Unsteady Continuous Adjoint for ODE

Andrew Ning

Let's consider the simplest case of an ODE-constrained optimization problem. We will assume an explicit ODE of the form:

$$\dot{y} = f(y, \theta, t) \tag{1}$$

where y is the state, θ are the model parameters, and t is time. We will express this as a residual:

$$r(y, \dot{y}, \theta, t) = \dot{y} - f(y, \theta, t) = 0 \tag{2}$$

which could easily generalize to the implicit case.

Our objective, or loss function, we will assume is a function of just the final state: $L(y_f)$, which would be the natural choice in a neural ODE context, but in a more general problem L can be a function of all states and even the parameters. We will also assume that our initial condition is a fixed constant: $y(t_i) = y_0$ In the general case it could be a (explicit or implicit) function of the parameters. In summary:

$$\begin{aligned}
\min_{\theta} \quad & L(y_f) \\
s.t. \quad & \dot{y} - f(y, \theta, t) = 0 \\
with \quad & y(t_i) = y_0
\end{aligned} \tag{3}$$

We want the total derivative $dL/d\theta$. We can form a Lagrangian, which needs to be an integral in this case since we are treating it as a continuous time problem:

$$\mathcal{L} = L(y_f) + \int_{t_i}^{t_f} \lambda(t)(\dot{y} - f(y, \theta, t))dt \tag{4}$$

Our "constraints" are always satisfied in this case, by design, so the second term is always zero. That means that $d\mathcal{L}/d\theta = dL/d\theta$, and it also means we have a degree of freedom in choosing $\lambda(\theta)$. So, in effect we've just added some clever term, that hasn't actually changed the value of our loss function. The goal will be to choose λ to eliminate any dependence on total derivatives (hard to compute) leaving only partial derivatives (easy), somewhat analogous to the steady adjoint case.

Let's take derivatives of the Lagrangian then, since that will be equivalent to the derivative we are after:

$$\frac{dL}{d\theta} = \frac{d\mathcal{L}}{d\theta} = \frac{dL}{dy_f}\frac{dy_f}{d\theta} + \int_{t_i}^{t_f} \lambda(t) \left[\frac{d\dot{y}}{d\theta} - \frac{\partial f}{\partial y}\frac{dy}{d\theta} - \frac{\partial f}{\partial \theta} \right] dt \tag{5}$$

We want to express the $d\dot{y}/d\theta$ term in terms of $dy/d\theta$ so we can consolidate. We'll remove it through integration by parts. Let's examine just that term. We will first pull out the time derivative to make it clearer how we are integrating by parts:

$$\int_{t_i}^{t_f} \lambda(t) \frac{d\dot{y}}{d\theta} dt = \int_{t_i}^{t_f} \lambda(t) \frac{d}{dt} \left(\frac{dy}{d\theta} \right) dt \tag{6}$$

$$= \lambda(t) \frac{dy}{d\theta} \bigg|_{t_i}^{t_f} - \int_{t_i}^{t_f} \frac{dy}{d\theta} \dot{\lambda} dt \tag{7}$$

We now substitute that in (while expanding the first term):

$$\frac{dL}{d\theta} = \frac{dL}{dy_f}\frac{dy_f}{d\theta} + \lambda(t_f)\frac{dy_f}{d\theta} - \cancel{\lambda(t_i)\frac{dy_i}{d\theta}} - \int_{t_i}^{t_f} \frac{dy}{d\theta}\dot{\lambda} - \lambda(t)\left[\frac{\partial f}{\partial y}\frac{dy}{d\theta} + \frac{\partial f}{\partial \theta}\right] dt \tag{8}$$

Notice that we can eliminate the derivative of the initial condition, since it is constant in our formulation. Under the integral we will group the terms that depend on $dy/d\theta$. That is a total derivative we want to eliminate. We also group the first two terms, which have a total derivative as well.

$$\frac{dL}{d\theta} = \left(\frac{dL}{dy_f} + \lambda(t_f) \right) \frac{dy_f}{d\theta} - \int_{t_i}^{t_f} \frac{dy}{d\theta} \left[\dot{\lambda} + \lambda \frac{\partial f}{\partial y} \right] + \lambda \frac{\partial f}{\partial \theta} dt \tag{9}$$

Since we free to choose λ we are going to choose it so that all of our terms with total derivatives are eliminated. That means we want the term in brackets to be zero. Thus, we need to solve λ such that:

$$\dot{\lambda} + \lambda \frac{\partial f}{\partial y} = 0 \tag{10}$$

We also choose $\lambda(t_f)$ such that first term goes to zero, which eliminates the other problematic total derivative. In other words:

$$\lambda(t_f) = -\frac{dL}{dy_f} \tag{11}$$

(generally a very simple derivative). What remains is:

$$\frac{dL}{d\theta} = -\int_{t_i}^{t_f} \lambda \frac{\partial f}{\partial \theta} dt \tag{12}$$

So the procedure is to solve the original ODE $\dot{y} = f(y, \theta, t)$ normally (forward pass to obtain y), then we solve the following ODE backwards from t_f to t_i for λ:

$$\dot{\lambda} = -\lambda \frac{\partial f}{\partial y} \tag{13}$$

where the "initial condition" is given in eq. (11). Now that we have λ we simply integrate eq. (12) to obtain our desired total derivatives. Note that the desired total derivative depends only on partial derivatives ($\partial f/\partial y$ and $\partial f/\partial \theta$) which are easy to compute.