My work on this summer research internship mainly revolves around the topic of optimization methods for large-scale machine learning, under the supervision of Professor Wai Hoi-To (SEEM).

1 Research Directions

Previous work on Neural Networks' quantization in its training phase (e.g. `BinaryConnect` that uses weights $\{-1, +1\}$ during forward and backward propagations) has empirically shown the effect as a regularizer while eliminating the need of precise multiplications during propagations. First, discretizations can be viewed as noise injections to the activations or weights, preserving the expected value of the discretized weight if controlled properly. Second, quantized weights are more friendly to the current hardware structures, which in turn speeds up the process of neural network training.

In this research period, we turn to investigate the robustness of a category of neural networks with quantized weights (QNNs), formulating them as an optimization problem, thus bounding the error caused by precision clipping. Our major work emphasizes on:

- Relaxation of assumptions for generalizations
- Neural network performance experiments

2 Project Timetable

Date	Work
Mar 30 - Jun 2	Topic Discussion and Research Direction Review (Meeting I)
Jun 3 - Jun 28	Reading: Quantized Neural Networks[1] (Meeting II)
Jun 21 - Ongoing	Part 1 - Proof of Convergence Guarantees Part 2 - Numerical Experiments (Meeting III, IV)

Note: 1. During Jun 3 - Jun 18, I was put under a service learning trip (serving the partial requirement of college GE course GEMC3001) to learn about non-governmental organizations' work in Greece, with the permission of my supervisor (by email). I spent some time familiarizing myself with [C1] and linear algebra (matrix 2-norm, Courant-Fischer Theorem, simple path counting).

3 Meeting Details

3.1 Meeting I (Mar 21, 06:30 PM) Prior to the meeting, Professor Wai suggested a feasible research direction, which could possibly ensure the best use of time, for this summer internship. In this meeting, Professor Wai provided a brief overview on the arise of optimization problems in machine learning models, and some new perspectives on the field of graph signal processing.

The general optimization problem for machine learning models can be formulated as below, which can be interpreted as finding a set of optimal weights $\mathbf{w} = (w_1, w_2, \ldots, w_d)$ for a particular prediction machine architecture (which formulates the function $h_{\mathbf{w}}$) to minimize the error (a measure defined by a function ℓ) between the predicted values $h_{\mathbf{w}}(\mathbf{x})$ and true values y:

$$\min_{\mathbf{w}\in\mathbb{R}^d} \quad \mathbb{E}_{(\mathbf{x},y)\sim p_{data}}[\ell(h_{\mathbf{w}}(\mathbf{x}), y)],$$

under the assumption that the samples used to train the model is unbiased to the underlying data distribution, i.e. $(\mathbf{x}, y) \sim p_{data}$.

The highlights are the prevalent optimization analyses involving the great use of convexity and differentiability. Three major optimization methods are introduced:

- Gradient Descendent
- Batch Gradient Descendent
- Stochastic Gradient Descendent

A deeper ponder which consists of more unknown factors will be the inference of graph networking structures driven by observed data, under the theory of graph signal processing. Different famous signal processing techniques can be adapted to develop graph learning strategies, including Fourier transform, low-pass filters and spectral decomposition, whose application scenarios span a vast range from social networks to neural sciences.

Works To Be Done

- Papers [A3], [C2], book [B1], etc. are assigned for reading.
- Select a solid topic that would be interesting for the self and for the work to delve into.
- Personal growth: What will be the future interest?
- Modify the research application proposal.

3.2 Meeting II (Jun 3, 10:00 AM) The second meeting discusses in-depth about the basic elements and theoretical guarantees of the series of gradient descendent methods, and provided a clear summary on the comparison of different optimization methods (See [A3]). Professor Wai also offered a research topic on the weight quantization of neural networks, and recommended to compare the existing optimization methods with current algorithm proposed (e.g. What is remained unclear? Are there any unnecessary assumptions?).

In this meeting, Professor Wai answered my concerns about the unfamiliarity on concepts such as Lipschitz continuity and matrix norms.

First, the simplest function, where the L-Lipschitz continuous condition is applicable on its real domain, is $f(x) = \frac{1}{2}x^T x$, as we note that

$$\nabla f = x, \|\nabla f(x) - \nabla f(y)\| = \|x - y\| \leq L \|x - y\|$$

holds whenever we pick a constant $L \geq 1$.

Second, the loss function for the logistic regression model

In short, Lipschitz continuity is not a strong assumption, and it is being used to bound the size of a gradient.

I also found it interesting to found the intrinsic relationship between eigenvalues and (some special) matrix norms, which is also being shown in this meeting. I then decided to read about spectral graph theory and hone my skills that are not well-covered in the first-year linear algebra course, which I didn't really understand how it matters. The main theorem is that a symmetric real matrix A has a matrix norm $\|A\|_2$ equal to its largest eigenvalue $\lambda_{max}(A)$.

Works To Be Done

- Cultivate the sense of analyzing several first-order and second-order optimization methods.
- Understand the `BinaryConnect` architecture and `ASkewSGD` algorithm.
- Discuss the results and shortcomings of the results shown in paper [A1] in the subsequent meeting.

3.3 Meeting III (Jun 21, 04:30 PM) In this meeting, we discussed the characteristics of the two papers [A1] and [A2]. [A2] provides an empirical understanding on the power of quantization whilst training neural networks as a regularizer, which stresses on the instrumental value of the method. As a comparison, [A1] has formulated an optimization problem as a suggestion of obtaining the error bound equipped with the ideal situation.

To get a better sense of writing and verifying proofs for the algorithm presented in [A1], Professor Wai suggested that I should implement the experiments and try to replicate the results shown in the paper. We also discussed about how the assumptions can be relaxed without using overgeneralized subgradient arguments.

Works To Be Done

- Experiment I - Implement the logistic regression model by stochastic gradient descendent algorithm and the `ASkewSGD` algorithm.
- Familiarize the definitions and meanings of the variables defined in the `ASkewSGD` model.
- Reconsider the convergence conditions for the algorithm.

3.4 Meeting IV (Jul 4, 10:30 AM) In this meeting, we exploited the technical aspects of the algorithm. Professor Wai suggested that several online tools could be used for NN performance plotting, and prevalent methods like Straight-Through Estimators (STEs) could be added for more comparison and for a more comprehensive explanation. Then, we have systematically compared the difference in application of quantization in different frameworks. We then turn to consider what makes the type of optimization problem formulated in [A1] not "easy". The reason is mainly due to (i) the disconnectedness of the feasible region $\mathcal{C}_\varepsilon$ controlled by the parameter ε used while pushing the weights towards the quantization constants, and (ii) the complexity of KKT point set, especially the normal cone of the constraint function g, $N_{\mathcal{C}_\varepsilon}$. Professor Wai has provided another important optimization technique which could possibly unravel and get over the obstacles aforementioned, which is to consider the dual optimization problem, which is likely to be accompanied with strong duality. To give a good start, Professor Wai also suggested me to read on Muehlebach and Jordan's paper, as the `ASkewSGD` algorithm is based on their work, but instead with convexity and deterministic conditions. This could be a great help for inspection about how `ASkewSGD` is proven.

Works To Be Done

- Compare the loss several QNN training methods: `BinaryConnect`, `Straight Through Estimators`, `Full-Precision`, and `ASkewSGD` (with constant/decreasing step sizes) with weights after quantization.
- Consider using tools (`plotly`, `wandb`) for plots.
- Read Muehlebach and Jordan's paper for simpler proofs with relaxed constraints.
- Difficulty: $\mathcal{N}_{C_\varepsilon}$ as in the KKT point set might be hard to deal with; The descendent direction $s_{\varepsilon,\alpha}(\widehat{\nabla\ell}, w)$ can be tricky when bounding the difference between the chosen direction and the stochastic gradient; The feasible region is smooth but is also broken down into pieces.
- Possible direction: Dive into solving the dual optimization problem, which will result in a gradient ascent algorithm.
- Noticing that we have constructed a function with "good" properties, we shall be able to provide better theoretical guarantees for its convergence.
- Distinguish the different directions for QNN training (with or without full-precision multiplications).

4 Acknowledgements

I truly appreciate Professor Wai Hoi-To's insightful advice and suggestions during the research period. His considerate pedagogical approach has greatly enhanced my understanding about the practical application of knowledge in research, and has contributed to my quicker sense to mathematical optimization problems.

5 References

(Major Texts)

[A1] L. Leconte, S. Schechtman and E. Moulines, (2023) ASkewSGD: An Annealed Interval-Constrained Optimisation Method to Train Quantized Neural Networks. In *Artificial Intelligence and Statistics 2023*, **206**:3644-3663.

[A2] M. Courbariaux, Y. Bengio and J.-P. David, (2015) BinaryConnect: Training Deep Neural Networks with Binary Weights During Propagations. In *Advances in Neural Information Processing Systems*, **28**:3123-3131.

[A3] L. Bottou, F. E. Curtis, J. Nocedal, (2018) Optimization Methods for Large-Scale Machine Learning. In *SIAM Review*, **60**(2):223-311.

[A4] M. Muehlebach, M. I. Jordan, (2022) On Constraitns in First-Order Optimization: A View from Non-Smooth Dynamical Systems. In *Journal of Machine Learning Research*, **23**:1-47.

[A5] J. Choi, P. I. Chuang, Z. Wang, S. Venkataramani, V. Srinivasan, K. Gopalakrishnan, (2018) Briding the Accuracy Gap for 2-bit Quantized Neural Networks (QNN). `arXiv:1807.06964v1`.

[A6] Y. Bai, Y.-X. Wang, E. Liberty, (2019) Proxquant: Quantized Neural Networks via Proximal Operators. In *The International Conference on Learning Representations 2019*.

[A7] B. Chimel, R. Banner, E. Hoffer, H. B. Yaacov, D. Soudry, (2023) Accurate Neural Training with 4-bit Matrix Multiplications at Standard Formats. In *The International Conference on Learning Representations 2023*.

[A8] P. Yin, J. Liu, S. Zhang, S. Osher, Y. Qi, J. Xin, (2019) Understanding Straight-Through Estimator in Training Activation Quantized Neural Nets. In *The International Conference on Learning Representations 2019*.

[A9] A. Gholami, S. Kim, Z. Dong, Z. Yao, M. W. Mahoney, K. Keutzer, (2021) A Survey of Quantization Methods for Efficient Neural Network Interface. `arXiv:2103.13630v3`.

(Associated Literature)

[B1] G. Lan, (2020) *First-order and Stochastic Optimization Methods for Machine Learning*. Switzerland: Springer Nature.

[B2] H.-T. Wai, (2024) *Lecture Notes of ESTR2520 - Optimization Methods*. Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong.

[B3] D. P. Bertsekas, (1999) *Nonlinear Programming (Second Edition)*. Massachusetts: Athena Scientific.

[B4] A. M.-C. So, (2021) *Handouts of ENGG5501 - Foundations of Optimization*. Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong.

Also read articles about spectral graph theory, graph signal processing.

[C1] J. Jiang, (2012) An Introduction to Spectral Graph Theory. *Research Experiences for Undergraduates Program (2012) of University of Chicago*.

[C2] X. Dong, D. Thanou, L. Toni, M. Bronstein, P. Frossard, (2020) Graph Signal Processing for Machine Learning: A eview and New Perspectives. In *IEEE Signal Processing Magazine*, **37**(6):117-127.