Forward Mode Differentiation of Differential Equation and Nonlinear Solvers: Notes from the Geilo Winter School 2025

Chris Rackauckas

January 2025

1 Introduction

2 Differentiation of Loss Functions Containing ODE Definitions

Why do we need to caculate the derivative of the simulator?

Let's assume that $S(p)$ is the solution to the simulator given parameters p. And let's assume that we are solving the inverse problem by minimizing a cost function $C(p) = \|S(p) - d\|$. Question: what is $\frac{dC}{dp}$?

$C(p) = \sum_{i=1}^{n} (S(p)_i - d_i)^2$

$\frac{dC}{dp} = \sum_{i=1}^{n} 2(S(p)_i - d_i)\frac{dS_i}{dp}$

where $\frac{dS}{dp}$ is the change in the simulation output when changing your parameters.

3 Forward Sensitivity Analysis

3.1 Forward Sensitivity Analysis of an ODE

$\frac{du(t,p)}{dt} = f(u(t,p),p,t)$

Differentiate both sides by p:

$\frac{d}{dt}\frac{du}{dp} = \frac{d}{dp}\frac{du(t,p)}{dt} = \frac{d}{dp}f(u(t,p),p,t) = \frac{\partial f}{\partial u}\frac{du}{dp} + \frac{\partial f}{\partial p}$

Now let's let $s = sensitivty = \frac{du(t,p)}{dp}$

$\frac{ds}{dt} = s' = \frac{\partial f(u(t,p),p,t)}{\partial u}s + \frac{\partial f(u(t,p),p,t)}{\partial p}$

$$u' = f(u,p,t)$$
$$s' = \frac{\partial f}{\partial u}s + \frac{\partial f}{\partial p}$$

System of differential equations known as the forward sensitivity equation.

For this to be an ODE, you need $s(0)$. What is $s(0)$?

$s(0) = \frac{du(0,p)}{dp}$ =the derivative of the initial conditions w.r.t. the parameters.

$\frac{d}{dt}\frac{dx}{d\alpha} = \frac{\partial f_1}{\partial x}\frac{dx}{d\alpha} + \frac{\partial f_1}{\partial \alpha}$

Take an ODE system of size n with p parameters, and turn it into a new ODE of size $(n+1)p$, what you get is a time series $\frac{du(t)}{dp}$.

Downsides? Cost $\mathcal{O}(np)$

Question: can you do this in $\mathcal{O}(n+p)$?

3.2 Forward Sensitivity Analysis of Nonlinear Solving

So assume you have a nonlinear system:

Solve for $u^*(p)$ s.t. $f(u^*(p), p) = 0$. So by definition:

$$f(u^*(p), p) = 0$$
$$\frac{d}{dp} f(u^*(p), p) = 0$$
$$\frac{\partial f}{\partial u}\frac{du^*}{dp} + \frac{\partial f}{\partial p} = 0$$
$$Ax = b$$

where $b = -\frac{\partial f}{\partial p}$, $A = \frac{\partial f}{\partial u}$

so then $\frac{du^*}{dp} = -\frac{\partial f}{\partial u}^{-1}|_{u^*}\frac{\partial f}{\partial p}$

$A(p)x = b$

$\frac{dx}{dp}$

Let $S(p)$ be the solution to solving an optimization problem with parameters p. What is $\frac{dS}{dp}$?

4 What is Forward-Mode Automatic Differentiation?

So forward mode automatic differentiation is a scheme for computing derivatives via dual number arithmatic. (Some formulations of forward mode AD do not use dual numbers, though this is equivalent to forms via computational graphs).

Definition: a dual number is a number $d = x + y\epsilon$ (where ϵ is a dimensional signifier, (x, y), x is known as the primal and y is the partial) which satisfies the algebra defined by:

$$f(d) = f(x) + yf'(x)\epsilon$$
$$\epsilon^2 = 0$$

Think of x as "carrying forward the original value", and y as "the derivative of the program w.r.t. x".

Let's check:

$f(x+1\epsilon) = f(x) + f'(x)\epsilon$

$$(x+y\epsilon)+(v+w\epsilon) = (x+v)+(y+w)\epsilon$$
$$(x+y\epsilon)(v+w\epsilon) = xv+(yv+wx)\epsilon$$

$g(f(d)) = g(f(x)+yf'(x)\epsilon) = g(f(x)) + g'(f(x))f'(x)y\epsilon$

$\|d\| = \|x\|$

$\|d\| = \|x\| + \|y\|$

4.1 What does automatic differentiation do to complex codes, like a solver for an ordinary differential equation?

Let's figure out why the freaky magic works by using Euler's method:

$u' = f(u)$

We solve this via the iteration:

$$u_{n+1} = u_n + hf(u_n)$$

Now let $u_n = x_n + y_n\epsilon$. Plug it in:

$$x_{n+1} + y_{n+1}\epsilon = x_n + y_n\epsilon + hf(x_n + y_n\epsilon,)$$
$$x_{n+1} + y_{n+1}\epsilon = x_n + y_n\epsilon + h(f(x_n) + y_n f'(x_n)\epsilon)$$

now I split by taking the terms with and without epsilon:

$$x_{n+1} = x_n + hf(x_n)$$
$$y_{n+1} = y_n + hy_n f'(x_n)$$

now let $h \to 0$, then this is Euler's method for solving what ODE?

$$x' = f(x)$$
$$y' = f'(x)y$$

Let's change notation: $x = u$, $y = s$ then...

$$u' = f(u)$$
$$s' = \frac{\partial f}{\partial u} s$$