Optimizing a Polynomial Function on a Quantum Simulator

Keren Li,[1, *] Shijie Wei,[1, *] Feihao Zhang,[1] Pan Gao,[1] Zengrong Zhou,[1] Tao Xin,[2] Xiaoting Wang,[3, †] and Guilu Long[1, ‡]

[1]*State Key Laboratory of Low-Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing 100084, China*
[2]*Department of Physics and Institute for Quantum Science and Engineering,*
Southern University of Science and Technology, Shenzhen 518055, China
[3]*The basic and Frontier Research Institute, Electronic Science and Technology University,Chengdu,610054,China*
(Dated: April 17, 2018)

Gradient descent method, as one of the major methods in numerical optimization, is the key ingredient in many machine learning algorithms. As one of the most fundamental way to solve the optimization problems, it promises the function value to move along the direction of steepest descent. For the vast resource consumption when dealing with high-dimensional problems, a quantum version of this iterative optimization algorithm has been proposed recently[arXiv:1612.01789]. Here, we develop this protocol and implement it on a quantum simulator with limited resource. Moreover, a prototypical experiment was shown with a 4-qubit Nuclear Magnetic Resonance quantum processor, demonstrating a optimization process of polynomial function iteratively. In each iteration, we achieved an average fidelity of 94% compared with theoretical calculation via full-state tomography. In particular, the iterative point gradually converged to the local minimum. We apply our method to multidimensional scaling problem, further showing the potentially capability to yields an exponentially improvement compared with classical counterparts. With the onrushing tendency of quantum information, our work could provide a subroutine for the application of future practical quantum computers.

Introduction — Machine learning has fundamentally transformed the way people think and behave. It is so pervasive that it affects our daily life in every aspect, from practical facial recognition to the astonishing artificial intelligence AlphaGo, which defeated the best human player in the game of Go. As each machine learning algorithm consists of three components: representation, evaluation and optimization[1], gradient decent algorithm, as one of the most fundamental approaches to solve optimization problems, lies at the heart of machine learning methods, such as regression, support vector machines, and deep neural networks[2]. However, when the size of problem grows, gradient algorithm consumes tremendous resource. Particularly, when dealing with large data, in many applications, current computational resources are often pushed to their limit.

As the feat raises hope to have a big impact on the machine learning, quantum computing handle information process based on quantum mechanics which can provide potentially significant improvements compared with classical computer[3–5]. In this community, energy is suddenly buzzing and the recent progress is immense[6–8]. As the result of both thriving area, quantum machine learning, as an interdisciplinary research between machine learning and quantum information, has undergone a flurry of developments in recent years[9–12]. Among them, quantum-enhanced machine learning, referring to quantum algorithms that solve tasks in machine learning, thereby promising enhancement of the performance of machine learning algorithms for problems beyond the reach of classical computing, has obtained many interesting achievements since the breakthrough quantum algorithm for solving linear equation system[13–17].

Inspired by the protocol which brilliantly maps the gradient operator into a matrix[18], in this paper, we propose a quantum computational framework of gradient descent algorithm to process the non-convex optimization problems. Compared with the previous method, which requires multiple copies depending on the order of objective function, we avoid the phase estimation and reduce the quantum memory consumption to only two copies of the quantum state. In the case that the coefficient matrix of the objective function can be decomposed efficiently, our method exponentially outperforms corresponding classical algorithms. Moreover, we experimentally demonstrate a polynomial function optimization process with a normalization constraint iteratively, implementing it on a 4-qubit nuclear magnetic resonance (NMR) quantum simulator. For the consequence of algorithm, we obtained the point which finally arrived in the vicinity of a local minimum. In addition, we present a potential application of our method to solve multidimensional scaling problems. In a sense, our work attributes the community of quantum enhanced machine learning and opens up a way of the potential application for future practical quantum computer .

Algorithm frame — The basic idea of classical gradient descent is sketched as follows. To obtain the minimal of an objective function, which is denoted as as a map $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, one could set an initial point $\mathbf{X}^{(0)} \in \mathbb{R}^N$ and then iteratively move to next point in the direction of the gradient of the function at the current point.

$$\mathbf{X}^{(t+1)} = \mathbf{X}^{(t)} - \eta \, \nabla f(\mathbf{X}^{(t)}), \tag{1}$$

where η is a positive learning rate. Now, we transform above process into a quantum version. For simplicity, define a homogeneous objective function as a polynomial of order $2p$ over $\mathbf{X} \in \mathbb{R}^N$[18] ,

$$f(\mathbf{X}) = \sum_{i_1,\ldots,i_{2p}=1}^{N} a_{i_1\ldots i_{2p}} x_{i_1} \ldots x_{i_{2p}}, \tag{2}$$

with real number coefficients $a_{i_1\ldots i_{2p}} \in \mathbb{R}$ and $\mathbf{X} = (x_1, \ldots, x_N)^T$.

As the classical gradient operation is a non-linear operation, to transform the gradient process into the quantum operations, first we map the inputs $\mathbf{X} = (x_1, \ldots, x_N)^T$ to a higher dimensional space and represent the objective function as[18]

$$f(\mathbf{X}) = \mathbf{X}^T \otimes \cdots \otimes \mathbf{X}^T \mathbf{A} \, \mathbf{X} \otimes \cdots \otimes \mathbf{X}. \qquad (3)$$

where $\mathbf{A}$ can be decomposed as a sum of tensor products of unitary matrices

$$\mathbf{A} = \sum_{\alpha=1}^{K} \mathbf{A}_1^\alpha \otimes \cdots \otimes \mathbf{A}_p^\alpha, \qquad (4)$$

where each $\mathbf{A}_i^\alpha$ is a $N \times N$ unitary matrix for $i = 1, \ldots, p$ and K is the number of terms in the sum needed to specify $\mathbf{A}$. Due to the above extension to higher Hilbert space as shown in Eq. (4), the gradient of the objective function at point $\mathbf{X}$ can be expressed as a linear process

$$\nabla f(\mathbf{X}) = \sum_{\alpha=1}^{K} \sum_{j=1}^{p} \Big(\prod_{\substack{i=1 \\ i \neq j}}^{p} \mathbf{X}^T \mathbf{A}_i^\alpha \mathbf{X} \Big) \mathbf{A}_j^\alpha \, \mathbf{X}. \qquad (5)$$

Define an operator

$$\mathbf{D} =: \sum_{\alpha} \sum_{j} \Big(\prod_{i \neq j} \mathbf{X}^T \mathbf{A}_i^\alpha \mathbf{X} \Big) \mathbf{A}_j^\alpha \qquad (6)$$

Thus, the gradient process can be interpreted as an evolution of $\mathbf{X}$ under operator $\mathbf{D}$. The iteration process can be expressed as

$$|\mathbf{x}^{(t+1)}\rangle = \Big(|\mathbf{x}^{(t)}\rangle - \mathbf{D}|\mathbf{x}^{(t)}\rangle \Big). \qquad (7)$$

Details about the key component $\mathbf{D}$ can be found in Appendix.A[19]. Here, we represent $\mathbf{X}$ as the quantum states $|\mathbf{x}^{(t+1)}\rangle$, which naturally introduce to a constraint $\mathbf{X}^T \mathbf{X} = 1$. In optimization area, minimizing (or maximizing) a polynomial function, subject to some suitable constraints, is a fundamental model. Specific on the optimization problems with normalization constraint which denoted as spherical constraint in optimization literature, applications spread from image and signal processing , speech recognition , biomedical engineering to material science and quantum mechanics[20].

The flow of the entire gradient descent algorithm is shown in Fig.1, and can be divided into $'Soul'$.

Superposition: The initial guess or the current point $\mathbf{X} = (x_1, \ldots, x_N)^T$ should be efficiently presented as a initial state $|\mathbf{x}^{(t)}\rangle$, with the ancillary registers evolving from the initialization state $|0\rangle|0\rangle^{T_1}$ to a specific superposition state $|\psi_s\rangle$

$$|\psi_s\rangle = \beta|0\rangle|0\rangle^{T_1} + \sum_{m=0}^{Kp-1} c_m|1\rangle|m\rangle \qquad (8)$$

where $c_m = \dfrac{\prod_i \mathbf{X}^T \mathbf{A}_i^\alpha \mathbf{X}}{\mathbf{X}^T \mathbf{A}_j^\alpha \mathbf{X}}$ and $\beta = \sum_{m=0}^{Kp-1} c_m + 1$.

Operations of 'A_m' series: After the ancillary system superposed, a series of control A_m operations are applied.
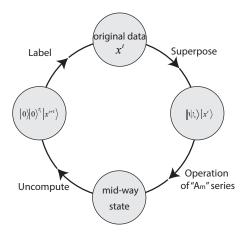

Figure 1. $Soul$ Flow, includes four parts: Superpose, Operation of 'A' series, Uncompute and Label.

The first ancillary system controlled operation $A_0^\dagger \otimes |0\rangle\langle 0|$ and the second ancillary system controlled operations $A_0 \otimes |0\rangle\langle 0|, A_1 \otimes |1\rangle\langle 1|, \ldots, A_{Kp-1} \otimes |Kp-1\rangle\langle Kp-1|$ are implemented successively on the working system $|\mathbf{x}^{(t)}\rangle$. This step performs the linear combination of unitary operators on $|\mathbf{x}^{(t)}\rangle$ to realize the iteration process described by Eq.7. And the unitary operator linear combination can be implemented by a standard method called duality quantum computing by adding ancillary qubits[21–24] . hence we got the mid-way state.

Uncompute: After the the operations of control-A_i, we uncompute the registers $\sqrt{\alpha}|0\rangle|0\rangle + \sum_{m=0}^{Kp-1} c_m|1\rangle|m\rangle$ and measure the two registers. If we obtain $|0\rangle|0\rangle^{T_1}$, our protocol successes and we have the state

$$\frac{1}{\beta} \Big(|\mathbf{X}^{(t)}\rangle - D|\mathbf{X}^{(t)}\rangle \Big) \qquad (9)$$

i.e. the new current point $\mathbf{X}^{(t+1)}$. If we make measurement directly, the probability of detecting the ancillary state $|0\rangle|0\rangle^{T_1}$ is

$$P_s = \| \, |\mathbf{X}^{(t)}\rangle - D|\mathbf{X}^{(t)}\rangle \, \|^2 / (\sum_{m=0}^{Kp-1} c_m + 1)^2$$

To Amplify the amplitude of the desired term before the measurement, the robust obvious amplitude amplification can be employed and about $O(\sqrt{Kp})$ repetitions are sufficient to get the satisfying result[25] .

Label: Finally, after measurement, we would label the result either as the new iteration point or the purpose state based on the previous set-up convergence condition. If the object function of such current point does not hit our pre-set threshold, this output x^{t+1} can be regard as the new current x^t and run the next iterative cycle. Otherwise, the iteration can be terminated and this x^{t+1} is the final result for this objective function.

$Complexity$ — (1) Gate complexity. At the $Superposition$ part of the $'Soul'$, method in reference[26] could help us pre-

pare an initial state in $O(log_2(N+KP+1))$ steps. Moreover, the controlled operations A_m at " Operations of 'A_m' series " part can be decomposed into $O(Kplog_2(N))$ basic gates [27]. In Appendix.A, the complexity of quantum circuit of obtaining the coefficient c_m is $O(Kplog_2(N))$. In summary, counting the repetitions of amplitude amplification, our algorithm uses roughly $O(K^{3/2}p^{3/2}log_2(N) + \sqrt{Kp}(log_2(N + KP + 1)))$ steps to create the next state. Thus, if d and K only grow polylogarithmically with N, our algorithm could achieve an exponential speedup over classical algorithms. (2) Memory consumption. The ancillary qubits in the whole algorithm are $2T_1+1$, where $T_1 = log_2(N)$ and two copies of $|X\rangle$ are needed. Thus, the total qubits resource is $2log_2(N) + 1$. The query complexity for the full simulation algorithm is $O(\sqrt{Kp}Kp + Kp)$.

Experimental implementation — In order to show the flow of *Soul*, a constricted bivariate quartic function(Eq.10) works as the objective function and approaches its local minimum iteratively:

$$min \quad f(x) = \frac{1}{2}x^T \otimes x^T \widehat{A}x \otimes x$$
$$s.t. \quad norm(x) = 1 \tag{10}$$

where x is a 2-dimension real vector. Since the normalization of x, it has only one-dimension freedom. However, as the growth of the size of x, a surge of information processing would be included. $\widehat{A}$ is the coefficient matrix and could be represented by a tensor products series $\widehat{A} = -\sigma_I \otimes \sigma_X + \sigma_X \otimes \sigma_Z$ ($\sigma_i(i = I, x, y, z)$ denote the Pauli matrices). Since the mission to find the minimum of this object function, tranditionally, a first-order iterative optimization algorithm— gradient descent is employed iteratively: 1) Calculating the gradient of the objective function at the current point x; 2)Obtaining the new current point x' by taking steps proportional to the negative of the gradient at the old current point x. Experimentally, those two steps can be replaced by the quantum version shown in Fig.2(b). In each iteration, from the $|0\rangle|0\rangle^2|x\rangle$, the new current point x' can be observed at the end of set of operations and the objective function can be thus optimized iteratively. As it is shown in Fig.2(b), each iteration for implementing gradient descent can be realized by 3 steps: initially, a four-qubit quantum register is required to save a random guess or the current point, $x = (x_1, x_2)^T$. Three qubits work as ancillary system, assisting x to form the wanted superposition state $|\psi_s\rangle|x\rangle$(as Eq.8 shows and T_1=2, Kp=4). After-that, a series of $control$-$A_i(i = 1...4)$ operation follows, as they came from the decomposition of D, to implement the gradient operation over the objective function. It is worth mentioning that the number of $control$-A_i only depends on the coefficient matrix $\widehat{A}$. Finally, after the uncompute operation, in the subspace of $|0\rangle_1|00\rangle_{23}$, the new current point x' can be obtained. if the object function of this current point does not hit our preset threshold, this output x' can be regard as the input x in the next iteration. Otherwise, the iteration can be terminated and this x' is the final result for this object function.

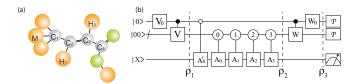All experiments were carried out on a Bruker DRX

Figure 2. (a) Molecule structure of 4-qubit sample : Crontonic acid. (b) Quantum circuit for an iteration to realize gradient descent algorithm. $|x\rangle$ denotes the initial state of work system, and ancillary system are $T_1 + 1$ qubits in the $|0\rangle|0\rangle^{T_1}$ state, where$T_1 = 2$ here. The squares represent unitary operations and the circles represent the state of the controlling qubit.

400MHz spectrometer at room temperature. As it is shown in Fig.2(a), a 4-qubit system is demanded, represented by the liquid ^{13}C-labeled Crontonic acid sample dissolved in d-chloroform [28]. Four carbon-13 nuclei spins(^{13}C) are denoted as four quantum bit. C_i(i=1..3) work as the ancillary system while C_4 are considered as the working system. Their internal Hamiltonian dominates the free evolution of this 4-qubit system:

$$\mathcal{H}_{int} = \sum_{j=1}^{4} \pi\nu_j\sigma_z^j + \sum_{j<k,=1}^{4} \frac{\pi}{2} J_{jk}\sigma_z^j\sigma_z^k, \tag{11}$$

where ν_j and J_{jk} are the resonance frequency of the *j*th spin and the *J*-coupling strength between spins *j* and *k*, respectively. Details of all parameterscan be found in Appendix.B[19]. In order to alter the evolution of the system, as the control field, the radio-frequency(r.f) pulses were introduced:

$$\mathcal{H}_{rf} =$$
$$-\frac{1}{2}\omega_1 \sum_{i=1}^{4}(\cos(\omega_{rf}t + \phi)\sigma_x^i + \sin(\omega_{rf}t + \phi)\sigma_y^i) \tag{12}$$

Among them, duration, intensity ω_1, phase ϕ and frequency ω_{rf} are four tunable parameters. By combining internal system (Eq.31) and r.f control field (Eq.12), the four-qubit universal quantum gates are theoretically achievable.

In the experiment, two points X_0^{S1} $(-0.38, 0.92)$ and X_0^{S2} $(0.86, 0.50)$ are considered as the initial guess of two set of experiment S1 and S2. Several iterations later, as intended, they will all converged to a point $X_{opt}(0.50, 0.86)$. We depict our experimental realization of Fig.2 as following:

Initialization — At room temperature, the 4-qubit quantum system is in the thermal equilibrium state. With the spatial average method[29], we drove this thermal equilibrium system to a pseudo-pure state(PPS). To measure its fidelity, we did a tomography and the result is about 99.01%. More information about how to prepare the PPS can be found in Appendix. B [19]. Then, $|0\rangle_1|00\rangle_{23}|x\rangle_4$ was prepared from this PPS, with simply a single-qubit rotation on 4th qubit, where x is either initial guess or new current point of last iteration x'.

Implementing gradient descent — As Eq.7 shows, the gradient descent operation is realized by evolving the system with

the D operator, where $D = \sum_{m=1}^{4} c_m \mathbf{A}_m$. In our experiment, to guarantee the experimental precision, we skipped the step to experimentally obtain c_m and got them by the theoretical simulation[19]. First three steps of *Soul* are required to achieve an iteration of the optimization algorithm. (S) Drive the system to the specific superposed state $\rho_1 = |\psi_s\rangle|x\rangle\langle x|\langle\psi_s|$. As the circuit in Fig.2(b) shows, this was realized by a combination of single-qubit rotation V_0 and control-V gate. Gradient ascent pulse engineering (GRAPE) was employed here, to generated a 20ms optimized pulse with the simulated fidelity over 99.9% [30]. (O) A combination of the addition or subtraction of the unitary operators A_i was implemented. It is worth mentioning that operator $A_0^{\dagger}$ is demanded for the sigh of gradient value. In circuit (Fig.2), the process is implemented from ρ_1 to ρ_2 and is also optimized with a 30ms GRAPE pulse, with simulated fidelity over 99.9%. (U) An uncompute operation was implemented afterwards, which is just an inverse operation of the first step and also implemented by an 20ms pulses with simulated fidelity of 99.9%. More information about operation V and W can be found in Appendix.B[19].

Measuring and labeling — Since only the state in subspace of $|0\rangle_1|00\rangle_{23}$ is necessary for obtaining the new iteration point x', a full tomography in such subspace was employed. All read out pulses are 0.9ms with 99.8% simulated fidelity. For the sake of experimental errors, mixed states were led in our results, however, the 2-dimension vector x' should be a pure real state. Hence, a purification step was added after this measurement and it is realized with the method of maximum likelihood[31]. (L) As the consequence of the output x', $|\phi_i^{S1}\rangle$ and $|\phi_i^{S2}\rangle$ (i=1...4) were found to be the closest to our experimental density matrices for two different case S1 and S2. According to pre-set threshold, this output x' can be labeled as the new current x to run the next iterative cycle or be the final result and the iteration terminated. On the other hand, to check the performance for the circuit, a 4-qubit tomography was also implemented here and 4-qubit state ρ_3 was obtained. For those 4-qubit states, they have the average of 94% fidelity and the detailed information are shown in Appendix.B[19].

For both case S1 and S2, convergence appears after 4 time iterations from the initial point $X_0^{S1}(-0.38, 0.92)$ and X_0^{S2} (0.86, 0.50). The measured results are shown in Fig.3. $|\phi_i^{S1}\rangle$ and $|\phi_i^{S2}\rangle$(i=1...4), as the outcome of each iteration, are plotted in the subfigure (a) and (b) respectively. On the other hand, theoretical simulation are provided as the comparison, whose inputs were chosen as the measurements of last experimental iteration. In addition, the analytical analyses can be given, simply by substituting $x_1 = \cos(\theta)$ and $x_2 = \sin(\theta)$, the object function is represented as

$$f(x) = -2\sin^3\theta\cos\theta$$

and reduced into one-dimensional unconstraint optimization problem, where local minimum lies at $\theta = 0$, $\pi/3$. Among them, $\theta = 0$ is unstable point while $\pi/3$ is stable local minimum. To make the results clearly, not only iteration points but also function value are shown in Fig.3(c). We alternate
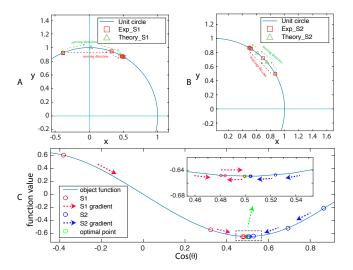

Figure 3. Theoretical simulation and experimental results: (a) and (b) shows the output x' in the iteration process using its two vector elements. ie. $|\phi_i^{S1}\rangle$ and $|\phi_i^{S2}\rangle$(i=1...4). Green triangles are the theoretical simulation results while red squares are experimental measured outputs. ps. they both begin with a same initial point X_0. In addition, the moving direction are also labeled by the dashed arrows coloured green or red. (c) is the 1-d depiction. Beginning with two initial points, for S1(coloured red) and S2(coloured blue), the iteration outputs become lower and lower, until slipping into the neighbour of local minimum. The dashed arrows shows the moving direction for each iteration and in the zoom-in figure, it shows they gradually converging to the optimal minimum point were x (or $cos\theta$) = 0.5.
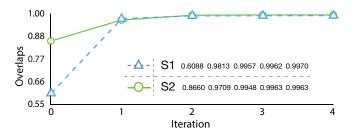

Figure 4. Overlaps of each iteration: S1 and S2 are represented respectively with triangles coloured with blue dashed line and circles colored with green solid line. In total, five overlaps are calculated as the first is the overlaps of initial guess. In subtable, the specific value of overlaps for each iteration are shown.

it into a 1-d depiction. Beginning with two initial points $\cos(\theta) = -0.38$(S1 and coloured red) and $\cos(\theta) = -0.86$ (S2 and coloured blue), respectively, the iteration outputs become lower and lower, until slipping into the neighbour of local minimum. To shows this convergence clearly, in Fig.4, relations between iteration times and overlaps were given. The value of vertical axis was defined as the overlaps between the optimal state and the output state after each iteration: $\langle\phi_o|\phi_i^j\rangle$ (i=0,1..4. j=S1,S2). As the horizontal axis is the number of iteration. It shows that the overlaps converges to 1 no matter the initial point is X_0^{S1} or X_0^{S2}. The specific value of the overlaps are shown in its subtable.

Application— Multidimensional scaling (MDS) is a technique, providing a visual representation of the pattern of proximities in a dataset. It is a common method of statistical analysis in sociology, quantitative psychology, marketing and so on. We apply our method to quantize a algorithm for fitting the simplest of multidimensional scaling models in majority applications.

Given a matrix $A = \delta_{ij}$ that is nonnegative symmetric with zero diagonal. A set of number δ_{ij} is the data collected in a classical multidimensional scaling problem and δ_{ij} is the dissimilarity between objects i and j. Representing n objects as n points via ignoring the objects size, the dissimilarity of objects i and j is approximately equal to the distance between points i and j. The goal is to find n points in m dimensions, denoted by $\mathbf{x}_1, \mathbf{x}_2, \cdot, \mathbf{x}_n$ to form a configuration with coordinates in an $n \times m$ matrix $\mathbf{X}$. When $m = 3$, it is reduced to a molecular conformation problem[32], which plays an important role in chemical and biological fields. Let $d_{ij}(\mathbf{X})$ denotes the Euclidean distances between the points $\mathbf{x}_i$ and $\mathbf{x}_j$. It follows that

$$d_{ij}^2(\mathbf{X}) = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) \tag{13}$$

We minimize the loss function, defined as

$$f(\mathbf{X}) = 1/2 \sum_i \sum_j w_{ij}(d_{ij}(\mathbf{X}) - \delta_{ij})^2 \tag{14}$$

where $W = w_{ij}$ is a symmetric weight matrix can be used to code various supplementary information. The purpose of this algorithm is to find the most suitable information visualization configuration. Now we map it to a quantum version. Firstly, the loss function is rewritten as[33]

$$f(\mathbf{X}) = 1/2 \sum_i \sum_j w_{ij} \delta_{ij}^2 - 2g(\mathbf{X}) + h^2(\mathbf{X}) \tag{15}$$

where

$$g(\mathbf{X}) = 1/2 \sum_i \sum_j w_{ij} \delta_{ij} d_{ij}(\mathbf{X}) \tag{16}$$

and

$$h^2(\mathbf{X}) = 1/2 \sum_i \sum_j w_{ij} d_{ij}^2(\mathbf{X}) \tag{17}$$

Thus, we only need to minimize $f'(\mathbf{X}) = -2g(\mathbf{X}) + h^2(\mathbf{X})$. $g(\mathbf{X})$ and $h^2(\mathbf{X})$ can be further expressed as a trace of some matrixes multiproduction. We have $g(\mathbf{X}) = Tr\mathbf{X}^T B(\mathbf{X})\mathbf{X}$ with $B(\mathbf{X}) = 1/2 \sum_i \sum_j w_{ij} A_{ij} k_{ij}(\mathbf{X})$, where

$$k_{ij}(\mathbf{X}) = 1/d_{ij}(\mathbf{X}), d_{ij}(\mathbf{X}) \neq 0 \tag{18}$$
$$k_{ij}(\mathbf{X}) = 0, otherwise. \tag{19}$$

Similarly, $h(\mathbf{X})^2 = Tr\mathbf{X}^T C(\mathbf{X})\mathbf{X}$ with $C(\mathbf{X}) = 1/2 \sum_i \sum_j w_{ij} A_{ij}$. Then, we have

$$f'(\mathbf{X}) = Tr\mathbf{X}^T D(\mathbf{X})\mathbf{X} \tag{20}$$

where $D(\mathbf{X}) = C(\mathbf{X}) - 2B(\mathbf{X})$. It should be noticed that here $\mathbf{X}$ is a $n \times m$ matrix. In order to represent $\mathbf{X}$ as quantum states, we map it to a sum of the m column vectors $\mathbf{X}_v$ of $\mathbf{X}$. Now, we can apply our quantum gradient algorithm to minimize the objective function

$$f'(\mathbf{X}) = \sum_m Tr\mathbf{X}_v^T D(\mathbf{X})\mathbf{X}_v \tag{21}$$

In this special case, the function order $p = 1$ and $D(x)$ is a symmetric matrix which likely to be decomposed efficiently. It potentially yields an exponential speed up over the classical algorithm in multidimensional scaling problems.

Conclusion — In summary, we have come up with an efficient quantum computational framework to implement the gradient algorithm using quantum machine. Compare with the previous protocol, our method reduce the resource consumption: in each iteration, only two copies of a quantum state are required to produce the next new quantum state. While, in their work, the protocol consumes multiple copies which is linearly depending the order of objective function. Besides, in our method, the complexity of required operations is polylog-arithmical to the dimension of space. Compared with classical algorithms, our algorithm achieves an exponential speedup in high-dimensional cases whose have simple decompositions of coefficient matrices. Moreover, we experimentally implement our protocol on an 4-qubit NMR quantum processor. Beginning with either initial point X_0^{S1} or X_0^{S2}, we optimized the objective function iteratively, final finding the point in the vicinity of the local minimum. Each iterative gradient descent algorithm was implemented with the circuit shown in Fig.2(b). Our experimental procedure and result, in particular the observation results of each iteration, demonstrates the practicability of the protocol to implement the optimization process.

Our method also provides potential applications in quantum controlling technology. For example in Eq.3, setting $p = 1$, the objective function reduces to a quadratic optimization problem in the form of $f(x) = x^T A x$. If the coefficient matrix A is restricted to a density matrix, the objective function represents the overlaps between A and xx^T. Thus, we can product a state x closely enough to a density matrix A by finding the maximum of $f(x)$. It can be used as a quantum method to prepare the specific state. In a word, this optimization approach could be exceptionally useful in a large system to achieve classical-unfriendly optimization problem, and is readily transferrable to other systems such as superconducting circuits ion trap quantum system, being an subroutine for future practical large-scale quantum computer.

* These authors contributed equally to this work.
† wangxiaoting528@gmail.com
‡ gllong@tsinghua.edu.cn

[1] Pedro Domingos. A few useful things to know about machine learning. *Commun. ACM*, 55(10):78–87, October 2012.

[2] Aurlien Gron. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., 1st edition, 2017.

[3] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, New York, NY, USA, 10th edition, 2011.

[4] P. W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, SFCS '94, pages 124–134, Washington, DC, USA, 1994. IEEE Computer Society.

[5] L. K. Grover. A fast quantum mechanical algorithm for database search. *eprint arXiv:quant-ph/9605043*, May 1996.

[6] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature (London)*, 549:242–246, September 2017.

[7] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin. Probing many-body dynamics on a 51-atom quantum simulator. *Nature (London)*, 551:579–584, November 2017.

[8] J. Zhang, G. Pagano, P. W. Hess, A. Kyprianidis, P. Becker, H. Kaplan, A. V. Gorshkov, Z.-X. Gong, and C. Monroe. Observation of a many-body dynamical phase transition with a 53-qubit quantum simulator. *Nature (London)*, 551:601–604, November 2017.

[9] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd. Quantum machine learning. *Nature (London)*, 549:195–202, September 2017.

[10] H.-K. Lau, R. Pooser, G. Siopsis, and C. Weedbrook. Quantum Machine Learning over Infinite Dimensions. *Physical Review Letters*, 118(8):080501, February 2017.

[11] D. Lu, K. Li, J. Li, H. Katiyar, A. J. Park, G. Feng, T. Xin, H. Li, G. Long, A. Brodutch, J. Baugh, B. Zeng, and R. Laflamme. Enhancing quantum control by bootstrapping a quantum processor of 12 qubits. *npj Quantum Information*, 3:45, October 2017.

[12] Jun Li, Xiaodong Yang, Xinhua Peng, and Chang-Pu Sun. Hybrid quantum-classical approach to quantum optimal control. *Phys. Rev. Lett.*, 118:150503, Apr 2017.

[13] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.*, 103:150502, Oct 2009.

[14] B. D. Clader, B. C. Jacobs, and C. R. Sprouse. Preconditioned quantum linear system algorithm. *Phys. Rev. Lett.*, 110:250504, Jun 2013.

[15] Nathan Wiebe, Daniel Braun, and Seth Lloyd. Quantum algorithm for data fitting. *Phys. Rev. Lett.*, 109:050505, Aug 2012.

[16] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. Quantum support vector machine for big data classification. *Phys. Rev. Lett.*, 113:130503, Sep 2014.

[17] S. Lloyd, M. Mohseni, and P. Rebentrost. Quantum principal component analysis. *Nature Physics*, 10:631–633, September 2014.

[18] P. Rebentrost, M. Schuld, L. Wossnig, F. Petruccione, and S. Lloyd. Quantum gradient descent and Newton's method for constrained polynomial optimization. *ArXiv e-prints*, December 2016.

[19] *See Supplemental Information for a detailed description of the theory and experiment.*

[20] Simai He, Zhening Li, and Shuzhong Zhang. Approximation algorithms for homogeneous polynomial optimization with quadratic constraints. *Mathematical Programming*, 125(2):353–383, Oct 2010.

[21] Long Gui-Lu. General quantum interference principle and duality computer. *Communications in Theoretical Physics*, 45(5):825, 2006.

[22] Gui Lu Long. Mathematical theory of the duality computer in the density matrix formalism. *Quantum Information Processing*, 6(1):49–54, Feb 2007.

[23] Shi-Jie Wei, Dong Ruan, and Gui-Lu Long. Duality quantum algorithm efficiently simulates open quantum systems. *Scientific Reports*, 6:30727 EP –, 07 2016.

[24] Shi-Jie Wei and Gui-Lu Long. Duality quantum computer and the efficient quantum simulations. *Quantum Information Processing*, 15(3):1189–1212, Mar 2016.

[25] D. W. Berry, A. M. Childs, R. Cleve, R. Kothari, and R. D. Somma. Simulating Hamiltonian Dynamics with a Truncated Taylor Series. *Physical Review Letters*, 114(9):090502, March 2015.

[26] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum Random Access Memory. *Physical Review Letters*, 100(16):160501, April 2008.

[27] Tao Xin, Shi Jie Wei, Julen S. Pedernales, Enrique Solano, and Gui Lu Long. Quantum simulation of quantum channels in nuclear magnetic resonance. *Phys.rev.a*, 96(6), 2017.

[28] Keren Li, Youning Li, Muxin Han, Sirui Lu, Jie Zhou, Dong Ruan, Guilu Long, Yidun Wan, Dawei lu, Bei Zeng, and Raymond Laflamme. Quantum spacetime on a quantum simulator. 12 2017.

[29] David G. Cory, Amr F. Fahmy, and Timothy F. Havel. Ensemble quantum computing by nmr spectroscopy. *Proceedings of the National Academy of Sciences*, 94(5):1634–1639, 1997.

[30] Navin Khaneja, Timo Reiss, Cindie Kehlet, Thomas Schulte-Herbrüggen, and Steffen J Glaser. Optimal control of coupled spin dynamics: design of nmr pulse sequences by gradient ascent algorithms. *Journal of Magnetic Resonance*, 172(2):296–305, 2005.

[31] J Řeháček, Z Hradil, and M Ježek. Iterative algorithm for reconstruction of entangled states. *Phys. Rev. A*, 63(4):040303, 2001.

[32] W Glunt, T L Hayden, and M Raydan. Molecular conformations from distance matrices. *Journal of Computational Chemistry*, 14(1):114–120, 2010.

[33] Jan De Leeuw. Convergence of the majorization method for multidimensional scaling. *Journal of Classification*, 5(2):163–180, 1988.

Supplementary information for "Optimizing a polynomial function in quantum simulator"

A. protocol details

Now, we discuss the details of our method. Carry on the Equation.22, defining the $\mathbf{D}$ operator

$$\mathbf{D} =: \sum_{\alpha} \sum_{j} \left(\prod_{i \neq j} \mathbf{X}^T \mathbf{A}_i^{\alpha} \mathbf{X} \right) \mathbf{A}_j^{\alpha} \tag{22}$$

define $Z = \mathbf{X}\mathbf{X}^T$, $b_j^{\alpha} = \mathbf{X}^T \mathbf{A}_j^{\alpha} \mathbf{X}$ and $Q_j^{\alpha} = \mathbf{A}_1^{\alpha} Z \mathbf{A}_2^{\alpha} Z \cdots \mathbf{A}_p^{\alpha}$. Then,

$$M^{\alpha} = \prod_i \mathbf{X}^T \mathbf{A}_i^{\alpha} \mathbf{X} = \mathbf{X}^T Q^{\alpha} \mathbf{X}$$

The operator $\mathbf{D}$ can be rewritten as

$$\mathbf{D} = \sum_{\alpha} \sum_{j} \left(\prod_i \mathbf{X}^T \mathbf{A}_i^{\alpha} \mathbf{X} \right) \frac{\mathbf{A}_j^{\alpha}}{\mathbf{X}^T \mathbf{A}_j^{\alpha} \mathbf{X}} \tag{23}$$

$$= \sum_{\alpha} \sum_{j} M^{\alpha} \frac{\mathbf{A}_j^{\alpha}}{b_j^{\alpha}} \tag{24}$$

$$= \sum_{m=1}^{Kp} c_m \mathbf{A}_m \tag{25}$$

where $c_m = M^{\alpha}/b_j^{\alpha}$.

The method to obtain the expected value M^{α} of Q_j^{α} and expected value b_j^{α} of $\mathbf{A}_j^{\alpha}$ are as follows.
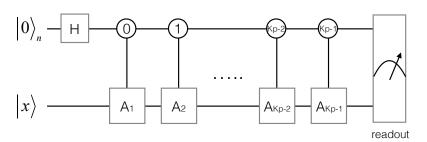

Figure 5. Quantum circuit for obtaining the expected value b_j^{α}. $|x\rangle$ denotes the initial state of work system, and ancillary system consists of n qubits in the $\|0\rangle_n$ state, where $n = T_1$. The squares represent unitary operations and the circles represent the state of the controlling qubit.

In Fig.5, the unitary matrix H represents the tenser products of T_1 number Hadamard gates. The process can be expressed as:

$$|0\rangle^{T_1} |X\rangle \rightarrow H |0\rangle^{T_1} |X\rangle$$

$$\rightarrow \frac{1}{\sqrt{2^{T_1}}} (|0\rangle A_1 |X\rangle + |1\rangle A_2 |X\rangle + \cdots).$$

$$= \frac{1}{\sqrt{2^{T_1}}} \left(\sum_{m=1}^{Kp} |m\rangle A_m |X\rangle \right) \tag{26}$$

When the ancillary system in state $|m\rangle$, we measure the work system via $|X\rangle\langle X|$ basis, we can obtain the expected value b_j^{α}. The expected value M^{α} can be calculated by $M^{\alpha} = \prod_i b_j^{\alpha}$.

Now, we go to the iteration part. We conbined the minus sign and the operator $\mathbf{D}$, into a unitary operator $\mathbf{A}_m$. The iteration process can be expressed as

$$|\mathbf{X}^{(t+1)}\rangle = |\mathbf{X}^{(t)}\rangle - \mathbf{D}|\mathbf{X}^{(t)}\rangle \tag{27}$$

$$= |\mathbf{X}^{(t)}\rangle + \sum_{m=0}^{Kp-1} c_m \mathbf{A}_m |\mathbf{X}^{(t)}\rangle.$$

We can prepare the following initial state firstly

$$|0\rangle|0\rangle^{T_1}|0\rangle \rightarrow (\sqrt{\beta}|0\rangle|0\rangle^{T_1} + \sum_{m=0}^{Kp-1} c_m|1\rangle|m\rangle)|\mathbf{X}^{(t)}\rangle. \tag{28}$$

Then, we perform the first ancillary qubit controlled operation $|0\rangle\langle 0| \otimes A_0^\dagger$ and the second ancillary system controlled operations $|0\rangle\langle 0| \otimes A_0, |1\rangle\langle 1| \otimes A_1, \ldots, |Kp-1\rangle\langle Kp-1| \otimes A_{Kp-1}$ on the work system $|\mathbf{x}^{(t)}\rangle$. This step transforms the initial state into

$$(\sqrt{\beta}|0\rangle|0\rangle^{T_1} + \sum_{m=0}^{Kp-1} c_m|1\rangle|m\rangle)|\mathbf{X}^{(t)}\rangle = \frac{1}{\sqrt{(\sum_{m=0}^{Kp-1} c_m + 1)}}|0\rangle|0\rangle^{T_1}|\mathbf{X}^{(t)}\rangle + \frac{\sqrt{\sum_{m=0}^{Kp-1} c_m}}{\sqrt{(\sum_{m=0}^{Kp-1} c_m + 1)}}|1\rangle|0\rangle^{T_1}|\mathbf{X}^{(t)}\rangle \tag{29}$$

$$\rightarrow \frac{1}{\sqrt{(\sum_{m=0}^{Kp-1} c_m + 1)}}|0\rangle|0\rangle^{T_1}|\mathbf{X}^{(t)}\rangle + \frac{\sqrt{\sum_{m=0}^{Kp-1} c_m}}{\sqrt{(\sum_{m=0}^{Kp} c_m + 1)}}|1\rangle|c_m\rangle A_m|\mathbf{X}^{(t)}\rangle$$

.

Finally, applying the gate sequence of preparing initial state in reverse and measuring the two registers. An output of $|0\rangle|0\rangle^{T_1}$ will result in the following state

$$\frac{1}{\sqrt{(\sum_{m=0}^{Kp-1} c_m + 1)}}|0\rangle|0\rangle^{T_1}|\mathbf{X}^{(t)}\rangle + \frac{\sqrt{\sum_{m=0}^{Kp-1} c_m}}{\sqrt{(\sum_{m=0}^{Kp} c_m + 1)}}|1\rangle|c_m\rangle A_m|\mathbf{X}^{(t)}\rangle \tag{30}$$

$$\rightarrow \frac{1}{\sum_{m=1}^{Kp} c_m + 1}\left(|\mathbf{X}^{(t)}\rangle + \sum_{m=1}^{Kp} c_m \mathbf{A}_m|\mathbf{X}^{(t)}\rangle\right),$$

which is proportional to $\mathbf{X}^{(t+1)}$. When we get the final result, we can multiply $\sum_{m=1}^{Kp} c_m + 1$ to obtain $|\mathbf{X}^{(t+1)}\rangle$.

B.   experimental process

1.   molecule

The demonstration of the whole algorithm were conducted on a four-qubit nuclear magnetic resonance (NMR) system. The four-qubit sample is ^{13}C-labeled crotonic acid dissolved in d6-acetone. Fig. 6 shows the structure of this molecule, where C_1 to C_4 are denoted as four qubits, representing the ancillary and working system. Throughout the entire experiments, M, H_1 and H_2 nuclei in the methyl group were decoupled. With the weak coupling assumption, the internal Hamiltonian of this liquid sample are written as:

$$\mathcal{H} = \sum_{j=1}^{4} \frac{1}{2}\omega_j \sigma_z^j + \sum_{j<k}^{4} \frac{\pi}{2} J_{jk} \sigma_z^j \sigma_z^k, \tag{31}$$

where ν_j is the chemical shift and J_{jk} is the J-coupling strength. All parameters were obtained on this Bruker DRX 400MHz spectrometer at room temperature (296.5K).

	C1	C2	C3	C4
C1	1705.5			
C2	41.64	14558.0		
C3	1.46	69.72	12330.5	
C4	7.04	1.18	72.36	16764.0
T2	0.84	0.92	0.66	0.79

Figure 6. Molecular structure of ^{13}C-labeled crotonic acid. C_1 is denoted as the working system while C_2, C_3 and C_4 are denoted as ancillary qubits. The chemical shifts and J-couplings (in Hz) are listed by the diagonal and off-diagonal elements, respectively. T_2 (in Seconds) are also shown at bottom. All parameters were obtained on this Bruker DRX 400MHz spectrometer at room temperature (296.5K).
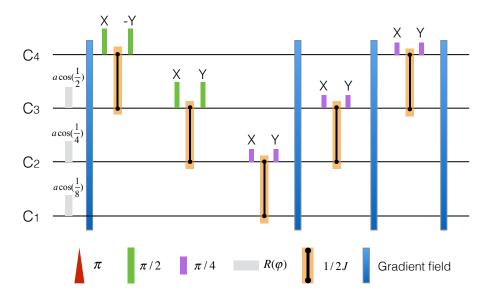

Figure 7. Circuit for creating the 4-qubit pseudo-pure state: via spatial average technique, the entire procedure includes local operations, five J-coupling evolutions, and four z-gradient pulses which is to destroy the unnecessary coherent terms. The time of each free evolutions is several ms, determined by 1/2J, i.e. the J-coupling strengths of the relevant spins.

2. Preparation of the pseudo-pure state

In our four-qubit NMR system, the thermal equilibrium state ρ_{eq} is,

$$\rho_{eq} = \frac{1-\epsilon}{16}\mathbb{I} + \epsilon(\gamma_{C1}\sigma_z^1 + \gamma_{C2}\sigma_z^2 + \gamma_{C3}\sigma_z^3 + \gamma_{C4}\sigma_z^4), \tag{32}$$

where $\epsilon \approx 10^{-5}$ polarization coefficient, $\mathbb{I}$ is a 16×16 identity matrix, and $\gamma_{Ci}(i = 1...4)$ are the gyromagnetic ratios for each carbon nuclei, respectively. As the identity part does not influence the unitary operations or measurements in NMR experiments, the original density matrix ρ_{eq} can be replaced by the deviation matrix, i.e,

$$\rho_{eq} = \sigma_z^1 + \sigma_z^2 + \sigma_z^3 + \sigma_z^4. \tag{33}$$

Our purpose is to create the pseudo-pure state

$$\rho_{0000} = \frac{1-\epsilon}{16}\mathbb{I} + \epsilon|0000\rangle\langle0000|, \tag{34}$$

Spatial average technique[1] was used here, as the pulse sequences are shown in Fig. 7. It includes local rotation and four z-gradient fields for destroying the unnecessary coherent terms. The entire procedure of the pseudo-pure state creation takes about 70ms with the simulated fidelity 99.8%. In experiment, the fidelity is around 99.01%.
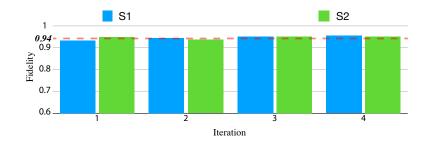
Figure 8. Quantum circuit for realising gradient descent algorithm. $|X\rangle$ denotes the initial state of work system, and ancillary system are $T_1 + 1$ qubits in the $|0\rangle|0\rangle^{T_1}$ state, where $T_1 = log_2(Kp)$. The squares represent unitary operations and the circles represent the state of the controlling qubit.

3. operators in expeimrnt

We depicts the details of the experimental circuits in this section, which implements one iteration to realize gradient descent algorithm. During the experiment, the unitary operators were arranged as follows. V^0 is a sigle qubit rotation and $W^0 = V^{0\dagger}$ and

$$V^0 = \begin{pmatrix} \frac{1}{\sqrt{(\sum_{m=1}^{Kp} c_m + 1)}} & \frac{\sqrt{\sum_{m=1}^{Kp} c_m}}{\sqrt{(\sum_{m=1}^{Kp} c_m + 1)}} \\ \frac{\sqrt{\sum_{m=1}^{Kp} c_m}}{\sqrt{(\sum_{m=1}^{Kp} c_m + 1)}} & -\frac{1}{\sqrt{(\sum_{m=1}^{Kp} c_m + 1)}} \end{pmatrix}. \tag{35}$$

V is a $Kp \times Kp$ unitary matrix, which maps $V|0\rangle^T = \frac{1}{\sqrt{\sum_{m=1}^{Kp} c_m}}(\sqrt{c_1}|0\rangle) + \sqrt{c_2}|1\rangle + ... + \sqrt{c_{Kp}}|Kp - 1\rangle$, where $T = 2$ and $Kp = 4$ in our experiment.. Using the Schmidt orthogonalization, we could find a unitary matrix implementing it. Meanwhile, $W = V^\dagger$. As for the series of A_i, since the decomposition $\widehat{A} = -\sigma_I \otimes \sigma_X + \sigma_X \otimes \sigma_Z$, $A_i(i = 1...4)$ equals to $-\sigma_I, \sigma_X, \sigma_X$ and σ_Z, respectively.

4. the results of each iteration

Each time we measure the output of the gradient descent implement circuit, the 4-qubit tomography was also employed and 4-qubit state ρ_3 was obtained. Theoretical states ρ_{th} are listed by Eq.31, where $|\mathbf{X}^{(t)}\rangle$ is the input state, i.e. the current point. we calculate the fidelity between the theoretical and experimental results[2]:

$$F = \text{tr}(\rho_{\text{th}}\rho_{\text{exp}})/\sqrt{\text{tr}(\rho_{\text{th}}^2)\,\text{tr}(\rho_{\text{exp}}^2)}. \tag{36}$$

Here, we write the states in terms of density matrices because the experimentally prepared state is mixed due to the experimental imperfections. For those 4-qubit states, since the decoherence effect introduced by the long algorithm operation time about 70ms and the inaccurate pulse from arbitrary wave generator, they have the average of 94% fidelity. Fig.8 has shown this consequence.

---

[1] David G. Cory, Amr F. Fahmy, and Timothy F. Havel. Ensemble quantum computing by nmr spectroscopy. *Proceedings of the National Academy of Sciences*, 94(5):1634–1639, 1997.
[2] Evan M. Fortunato, Marco A. Pravia, Nicolas Boulant, Grum Teklemariam, Timothy F. Havel, and David G. Cory. Design of strongly modulating pulses to implement precise effective hamiltonians for quantum information processing. *The Journal of Chemical Physics*, 116(17):7599–7606, 2002.