The adjoint state method without a Lagrangian

N. Anders Petersson

April 17, 2018

Consider the system of ODEs (the state equation),

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

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

$$A(\alpha) := i\left(H_0 + f(\alpha)H_c\right), \quad H_0 = H_0^*, \quad H_c = H_c^*,$$

where $f(\alpha) \in \mathbb{R}$ is a control function that depends on the parameter vector $\alpha \in \mathbb{R}^M$, with components α_k, $k = 1, 2, \ldots, M$.

We consider the minimization problem

$$\min_{\alpha} J(\alpha) := g(\Psi(\alpha)),$$

under the constraint that $\Psi = \Psi(\alpha)$ is a solution of (1) corresponding to the parameter α. Here, $g(\Psi)$ is a functional of Ψ. To demonstrate the technique, let the functional satisfy

$$g(\Psi) = \int_0^T |\Psi(\tau) - d(\tau)|^2 \, d\tau. \tag{2}$$

The function $d(t)$ is given, e.g. from measurments.

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

$$(u, v) = \int_0^T \langle u(\tau), v(\tau)\rangle_2 \, d\tau, \quad \langle u, v\rangle_2 = \sum_{j=1}^{D} \bar{u}_j v_j.$$

The functional (2) can then be written

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

Note that $d(t)$ is independent of α. Each component of the gradient of the cost function satisfies

$$\frac{\partial J}{\partial \alpha_k} = \left(\frac{\partial \Psi}{\partial \alpha_k}, \Psi - d\right) + \left(\Psi - d, \frac{\partial \Psi}{\partial \alpha_k}\right) = 2\,\mathrm{Re}\left(\frac{\partial \Psi}{\partial \alpha_k}, \Psi - d\right) \tag{3}$$

By differentiating the state equation (1) with respect to α_k and introducing the function $\Phi = \partial\Psi/\partial\alpha_k$

$$\dot{\Phi} + A(\alpha)\Phi = f(t), \quad t \in [0, T], \quad \Phi(0) = 0, \tag{4}$$

where $f(t) = -\partial A(\alpha)/\partial\alpha_k\, \Psi(\alpha)$. Thus, Φ satisfies a state equation with the forcing function f. Inserted into (3),

$$\frac{\partial J}{\partial \alpha_k} = 2\,\mathrm{Re}\,(\Phi, \Psi - d) \tag{5}$$

Consider the adjoint state equation,

$$-\dot{\lambda} + A^* \lambda = h(t), \quad T \geq t \geq 0, \quad \lambda(T) = 0, \tag{6}$$

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

$$(f, \lambda) = (\Phi, h). \tag{7}$$

Proof: From (4), $f = \dot{\Phi} + A(\alpha)\Phi$, which we insert into the left hand side of (7). By integration by parts in time,

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

The boundary term is zero because $\Phi(0) = 0$ and $\lambda(T) = 0$. From (6), $h = -\dot{\lambda} + A^*\lambda$, which proves the lemma. $\square$

The adjoint relation can be used to calculate the gradient of the cost function (5). The function $\Phi(t)$ satisfies (4) with forcing $f(t) = -\partial A(\alpha)/\partial\alpha_k \, \Psi(t)$. By taking the forcing in the adjoint equation (6) to be $h(t) = \Psi(t) - d(t)$, it becomes

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

Lemma 1 gives

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

The advantage of using the adjoint relation is that all components of the gradient can be calculated from $\Psi(t)$ and $\lambda(t)$, i.e., by solving one state equation and one adjoint state equation. If, in contrast, the original formula (5) is used it is necessary to solve $M+1$ state equations to obtain all components of the gradient.

For the quantum control problem we need to consider more general cost functionals, such as

$$g_2(\Psi) = \int_0^T w(\tau) \sum_{j=1}^{D} \left(|\Psi_j(\tau)|^2 - |d_j(\tau)|^2 \right)^2 \, d\tau, \quad w(t) \geq 0,$$

Here, $w(t)$ is a weight function that could be increasing in time or localized near $t = T$.