

# Solving the quantum control problem using the adjoint state method

N. Anders Petersson

July 22, 2019

## 1 ODE constrained minimization of a functional

The wave function  $\Psi(t)$  is governed by the normalized Schroedinger equation,

$$i\dot{\Psi} = H\Psi, \quad H = H^*, \quad t \geq 0,$$

where time has been scaled by  $\hbar$ . We write this system of ordinary differential equations (ODEs) in the form (the state equation),

$$\dot{\Psi} + A(t, \alpha)\Psi = 0, \quad t \in [0, T], \quad \Psi(0) = \Psi_0, \quad A^* = -A, \quad (1)$$

where  $\Psi(t) \in \mathbb{C}^D$ ,  $A \in \mathbb{C}^{D \times D}$ . Let

$$A(t, \alpha) := iH(t, \alpha) = i(H_0 + p(t, \alpha)H_{c1} + iq(t, \alpha)H_{c2}), \quad (2)$$

$$H_0 = H_0^*, \quad H_{c1} = H_{c1}^*, \quad H_{c2} = -H_{c2}^*. \quad (3)$$

Here,  $p \in \mathbb{R}$  and  $q \in \mathbb{R}$  are given control functions that depend on time and the parameter vector  $\alpha \in \mathbb{R}^M$ , with components  $\alpha_k$ ,  $k = 1, 2, \dots, M$ .

The solution of the system of ODEs (1) depends implicitly on the parameter vector  $\alpha$  through the matrix  $A(t, \alpha)$ . Thus,  $\Psi(t) = \Psi^\alpha(t)$ .

The quantum control problem can be cast into a minimization problem for a real-valued functional  $J(\alpha)$  under the constraint that  $\Psi^\alpha$  satisfies (1). To illustrate the idea, consider the simple functional

$$J(\alpha) := g(\Psi^\alpha) = \int_0^T |\Psi^\alpha(\tau) - d(\tau)|^2 d\tau, \quad (4)$$

where the function  $d(t)$  is given. We want to find

$$\alpha : \min_{\alpha} J(\alpha), \quad \Psi^\alpha(t) \text{ satisfies (1)}$$

We define a scalar product for functions  $u$  and  $v$  in  $\mathbb{C}^D \times [0, T]$ ,

$$(u, v) = \int_0^T \langle u(\tau), v(\tau) \rangle d\tau, \quad \langle u, v \rangle = \sum_{j=0}^{D-1} \bar{u}_j v_j.$$

The functional (4) can then be written

$$g(\Psi^\alpha) = (\Psi^\alpha - d, \Psi^\alpha - d).$$

Note that  $d(t)$  is independent of  $\alpha$ . Each component of the gradient of the objective functional satisfies

$$\frac{\partial J}{\partial \alpha_k} = \left( \frac{\partial \Psi^\alpha}{\partial \alpha_k}, \Psi^\alpha - d \right) + \left( \Psi^\alpha - d, \frac{\partial \Psi^\alpha}{\partial \alpha_k} \right) = 2 \operatorname{Re} \left( \frac{\partial \Psi^\alpha}{\partial \alpha_k}, \Psi^\alpha - d \right) \quad (5)$$

By differentiating the state equation (1) with respect to  $\alpha_k$  and introducing the function  $\Phi_k(t) = \partial \Psi^\alpha / \partial \alpha_k(t)$ ,

$$\dot{\Phi}_k + A(t, \alpha) \Phi_k = f_k(t, \alpha), \quad t \in [0, T], \quad \Phi(0) = 0, \quad f_k(t, \alpha) := -\frac{\partial A}{\partial \alpha_k} \Psi^\alpha(t). \quad (6)$$

Thus,  $\Phi_k$  satisfies a state equation with the forcing function  $f_k$ . Inserted into (5), the formula for the gradient becomes

$$\frac{\partial J}{\partial \alpha_k} = 2 \operatorname{Re} (\Phi_k, \Psi^\alpha - d) \quad (7)$$

Using the above approach, we can compute the gradient with respect to all  $M$  parameters by solving the state equation (1) for  $\Psi^\alpha$ , followed by solving (6) for  $\Phi_k$ ,  $k = 1, 2, \dots, M$ . The gradient is finally obtained by evaluating (7). Hence, each evaluation of the gradient requires  $M + 1$  systems of ODEs to be solved. We proceed by deriving a significantly more efficient alternative based on solving the adjoint system of ODEs.

### 1.1 The adjoint problem

Consider the adjoint state equation,

$$-\dot{\lambda} + A^* \lambda = h(t), \quad T \geq t \geq 0, \quad \lambda(T) = 0, \quad (8)$$

where the functions  $\lambda(t)$  and  $h(t)$  are in  $\mathbb{C}^D$ . Note that the adjoint equation is solved backwards in time from the terminal condition  $\lambda(T) = 0$ .

**Lemma 1** (Adjoint relation). *Let  $\Phi(t)$  be the solution of the state equation (6) with forcing function  $f(t)$  and let  $\lambda(t)$  be the solution of the adjoint state equation (8) with forcing function  $h(t)$ . The solutions and forcing functions satisfy the adjoint relation*

$$(f, \lambda) = (\Phi, h). \quad (9)$$

**Proof:** From (6),  $f = \dot{\Phi} + A(\alpha)\Phi$ , which we insert into the left hand side of (9). By integration by parts in time,

$$(\dot{\Phi} + A\Phi, \lambda) = \langle \Phi(\tau), \lambda(\tau) \rangle \Big|_0^T - (\Phi, \dot{\lambda}) + (\Phi, A^* \lambda) = (\Phi, -\dot{\lambda} + A^* \lambda).$$

The boundary term is zero because  $\Phi(0) = 0$  and  $\lambda(T) = 0$ . From (8),  $h = -\dot{\lambda} + A^*\lambda$ , which proves the lemma.  $\square$

The adjoint relation can be used to calculate the gradient of the objective functional (7). The function  $\Phi_k(t)$  satisfies (6) with forcing  $f_k(t) = -\partial A/\partial \alpha_k \Psi^\alpha(t)$ . By taking the forcing in the adjoint equation (8) to be  $h(t) = \Psi^\alpha(t) - d(t)$ , it becomes

$$-\dot{\lambda} + A^*\lambda = \Psi(t) - d(t), \quad T \geq t \geq 0, \quad \lambda(T) = 0. \quad (10)$$

Lemma 1 gives

$$\frac{\partial J}{\partial \alpha_k} = 2 \operatorname{Re}(\Phi_k, h) = 2 \operatorname{Re}(f_k, \lambda) = -2 \operatorname{Re}\left(\frac{\partial A}{\partial \alpha_k} \Psi^\alpha, \lambda\right).$$

The advantage of using the adjoint relation is that all components of the gradient can be calculated from  $\Psi^\alpha(t)$  and  $\lambda(t)$ , i.e., by solving one state equation and one adjoint state equation, we can compute the gradient with respect to any number of parameters,  $M$ .

## 2 The quantum control problem

The solution of the state equation (1) can formally be written

$$\Psi(t) = U(t, \alpha) \Psi(0), \quad U(t, \alpha) = \exp\left(-i \int_0^t H(\tau, \alpha) d\tau\right).$$

Because  $H = H^*$ , the solution operator is unitary,  $U^*U = I$ .

A quantum gate uses control functions to transform a basis of initial data

$$\Psi(0) := [\Psi_0(0), \Psi_1(0), \dots, \Psi_{D-1}(0)] = \begin{bmatrix} 1 & 0 & & & \\ 0 & 1 & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & 0 & 1 & 0 \\ & & & 0 & 1 \end{bmatrix}$$

to some desired target state

$$\Psi^t(T) := [\Psi_0^t(T), \Psi_1^t(T), \dots, \Psi_{D-1}^t(T)] =: G_t \Psi(0) = G_t, \quad G_t^* G_t = I.$$

Here, the complex  $D \times D$  matrix  $G_t$  represents the target gate. Because the solution operator  $U(t, \alpha)$  is unitary for all  $\alpha$ , assuming that the control problem has a solution, the target state must be a unitary transformation of the initial data, implying that  $G_t$  is a unitary matrix. Examples of quantum gates include the Hadamard, Pauli, controlled NOT (CNOT) and the Toffoli (CCNOT) gates.

Given the parameter vector  $\alpha$ , we can integrate the state equation (1) for each set of initial data in  $\Psi(0)$ . The resulting transformation is denoted

$$\Psi^\alpha(T) = G_\alpha \Psi(0) = G_\alpha.$$

## 2.1 Objective functionals

The objective functional must measure the distance between a target gate matrix,  $G_t$ , and the realized gate matrix,  $G_\alpha$ . If  $G_\alpha = G_t$ , then the Frobenious matrix norm satisfies

$$\|G_t\|_F^2 = \langle G_t, G_t \rangle_F = \text{Tr}(G_t^* G_t) = \text{Tr}(I) = D.$$

Furthermore,

$$|\langle G_\alpha, G_t \rangle_F| \leq \|G_\alpha\|_F \|G_t\|_F = D.$$

These observations indicate that the functional

$$c_1 := 1 - \frac{1}{D} |\text{Tr}(G_\alpha^* G_t)| = 1 - \frac{1}{D} |\langle G_\alpha, G_t \rangle_F| \quad (11)$$

could be an appropriate candidate for the quantum control problem. However, consider the (slightly weird) one-dimensional case ( $D = 1$ ) with  $G_t = 1$  and  $G_\alpha = x$ . Then  $G_\alpha^* G_t = x^*$  and  $|\text{Tr}(G_\alpha^* G_t)| = |x|$ . This functional is not differentiable at  $x = 0$  and might cause convergence problems with gradient-based minimization algorithms.

A better choice might be

$$c_2 := 1 - \frac{1}{D^2} |\text{Tr}(G_\alpha^* G_t)|^2 = 1 - \frac{1}{D^2} \langle G_\alpha, G_t \rangle_F \langle G_t, G_\alpha \rangle_F, \quad (12)$$

but it is not obvious what properties of  $G_\alpha$  it actually measures. To gain more insight, consider the functional

$$\begin{aligned} c_3 &:= \|G_\alpha - G_t\|_F^2 = \langle G_\alpha - G_t, G_\alpha - G_t \rangle_F = \\ &\quad \langle G_\alpha, G_\alpha \rangle_F - \langle G_\alpha, G_t \rangle_F - \langle G_t, G_\alpha \rangle_F + \langle G_t, G_t \rangle_F = \\ &\quad 2D - 2\text{Re} \langle G_\alpha, G_t \rangle_F = 2D \left( 1 - \frac{1}{D} \text{Re} \langle G_\alpha, G_t \rangle_F \right). \end{aligned} \quad (13)$$

This functional clearly has a global minima at  $G_\alpha = G_t$ . It shares similarities with (11), but instead of taking the magnitude of the trace, we here take its real part. It may be more convenient to work with the expression on left hand side, because the Frobenious matrix scalar product satisfies

$$\langle A, B \rangle_F = \text{Tr}(A^* B) = \sum_{j=0}^{D-1} \langle A \mathbf{e}_j, B \mathbf{e}_j \rangle,$$

where  $\mathbf{e}_j$  is the  $j^{\text{th}}$  unit vector. By taking  $A = B = G_\alpha - G_t$ ,

$$(G_\alpha - G_t) \mathbf{e}_j = G_\alpha \mathbf{e}_j - G_t \mathbf{e}_j = \Psi_j^\alpha(T) - \Psi_j^t(T).$$

Thus,

$$c_3 = \sum_{j=0}^{D-1} |(G_\alpha - G_t) \mathbf{e}_j|^2 = \sum_{j=0}^{D-1} |\Psi_j^\alpha(T) - \Psi_j^t(T)|^2.$$

We can generalize the  $c_3$  functional to be

$$g_3(\Psi^\alpha) = \sum_{j=0}^{D-1} \int_0^T w(\tau) |\Psi_j^\alpha(\tau) - d_j(\tau)|^2 d\tau, \quad w(t) \geq 0, \quad d_j(t) = \Psi_j^t(t). \quad (14)$$

Here,  $w(t)$  is a weight function with  $w(0) = 0$ ,  $w(T) = 1$ . It is monotonically increasing in time and localized near  $t = T$ . We note that this functional is a vectorized version of the simple functional (4). The gradient of  $g_3$  with respect to all components of  $\alpha$  can therefore be calculated by solving one state and one adjoint wave equation for each  $j$ .

We can rewrite the  $c_2$  functional (12) using the same technique,

$$c_2 = 1 - \frac{1}{D^2} \left( \sum_{j=0}^{D-1} \langle \Psi_j^\alpha(T), d_j(T) \rangle \right) \left( \sum_{j=0}^{D-1} \overline{\langle \Psi_j^\alpha(T), d_j(T) \rangle} \right),$$

which can be generalized into

$$g_2(\Psi^\alpha) = \int_0^T w(\tau) \left( 1 - \frac{1}{D^2} \left( \sum_{j=0}^{D-1} \langle \Psi_j^\alpha(\tau), d_j(\tau) \rangle \right) \left( \sum_{j=0}^{D-1} \overline{\langle \Psi_j^\alpha(\tau), d_j(\tau) \rangle} \right) \right) d\tau, \quad (15)$$

The objective functional that was implemented last spring is

$$g_4(\Psi^\alpha) := \sum_{j=0}^{D-1} \int_0^T w(\tau) (|\Psi_j^\alpha(\tau)|^2 - |d_j(\tau)|^2)^2 d\tau, \quad w(t) \geq 0,$$

It does not consider the relative phases of  $\Psi_j$  and would probably not be very useful in practice.

**Remark 1.** *In many cases the absolute phase of the gate matrix is irrelevant. This means that the objective functional should be invariant to transformations  $G_t \rightarrow \exp(i\beta)G_t$ , for any real phase angle  $\beta$ . Of the functionals discussed above, only  $g_2$  and  $c_1$  satisfy that criteria.*