Int J Theor Phys (2018) 57:862–876
https://doi.org/10.1007/s10773-017-3619-9

Fidelity-Based Ant Colony Algorithm with Q-learning of Quantum System

Qin Liao[1] · Ying Guo[1] · Yifeng Tu[1] · Hang Zhang[1]

Received: 20 July 2017 / Accepted: 29 November 2017 / Published online: 11 December 2017

Abstract Quantum ant colony algorithm (ACA) has potential applications in quantum information processing, such as solutions of traveling salesman problem, zero-one knapsack problem, robot route planning problem, and so on. To shorten the search time of the ACA, we suggest the fidelity-based ant colony algorithm (FACA) for the control of quantum system. Motivated by structure of the Q-learning algorithm, we demonstrate the combination of a FACA with the Q-learning algorithm and suggest the design of a fidelity-based ant colony algorithm with the Q-learning to improve the performance of the FACA in a spin-1/2 quantum system. The numeric simulation results show that the FACA with the Q-learning can efficiently avoid trapping into local optimal policies and increase the speed of convergence process of quantum system.

Keywords Fidelity · Ant colony algorithm · Q-learning · Quantum computation

1 Introduction

There are two kinds of control strategies for the optimum designs, i.e., the open-loop control and the closed-loop control. Because there is a simple controller in the open-loop control, which is convenient to be implemented, the open-loop control strategy was initially adopted for solving the optimizing problems in quantum system [1]. It must satisfy the following constraints: i) The system Hamiltonian H_0 is known to high precision. ii) The multidimensional Schrödinger equation $i\hbar\frac{\partial\psi}{\partial t} = [H_0 - \mu\varepsilon(t)]\psi$ may be reliably solved. iii) The resultant control field design $\varepsilon(t)$ may be faithfully reproduced. Unfortunately, it

✉ Ying Guo
yingguo@csu.edu.cn

[1] School of Information Science and Engineering, Central South University, Changsha, 410083, China

is difficult to practically implement such an open-loop control strategy that satisfies the above-mentioned assumptions, leading to the quantum closed-loop control strategy in the literature [2].

The traditional closed-loop learning control strategy was initially proposed by Rabitz [3]. It is an effective control method for the quantum control, which can optimize the system performance through learning from experience by searching for the best control strategy. It involves three elements, i.e., i) an input initial guess, ii) the laboratory generation of the control applying to the sample and subsequently observed for its impact, and iii) a learning algorithm that considers the prior experiments and suggests the form of the next control input. Designing the appropriate quantum learning control algorithm is an important task for the closed-loop learning control strategy, which includes the gradient algorithm, the genetic algorithm, the linear mapping algorithm, the nonlinear mapping algorithm, the reinforcement learning algorithm, and so on.

There is strong robustness required for the gradient algorithm [3], whereas the gradient $\delta J/\delta\varepsilon(t)$ must be measured with the inherent errors. The genetic algorithm was widely used in quantum control, due to its strong optimization ability and powerful search ability [4]. However a potential difficulty with this algorithm is that it has to work with large numbers of control variables. The input-output relationship of controlled system can be fully considered in the linear mapping algorithm and be simplified to the linear maps. So the linear mapping algorithm can be effectively used in the closed-loop learning control. Because the quantum control systems may be highly nonlinear, the practical application of the linear mapping algorithm has been limited. Under such circumstances, the nonlinear mapping algorithm has been development [5, 6]. The integration of the genetic algorithm and the nonlinear algorithm can solve the problem of the chemical quantum system. The reinforcement learning algorithm makes a good tradeoff between exploration and exploitation and hence can speedup the learning rate. However, because of the implementing constraints of the Grover algorithm, the reinforcement learning algorithm can't be used for the practical quantum system.

To strengthen the searching speed, the ACA was proposed by Marco Dorigo motivated by the behavior that ants find the optimal path in the feeding process [7]. It is essentially a complex intelligent system. The ACA is a probability algorithm, which simulates the foraging behavior of ants to find an optimal path [8–10]. The ACA-based optimality has been deeply studied in the intelligent control and industrial applications [11, 12]. However, there exist several issues in developing the ACA-based applications, including tradeoff between exploration and exploitation, search efficiency and search quality [13, 14]. After that, the FACA was proposed to solve the above-mentioned problems, only focusing on the transition probability with fidelity-directed exploration strategy in quantum systems [15, 16]. Although the FACA can optimize the traditional ACA, the parameter selection process is an intolerable tedious task since it requires a large number of experiments to compare and determine the parameter values. On one hand, the parameter selections of ACA are sensitive and rely on the characteristics of the searching goal. On the other hand, the parameters cannot be transferred from a problem to another. In order to simplify the parameter selection of the ACA and reduce the human interaction with the intelligence, the Q-learning, as a kind of reinforcement learning algorithm, is used for the FACA [17, 18], which can fling off environmental restrictions. Through the Q-function training, the agent can determine the optimal strategy by learning in the case of unknown environment models [19, 20]. Because the self-learning of the agent replaces the complicated parameter selection process [21], the intellectuality and stability of ACA will be improved.

Motivated by unique characteristics of quantum systems that make open loop strategies competitive [22, 23], we propose an algorithm to increase the transition probability of the ACA in the complex quantum system [24, 25]. For the given set of control fields, we apply the FACA to control laws for solving a class of quantum control problems. In quantum information processing, the closeness between two quantum states can be measured by fidelity [26]. Higher fidelity between the two states represents that their similarity is also higher [27], thus we can use the fidelity to replace the heuristic function between quantum states [28]. The fidelity can be sent back to the transition probability, so we can use it to speedup the solving process to avoid getting lost and getting local optimal solution. Numerical simulations show that the FACA with the Q-learning enables the performance improvement of quantum control systems.

The rest of the paper is organized as follows. In Section 2, we demonstrate the structure of the ACA. In Section 3, the FACA is designed with the Q-learning in quantum system. In Section 4, we illustrate the applications of the FACA with the Q-learning in the control of a typical class of quantum systems(i.e. spin-1/2 system). In Section 5, we analyze the performance of the proposed algorithm by illustrating the convergence of the FACA with the Q-learning. Finally, conclusion is drawn in Section 6.

2 Ant Colony Algorithm

To illustrate the mathematical model of ant colony algorithm, the classical combination optimization problem is suggested to determine a shortest path passing through each node one time for n selectable nodes.

We assume d_{ij} is the distance between node i and node j, τ_{ij} is the pheromone intensity of path (i, j), and η_{ij} is the heuristic function, which is the visibility of path and reflects the degree of inspiration from node i to node j. In this section, the value of the heuristic function is given by $\frac{1}{d_{ij}}$.

The practical ACA can acquire an optimal control strategy of pheromone which is a response to environment or system. At the initial moment, the pheromone of each path is equal to a constant C, i.e., $\tau_{ij}(0) = C$. For the k^{th} ant, it selects the next node according to the amount of pheromone. The transition probability from node i to node j is given by

$$P_{ij}^{k}(t) = \begin{cases} \dfrac{\tau_{ij}^{\alpha}(t)\eta_{ij}^{\beta}(t)}{\sum\limits_{s \in allowed_k} \tau_{is}^{\alpha}(t)\eta_{is}^{\beta}(t)} & , \ j \in allowed_k \\ 0 & , \ otherwise, \end{cases} \tag{1}$$

where α is the information heuristic factor that reflects the role of the pheromone accrued by the walking ants, β is the expected heuristic value that represents the effect of the visibility of the heuristic information, and $allowed_k = \{1, 2, \cdots, n\}$ represents the possible nodes that the k^{th} ant can select.

Different from the real ants, the artificial ant system owns memory function. In order to satisfy the constraint, each ant cannot pass a node twice in one loop, where the table $tabu_k$ is designed to record the node that the ant has passed in current time. In this loop the ant is not allowed to go through the node again. In the next loop, the table $tabu_k$ is cleared and the ant chooses the next node again. Namely, $allowed_k - tabu_k$ shows the remaining nodes can be selected by the ants in the next step.

After repeating the above-mentioned processing n times, and the amount of pheromone for each path is updated according to the constraints

$$\tau_{ij}(t+n) = \rho\tau_{ij}(t) + \triangle\ \tau_{ij}(t, t+n), \tag{2}$$

$$\triangle\ \tau_{ij}(t, t+n) = \sum_{k=1}^{m} \triangle\ \tau_{ij}^{k}(t, t+n), \tag{3}$$

where ρ is the persistence coefficient of pheromone, $\triangle\ \tau_{ij}(t, t+n)$ is the increment of pheromone on the path (i, j), $\triangle\ \tau_{ij}^{k}(t, t+n)$ is the amount of pheromone left by ant k on the path (i, j) given by

$$\triangle\ \tau_{ij}^{k}(t, t+n) = \begin{cases} Q\ ,\ if\ \ pass\ \ path\,(\mathrm{i,j}) \\ 0\ \ \ ,\ otherwise, \end{cases} \tag{4}$$

where Q denotes the intensity of pheromone that affects the converging rate of algorithm. The procedure of a classical ant colony algorithm is shown as Algorithm 1.

Algorithm 1 A Classical Ant Colony Algorithm

```
Initialize τ_ij arbitrarily
Initialize P_ij^k
The m ants were randomly placed in n nodes
repeat
    for i = 1 to n do
        for k = 1 to b_i(t) do
            Choose state j with probability p_ij^k, and move ant k to node j
            Add the node j to tabu table
            P_ij^k ← formula(1)
            τ_ij^k ← ρτ_ij(t) + △ τ_ij(t, t+n)
        end for
    end for
until the target state is reached
Output the shortest path
```

3 The Improved Ant Colony Algorithm

3.1 The Fidelity-based Ant Colony Algorithm

In the traditional ACA, mostly, the reciprocal of the distance between two points is used as the heuristic function. Although the traditional ACA can make the system converge to a certain degree, it needs longer searching time than others and tends to be precocial and stagnant in some station. Thus, if we can extract more information from the system structure or system behavior, the system will converge more easily. Since the concept of fidelity is widely used in quantum and it reflects the internal connection between the input signal and the output signal. Moreover, it could potentially handle the complex quantum problems. Here, we develop a FACA method for control of quantum systems, which can also be applied to some other complex problems and can accelerate the convergence speed of control system.

The state transition updating rule of fidelity-based ACA for $P_{ij}^{k}(t)$ can be expressed as

$$P_{ij}^{k}(t) = \begin{cases} \dfrac{\tau_{ij}^{k}(t)\ F_{(i,target)}^{\beta}(t)}{\sum_{s \in allowed_k} \tau_{is}^{k}(t)\ F_{(i,target)}^{\beta}(t)} & , \ j \in allowed_k \\ 0 & , \ otherwise. \end{cases} \tag{5}$$

The specification of the fidelity $F_{(i,target)}$ is defined regarding the objective of the learning control task. In this paper, a fidelity of quantum pure states is adopted for the control of quantum systems. The pheromone updating rule is the same as formulas (2), (3) and (4). The procedure of the fidelity-based ACA algorithm is shown as Algorithm 2.

Algorithm 2 The Fidelity-based Ant Colony Algorithm

```
Initialize τ_ij arbitrarily
Initialize P_ij^k
The m ants were randomly placed in n states
repeat
  for i = 1 to n do
    for k = 1 to b_i(t) do
      Choose state j with probability p_ij^k, and move ant k to state j
      Add the state j to tabu table
      P_ij^k ← formula(5)
      τ_ij^k ← ρτ_ij(t) + △τ_ij(t, t+n)
    end for
  end for
until the target state is reached
Output the shortest path
```

3.2 The Fidelity-based Ant Colony Algorithm with Q-learning

Although the FACA can optimize the traditional ACA, the parameter selection process is complex and often needs a large number of experiments to determine the value of parameter. On the one hand, the parameters of ACA are sensitive and rely on the characteristics of the research question. On the other hand, parameters cannot be transferred from a problem to another. In order to simplify the selection problem of ACA parameters, reduce human interaction and improve the intelligence of the algorithm, we apply the Q-learning into the ACA and propose the fidelity-based ACA with Q-learning. Q-learning is a kind of reinforcement learning algorithm that can fling off environmental restrictions. Through Q function training, the Agent can determine the optimal strategy by learning in the case of unknown environment model. The self-learning of agent enhances the intellectuality and stability of ACA, since it replaces the parameter selection process which is tedious and needs human intervention.

The updating rule of fidelity-based ACA with Q-learning for P_{ij}^{k} is the same as formula (5). After the ant complete a cycle, the pheromone is updated as follows:

$$\tau_{ij}(t+n) = (1-\delta)\tau_{ij}(t) + \delta(\triangle\ \tau_{ij}(t, t+n) + \gamma \max_{s \in allowed} \tau_{is}(t)), \tag{6}$$

where δ is the learning rate in Q-learning, γ is a discount factor and the $\triangle\ \tau_{ij}(t, t+n)$ is the same as formula (3). The procedure of the fidelity-based ACA with Q-learning algorithm is shown as Algorithm 3.

Algorithm 3 The Fidelity-based Ant Colony Algorithm with Q-learning Algorithm

Initialize τ_{ij} arbitrarily
Initialize P_{ij}^k
The m ants were randomly placed in n states
repeat
  for $i = 1$ **to** n **do**
    for $k = 1$ **to** $b_i(t)$ **do**
      Choose state j with probability p_{ij}^k, and move ant k to state j
      Add the state j to tabu table
      $P_{ij}^k \leftarrow formula(5)$
      $\tau_{ij}^k \leftarrow (1-\delta)\tau_{ij}(t) + \delta(\triangle\ \tau_{ij}(t, t+n) + \gamma \max_{s \in allowed} \tau_{is}(t))$
    end for
  end for
until the target state is reached
Output the shortest path

4 The Improved Ant Colony Algorithms for Control of Quantum System Control

4.1 Learning Control of Quantum System

Learning control is an efficient way for the control of quantum system. In learning control, a control rule can be learned from the experience and an optimal control strategy can be gained by iteration. Here, we concentrate on the control problem of quantum pure state transition of N-level. The eigenstates of the free Hamiltonian H_0 of an N-level quantum system can be denoted as $D = |\phi_i\rangle_{i=1}^N$. The state $|\psi(t)\rangle$ can be expanded into a set of eigenstates in the set D, we have

$$|\psi(t)\rangle = \sum_{i=1}^{N} c_i(t)|\phi_i\rangle, \tag{7}$$

where $c_i(t)$ are complex numbers which satisfy $\sum_{i=1}^N |c_i(t)|^2 = 1$.

Definition 1 (*Fidelity of Quantum Pure States*): The fidelity between two pure states $|\psi^a\rangle = \sum_{i=1}^N c_i^a|\phi_i\rangle$ and $|\psi^b\rangle = \sum_{i=1}^N c_i^b|\phi_i\rangle$ is define as

$$F(|\psi^a\rangle, |\psi^b\rangle) = |\langle\psi^a|\psi^b\rangle| = |\sum_{i=1}^{N}(c_i^a)^* c_i^b|, \tag{8}$$

where $(c_i^a)^*$ is the complex conjugate of c_i^a.

Assuming that a quantum system including the internal Hamiltonian is H_0, the initial state is $\psi_0(x)$ and the external effects which coupled into a operator H_1. The Schrödinger equation then can be expressed as follows

$$\begin{cases} i\hbar\frac{\partial}{\partial t}\psi(x,t) = H_0\psi(x,t) + \varepsilon(t)H_1\psi(x,t) \\ \psi(x, t=0) = \psi_0(x), \end{cases} \tag{9}$$

where $[H_0, H_1] = H_0H_1 - H_1H_0 \neq 0$. The $\psi(x,t)$ are limited in a finite dimensional space. H_0 and H_1 correspond to matrices A and B respectively. C is the coefficient matrix of $\psi(x,t)$ expanded. Formula (9) can be expressed as follows:

$$\begin{cases} i\hbar\dot{C}(t) = [A + \varepsilon(t)B]C(t) \\ C(t=0) = C_0, \end{cases} \tag{10}$$

where $i = \sqrt{-1}$, $C(t) = (c_i(t))_{i=1}^N$, $C_0 = (c_{0i})_{i=1}^N$, $c_{0i} = \langle\phi_i|\psi_0\rangle$, $\sum_{i=1}^N |c_{0i}| = 1$, $\hbar$ is the reduced Planck constant. Here, we assume that matrix A is diagonal, matrix B is Hermitian and $[A, B] = AB - BA \neq 0$. The task of learning control is to find a control sequence to drive the quantum system from an initial state $|\psi_0\rangle$ to a target state $|\psi_f\rangle$. When an operator U is applied to the system at time $t = t_2$, the state transforms from state $|\psi(t_1)\rangle$ to state $|\psi(t_2)\rangle$ with the initial condition $|\psi(t_1)\rangle$ at time $t = t_1$.

4.2 Quantum Controlled Transition Landscapes

The aim of learning control of quantum systems is to find an optimal control strategy, which can manipulate the dynamics of physical processes on the atomic and molecular scales. Quantum control landscapes lay a theoretical foundation for analyzing learning control of quantum systems in recent year. A control landscape is defined as the mao between the time-dependent control Hamiltonian and associated values of the control performance functional. Most quantum control problems can be formulated as the maximization of an objective performance function.

Although quantum control applications may contain a variety of objectives, most of them correspond to maximizing the probability of transition from an initial state $|\psi_0\rangle$ to a desired final state $|\psi_f\rangle$. For the state transition problem with $t \in [0, T]$, we define the quantum controlled transition landscape [27] as

$$J(\varepsilon) = tr(U_{(\varepsilon,T)}|\psi_0\rangle\langle\psi_0|U^{\dagger}_{(\varepsilon,T)}|\psi_f\rangle\langle\psi_f|), \tag{11}$$

where $tr(\cdot)$ is the trace operator and $U^{\dagger}$ is the adjoint of U. The objective of the learning control system is to find a global optimal control strategy ε^*, which satisfies

$$\varepsilon^* = argmax_{\varepsilon} J(\varepsilon). \tag{12}$$

If the dependence of $U_{(T)}$ on ε is suppressed, (11) can be re-written as

$$J(U) = tr(U_{(T)}|\psi_0\rangle\langle\psi_0|U^{\dagger}_{(T)}|\psi_f\rangle\langle\psi_f|). \tag{13}$$

Equations (11) and (13) are called the dynamic control landscape (denoted as $J_D(\varepsilon)$ instead) and the kinematic landscape (denoted as $J_K(U)$ instead) respectively.

The characteristic of the existence or absence of traps is most important for exploring the quantum control landscape with a learning control algorithm, which can be studied using critical points. A dynamic critical point is defined by

$$\bigtriangledown J_D(\varepsilon) = \delta J_D(\varepsilon)/\delta_{\varepsilon} = 0 \tag{14}$$

and a kinematic critical point is defined as

$$\bigtriangledown J_K(U) = \delta J_K(U)/\delta_U = 0, \tag{15}$$

where ∇ denotes gradient. By the chain rule, we have

$$\nabla J_D(\varepsilon) = \frac{\delta J_K(U)}{\delta U_{(\varepsilon,T)}} \frac{\delta U_{(\varepsilon,T)}}{\delta_\varepsilon} = \nabla J_K(U) \frac{\delta U_{(\varepsilon,T)}}{\delta_\varepsilon}. \tag{16}$$

According to the results in Refs. [28], we can summarize the properties of quantum controlled transition landscape as *Theorem 1*.

Theorem 1 *For the quantum control problem defined with the dynamic control landscape* (11) *and the kinematic control landscape* (13) *respectively, the properties of the solution sets of the quantum controlled transition landscape are listed as follows.*

1) *The kinematic control landscape is free of traps (i.e., all critical points of $J_K(U)$ are either global maxima or saddles) if the operator U can be any unitary operator (i.e., the system is completely controllable).*
2) *The dynamic control landscape is free of traps if the operator U can be any unitary operator and the Jacobian $\delta U_{(\varepsilon,T)}/\delta_\varepsilon$ has full rank at any ε.*

For detailed proof and discussion about Theorem 1, please refer to Refs. [27, 28].

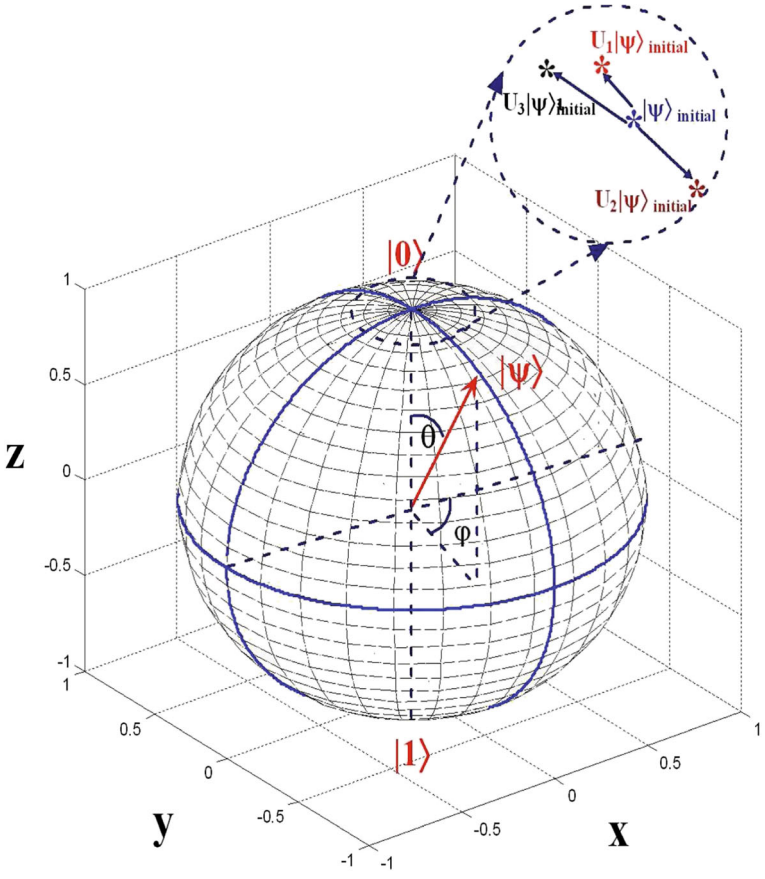

Fig. 1 Demonstration of a spin-(1/2) system with a Bloch sphere in a 3-D Cartesian coordinates and the state transitions for an initial quantum state $|\psi\rangle_{initial}$ using different one-step controls(U_1, U_2, U_3)

4.3 Example: Controlling of a Spin-(1/2) Quantum System

The spin-(1/2) system is a typical 2-level quantum system, which has important theoretical implications and practical applications. Its Bloch vector can be visualized on a 3-D Bloch sphere as shown in Fig. 1. The state of the spin-(1/2) quantum system $|\psi\rangle$ is represented as

$$|\psi\rangle = cos\frac{\theta}{2}|0\rangle + e^{i\varphi}sin\frac{\theta}{2}|1\rangle, \tag{17}$$

where $\theta \in [0, 2\pi]$ and $\varphi \in [0, 2\pi]$ are polar angle and azimuthal angle respectively, which specify a point $\vec{a} = (x, y, z) = (sin\theta cos\varphi, sin\theta sin\varphi, cos\theta)$ on the unit sphere in $\mathbf{R}^3$.

At each control step, the permitted controls for every point are U_1 (no control input), U_2 (a positive pulse control), and U_3 (a negative pulse control). Figure 1 shows a sketch map of one step control effects on the evolution of the quantum system.

The propagators $U_i, i = 1, 2, 3$ are listed as follows:

$$U_1 = e^{-iI_z\frac{\pi}{15}}, \tag{18}$$

$$U_1 = e^{-i(I_z+0.5I_x)\frac{\pi}{15}}, \tag{19}$$

$$U_1 = e^{-i(I_z-0.5I_x)\frac{\pi}{15}}, \tag{20}$$

where

$$I_z = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{21}$$

$$I_x = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{22}$$

Now the objective of control is to control the spin-(1/2) system from the initial state $(\theta = (\pi/60), \varphi = (\pi/30))$ to the target state $(\theta = (41\pi/60), \varphi = (29\pi/30))$ with

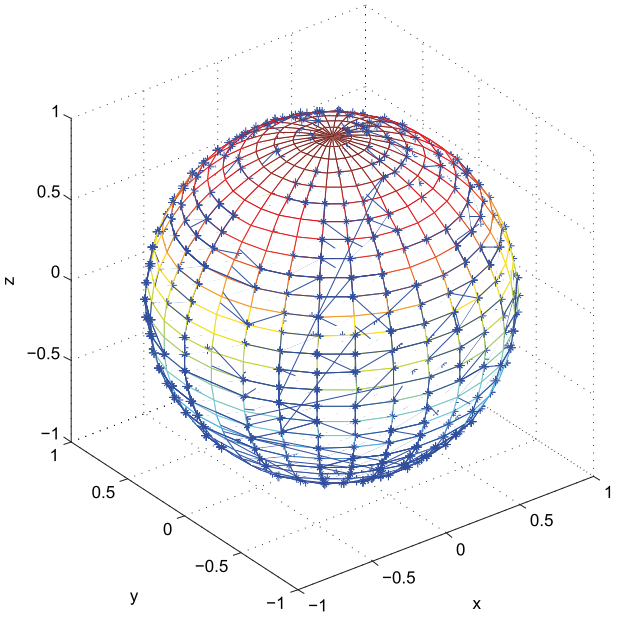

Fig. 2 Demonstration of all control processes

minimized control steps. Figures 2 shows the crawl path of each ant, namely the control process. We apply the fidelity-based ACA with Q-learning and fidelity-based ACA to this control problem respectively. Now we reformulate the problem of controlling a quantum system from an initial state $|\psi_{initial}\rangle$ to a desired target state $|\psi_{target}\rangle$ as follows: $S = i = |\psi_i\rangle$, $(i = 1, 2, ..., n)$, and this state can be obtained through some operation. The parameter settings for these algorithms are $\rho = 1 - \delta = 0.7$, $\alpha = 1$, $\beta = 10$ and $\gamma = 0.99$.

Figures 3 and 4 show the control performance of FACA and fidelity-based ACA with Q-learning respectively. There are many paths from the initial state to the target state, but there is only one shortest path in each experiment. Hundreds of times of the process finding the optimal path and it maintains similar performance. We provide the shortest path of the experiment. The experimental results show that fidelity-based ACA with Q-learning outperforms fidelity-based ACA. It is clear that the FACA can accelerate the convergence of the system. The control result of the fidelity-based ACA with Q-learning is better than FACA.

Figures 5 and 6 show the convergence of the FACA and fidelity-based ACA with Q-learning respectively. From the picture, we can know that the FACA and fidelity-based ACA with Q-learning can converge to an particular value. That is the FACA and fidelity-based ACA with Q-learning can obtain a certain control sequence to the spin-(1/2) system and the control sequence needs less than 50 steps. The fidelity-based ACA with Q-learning can find the optimal control sequence after about 65 generations, but the fidelity-based ACA needs about 115 generations.

The Fig. 7 shows the average mobile steps of the FACA and fidelity-based ACA with Q-learning. Although the fidelity-based ACA and fidelity-based ACA with Q-learning all can get a optimal control sequence, the average mobile steps of fidelity-based ACA is about 200 steps and the fidelity-based ACA with Q-learning is about 50 steps. As a result, the

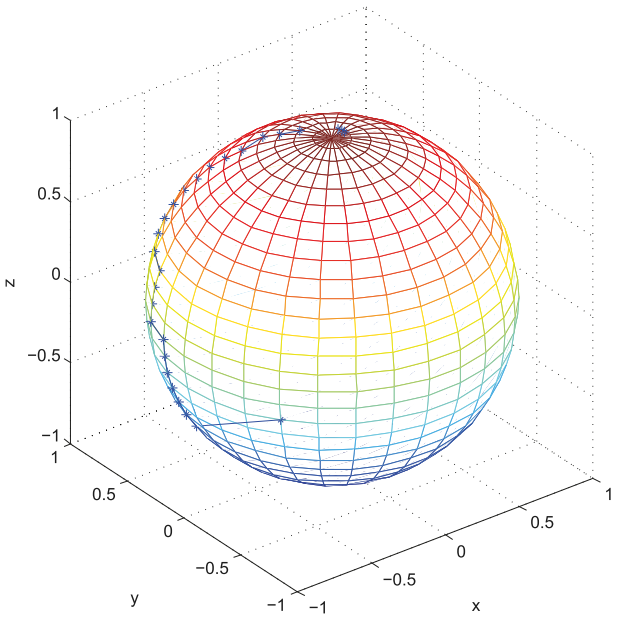

Fig. 3 The control result with an optimal control sequence of fidelity-based ACA

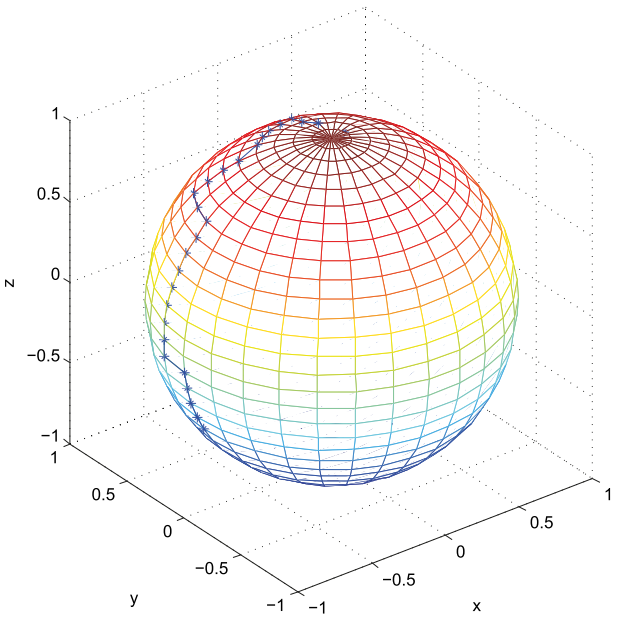

Fig. 4 The control result with an optimal control sequence of fidelity-based ACA with Q-learning

fidelity-based ACA with Q-learning has better performance and it can accelerate the speed of convergence.

5 The Convergence of the Improved ant Colony Algorithms

Assuming that there are m lines between state A and state B, for instance, AC_1B, AC_2B, ..., AC_MB. The fidelity of corresponding lines are F_1, F_2, ..., F_m respectively and

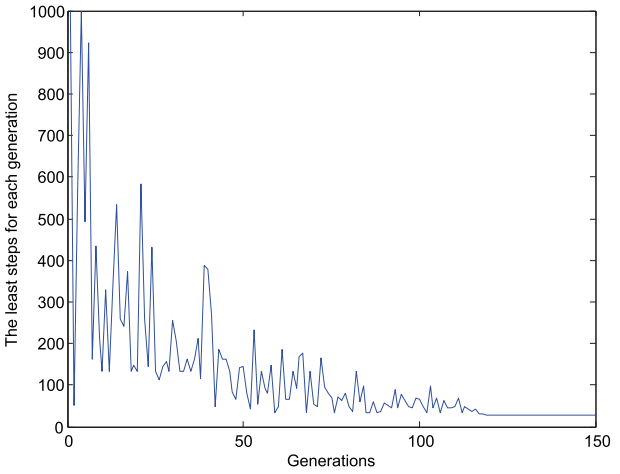

Fig. 5 The convergence of fidelity-based ACA

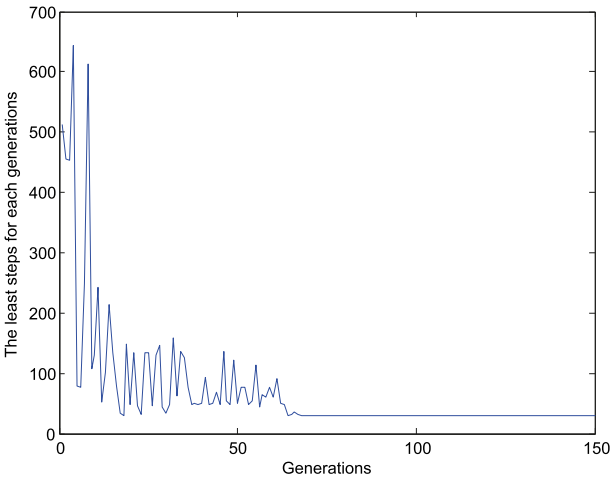

Fig. 6 The convergence of fidelity-based ACA with Q-learning

they meet the condition: $F_1 \geq F_2 \geq ... \geq F_m$. n ants crawl back and forth between A and B. According to FACA, if most of ants select the path AC_1B, then the path AC_1B is regarded as the optimal path. The FACA algorithm is convergent when the optimal path AC_1B is selected with close to 1.

Suppose that q_{ik} is the rest of average pheromone after the ant run k times. p_{ik} is the average probability of the path AC_iB selected after the ant runs k times. At the initial moment, the pheromone is $C(C$ is a constant) on each path. We can define a theorem as follows:

Theorem 2 *If* $\alpha \geq 0$, $\beta \geq 0$ *and* $F_1 \geq F_2 \geq ... \geq F_m$, *then* $q_{1k} \geq q_{2k} \geq ... \geq q_{mk}$ *and* $p_{1k} \geq p_{2k} \geq ... \geq p_{mk}$.

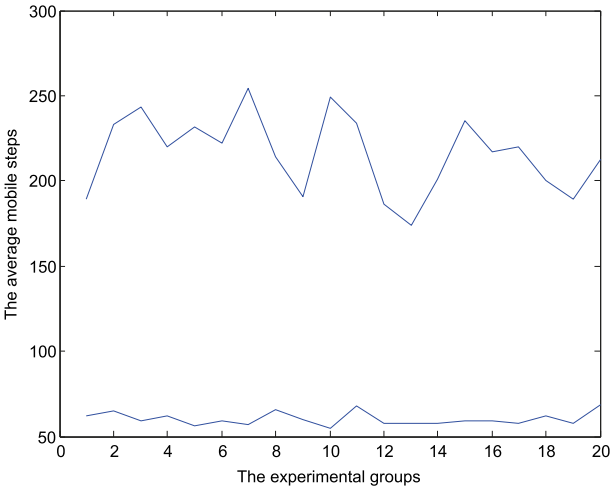

Fig. 7 The average mobile steps of fidelity-based ACA with Q-learning (the lower curve) and fidelity-based ACA (the higher curve)

Proof If $\alpha \geq 0$, $\beta \geq 0$, and $F_1 \geq F_2 \geq ... \geq F_m$, according to the formula (5), the $p_{i0} = \frac{C^\alpha F_i^\beta}{\sum_{j=1}^{m}(C^\alpha F_j^\beta)}$ can be obtained, and $p_{10} \geq p_{20} \geq ... \geq p_{m0}$ can be deduced. Thus $q_{11} \geq q_{21} \geq ... \geq q_{m1}$ according to $q_{i1} = \rho C + np_{i0}Q$. Similarly, the basis of $p_{i1} = \frac{q_{i1}^\alpha F_i^\beta}{\sum_{j=1}^{m}(q_{j1}^\alpha F_j^\beta)}$, $p_{11} \geq p_{21} \geq ... \geq p_{m1}$ can be deduced. We have, $q_{ik} = \rho q_{i(k-1)} + np_{i(k-1)}Q$ and $p_{ik} = \frac{q_{ik}^\alpha F_i^\beta}{\sum_{j=1}^{m}(q_{jk}^\alpha F_j^\beta)}$.

Through mathematical induction, we can know that when $\alpha \geq 0, \beta \geq 0$, then $q_{1k} \geq q_{2k} \geq \cdots \geq q_{mk}$ and $p_{1k} \geq p_{2k} \geq \cdots \geq p_{mk}$. □

Theorem 3 *shows that the average pheromone of path AC_1B is the greatest after the ants run one time, so the probability of path AC_1B selected is the optimal.*

Theorem 4 *If $\alpha \geq 1, \beta \geq 0$, then $p_{1k} > p_{1k-1}$.*

Proof Because $p_{1k} = \frac{q_{1k}^\alpha F_1^\beta}{\sum_{j=1}^{m} q_{jk}^\alpha F_j^\beta} = \frac{1}{1+(\frac{F_2}{F_1})^\beta(\frac{q_{2k}}{q_{1k}})^\alpha+(\frac{F_3}{F_1})^\beta(\frac{q_{3k}}{q_{1k}})^\alpha+\cdots+(\frac{F_m}{F_1})^\beta(\frac{q_{mk}}{q_{1k}})^\alpha}$.

so $\frac{1}{p_{1,k}} - \frac{1}{p_{1,k-1}} = (\frac{F_2}{F_1})^\beta[(\frac{q_{2,k}}{q_{1,k}})^\alpha - (\frac{q_{2,k-1}}{q_{1,k-1}})^\alpha] + (\frac{F_3}{F_1})^\beta[(\frac{q_{3,k}}{q_{1,k}})^\alpha - (\frac{q_{3,k-1}}{q_{1,k-1}})^\alpha] + \cdots + (\frac{F_m}{F_1})^\beta[(\frac{q_{m,k}}{q_{1,k}})^\alpha - (\frac{q_{m,k-1}}{q_{1,k-1}})^\alpha]$.

And $(\frac{q_{j,k}}{q_{1,k}}) - (\frac{q_{j,k-1}}{q_{1,k-1}}) = \frac{\rho q_{j,k-1}+np_{j,k-1}Q}{\rho q_{1,k-1}+np_{1,k-1}Q} - \frac{q_{j,k-1}}{q_{1,k-1}} = \frac{nQ(p_{j,k-1}q_{1,k-1}-p_{1,k-1}q_{j,k-1})}{(\rho q_{1,k-1}+nQp_{1,k-1})q_{1,k-1}}$.

Requesting $\frac{q_{j,k}}{q_{1,k}} - \frac{q_{j,k-1}}{q_{1,k-1}} < 0$, that is to say $p_{j,k-1}q_{1,k-1} < p_{1,k-1}q_{j,k-1}$.

The following formula can be calculated by substituting the expressions of $p_{j,k-1}$ and $p_{1,k-1}$:

$$q_{j,k-1}^\alpha q_{1,k-1} F_j^\beta < q_{1,k-1}^\alpha q_{j,k-1} F_1^\beta.$$

The above formula can be simplified as follows

$$q_{j,k-1}^{\alpha-1} F_j^\beta < q_{1,k-1}^{\alpha-1} F_1^\beta.$$

Owing to $q_{j,k-1} < q_{1,k-1}$ and $F_1 > F_j$, so if $\alpha \geq 1, \beta \geq 0$, then $q_{j,k-1}^{\alpha-1} F_j^\beta < q_{1,k-1}^{\alpha-1} F_1^\beta$.

Thus $\frac{1}{p_{1,k}} - \frac{1}{p_{1,k-1}}$, that is $p_{1,k} > p_{1,k-1}$. □

Theorem 5 *shows that the average probability of the path AC_1B selected is larger and larger as time goes on.*

Theorem 6 *when $\alpha \geq 1, \beta \geq 0$, then* $\lim_{k\to\infty} p_{1k} = 1$

Proof According to $\frac{q_{j,k}}{q_{1,k}} < \frac{q_{j,k-1}}{q_{1,k-1}}$, we know that the value of $\frac{q_{j,k}}{q_{1,k}}$ gets smaller and smaller, thus we have $\frac{q_{j,k}}{q_{1,k}} \to \infty$, when $k \to \infty$.

$$\lim_{k\to\infty} p_{1k} = \lim_{k\to\infty}\left[1 + (\frac{F_2}{F_1})^\beta(\frac{q_{2k}}{q_{1k}})^\alpha + (\frac{F_3}{d_1})^\beta(\frac{q_{3k}}{q_{1k}})^\alpha + \cdots + (\frac{F_m}{F_1})^\beta(\frac{q_{mk}}{q_{1k}})^\alpha\right]^{-1} = 1. \square$$

It is worth noticing that, $\alpha \geq 1, \beta \geq 0$ is the sufficient condition but not the necessary condition. *Theorem 4* shows that as time goes on, the average probability of the path

AC_1B selected is close to 1. Thus, the convergence of the fidelity-based ACA is proved. The convergence of fidelity-based ACA with Q-learning is the same as that of fidelity-based ACA, because the difference only lies in the update mode of pheromone which does not effect the convergence of the algorithm.

6 Conclusion

In this paper, a fidelity-based ACA is presented for the control design of quantum system. To improve the performance of fidelity-based ACA, a fidelity-based ACA with Q-learning is introduced. In this improved algorithm, the fidelity information can be extracted from the system structure or the system behavior. The aim is to design a good exploration strategy for a better tradeoff between exploration and exploitation, and to speed up the convergence as well. Experimental results show that fidelity-based ACA with Q-learning is superior to the fidelity-based ACA. The control problems of a spin-(1/2) system is adopted to demonstrate the performance of the fidelity-based ACA with Q-learning. In the future, our work will focus on improving the fidelity-based ACA by combining with other algorithms.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 61379153, 61401519, 61572927).

References

1. Rabitz, H.: Algorithms for closed loop control of quantum dynamics. In: Proceedings of the IEEE conference on decision and control, vol. 1, pp. 937–942 (2000)
2. Korotkov, A.N.: Continuous quantum measurement of a double dot[J]. Phys. Rev. B **60**(8) (1999)
3. Judson, R.S., Rabitz, H.: Teaching lasers to control molecules. Phys. Rev. Lett. **68.10**, 1500–1503 (1992)
4. Besten, M.D., Stutzle, T.: An ant colony optimization application to the single machine total weighted tardiness problem. In: Proceedings of ants (2001)
5. Hao, X., Liu, L., Wu, Y.: Positive solutions for nonlinear fractional semipositone differential equation with nonlocal boundary conditions. J. Nonlinear Sci. Appl. **9**(6), 3992–4002 (2016)
6. Blum, C.: Beam-ACO-hybridizing ant colony optimization with beam search: an application to open shop scheduling. Comput. Oper. Res. **32.6**, 1565–1591 (2005)
7. Sttzle, T.: Ant colony optimization. Alphascript Publishing **28.3**, 1155–1173 (2010)
8. Xu, F., Zhang, X., Wu, Y., Liu, L.: The optimal convergence rates for the multi-dimensional compressible viscoelastic flows. Z. Angew. Math. Mech. **96**(12), 1490–1504 (2016)
9. Di Caro, G.: Ant Colony Optimization and its Application to Adaptive Routing in Telecommunication Networks[J]. Faculte Desences Appliquees (2004)
10. Korb, O., Sttzle, T., Exner, T.E.: PLANTS: application of ant colony optimization to structure-based drug design. ant colony optimization and swarm intelligence, pp. 247–258. Springer, Berlin (2006)
11. Chen, C. et al.: Fidelity-based probabilistic q-learning for control of quantum systems. IEEE Transactions on Neural Networks & Learning Systems **25.5**, 920–933 (2014)
12. Wang, L., Niu, Q., Fei, M.: A novel quantum ant colony optimization algorithm and its application to fault diagnosis. Trans. Inst. Meas. Control. **30.4**, 313–329 (2008)
13. Li, W., Yin, Q., Zhang, X.: Continuous quantum ant colony optimization and its application to optimization and analysis of induction motor structure. In: 2010 IEEE Fifth international conference on Bio-inspired computing: theories and applications (BIC-TA), pp. 313–317. IEEE, Piscataway (2010)
14. Li, P., Song, K., Yang, E.: Quantum ant colony optimization with application. In: 2010 sixth international conference on natural computation (ICNC), pp. 2989–2993. IEEE, Piscataway (2010)
15. Chen, C.L., Dong, D.Y., Chen, Z.H.: Quantum computation for action selection using reinforcement learning. International Journal of Quantum Information **4.6**, 1071–1083 (2011)
16. Dong, D., Chen, C., Chen, Z.: Quantum reinforcement learning. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society **38.5**, 1207–1220 (2008)

17. Deng, M., Inoue, A., Kawakami, S.: Optimal path planning for material and products transfer in steel works using ACO. In: 2011 international conference on advanced mechatronic systems (ICAMechS), pp. 47–50. IEEE, Piscataway (2011)
18. Paul, T.: Quantum computation and quantum information. Am. J. Phys. **70.5**, 558–559 (2000)
19. Wiseman, H.M., Milburn, G.J.: Quantum measurement and control. In: Wiseman, H.M., Milburn, G.J. (eds.) Quantum Measurement & Control, vol. 11.1, pp. 313–315. Cambridge University Press, Cambridge (2010)
20. Dong, D., Petersen, I.R.: Sliding mode control of two-level quantum systems. Automatica A Journal of Ifac the International Federation of Automatic Control **48.5**, 725–735 (2012)
21. Chen, C. et al.: Control design of uncertain quantum systems with fuzzy estimators. IEEE Trans. Fuzzy Syst. **20.5**, 820–831 (2012)
22. Dong, D., Petersen, I.R.: Quantum control theory and applications: a survey. IET Control Theory & amp; Applications **4.12**, 2651–2671 (2010)
23. Altafini, C., Ticozzi, F.: Modeling and control of quantum systems: an introduction. IEEE Trans. Autom. Control **57.8**, 1898–1917 (2012)
24. Qi, B., Pan, H., Guo, L.: Further results on stabilizing control of quantum systems. IEEE Trans. Autom. Control **58**, 1349–1354 (2013)
25. Dong, D., Petersen, I.R., Rabitz, H.: Sampled-data design for robust control of a single qubit. IEEE Trans. Autom. Control **58.10**, 2654–2659 (2013)
26. Cozzini, M., Ionicioiu, R., Zanardi, P.: Quantum fidelity and quantum phase transitions in matrix product states. Phys. Rev. B **76.10**, 3398–3407 (2006)
27. Rabitz, H.A., Hsieh, M.M., Rosenthal, C.M.: Quantum optimally controlled transition landscapes. Science **303.7**, 1998–2001 (2004)
28. Chakrabarti, R., Rabitz, H.: Quantum control landscapes. Int. Rev. Phys. Chem. **26.4**, 671–735 (2007)