Algorithmic Differentiation

Manuel Alberto Parra Díaz
MSc. Mathematics

Faculty of Economics
Universidad del Rosario

August 11, 2022

Presentation Outline

Some ways to get the partial derivatives of functions are

- Manual
- Numerical
- Symbolic
- Automatic

The Algortihmic Differentiation techniques have been used in different fields where the computation of function derivatives is crucial.

In quantitative finance, the computation of a large number of sensitivities is made often with numerical approximations (finite differences).

However, the AD techniques appear as a solution to find better approximations with lower computational cost.

Definition 2.1

A function

$$f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$$
$$x \longrightarrow f(x)$$

is said to be *diferentiable* at a point $x_0 \in \mathbb{R}^m$ if f is defined on that point and there exists a linear function $Df(x_0) : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ such that

$$\lim_{\epsilon \to 0} \frac{f(x_0 + \epsilon) - f(x_0) - Df(x_0)(\epsilon)}{|\epsilon|} = 0$$

The linear function $Df(x_0)$ is called the *derivative* of f in x_0.

Definition 2.2

Let $f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ be a differentiable function at a point $x_0 \in \mathbb{R}^m$ and e_i be the base vector of $\mathbb{R}^m$ in the i-th dimension $(1 \leq i \leq m)$.

The *partial derivative* of f is defined by

$$\lim_{\epsilon \to 0} \frac{f(x_0 + \epsilon e_i) - f(x_0)}{\epsilon}.$$

If the main function of the model is “simple”, the previous definition of partial derivatives is easily applied through differentiation rules (symbolic computation). But working with more complex functions implies the use of composition rules that are more difficult to program.

Definition 2.3

Let $g : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ and $f : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ be two functions. The function called f *compounded* g, is the function

$$\begin{aligned} f \circ g : \mathbb{R}^m &\longrightarrow \mathbb{R}^p \\ x &\longrightarrow f(g(x)) \end{aligned}$$

Theorem 2.4 (Chain Rule)

Let $g : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ and $f : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ be two functions. If g is differentiable in x_0 and f is differentiable in $g(x_0)$, then f compounded g is diferentiable in x_0 and

$$D(f \circ g)(x_0) = Df(g(x_0)) \cdot Dg(x_0).$$

$$x_0 \xrightarrow{g} g(x_0) \xrightarrow{f} f(g(x_0))$$

Definitions

A main assumption in this theory is that, all computer programs use a finite set of elementary operations as defined by the programming language. And all the other values or functions are just compositions of these elementary functions, its derivatives are known and the others use Chain rule.

Definition 2.5

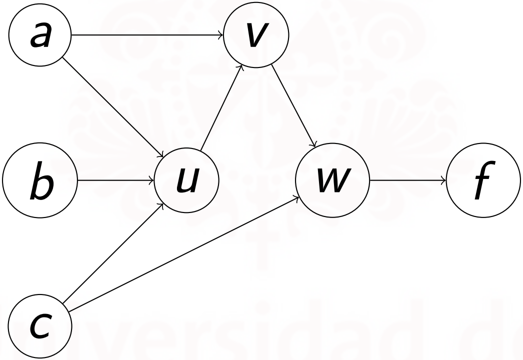
Algorithmic Differentiation is a chain-rule-based technique for evaluating the derivatives with respect to the input variables of functions defined by a high-level language computer program.

There are two modes of operation for Algorithmic Differentiation:

1. **Forward or Tangent Linear Mode**
   It is used when f has a large number of outputs and a small number of inputs. Here, the derivatives are propagated by throghout the computation using the Chain Rule.
2. **Reverse or Adjoint Mode**
   It is used when f has a small number of outputs and a large number of inputs. Here, we calculate the derivatives for all intermediate variables backwards through the computation.

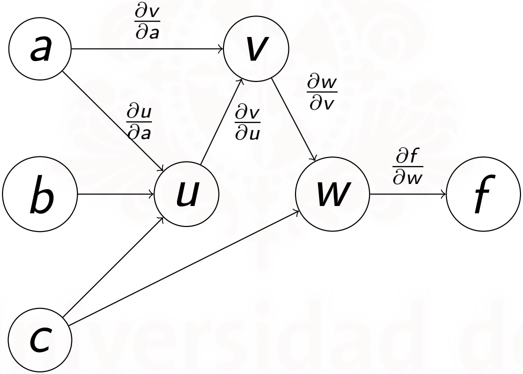
- Adjoint mode cost is smaller than tangent linear because of its independency of the number of inputs, whereas in tangent linear mode it increase linearly with the number of inputs.
- The cost of adjoint mode is also smaller than 5 times the computation cost of a regular run (finite differences). [3]
- Adjoint mode computes the derivatives exactly (up to machine precision) while finite differences incur truncation errors.

Tangent Linear Mode



$$¿\frac{\partial f}{\partial a}?$$

Tangent Linear Mode



$$¿\frac{\partial f}{\partial a}?$$

$$(1)\ \frac{\partial u}{\partial a}$$

$$(2)\ \frac{\partial u}{\partial a} \cdot \frac{\partial v}{\partial u} + \frac{\partial v}{\partial a}$$

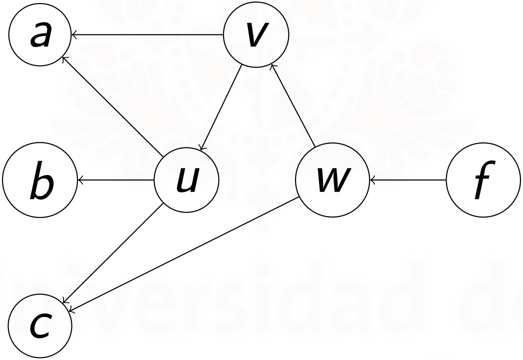
$$(3)\ \left(\frac{\partial u}{\partial a} \cdot \frac{\partial v}{\partial u} + \frac{\partial v}{\partial a}\right) \frac{\partial w}{\partial v}$$

$$(4)\ \left(\frac{\partial u}{\partial a} \cdot \frac{\partial v}{\partial u} + \frac{\partial v}{\partial a}\right) \frac{\partial w}{\partial v} \cdot \frac{\partial f}{\partial w}$$

$$(5)\ \frac{\partial f}{\partial a}$$

This process can be followed to complete the computation of the other derivatives.

Adjoint Mode



In the previous method, we compute

$$
\begin{aligned}
du &= \frac{\partial u}{\partial a} da + \frac{\partial u}{\partial b} db + \frac{\partial u}{\partial c} dc \\
dv &= \frac{\partial v}{\partial a} da + \frac{\partial v}{\partial u} du \\
dw &= \frac{\partial w}{\partial v} dv + \frac{\partial w}{\partial c} dc \\
df &= \frac{\partial f}{\partial w} dw
\end{aligned}
$$

to obtain the partial derivatives of the function respect to the inputs. But, we can go backwards defining **Adjoints** that store the dependencies in the function and its intermediate variables.

$\overline{x}$ denotes the adjoint to the intermediate variable x of a function and is defined by

$$\overline{x} = \sum_{y \in A} \overline{y} \frac{\partial y}{\partial x}$$

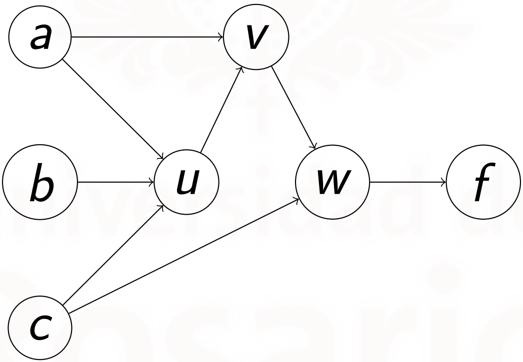
where A is the set of all intermediate variables that have x as an input.

Compute the gradient of $f(a,b,c) = (w - w_0)^2$ where,

$$
\begin{aligned}
w &= \ln\left(v^2 + 1\right) + \cos\left(c^2 - 1\right) \\
v &= \exp\left(u^2 - 1\right) + a^2 \\
u &= \sin\left(ab\right) + cb^2 + a^3c^2
\end{aligned}
$$

Compute the gradient of $f(a,b,c) = (w - w_0)^2$ where,

$$
\begin{aligned}
w &= \ln\left(v^2 + 1\right) + \cos\left(c^2 - 1\right) \\
v &= \exp\left(u^2 - 1\right) + a^2 \\
u &= \sin\left(ab\right) + cb^2 + a^3c^2
\end{aligned}
$$



$$
\begin{aligned}
\frac{\partial u}{\partial a} &= b\cos(ab) + 3a^2c^2 \\
\frac{\partial v}{\partial u} &= 2u\exp(u^2 - 1) \\
\frac{\partial v}{\partial a} &= 2a \\
\frac{\partial w}{\partial v} &= \frac{2v}{v^2+1} \\
\frac{\partial f}{\partial w} &= 2(w - w_0)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f}{\partial a} &= \left(\frac{\partial u}{\partial a}\frac{\partial v}{\partial u} + \frac{\partial v}{\partial a}\right)\frac{\partial w}{\partial v}\frac{\partial f}{\partial w} \\
&= ((b\cos(ab) + 3a^2c^2)(2u\exp(u^2-1)) + 2a)\cdot\frac{2v}{v^2+1}\cdot 2(w - w_0)
\end{aligned}
$$

$$\frac{\partial u}{\partial b} = a\cos(ab) + 2bc$$
$$\frac{\partial v}{\partial u} = 2u\exp(u^2 - 1)$$
$$\frac{\partial w}{\partial v} = \frac{2v}{v^2+1}$$
$$\frac{\partial f}{\partial w} = 2(w - w_0)$$

$$\begin{aligned}\frac{\partial f}{\partial b} &= \frac{\partial u}{\partial b}\frac{\partial v}{\partial u}\frac{\partial w}{\partial v}\frac{\partial f}{\partial w} \\ &= (a\cos(ab) + 2bc)\cdot 2u\exp(u^2-1)\cdot\frac{2v}{v^2+1}\cdot 2(w-w_0)\end{aligned}$$

$$\frac{\partial u}{\partial c} = b^2 + 2a^3c$$
$$\frac{\partial v}{\partial u} = 2u\exp\left(u^2 - 1\right)$$
$$\frac{\partial w}{\partial v} = \frac{2v}{v^2+1}$$
$$\frac{\partial w}{\partial c} = -2c\sin\left(c^2 - 1\right)$$
$$\frac{\partial f}{\partial w} = 2(w - w_0)$$

$$\begin{aligned}\frac{\partial f}{\partial c} &= \left(\frac{\partial u}{\partial c}\frac{\partial v}{\partial u}\frac{\partial w}{\partial v} + \frac{\partial w}{\partial c}\right)\frac{\partial f}{\partial w} \\ &= \left((b^2 + 2a^3c)\cdot 2u\exp\left(u^2-1\right)\cdot\frac{2v}{v^2+1} - 2c\sin\left(c^2-1\right)\right)2(w-w_0)\end{aligned}$$

In the previous method, we get that

$$
\begin{aligned}
du &= (a\cos(ab) + 2bc)da + (a\cos(ab) + 2bc)db + (b^2 + 2a^3c)dc \\
dv &= 2a\,da + 2u\exp(u^2 - 1)\,du \\
dw &= \frac{2v}{v^2+1}\,dv - 2c\sin(c^2 - 1)\,dc \\
df &= 2(w - w_0)\,dw
\end{aligned}
$$

Now, we can define the adjoints:

$$\overline{w} = \sum_{y \in A_w} \overline{y} \frac{\partial y}{\partial w} = \overline{f} \frac{\partial f}{\partial w} = 2(w - w_0)\overline{f}$$

Now, we can define the adjoints:

$$\overline{w} = \sum_{y \in A_w} \overline{y}\frac{\partial y}{\partial w} = \overline{f}\frac{\partial f}{\partial w} = 2(w - w_0)\overline{f}$$

$$\overline{v} = \sum_{y \in A_v} \overline{y}\frac{\partial y}{\partial v} = \overline{w}\frac{\partial w}{\partial v} = \frac{2v}{v^2 + 1}\overline{w}$$

Now, we can define the adjoints:

$$\overline{w} = \sum_{y \in A_w} \overline{y} \frac{\partial y}{\partial w} = \overline{f} \frac{\partial f}{\partial w} = 2(w - w_0)\overline{f}$$

$$\overline{v} = \sum_{y \in A_v} \overline{y} \frac{\partial y}{\partial v} = \overline{w} \frac{\partial w}{\partial v} = \frac{2v}{v^2 + 1} \overline{w}$$

$$\overline{u} = \sum_{y \in A_u} \overline{y} \frac{\partial y}{\partial u} = \overline{v} \frac{\partial v}{\partial u} = 2u \exp\left(u^2 - 1\right)\overline{v}$$

$$\bar{a} = \sum_{y \in A_a} \bar{y} \frac{\partial y}{\partial a} = \bar{v} \frac{\partial v}{\partial a} + \bar{u} \frac{\partial u}{\partial a} = 2a\bar{v} + (a \cos(ab) + 2bc)\bar{u}$$

$$\bar{a} = \sum_{y \in A_a} \bar{y}\frac{\partial y}{\partial a} = \bar{v}\frac{\partial v}{\partial a} + \bar{u}\frac{\partial u}{\partial a} = 2a\bar{v} + (a\cos(ab) + 2bc)\bar{u}$$

$$\bar{b} = \sum_{y \in A_b} \bar{y}\frac{\partial y}{\partial b} = \bar{u}\frac{\partial u}{\partial b} = (a\cos(ab) + 2bc)\bar{u}$$

$$\bar{a} = \sum_{y \in A_a} \bar{y} \frac{\partial y}{\partial a} = \bar{v} \frac{\partial v}{\partial a} + \bar{u} \frac{\partial u}{\partial a} = 2a\bar{v} + (a\cos(ab) + 2bc)\bar{u}$$

$$\bar{b} = \sum_{y \in A_b} \bar{y} \frac{\partial y}{\partial b} = \bar{u} \frac{\partial u}{\partial b} = (a\cos(ab) + 2bc)\bar{u}$$

$$\bar{c} = \sum_{y \in A_c} \bar{y} \frac{\partial y}{\partial c} = \bar{u} \frac{\partial u}{\partial c} + \bar{w} \frac{\partial w}{\partial c} = (b^2 + 2a^3c)\bar{u} - 2c\sin(c^2 - 1)\bar{w}$$

$$
\begin{aligned}
\bar{a} &= 2a\bar{v} + (a\cos(ab) + 2bc)\bar{u} \\
&= 2a\bar{v} + (a\cos(ab) + 2bc)\cdot 2u\exp(u^2 - 1)\bar{v} \\
&= \left(2a + (a\cos(ab) + 2bc)\cdot 2u\exp(u^2 - 1)\right)\bar{v} \\
&= \left(2a + (a\cos(ab) + 2bc)\cdot 2u\exp(u^2 - 1)\right)\frac{2v}{v^2+1}\bar{w} \\
&= \left(2a + (a\cos(ab) + 2bc)\cdot 2u\exp(u^2 - 1)\right)\frac{2v}{v^2+1}2(w - w_0)\bar{f} \\
&= \frac{\partial f}{\partial a}\bar{f}
\end{aligned}
$$

List of References

[1] M. Henrard. *Algorithmic differentiation in finance explained*. 2017.

[2] C. Homescu. "Adjoints and automatic (algorithmic) differentiation in computational finance". In: (2011).

[3] Griewank & Walther. "Evaluating derivatives: principles and techniques of algorithmic differentiation". In: (2008).