A note on state preparation for quantum machine learning

Zhikuan Zhao,[1] Vedran Dunjko,[2] Jack K. Fitzsimons,[3] Patrick Rebentrost,[4] and Joseph F. Fitzsimons[1,5,*]

[1] *Singapore University of Technology and Design, 8 Somapah Road, Singapore 487372*
[2] *Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Str. 1, 85748 Garching, Germany*
[3] *Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK*
[4] *Xanadu, 372 Richmond Street W, Toronto, Ontario M5V 1X6, Canada*
[5] *Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543*

The intersection between the fields of machine learning and quantum information processing is proving to be a fruitful field for the discovery of new quantum algorithms, which potentially offer an exponential speed-up over their classical counterparts. However, many such algorithms require the ability to produce states proportional to vectors stored in quantum memory. Even given access to quantum databases which store exponentially long vectors, the construction of which is considered a one-off overhead, it has been argued that the cost of preparing such amplitude-encoded states may offset any exponential quantum advantage. Here we argue that specifically in the context of machine learning applications it suffices to prepare a state close to the ideal state only in the ∞-norm, and that this can be achieved with only a constant number of memory queries.

In recent years, machine learning has emerged as a rich area for showing quantum speed-ups [1–6], based in part on the seminal quantum algorithm of Harrow, Hassidim and Lloyd [7] for solving systems of linear equations. This algorithm provides a way to generate quantum states proportional to $A^{-1}\mathbf{x}$, in time logarithmic in the length of the vector $\mathbf{x}$ (even with exponential precision [8]), and consequently to estimate the quantities of the form $\mathbf{y}^T A^{-1}\mathbf{x}$ as efficiently [5]. However, as Aaronson pointed out [9], this approach comes with several caveats relating to the sparsity and conditioning of A, and the structure of $\mathbf{x}$. These issues persist even under the reasonable assumption that the loading and construction of the required databases constitute a one-off overhead, and does not contribute to the computational complexity analysis of data processing.

Here we focus on the task of preparing a state

$$|\mathbf{x}\rangle = |\mathbf{x}|_2^{-\frac{1}{2}} \sum_{i=1}^{D} x_i |i\rangle, \tag{1}$$

given some vector $\mathbf{x} \in \mathbb{R}^D$ stored in quantum random access memory (QRAM) [10]. QRAM allows for data to be read from different memory locations in quantum superposition, enabling the operation

$$\sum_{i,j} \alpha_{ij} |i\rangle|j\rangle \xrightarrow{\text{QRAM}} \sum_{i,j} \alpha_{ij} |i\rangle|j + m_i\rangle \tag{2}$$

where m_i is the ith entry stored in memory. QRAM provides a way to probabilistically produce $|\mathbf{x}\rangle$ for any $\mathbf{x}$ stored in memory. To produce $|\mathbf{x}\rangle$ for any $\mathbf{x}$, one can start with the state $D^{-\frac{1}{2}} \sum_i |i\rangle|0\rangle$ as the query state for the QRAM to obtain $D^{-\frac{1}{2}} \sum_i |i\rangle|x_i\rangle$. An ancillary qubits is prepared in state $|0\rangle$ and then conditionally rotated based on the second register to obtain $D^{-\frac{1}{2}} \sum_i |i\rangle|x_i\rangle \left(\sqrt{1-|x_i|^2}|0\rangle + x_i|1\rangle\right)$, where we have assumed for simplicity that $\mathbf{x}$ is normalised such that $|x_i| \leq 1$. Performing a second QRAM call to uncompute the registers $|x_i\rangle$, and post-selecting onto $|1\rangle$ results in the desired state $|\mathbf{x}\rangle$.

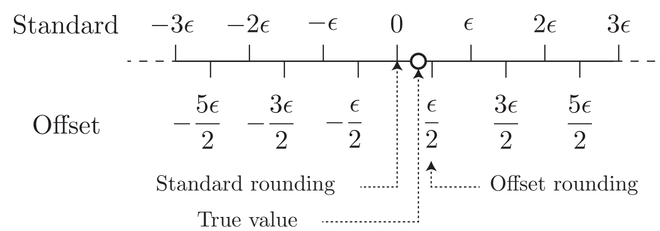

FIG. 1: Numerical rounding conventions. In the standard rounding convention scalar values are rounded to the nearest integer multiple of precision ϵ. Alternatively, we can consider an offset rounding convention, where the rounding is to the nearest half-integer multiple of ϵ. As a result, this numbering convention does not contain an exact representation of 0. In either scheme, the rounded value is always within $\frac{\epsilon}{2}$ of the true value.

One significant barrier to the QRAM approach to state preparation is the probability of projecting onto the correct subspace in the final step, which is given by $D^{-1}\sum_i |x_i|^2$. When the entries of $\mathbf{x}$ are of similar magnitude, $|\mathbf{x}\rangle$ can be prepared using constant queries. However, in the case where a few entries are much larger than the rest, the lower bounds on unordered search [11] imply that the corresponding state requires $\Omega\left(\sqrt{D}\right)$ QRAM queries [12]. This argument can be extended to the case where it is only necessary to prepare any $|\mathbf{x}'\rangle$ such that $|\mathbf{x}' - \mathbf{x}|_2$ is sufficiently small.

However, note that data processing in machine learning always implicitly assumes robustness with respect to the ∞-norm induced distance, which measures just the largest entry-wise perturbation. Specifically, any digital data processing with floating-point arithmetic only makes sense if the overall results of such a computation maintain its validity when the features of the input vector had been perturbed below the machine precision. This is also practically reasonable, since real world data comes from measurements of finite precision. As such, the corresponding constraint is only that $|\mathbf{x}' - \mathbf{x}|_\infty \leq \epsilon$, rather than requiring closeness in the 2-norm.

The tolerance against small ∞-norm perturbations allows us to work with the vector entries x_i' which are half-integer multiples of the base precision ϵ. Such that $\mathbf{x}'$ is chosen to be the closest representable vector to $\mathbf{x}$ (as shown in Figure 1), which satisfies $|\mathbf{x}' - \mathbf{x}|_\infty \leq \frac{\epsilon}{2}$ and the distance from the true value of the data is less than ϵ. This offset rounding can be implemented in the loading of the QRAM, or effectively realised at the controlled rotation stage, as discussed earlier. The aforementioned robustness requirement in data processing guarantees the analysis of results are insensitive to such offset rounding. The key benefit in using the offset rounding is that, since the exact representation of 0 is not included in this offset rounding convention, the probability of the final projection step succeeding is at worst $\frac{\epsilon^2}{4}$, and hence independent of D. This success probability can be further improved to $\mathcal{O}(\epsilon)$ using fixed-point quantum amplitude amplification [13]. In the cases where a systematic perturbation of data-vectors, by utilizing, say, a positive sign offset $(+\epsilon/2)$ to data-points is undesirable, one can opt for a near-white noise offset using either a suitable pseudo-random number generator seeded by the memory location being queried or by adding data random data stored elsewhere in QRAM. Furthermore, while ϵ need not be on the order of machine precision, since any ϵ which is small compared to accuracy to which the input data is known will lead to negligible error, and hence low precision data will have a more efficient loading procedure then high precision data. Critically, however, the number of QRAM queries necessary to successfully prepare the state always has an upper bound that is independent from the database size.

In summary, any application which is robust under small ∞-norm perturbations allows for efficient state preparation. This suggests that the caveat related to state preparation raised by Aaronson [9] can generally be overcome in the context of machine learning, due to the natural robustness assumption. This feature is, however, not necessarily shared by computational physics, or other numerical mathematics applications.

Acknowledgements— JFF and ZZ acknowledge support from Singapore's Ministry of Education and National Research Foundation. JFF acknowledges support from the Air Force Office of Scientific Research under AOARD grant FA2386-15-1-4082. This material is based on research funded in part by the Singapore National Research Foundation under NRF Award NRF-NRFF2013-01.

* Electronic address: joseph˙fitzsimons@sutd.edu.sg

[1] E. Aïmeur, G. Brassard, and S. Gambs, in *Advances in Artificial Intelligence* (Springer, 2006), pp. 431–442.

[2] S. Lloyd, M. Mohseni, and P. Rebentrost, arXiv preprint arXiv:1307.0411 (2013).

[3] P. Rebentrost, M. Mohseni, and S. Lloyd, Physical Review Letters **113**, 130503 (2014).

[4] M. Schuld, I. Sinayskiy, and F. Petruccione, in *PRICAI 2014: Trends in Artificial Intelligence* (Springer, 2014), pp. 208–220.

[5] Z. Zhao, J. K. Fitzsimons, and J. F. Fitzsimons, arXiv preprint arXiv:1512.03929 (2015).

[6] V. Dunjko and H. J. Briegel, Reports on Progress in Physics (2018).

[7] A. W. Harrow, A. Hassidim, and S. Lloyd, Physical Review Letters **103**, 150502 (2009).

[8] A. M. Childs, R. Kothari, and R. D. Somma, SIAM Journal on Computing **46**, 1920 (2017).

[9] S. Aaronson, Nature Physics **11**, 291 (2015).

[10] V. Giovannetti, S. Lloyd, and L. Maccone, Physical Review Letters **100**, 160501 (2008).

[11] M. Boyer, G. Brassard, P. Høyer, and A. Tapp, arXiv preprint quant-ph/9605034 (1996).

[12] A. N. Soklakov and R. Schack, Physical Review A **73**, 012307 (2006).

[13] T. J. Yoder, G. H. Low, and I. L. Chuang, Physical review letters **113**, 210501 (2014).