

# Learning-Based Quantum Control for Optimal Pure State Manipulation

**DOI:**
10.1109/LCSYS.2024.3409671

**Document Version**
Accepted author manuscript

# Learning-Based Quantum Control for Optimal Pure State Manipulation

Anthony Siming Chen[1], *Member, IEEE*, Guido Herrmann[1], *Senior Member, IEEE*, Kyriakos G. Vamvoudakis[2], *Senior Member, IEEE*, Jayadev Vijayan[3]

***Abstract*— In this paper, we propose an adaptive critic learning approach for two classes of optimal pure state transition problems for closed quantum systems: i) when the target state is an eigenstate, and ii) when the target state is a superposition pure state. First, we describe a finite-dimensional quantum system based on the Schrodinger equation with the action of control fields. Then, we consider the target state to be i) an eigenstate of the internal Hamiltonian and ii) an arbitrary pure state via a unitary transformation. Meanwhile, the quantum state manipulation is formulated as an optimal control problem for solving the complex partial differential Hamilton-Jacobi-Bellman (HJB) equation, of which the control solution is found using continuous-time Q-learning of an adaptive critic. Finally, numerical simulation for a spin-1/2 particle system demonstrates the effectiveness of the proposed approach.**

***Index Terms*— Adaptive optimal control, quantum control, Q-learning, Schrödinger equation.**

## I. Introduction

In the rapidly-growing field of quantum control, the pursuit of state manipulation over quantum systems has emerged as a fundamental task with far-reaching implications [1], [2]. For closed quantum systems, Lyapunov control methods have been extensively studied [3]–[6], while optimal control techniques have found useful application in governing quantum phenomena within physical chemistry [7]–[10]. Recently, of particular significance is the quest for *model-free* reinforcement-learning-based methods [11]–[13] in the field of quantum control. Reinforcement learning or approximate dynamic programming (ADP) addresses the problem of how an agent/controller can learn to approximate an optimal policy/control while interacting with its environment. Among the ADP algorithms [14]–[17], Q-learning is a *model-free* feedback-based approach and works well even when the system model is unknown or with uncertainties [18]–[20].

This work was supported in part, by NSF under grant Nos. CAREER CPS-1851588, CPS-2227185, and SATC-1801611. J. Vijayan acknowledges support from the Dame Kathleen Ollerenshaw Fellowship scheme of the University of Manchester.

[1]Anthony Siming Chen and Guido Herrmann are with the Control Systems and Robotics Group at the Department of Electrical and Electronic Engineering, The University of Manchester, Manchester, M13 9PL, United Kingdom (e-mail: siming.chen@manchester.ac.uk, guido.herrmann@manchester.ac.uk).

[2]Kyriakos G. Vamvoudakis is with the Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332, United States (e-mail: kyriakos@gatech.edu).

[3]Jayadev Vijayan is with the Photon Science Institute, Department of Electrical and Electronic Engineering, The University of Manchester, Manchester M13 9PY, UK (e-mail: jayadev.vijayan@manchester.ac.uk).

This paper proposes an adaptive critic learning approach for pure state manipulation of closed-quantum systems. The novelty and contributions of this work include:

1) Two different formulations of the Schrödinger equation with the action of control fields are summarized in simple forms to allow optimal pure state control design for a closed quantum system.
2) A *complex* partial differential HJB equation is developed as a necessary and sufficient condition for the optimal control of pure state manipulation.
3) A continuous-time Q-learning theory [20] is *refined* and *applied* to form an adaptive critic for solving the quantum optimal control problem online in a *model-free* manner with validated simulation results.

The motivation of this work is to develop a model-free reinforcement learning approach for quantum pure state manipulation. This paper is organized as follows. Section II provides the background on pure state manipulation. Section III formulates the optimal control problem as solving the HJB equation. Section IV presents the control design via Q-learning. Section V demonstrates the simulated controller.

## II. Background

Here, we formulate the closed quantum system with action of control fields based on the Schrödinger equation for two cases of pure state manipulation: i) when the target state is an eigenstate, and ii) when the target state is an arbitrary pure state, which is any superposition of the eigenstates.

### A. Schrödinger Equation of Closed Quantum Systems

The quantum states of particles at time $t$ can be described through the time dependent Schrödinger equation, which provides analytical solutions that precisely determine the temporal evolution of the state:

$$i\hbar\frac{\partial}{\partial t}\left|\Psi(t)\right\rangle = \hat{H}_0\left|\Psi(t)\right\rangle, \quad \left|\Psi(0)\right\rangle = \left|\Psi_0\right\rangle, \tag{1}$$

where the quantum state $\left|\Psi(t)\right\rangle \in \mathbb{C}^n$ is a wave function in a $n$-dimensional Hilbert space; $\hat{H}_0 \in \mathbb{C}^{n\times n}$ is an observable, the internal or unperturbed Hamiltonian operator of the system; $\hbar = 1.05457\times10^{-34}\ m^2kg/s$ is the reduced Planck's constant (see [21] for Dirac's bra-ket notation and a comprehensive introduction to quantum mechanics).

With the action of external control field $u_k(t) \in \mathbb{R}$, the closed quantum system can be written as

$$i\hbar\frac{\partial}{\partial t}\left|\Psi(t)\right\rangle = (\hat{H}_0 + \hat{H}_c(t))\left|\Psi(t)\right\rangle, \tag{2}$$

with $\hat{H}_c(t) = \Sigma_{k=1}^{m} \hat{H}_k u_k(t)$ being the external/control Hamilonian for $k = 1, 2, \cdots, m$ and $m \in \mathbb{N}^+$. Note that the Schrödinger equation (2) is a bilinear model. For simplicity in theoretical analysis, $\hbar$ is often grouped into the Hamiltonian or set to be 1. Therefore, considering $\hat{H} = \hat{H}_0 + \hat{H}_c$, we can rearrange (2) into a compact form:

$$i|\dot{\Psi}\rangle = \hat{H}\,|\Psi\rangle\,. \tag{3}$$

**Assumption 1** *The system* (2) *or* (3) *is a finite-dimensional quantum system, i.e.,* $n \in \mathbb{N}^+ < \infty$. *Both* $\hat{H}_0$ *and* $\hat{H}_k$ *are linear Hermitian operators independent of time* $t$, *i.e.,* $\hat{H}_0^\dagger = \hat{H}_0$ *and* $\hat{H}_k^\dagger = \hat{H}_k$. *The external control fields* $u_k(t)$ *is a real, scalar, realizable function. The Hamiltonians are non-degenerate throughout this paper, i.e., all of the eigenvalues are distinct (See [22] for treatment to degenerate cases).* ◇

### B. Eigenstate Manipulation: Fictitious Control $\omega$ Approach

For quantum control in quantum chemistry, the target state is usually an eigenstate of the internal Hamiltonian $\hat{H}_0$, i.e., the eigen-equation of $\hat{H}_0$ is given by

$$\hat{H}_0\,|\Psi_k\rangle = \lambda_k\,|\Psi_k\rangle\,, \tag{4}$$

with $\lambda_k$ is the eigenvalue of $\hat{H}_0$ corresponding to the eigenstate $|\Psi_k\rangle$. The set of eigenstates $|\Psi_k\rangle$ associated with a particular observable forms a basis for the Hilbert space, hence the following equation holds:

$$|\Psi\rangle = \Sigma_n c_n\,|\Psi_k\rangle\,, \tag{5}$$

where $c_n \in \mathbb{C}$ and $\Sigma_n |c_n| = 1$. $|c_n|$ represents the quantum probability in the eigenstate $|\Psi_k\rangle$, therefore $|\Psi\rangle$ lives on the unit sphere of $\mathbb{C}$. The probability of the state $|\Psi\rangle$ is equal to the probability of the state $e^{i\theta(t)}\,|\Psi\rangle$. In reality, $|\Psi\rangle$ and $e^{i\theta(t)}\,|\Psi\rangle$ describe the same physical state for any global phase factor $\theta(t) \in \mathbb{R}$, which is a time-dependent real function. Inspired by this non-trivial geometry [3], we introduce a second control $\omega(t)$ corresponding to $\dot{\theta}(t)$ :

$$i\frac{\partial}{\partial t}\,|\Psi(t)\rangle = (\hat{H}_0 + \Sigma_{k=1}^{m}\hat{H}_k u_k(t) + \omega(t)I)\,|\Psi(t)\rangle\,, \tag{6}$$

with the new control $\omega(t) \in \mathbb{R}$ that functions as a gauge degree of freedom, offering flexibility in its selection without altering the physical quantities associated to $|\Psi(t)\rangle$. With such additional fictitious control $\omega$, the solution of (6) is equal to the solution of (2) multiplied by a global phase factor $e^{-i\omega t}$. This formulation is advantageous in analyzing eigenstate manipulation. For example, assume that the target state $|\Psi_d\rangle$ is an eigenstate and satisfies

$$\hat{H}_0\,|\Psi_d\rangle = \lambda_d\,|\Psi_d\rangle \tag{7}$$

and substituting $\omega = -\lambda_d$, we have

$$(\hat{H}_0 + \omega(t)I)\,|\Psi(t)\rangle = 0. \tag{8}$$

### C. Pure State Manipulation: Unitary Operator U Approach

The above formulation is obtained under the assumption that the target state is an eigenstate, this appears in the case of population transfer in chemical reactions. However, the target state is often an arbitrary pure state, a superposition state, in physical application fields. Inspired by the unitary evolution operator [7] (also known as *complete control* in [23]) and the general-complex *gauge transformation* in [24], we show how to apply the macroscopic field to the pure-state quantum system via a unitary operator $U(t)$. Consider the quantum system (2) or (3), to make the system state reach the arbitrary pure state, we need to eliminate the drift term, i.e., the internal Hamiltonian $\hat{H}_0$. Expressing the internal Hamiltonian in terms of its eigenvalue:

$$\hat{H}_0 = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n), \tag{9}$$

we define a unitary operator

$$U(t) = \mathrm{diag}(e^{-i\lambda_1 t}, e^{-i\lambda_2 t}, \cdots, e^{-i\lambda_n t}). \tag{10}$$

Substituting the unitary transformation with the transformed state $|\Psi^u(t)\rangle$:

$$|\Psi(t)\rangle = U(t)\,|\Psi^u(t)\rangle\,, \tag{11}$$

into the (2), we obtain

$$i\frac{\partial}{\partial t}\,|\Psi^u\rangle = (U^\dagger\hat{H}_0 U + \Sigma_{k=1}^{m} U^\dagger\hat{H}_k U u_k - iU^\dagger\dot{U})\,|\Psi^u\rangle\,. \tag{12}$$

Since $U^\dagger\hat{H}_0 U = \hat{H}_0$ and $-iU^\dagger\dot{U} = \hat{H}_0$, the drift term $\hat{H}_0$ is eliminated and we have

$$i\frac{\partial}{\partial t}\,|\Psi^u(t)\rangle = (\Sigma_{k=1}^{m}\hat{H}_k^u(t)u_k(t))\,|\Psi^u(t)\rangle\,, \tag{13}$$

with $\hat{H}_k^u(t) = U^\dagger(t)\hat{H}_k U(t)$. Hence, similar to (5), the following equation holds:

$$|\Psi^u\rangle = \Sigma_n e^{i\lambda_n t}c_n\,|\Psi_k^u\rangle\,, \tag{14}$$

with $|e^{i\lambda_n t}c_n| = |c_n|$.

**Problem Formulation** *Finding a set of real, scalar, realizable control* $u_k$ *with* $k = 1, 2, \cdots, m$ *that manipulates the state,* $|\Psi(t)\rangle$ *of* (6) *or* $|\Psi^u(t)\rangle$ *of* (13), *to a target state* $|\Psi_d\rangle$ *by minimising an infinite-horizon cost* $V(|\Psi\rangle, |\Psi_d\rangle)$. ◇

**Remark 1** *The two formulations above:* (6) *and* (13), *are suitable for different use cases. The fictitious control approach can adjust the global phase of the controlled state using an imaginary real control* $\omega(t)$, *while the unitary operator approach can change the local phase by applying* $U(t)$ *that eliminates the drift term* $\hat{H}_0$. *Both approaches do not change the probability distribution of the system. The unitary operator approach* (13) *tends to be more complex when implementing the control law because* $\hat{H}_k^u(t)$ *is time-dependent while a time-invariant* $\hat{H}_k$ *is used in* (6). ◇

## III. HJB Equation

For quantum systems (6), (13) and a target state $|\Psi_d\rangle$, the objective is to minimize an infinite-horizon integral cost:

$$V(|\Psi\rangle, |\Psi_d\rangle, u_k) = \int_t^\infty r_{\mathrm{utl}}(|\Psi\rangle, |\Psi_d\rangle, u_k)\mathrm{d}\tau, \tag{15}$$

This article has been accepted for publication in IEEE Control Systems Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2024.3409671

by finding the optimal control $u_k(t) = u_k^\star$ such that

$$u_k^\star(|\Psi\rangle, |\Psi_d\rangle) = \arg\min_{u_k} V(|\Psi\rangle, |\Psi_d\rangle). \tag{16}$$

For the system (6), the *utility* (known as the *reward* in reinforcement learning) of the bilinear problem is selected as quartic on state and quadratic on control [6, Eqn. (4.25)]:

$$r_{\text{utl}} = \Sigma_{k=1}^m \frac{1}{r_k}[\langle\Psi(t) - \Psi_d| S\hat{H}_k |\Psi(t)\rangle]^2 + \langle\mu| R |\mu\rangle, \tag{17}$$

with a control vector $\mu = [u_1, u_2, \cdots, u_m]^\mathsf{T}$, a positive definite matrix $S$ and a positive definite $R = \text{diag}(r_1, r_2, \cdots, r_m)$. $S$ and $R$ are the reward weights and should be chosen to effectively balance the trade-offs between state performance and control efforts. Note that all $r_k$ must be strictly positive (the control field is penalized so that an optimal solution exists). For system (13), the utility function is by replacing $\hat{H}_k$ with $\hat{H}_k^u(t)$ in (17). In the following discussion, we will not distinguish $\hat{H}_k$ and $\hat{H}_k^u(t)$ for the sake of simplicity.

The optimal control problem can be addressed by formulating a general solution represented by a partial differential equation governing the optimal cost function, denoted as $V^\star(|\Psi\rangle, |\Psi_d\rangle) = \min_{u_k(t)} V(|\Psi\rangle, |\Psi_d\rangle)$. Define the Hamiltonian of the problem as

$$\begin{aligned}\mathcal{H}(|\Psi\rangle, |\Psi_d\rangle, u_k, \nabla V_\Psi^\star) :=& \\ r_{\text{utl}}(|\Psi\rangle, |\Psi_d\rangle, u_k) &+ \langle\nabla V_\Psi^\star| i\frac{\partial}{\partial t} |\Psi\rangle,\end{aligned} \tag{18}$$

with the gradient vector $\nabla V_\Psi^\star = \partial V^\star / \partial |\Psi\rangle$. The optimal cost function $V^\star(|\Psi\rangle, |\Psi_d\rangle)$ satisfies the *Hamilton-Jacobi-Bellman* (HJB) equation

$$0 = \nabla V_t^\star + \min_{u_k} \mathcal{H}(|\Psi\rangle, |\Psi_d\rangle, u_k, \nabla V_\Psi^\star), \tag{19}$$

where $\nabla V_t^\star = \partial V^\star / \partial t = 0$ as the optimal cost is not an explicit function of time. The optimal control $u_k^\star$ can be determined by setting

$$\frac{\partial \mathcal{H}(|\Psi\rangle, |\Psi_d\rangle, u_k, \nabla V_\Psi^\star)}{\partial u_k} = 0, \tag{20}$$

so that

$$u_k^\star = -\frac{1}{2r_k} \langle\nabla V_\Psi^\star| \hat{H}_k |\Psi\rangle, \tag{21}$$

for $k = 1, 2, \cdots, m$ and $m \in \mathbb{N}^+$. We can conclude that the optimal control $u_k^\star$ can be determined by solving the HJB equation (19) for $\nabla V_\Psi^\star$.

While reinforcement learning is traditionally known for its model-free features, it can be effectively combined with models to tackle challenging optimal control problems such as solving the HJB equation in control theory. To prepare the formulation of optimal control for the Q-learning approach later, we can rewrite the optimal control (24) by separating the real part and the imaginary part of the coefficient matrices and state variables. Let $|\Psi\rangle = [\psi_1 + i\psi_{n+1}, \psi_2 + i\psi_{n+2}, \cdots, \psi_n + i\psi_{2n}]^\mathsf{T}$, the system (6) becomes

$$\begin{aligned}\dot{\psi}(t) =& \Big(\begin{bmatrix} \Im(\hat{H}_0 + \omega I) & \Re(\hat{H}_0 + \omega I) \\ -\Re(\hat{H}_0 + \omega I) & \Im(\hat{H}_0 + \omega I) \end{bmatrix} \\ &+ \Sigma_{k=1}^m \begin{bmatrix} \Im(\hat{H}_k) & \Re(\hat{H}_k) \\ -\Re(\hat{H}_k) & \Im(\hat{H}_k) \end{bmatrix} u_k(t)\Big)\psi(t)\end{aligned} \tag{22}$$

or simply

$$\dot{\psi}(t) = (F + \Sigma_{k=1}^m G_k u_k(t))\psi(t). \tag{23}$$

Then we have

$$u_k^\star = -\frac{1}{2r_k}(\psi - \psi_d)^\mathsf{T} S G_k \psi \tag{24}$$

and again, one can obtain the optimal control for the system (13) by replacing $\hat{H}_k$ with $\hat{H}_k^u(t)$ in $G_k$ in (24).

**Remark 2** *Note that the Hamiltonian $\mathcal{H}$ (18) is in a different context from the previous Hamiltonian $\hat{H}$ in the Schrödinger equation (3). $\hat{H}$ is an observable while $\mathcal{H}$ is an instantaneous increment of the Lagrangian of the optimal control problem that is to be optimized. Both Hamiltonians are inspired by but distinct from the Hamiltonian of classical mechanics. Moreover, note that $R$ needs to be strictly positive definite to avoid the singularity in (24).* ◇

## IV. Quantum Control Design

This section presents the design of the optimal quantum controller via continuous-time Q-learning to solve the *HJB equation* (19) in real-time.

### A. Quantum Q-function

We refine the theory in [20] regarding the continuous-time Q-learning algorithm and then extend it to the quantum control problem. Define an action-dependent version of cost function $Q(|\Psi\rangle, |\Psi_d\rangle, u_k)$ such that $Q^\star(|\Psi\rangle, |\Psi_d\rangle, u_k^\star) = V^\star(|\Psi\rangle, |\Psi_d\rangle)$. For the quantum systems (6) and (13), a *Q-function* can be explicitly defined by adding the *right-hand side* of a *generalized-HJB* (GHJB) equation:

$$0 = \nabla V_t + \mathcal{H}(|\Psi\rangle, |\Psi_d\rangle, u_k, \nabla V_\Psi), \tag{25}$$

for $\nabla V_t^\star$ and $\nabla V_\Psi^\star$, onto the optimal cost $V^\star$ as

$$Q(|\Psi\rangle, |\Psi_d\rangle, u_k) := V^\star + \nabla V_t^\star + \mathcal{H} \tag{26}$$

which can be approximated via parameterization [20].

**Lemma 1** *The Q-function defined in (26) is positive definite with the optimization $Q^\star(|\Psi\rangle, |\Psi_d\rangle, u_k^\star) = \min_{u_k} Q(|\Psi\rangle, |\Psi_d\rangle, u_k)$. The optimal Q-function $Q^\star(|\Psi\rangle, |\Psi_d\rangle, u_k^\star)$ has the same optimal value as $V^\star(|\Psi\rangle, |\Psi_d\rangle)$ for the cost $V(|\Psi\rangle, |\Psi_d\rangle)$ defined in (15), i.e., $Q^\star(|\Psi\rangle, |\Psi_d\rangle, u_k^\star) = V^\star(|\Psi\rangle, |\Psi_d\rangle)$ when applying the optimal control $u_k^\star$.* ◇

**Proof.** Refer to [20, Lemma 3] for a similar proof using the HJB equation (19). □

**Remark 3** *We refined the theory in our earlier work [20] in terms of justifying the definition of a Q-function. In this paper, we add the right-hand side of the GHJB equation (25) for $\nabla V_t^\star$ and $\nabla V_\Psi^\star$ to the optimal cost instead of adding only the Hamitonian as per [20]. Though the two definitions are equivalent for $\nabla V_t^\star = 0$, the definition here is more proper by invoking the instrumental lemma from [25, p.441]. The GHJB equation gives the cost of an arbitrary control. It shows that the Hamiltonian is quadratic in the control deviation from the optimal control, so the Q-function contains such quantified deviation.* ◇

This article has been accepted for publication in IEEE Control Systems Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2024.3409671

### B. Adaptive Critic Learning

We approximate the Q-function (26) using a critic network by

$$Q(|\Psi\rangle, |\Psi_d\rangle, u_k) = W^{\mathsf{T}}\Phi(|\Psi\rangle, |\Psi_d\rangle, u_k) + \varepsilon, \tag{27}$$

where $\Phi(|\Psi\rangle, |\Psi_d\rangle, u_k)$ denotes the activation function vector with the number $N$ of nodes in the hidden layer; $W$ is the weight vector; $\varepsilon(|\Psi\rangle, |\Psi_d\rangle, u_k)$ is the network approximation error; and $W^{\mathsf{T}}\Phi(|\Psi\rangle, |\Psi_d\rangle, u_k)$ can be explicitly expressed according to three components $F_\Psi(|\Psi\rangle, |\Psi_d\rangle)$, $F_{\Psi u}(|\Psi\rangle, |\Psi_d\rangle, u_k)$, and $F_u(u_k)$ in (26) as

$$W^{\mathsf{T}}\Phi = \begin{bmatrix} W_\Psi^{\mathsf{T}} & W_{\Psi u}^{\mathsf{T}} & r_k \end{bmatrix} \begin{bmatrix} \Phi_\Psi(|\Psi\rangle, |\Psi_d\rangle) \\ \Phi_{\Psi u}(|\Psi\rangle, |\Psi_d\rangle)u_k \\ \Phi_u(u_k) \end{bmatrix}. \tag{28}$$

The regressor $\Phi(|\Psi\rangle, |\Psi_d\rangle, u_k)$ is selected to provide a complete independent basis such that $Q(|\Psi\rangle, |\Psi_d\rangle, u_k)$ is uniformly bounded, e.g., $\Phi$ can be chosen as the power series (polynomial) or radial basis functions (e.g., sigmoid, tanh) of the signals $|\Psi\rangle, |\Psi_d\rangle, u_k$ [25]. Recalling from the Weierstrass higher-order approximation theorem [26], the approximation error $\varepsilon(|\Psi\rangle, |\Psi_d\rangle, u_k)$ is bounded for a fixed $N$ and as the number of nodes $N \to \infty$, we have $\varepsilon(|\Psi\rangle, |\Psi_d\rangle, u_k) \to 0$.

Now we derive the Bellman equation in terms of the Q-function to update the critic. Consider the Bellman's principle of optimality [15], the instrumental Lemma [25, p.441], and the *Lemma 1* above, we can derive

$$\begin{aligned} &\overbrace{-\int_{t-T}^{t} r_{\text{utl}}(|\Psi\rangle, |\Psi_d\rangle, u_k)\mathrm{d}\tau}^{-\rho(|\Psi\rangle, |\Psi_d\rangle, u_k)} = Q^\star(|\Psi(t)\rangle, u_k^\star(t)) \\ &\qquad - Q^\star(|\Psi(t-T)\rangle, u_k^\star(t-T)) + \xi(t) \\ &= \underbrace{W^{\mathsf{T}}\Phi(|\Psi(t)\rangle, u_k^\star(t)) - W^{\mathsf{T}}\Phi(|\Psi(t-T)\rangle, u_k^\star(t-T))}_{W^{\mathsf{T}}\Delta\Phi(|\Psi\rangle, |\Psi_d\rangle, u_k^\star)} \\ &\quad + \underbrace{\Delta\varepsilon + \xi(t-T,t)}_{\varepsilon_B}, \end{aligned} \tag{29}$$

with $\Delta\varepsilon$ being the residual network approximation

$$\Delta\varepsilon := \varepsilon(|\Psi(t)\rangle, u_k^\star(t)) - \varepsilon(|\Psi(t-T)\rangle, u_k^\star(t-T)) \tag{30}$$

and $\xi(t-T,t)$ being a *residual control error* as

$$\begin{aligned} \xi(t-T,t) := &-\int_{t-T}^{t} r_k(u_k(\tau) - u_k^\star(\tau))^2\mathrm{d}\tau \\ &+ r_k(u_k(t-T) - u_k^\star(t-T))^2 \\ &- r_k(u_k(t) - u_k^\star(t))^2, \end{aligned} \tag{31}$$

the integral reinforcement $\rho(|\Psi\rangle, |\Psi_d\rangle, u_k)$, the regressor difference $\Delta\Phi(t) = \Phi(|\Psi(t)\rangle, u_k^\star(t)) - \Phi(|\Psi(t-T)\rangle, u_k^\star(t-T))$, and the Bellman error $\varepsilon_B = \Delta\varepsilon + \xi$ with $\Delta\varepsilon = \varepsilon(|\Psi(t)\rangle, u_k^\star(t)) - \varepsilon(|\Psi(t-T)\rangle, u_k^\star(t-T))$ being bounded for bounded $\varepsilon$. The Bellman equation (29) forms the basis for adaptive critic design.

**Remark 4** *The Bellman equation* (29) *is different from [20] as it also considers the residual control error $\xi(t-T,t)$ due to the difference in values between the current control $u_k$ and its optimal control $u_k^\star$ for $\rho(t-T,t)$, $Q(t-T)$, and $Q(t)$. The uniform ultimate boundedness of such residual error $\psi(t-T,t)$ is proved later in the main theorem by using the delay-dependent Lyapunov functions.* ◇

Define two auxiliary variables $\mathcal{P}(t)$ and $\mathcal{Q}(t)$ by low-pass filtering the variables in (29) as

$$\begin{cases} \dot{\mathcal{P}}(t) = -\ell\mathcal{P}(t) + \Delta\Phi(t)\Delta\Phi(t)^{\mathsf{T}}, & \mathcal{P}(0) = 0, \\ \dot{\mathcal{Q}}(t) = -\ell\mathcal{Q}(t) + \Delta\Phi(t)\rho, & \mathcal{Q}(0) = 0, \end{cases} \tag{32}$$

with a filter parameter $\ell > 0$. The critic network can be written as

$$\hat{Q}(|\Psi\rangle, |\Psi_d\rangle, u) = \hat{W}^{\mathsf{T}}\Phi(|\Psi\rangle, |\Psi_d\rangle, u_k), \tag{33}$$

where $\hat{W}$ and $\hat{Q}(|\Psi\rangle, |\Psi_d\rangle, u_k)$ denote the current estimate of $W$ and $Q(|\Psi\rangle, |\Psi_d\rangle, u_k)$, respectively.

We design the adaptation law to update $\hat{W}$ such that

$$\dot{\hat{W}}(t) = -\Gamma(\mathcal{P}(t)\hat{W}(t) + \mathcal{Q}(t)), \tag{34}$$

where the positive-definite diagonal matrix $\Gamma$ is an adaptive learning gain.

### C. Quantum Control Synthesis

We reconstruct the optimal control $u_k^\star$ from (24) based on the parameterization of $Q(|\Psi\rangle, |\Psi_d\rangle, u_k)$ such that

$$u_k^\star = -\frac{1}{2r_k}W_{\Psi u}^{\mathsf{T}}\Phi_{\Psi u} + \varepsilon_u, \tag{35}$$

where $\varepsilon_u$ is a bounded approximation error due to $\varepsilon$, $W_{\Psi u}^{\mathsf{T}}\Phi_{\Psi u}$ accounts for the term $(\psi - \psi_d)^{\mathsf{T}}SG_k\psi$. One can determine the optimal control directly using the adaptive critic (33) if the weight $\hat{W}$ converges to the actual weight $W$. Therefore, the control command (actor) is

$$u_k = -\frac{1}{2r_k}\hat{W}_{\Psi u}^{\mathsf{T}}\Phi_{\Psi u}. \tag{36}$$

The main result of the learning-based quantum control design is summarized in the following theorem.

**Theorem 1** *Given the quantum systems* (6) *and* (13) *with the cost* (15) *and Q-function* (26)*, if the regressor signal $\Delta\Phi(t)$ is persistently excited*[1]*, the adaptive critic network* (33) *with the adaptation law* (34) *leads to a learning-based controller* (36) *so that the adaptive critic weight estimation error $\tilde{W} = \hat{W} - W$ and the quantum state $|\Psi\rangle$ will converge to a compact set around $|\Psi_d\rangle$ and the controller (actor) $u_k$ will converge to a small bounded set around its optimal control value $u_k^\star$. Moreover, in the extreme case that there is no network approximation error, i.e., $\varepsilon = 0$, then the convergence is exponential.* ◇

**Proof.** To begin with, we analyze the effect of the Bellman error $\varepsilon_B$ over the low-pass filtering dynamics (32). It can be deducted that $\mathcal{P}\hat{W} + \mathcal{Q} = \mathcal{P}\tilde{W} - (\mathcal{P}W + \Lambda) = \mathcal{P}\tilde{W} - \Lambda$. This leads to $\Lambda(t) = -\int_0^t e^{-\ell(t-\tau)}\Delta\Phi(\tau)\varepsilon_B(\tau)\mathrm{d}\tau$, $\Lambda(0) = 0$,

[1]The *persistency of excitation* (PE) condition is a common assumption for convergence of the parameter estimates in adaptive control. This can be verified online by checking the minimum eigenvalue of the auxiliary matrix $\mathcal{P}(t)$ being strictly positive as shown in [20, Lemma 1].

This article has been accepted for publication in IEEE Control Systems Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCSYS.2024.3409671

where $\varepsilon_B = \Delta\varepsilon + \xi$. For simplicity, we split the term into two parts written as $\Lambda(t) = \xi_1 + \xi_2$ with

$$\xi_1(t) = -\int_0^t e^{-\ell(t-\tau)}\Delta\Phi(\tau)\Delta\varepsilon(\tau)\mathrm{d}\tau,\ \xi_1(0) = 0, \quad (37)$$

$$\xi_2(t) = -\int_0^t e^{-\ell(t-\tau)}\Delta\Phi(\tau)\xi(\tau)\mathrm{d}\tau, \quad \xi_2(0) = 0. \quad (38)$$

Using the idea of delay-dependent stability [27], we design a Lyapunov function candidate $\mathscr{L}$ as

$$\mathscr{L} = \mathcal{L}_1 + k_2\mathcal{L}_2 + k_3\mathcal{L}_3 + k_4\mathcal{L}_4 + k_5\mathcal{L}_5 + k_6\mathcal{L}_6, \quad (39)$$

where sub-Lyapunov functions $\mathcal{L}_1 = \frac{1}{2}\tilde{W}^\mathsf{T}\Gamma^{-1}\tilde{W}$, $\mathcal{L}_2 = Q^\star(|\Psi\rangle, |\Psi_d\rangle, u_k^\star)$, $\mathcal{L}_3 = \frac{1}{2}\xi_1^\mathsf{T}\xi_1$, $\mathcal{L}_4 = \frac{1}{2}\xi_2^\mathsf{T}\xi_2$, $\mathcal{L}_5 = \int_{-T}^0 \tilde{W}^\mathsf{T}(t+\tau)\tilde{W}(t+\tau)\mathrm{d}\tau$, $\mathcal{L}_6 = \int_{-T}^0\int_{t+\theta}^t \tilde{W}^\mathsf{T}(\tau)\tilde{W}(\tau)\mathrm{d}\tau\mathrm{d}\theta$, and $k_2$, $k_3$, $k_4$, $k_5$, and $k_6$ are some positive constants. For the first term, considering the *Young's inequality* with $\eta$: $||a||||b|| \le \frac{1}{2\eta}||a||^2 + \frac{\eta}{2}||b||^2$ (valid for every $\eta > 0$) and the *PE condition*: $\lambda_{\min}(\mathcal{P}(t)) > \sigma > 0$, $\forall t \ge 0$, using (32)(34) and [20, Lemma 3], the derivative of $\mathcal{L}_1$ can be written as

$$\begin{aligned}\dot{\mathcal{L}}_1 &= \tilde{W}^\mathsf{T}\Gamma^{-1}\dot{\tilde{W}} = \tilde{W}^\mathsf{T}(\mathcal{P}\hat{W} + \mathcal{Q}) = -\tilde{W}^\mathsf{T}(\mathcal{P}\tilde{W} - \Lambda)\\ &\le -\sigma||\tilde{W}||^2 + ||\xi_1 + \xi_2||\ ||\tilde{W}||\\ &\le -(\sigma - \frac{1}{2\eta_1} - \frac{1}{2\eta_2})||\tilde{W}||^2 + \frac{\eta_1}{2}||\xi_1||^2 + \frac{\eta_2}{2}||\xi_2||^2,\end{aligned} \quad (40)$$

where $\eta_1 > 0$, $\eta_2 > 0$ are properly chosen constants such that $\sigma - \frac{1}{2\eta_1} - \frac{1}{2\eta_2} > 0$ holds. Similarly, analyzing the derivative of each term in $\mathscr{L}$ (39), the derivative $\dot{\mathscr{L}}$ can be written as

$$\begin{aligned}\dot{\mathscr{L}} \le &-\alpha_1||\tilde{W}(t)||^2 - \alpha_2||\xi_1||^2 - \alpha_3||\xi_2||^2 - \alpha_4\langle\Psi|\Psi\rangle\\ &- \alpha_5||\tilde{W}(t-T)||^2 - \alpha_6\int_{t-T}^t ||\tilde{W}(\tau)||^2\mathrm{d}\tau + \beta,\end{aligned} \quad (41)$$

where $\alpha_i$ $(i = 1, 2, ..., 6)$ are the positive scalars with properly chosen $k_i$ and $\eta_i$; $\beta$ is a bounded constant that characterizes the effect of the network approximation error $\varepsilon$. According to the Lyapunov theorem, $|\Psi(t)\rangle$, $\tilde{W}(t)$, $\xi_1(t)$, and $\xi_2(t)$ are uniformly ultimately bounded. Moreover,

$$|u_k - u_k^\star| \le \frac{1}{2r_k}||\Phi_{\Psi u}(x)||\ ||\tilde{W}_{\Psi u}|| + |\varepsilon_u|, \quad (42)$$

remains bounded. If $\varepsilon = 0$, we have $\beta = 0$ hence $|u_k - u_k^\star|$ will exponentially converge to zero. □

## V. Simulations

In this section, we apply the adaptive critic learning approach via numerical simulations to a two-level quantum system, i.e., a spin-1/2 particle, for two cases: (i) the target state is an eigenstate, and (ii) the target state is a superposition state. This example is representative because all the fermions: proton, neutron, electron, and quarks, have net spin-1/2, which may be used to constitute qubit to achieve necessary state manipulation in quantum communication and quantum computing.

The spin-1/2 particle with its spin in $\sigma_z$ is described by the following Schrödinger equation:

$$i\frac{\partial}{\partial t}|\Psi(t)\rangle = (\hat{H}_0 + \hat{H}_1 u_1(t))|\Psi(t)\rangle, \quad (43)$$

with the internal Hamiltonian $\hat{H}_0$ and the control Hamiltonian $\hat{H}_1$ (for $m = 1$) being

$$\hat{H}_0 = \sigma_z = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}, \quad \hat{H}_1 = \sigma_y = \begin{bmatrix}0 & -i\\ i & 0\end{bmatrix}. \quad (44)$$

Consider the two available basis states of the system as $|0\rangle$ and $|1\rangle$, the general state can be expressed as a superposition of these states with probability amplitude $c_1, c_2 \in \mathbb{C}$:

$$|\Psi\rangle = c_1|0\rangle + c_2|1\rangle, \quad (45)$$

with $|c_1|^2 + |c_2|^2 = 1$ for the states $|\Psi\rangle$ being normalised, i.e., pure states.

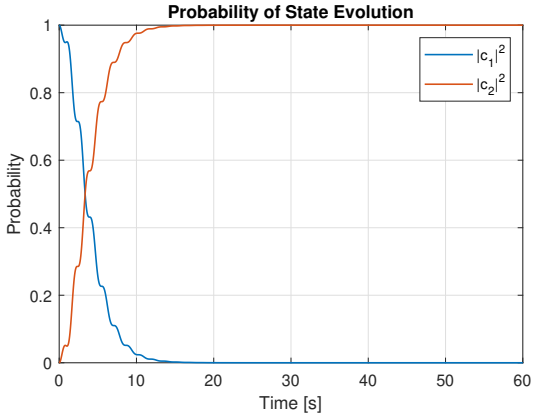


Fig. 1. Probability of states: Eigenstate manipulation.

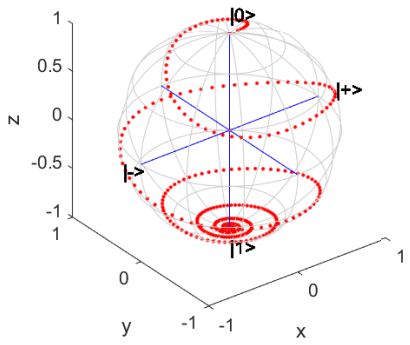


Fig. 2. Eigenstate manipulation on a Bloch sphere.

### A. Eigenstate Manipulation

Given that the target state is an eigenstate, we use the first formulation given in (6) of the quantum system with $\omega = -\lambda_d$ being a constant. Let the initial state of the system be $|\Psi_0\rangle = |0\rangle = [1,\ 0]^\mathsf{T}$ and the target state be $|\Psi_d\rangle = |1\rangle = [0,\ 1]^\mathsf{T}$. We can rewrite the quantum state by separating the real part and the imaginary part:

$$|\Psi(t)\rangle = \begin{bmatrix}\psi_1(t) + i\psi_3(t)\\ \psi_2(t) + i\psi_4(t)\end{bmatrix}. \quad (46)$$

Therefore, we have $\psi(0) = [1,\ 0,\ 0,\ 0]^\mathsf{T}$ and $\psi_d = [0,\ 1,\ 0,\ 0]^\mathsf{T}$, correspondingly, $|c_1|^2 = \psi_1^2 + \psi_3^2$ and $|c_2|^2 = \psi_2^2 + \psi_4^2$. The parameters are set as follows: the period $T = 0.5$ s, the filter parameter $\ell = 1$, the adaptive learning gain $\Gamma = 5I$, the regressor $\Phi$ is the power series of $(|\Psi\rangle, |\Psi_d\rangle, u_k)$

up to the fourth order. Fig. 1 presents the probability of the state evolution for eigenstate manipulation and Fig. 2 depicts its state evolution on a Bloch sphere. It is easily verified that $|c_1|^2 + |c_2|^2 = 1$ for any arbitrary time instance. The state reaches the target state around $t = 15$ s.

### B. Arbitrary Pure State Manipulation

When the target state is an arbitrary pure state, we use the second formulation (13) of the quantum system with a unitary operator $U(t)$. Let the initial state of the system be $|\Psi_0\rangle = |0\rangle = [1, \ 0]^\mathsf{T}$ and the target state be a superposition state $|\Psi_d\rangle = |+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$. We choose the unitary operator $U(t) = \text{diag}(e^{-it}, e^{it})$. Separating the real part and the imaginary part as (46), we have $\psi(0) = [1, \ 0, \ 0, \ 0]^\mathsf{T}$ and $\psi_d = [\frac{1}{\sqrt{2}}, \ \frac{1}{\sqrt{2}}, \ 0, \ 0]^\mathsf{T}$, correspondingly, $|c_1|^2 = \psi_1^2 + \psi_3^2$ and $|c_2|^2 = \psi_2^2 + \psi_4^2$. Fig. 3 presents the probability of the state evolution for superposition state manipulation and Fig. 4 depicts its quantum state evolution on a Bloch sphere. Again, it is easily verified that $|c_1|^2 + |c_2|^2 = 1$ for any arbitrary time instance. The state reaches the superposition target state around $t = 25$ s.

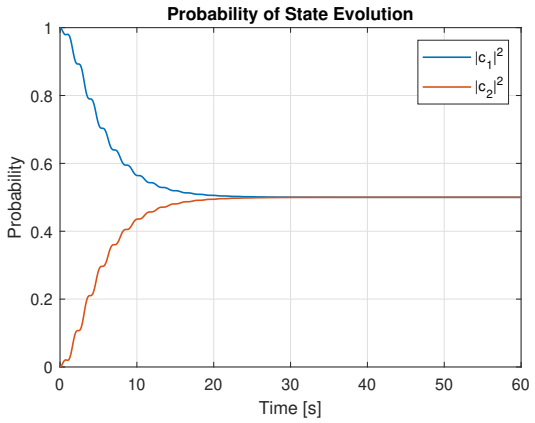


Fig. 3. Probability of states: Arbitrary pure target state.

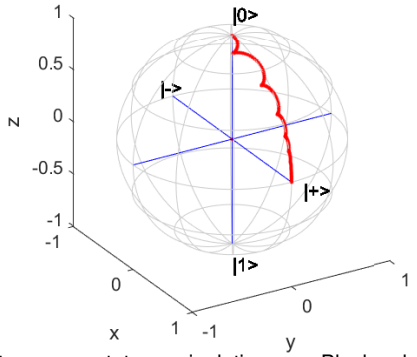


Fig. 4. Arbitrary pure state manipulation on a Bloch sphere.

## VI. Conclusion

In this paper, the proposed adaptive critic learning approach can manipulate pure states online in a *model-free* manner. Nevertheless, it is a data-driven approach that requires full-state feedback. For practical feasibility, the proposed control requires pre-solving the Schrödinger equation because measurements of a quantum system inevitably interfere with its state (decoherence). Future research will extend the results to open quantum systems with coherent control.

## References

[1] D. Dong and I. R. Petersen, "Quantum control theory and applications: a survey," *IET Control Theory & Applications*, vol. 4, no. 12, pp. 2651–2671, 2010.
[2] C. Altafini and F. Ticozzi, "Modeling and control of quantum systems: An introduction," *IEEE Trans. Autom. Contr.*, vol. 57, no. 8, pp. 1898–1917, 2012.
[3] M. Mirrahimi, P. Rouchon, and G. Turinici, "Lyapunov control of bilinear Schrödinger equations," *Automatica*, vol. 41, no. 11, pp. 1987–1994, 2005.
[4] S. Kuang and S. Cong, "Lyapunov control methods of closed quantum systems," *Automatica*, vol. 44, no. 1, pp. 98–108, 2008.
[5] K. Beauchard, J. M. Coron, M. Mirrahimi, and P. Rouchon, "Implicit lyapunov control of finite dimensional Schrödinger equations," *Systems & Control Letters*, vol. 56, no. 5, pp. 388–395, 2007.
[6] S. Cong, *Control of quantum systems: theory and methods*. John Wiley & Sons, 2014.
[7] D. D'alessandro and M. Dahleh, "Optimal control of two-level quantum systems," *IEEE Trans. Autom. Contr.*, vol. 46, no. 6, p. 866, 2001.
[8] L. Magrini, P. Rosenzweig, C. Bach, A. Deutschmann-Olek, S. G. Hofer, S. Hong, N. Kiesel, A. Kugi, and M. Aspelmeyer, "Real-time optimal quantum control of mechanical motion at room temperature," *Nature*, vol. 595, no. 7867, pp. 373–377, 2021.
[9] H. Ding and G. Zhang, "Quantum coherent feedback control with photons," *IEEE Trans. Autom. Contr.*, vol. 69, no. 2, p. 856, 2023.
[10] S. Wang, C. Ding, Q. Fang, and Y. Wang, "Quantum robust optimal control for linear complex quantum systems with uncertainties," *IEEE Trans. Autom. Contr.*, vol. 68, no. 11, pp. 6967–6974, 2023.
[11] D. Dong and I. R. Petersen, "Machine learning for quantum control," in *Learning and Robust Control in Quantum Technology*. Springer, 2023, pp. 93–140.
[12] V. Sivak, A. Eickbusch, H. Liu, B. Royer, I. Tsioutsios, and M. Devoret, "Model-free quantum control with reinforcement learning," *Physical Review X*, vol. 12, no. 1, p. 011059, 2022.
[13] O. Shindi, Q. Yu, P. Girdhar, and D. Dong, "Model-free quantum gate design and calibration using deep reinforcement learning," *IEEE Trans. Artif. Intell.*, vol. 5, no. 1, pp. 346–357, 2023.
[14] R. S. Sutton, A. G. Barto, and R. J. Williams, "Reinforcement learning is direct adaptive optimal control," *IEEE Control Systems*, vol. 12, no. 2, pp. 19–22, 1992.
[15] D. Vrabie, K. G. Vamvoudakis, and F. L. Lewis, "Optimal adaptive control and differential games by reinf. learn. princ." *IET*, 2012.
[16] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, and F. L. Lewis, "Optimal and autonomous control using reinforcement learning: A survey," *IEEE Trans. on Neur. Net. and Learn. Sys.*, vol. 29, no. 6, p. 2042, 2018.
[17] R. Kamalapurkar, P. Walters, J. Rosenfeld, and W. Dixon, *Reinforcement learning for optimal feedback control*. Springer, 2018.
[18] D. Dong and I. R. Petersen, *Learning and Robust Control in Quantum Technology*. Springer Nature, 2023.
[19] C. Chen, D. Dong, H.-X. Li, J. Chu, and T.-J. Tarn, "Fidelity-based probabilistic Q-learning for control of quantum systems," *IEEE Trans. Neur. Net. and Learn. Sys.*, vol. 25, no. 5, pp. 920–933, 2013.
[20] A. S. Chen and G. Herrmann, "Adaptive optimal control via continuous-time Q-learning for unknown nonlinear affine systems," in *IEEE 58th Conf. on Decis. and Contr. (CDC)*, 2019, pp. 1007–1012.
[21] P. A. M. Dirac, *The principles of quantum mechanics*. Oxford university press, 1981, no. 27.
[22] S. Cong, F. Meng, and S. Kuang, "Quantum lyapunov control based on the average value of an imaginary mechanical quantity," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 9991–9997, 2014.
[23] G. Harel and V. Akulin, "Complete control of Hamiltonian quantum systems: Engineering of floquet evolution," *Physical Review Letters*, vol. 82, no. 1, p. 1, 1999.
[24] U. Boscain, T. Chambrion, and J.-P. Gauthier, "On the K+P problem for a three-level quantum system: Optimality implies resonance," *Journal of Dynamical and Control Systems*, vol. 8, pp. 547–572, 2002.
[25] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal control*. John Wiley & Sons, 2012.
[26] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach," *Automatica*, vol. 41, no. 5, pp. 779–791, 2005.
[27] J. P. Richard, "Time-delay systems: an overview of some recent advances and open problems," *Automatica*, vol. 39, no. 10, pp. 1667–1694, 2003.