Quantum regularized least squares solver with parameter estimate

Changpeng Shao¹ · Hua Xiang²,³

Received: 4 October 2019 / Accepted: 8 February 2020 / Published online: 14 February 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract
For ill-conditioned least squares problems, regularization techniques such as Tikhonov regularization are the key ingredients to obtain meaningful solutions, and the determination of the proper regularization parameter is the most difficult step. This paper focuses on the choice of regularization parameter on a quantum computer. We combine the classical L-curve or Hanke–Raus rule with the HHL algorithm and quantum amplitude estimation that compute the regularized solution and the corresponding residual and their norms. When a series of regularization parameters are tested, we then apply Grover’s search algorithm to find the best one that gives the meaningful solution. This yields a quadratic speedup in the number of regularization parameters.

Keywords Tikhonov regularization · Regularization parameter estimate · Quantum algorithm · Grover’s search · L-curve · Hanke–Raus rule

1 Introduction
For the ill-conditioned linear system \(Ax = b\), or the least squares problem (LSP) \(\min_x \|Ax - b\|\), that arises from many areas of scientific computing like the discretization of the inverse problem [1–3], the condition number of \(A\) is large. The solution is sensitive and easily contaminated by the data perturbation, so some regularization techniques are necessary in order to obtain a meaningful solution. Tikhonov regularization [4] is one of the most popular and effective techniques. It converts the original linear system into the following regularized LSP
\[
\min_x \|Ax - b\|^2 + \mu^2 \|x\|^2,
\]

(1)

where the constant \(\mu\) is the so-called regularization parameter. By introducing the regularization parameter \(\mu\), one can make a comprise between the sensitivity of the problem and the perturbation of the measured data and thus greatly reduce the effects caused by the noise data. The Tikhonov regularization also provides the mathematical foundations for machine learning [5]. It can be of the following more general form [6]

\[
\min_x \|Ax - b\|^2 + \mu^2 \|Lx\|^2,
\]

where the matrix \(L\) arises from the discrete approximation to some differential operator, for example the Laplacian or the gradient operator.

The LSP (1) can be solved by the singular value decomposition (SVD). Suppose that we have the SVD of matrix \(A \in \mathbb{R}^{m \times n}(m \geq n)\), and it reads \(A = U \Sigma V^\dagger\), where \(U = (u_1, \ldots, u_m)\), \(V = (v_1, \ldots, v_n)\) are unitary matrices, and \(\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n) \in \mathbb{R}^{m \times n}\) with \(\sigma_i (i = 1, 2, \ldots, n)\) being the singular values. Then the solution of LSP (1), i.e., the Tikhonov regularized solution \(x_\mu\), can be expressed as

\[
x_\mu = \sum_{i=1}^{n} f_i \frac{u_i^\dagger b}{\sigma_i} v_i,
\]

(2)

where \(f_i = \sigma_i^2 / (\sigma_i^2 + \mu^2)\) is the Tikhonov filter factor [3].

The success of regularization methods highly depends on the right choice of regularization parameter \(\mu\), which is our focus of this paper. There are several popular techniques in the literature to determine effective regularization parameters. When the noise level is unknown, we may use some heuristic methods, such as the L-curve method [7,8], the Hanke–Raus rule [9], the generalized cross-validation (GCV) function [10] and the quasi-optimality criterion. When the noise level is known, the discrepancy principle, the monotone error rule, and the balancing principle can be applied (see [11] and the references therein).

The regularization parameter estimate is critical and time-consuming classically. In this paper, we will apply quantum linear algebraic techniques to accelerate the estimation of regularization parameter. We focus on two heuristic methods: the L-curve and the Hanke–Raus rule. The main difficulty of these two methods lies on the estimations of the solution norm \(\|x_\mu\|\) and the residual norm \(\|Ax_\mu - b\|\). On a quantum computer, under certain conditions, we can solve LSP (1) exponentially faster than a classical computer, such as by the HHL algorithm [12]. Then by the amplitude estimation technique [13], we can efficiently estimate \(\|x_\mu\|\) as well as \(\|Ax_\mu - b\|\) within a given precision. Applying the HHL algorithm to LSP (1) needs the same assumptions as those for solving \(Ax = b\); however, the regularized LSP has a smaller condition number. To find the best regularization parameter, a number of parameters need to be tested. That is, a series of the estimations of \(\|x_{\mu_j}\|\) and \(\|Ax_{\mu_j} - b\|\) (\(j = 1, \ldots, p\)) should be performed. Due to quantum parallelism feature, we can implement this procedure in parallel. Together with the quantum minimum finding algorithm [14], we can achieve quadratic speedup at the number of tested parameters. Roughly, the

\(\text{Springer}\)
quantum algorithm costs $O(\kappa^2 \sqrt{p} (\log n)/\epsilon^2)$, while the classical methods need at least $O(\sqrt{\kappa} p n \log 1/\epsilon)$.

1.1 Related works

Quantum algorithm to solve LSP was first studied in [15] in the sparse case based on the HHL algorithm. Extension to the non-sparse case (low-rank assumption maybe needed) was given in [16] by using the quantum principal component analysis. In [17,18], quantum algorithms to weighted LSP were studied, which were based on block-encoding and singular value estimation, respectively. The above results only return quantum state of the optimal solution of the LSP. In [19], Wang proposed a quantum algorithm that return the optimal parameters in the classical form. Solving the regularized LSP with a fixed regularization parameter $\mu$ has been considered in [20,21], based on the SVD of $A$. About the determination of the regularization parameter, in [22] Yu et al. proposed the quantum algorithm based on $K$-fold cross-validation. The complexity is $O(\kappa^5 p \|A\|_{\max}^2 (\poly \log n)/\epsilon^4)$. In comparison, we use a different strategy and our quantum algorithm has better dependence on the parameters.

The rest of this paper is organized as follows. In Sect. 2, we introduce the amplitude estimation and its generalization, which will be used for the norm estimation in the determination of regularization parameter. In Sect. 3, we consider the quantum solver based on HHL for the regularized least squares problem with a fixed regularization parameter. It is a preparation for the quantum algorithm with variable regularization parameters. We propose two quantum algorithms to speed up the regularization parameter estimate in Sect. 4. Some discussions and concluding remarks are given in Sects. 5 and 6, respectively.

2 Amplitude estimation and its generalization

Let $|\phi\rangle = \cos \theta |0\rangle |u\rangle + \sin \theta |1\rangle |v\rangle$ be a quantum state that can be prepared in time $O(T)$. Using quantum phase estimation (QPE) and Grover’s algorithm, we can obtain $\epsilon$-approximations of $\pm \theta$ on a quantum computer with high success probability close to 1 in time $O(T/\epsilon)$. But the query complexity is $O(1/\epsilon)$. As a result, we can estimate the probabilities $|\sin \theta|^2$ and $|\cos \theta|^2$ efficiently. This method is known as amplitude estimation [13]. It plays an important role in this paper, so in the following, we first briefly review this method and then do some modifications for further applications.

Assume that there is a unitary operator $U$ with implementation time $O(T)$ such that $|\phi\rangle = U|0\rangle^\otimes k$. Let $Z$ be the 2-dimensional Pauli-Z matrix that maps $|0\rangle$ to $|0\rangle$ and $|1\rangle$ to $-|1\rangle$. Denote

$$G = (2|\phi\rangle \langle \phi| - I)(Z \otimes I) = U (2|0\rangle^\otimes k \langle 0|^\otimes k - I) U^\dagger (Z \otimes I),$$

(3)
which is similar to the rotation used in Grover’s algorithm. We can check that

\[
G = \begin{pmatrix}
\cos 2\theta & -\sin 2\theta \\
\sin 2\theta & \cos 2\theta
\end{pmatrix}
\]

in the space spanned by \{\langle 0|u\rangle, \langle 1|v\rangle\}. The eigenvalues of \(G\) are \(e^{\pm i2\theta}\) and the corresponding eigenvectors are \(|w_\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm i|1\rangle)|v\rangle\).

To apply QPE, we choose the initial state as \(|\phi\rangle|0\rangle^{\otimes l}\), where \(l = O(\log 1/\delta\epsilon)\), and \(\delta\) is the failure probability [23]. It can be rewritten as

\[
|\phi\rangle|0\rangle^{\otimes l} = \frac{1}{\sqrt{2}}(e^{i\theta}|w_-\rangle + e^{-i\theta}|w_+\rangle)|0\rangle^{\otimes l}.
\]

By QPE, with probability close to \(1 - \delta\), we will obtain the following state

\[
\frac{1}{\sqrt{2}}(e^{i\theta}|w_-\rangle|y\rangle + e^{-i\theta}|w_+\rangle| - y\rangle)
\]

in time \(O(T/\epsilon \delta)\), where \(y \in \mathbb{Z}_{2^l}\) satisfies \(|\theta - y\pi/2^l| \leq \epsilon\). Here we neglect the circuit complexity, which equals \(O(l^2) = O((\log 1/\delta\epsilon)^2)\), to implement the Hadamard transformation and the quantum Fourier transform in QPE. Performing a measurement on (4), we can get an \(\epsilon\)-approximation of \(\theta\) or \(-\theta\) with probability at least \(1 - \delta\). From the approximation of \(\pm \theta\), we can estimate the probabilities \(|\sin \theta|^2\) and \(|\cos \theta|^2\) efficiently. Generally, \(\delta\) is chosen as a small constant and thus can be ignored in the complexity analysis. Figure 1 shows the quantum circuit of amplitude estimation, and its complexity is mainly determined by the circuit complexity for implementing \(G\). Also from this circuit, it is easy to determine the query complexity of this algorithm, which equals the number of \(G\) used in the circuit. Thus the query complexity is \(O(2^l) = O(1/\epsilon)\). Since amplitude estimation is an application of QPE, for further reference we denote Fig. 1 by QPE\(_G\).

To further apply the information of \(\theta\) to solve other problems, such as the finding of the best regularization parameter studied in this paper, instead of performing measurements in the state (4) to get \(\tilde{\theta} = y\pi/2^l\), it is more useful to generate the following quantum state

\[
|\phi\rangle|g(\tilde{\theta})\rangle,
\]

\(\tilde{\theta} \in \mathbb{R}\).
where $g$ is an even function such that $U_g : |x, y⟩ \mapsto |x, y \oplus g(x)⟩$ is efficiently implemented. The quantum state (5) is obtained by adding a register to store $g(\tilde{\theta})$ and undoing the QPE to (4). If $g$ is an elementary function, such as linear or cosine function, then $U_g$ can be efficiently implemented on a quantum computer [24]. All the functions we encountered in this paper are elementary, so in the following, we always assume that $U_g$ is available. Consequently, we have

**Proposition 1** Let $|\phi⟩ = \cos \theta|0⟩|u⟩ + \sin \theta|1⟩|v⟩$ be a quantum state that can be prepared in time $O(T)$. Let $f$ be a univariate function, then the following unitary transformation

$$|\phi⟩|0⟩ \mapsto |\phi⟩ \left| f\left(\cos \tilde{\theta}\right)\right⟩$$

(6)

can be achieved in time $O(T/\epsilon)$, where $|\theta − \tilde{\theta}| \leq \epsilon$. The query complexity is $O(1/\epsilon)$.

Based on Fig. 1 to implement QPE, we can implement procedure (6) as Fig. 2. It is a quantum algorithm to evaluate $f$ on the amplitude of $|0⟩$ of $|\phi⟩$ and denoted by $\text{QEval}_{|\phi⟩, f}$ for short.

The unitary procedure (6) can be implemented in parallel due to quantum superposition and parallelism as the following corollary states.

**Corollary 1** Let $|\phi_j⟩ = \cos \theta_j|0⟩|u⟩ + \sin \theta_j|1⟩|v⟩$, ($j = 1, \ldots, p$) be $p$ quantum states that are prepared in time $O(T)$. Let $\sum_{j=1}^{p} \alpha_j |j⟩$ be a given quantum state and $f$ be a univariate function. Then the following unitary transformation

$$\sum_{j=1}^{p} \alpha_j |j⟩|\phi_j⟩|0⟩ \mapsto \sum_{j=1}^{p} \alpha_j |j⟩|\phi_j⟩ \left| f\left(\cos \tilde{\theta}_j\right)\right⟩$$

(7)

can be achieved in time $O(T/\epsilon)$, where $|\theta_j − \tilde{\theta}_j| \leq \epsilon$. The query complexity is $O(1/\epsilon)$.

Combining Corollary 1 and the quantum minimum finding algorithm (QMF) [14], if $\alpha_j = 1/\sqrt{p}$ for all $j$ in (7), then we can find $j_0 = \arg \min_j f(\cos \theta_j)$ with gate complexity $O\left(\sqrt{p}T/\epsilon\right)$ and query complexity $O\left(\sqrt{p}/\epsilon\right)$. This result will be applied later to find the best regularization parameter with a quadratic speedup at $p$ over the classical methods. The following QMF algorithm is a direct modification of [14]. It can help to better understand the construction of quantum algorithms of finding the best regularization parameter.
Algorithm 1 Quantum minimum finding (QMF) [14]

1: Choose a threshold index $1 \leq k \leq p$ by performing a measurement on (7).
2: Repeat until the total running time is more than $22.5 \sqrt{p} + 1.4(\log p)^2$.
   (a) Initialize the memory as
   $$\frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle|\phi_j\rangle|f(\cos \theta_j)\rangle.$$  
   Mark every item $j$ for which $f(\cos \theta_j) < f(\cos \theta_k)$.
   (b) Apply the amplitude amplification to improve the probability of marked items.
   (c) Observe the first register: let $k'$ be the outcome. If $f(\cos \theta_{k'}) < f(\cos \theta_k)$, then set $k'$ to be the new threshold index $k$.
3: Return the index $k$.

3 Quantum regularized LSP algorithm

For the LSP, we need a quantum solver for a system of linear equations, such as the HHL algorithm [12], the SVE [25], the blocked-encoding framework [17] and other extensions [26–28]. All these methods are affected by the condition number of the underlying linear system. Since solving linear system is not the main concern of this paper, we only focus on using the HHL algorithm to solve linear systems. Other approaches may be applied to improve the efficiency of our quantum algorithms under different assumptions.

The HHL algorithm solves the linear system $Ax = b$, or equivalently $\begin{pmatrix} 0 & A \\ A^\dagger 0 \end{pmatrix}$, and outputs the quantum state $|x\rangle$ proportional to the solution. We can get the expectation value of a certain operator $M$ associated with $x$, i.e., $x^\dagger M x$, by swap test [29]. Similarly, given another quantum state $|c\rangle$, we can obtain an estimation of $\langle x | c \rangle$.

It is easy to verify that the LSP (1) has the following equivalent form

$$\min_x \left\| \begin{pmatrix} A \\ \mu I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|^2.$$  

Our LSP solver starts from the SVD of the extended matrix $A_\mu := \begin{pmatrix} A \\ \mu I \end{pmatrix}$. Denote the condition number of $A$ and $A_\mu$ as $\kappa$ and $\kappa_\mu$, respectively. Using the SVD of $A$, we can easily derive that the eigenvalues of $A_\mu^\dagger A_\mu$ are $\sigma_i^2 + \mu^2$ ($i = 1, \ldots, n$), so
the singular values of $A_\mu$ are $\sqrt{\sigma_i^2 + \mu^2}$ ($i = 1, \ldots, n$). Moreover, we can compute a complete SVD of $A_\mu$ as follows. By the SVD of $A$, we have
\[
\begin{pmatrix}
  A \\
  \mu I
\end{pmatrix} = \begin{pmatrix}
  U & 0 \\
  0 & V
\end{pmatrix} \begin{pmatrix}
  \Sigma \\
  \mu I
\end{pmatrix} V^\dagger.
\]

Applying the Givens rotations to $\begin{pmatrix}
  \Sigma \\
  \mu I
\end{pmatrix}$, we obtain the singular values of $A_\mu$: $\sqrt{\sigma_i^2 + \mu^2}$ ($i = 1, \ldots, n$) and can also derive the singular vectors. We will not go into details on this point here.

Let $\sigma_{\text{max}}, \sigma_{\text{min}}$ be the largest and the smallest nonzero singular value of $A$, respectively. If $A$ is of full column rank, then $\sigma_n = \sigma_{\text{min}} \neq 0$, and
\[
\kappa_\mu = \sqrt{\frac{\sigma_{\text{max}}^2 + \mu^2}{\sigma_n^2 + \mu^2}} = \sqrt{\frac{\kappa^2 + (\mu/\sigma_{\text{min}})^2}{1 + (\mu/\sigma_{\text{min}})^2}}. 
\]
For the ill-conditioned case where $\kappa \gg 1$, the regularization parameter $\mu$ is chosen such that $\mu \gg \sigma_{\text{min}}$. Then generally we have $\kappa_\mu \ll \kappa$. So we may have a much smaller condition number for solving the LSP (1) based on $A_\mu$, rather than that based on $A$.

The minimization problem (8) can be solved by applying HHL algorithm to the following linear system
\[
\begin{pmatrix}
  0 & A_\mu \\
  A_\mu^\dagger & 0
\end{pmatrix} \begin{pmatrix}
  0 \\
  x
\end{pmatrix} = \begin{pmatrix}
  b \\
  0
\end{pmatrix}. 
\]

Denote $\tilde{A}_\mu = \begin{pmatrix}
  0 & A_\mu \\
  A_\mu^\dagger & 0
\end{pmatrix}$ and $\tilde{b} = \begin{pmatrix}
  b \\
  0
\end{pmatrix}$. Since
\[
\tilde{A}_\mu = \begin{pmatrix}
  0 & 0 & 0 & \mu I \\
  0 & 0 & A & 0 \\
  A_\mu^\dagger & 0 & 0 & 0 \\
  A_\mu^\dagger & 0 & 0 & \mu I
\end{pmatrix} = \begin{pmatrix}
  0 & 0 & 0 \\
  0 & 0 & A \\
  0 & 0 & \mu I \\
  0 & 0 & \mu I
\end{pmatrix}
\]
the Hamiltonian simulation of $\tilde{A}_\mu$ is equivalently efficient as $\tilde{A} = \begin{pmatrix}
  0 & A \\
  A^\dagger & 0
\end{pmatrix}$ in HHL algorithm [30]. So the assumption about Hamiltonian simulation to solve (11) is the same as HHL algorithm. In the following, we simply assume that the Hamiltonian simulation is efficient.

Suppose the SVD of $\tilde{A}_\mu = \sum_j \tilde{\sigma}_j |\tilde{u}_j \rangle \langle \tilde{u}_j |$, and the smallest nonzero singular value is represented by $\tilde{\sigma}_{\text{min}}$. We formally decompose $|\tilde{b}\rangle = \sum_j \tilde{\beta}_j |\tilde{u}_j \rangle$. Then using a similar
procedure to HHL algorithm (see appendix A), we obtain

$$|X_\mu\rangle = \sum_j \tilde{\beta}_j |\tilde{u}_j\rangle \left[ \tilde{C}\tilde{\sigma}_{j}^{-1}|0\rangle + \sqrt{1-\tilde{C}^2\tilde{\sigma}_{j}^{-2}}|1\rangle \right]$$

$$= \tilde{C}\|x_\mu\||x_\mu\rangle|0\rangle + P_1|\phi_1\rangle|1\rangle,$$

(12)

where $\tilde{C}$ is a lower bound of $\tilde{\sigma}_{\text{min}}$ and $|\phi_1\rangle$ is the state proportional to $\sum_j \tilde{\beta}_j \sqrt{1-\tilde{C}^2\tilde{\sigma}_{j}^{-2}}|\tilde{u}_j\rangle$ with amplitude $P_1$.

When solving linear systems on a quantum computer, the query and gate complexity differ on a logarithm term [26]. More precisely, let $O_A$ and $O_b$ be the oracles to query the matrix $A$ and the vector $b$. More precisely, if the $(j, k)$th entry of $A$ is $a_{ij}$, the $j$th entry of $b$ is $b_j$, then $O_A : |j, k, z\rangle \mapsto |j, k, z \oplus a_{jk}\rangle$ and $O_b : |j, z\rangle \mapsto |j, z \oplus b_j\rangle$. The oracle $O_A$ also performs: $O_A : |j, l\rangle \mapsto |j, \nu(j, l)\rangle$, where $\nu(j, l)$ computes the row index of the $l$th nonzero entry of the $j$th column of $A$. With $O_A$, it is easy to build an oracle to query the entries of $A_\mu$. Then the query complexity to obtain state (12) is $O(\kappa_\mu/\epsilon)$, while the gate complexity is $O(\kappa_\mu(\log n)/\epsilon)$. The derivation of the complexities is the same as the HHL algorithm. The HHL algorithm has a quadratic dependence on the condition number, which arises from two resources. One is from estimating $\tilde{\sigma}_{j}^{-1}$, and the other one is from the success probability. However, we do not perform any measurement here, so the complexity is linear in the condition number. We will see in the next section that the quantum state (12) is needed for locating the best regularization parameter.

To better understand the combination of HHL algorithm and amplitude estimation in the next section, in the end of this section we plot the quantum circuit of HHL algorithm. The HHL algorithm is an application of quantum phase estimation together with a control rotation. The circuit of quantum phase estimation is depicted in Fig. 1. In the HHL algorithm, $G = \exp(-i\tilde{t}_0 A_\mu)$, where $t_0 = O(\kappa_\mu/\epsilon)$ is chosen to suppress the singular values of $t_0 A_\mu$ smaller than 1, and approximate the inverse of singular values to precision $\epsilon$. As for the control rotation, assume that in binary form $\theta = \arccos(\tilde{C}/\tilde{\sigma}_j) \approx 2\pi \sum_{k=0}^{d-1} \tilde{\sigma}_{j,k}2^{-k}$, where $\tilde{\sigma}_{j,k} \in \{0, 1\}$ and $d$ is the length of binary expanding, then the quantum circuit of control rotation can be described by Fig. 3.

![Fig. 3 QCR: quantum circuit of control rotation. $R(\alpha)$ is the two-dimensional rotation with rotation angle $\alpha$ and $\theta = \arccos(\tilde{C}/\tilde{\sigma}_j)$](image-url)
Define $f(z) = \arccos(\tilde{C}/z)$, then the quantum circuit of HHL algorithm is plotted in Fig. 4. But here we call it $\text{QSol}_{\{x_{\mu}\}}$, which generates the solution state (12) of the LSP.

### 4 Regularization parameter choice

A key issue for the success of the Tikhonov regularization is to determine a reasonable regularization parameter $\mu$. Classically, we choose a range of parameters $\mu$. For each parameter $\mu$, we solve the problem (1) and obtain the corresponding solution $x_{\mu}$. From such a series of solutions, we take a proper strategy to find the possible best regularization parameter. Here we consider the quantum implementations of two typical heuristic methods, which use the L-curve and the Hanke–Raus rule, respectively. In L-curve or Hanke–Raus rule, we need to estimate the norm of the solution $\|x_{\mu}\|$ and the norm of the residual $\|Ax_{\mu} - b\|$.

A main technique we will use is the amplitude estimation, which is discussed in Sect. 2. Note that to apply HHL algorithm to solve the linear system (11), we only generate state (12) without measurement. Thus, (i) the norm of the solution is included in the amplitude of the first term of state (12). We can apply amplitude estimation to estimate it. (ii) The same technique applies to estimate $\|Ax_{\mu} - b\|$ if we can construct a quantum state such that $\|Ax_{\mu} - b\|$ appears in the amplitude of a term. This quantum state is obtained by first multiplying $A$ on (12) and then computing its linear combination with $|b\rangle$. In the following, we show the details of this idea.

First, we estimate the solution norm $\|x_{\mu}\|$. Applying the amplitude estimation to state (12), we obtain an $\alpha$ such that

$$
|\tilde{C}\|x_{\mu}\| - \alpha| \leq \epsilon_0,
$$

that is, $||x_{\mu}|| - \alpha/\tilde{C}| \leq \epsilon_0/\tilde{C}$. Note that the amplitude estimation involves the unitary operator $G$ defined in (3), which is further determined by $U$. In this case, $U$ is the unitary given in Fig. 4, and $\cos \theta = \tilde{C}\|x_{\mu}\|$.

This step costs $O(\kappa\mu(\log n)/\epsilon\epsilon_0)$. To make the error small in size $\epsilon$, we choose $\epsilon_0 = \tilde{C}\epsilon$. Finally, the complexity to get an $\epsilon$-approximation of $\|x_{\mu}\|$ is

$$
O\left(\kappa\mu(\log n)/\epsilon^2\tilde{\sigma}_{\text{min}}\right).
$$

![Fig. 4 QSol\(_{\{x_{\mu}\}}\): quantum circuit to generate the solution state (12), where $G = \exp(-i\tilde{A}_\mu t)$ and QPE\(_G\) is the quantum circuit shown in Fig. 1. QCR refers to the quantum circuit of Fig. 3. The outputs of the first and fourth line constitute state (12)](image)
Next, we consider how to estimate the norm \( \|Ax_\mu - b\| \). As a generalization of HHL algorithm, the quantum state \(|x_\mu\rangle\) given in (12) can be multiplied by \( A \). Then we get the following quantum state (see appendix B for more details)

\[
|\psi\rangle = C\|x_\mu\| A|x_\mu\rangle |0\rangle + |0\rangle^\perp,
\]

where \( C = \tilde{C}/\hat{C} \) and \( \hat{C} \) is a upper bound of \( \sigma_{\text{max}} \). Based on the analysis in appendix B, we can construct a unitary operator to prepare \(|\psi\rangle\) from \(|0\rangle\), which consists of two stages. The first stage is in fact \( \text{QSol} |X_\mu\rangle \) in Fig. 4, which prepares state (12). The second stage implements the multiplication of \( A \) on state (12), which is achieved by substituting \( G = \exp(-it_0\tilde{A}) \), and \( f(z) = \arccos(z/\sigma_{\text{max}}) \) in Fig. 4.

To estimate \( \|Ax_\mu - b\| \) by amplitude estimation, we need to prepare a quantum state proportional to \( \|x_\mu\| A|x_\mu\rangle - |b\rangle \). By definition \( \|Ax_\mu - b\| = \|\|x_\mu\| A|x_\mu\rangle - |b\rangle\| \).

Here we assume that \( b \) has unit norm, which is a reasonable assumption when solving linear systems on a quantum computer. The required quantum state can be obtained by the following steps.

**Step 1:** Assume that \(|b\rangle\) is efficiently prepared, then we can prepare the following initial state efficiently

\[
\frac{1}{\sqrt{2}} (|\psi\rangle |0\rangle - |b, 0\rangle |1\rangle) |0\rangle.
\]

**Step 2:** Set \( \tau = \min\{1, C\} \). Here we do not need \( \tau \) equals \( \min\{1, C\} \) exactly. A low bound of \( \min\{1, C\} \) still works. Apply control rotations to (15) as follows. (i) \( R_0 \): If the second register is \(|0\rangle\), then change the last qubit \(|0\rangle\) into \( \tau C^{-1}|0\rangle + \sqrt{1 - \tau^2 C^{-1}|1\rangle} \); (ii) \( R_1 \): If the second register is \(|1\rangle\), then transform the last qubit \(|0\rangle\) into \( \tau |0\rangle + \sqrt{1 - \tau^2 |1\rangle} \). So we obtain

\[
\frac{1}{\sqrt{2}} |\psi\rangle |0\rangle \left[ \tau C^{-1}|0\rangle + \sqrt{1 - \tau^2 C^{-1}|1\rangle} \right] \\
- \frac{1}{\sqrt{2}} |b, 0\rangle |1\rangle \left[ \tau |0\rangle + \sqrt{1 - \tau^2 |1\rangle} \right].
\]

**Step 3:** Apply the Hadamard gate to the second register of (16) to generate

\[
|R_\mu\rangle = \frac{1}{\sqrt{2}} |\psi\rangle |+\rangle \left[ \tau C^{-1}|0\rangle + \sqrt{1 - \tau^2 C^{-1}|1\rangle} \right] \\
- \frac{1}{\sqrt{2}} |b, 0\rangle |-\rangle \left[ \tau |0\rangle + \sqrt{1 - \tau^2 |1\rangle} \right] \\
= \frac{\tau}{2} (\|x_\mu\| A|x_\mu\rangle - |b\rangle) |0, 0, 0\rangle + |0, 0, 0\rangle^\perp.
\]

It costs \( O(\kappa_\mu (\log n)/\epsilon) \) to obtain the above quantum state. The corresponding query complexity is \( O(\kappa_\mu /\epsilon) \). The quantum circuit of the above procedure is depicted in Fig. 5. We call it \( \text{QRes}_{|R_\mu\rangle} \), a quantum circuit to generate the quantum state (17) containing the residual.
When applying the amplitude estimation, the unitary operator $U$ used in Eq. (3) now equals the unitary drawn in Fig. 5, and $\cos \theta = \frac{\tau}{2} \| |x_\mu\| A|x_\mu\rangle - |b\rangle \|. Therefore, by amplitude estimation, we get a value $\beta$ in time $O(\kappa_\mu (\log n)/\epsilon \epsilon_1)$, such that

$$\left| \frac{\tau}{2} \| |x_\mu\| A|x_\mu\rangle - |b\rangle \| - \beta \right| \leq \epsilon_1.$$ 

To get an $\epsilon$-approximate of $\| |x_\mu\| A|x_\mu\rangle - |b\rangle \|$, we set $\epsilon_1 = \epsilon \tau / 2$. Generally $\tau = C$ by definition, then we have $\epsilon_1 = \epsilon C/2 \leq \epsilon \tilde{\sigma}_{\min}/2\sigma_{\max}$. Hence, the gate complexity to get an $\epsilon$-approximation of $\| |x_\mu\| A|x_\mu\rangle - |b\rangle \|$ reads

$$O \left( \kappa_\mu (\log n)\sigma_{\max}/\epsilon^2 \tilde{\sigma}_{\min} \right),$$

and the query complexity is $O(\kappa_\mu \sigma_{\max}/\epsilon^2 \tilde{\sigma}_{\min})$.

As shown in Sect. 3, $\hat{\sigma}_j = \sqrt{\sigma_j^2 + \mu^2}$ and $\mu \ll 1$ generally. The singular values of $A$ usually contain small magnitudes, so we just assume that $1/\kappa \leq \sigma_j < 1$ for all $j$. Thus, $\max \hat{\sigma}_j = O(1)$, which implies that $1/\tilde{\sigma}_{\min} \approx \kappa_\mu$. Concluding the above analysis, we have

**Theorem 1** Suppose the Hamiltonian simulation of $A$ and the preparation of the state $|b\rangle$ are efficient. Let $x_\mu$ be the solution of the regularized least squares problem (1). Then we can compute the $\epsilon$-approximations of $\| x_\mu \|$ and $\| Ax_\mu - b \|$ by using $O(\kappa_\mu^2/\epsilon^2)$ queries of $O_A$ and $O_b$, where $\kappa_\mu$ is the condition number of $A_\mu$. The gate complexity is $O(\kappa_\mu^2 (\log n)/\epsilon^2)$.

Note that the assumptions of the above theorem are the same as those of HHL algorithm. If we apply other quantum linear solvers, then the assumptions should be adjusted accordingly.

### 4.1 The L-curve

The so-called L-curve method uses a plot of the norms of the regularized solution versus the corresponding residual norms, i.e., the plot of $(\|Ax_\mu - b\|, \|x_\mu\|)$ over

---

1 Actually $\tau = C$, since if $\mu$ is small, then $C = \tilde{\sigma}_{\min}/\sigma_{\max} = \sqrt{\sigma_{\min}^2 + \mu^2}/\sigma_{\max} \leq 1$ generally.
Fig. 6 The L-shaped curve, where $\|Ax_\mu - b\|$ versus $\|x_\mu\|$ are in log–log scale

Table 1 The cost of each step of Algorithm 2

| #Step | Costs | Comments |
|-------|-------|----------|
| 1     | $O(\log p)$ | The application of Hadamard operator |
| 2     | $O \left( \left( \max_j \kappa_{\mu_j} \right) (\log n)/\epsilon \right)$ | The construction of (12) and (17) |
| 3     | $O \left( \left( \max_j \kappa_{\mu_j} \right)^2 (\log n)/\epsilon^2 \right)$ | The amplitude estimation is accomplished in parallel, by Corollary 1 and Theorem 1 |
| 5     | $O \left( \sqrt{p} \left( \max_j \kappa_{\mu_j} \right)^2 (\log np)/\epsilon^2 \right)$ | The QMF algorithm to find the best $\mu_j$ can achieve a quadratic speedup in $p$ |

A range of $\mu$. It gives an insight into the regularizing properties of the underlying regularization method and helps to choose an appropriate regularization parameter. To be precisely, we set a series of parameters, for example, $\mu_j = \rho^j$ ($j = 1, \ldots, p$) with $\rho < 1$, and plot $\|Ax_\mu - b\|$ versus $\|x_\mu\|$. If there is a corner on the L-curve, then we can take the corresponding parameter $\mu$ as the desired regularization parameter. An illustrative example\(^2\) for L-curve is given in Fig. 6.

A naive strategy suggests us to calculate all the values of ($\|Ax_{\mu_j} - b\|$, $\|x_{\mu_j}\|$) for $j = 1, \ldots, p$, at the cost $O(p(\max_j \kappa_{\mu_j})^2(\log n)/\epsilon^2)$ by Theorem 1. We then use these $p$ pairs of data to plot the L-curve and locate its corner (see Fig. 6). Classically, if we use the conjugate gradient method to solve the linear system, then the complexity is $O(psn(\max_j \sqrt{\kappa_{\mu_j}}) \log(1/\epsilon))$, where $s$ is the sparsity of $A$. So this naive quantum

\(^2\) The linear system $Ax = b$ arises from the example SHAW in the regularization tools [31], where the matrix $A$ is of order 1000 and right-hand size $b$ has a noise level $10^{-2}$. Each cross-marker corresponds to a certain $\mu$. 

algorithm achieves exponential speedup in $n$ due to the HHL algorithm, but no speedup at $p$.

Based on the QMF (Algorithm 1), we can improve the naïve quantum algorithm further to achieve a quadratic speedup in $p$. The appropriate regularization parameter is one of $\mu_j$ that corresponds to the corner of L-curve. For most cases, the point $(\|Ax_{\mu_j} - b\|, \|x_{\mu_j}\|)$ with minimal $f(\mu_j)$ is the corner, where $f(\mu_j) = (\log \|Ax_{\mu_j} - b\|)^2 + (\log \|x_{\mu_j}\|)^2$. Otherwise, a suitable translation is applied. Here we use this simplified version of L-curve method, that is, the best regularization parameter satisfies the condition that it minimizes $(\log \|Ax_{\mu_j} - b\|)^2 + (\log \|x_{\mu_j}\|)^2$. We can find a proper regularization parameter by computing $(\|Ax_{\mu_j} - b\|, \|x_{\mu_j}\|)$ in parallel and utilizing the QMF algorithm. The whole procedure goes as follows.

**Algorithm 2 Quantum regularization parameter estimate based on the L-curve**

1. Prepare the state

$$\frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle|0\rangle|0\rangle|0\rangle|0\rangle.$$

2. Denote the quantum states (12) and (17) for $\mu_j$ as $|X_j\rangle$ and $|R_j\rangle$ respectively. Then prepare them in parallel by the control operation, so we have

$$\frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle|X_j\rangle|0\rangle|0\rangle|R_j\rangle.$$

3. Apply (7) to estimate $\|x_{\mu_j}\|$ and $\|Ax_{\mu_j} - b\|$ in parallel, which yields

$$\frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle|X_j\rangle|\|x_{\mu_j}\||0\rangle|\|Ax_{\mu_j} - b\||R_j\rangle. \tag{19}$$

4. Apply the oracle $O : |a, 0, b\rangle \mapsto |a, \log^2 a + \log^2 b, b\rangle$ to the above state to get

$$\frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle|X_j\rangle|\|x_{\mu_j}\||f(\mu_j)|\|Ax_{\mu_j} - b\||R_j\rangle.$$

5. Apply the QMF algorithm to seek the minimum of $\{f(\mu_j) : j = 1, \ldots, p\}$ and then find the best $\mu_j$.

Regarding the complexity, we list the cost of each step in Table 1. Note that in step 4, the oracle is generated by a simple function, which has an efficient quantum circuit. So we will not count it into the complexity analysis. The final result shows that the total cost (gate complexity) is $O(\sqrt{p}(\max_j \kappa_{\mu_j})^2(\log np)/\epsilon^2)$. Algorithm 1 follows
Fig. 7 Quantum circuit of L-curve: $\text{QSol}|x_j\rangle$ and $\text{QRes}|R_j\rangle$, respectively, refer to the quantum circuit shown in Figs. 4 and 5 to generate $|x_j\rangle$ and $|R_j\rangle$ with regularization parameter $\mu_j$. $\text{QEval}|x_j\rangle,f_0$ and $\text{QEval}|R_j\rangle,f_1$ are the quantum circuits shown in Fig. 2 with $f_0(z) = z/\widetilde{C}$ and $f_1(z) = 2z/\min\{1, C\}$. The oracle $O : |a, 0, b\rangle \mapsto |a, \log^2 a + \log^2 b, b\rangle$ is the oracle to implement the function $f(a, b) = \log^2 a + \log^2 b$. QMF refers to the quantum minimum finding algorithm.

Fig. 8 Quantum circuit of Hanke–Raus method, where $O_\mu : |j, 0\rangle \mapsto |j, \mu_j\rangle$ is the oracle to query the testing regularization parameters and $O : |a, 0, b\rangle \mapsto |a, b/\sqrt{a}, b\rangle$ is the oracle to implement the function $f(a, b) = b/\sqrt{a}$ when $a \neq 0$. $\text{QRes}|R_j\rangle$ is the quantum circuit shown in Fig. 5 to generate $|R_j\rangle$ with regularization parameter $\mu_j$, and $\text{QEval}|R_j\rangle,f_1$ is the quantum circuit shown in Fig. 2 with $f_1(z) = 2z/\min\{1, C\}$. QMF is the quantum minimum finding algorithm.

the idea of Grover’s algorithm, whose query complexity is $O(\sqrt{p})$. Together with the query complexity of solving linear systems, the query complexity of Algorithm 2 equals $O(\sqrt{p}(\max_j \kappa_{\mu_j})^2/\epsilon^2)$. The quantum circuit to implement the Algorithm 2 is given in Fig. 7, where we assume that $p = 2^q$ for convenience.

4.2 The Hanke–Raus rule

The Hanke–Raus rule [9] aims to find the best regularization parameter by minimizing

$$\frac{\|Ax_\mu - b\|}{\sqrt{\mu}}. \quad (20)$$
Similar to the L-curve method, we set a series of testing regularization parameters \( \mu_1, \ldots, \mu_p \) and then find the best one such that it minimizes (20). The complexity of the classical algorithm is still \( O(psn(\max_j \sqrt{\kappa_{\mu_j}}) \log(1/e)) \) if we use the conjugate gradient method to solve the linear systems. Similar to the quantum algorithm based on L-curve, quantum Hanke–Raus method also achieves quadratic speedup at \( p \) and exponential speedup at \( n \). The algorithm reads as follows.

**Algorithm 3 Quantum regularization parameter estimate based on the Hanke-Raus rule**

1: Prepare the state

\[
\frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle |\mu_j\rangle |0\rangle |0\rangle.
\]

2: Denote the quantum state (17) for \( \mu_j \) as \( |R_j\rangle \), then prepare

\[
\frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle |\mu_j\rangle |0\rangle |R_j\rangle.
\]

3: Apply (7) to estimate \( \|Ax_{\mu_j} - b\| \) in parallel, which yields

\[
\frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle |\mu_j\rangle |0\rangle \|Ax_{\mu_j} - b\| |R_j\rangle.
\]

4: Apply an oracle, which is defined as \( |a, 0, b\rangle \mapsto |a, b/\sqrt{a}, b\rangle \), to prepare

\[
\frac{1}{\sqrt{p}} \sum_{j=1}^{p} |j\rangle |\mu_j\rangle \left( \frac{\|Ax_{\mu_j} - b\|}{\sqrt{\mu_j}} \right) \|Ax_{\mu_j} - b\| |R_j\rangle.
\]

5: Apply the QMF algorithm to find \( \mu_j \) such that it gives the minimal \( \|Ax_{\mu_j} - b\|/\sqrt{\mu_j} \).

The complexity analysis is the same as that of L-curve method. Figure 8 shows the quantum circuit of the above procedures in Algorithm 3.

5 Discussion

The discussion till now only focuses on the determination of Tikhonov regularization parameter, but it is also applicable to the truncated SVD method (TSVD), where just the largest \( r \) singular values are kept while the other small ones are neglected. Using
the resulting best rank-\( r \) approximation of \( A \), the TSVD regularized solution \( x_r \) is given by

\[
x_r = \sum_{i=1}^{r} \frac{u_i^\dagger b}{\sigma_i} v_i,
\]

(21)

where \( r \) acts as the truncation parameter and is chosen to discard the noise-dominated small singular values. From another viewpoint, (2) can be reduced to the TSVD solution (21) by replacing the filter factors \( f_i \) in (2) with 0’s and 1’s appropriately.

The complexity to find the best regularization parameter by L-curve or Hanke–Raus rule depends on the condition number \( \kappa_{\mu} \) of \( A_{\mu} \). For an ill-posed problem, the condition number of \( A \) is large. For a properly chosen parameter \( \mu \), the condition number \( \kappa_{\mu} \) can be much smaller. However, \( \kappa_{\mu} \) can be still large if \( \mu \) is chosen improperly; as a result, the HHL solver runs very slowly. For the practical implementation, it is reasonable to set a threshold \( \tau = O(\text{poly log } n) \) for the runtime of HHL. If the runtime is larger than \( \tau \), then we conclude that the parameter is not proper and terminate the algorithm. In L-curve or Hanke–Raus rule, the runtime changes gradually according to the regularization parameters \( \mu_1, \ldots, \mu_p \). When the regularization method works, then solving LSP based on the best regularization parameter, say \( \mu_{\text{opt}} \), runs in a reasonable time. The runtime for other regularization parameters close to \( \mu_{\text{opt}} \) is also acceptable. As for the bad regularization parameters, perhaps \( \mu_1 \) or \( \mu_p \), the runtime is much longer than \( \mu_{\text{opt}} \). Therefore, setting a threshold on the runtime is reasonable for practical implementation.

After finding the best regularization parameter by our algorithms, we then obtain a regularized least squares solution \( |x\rangle \) in the quantum state. To read out the classical information of \( x \) from its quantum state is not easy. We can use statistical sampling and quantum state tomography to extract a concise representation for the quantum state \( |x\rangle \) by \( M \) parameters [15], where \( M = O(\text{poly log } n) \). Compressed sensing [32–34] can be used to reconstruct the state within error \( O(\epsilon) \) after \( O(M^2 \log^2 M / \epsilon^2) \) measurements [15]. Furthermore, for many problems in machine learning, such as big data classification, the quantum state of \( x \) is already enough.

Our quantum algorithms are based on the HHL algorithm and hence have the same limitations as the HHL, for example, requiring the sparsity of coefficient matrix to guarantee an efficient simulation of Hamiltonian, together with an efficient preparation of quantum states such as \( |b\rangle \). We can apply other efficient quantum linear solvers [17,25] to overcome the restriction on Hamiltonian simulation and to improve the dependence of the complexity of our quantum algorithms on condition number and precision.

6 Conclusions

The LSP solver is a basic engine in big data and machine learning. To obtain a meaningful solution for an ill-posed problem, a regularization technique is necessary. The determination of regularization parameter is the most important, but also the most
time-consuming problem. In this work, based on L-curve and Hanke–Raus rule, we proposed two quantum algorithms to solve the regularization parameter estimate problem. The result shows that quantum computer can achieve quadratic speedup in the number of given regularization parameters and exponential speedup in the dimension of problem size. The former results from the QMF algorithm, and the latter is due to the HHL algorithm.

**Acknowledgements**

HX is supported by the Natural Science Foundation of China under Grants 11571265 and NSFC-RGC No. 11661160107. CS was supported by the QuantERA ERA-NET Cofund in Quantum Technologies implemented within the European Union’s Horizon 2020 Programme (QuantAlgo Project), and EPSRC Grants EP/L021005/1 and EP/R043957/1. No new data were created during this study.

**Appendix A: The procedure of HHL algorithm to generate state (12)**

Suppose the SVD of \( \widetilde{A}_\mu = \sum_j \tilde{\sigma}_j \tilde{u}_j \langle \tilde{u}_j | \) and the smallest nonzero singular value is represented by \( \tilde{\sigma}_{\text{min}} \). We formally decompose \( | \tilde{b} \rangle = \sum_j \tilde{\beta}_j | \tilde{u}_j \rangle \). Then by QPE, we obtain

\[
\sum_j \tilde{\beta}_j | \tilde{u}_j \rangle | \tilde{\sigma}_j \rangle.
\]

Denote \( \tilde{C} \) as a lower bound of \( \sigma_{\text{min}} \), and define \( f(z) = \arccos(\tilde{C}/z) \), then apply \( U_f \) to the above state to get

\[
\sum_j \tilde{\beta}_j | \tilde{u}_j \rangle | \tilde{\sigma}_j \rangle | \arccos \left( \frac{\tilde{C}}{\tilde{\sigma}_j} \right) \rangle.
\]

Now use control rotation to generate

\[
\sum_j \tilde{\beta}_j | \tilde{u}_j \rangle | \tilde{\sigma}_j \rangle \left[ \frac{\tilde{C}}{\tilde{\sigma}_j} | 0 \rangle + \sqrt{1 - \left( \frac{\tilde{C}}{\tilde{\sigma}_j} \right)^2} | 1 \rangle \right].
\]

At last, undo \( U_f \) and QPE, then we have

\[
\sum_j \tilde{\beta}_j | \tilde{u}_j \rangle \left[ \tilde{C} \tilde{\sigma}_j^{-1} | 0 \rangle + \sqrt{1 - \tilde{C}^2 \tilde{\sigma}_j^{-2}} | 1 \rangle \right] = \tilde{C} \| x_\mu \| | x_\mu \rangle | 0 \rangle + P_1 | \phi_1 \rangle | 1 \rangle.
\]

**Appendix B: Multiply A on state (12)**

Denote the SVD of \( A = \sum_j \sigma_j | u_j \rangle \langle v_j | \), then the eigenvalue decomposition of \( \widetilde{A} = \begin{pmatrix} 0 & A \\ A^\dagger & 0 \end{pmatrix} \) is \( \sum_j \pm \sigma_j | w_j^\pm \rangle \langle w_j^\pm | \), where \( | w_j^\pm \rangle = \frac{1}{\sqrt{2}} (| 0 \rangle | u_j \rangle \pm | 1 \rangle | v_j \rangle) \).
Formally, we can rewrite $|1, x_\mu\rangle = \sum_j x_{\mu,j} |1, v_j\rangle = \sum_j x_{\mu,j} \sqrt{2} (|w_j^+\rangle - |w_j^-\rangle)$.

Then

$$|1, X_\mu\rangle = \tilde{C} \|x_\mu\| \sum_j \frac{x_{\mu,j}}{\sqrt{2}} (|w_j^+\rangle - |w_j^-\rangle) |0\rangle + P_1 |0, \phi_1\rangle |1\rangle.$$ 

By viewing $|0\rangle$ in the third register as control qubit, we perform QPE to $e^{i0\tilde{A}}$ with initial state $|1, X_\mu\rangle$, then we obtain

$$\tilde{C} \|x_\mu\| \sum_j \frac{x_{\mu,j}}{\sqrt{2}} (|w_j^+\rangle |\sigma_j\rangle - |w_j^-\rangle - \sigma_j) |0\rangle + P_1 |0, \phi_1\rangle |1\rangle.$$ 

Denote $f(z) = \arccos(z/\tilde{C})$, where $\tilde{C}$ is a upper bound of $\sigma_{\max}$. Apply $U_f$ to the above state to prepare

$$\tilde{C} \|x_\mu\| \sum_j \frac{x_{\mu,j}}{\sqrt{2}} (|w_j^+\rangle |\sigma_j\rangle - |w_j^-\rangle - \sigma_j) \arccos \left( -\frac{\sigma_j}{\tilde{C}} \right) |0\rangle + P_1 |0, \phi_1\rangle |1\rangle.$$ 

Now apply control rotation to generate

$$\tilde{C} \|x_\mu\| \sum_j \frac{x_{\mu,j}}{\sqrt{2}} \left( |w_j^+\rangle |\sigma_j\rangle \arccos \left( \frac{\sigma_j}{\tilde{C}} \right) \right) \otimes \left[ \frac{\sigma_j}{\tilde{C}} |0\rangle + \sqrt{1 - \frac{\sigma_j^2}{\tilde{C}^2}} |1\rangle \right]$$

$$- |w_j^-\rangle |\sigma_j\rangle \arccos \left( -\frac{\sigma_j}{\tilde{C}} \right) \otimes \left[ -\frac{\sigma_j}{\tilde{C}} |0\rangle + \sqrt{1 - \frac{\sigma_j^2}{\tilde{C}^2}} |1\rangle \right] |0\rangle + P_1 |0, \phi_1\rangle |1\rangle.$$ 

Finally, undo $U_f$ and QPE we obtain

$$\frac{\tilde{C}}{\tilde{C}^\prime} \|x_\mu\| \sum_j \frac{x_{\mu,j}}{\sqrt{2}} \sigma_j (|w_j^+\rangle + |w_j^-\rangle) |0, 0\rangle + \text{orthogonal terms}$$

$$= \frac{\tilde{C}}{\tilde{C}^\prime} \|x_\mu\| \sum_j x_{\mu,j} \sigma_j |0, u_j\rangle |0, 0\rangle + \text{orthogonal terms}$$

$$= \frac{\tilde{C}}{\tilde{C}^\prime} \|x_\mu\| |0\rangle A |x_\mu\rangle |0, 0\rangle + \text{orthogonal terms}.$$ 

Therefore, there is a unitary that maps $|1, X_\mu\rangle$ to state (22). The first qubit $|0\rangle$ is ancillary.

References

1. Banks, H., Kunisch, K.: Parameter Estimation Techniques for Distributed Systems. Birkhäuser, Boston (1989)
2. Engl, H.W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer, Dordrecht (1996)
3. Hansen, P.C.: Discrete Inverse Problems, Insight and Applications. SIAM, Philadelphia (2010)
4. Tikhonov, A., Arsenin, V.: Solutions of Ill-Posed Problems. Wiley, New York (1977)
5. Haykin, S.O.: Neural Networks and Learning Machine, 3rd edn. Pearson, London (2008)
6. Xiang, H., Zou, J.: Randomized algorithms for large-scale inverse problems with general Tikhonov regularizations. Inverse Probl. 31, 085008 (2015)
7. Hansen, P.C.: Analysis of discrete ill-posed problems by means of the L-curve. SIAM Rev. 34, 561–580 (1992)
8. Hansen, P.C., O’Leary, D.P.: The use of the L-curve in the regularization of discrete ill-posed problems. SIAM J. Sci. Comput. 14, 1487–1503 (1993)
9. Hanke, M., Raus, T.: A general heuristic for choosing the regularization parameter in ill-posed problems. SIAM J. Sci. Comput. 17, 956–972 (1996)
10. Golub, G.H., Heath, M., Wahba, G.: Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21, 215–223 (1979)
11. Xiang, H., Zou, J.: Regularization with randomized SVD for large-scale discrete inverse problems. Inverse Probl. 29, 085008 (2013)
12. Harrow, A.W., Hassidim, A., Lloyd, S.: Quantum algorithm for solving linear systems of equations. Phys. Rev. Lett. 15(103), 150502 (2009)
13. Brassard, G., Høyer, P., Mosca, M.: Quantum amplitude amplification and estimation. Quantum Comput. Quantum Inf. 305, 53–74 (2002)
14. Dürr, D., Høyer, P.: A quantum algorithm for finding the minimum. arXiv:quant-ph/9607014 (1996)
15. Wiebe, N., Braun, D., Lloyd, S.: Quantum algorithm for data fitting. Phys. Rev. Lett. 109, 050505 (2012)
16. Schuld, M., Sinayskiy, I., Petruccione, F.: Prediction by linear regression on a quantum computer. Phys. Rev. A 94, 022342 (2016)
17. Chakraborty, S., Gilyén, A., Jeffery, S.: The power of block-encoded matrix powers: improved regression techniques via faster Hamiltonian simulation. in: Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP 2019), pp. 33:1–33:14 (2019)
18. Kerenidis, I., Prakash, A.: Quantum gradient descent for linear systems and least squares. arXiv:1704.04992v3 (2017)
19. Wang, G.M.: Quantum algorithm for linear regression. Phys. Rev. A 96, 012335 (2017)
20. Liu, Y., Zhang, S.Y.: Fast quantum algorithms for least squares regression and statistic leverage scores. Theoret. Comput. Sci. 657, 38–47 (2017)
21. Li, G.X., Wang, Y.L., Luo, Y., Feng, Y.: Quantum data fitting algorithm for non-sparse matrices. arXiv:1907.06949 (2019)
22. Yu, C.H., Gao, F., Wen, Q.Y: An improved quantum algorithm for ridge regression. arXiv:1707.09524 (2017)
23. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000)
24. Rieffel, E., Polak, W.: Quantum Computing—A Gentle Introduction. The MIT Press, Cambridge (2011)
25. Wossnig, L., Zhao, Z.K., Prakash, A.: A quantum linear system algorithm for dense matrices. Phys. Rev. Lett. 120, 050502 (2018)
26. Childs, A.M., Kothari, R., Somma, R.D.: Quantum linear systems algorithm with exponentially improved dependence on precision. SIAM J. Comput. 46, 1920–1950 (2017)
27. Clader, B.D., Jacobs, B.C., Sprouse, C.R.: Preconditioned quantum linear system algorithm. Phys. Rev. Lett. 110, 250504 (2013)
28. Shao, C., Xiang, H.: Quantum circulant preconditioner for a linear system of equations. Phys. Rev. A 98, 062321 (2018)
29. Buhrman, H., Cleve, R., Watrous, J., de Wolf, R.: Quantum fingerprinting. Phys. Rev. Lett. 87, 167902 (2001)
30. Childs, A.M., Maslov, D., Nam, Y., Ross, N.J., Su, Y.: Toward the first quantum simulation with quantum speedup. arXiv:1711.10980 (2017)
31. Hansen, P.C.: Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems. Numer. Algorithms 6, 1–15 (1994)
32. Gross, D., Liu, Y.K., Flammia, S.T., Becker, S., Eisert, J.: Quantum state tomography via compressed sensing. Phys. Rev. Lett. 105, 150401 (2010)
33. Shabani, A., Kosut, R.L., Mohseni, M., Rabitz, H., Broome, M.A., Almeida, M.P., Fedrizzi, A., White, A.G.: Efficient measurement of quantum dynamics via compressive sensing. Phys. Rev. Lett. 106, 100401 (2011)

34. Shabani, A., Mohseni, M., Lloyd, S., Kosut, R.L., Rabitz, H.: Estimation of many-body quantum Hamiltonians via compressive sensing. Phys. Rev. A 84, 012107 (2011)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.