

Adjoint Equation

General problem

Looking at an optimal control problem

$$\min_{y,u} J(y,u) \text{ subject to } E(y,u) = 0$$

Where $u \in L^2(\Omega)$ and $y \in H^1(\Omega)$, with $\Omega = (0, T)$. I have chosen these spaces because they are Hilbert spaces, however I will for the most part ignore them. J is a functional on $L^2(\Omega) \times H^1(\Omega)$ and E is an operator on $H^1(\Omega)$. Both J and E need certain properties described elsewhere.

Differentiating J is required for solving the problem. To do this we reduce J to $\hat{J}(u) = J(y(u), u)$ and compute its gradient in direction $s \in L^2(\Omega)$. Will use the notation: $\langle \hat{J}'(u), s \rangle$ for the gradient.

$$\begin{aligned} \langle \hat{J}'(u), s \rangle &= \left\langle \frac{\partial J(y(u), u)}{\partial u}, s \right\rangle \\ &= \left\langle \frac{\partial y(u)}{\partial u}^* J_y(y(u), u), s \right\rangle + \langle J_u(y(u), u), s \rangle \\ &= \langle y'(u)^* J_y(u), s \rangle + \langle J_u(u), s \rangle \end{aligned}$$

Here $\langle \cdot, \cdot \rangle$ is the L^2 inner product. The difficult term in the expression above is $y'(u)^*$, so lets first differentiate $E(y(u), u) = 0$ with respect to u , and try to find an expression for $y'(u)^*$:

$$\begin{aligned} \frac{\partial}{\partial u} E(y(u), u) = 0 &\Rightarrow E_y(y(u), u) y'(u) = -E_u(y(u), u) \\ &\Rightarrow y'(u) = -E_y(y(u), u)^{-1} E_u(y(u), u) \\ &\Rightarrow y'(u)^* = -E_u(y(u), u)^* E_y(y(u), u)^{-*} \end{aligned}$$

By inserting our new expression for $y'(u)^*$ into $y'(u)^* J_y(u)$, we get:

$$\begin{aligned} y'(u)^* J_y(u) &= -E_u(y(u), u)^* E_y(y(u), u)^{-*} J_y(u) \\ &= -E_u(y(u), u) \lambda \end{aligned}$$

λ is here the solution of the adjoint equation

$$E_y(y(u), u)^* \lambda = J_y(u)$$

If we can solve this equation for λ , the gradient of \hat{J} will be given by the following formula:

$$\langle \hat{J}'(u), s \rangle = \langle -E_u(y(u), u) \lambda, s \rangle + \langle J_u(u), s \rangle \quad (1)$$

Specific problem and differentiation of the operators

We now look at an example of the above problem, and try to derive the adjoint equation and the gradient. Let J be defined as:

$$J(y, u) = \frac{1}{2} \int_0^T u^2 dt + \frac{1}{2} (y(T) - y^T)^2 \quad (2)$$

and let our ODE constraint be:

$$\begin{cases} E(y, u) = y' - \alpha y - u \\ y(0) = y_0 \end{cases} \quad (3)$$

Before we derive the adjoint equation, lets find E_u , E_y , $\langle J_u(u), s \rangle$ and J_y with respect to our E and J .

$$E_u(y(u), u) = -1$$

$$E_y(y(u), u) = \frac{\partial}{\partial t} - \alpha + \delta_0, \text{ where } \delta_0 \text{ is evaluation at } 0$$

$$\langle J_u(u), s \rangle = \int_0^T u(t)s(t)dt$$

Lets be more thorough with J_y , which is the right hand side in the adjoint equation.

$$\begin{aligned} J_y(y(u), u) &= \frac{\partial}{\partial y} \left(\frac{1}{2} \int_0^T u^2 dt + \frac{1}{2} (y(T) - y^T)^2 \right) \\ &= \frac{\partial}{\partial y} \frac{1}{2} (y(T) - y^T)^2 \\ &= \frac{\partial}{\partial y} \frac{1}{2} \left(\int_0^T \delta_T (y - y^T) dt \right)^2 \\ &= \delta_T \int_0^T \delta_T (y(t) - y^T) dt \\ &= \delta_T (y(T) - y^T) = L \end{aligned}$$

Here I have use some dirac-delta tricks, that may not be valid, but the result is probably correct. By $\delta_T(y(T) - y^T)$, I mean evaluation at time T , of the constant function $y(T) - y^T$.

Deriving the adjoint equation

We have $E_y(y(u), u) = \frac{\partial}{\partial t} - \alpha + \delta_0$, but for the adjoint equation we need to find E_y^* . To derive the adjoint of E_y , we will apply it to a function v and then take the L^2 inner product with another function w . The next step is then to try to "move" the operator E_y from v to w . As becomes clear below,

partial integration is the main trick to achieve this:

$$\begin{aligned}
\langle E_y v, w \rangle &= \int_0^T (v'(t) - \alpha v(t) + \delta_0 v(t)) w(t) dt \\
&= \int_0^T v'(t) w(t) dt - \alpha \int_0^T v(t) w(t) dt + v(0) w(0) \\
&= - \int_0^T v(t) w'(t) dt + v(t) w(t) \Big|_0^T - \alpha \langle v, w \rangle + v(0) w(0) \\
&= - \int_0^T v(t) w'(t) dt - \alpha \langle v, w \rangle + v(T) w(T) \\
&= \langle v, Pw \rangle
\end{aligned}$$

Where $P = -\frac{\partial}{\partial t} - \alpha + \delta_T$. This means that $E_y^* = P$, and we now have the left hand side in the adjoint equation. The right hand side is $J_y(y(u), u) = L$, which we have already found. If we write the adjoint equation on variational form it will look like this: $\langle P\lambda, w \rangle = \langle L, w \rangle$. To get back to standard ODE form, we can do some manipulation:

$$\begin{aligned}
\langle -\lambda' - \alpha\lambda + \delta_T \lambda, w \rangle &= \langle \delta_T(y(T) - y^T), w \rangle \\
\langle -\lambda' - \alpha\lambda, w \rangle &= \langle \delta_T(y(T) - y^T - \lambda), w \rangle
\end{aligned}$$

The right hand side is point evaluation at $t = T$, while the left hand side is an expression for all t . This finally gives us our adjoint equation:

$$\begin{cases} -\lambda'(t) - \alpha\lambda(t) = 0 \\ \lambda(T) = y(T) - y^T \end{cases} \quad (4)$$

This is a simple and easily solvable ODE.

Expression for the gradient

We now have all the ingredients for finding an expression for the gradient of \hat{J} . If we remember that $\langle \hat{J}'(u), s \rangle = \langle y'(u)^* J_y(u), s \rangle + \langle J_u(u), s \rangle$, and all the different expressions for all the terms we calculated, we find:

$$\begin{aligned}
\langle \hat{J}'(u), s \rangle &= \langle y'(u)^* J_y(u), s \rangle + \langle J_u(u), s \rangle \\
&= \langle -E_u^* \lambda, s \rangle + \langle J_u(u), s \rangle \\
&= \langle -(-1)^* \lambda, s \rangle + \langle u, s \rangle \\
&= \langle \lambda + u, s \rangle \\
&= \int_0^T (\lambda(t) + u(t)) s(t) dt
\end{aligned}$$

Note that the adjoint of a constant is just the constant itself.

Simple example

Let $T = y_T = y_0 = \alpha = 1$ and assume that we want to find the gradient of \hat{J} at $u(t) = 0$. We then have:

$$J(y, u) = \frac{1}{2} \int_0^1 u^2 dt + \frac{1}{2} (y(T) - 1)^2 \quad (5)$$

and

$$\begin{cases} E(y, u) = y' - y + u \\ y(0) = 1 \end{cases} \quad (6)$$

Since $u = 0$, we easily find $y(t) = e^t$. This gives us the adjoint equation:

$$\begin{cases} -\lambda'(t) - \lambda(t) = 0 \\ \lambda(T) = e - 1 \end{cases} \quad (7)$$

This is again a simple equation which yields $\lambda(t) = (e - 1)e^{1-t}$. The gradient of \hat{J} is then:

$$\langle \hat{J}'(u), s \rangle = \int_0^1 (e - 1)e^{1-t} s(t) dt$$

Discretization

Let us discretize our interval $[0, T]$ using $N + 1$ points where

$$x_n = n\Delta t, \quad i = 0, \dots, N \quad \text{and} \\ \Delta t = \frac{T}{N}$$

We also let $y^n = y(x^n)$ and $u^n = u(x^n)$. The integrals in our functional and its gradient we evaluate using the trapezoidal rule, and we discretize our ODE $E(y, u) = 0$ and the adjoint equation using the Backward Euler scheme. For $E(y, u) = 0$ we get :

$$\begin{aligned} \frac{y^n - y^{n-1}}{\Delta t} &= \alpha y^n + u^n \\ (1 - \alpha\Delta t)y^n &= y^{n-1} + \Delta t u^n \\ y^n &= \frac{y^{n-1} + \Delta t u^n}{1 - \alpha\Delta t} \end{aligned}$$

Here the initial condition $y^0 = y_0$ is known. For the adjoint equation the initial condition is $\lambda^N = y^N - y^T$, and the Backward Euler scheme gives us:

$$\begin{aligned} -\frac{\lambda^n - \lambda^{n-1}}{\Delta t} - \alpha\lambda^n &= 0 \\ \lambda^{n-1} - \lambda^n &= \Delta t \alpha \lambda^n \\ \lambda^{n-1} &= (1 + \Delta t \alpha) \lambda^n \end{aligned}$$

The discrete gradient

So we now have a way of solving our ODEs numerically. In the continuous case the gradient was $\int_0^T (\lambda(t) + u(t))s(t)dt$, however in the discrete case, \hat{J} is a function dependent on the $N + 1$ values of u . This would suggest that the gradient of \hat{J} should be a vector of size $N + 1$. The thing that makes the most sense to me is to insert the unit vectors of \mathbb{R}^{N+1} into our continuous gradient, and then evaluate the integral using the trapezoidal rule. Based on experiments using finite difference for calculating \hat{J} , this approach works for $n \neq 0$ and $n \neq N$. Without more explanation I will assert that our discrete gradient $\hat{J}'_{\Delta t}(u)$ looks like this:

$$\begin{aligned}\hat{J}'_{\Delta t}(u)^n &= \Delta t(u^n + \lambda^n) \text{ when } n = 1, \dots, N - 1 \\ \hat{J}'_{\Delta t}(u)^0 &= \Delta t \frac{1}{2} u^0 \\ \hat{J}'_{\Delta t}(u)^N &= \Delta t \left(\frac{1}{2} u^N + \lambda^N \right)\end{aligned}$$

Lets try to understand what happens for $n = 0$ and $n = N$. For $n = 0$ we see that there is no λ term. This is because y does not depend on $u(0)$. The reason that this matters in the discrete case and not in the continuous case, is that the point $t = 0$ has measure zero, and the continuous gradient is an integral. We also notice the $\frac{1}{2}$ term in front off $\Delta t u^0$. This comes from our numerical integration using the trapezoidal rule. To make this clear lets state the trapezoidal rule:

$$\int_0^1 f(t)dt \approx \Delta t \left[\frac{f^0 + f^N}{2} + \sum_{n=1}^{N-1} f^n \right]$$

Looking at this expression we also understand the $n = N$ case. Also note that integrating over $(\lambda + \frac{1}{2}u)e^N$ using the trapezoidal rule would give us an extra factor of $\frac{1}{2}$ that is not there when we use finite difference for $\hat{J}'(u)$. This Will all be demonstrated below. One could perhaps derive the discrete results by translating the functional and the ODE to discrete setting , where you exchange $L^2(\Omega)$ with \mathbb{R}^{N+1} , but I will not do this now.

Testing numerics with simple example

Want to test the numerical adjoint to the exact adjoint for the simple example I did above, i.e. $T = y_T = y_0 = \alpha = 1$ and $u = 0$. This gave us the the following solution to our adjoint equation: $\lambda(t) = (e - 1)e^{1-t}$. Using the finite difference schemes I derived above, I calculated the maximum difference between the exact and the numerical adjoint for $N = \{50, 100, 500, 1000\}$ points. The results of the experiment is added in the table below:

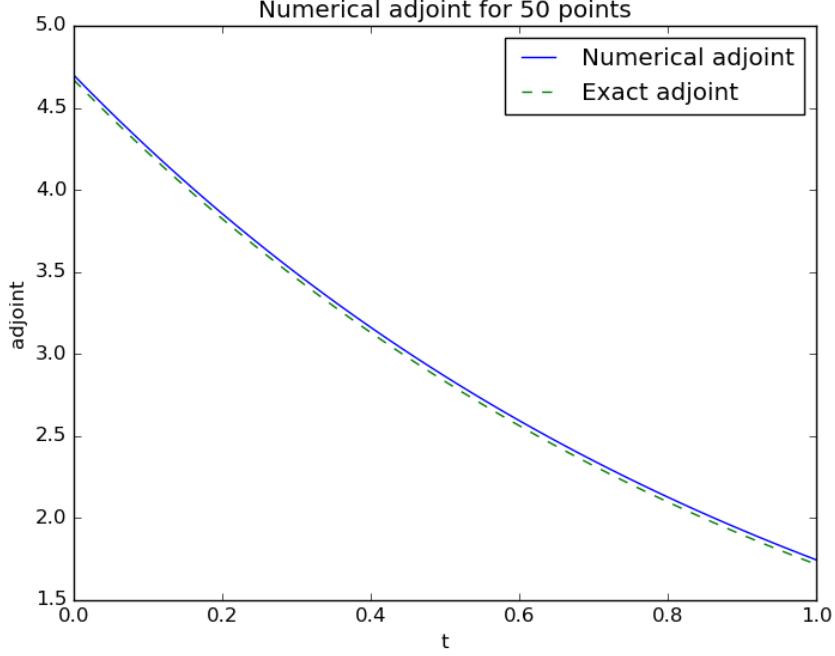


Figure 1: Adjoint for $u = 0$ and $T = y_T = y_0 = \alpha = 1$

N	50	100	500	1000
$\max(\lambda^n - \lambda(t^n))$	0.0317	0.0156	0.0031	0.0015

Using least squares we can find the convergence rate of $|\lambda^n - \lambda(t^n)|$ in Δt , and as expected using simple backward Euler, we get linear convergence:

$$|\lambda^n - \lambda(t^n)|_\infty \leq \Delta t C$$

In this case $C \approx 1.7$. I have also added a plot of the exact and numerical adjoints for $N = 50$.

Testing gradient using finite difference

Now lets try to test the claims I made earlier about the discrete gradient. I will approximate the gradient using finite difference in the following way:

$$\hat{J}'(u)^n \approx \frac{\hat{J}(u + \epsilon^n) - \hat{J}(u)}{\epsilon}$$

$\epsilon^n = \epsilon e^n \in \mathbb{R}^{N+1}$ with $\epsilon > 0$ small, and e^n the unit vector

As always I let $T = y_T = y_0 = \alpha = 1$, however this time I choose $u(t) = e^t + t$. I then define the relative error E between the discrete adjoint gradient $\hat{J}'_{\Delta t}(u)$

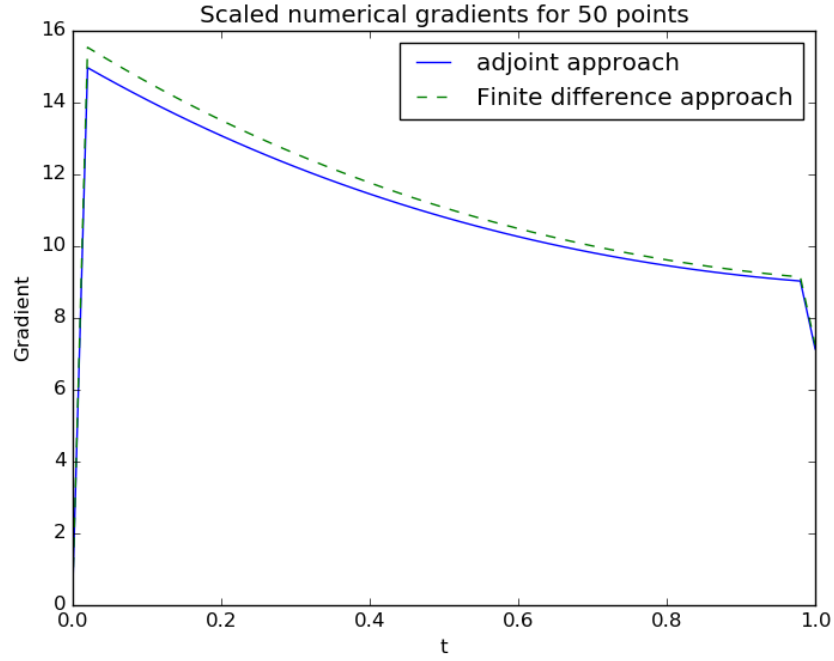


Figure 2: "Relative" gradients for $u = 0$ and $T = y_T = y_0 = \alpha = 1$

and the finite difference gradient $\hat{J}'_\epsilon(u)$ defined as:

$$E = \left| \frac{\hat{J}'_{\Delta t}(u) - \hat{J}'_\epsilon(u)}{\Delta t} \right|_\infty$$

I use this error to test the gradients for different N . The result is given in a table below, and I have also added a plot. Note that I as last time calculated convergence rate using least squares, and the result was: $E \leq \Delta t C$, with $C \approx 27$.

N	50	100	500	1000
E	0.5658	0.2816	0.0561	0.0281