# Derivative Calculation I

- Automatic differentiation (AD) is a method to calculate the derivatives
- In this set of slides, we will introduce the two modes of automatic differentiation: forward mode and reverse mode
- Most materials in this section are from Baydin et al. (2018)
- We consider the same example function

$$f(x_1, x_2) = \log x_1 + x_1 x_2 - \sin x_2$$

# Forward Mode of AD I

• The following table demonstrates the forward calculation of f(2,5)

	$x_1$		= 2
	<i>x</i> <sub>2</sub>		= 5
	<i>v</i> <sub>1</sub>	$= \log x_1$	$= \log 2$
	<i>v</i> <sub>2</sub>	$= x_1 \times x_2$	$= 2 \times 5$
	<i>V</i> 3	$= \sin x_2$	$= \sin 5$
	<i>V</i> 4	$= v_1 + v_2$	= 0.693 + 10
	<i>V</i> 5	$= v_4 - v_3$	= 10.693 + 0.959
↓	y	$= v_5$	= 11.652

イロト 不得 トイヨト イヨト

# Forward Mode of AD II

- Variables  $x_1$  and  $x_2$  are the input nodes
- Each v<sub>i</sub>, i = 1, 2, 3, ..., is related to a simple operation
- As shown in the table, a sequence of simple operations leads to the function evaluation
- The table corresponds to the following computational graph

イロト イヨト イヨト ・

#### Forward Mode of AD III



< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

# Forward Mode of AD IV

- For example, x<sub>1</sub> is used in calculating v<sub>1</sub> and v<sub>2</sub>. So x<sub>1</sub> has two links to other nodes
- This computational graph tells us the dependencies of variables
- Remember we would like to calculate the derivative
- In particular, we check the partial derivative with respect to *x*<sub>1</sub>

$$\frac{\partial y}{\partial x_1} = \frac{\partial v_5}{\partial x_1}$$

# Forward Mode of AD V

• Here, for any variable v, we denote

$$\dot{\mathbf{v}} = rac{\partial \mathbf{v}}{\partial x_1}$$

• We will explain that using the chain rule can give us the following forward derivative calculation

< □ > < □ > < □ > < □ > < □ > < □ >

# Forward Mode of AD VI

$$\begin{array}{rclrcl}
\dot{x}_1 &= \partial x_1 / \partial x_1 &= 1 \\
\dot{x}_2 &= \partial x_2 / \partial x_1 &= 0 \\
\hline \dot{v}_1 &= \dot{x}_1 / x_1 &= 1/2 \\
\dot{v}_2 &= \dot{x}_1 \times x_2 + \dot{x}_2 \times x_1 &= 1 \times 5 + 0 \times 2 \\
\dot{v}_3 &= \dot{x}_2 \times \cos x_2 &= 0 \times \cos 5 \\
\dot{v}_4 &= \dot{v}_1 + \dot{v}_2 &= 0.5 + 5 \\
\dot{v}_5 &= \dot{v}_4 - \dot{v}_3 &= 5.5 - 0 \\
\hline \dot{y} &= \dot{v}_5 &= 5.5
\end{array}$$

• The table starts from  $\dot{x}_1 = 1$  and  $\dot{x}_2 = 0$ 

• Based on  $\dot{x}_1$  and  $\dot{x}_2$ , we can calculate other values

イロト イポト イヨト イヨト

## Forward Mode of AD VII

• For example,

$$\begin{aligned} v_1 &= \log x_1 \\ \frac{\partial v_1}{\partial x_1} &= \frac{\partial (\log x_1)}{\partial x_1} \times \frac{\partial x_1}{\partial x_1} \\ &= \frac{1}{x_1} \times \frac{\partial x_1}{\partial x_1} = \frac{\dot{x}_1}{x_1} \end{aligned}$$

• Clearly, the chain rule plays an important role here. The calculation of other  $\dot{v}_i$  is similar

イロト 不得 トイヨト イヨト

# Reverse Mode of AD I

- Next we discuss the reverse mode of automatic differentiation
- We denote

$$\bar{\mathbf{v}} = \frac{\partial \mathbf{y}}{\partial \mathbf{v}}$$

• Note that earlier, in the forward mode, we considered

$$\dot{\mathbf{v}} = \frac{\partial \mathbf{v}}{\partial x_1},$$

so the focus is on the derivatives of all variables with respect to one input variable

### Reverse Mode of AD II

- In contrast, the reverse mode focuses on the partial derivatives of one output with respect to all variables
- Here, we illustrate the calculation of

$$\frac{\partial y}{\partial x_2} = \bar{x}_2$$

 By checking the variable x<sub>2</sub> in the computational graph, we see that variable x<sub>2</sub> affects y by affecting v<sub>2</sub> and v<sub>3</sub>

## Reverse Mode of AD III



э

イロト イポト イヨト イヨト

#### Reverse Mode of AD IV

• By the chain rule, we have

$$\frac{\partial y}{\partial x_2} = \frac{\partial y}{\partial v_2} \frac{\partial v_2}{\partial x_2} + \frac{\partial y}{\partial v_3} \frac{\partial v_3}{\partial x_2}$$

or

$$ar{x}_2 = ar{v}_2 rac{\partial v_2}{\partial x_2} + ar{v}_3 rac{\partial v_3}{\partial x_2}$$

- Assume that  $\bar{v}_2$  and  $\bar{v}_3$  have been available
- Then we need to calculate

$$\frac{\partial v_2}{\partial x_2}$$
 and  $\frac{\partial v_3}{\partial x_2}$ 

### Reverse Mode of AD V

• From the relation  $v_3 = \sin x_2$ , we know

$$\frac{\partial v_3}{\partial x_2} = \cos(x_2).$$

Similarly, from  $v_2 = x_1 \times x_2$ , we have

$$\frac{\partial v_2}{\partial x_2} = x_1$$

イロト 不得 トイヨト イヨト 二日

### Reverse Mode of AD VI

• Then the evaluation of  $\bar{x}_2$  is done in two steps:

$$ar{x}_2 \leftarrow ar{v}_3 rac{\partial v_3}{\partial x_2} \ ar{x}_2 \leftarrow ar{x}_2 + ar{v}_2 rac{\partial v_2}{\partial x_2}$$

• These steps are part of the sequence of a reverse traversal, shown in the following table

イロト 不得 ト イヨト イヨト

### Reverse Mode of AD VII

$$\begin{array}{|c|c|c|c|c|c|c|} \hline \bar{x}_1 &= \bar{x}_1 &= 5.5 \\ \hline \bar{x}_2 &= \bar{x}_2 &= 1.716 \\ \hline \bar{x}_1 &= \bar{x}_1 + \bar{v}_1 \frac{\partial v_1}{\partial x_1} &= \bar{x}_1 + \bar{v}_1 / x_1 &= 5.5 \\ \hline \bar{x}_2 &= \bar{x}_2 + \bar{v}_2 \frac{\partial v_2}{\partial x_2} &= \bar{x}_2 + \bar{v}_2 \times x_1 &= 1.716 \\ \hline \bar{x}_1 &= \bar{v}_2 \frac{\partial v_2}{\partial x_1} &= \bar{v}_2 \times x_2 &= 5 \\ \hline \bar{x}_2 &= \bar{v}_3 \frac{\partial v_3}{\partial x_2} &= \bar{v}_3 \times \cos x_2 &= -0.284 \\ \hline \bar{v}_2 &= \bar{v}_4 \frac{\partial v_4}{\partial v_1} &= \bar{v}_4 \times 1 &= 1 \\ \hline \bar{v}_1 &= \bar{v}_4 \frac{\partial v_4}{\partial v_1} &= \bar{v}_5 \times (-1) &= -1 \\ \hline \bar{v}_4 &= \bar{v}_5 \frac{\partial v_5}{\partial v_4} &= \bar{v}_5 \times 1 &= 1 \\ \hline \hline \bar{v}_5 &= \bar{y} &= 1 \end{array}$$

イロト イポト イヨト イヨト 二日

#### Reverse Mode of AD VIII

• We begin with

$$ar{v}_5 = rac{\partial y}{\partial v_5} = rac{\partial y}{\partial y} = 1$$

• By the sequence of a reverse traversal, in the end, we have

$$\frac{\partial y}{\partial x_1} = \bar{x}_1$$
$$\frac{\partial y}{\partial x_2} = \bar{x}_2$$

(日)

# Reverse Mode of AD IX

• As we mentioned earlier, the reverse mode can obtain the derivatives with respect to all variables at the same time



A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind. Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research*, 18(153):1–43, 2018.

3

イロト イヨト イヨト イヨト