AlexNet-like architecture and the second a state-of-the-art residual network (RESNET) model, both implemented in TensorFlow. The two DCNNs yielded classification scores of $94.54 \pm 0.7\%$ and $98.8 \pm 0.1\%$, respectively. As a control, we showed that both DCNN architectures produced near-state-of-the-art classification performance ($\sim 99\%$) on the original MNIST images. To obtain a set of optimized features for inferring sparse representations of the reduced dimensional MNIST dataset, we imprinted on a random set of 47 image patches followed by an off-line unsupervised learning algorithm using stochastic gradient descent to optimize for sparse coding. Our single-layer of sparse coding matched the stride and patch size of the first convolutional layer of the AlexNet-like DCNN and contained 47 fully-connected features, 47 being the maximum number of dictionary elements that could be embedded onto the D-Wave 2X hardware. When the sparse representations inferred by the D-Wave 2X were passed to a linear support vector machine, we obtained a classification score of 95.68%. We found that the classification performance supported by quantum inference was maximal at an optimal level of sparsity corresponding to a critical value of the sparsity/reconstruction error trade-off parameter that previous work has associated with a second order phase transition, an observation supported by a free energy analysis of D-Wave energy states. We mimicked a transfer learning protocol by feeding the D-Wave representations into a multilayer perceptron (MLP), yielding 98.48% classification performance. The classification performance supported by a single-layer of quantum inference was superior to that supported by a classical matching pursuit algorithm set to the same level of sparsity. Whereas the classification performance of both DCNNs declined as the number of training examples was reduced, the classification performance supported by quantum inference was insensitive to the number of training examples. We thus conclude that quantum inference supports classification of reduced dimensional MNIST images exceeding that of a size-matched AlexNet-like DCNN and nearly equivalent to a state-of-the-art RESNET DCNN.**

***Index Terms*—Sparse coding, Neuromorphic computing, MNIST, Quantum annealing D-Wave 2X, Deep Convolutional Neural Networks, Autoencoder**

I. Introduction

Deep learning has yielded impressive advances across a variety of machine learning tasks such as Alpha Go Zero [1] and the ImageNet challenge [2]. However, deep neural networks can be spoofed by adversarial examples [3], possibly due to an underlying reliance on low-level image statistics [4]. Moreover, the ability of deep neural networks to learn by analogy [5] as well as the importance of network depth in constructing abstract semantically-meaningful representations [6] has been called into question. Unsupervised learning, on the other hand, particularly Boltzmann Machines [7] and sparse coding paradigms [8], seek to learn joint distributions or causes directly from unlabeled data and thus may be less prone to some of the issues that have plagued task-specific deep learning approaches. For example, features optimized in an entirely unsupervised manner for sparse coding can nonetheless support performance on image classification tasks that is only slightly below that achieved by standard deep neural network architectures such as AlexNet [9]. Sparse inference in these examples, however, relies on a convex approximation to the desired solution. Optimal sparse inference, especially when derived from a highly overcomplete dictionary, is NP-hard, making it difficult to fully assess the potential of such approaches with existing algorithms. Quantum computers offer a possible strategy for rapidly obtaining good solutions to NP-hard sparse inference problems. Previous research [10] demonstrated that models of sparse inference based on lateral inhibition between binary neurons can be directly implemented on the D-Wave 2X Quantum Annealing Computer. Here, we show that on a suitably low-dimensional problem that fits on available quantum annealing architectures, a single layer of features optimized in an unsupervised manner for sparse inference followed by a multi-layer perceptron (MLP) nearly matches the classification performance of a state-of-the-art RESNET deep convolutional neural network (DCNN) trained in a fully supervised manner for the specific task in question.

Our paper is organized as follows. Sec. II, Sec. III, and Sec. IV introduce concepts, methodology, and reduced dimensional images for mapping a binary sparse coding classification problem onto the D-Wave 2X machine. Classification based on D-Wave generated sparse representations are compared

with with the state-of-the-art DCNN classifiers in Sec. V and Sec. VI. Evidence for a phase-transition based on a free energy analysis of solutions from the D-Wave quantum annealer is presented in Sec. VII.

II. Sparse-coding

The hypothesis that neurons encode stimuli by inferring sparse representations explains many of the response properties of simple cells in the mammalian primary visual cortex [8], [11]. Given an overcomplete, non-orthonormal basis $\{\phi_i\}$, inferring a sparse representation involves finding the minimal set of non-zero activation coefficients $\boldsymbol{a}$ that accurately reconstruct a given input signal $\boldsymbol{X}$, corresponding to a minimum of the following energy function:

$$E(\boldsymbol{X}, \boldsymbol{\phi}, \boldsymbol{a}) = \min_{\{\boldsymbol{a}\}}[\,\frac{1}{2}||\boldsymbol{X} - \boldsymbol{\phi}\boldsymbol{a}||^2 + \lambda||\boldsymbol{a}||_0\,] \qquad (1)$$

where λ is a trade-off parameter that determines the balance between reconstruction error of the original input image $\boldsymbol{X}$ and the number of non-zero (sparse) activation coefficients. A larger λ encourages sparser solutions. We define $\gamma = \frac{\text{rank}(\boldsymbol{a})}{\text{rank}(\boldsymbol{X})}$ as the overcompleteness of the basis $\{\phi_i\}$. γ typically is chosen such that $\gamma \gg 1$ which means that in general there will exist many solutions which achieve a similarly small reconstruction error $||\boldsymbol{X} - \boldsymbol{\phi}\boldsymbol{a}||^2$ and our task is to find the sparsest one. The energy function Eq. (1) is non-convex and contains multiple local minima, so that finding a sparse representation falls into an NP-hard complexity class of decision problems [12].

III. Quantum inference

A. Mapping Sparse coding problem on D-Wave 2X

The D-Wave 2X [13] consists of 1152 quantum bits (qubits) arranged into 12x12 unit cells, forming a Chimera structure with dimensions 12x12x8. Sparse interactions between qubits are restricted to the 16 connections within a unit cell and the 16 connections between nearest-neighboring unit cells [13].

In detail, each unit cell contains 4 qubits aligned along a horizontal axis and 4 qubits aligned vertically. Within a unit cell, the 4 qubits of a given orientation can only can interact with the 4 qubits with the opposite orientation (details see [10]). Between unit cells, interactions are only allowed between nearest-neighbors and even between nearest-neighbors the allowed connections are restricted according to relative orientation. A vertically (horizontally) oriented qubit can only connect to the two vertically (horizontally) oriented qubits at the same relative position in the nearest-neighboring unit cells immediately above (left) and below (right) along one column of the 12x12 grid. Thereby, in a chimera graph, one qubit can interact with *at most* 6 other qubits.

The D-Wave 2X [13] finds optimal solutions to a (discrete) Ising system consisting of N_q binary variables via quantum annealing. Such N_q-body systems can be described by the following classical Hamiltonian:

$$H(\boldsymbol{h}, \boldsymbol{Q}, \boldsymbol{a}) = \sum_{i}^{N_q} h_i a_i + \sum_{i<j}^{N_q} Q_{ij} a_i a_j \qquad (2)$$

with binary activation coefficients $a_i = \{0, 1\}\,\forall i \in (1, 2, 3, ..., N_q)$. This objective function defines a quadratic unconstrained binary optimization (QUBO) problem. We cast our sparse coding problem, Eq. (1), into QUBO form, Eq. (2), by the transformations [14]:

$$\begin{aligned} h_i &= (-\boldsymbol{\phi}^T\boldsymbol{X} + (\lambda + \frac{1}{2}))_i, \\ Q_{ij} &= (\boldsymbol{\phi}^T\boldsymbol{\phi})_{ij}. \end{aligned} \qquad (3)$$

In Eq. (3), the bias term $\boldsymbol{h}$ (elements h_i) in the Ising model is proportional to the weighted input $\boldsymbol{\phi}^T\boldsymbol{X}$ while the coupling term $\boldsymbol{Q}$ (elements Q_{ij}) corresponds to lateral competition (see also [15]) between qubits given by the interaction matrix $\boldsymbol{\phi}^T\boldsymbol{\phi}$ [with self-interaction excluded and $\boldsymbol{Q}$ being symmetric i.e. $Q_{ij} = Q_{ji}\ \forall i \neq j$, see Eq. (2)]. The trade-off parameter λ corresponds to a uniform applied field that biases each qubit to be in the zero state.

From Eq. (3), we infer a sparse representation to Eq. (1) on D-Wave 2X hardware by associating each neuron with a single feature ϕ_i, represented as a binary logical qubit, with logical qubits embedded on the D-Wave physical chimera as follows.

B. Embedding

Despite the sparsity of physical connections on the D-Wave, it is nonetheless possible to construct graphs with arbitrarily dense connectivity by employing "embedding" techniques. Embedding works by chaining together physical qubits so as to extend the effective connectivity but at the cost of reducing the total number of available logical qubits. Because logical qubits do not need to follow the connection rules that physical qubits do, it is possible to implement general QUBO problems with arbitrarily dense connectivity on the sparsely connected D-Wave chimera. The D-Wave API provides a heuristic algorithm that searches for an optimal embedding that minimizes the number of physical qubits that are chained together.

The exact mapping of a spin glass problem onto the physical D-Wave 2X chimera, including defects, can typically contain approximately $N_q \sim 1000$ spins (qubits) with > 3000 local spin-spin interactions. In contrast, embedding an arbitrary QUBO problem onto the same 2X chimera typically allows no more than $N_q \sim 47$ nodes (logical qubits) but these nodes may be fully connected. Thus, embedding effectively trades qubits for connectivity.

Because a logical qubit is assigned only two possibilities 0 and 1, each neuron is treated as a "quantum object" bearing two possibilities: firing with maximum activation with coefficient 1 and silent with coefficient 0. Because each neuron is a quantum object, the state of any neuron is described in general by a superposition of 1 and 0, in which the neuron is both active and non-active at the same time, a logical impossibility for any classical system. Due to this quantum superposition, the annealing processing in D-Wave allows to explore the entire energy landscape at once.

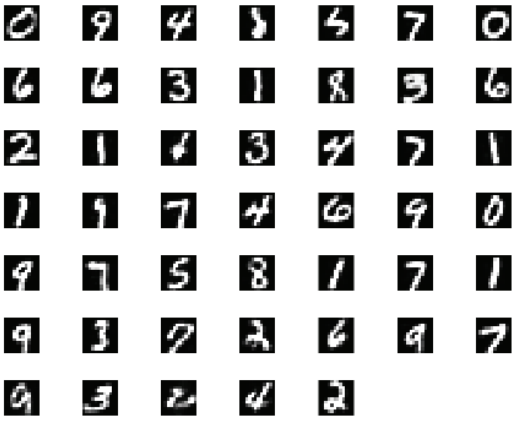


Fig. 1: 47 12x12 randomly selected reduced dimensional images reconstructed through a bottleneck autoencoder, yielding an undercomplete dictionary of 47 features for generating sparse representations of the reduced dimensional MNIST images. To construct an overcomplete dictionary, we randomly selected a set of 47 6x6 patches and divided each reduced dimensional MNIST image into overlapping 6x6 patches with a stride of 2 and separately generated a sparse representation of each 6x6 patch. The resulting 47 imprinted features, whether 12x12 or 6x6, were further refined using unsupervised SGD to optimize for sparse reconstruction.

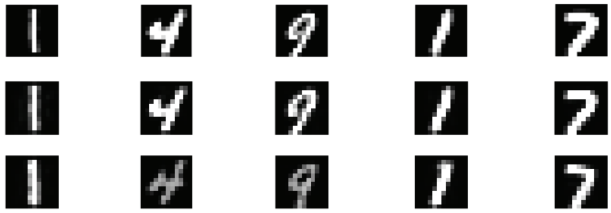


Fig. 2: Top: Original MNIST images downsampled to 12x12, Middle: Reconstructed images from bottleneck autoencoder, Bottom: D-Wave reconstruction using undercomplete dictionary based on randomly selected imprinted features.

IV. D-Wave inputs: reduced dimensional images

We first downsampled original 28x28 MNIST images to 12x12 and trained a bottleneck autoencoder [16] (using TensorFlow) that consisted of 47 neurons at its narrowest point. The resulting autoencoder generated images, examples of which are shown in Fig. 1, constitute a database of images with greatly reduced dimensionality suitable for analysis by the D-Wave 2X. Constructing a dictionary based on a randomly chosen set of 47 12x12 reduced dimensional MNIST images results in undercomplete basis $\gamma = 47/144$ (≈ 0.33) $\ll 1$. We used this undercomplete basis to demonstrate that it was possible to obtain sparse representations of reduced dimensional MNIST images on the D-Wave 2X (Fig. 2). Using an undercomplete basis however reduces the advantages of sparse inference as the resulting representations are unlikely to be sparse.

V. Classification performance: Undercomplete dictionary

A. Image classification

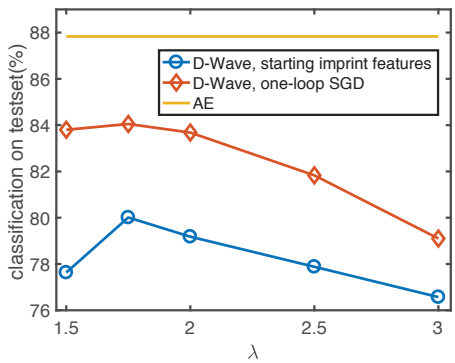
Due to computational limitations, it was not possible to perform all of the analyses presented in this paper using an overcomplete dictionary for generating sparse representations. We therefore first present benchmark performance using an undercomplete dictionary. Benchmark results obtained using an undercomplete dictionary helps to illustrate the advantages of sparse inference on the D-Wave quantum annealing machine when using an overcomplete basis.

Recall that reduced dimensional images were obtained by first downsampling to 12x12 and then reconstructing through a bottleneck autoencoder employing a hidden layer containing 47 feature vectors, thereby "compressing" the 144 pixels in the downsampled images to a 47 dimensional manifold. Thus, using a dictionary containing 47 features might be considered complete, although nominally $\gamma = 47/144$.

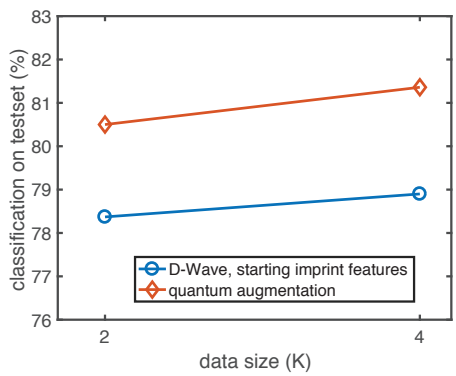
Fig. 3(a) shows classification results on reduced dimensional 12x12 MNIST images using an undercomplete dictionary for sparse coding containing 47 randomly selected imprinted features. The sparse representations generated by the D-Wave 2X for the complete set of 55 K reduced dimensional MNIST images was passed to a linear SVM. As a function of the trade-off parameter λ, classification performance using an undercomplete basis peaked at approximately 80%. Here, sparsity was approximately $\{16\%, 13\%, 12\%, 9\%, 5\%\}$ for λ of $\{1.5, 1.7, 2.0, 2.5, 3.0\}$, respectively. By comparison, simply passing the latent representations from the hidden layer of the autoencoder to an SVM yielded a classification score nearly 88% [see Fig. 3(a), note that a bottleneck autoencoder is independent of λ]. Thus, when using an undercomplete basis, we find no advantage to generating sparse representations on the D-Wave 2X.

Randomly selected imprinted features are unlikely to be optimal for sparse coding, which likely reduced the classification performance using the resulting sparse representations. We therefore conducted unsupervised training to optimize our undercomplete basis for sparse coding. Our unsupervised training procedure used a local Hebbian rule with momentum for computing weight updates to optimize our initially imprinted features ϕ.

The SGD loop required generating sparse representations on the D-Wave 2X for a randomly chosen mini-batch consisting of 10K reduced dimensional MNIST images, from which we computed an average weight update for the undercomplete dictionary ϕ. This cycle was repeated for approximately a few thousand steps until the loss function saturated. After unsupervised optimization, we obtained an approximately 4%



(a)



(b)

Fig. 3: (Color online) (a): Image classification obtained by feeding the sparse representations of the downsampled autoencoded MNIST images generated by the D-Wave 2X before (blue curve) and after unsupervised feature optimization (red curve) to a linear SVM. Dataset contains 55 K images. The classification result obtained using the autoencoder latent representations as input to a linear SVM was 87.83% (yellow line). (b): Image classification on up to $4,000$ sparse representations of downsampled autoencoded MNIST images generated by the D-Wave 2X using the lowest-energy state (blue circles) and quantum augmentation using the 30 lowest energy level states (red diamonds). Training and test sets were divided in a ratio of 5/1, respectively. Dataset size is in units of $1,000$ images.

increase in the peak classification accuracy [red curve in Fig. 3(a)].

B. Quantum augmentation

Quantum annealing on the D-Wave is able to return multiple sparse solutions corresponding to a set of good (local) minima given an input $\{\boldsymbol{h}, \boldsymbol{Q}\}$. We examined whether these multiple sparse solutions could support improved image classification by treating these distinct representations analogously to standard image augmentation procedures. Due to the cost of computation for this task, we present the classification scores on only 2 K and 4 K "quantum augmentation" images (with the ratio between training/test sets 5/1) of the reduced dataset for the "best" $\lambda = 1.7$ obtained from Fig. 3(a). We show in Fig. 3(b) that a quantum augmentation consisting of 30 sparse solutions extracted from the 2X hardware with an undercomplete dictionary gains about $\sim 2\%$ in classification in this case.

VI. Classification performance: Overcomplete dictionary

The above classification results obtained via quantum inference used a single 1x1x47 feature map fed directly into a linear SVM. In this section, we generate sparse representations on the D-Wave 2X using an overcomplete dictionary by breaking the 12x12 reduced-dimensional inputs into overlapping 6x6 tiles.

A. Patch, stride

To construct an overcomplete dictionary (i.e. $\gamma > 1$), we represent each image as a set of overlapping patches. Sliding a 6x6 patch throughout the entire image of total size 12x12 with a stride $s = 2$, we obtain a set of 4x4 (= 16) patches for each image. The dictionary, which now consists of 47 randomly selected 6x6 patches, has overcompleteness of $\gamma = 47/36 = 1.31$. Consequently, the set of feature maps now increases from 1x1x47 to 4x4x47.

B. Evaluating classification performance

To determine an optimal value for the trade-off parameter λ, we study the four different cases of $\lambda = \{0.5, 0.7, 1.0, 1.5\}$. This corresponds, respectively, to sparsity of $\{14\%, 12\%, 10\%,$ and $8\%\}$ on average. For classification, we employed the "feature extraction" technique (see for example Ref. [9]) to obtain a sparse representation for each image by concatenating the binary sparse coefficients for each 6x6 image patch into a 4x4x47 array. We used the LIBLINEAR package [17] to train a linear SVM classifier on this 10-category image classification task using the combined sparse coefficients. For the first 20K images, because each 12x12 image has 16 6x6 overlapping tiles, there were 16x20K (= 320K) input Hamiltonians to solve using the D-Wave 2X. As a function of λ, we obtained the classification scores shown in Fig. 4. By replacing the undercomplete dictionary applied to the entire image with an overcomplete dictionary applied to each image patch, we improved our classification results based on quantum inference substantially. We obtain a peak behavior for $\lambda_c = 0.7$ which translates into $\sim 12\%$ of sparsity, corresponding to an average of $5 - 6$ active neurons to reconstruct each 6x6 tile.

Image classification: D-Wave 2X vs. DCNN

As an initial benchmark, we started with an AlexNet-like DCNN architecture that approximately matched the size of the latent representations generated by the D-Wave 2X. The AlexNet-like DCNN possessed an initial convolutional layer with a stride of 2 and 12 features, yielding a set 12 4x4 feature maps, each being the same size as the feature maps generated by the D-Wave 2X. A subsequent set of 16 5x5 feature maps was followed by a final all-to-all layer using 47 features, the

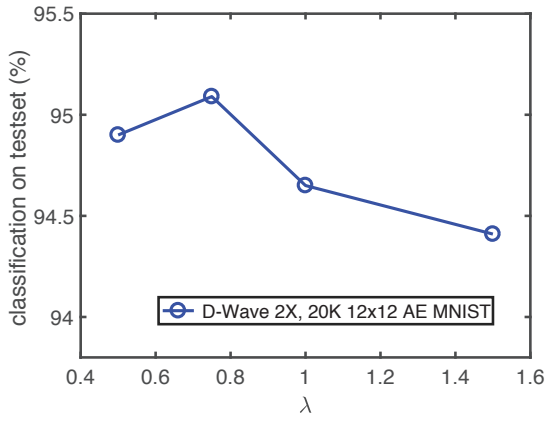


Fig. 4: (Color online) Image classification using an overcomplete dictionary applied separately to overlapping 6x6 image patches on the D-Wave 2X for the first 20 K images of our 12x12 autoencoded MNIST dataset plotted as a function of λ. Training and test sets were divided in a ratio of 5/1, respectively. This procedure established a best value of λ.

same number of feature maps on the D-Wave 2X, which in turn fed into a SoftMAX classifier (cross-entropy loss).

Fig. 5 shows that quantum inference enabled superior classification performance compared to an AlexNet-like DCNN. When applied to the full reduced dimensional MNIST dataset, the classification accuracy of the AlexNet-like DCNN was $94.54 \pm 0.7\%$ (green connected circles with error bars). The same AlexNet-like DCNN produced near-state-of-the-art results of $\sim 98.9\%$ on the original MNIST 28x28 pixel dataset (not plotted), affirming the validity of our DCNN implementation. Likewise, if we remove the autoencoder procedure and use the downsampled MNIST images to train the AlexNet-like DCNN model, we obtained an accuracy of 96.45% (not plotted), consistent with the expectation that the autoencoder destroys some information. Sparse representations generated by the D-Wave 2X, when fed to a linear SVM classifier [17], produced a slightly higher classification score than the AlexNet-like DCNN, 95.68% (purple circles) and exhibited almost no variation between reinitialized runs. Additionally, classification by the AlexNet-like DCNN declined as the size the training set was reduced whereas classification based on quantum inference showed no such degradation. Matching pursuit [18], a classical technique for solving sparse coding problems, produced a lower classification score than did quantum inference when the former was configured to match the same sparsity as the solutions generated by the D-Wave ($\sim 12 - 14\%$).

We further examined a transfer learning procedure where the D-Wave sparse representations were fed into a multilayer perceptron (MLP), yielding a classification of $\sim 98.48\%$ (red circles). Feeding the sparse representations produced by a classical matching pursuit algorithm into an MLP yielded a lower classification score of $\sim 96.34\%$ (yellow diamonds),

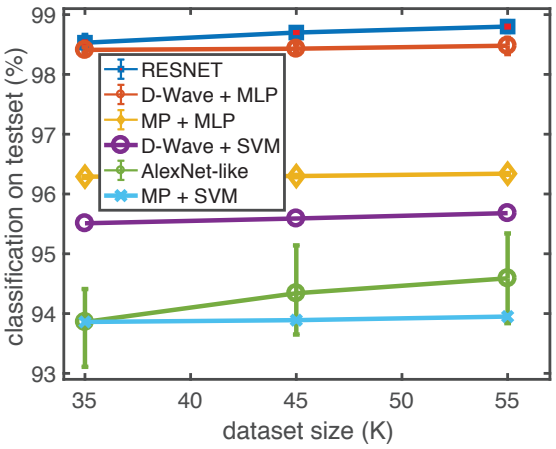


Fig. 5: (Color online) Image classification on the D-Wave 2X, near state-of-the-art DCNN (AlexNet-like) built with TensorFlow, and matching pursuit for our reduced dimensional 12x12 autoencoded MNIST dataset plotted as a function of dataset size. Training and test sets in each case are in units of 1,000 images and divided in a ratio of 5/1, respectively. We also show classification results using state-of-the-art RESNET for our customized MNIST images.

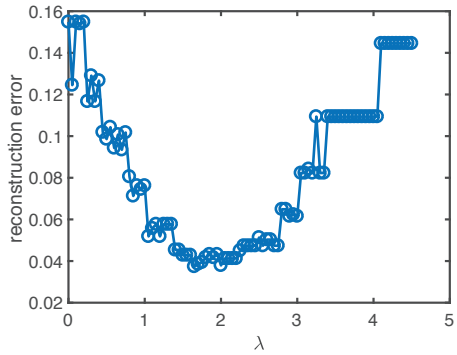
suggesting that quantum inference may yield superior solutions compared to classical approaches.

The AlexNet-like DCNN was tailored to approximately match the size of the sparse representations generated by the D-Wave 2X. We also tested a state-of-the-art (RESNET) DCNN and obtained a classification of $\sim 98.8\%$ (blue squares) on the reduced MNIST database. While RESNET yielded a slightly higher classification score using all of the available training data, this advantage disappeared as the amount of training data was reduced. Using a reduced dimensional dataset that fits onto the D-Wave 2X hardware thus reveals the benefits of quantum inference for image classification.

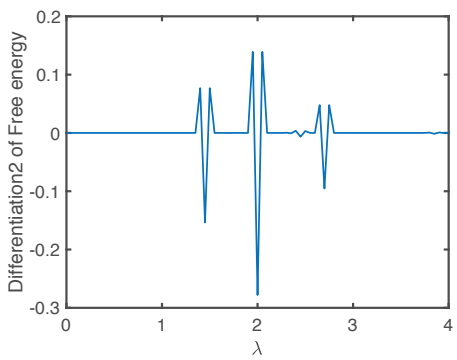
VII. Phase-transition in a Sparse-coding model

Previous work [19] has revealed the existence of a 2nd order phase transition in the sparse representations generated by Locally Competitive Algorithm (LCA) implemented using continuously-valued neurons simulated on a conventional digital computer. Here, we search of evidence of a phase transition in the binary sparse coding problem solved on the D-Wave 2X by assuming a thermal Boltzmann distribution for the spectrum of energy states generated by different annealing runs.

We consider the sparse representation of one randomly-chosen 12x12 reduced dimensional MNIST image. To explore statistical properties of the sparse representations inferred on a D-Wave hardware, we treat the trade-off parameter λ as an "effective" (magnetic) field, where λ "couples" to an Ising model through the qubit-field interaction in $\boldsymbol{h}$. Next, the



(a)



(b)

Fig. 6: (Color online) Top: Reconstruction error in a denoising sparse coding system for one autoencoded-12x12 image plotted as a function of sparsity threshold. Bottom: Second derivative of the free energy with respect to λ of the system described in the top plot at temperature $k_B T = 10^{-2}$ (energy unit).

energy spectrum $\{E_i\}$ is obtained using different annealing runs, corresponding to parameter deviations in $\{\boldsymbol{h}, \boldsymbol{Q}\}$. We executed multiple runs and obtained over ten thousand sparse solutions along with the corresponding energy spectrum for a particular reduced dimensional MNIST image. We used that energy spectrum to examine the 2nd derivative of free energy, defined as $F = -\frac{1}{\beta} \log Z$, with respect to the "field" λ in Fig. 6. Here, $Z = \sum_{i=1}^{N_s} e^{-\beta E_i}$ is defined as partition function ([20]), $\beta = 1/k_B T$ with k_B is Boltzmann constant and T the temperature of the embedded sparse coding model. N_s is the total number of N_q-body states obtained in this case by solving Eqs. (1) and (3) to yield the corresponding energy set $\{E_i\}$. The reconstruction error calculated using the lowest-energy level as a function of sparsity λ, presented in Fig. 6(a), shows that there exists a critical $\lambda_c \approx 2$ that yields the best denoising result (i.e. yields the most accurate reconstruction of the input). This optimal behavior coincides with sharp discontinuity around $\lambda \approx 2$ in the second derivative of F with respect to λ in Fig. 6(b). The critical point is consistent with the optimal values of λ found above.

VIII. Conclusions

We have explored classification performance on dimensionally-reduced binary MNIST images using sparse representations generated by the D-Wave 2X quantum annealing computer at a putative critical value of sparsity. Given the limited number of qubits available on the D-Wave 2X, we first used a bottleneck autoencoder to reduce the intrinsic dimensionality of MNIST images, which originally consisted of 28x28=784 binary values. A bottleneck autoencoder is trained in a standard manner using backprop and SGD to reconstruct, as accurately as possible, input images that have been forced through a narrow waist, corresponding to a latent representation whose dimensionality (number of neurons in the layer comprising the waist of the hourglass) is less than the dimensionality of the original image.

To investigate how quantum inference might contribute to classification performance, we added lateral inhibition between 47 features obtained by randomly choosing (imprinting) from a subset of reduced 12x12 autoencoder-generated MNIST images, allowing us to infer a sparse binary representation for each reduced-dimensional image via quantum inference. Although the reduced database itself was constructed to live on a 47-dimensional manifold, local regions of the manifold are likely to be characterized by an even smaller number of basis elements. In principle, this reduced basis can be inferred via quantum inference. As a target benchmark, we trained standard deep neural network classifiers, implemented in TensorFlow, on our reduced dimensional database. A size matched deep learning architecture yielded a classification score of $94.54 \pm 0.7\%$ compared to 98.48% obtained by feeding sparse representations inferred on the D-Wave 2X into an MLP. Classification performance of a state-of-the-art model (RESNET) was only slightly higher. The RESNET model however contains orders of magnitude more neurons and free parameters compared to quantum inference. As a result, the RESNET model was more susceptible to degradation as the number of training examples was reduced. We showed that quantum inference on the D-Wave 2X, using approximately 47 binary neurons (logical qubits) and only a single layer of convolutional kernels optimized for sparse reconstruction using an unsupervised training procedure, could match the classification performance obtained by a state-of-the-art deep neural network as the size of the training set was reduced.

Quantum inference on the D-Wave hardware allowed us to examine a postulated thermodynamical phase-transition in a sparse coding system by analyzing a large number of energy states obeying a Boltzmann distribution. We found a critical sparsity λ_c as the minimal point in the reconstruction error and this optimal λ_c that appears as sharp discontinuity in the 2nd differentiation of the thermodynamically averaged free energy produces the smallest image classification error.

Acknowledgment

We acknowledge Marcus Daniels and Jack Raymond for helpful discussions.

REFERENCES

[1] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, "Mastering the game of Go without human knowledge," Nature, vol. 550, pp. 354–359, April 2017.

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A, Karpathy, A, Khosla, M. Bernstein, A. C. Berg, and Li Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," arXiv:1409.0575 [cs.CV], September 2014.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in Neural Information Processing Systems 27, NIPS, December 2014.

[4] J. Jo and Y. Bengio, "Measuring the tendency of CNNs to Learn Surface Statistical Regularities," arXiv:1711.11561 [cs.LG], November 2017.

[5] M. Ricci, J. Kim, and T. Serre, "Not-So-CLEVR: Visual Relations Strain Feedforward Neural Networks," arXiv:1802.03390 [cs.CV], February 2018.

[6] R. Shwartz-Ziv and N. Tishby "Opening the Black Box of Deep Neural Networks via Information," arXiv:1703.00810 [cs.LG], April 2017.

[7] G. E. Hinton, "A Practical Guide to Training Restricted Boltzmann Machines," in Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol. 7700. Springer, Berlin, Heidelbergpp. pp. 599-619, 2012.

[8] B. A. Olshausen and D. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," Nature, vol. 381, pp 607–609, June 1996.

[9] A. Coates and A. Y. Ng., "The Importance of Encoding versus Training with Sparse Coding and Vector Quantization," in Proceedings of The 28th International Conference on Machine Learning, ICML, pp. 921-928, June 2011.

[10] N. T. T. Nguyen, A. E. Larson, and G. T. Kenyon, "Generating sparse representations using quantum annealing: Comparison to classical algorithms," in Proceedings of the 2017 IEEE International Conference on Rebooting Computing (ICRC), pp. 146–151, November 2017.

[11] B. A. Olshausen and D. Field, "Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?" Vision Res., vol. 37, pp. 3311–3325, December 1997.

[12] B. Natarajan, "Sparse approximate solutions to linear systems," SIAM Journal on Computing, vol. 24 , pp. 227–234, April 1995.

[13] D-Wave system: http://www.dwavesys.com/.

[14] N. T. T. Nguyen and G. T. Kenyon, "Solving sparse representation for object classification using quantum D-Wave 2X machine," in the Proceedings of The First International Workshop on Post Moore's Era Supercomputing, pp. 43–44, November 2016.

[15] C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. A. Olshausen, "Sparse Coding via Thresholding and Local Competition in Neural Circuits," Neural Computation, vol. 20, pp. 2526–2563, October 2008.

[16] G. E. Hinton, R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," Science, vol. 313 (5786), pp. 504–507, July 2006.

[17] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, Software available at http://www.csie.ntu.edu.tw/~cjlin/liblinear, "LIBLINEAR: A Library for Large Linear Classification," Journal of Machine Learning Research, vol. 9, pp. 1871–1874, August 2008.

[18] S. G. Mallat and Z. Zhang, "Matching Pursuits with Time-Frequency Dictionaries," IEEE Transactions on Signal Processing, vol. 41, pp. 3397-3415, December 1993.

[19] J. Carroll, N. Carlson, and G. T. Kenyon, "Phase Transitions in Image Denoising via Sparsely Coding Convolutional Neural Networks ," arXiv 1710.09875, October 2017.

[20] R. K. Pathria and P. D. Beale, Statistical Mechanics, 3rd Ed, Elsevier, February 2011.