

Penalty gradient and non-linear ODEs

We have the penalized functional

$$J(y, u, \lambda) = \int_0^T u^2 dt + \frac{1}{2}(y_n(T) - y_T)^2 + \frac{\mu}{2} \sum_{i=1}^n (y_{i-1}(T_i) - \lambda_i)^2$$

Our state equation is solved separately on $n + 1$ intervals:

$$\begin{aligned} \frac{\partial}{\partial t} y_i &= y_i + u \text{ for } t \in [T_i, T_{i+1}] \\ y_i(T_i) &= \lambda_i \end{aligned}$$

here $i = 0, \dots, n$, $\lambda_0 = y_0$ and $0 = T_0 < T_1 < \dots < T_n < T_{n+1} = T$. Want the gradient of the reduced functional:

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

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

$$E^i(y, u, \lambda) = \frac{\partial}{\partial t} y_i - y_i - u + \delta_{T_i}(y_i - \lambda_i)$$

Lets differentiate E :

$$\begin{aligned} E_y^i &= \frac{\partial}{\partial t} - 1 + \delta_{T_i} \\ E_u^i &= -1 \\ E_{\lambda_i}^i &= -\delta_{T_i} \end{aligned}$$

Lets differentiate J :

$$\begin{aligned} \langle J_u, s \rangle &= \int_0^T u s \, dt \\ J_{\lambda_i} &= -\mu(y_{i-1}(T_i) - \lambda_i) \\ J_y &= \delta_{T_{n+1}}(y_n(T_{n+1}) - y_T) + \mu \sum_{i=1}^n \delta_{T_i}(y_{i-1}(T_i) - \lambda_i) \end{aligned}$$

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

$$\begin{aligned} \int_{T_i}^{T_{i+1}} E_y^i w \, v \, dt &= \int_{T_i}^{T_{i+1}} \left(\frac{\partial}{\partial t} w - w \right) v \, dt + w(T_i)v(T_i) \\ &= \int_{T_i}^{T_{i+1}} -\left(\frac{\partial}{\partial t} v + v \right) w \, dt + w(T_{i+1})v(T_{i+1}) \\ &= \int_{T_i}^{T_{i+1}} \left(-\frac{\partial}{\partial t} - 1 + \delta_{T_{i+1}} \right) v \, w \, dt \end{aligned}$$

this means that $(E_y^i)^* = -\frac{\partial}{\partial t} - 1 + \delta_{T_{i+1}}$. This gives us the following expressions for the adjoint equations:

$i = n$ case:

$$\begin{aligned} -\frac{\partial}{\partial t} p_n &= p_n \\ p_n(T_{n+1}) &= y_n(T_{n+1}) - y_T \end{aligned}$$

$i \neq n$ cases:

$$\begin{aligned} -\frac{\partial}{\partial t} p_i &= p_i \\ p_i(T_{i+1}) &= \mu(y_i(T_{i+1}) - \lambda_{i+1}) \end{aligned}$$

Lets put everything into our expression for or gradient:

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

Augmented Lagrange

An alternative to the penalty method is the augmented Lagrange method. This method is similar to the penalty method, but it is apparently more stable. The main difference is in the functional, which now reads as:

$$\begin{aligned} J(y, u, \lambda) &= \int_0^T u^2 dt + \frac{1}{2} (y_n(T) - y_T)^2 + \frac{\mu}{2} \sum_{i=1}^n (y_{i-1}(T_i) - \lambda_i)^2 + \sum_{i=1}^n \Gamma_i (y_{i-1}(T_i) - \lambda_i) \\ &= \int_0^T u^2 dt + \frac{1}{2} (y_n(T) - y_T)^2 + (y_{i-1}(T_i) - \lambda_i) \sum_{i=1}^n (\Gamma_i + \frac{\mu}{2} (y_{i-1}(T_i) - \lambda_i)) \end{aligned}$$

The difference is that we add non squared penalty terms with a Lagrange multiplication factor Γ_i . Since we change the functional, its derivative changes, however the only the J_y and J_{λ_i} terms change:

$$\begin{aligned} J_{\lambda_i} &= -\mu(y_{i-1}(T_i) - \lambda_i) - \Gamma_i \\ J_y &= \delta_{T_{n+1}}(y_n(T_{n+1}) - y_T) + \sum_{i=1}^n \delta_{T_i} [\mu(y_{i-1}(T_i) - \lambda_i) + \Gamma_i] \end{aligned}$$

This changes both the adjoint equations and the gradient. For the adjoint equations the change happens in the 'initial' condition:

$$\begin{aligned} -\frac{\partial}{\partial t} p_i &= p_i \\ p_i(T_{i+1}) &= \mu(y_i(T_{i+1}) - \lambda_{i+1}) + \Gamma_{i+1} \end{aligned}$$

For the gradient the change happens in the λ part of the control:

$$\begin{aligned} \langle \hat{J}'(u, \lambda), (s, l) \rangle &= \langle -(E_u + E_\lambda)p, (s, l) \rangle + \langle J_u + J_\lambda, (s, l) \rangle \\ &= \int_0^T (u + p)s \, dt + \sum_{i=1}^n [(p_i(T_i) - p_{i-1}(T_i)) - \Gamma_i] l_i \end{aligned}$$

The Γ need to be updated for each iteration, i.e. each time we change μ .

Non-linear ODEs

We want to solve the ODE constrained optimization problem:

$$\min_u J(u, y(u)) \text{ with } E(u, y) = 0$$

For the most part, we have let the ODE $E(u, y) = 0$ be linear both in y and u , and on the form:

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

Now I want to comment on what happens, when we let E be non-linear in y . Let us then have the following equation:

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

Here $F : \mathbb{R} \rightarrow \mathbb{R}$, is some differentiable function. We then get $E_y = \frac{\partial}{\partial t} - F'(y)$. To derive the adjoint equation, we need to the the adjoint of the operator E_y . This is problematic since it depends on y , however since we need to solve the state equation before the adjoint, we can think of $F'(y)$ as a function of t . Using this linearisation, we get the following adjoint equation:

$$\begin{cases} \lambda'(t) = F'(y(t))\lambda(t) \\ \lambda(T) = y(T) - y_T \end{cases}$$

The "initial" condition is derived in the usual way assuming:

$$J(y, u) = L(u) + (y(T) - y_T)^2$$

Where L is some functional.

Non-quadratic y-term in functional

To make the problem a bit more complicated, lets change the functional in its last term:

$$J(y, u) = \int_0^T u^2 dt + \frac{1}{3}(y(T) - y_T)^3$$

This change results in a different J_y , and therefore a changed adjoint equation. The derivative of J with respect to y is now:

$$J_y = (y(T) - y_T)^2 \delta_T$$

This changes the 'initial' condition of the adjoint equation. We now get the following ODE:

$$\begin{aligned} -\frac{\partial}{\partial t} p(t) &= p(t) \\ p(T) &= (y(T) - y_T)^2 \end{aligned}$$