1 Optimal control with ODE constraints

Lets try to derive the adjoint equation and the gradient, when we let $E(y, u)$ be the following ODE:

$$\begin{cases} y'(t) = \alpha y(t) + u(t), \ t \in (0, T) \\ y(0) = y_0 \end{cases}$$

We also choose the functional to be

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

Before we calculate the different terms in the gradient, we want to fit our ODE into an expression E. we do this by "moving" the initial condition into the equation:

$$E(y, u) = y' - \alpha y - u + \delta_0(y - y_0)$$

Here the δ_0 means evaluation at 0. Now 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.

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

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:

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

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: $\langle Pp, w \rangle = \langle L, w \rangle$. To get back to standard ODE form, we can do some manipulation:

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

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:

$$
\begin{cases}
-p'(t) = \alpha p(t) \\
p(T) = y(T) - y^T
\end{cases}
$$

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:

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

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

Parallelizeing in time using the penalty method

To find the above gradient, we must solve first the state equation forward in time and then the adjoint equation backwards in time. One way of speeding things up is to parallelize the solvers by partitioning the time interval and then solving the equation separately on each partitioned interval. If we split the interval $[0, T]$ into m parts we need to solve m state equations on the following form:

$$\begin{cases} \frac{\partial}{\partial t} y_i(t) = \alpha y_i(t) + u(t) \text{ for } t \in [T_{i-1}, T_i] \\ y_i(T_{i-1}) = \lambda_{i-1} \end{cases}$$

here $i = 1, ..., m$, $\lambda_0 = y_0$ and $0 = T_0 < T_1 < \cdots < T_m = T$. Since the equation on each interval depends on the equation in the previous interval, we need a trick, to get everything to hang together. We do this using the penalty method, which means adding a penalty to the functional. The new functional now looks like this:

$$J(y, u, \lambda) = \int_0^T u^2 dt + \frac{1}{2}(y_m(T) - y^T)^2 + \frac{\mu}{2} \sum_{i=1}^{m-1} (y_i(T_i) - \lambda_i)^2$$

This means that the problem now is to minimize J with respect to both u and λ, which means that the reduced functional depends on both u and λ. Since we change the functional and the equation, both the adjoint equation and the gradient changes. Lets try to derive the new adjoint equations and the new gradient. The gradient of the reduced functional now looks like the following:

$$\langle \hat{J}'(u, \lambda), (s, l) \rangle = \langle \frac{\partial y(u, \lambda)}{\partial(u, \lambda)}^* J_y(y(u, \lambda), u, \lambda), (s, l) \rangle + \langle J_u + J_\lambda, (s, l) \rangle$$
$$= \langle -(E_u + E_\lambda)p, (s, l) \rangle + \langle J_u + J_\lambda, (s, l) \rangle$$

Where p is the solution of the adjoint equation $E_y^* p = J_y$, and E is our ODEs:

$$E^i(y, u, \lambda) = \frac{\partial}{\partial t} y_i - \alpha y_i - u + \delta_{T_{i-1}}(y_i - \lambda_{i-1})$$

Lets differentiate E:

$$E_y^i = \frac{\partial}{\partial t} - \alpha + \delta_{T_{i-1}}$$
$$E_u^i = -1$$
$$E_{\lambda_{i-1}}^i = -\delta_{T_{i-1}} \text{ for } i = 2, ..., m$$

Lets differentiate J:

$$\langle J_u, s \rangle = \int_0^T us \, dt$$

$$J_{\lambda_i} = -\mu(y_i(T_i) - \lambda_i)$$

$$J_y = \delta_{T_m}(y_n(T_m) - y_T) + \mu \sum_{i=1}^{m-1} \delta_{T_i}(y_i(T_i) - \lambda_i)$$

We also need $(E_y^i)^*$

$$\int_{T_{i-1}}^{T_i} E_y^i w \, v \, dt = \int_{T_{i-1}}^{T_i} (\frac{\partial}{\partial t} w - \alpha w) \, v \, dt + w(T_{i-1})v(T_{i-1})$$

$$= \int_{T_{i-1}}^{T_i} -(\frac{\partial}{\partial t} v + \alpha v) \, w \, dt + w(T_i)v(T_i)$$

$$= \int_{T_{i-1}}^{T_i} (-\frac{\partial}{\partial t} - \alpha + \delta_{T_i})v \, w \, dt$$

this means that $(E_y^i)^* = -\frac{\partial}{\partial t} - \alpha + \delta_{T_i}$. This gives us the following expressions for the adjoint equations:

$i = m$ case:

$$-\frac{\partial}{\partial t} p_m = \alpha p_m$$

$$p_m(T_m) = y_m(T_m) - y_T$$

$i \neq m$ cases:

$$-\frac{\partial}{\partial t} p_i = p_i$$

$$p_i(T_i) = \mu(y_i(T_i) - \lambda_i)$$

Lets put everything into our expression for or gradient:

$$\langle \hat{J}'(u, \lambda), (s, l) \rangle = \langle -(E_u + E_\lambda)p, (s, l) \rangle + \langle J_u + J_\lambda, (s, l) \rangle$$

$$= \langle (p + \sum_{i=1}^{m-1} \delta_{T_i} p_{i+1}), (s, l) \rangle + \int_0^T us \, dt - \mu \sum_{i=1}^{m-1} (y_i(T_i) - \lambda_i)l_i$$

$$= \int_0^T (u + p)s \, dt + \sum_{i=1}^{m-1} (p_{i+1}(T_i) - \mu(y_i(T_i) - \lambda_i))l_i$$

$$= \int_0^T (u + p)s \, dt + \sum_{i=1}^{m-1} (p_{i+1}(T_i) - p_i(T_i))l_i$$