1 General Problem

Looking at an optimal control problem

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

Where $u \in U$ is the control and $y \in Y$ is the state that depends on u. Usually u and y are functions, and U and Y are either Hilbert or Banach spaces. I will not go into detail about these spaces, but they are mostly chosen to fit the differential equation E, which is an operator on $U \times Y$.

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 $s \in U$. Will use the notation: $\langle \hat{J}'(u), s \rangle$ for the gradient.

$$\langle \hat{J}'(u), s \rangle = \langle DJ(y(u), u), s \rangle$$
$$= \langle \frac{\partial y(u)}{\partial u}^* J_y(y(u), u), s \rangle + \langle J_u(y(u), u), s \rangle$$
$$= \langle y'(u)^* J_y(u), s \rangle + \langle J_u(u), s \rangle$$

Here $\langle \cdot, \cdot \rangle$ is the U 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 $y'(u)^*$:

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

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

$$y'(u)^* J_y(u) = -E_u(y(u), u)^* E_y(y(u), u)^{-*} J_y(u)$$
$$= -E_u(y(u), u)^* p$$

p is here the solution of the adjoint equation

$$E_y(y(u), u)^* p = J_y(u)$$

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

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

2 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

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

Theorem 1. *The adjoint equation of the problem (2) is:*

$$-p'(t) = \alpha p(t)$$
$$p(T) = y(T) - y^T$$

Proof. Before we calculate the different terms used to derive the adjoint equation, we want to fit our ODE into an expression E. We do this by writing up the weak formulation of the equation:

Find $y \in L^2$ such that

$$L[y, \phi] = \int_0^T -y(t)\phi'(t) - \alpha y(t)\phi(t)dt - y_0\phi(0) + y(T)\phi(T) - \int_0^T u(t)\phi(t) = 0$$

$$\forall \ \phi \in C^\infty((0, T))$$

To derive the adjoint we need E_y and J_y. For E_y we define $(\cdot, \cdot)$ to be the L^2 inner product over $(0, T)$. This gives us:

$$E_y = L_y[\cdot, \phi] = (\cdot, (\frac{\partial}{-\partial t} - \alpha + \delta_T)\phi)$$

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

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

2

We have $E_y = (\cdot, (-\frac{\partial}{\partial t} - \alpha + \delta_T)\phi)$, but for the adjoint equation we need to find E_y^*. To derive the adjoint of E_y, we will insert two functions v and w into $L_y[v, w]$, and try to change the places of v and w.

$$
\begin{aligned}
E_y = L_y[v, w] &= \int_0^T -v(t)(w'(t) + \alpha w(t))dt + v(T)w(T) \\
&= \int_0^T w(t)(v'(t) - \alpha v(t))dt + v(T)w(T) - v(T)w(T) + v(0)w(0) \\
&= \int_0^T w(t)(v'(t) - \alpha v(t))dt + v(0)w(0) \\
&= L_y^*[w, v] = E_y^*
\end{aligned}
$$

If we multiply J_y with a test function ψ and set $L_y^*[p, \psi] = (J_y, \psi)$, we get the following equation:

Find p such that

$$
\int_0^T p(t)\psi'(t) - \alpha p(t)\psi(t)dt + p(0)\psi(0) = (y(T) - y^T)\psi(T) \ \forall \ \psi \in C^\infty((0, T))
$$

If we multiply then do partial integration, we get:

Find p such that

$$
\int_0^T (-p'(t) - \alpha p(t))\psi(t)dt + p(T)\psi(T) = (y(T) - y^T)\psi(T) \ \forall \ \psi \in C^\infty((0, T))
$$

Using this we get the strong formulation:

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

$\square$

With the adjoint we can find the gradient of $\hat{J}$. Lets state the result first.

Theorem 2. *The gradient of the reduced functional $\hat{J}$ with respect to u is*

$$
\hat{J}'(u) = u + p
$$

Proof. Firstly we need J_u and E_u^*:

$$
J_u = u
$$
$$
E_u = L_u[\cdot, \phi] = -(\cdot, \phi)
$$

Since $L_u[\cdot, \phi]$ is symmetric, $E_u^* = E_u$, and strongly formulated, $E_u = -1$. The expression for the gradient is then simply:

$$
\begin{aligned}
\hat{J}'(u) &= -E_u^* p + J_u \\
&= p + u
\end{aligned}
$$

3

the directional derivative $\langle \hat{J}'(u), s \rangle$, will therefore be:

$$\langle \hat{J}'(u), s \rangle = \int_0^T (p(t) + u(t))s(t)dt$$

□

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:

$$J(y, u, \lambda) = \int_0^T u^2 dt + \frac{1}{2}(y_m(T) - y^T)^2 + \frac{\mu}{2} \sum_{i=1}^{m-1} (y_i(T_i) - \lambda_i)^2 \quad (3)$$

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:

$$\langle \hat{J}'(u, \lambda), (s, l) \rangle = \langle \frac{\partial y(u, \lambda)}{\partial(u, \lambda)}^* J_y(y(u, \lambda), u, \lambda), (s, l) \rangle + \langle J_u + J_\lambda, (s, l) \rangle \quad (4)$$

$$= \langle -(E_u + E_\lambda)p, (s, l) \rangle + \langle J_u + J_\lambda, (s, l) \rangle \quad (5)$$

Where p is the solution of the adjoint equation $E_y^* p = J_y$, and E is the collection of the interval specific state equations E^i. Since we now have separate state equations on each interval, the adjoint equation is also a collection of equations specific to each interval. I now state the new adjoints:

4

Theorem 3. *The adjoint equation on interval $[T_{m-1}, T_m]$ is:*

$$-\frac{\partial}{\partial t}p_m = \alpha p_m$$
$$p_m(T_m) = y_m(T_m) - y_T$$

On $[T_{i-1}, T_i]$ the adjoint equation is:

$$-\frac{\partial}{\partial t}p_i = p_i$$
$$p_i(T_i) = \mu(y_i(T_i) - \lambda_i)$$

Proof. Lets begin as we did for the non-penalty approach, by writing up the weak formulation of the state equations:

$$\text{Find } y_i \in L^2(T_{i-1}, T_i) \text{ such that}$$

$$L^i[y_i, \phi] = \int_{T_{i-1}}^{T_i} -y_i(t)(\phi'(t) + \alpha\phi(t)) + u(t)\phi(t)dt - \lambda_{i-1}\phi(T_{i-1}) + y_i(T_i)\phi(T_i) = 0$$

$$\forall \ \phi \in C^\infty((T_{i-1}, T_i))$$

To find the adjoint equations we want to differentiate the E^is and the functional J with respect to y. To make notation easier, let $(\cdot, \cdot)_i$ be L^2 inner product of the interval $[T_{i-1}, T_i]$.

$$E_y^i = L_y^i[\cdot, \phi] = (\cdot, -(\frac{\partial}{\partial t} + \alpha - \delta_{T_i})\phi)$$

Lets differentiate J:

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

Since y really is a collection of functions, we can differentiate J with respect to y_i. This gives us:

$$J_{y_m} = \delta_{T_m}(y_n(T_m) - y_T)$$
$$J_{y_i} = \mu\delta_{T_i}(y_i(T_i) - \lambda_i) \ i \neq m$$

We will now find the adjoint equations, by finding the adjoint of the E_y^is. This is done as above, by inserting two functions v, w into $L_y^i[v, w]$, and then moving the derivative form w to v.

$$E_y^i = L_y^i[v, w] = \int_{T_{i-1}}^{T_i} -v(t)(w'(t) + \alpha w(t))dt + v(T_i)w(T_i)$$

$$= \int_{T_{i-1}}^{T_i} w(t)(v'(t) - \alpha v(t))dt + v(T_i)w(T_i) - v(T_i)w(T_i) + v(T_{i-1})w(T_{i-1})$$

$$= \int_{T_{i-1}}^{T_i} w(t)(v'(t) - \alpha v(t))dt + v(T_{i-1})w(T_{i-1})$$

$$= (L_y^i)^*[w, v]$$

5

this means that $(E_y^i)^* = (L_y^i)^*[\cdot, \psi]$. The weak form of the adjoint equations is then found, by setting setting $(L_y^i)^*[p, \psi] = (J_{y_i}, \psi)_i$. This gives to cases:

$i = m$ case:

Find $p_m \in L^2(T_{m-1}, T_m)$ such that $\forall \ \psi \in C^\infty((T_{m-1}, T_m))$

$$\int_{T_{m-1}}^{T_m} p_m(t)\psi'(t) - \alpha p_m(t)\psi(t)dt + p_m(T_{m-1})\psi(T_{m-1}) = (y(T_m) - y^T)\psi(T_m)$$

$i \neq m$ cases:

Find $p_i \in L^2(T_{i-1}, T_i)$ such that $\forall \ \psi \in C^\infty((T_{i-1}, T_i))$

$$\int_{T_{i-1}}^{T_i} p_i(t)\psi'(t) - \alpha p_i(t)\psi(t)dt + p_i(T_{i-1})\psi(T_{i-1}) = \mu(y_i(T_i) - \lambda_i)\psi(T_i)$$

If we want to go back to the strong formulation, we do partial integration, and get:

$i = m$ case:

Find $p_m \in L^2(T_{m-1}, T_m)$ such that $\forall \ \psi \in C^\infty((T_{m-1}, T_m))$

$$\int_{T_{m-1}}^{T_m} -p_m'(t)\psi(t) - \alpha p_m(t)\psi(t)dt + p_m(T_m)\psi(T_m) = (y(T_m) - y^T)\psi(T_m)$$

$i \neq m$ cases:

Find $p_i \in L^2(T_{i-1}, T_i)$ such that $\forall \ \psi \in C^\infty((T_{i-1}, T_i))$

$$\int_{T_{i-1}}^{T_i} -p_i('t)\psi(t) - \alpha p_i(t)\psi(t)dt + p_i(T_i)\psi(T_i) = \mu(y_i(T_i) - \lambda_i)\psi(T_i)$$

This gives us the ODEs we wanted. □

With the adjont equations we can find the gradient.

Theorem 4. *The gradient of (3), $\hat{J}'$, with respect to the control $(u, (\lambda_1, ..., \lambda_{m-1}))$ is:*

$$\hat{J}'(u, \lambda) = (u + p, p_2(T_1) - p_1(T_1), ..., p_m(T_{m-1}) - p_m(T_{m-1}))$$

and the directional derivative with respect to L^2-norm in direction (s, l) is:

$$\langle \hat{J}'(u, \lambda), (s, l) \rangle = \int_0^T (u + p)s \ dt + \sum_{i=1}^{m-1} (p_{i+1}(T_i) - p_i(T_i))l_i$$

Proof. If we first find E_u^*, E_λ^*, J_u and J_λ find the gradient by simply inserting these expression into (5). We can begin with the E terms:

$$E_u = L_u[\cdot, \phi] = -(\cdot, \phi)$$
$$E_{\lambda_{i-1}}^i = L_{y_{i-1}}^i[\cdot, \phi] = -(\cdot, \delta_{T_{i-1}}\phi)_i$$

Notice that both of these forms are symmetric, and we therefore don't need to do more work to find their adjoints, they are however derived from the weak formulation, and it might therefore be easier to translate these forms to their strong counterpart:

$$E_u - 1$$
$$E_{\lambda_{i-1}}^i = -\delta_{T_{i-1}}$$

Then lets differentiate J:

$$J_u = u$$
$$J_\lambda = -\mu \sum_{i=1}^{m-1} (y_i(T_i) - \lambda_i)$$

Let us now insert these into (5), and firstly find the directional derivative:

$$
\begin{aligned}
\langle \hat{J}'(u, \lambda), (s, l) \rangle &= \langle -(E_u + E_\lambda)p, (s, l) \rangle + \langle J_u + J_\lambda, (s, l) \rangle \\
&= \langle (p + \sum_{i=1}^{m-1} \delta_{T_i} p_{i+1}), (s, l) \rangle + \int_0^T us \ dt - \mu \sum_{i=1}^{m-1}(y_i(T_i) - \lambda_i)l_i \\
&= \int_0^T (u + p)s \ dt + \sum_{i=1}^{m-1}(p_{i+1}(T_i) - \mu(y_i(T_i) - \lambda_i))l_i \\
&= \int_0^T (u + p)s \ dt + \sum_{i=1}^{m-1}(p_{i+1}(T_i) - p_i(T_i))l_i
\end{aligned}
$$

Here we use that $p_i(T_i) = \mu(y_i(T_i) - \lambda_i)$. We also see from this, that the gradient has the form we stated above. $\qquad\square$

3 Burgers Equation

Lets look at the optimal control with PDE constraint problem, where the equation is the burgers equation:

$$u_t + uu_x - \nu u_{xx} = 0 \text{ for } (x, t) \in \Omega \times (0, T) \tag{6a}$$
$$u(x, t) = h(x, t) \text{ for } (x, t) \in \partial\Omega \times (0, T) \tag{6b}$$
$$u(x, 0) = g(x) \text{ for } x \in \Omega \tag{6c}$$

Here $\Omega = (a, b)$. The functional is on the form:

$$J(u(g), g) = \int_0^T \int_\Omega u(x, t)^2 dx dt \tag{7}$$

We are actually more interested in the weak formulation of (6a). To simplify notation, let $(\cdot, \cdot)$ be L^2 inner product over Ω.

$$\textit{Find } p(t, x) \in L^2((0, T), H_0^1(\Omega)) \textit{ such that } \forall \phi \in L^2((0, T), H_0^1(\Omega)) \tag{8a}$$

$$(\dot{u}, \phi) - (uu_x, \phi) + \nu(u_x, \phi_x) = \int_{\partial\Omega} h\phi dx \tag{8b}$$

Theorem 5. *If we assume $h(x, t) = 0$, the adjoint equation of (7)-(8b) is:*

$$\textit{Find } p(t, x) \in L^2((0, T), H_0^1(\Omega)) \textit{ such that } \forall \psi \in L^2((0, T), H_0^1(\Omega))$$

$$- (\dot{p}, \psi) - (up_x, \psi) + \nu(p_x, \psi_x) = 2(u, \psi)$$

$$p(T) = 0$$

The gradient of the reduced functional (7) with respect to g is:

$$\hat{J}'(g) = p(0)$$

Proof. We want to minimize J with respect to the initial condition g. If we differentiate J with respect to g, we get:

$$\hat{J}'(g)(s) = \langle u'(g)^* J_u, s \rangle$$
$$= \langle -E_g p, s \rangle$$

where p is the solution of the adjoint equation:

$$E_u^* p = J_u$$

As always, E refers to the state equation, which in our case is Burgers equation (8b). To derive the adjoint it is useful to change the formulation of (8b), to also be an integral over time. We then get the following representation of Burgers equation E:

Find u s.t $\forall \phi \in C^\infty$ (9)

$$L[u, \phi] = \int_0^T -(u, \phi_t) + (uu_x, \phi) + \nu(\nabla u, \nabla \phi) dt + (u(T), \phi(T)) - (g, \phi(0)) = 0$$
(10)

If we differentiate the L form with respect to u, we get:

$$L_u[\cdot, \phi] = \int_0^T -(\cdot, \phi_t) + +(u\frac{\partial}{\partial x}\cdot, \phi) + (u_x\cdot, \phi) + \nu(\nabla\cdot, \nabla\phi) dt + (\delta_{t=T}\cdot, \phi(T))$$

We want to derive the adjoint, and we do this in the usual way, by inserting v, w into the equation, and trying to move the derivative from v to w. The only terms that change, are (v, w_t) and (uv_x, w), so i will do these separately:

$$-\int_0^T (v, w_t)dt = \int_0^T (v_t, w)dt + (v(0), w(0)) - (v(T), w(T))$$

$$\int_0^T (uv_x, w) = \int_0^T \int_\Omega u(x,t)v_x(x,t)w(x,t)dxdt$$

$$= -\int_0^T \int_\Omega v(x,t)(u_x(x,t)w(x,t) + u(x,t)w_x(x,t))dxdt + \int_0^T \int_{\partial\Omega} \frac{\partial v}{\partial n}uwdSdt$$

$$= -\int_0^T (u_x v, w) + (uv, w_x)dt$$

If now insert the expressions above into $L_u[v, w]$, we get:

$$L_u[v, w] = \int_0^T (v_t, w) - (uv, w_x) + \nu(v_x, w_x)dt + (v(0), w(0))$$

$$= L_u^*[w, v]$$

The adjoint equation will be $L[p, \psi] = \int_0^T (J_u, \psi)dt$. From (7) it is easy to see that $J_u = 2u$, which means that the adjoint equation looks like this:

Find p such that $\forall \psi \in C^\infty$

$$\int_0^T (p, \psi_t) - (up_x, \psi) + \nu(p_x, \psi_x)dt + (p(0), \psi(0)) = 2\int_0^T (u, \psi)dt$$

If we move the time derivative from ψ to p, we instead get:

Find p such that $\forall \psi \in C^\infty$

$$\int_0^T -(p_t, \psi) - (up_x, \psi) + \nu(p_x, \psi_x)dt + (p(T), \psi(T)) = 2\int_0^T (u, \psi)dt$$

The only thing we need to do to find the gradient, is to find E_g. From (10), we see that:

$$L_g[\cdot, \phi] = -(\delta_{t=0}\cdot, \phi(0))$$

This means that applying E_g to the solution of the adjoint equation p is the same as evaluating p at zero and multiplying with -1. This means that:

$$\hat{J}'(g) = -E_g p = p(0)$$

$\square$

m time intervals

Divide $[0, T]$ into m intervals $[T_{i-1}, T_i]$, where $0 = T_0 < T_1 < \cdots < T_m = T$. We then solve burgers equation for $\{u^i(x)\}_{i=1}^m$ on each interval with initial conditions $\{\lambda_i(x)\}_{i=0}^{m-1}$, where $\lambda_0(x) = g(x)$. As for the two interval case, we need to add penalty terms to the functional, to solve the optimization problem. The penalized functional looks like this:

$$J(u(g), g, \lambda) = \int_0^T \int_\Omega u(x, t)^2 dx dt + \frac{\mu}{2} \sum_{i=1}^{m-1} \int_\Omega (u^i(x, T_i) - \lambda_i(x))^2 dx$$

The state equation on each interval looks like:

$$u_t^i + u^i u_x^i - \nu u_{xx}^i = 0 \text{ for } (x, t) \in \Omega \times (T_{i-1}, T_i)$$
$$u^i(x, t) = h(x, t) \text{ for } (x, t) \in \partial\Omega \times (T_{i-1}, T_i)$$
$$u^i(x, 0) = \lambda_{i-1}(x) \text{ for } x \in \Omega$$

If write this equation as en operator E^i, we get:

$$E^i = u_t^i + u^i u_x^i - \nu u_{xx}^i + \delta_{t=T_{i-1}}(u^i - \lambda_{i-1}) + \delta_{\partial\Omega}(u^i - h)$$

Now lets look at the gradient. The gradient expression for m intervals depends on both g and λ, which is now a vector $\lambda = (\lambda_1, ..., \lambda_{m-1})$. First the gradient:

$$\langle \hat{J}'(g, \lambda), (s, l) \rangle = \langle -(E_g + E_\lambda)p, (s, l) \rangle + \langle J_g + J_\lambda, (s, l) \rangle$$

Again p is the solutions of the adjoint equations $(E_u^i)^* p^i = J_{u^i}$. Lets state the values of the different components in the gradient:

$$E_u^i = \frac{\partial}{\partial t} + u_x^i + u^i \frac{\partial}{\partial x} - \nu\Delta + \delta_{t=T_{i-1}} + \delta_{\partial\Omega}$$
$$E_g^1 = -\delta_{t=0}$$
$$E_{\lambda_i}^i = -\delta_{t=T_i} \ i \neq 1$$
$$J_g = 0$$
$$J_u = 2u + \mu \sum_{i=1}^{m-1} (u^i - \lambda_i)\delta_{t=T_i}$$
$$\langle J_{\lambda_i}, l \rangle = -\mu \int_\Omega (u^i(x, T_i) - \lambda_i(x)) l_i(x) dx$$

This gives us the following adjoint equations:

$$-p_t^i - u^i p_x^i - \nu p_{xx}^i = 2u^i \text{ for } (x, t) \in \Omega \times (T_{i-1}, T_i)$$
$$p^i(x, t) = 0 \text{ for } (x, t) \in \partial\Omega \times (T_{i-1}, T_i)$$
$$p^m(x, T) = 0 \text{ for } x \in \Omega$$
$$p^i(x, T_i) = \mu(u^i(x, T_i) - \lambda_i(x)) \text{ for } x \in \Omega \text{ and } i \neq m$$

We can also show that the gradient looks like the following:

$$\langle \hat{J}'(g, \lambda), (s, l) \rangle = \langle -(E_g + E_\lambda)p, (s, l) \rangle + \langle J_g + J_\lambda, (s, l) \rangle$$

$$= \int_\Omega p^1(x, 0)s(x)dx + \sum_{i=1}^{m-1} \int_\Omega (p^{i+1}(x, T_i) - p^i(x, T_i))l_i(x)dx$$