The adjoint state method

Method of Lagrange multipliers

N. Anders Petersson

Lawrence Livermore National Laboratory[1]

June 19, 2019

The control problem

Consider the system of ODEs (the state equation),

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

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 real parameters α_k, $k = 1, 2, \ldots, M$.

We consider minimizing a real-valued cost function

$$J(\alpha) := g(\Psi(\alpha))$$

under the constraint that $\Psi = \Psi(\alpha)$ satisfies (1). E.g.,

$$g(\Psi) = \int_0^T w(\tau)|\Psi(\tau) - d(\tau)|^2 \, d\tau, \quad w(\tau) \geq 0.$$

The Lagrange multiplier method

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

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

Let $\tilde{\Psi}(t)$ and $\tilde{\lambda}(t)$ be in $\mathbb{C}^D \times [0, T]$ (independent of α). Define the Lagrangian

$$\mathcal{L}(\tilde{\Psi}, \tilde{\lambda}, \alpha) := g(\tilde{\Psi}) - (\tilde{\lambda}, \dot{\tilde{\Psi}} + A(\alpha)\tilde{\Psi}).$$

The function Ψ minimizes $g(\tilde{\Psi})$ under the constraint (1) if (Ψ, λ) is a saddle point of the Lagrangian $\mathcal{L}$,

$$\frac{\partial \mathcal{L}}{\partial \tilde{\lambda}}(\Psi, \lambda, \alpha) = 0, \quad \frac{\partial \mathcal{L}}{\partial \tilde{\Psi}}(\Psi, \lambda, \alpha) = 0.$$

The adjoint state equation

The relation $\partial \mathcal{L} / \partial \tilde{\lambda} = 0$ gives the state equation (1) for $\Psi(t)$. To expose how $\mathcal{L}$ depends on $\tilde{\Psi}$, we first integrate by parts,

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

Thus,

$$\mathcal{L}(\tilde{\Psi}, \tilde{\lambda}, \alpha) := g(\tilde{\Psi}) - \langle \tilde{\lambda}(\tau), \tilde{\Psi}(\tau) \rangle \Big|_0^T - (-\dot{\tilde{\lambda}} + A^* \tilde{\lambda}, \tilde{\Psi}).$$

and $\partial \mathcal{L} / \partial \tilde{\Psi} = 0$ gives the adjoint state equation

$$-\dot{\lambda} + A^* \lambda = \frac{\partial g}{\partial \Psi}, \quad T \geq t \geq 0, \quad \lambda(T) = 0, \qquad (2)$$

which is solved backwards in time.

The gradient of the cost function

If $\tilde{\Psi} = \Psi(\alpha)$, it satisfies the state equation (1), and

$$\mathcal{L}(\tilde{\Psi}, \tilde{\lambda}, \alpha) = g(\Psi(\alpha)) := J(\alpha).$$

Because $\partial\mathcal{L}/\partial\tilde{\Psi} = 0$ for $\tilde{\Psi} = \Psi(\alpha)$

$$\frac{\partial J}{\partial \alpha_k} = \frac{\partial \mathcal{L}}{\partial \tilde{\Psi}}\frac{\partial \Psi}{\partial \alpha_k} + \frac{\partial \mathcal{L}}{\partial \alpha_k} = \frac{\partial \mathcal{L}}{\partial \alpha_k}.$$

and

$$\frac{\partial \mathcal{L}}{\partial \alpha_k} = -(\lambda, \frac{\partial A}{\partial \alpha_k}\Psi).$$

Because the state equation is skew-symmetric, it is reversible in time, and the gradient of J follow by backwards accumulation in time.

Direct calculation of the gradient

A direct calculation of the gradient requires $\partial \Psi / \partial \alpha_k$. In our example above, $J(\alpha) = g(\Psi(\alpha))$ with

$$g(\Psi) = \int_0^T w(\tau)|\Psi(\tau) - d(\tau)|^2 \, d\tau, \quad w(\tau) \geq 0,$$

and

$$\frac{\partial J}{\partial \alpha_k} = 2\int_0^T w(\tau)|\Psi(\tau) - d(\tau)|\frac{\partial \Psi}{\partial \alpha_k}(\tau)\, d\tau$$

Differentiating the state equation (1) wrt α_k gives

$$\frac{\partial \dot{\Psi}}{\partial \alpha_k} + A(\alpha)\frac{\partial \Psi}{\partial \alpha_k} = -\frac{\partial A(\alpha)}{\partial \alpha}\Psi, \quad \frac{\partial \Psi}{\partial \alpha_k}(0) = 0.$$

Hence, one state equation must be solved for each component of the gradient.