Towards provably efficient quantum algorithms for nonlinear dynamics and large-scale machine learning models

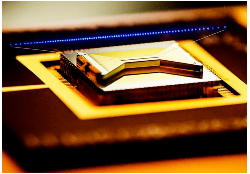
Jin-Peng Liu
Postdoctoral Fellow, Berkeley and MIT
Incoming Assistant Professor, Tsinghua

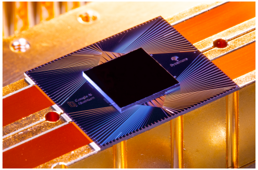
Oct 2023

Tsinghua University

Table of Contents

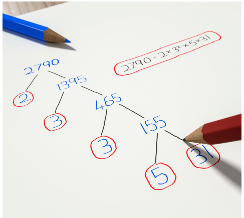
Quantum Computers

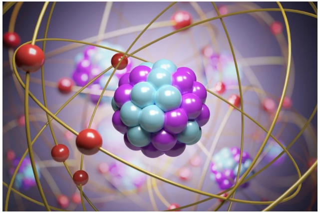


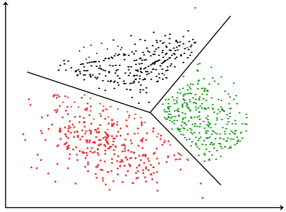


Quantum computers can outperform classical ones for certain tasks. Potential exponential speedups for

- integer factoring;
- quantum physics;
- linear algebra problems.

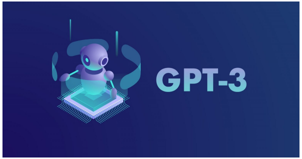






Quantum Revolution and Challenge

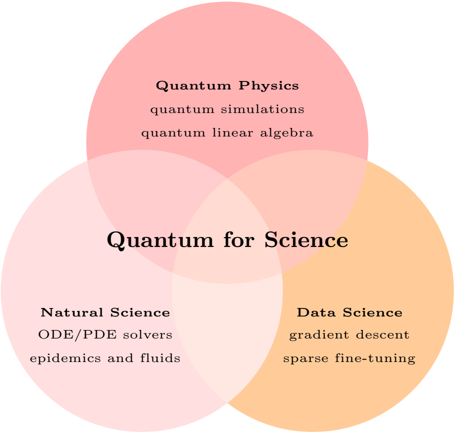
Potential killer-apps: large-scale computational models, e.g. epidemic spreading, climate change, and training large language models?

Challenges:

- From linear quantum mechanics to nonlinear realistic systems.
- End-to-end design: interface between classical and quantum data.

From “quantum for quantum” toward “quantum for science”, by cracking the nonlinear problem and end-to-end setting.



Quantum Simulation

Given a description of an s-sparse $n \times n$ Hamiltonian system

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}|\Psi(t)\rangle = H|\Psi(t)\rangle, \quad |\Psi(0)\rangle = |\Psi_{\text{in}}\rangle, \tag{1}$$

produce the final state

$$|\Psi(t)\rangle = e^{-iHt}|\Psi(0)\rangle \approx U_K \cdots U_2 U_1 |\Psi(0)\rangle. \tag{2}$$

- Complexity: $sT\,\mathrm{poly}(\log n, \log(1/\epsilon))$[1].
- Applications to quantum physics and quantum chemistry.

[1] [LLoyd 96; Berry et al. 15; Low, Chuang 17]

Quantum Linear System Algorithms

Given a description of an s-sparse $n \times n$ linear system

$$Ax = b, \tag{3}$$

produce a quantum encoding of the solution proportional to $x = A^{-1}b$.

- Complexity: $s\kappa \operatorname{poly}(\log n, \log(1/\epsilon))$[1], $\kappa = \|A\| \cdot \|A^{-1}\|$.
- Generalization to linear algebra problems.

[1][Harrow, Hassidim, Lloyd 09; Ambainis 12; Childs, Kothari, Somma 15]

Given a description of an s-sparse $n \times n$ linear ODE system

$$\frac{\mathrm{d}u(t)}{\mathrm{d}t} - A(t)u(t) = f(t), \quad u(0) = u_{\text{in}}, \tag{4}$$

produce a quantum encoding of the solution proportional to $u(T)$.

- Complexity: $sTq\,\mathrm{poly}(\log n, \log(1/\epsilon))$[1], $q = \|u_{\text{in}}\|/\|u(T)\|$.
- Applied to PDEs, open quantum systems, fast-forwarding and lower bounds.

[1] [Berry 14; Berry et al. 17; Childs, Liu 19; Krovi 22; Berry, Costa 22]

It is natural to simulate quantum (unitary) dynamics.

It takes extra cost to simulate linear and non-quantum (non-unitary) dynamics, e.g. linear combinations of unitaries.

It is exponentially expensive to simulate nonlinear dynamics! Longstanding open problem.

Inefficient quantum algorithm

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \sum_{j,k=1}^{n} \alpha_{jk}^{(i)} u_j u_k \approx \frac{u_i(t+\Delta t) - u_i(t)}{\Delta t}. \tag{5}$$

Consider $|\phi_t\rangle = \sum_j u_j |j\rangle$, and use $|\phi_t\rangle|\phi_t\rangle = \sum_{jk} u_j u_k |jk\rangle$ to generate $|\phi_{t+\Delta t}\rangle$.

No-cloning theorem: need to maintain totally $2^{O(T)}$ multiple copies of $|\phi_0\rangle$ for one $|\phi_T\rangle$[1].

Quantum lower bound

Nonlinearity implies poly-time solution for #P and PSPACE problems[2].

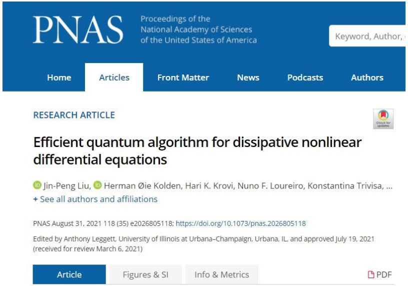
[1] [Leyton, Osborne 08]

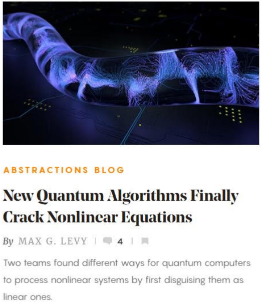
[2] [Abrams, Lloyd 98; Aaronson 05; Childs, Young 16]

Our Contributions

1 Efficient Quantum Algorithm for Nonlinear Dynamics

“The research Liu conducted at UMD on quantum algorithms for nonlinear differential equations exponentially improves over the previous best quantum algorithm in 15 years.”–Andrew Childs, June 2023





“While these are significant steps, they are still among the first in cracking nonlinear systems. More researchers will likely analyze and refine each method —even before the hardware needed to implement them becomes a reality.”–John Preskill, Jan 2021

Linear approximation approach

- Carleman Koopman, non-Hermitian Hamiltonian, homotopy, etc[1].
- n-dim nonlinear ODEs $\to$ $\mathrm{poly}(n)$-dim linear ODEs with truncation.
- Need to control the truncation error.

Phase space representation approach

- Liouville, Koopman, level set, etc[2].
- n-dim nonlinear ODEs $\to$ $\mathbb{R}^n$-dim linear PDEs ($\exp(n)$ grids after discretization).
- Hard to obtain exponential quantum speedup.

---

[1] [Liu et al. 21; Lloyd et al. 21; Engel et al. 21; Xue et al. 21; Ameri et al. 22; Itani, Succi 22; Li et al. 23]

[2] [Joseph 21; Dodin and Startsev 21; Jin and Liu 22; Jin, Liu, Yu 22]

Quantum Nonlinear ODE Problem

$$\frac{\mathrm{d}u}{\mathrm{d}t} = F_2 u^{\otimes 2} + F_1 u + F_0(t), \quad u(0) = u_{\text{in}}. \tag{6}$$

F_1 is dissipative: $\mathrm{Re}(\lambda_n) \leq \cdots \leq \mathrm{Re}(\lambda_1) < 0$, e.g. Burgers equation

$$\partial_t u = -\frac{1}{2}\partial_x u^2 + \nu \partial_x^2 u + f. \tag{7}$$

We define

$$R = \frac{1}{|\mathrm{Re}(\lambda_1)|}\left(\|u_{\text{in}}\|\|F_2\| + \frac{\|\|F_0\|\|}{\|u_{\text{in}}\|}\right). \tag{8}$$

It quantifies nonlinear and inhomogeneous strengths relative to dissipation.

R varies when it is generalized to polynomial nonlinearities.

First Poly-time Quantum Algorithm

Theorem 1 of [Liu et al. 21]

Assume $R < 1$. Let $q = \|u_{\text{in}}\| / \|u(T)\|$. There is a quantum algorithm that outputs the quantum state $|u(T)\rangle$ proportional to $u(T)$ within ϵ, with complexity

$$\frac{sT^2q}{\epsilon} \operatorname{poly}(\log n). \tag{9}$$

- Improved complexity: $sTq \operatorname{poly}(\log n, \log(1/\epsilon))$, and better convergence criteria[1].
- Applied to classical nonlinear control and classical computational fluid dynamics[2].

[1] [Krovi 21; An et al. 22]

[2] [Foret, Schilling 21; Itani, Succi 22; Itani, Sreenivasan, Succi 23; Li et al. 23]

Quantum Carleman Linearization

Considering a 1-dim quadratic ODE: $\frac{\mathrm{d}u}{\mathrm{d}t} = au^2 + bu + c$.
Naive linearization $u^2 \approx u(0)u$ does not work in long time.

Carleman linearization

- Embed an n-dim nonlinear ODEs to an infinite-dim linear ODEs.
- Truncate the dimension to obtain a $\mathrm{poly}(n)$-dim system, for which we prove the convergence when $R < 1$.
- Develop a quantum algorithm for the linearized ODEs with complexity $\mathrm{poly}(\log n)$.

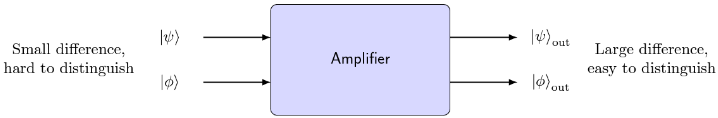
1-dim example: $\frac{\mathrm{d}u}{\mathrm{d}t} = au^2 + bu + c$, $\frac{\mathrm{d}u^2}{\mathrm{d}t} = 2au^3 + 2bu^2 + cu$, ..., $\frac{\mathrm{d}u^N}{\mathrm{d}t} \approx Nbu^N + Ncu^{N-1}$,
giving linear ODEs with observable variables $y_j \approx u^j$.

Quantum Lower Bound for Nonlinear ODEs

Theorem 2 of [Liu et al. 21]

Assume $R \geq \sqrt{2}$. Then there is an instance of the quantum quadratic ODE problem such that any quantum algorithm must have worst-case time complexity exponential in T.

- Hardness of distinguishing nonorthogonal quantum states.
- Butterfly effect: a small initial divergence results in a large violation.

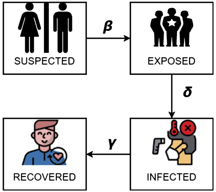


A recent paper give a tighter instance with $R \geq 1$ and close the gap[1].

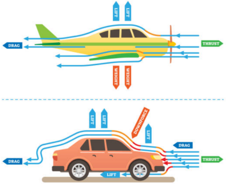
[1] [Lewis, Eidenbenz, Nadiga, Subasi 23]

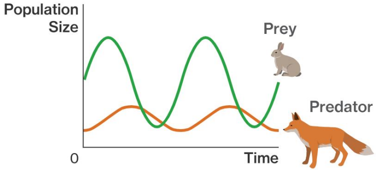
Applications in Scientific Computation

1 Efficient Quantum Algorithm for Nonlinear Dynamics

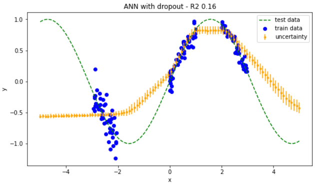


- Biology/epidemiology: SEIR model
- Fluid and plasma dynamics





- Nonlinear control: find attractors
- Uncertainty quantification



Reaction-diffusion system

$$\frac{\partial u}{\partial t}(x,t) = D\Delta u(x,t) + f(u(x,t)), \quad x \in [0,1]^d. \tag{10}$$

- $f(u) = u - u^2$: biological/ecological networks, spin glasses.
- $f(u) = u - u^3$: phase separation, Ginzburg-Landau theory.
- $f(u) = \sum_K a_k u^k$: disorder systems, branching processes.

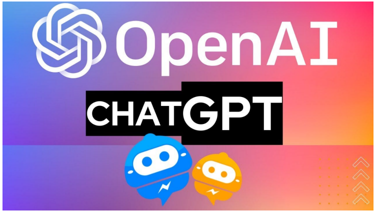
We develop quantum algorithm for $|u\rangle$ and estimate kinetic energy with complexity $O(T^2 d^2/\epsilon)$[1], a potential exp speedup compared to classical PDE solvers.

Improved analysis may indicate that linearization works better for gradient flows.

[1] [An et al. 22]

Table of Contents

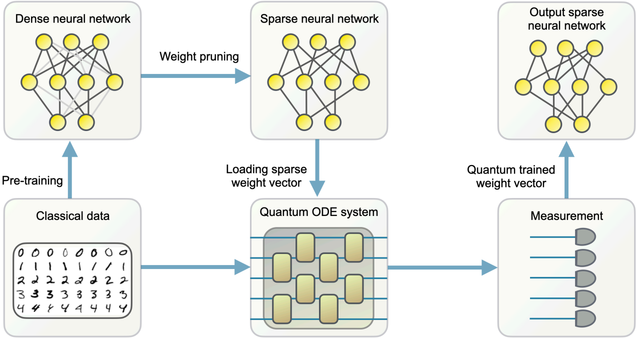
ChatGPT and DALL·E 3



Training language models like GPT-3 with over 10^{11} parameters is costly: 1.2×10^7 dollars and over 500 tons of CO_2 are produced.

$N = 10^{11}$, $\log N = 36.54$. Quantum: 0.001s v.s. Classical: 1 month. Is it possible?

Illustration of Model



Towards provably efficient quantum algorithms for large-scale machine-learning models

Junyu Liu,[1,2,3,4,5,6] Minzhao Liu,[7,8] Jin-Peng Liu,[9,10,11] Ziyu Ye,[2]
Yunfei Wang,[12] Yuri Alexeev,[8,2,3] Jens Eisert,[13] and Liang Jiang[1,3]

Efficient Quantum Algorithm

Theorem 1 and 2 of [Liu et al. 23]

For a sparse and almost dissipative machine learning model with sparsity s, size n, running in T iterations, a quantum algorithm for approximating the training dynamics takes

$$s^2T^2\,\mathrm{poly}(\log n). \tag{11}$$

Moreover, the algorithm outputs s-sparse weight vectors with tomographic cost $O(s^3/\epsilon^2)$.

- In sparse training, the sparsity doesn't scale with the size of the model.
- For Hessian of gradient dynamics, dissipative modes $\gg$ divergent modes.

Given loss function $\mathcal{L}_{\mathcal{A}}$, input weights $\theta(0)$, training set $\mathcal{A}$, we consider the gradient descent on the MSE loss

$$\theta(t+1) = \theta(t) - \eta \frac{\mathrm{d}\mathcal{L}_{\mathcal{A}}}{\mathrm{d}t}, \quad \mathcal{L}_{\mathcal{A}} = \frac{1}{2} \sum_{\alpha \in \mathcal{A}} |z^{(L)}(x_\alpha; \theta) - y|^2. \tag{12}$$

Multilayer perceptron (MLP) model

For $l \in [L]_0$, the l-th layer preactivation is defined as

$$z^{(l+1)}(x) = b^{(l+1)} + W^{(l+1)} \sigma\left(z^{(l)}(x)\right), \tag{13}$$

where x is the input data vector, $W^{(l)}$ and $b^{(l)}$ are the l-th layer trainable weights and biases, vectorized as θ, and σ is a polynomial activation.

Quantum data upload

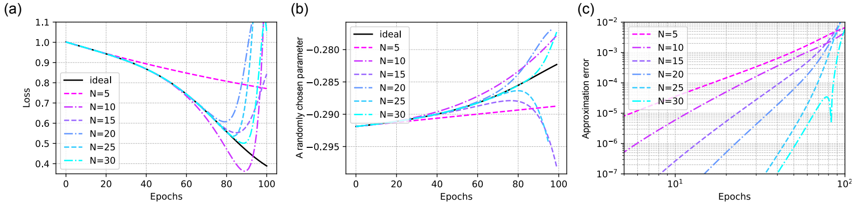
Quantum state preparation to upload s-sparse initial $|\theta(0)\rangle$ with $O(s)$; or QRAM for dense parameters.

Quantum training process

For almost dissipative traing dynamics, we develop a probabilistic Carleman linearization and truncated quantum linear system algorithm with $O(s^2T^2)$.

Quantum data download

Shadow tomography to download s-sparse trained $|\theta(T)\rangle$ with $O(s^3/\epsilon^2)$; or QRAM for dense parameters.



Single-hidden-layer neural network

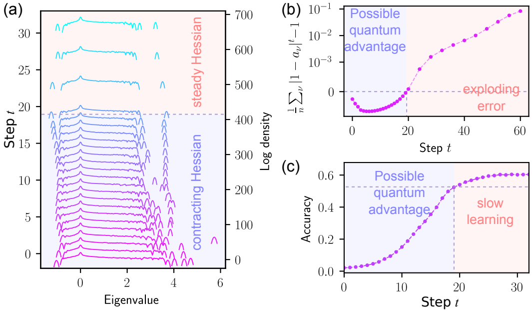
We examine a simple neural network (1 input, 2 hidden units, 1 output), applied to Iris data set

$$z_\alpha = \theta_5(x_\alpha\theta_1 + \theta_3)^2 + \theta_6(x_\alpha\theta_2 + \theta_4)^2. \tag{14}$$

The error scales exponentially in N, and has a power law scaling of time.

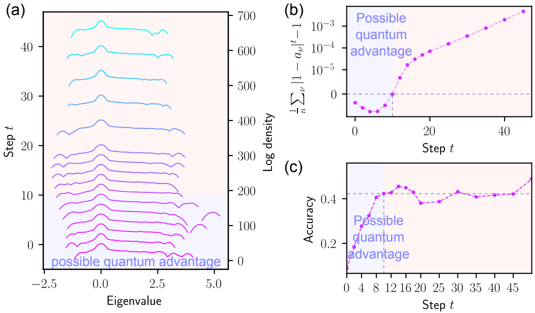
Classical Experiments

2 Efficient Quantum Algorithm for Large-scale Machine Learning Models



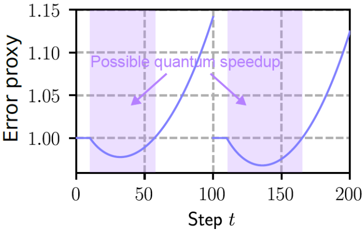
Sparse ResNet for vision

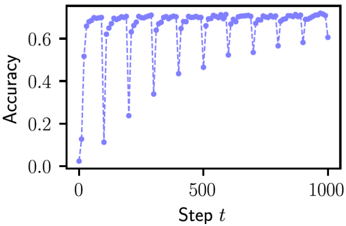
ResNet with depth 32 and 7×10^6 parameters, by sparse pruning and linearization, applied to CIFAR-100 data set. We record the Hessian spectra during sparse training to track the error propagation.



Sparse ResNet for vision

ResNet with depth 422 and 1.03×10^8 parameters, by sparse pruning and linearization, applied to CIFAR-100 data set. Possible quantum advantage in early-stage training process.





Sparse ResNet for vision

To reset the error proxy as zero, we download the quantum trained parameters sparsely and re-upload to the quantum computer to continue training every 100 steps.

Information Bottleneck Theory

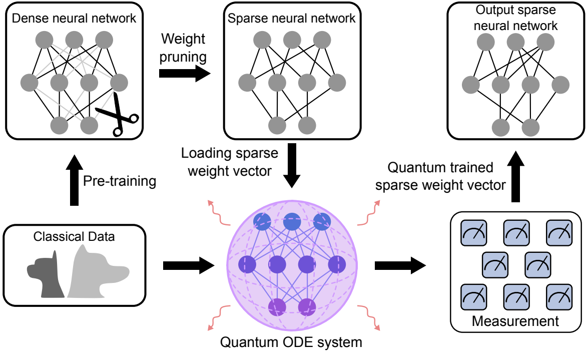
- When a system is actually learning knowledge, the Fisher information matrix (and second-order Hessian) has more positive modes, i.e. dissipative.
- Unitary models like variational quantum algorithms have numerous saddle points and are possibly hard to train.

Lottery Ticket Hypothesis

- A dense network contains good enough sparse subnetworks (winning ticket).
- In pre-training, data are typically dense; in fine-tuning, most parameters are reasonably pruned to zero throughout the training.

Quantum-enhanced Machine Learning

2 Efficient Quantum Algorithm for Large-scale Machine Learning Models



- First efficient quantum algorithm for training classical neural networks.
- 103M parameter classical experiments to validate possible quantum advantage.

Future Improvements

- Analysis of linearization beyond polynomial nonlinearities.
- Better criteria with weak/no dissipation.
- Connections to other PDE/ML models, such as diffusion models.
- Develop near-term or early fault-tolerant quantum algorithms.
- Estimate detailed running costs and resource counts.

Current and Future Research

2 Efficient Quantum Algorithm for Large-scale Machine Learning Models