Automatic Differentiation

How to evaluate derivatives?

1. Symbolic Differentiation
   - Using pen-and-paper (or Maple/Wolfram-Alpha) to get algebraic representation (=formula) for derivative
   - ... and code it (although Maple can also generate python code)
   - Gives exact derivatives

How to evaluate derivatives?

1. Symbolic Differentiation
   - Using pen-and-paper (or Maple/Wolfram-Alpha) to get algebraic representation (=formula) for derivative
   - ... and code it (although Maple can also generate python code)
   - Gives exact derivatives

2. Finite Differences

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) \approx \frac{f(x+\epsilon) - f(x)}{\epsilon}$$

   - Only get approximate values

How to evaluate derivatives?

1. Symbolic Differentiation
   - Using pen-and-paper (or Maple/Wolfram-Alpha) to get algebraic representation (=formula) for derivative
   - ... and code it (although Maple can also generate python code)
   - Gives exact derivatives

2. Finite Differences

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) \approx \frac{f(x+\epsilon) - f(x)}{\epsilon}$$

   - Only get approximate values

3. Automatic Differentiation
   - Able to obtain exact derivatives for less work than either of the above methods
   - Popularised by Andreas Griewank (1989)
   - building on earlier work by

     R.E. Wengert (1964) and Seppo Linnainmaa ($\approx$ 1960)

Evaluating derivatives: work required (for $f : \mathbb{R}^n \to \mathbb{R}$)

2. Finite Differences (approximate):
   $(n+1)\times$"cost of function evaluation"

Evaluating derivatives: work required (for $f : \mathbb{R}^n \to \mathbb{R}$)

1. Symbolic Differentiation (exact):
   - Work to obtain symbolic derivatives

2. Finite Differences (approximate):
   $(n+1)\times$"cost of function evaluation"

Evaluating derivatives: work required (for $f : \mathbb{R}^n \to \mathbb{R}$)

1. Symbolic Differentiation (exact):
   - Work to obtain symbolic derivatives
   - $+$ evaluation of derivatives:
     $\approx n(?)\times$"cost of function evaluation"
     (depending on function and cleverness of implementation)

2. Finite Differences (approximate):
   $(n+1)\times$"cost of function evaluation"

Evaluating derivatives: work required (for $f : \mathbb{R}^n \to \mathbb{R}$)

1. Symbolic Differentiation (exact):
   - Work to obtain symbolic derivatives
   - $+$ evaluation of derivatives:
     $\approx n(?) \times$"cost of function evaluation"
     (depending on function and cleverness of implementation)
2. Finite Differences (approximate):
   $(n+1) \times$"cost of function evaluation"
3. Automatic Differentiation (exact):
   Less than $5 \times$"cost of function evaluation"
   (independent of the dimension of the gradient)

Evaluating derivatives: work required (for $f : \mathbb{R}^n \to \mathbb{R}$)

1. Symbolic Differentiation (exact):
   - Work to obtain symbolic derivatives
   - $+$ evaluation of derivatives:
     $\approx n(?) \times$"cost of function evaluation"
     (depending on function and cleverness of implementation)

2. Finite Differences (approximate):
   $(n+1)\times$"cost of function evaluation"

3. Automatic Differentiation (exact):
   Less than $5\times$"cost of function evaluation"
   (independent of the dimension of the gradient)

So how does that work?

AD: Function evaluation via Expression Tree

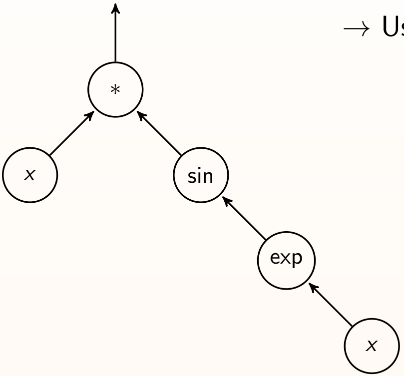
How do you evaluate

$$f(x) = x \sin(e^x) \text{ for } x = 2 \text{ ?}$$

AD: Function evaluation via Expression Tree

How do you evaluate

$$f(x) = x \sin(e^x) \text{ for } x = 2 \text{ ?}$$

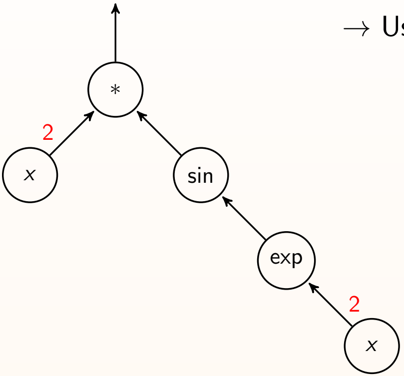


$\rightarrow$ Use expression tree

AD: Function evaluation via Expression Tree

How do you evaluate

$$f(x) = x\sin(e^x) \text{ for } x = 2 \text{ ?}$$

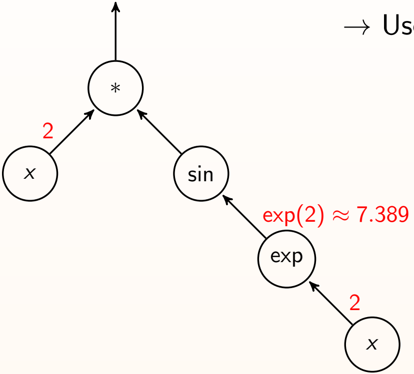


$\rightarrow$ Use expression tree

AD: Function evaluation via Expression Tree

How do you evaluate

$$f(x) = x\sin(e^x) \text{ for } x = 2 \text{ ?}$$

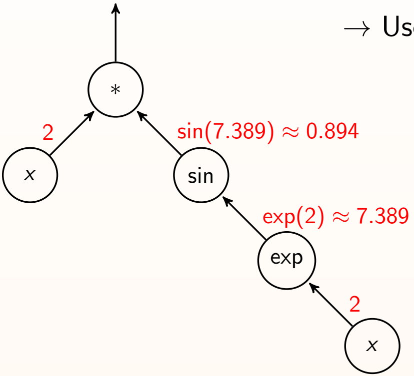


$\rightarrow$ Use expression tree

AD: Function evaluation via Expression Tree

How do you evaluate

$$f(x) = x\sin(e^x) \text{ for } x = 2 \text{ ?}$$

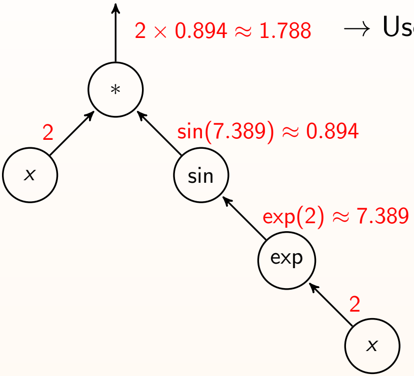


$\rightarrow$ Use expression tree

AD: Function evaluation via Expression Tree

How do you evaluate

$$f(x) = x \sin(e^x) \text{ for } x = 2 \text{ ?}$$

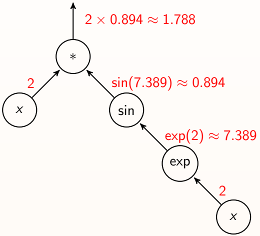


$\rightarrow$ Use expression tree

AD: Function evaluation via Expression Tree

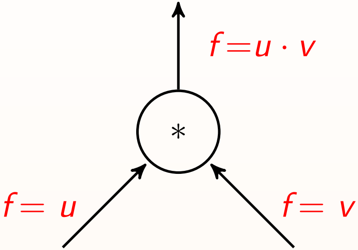
How do you evaluate

$$f(x) = x \sin(e^x) \text{ for } x = 2 \text{ ?}$$

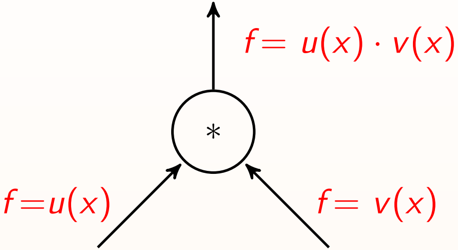


- Evaluate expression tree from bottom to top
- Every (type of) node has an associated evaluation rule
- Can use the same principle to evaluate derivatives on the same tree:
  $\rightarrow$ associate a derivative evaluation rule with every node

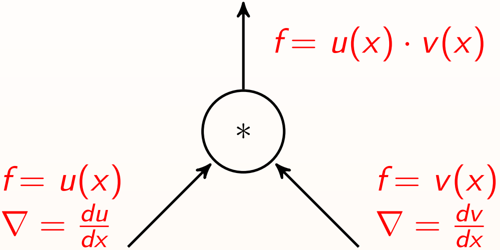
Evaluation rules for "$*$" node



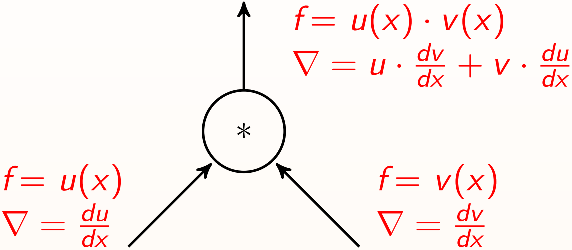
Evaluation rules for "$*$" node



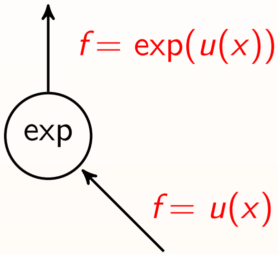
Evaluation rules for "$*$" node



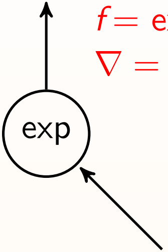
Evaluation rules for "$*$" node



Automatic Differentiation



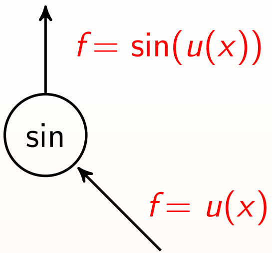
Automatic Differentiation



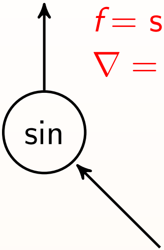
$f = \exp(u(x))$
$\nabla = \frac{du}{dx} \cdot \exp(u(x))$

$f = u(x)$
$\nabla = \frac{du}{dx}$

Automatic Differentiation



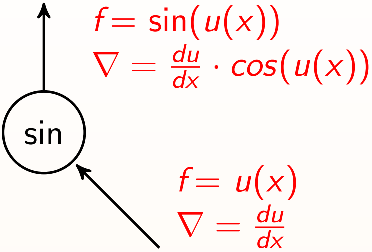
Automatic Differentiation



$$f = \sin(u(x))$$
$$\nabla = \frac{du}{dx} \cdot cos(u(x))$$

$$f = u(x)$$
$$\nabla = \frac{du}{dx}$$

Automatic Differentiation

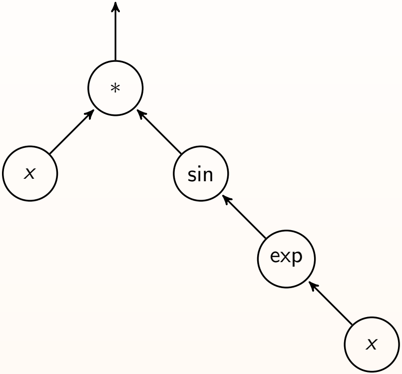


- Can use this principle to propagate up derivatives for any type of node
- For function of several variables have to propagate up partial derivatives w.r.t. each variable separately.

Automatic Differentiation: Evaluate $f(x)$ & $f'(x)$

How do you evaluate

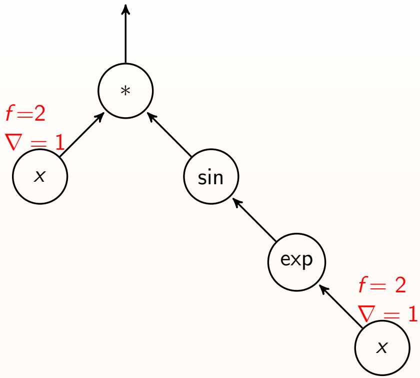
$$f(x) = x\sin(e^x) \text{ and } f'(x) = ?, \text{ for } x = 2 \text{ ?}$$



Automatic Differentiation: Evaluate $f(x)$ & $f'(x)$

How do you evaluate

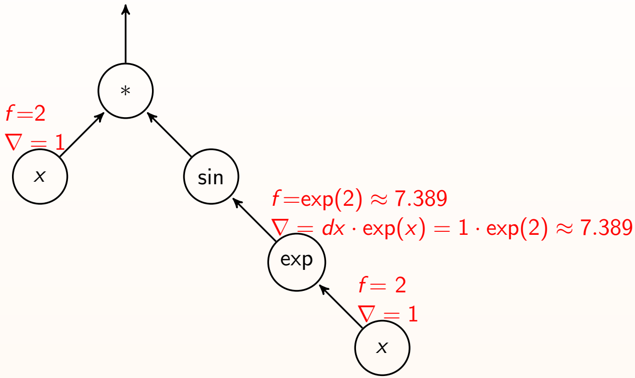
$$f(x) = x \sin(e^x) \text{ and } f'(x) = ?, \text{ for } x = 2 \text{ ?}$$



Automatic Differentiation: Evaluate $f(x)$ & $f'(x)$

How do you evaluate

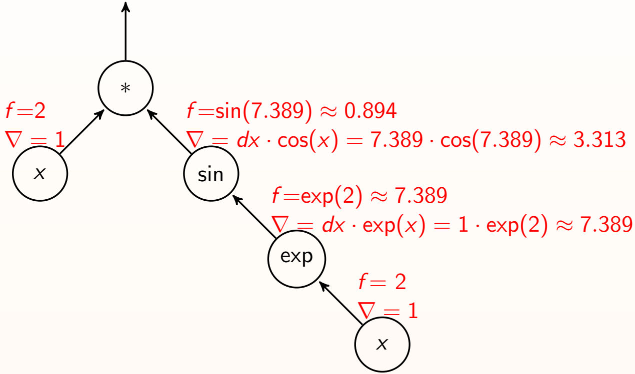
$$f(x) = x\sin(e^x) \text{ and } f'(x) = ?, \text{ for } x = 2 \text{ ?}$$



Automatic Differentiation: Evaluate $f(x)$ & $f'(x)$

How do you evaluate

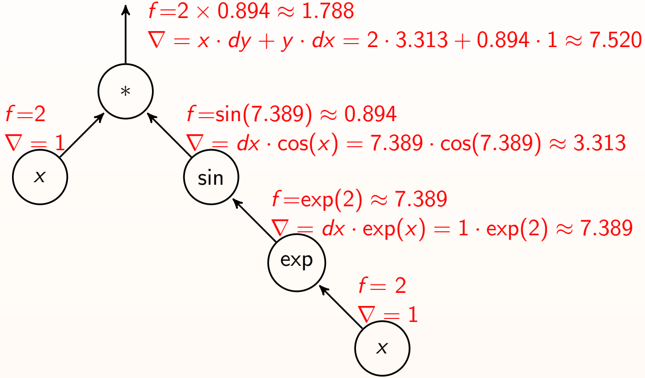
$$f(x) = x\sin(e^x) \text{ and } f'(x) = ?, \text{ for } x = 2 \text{ ?}$$



Automatic Differentiation: Evaluate $f(x)$ & $f'(x)$

How do you evaluate

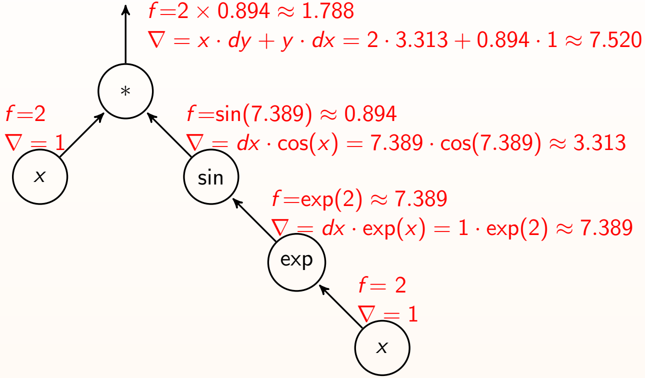
$$f(x) = x \sin(e^x) \text{ and } f'(x) = ?, \text{ for } x = 2 \text{ ?}$$



Automatic Differentiation: Evaluate $f(x)$ & $f'(x)$

How do you evaluate

$$f(x) = x\sin(e^x) \text{ and } f'(x) = ?, \text{ for } x = 2 \text{ ?}$$



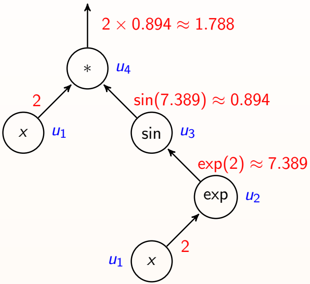
Note:

- We have never (explicitly) worked out $f'(x)$
- but still evaluated $f'(2)$!

Automatic Differentiation: Application of Chain Rule

The evaluation of the expression tree can be represented by the following chain of calculations

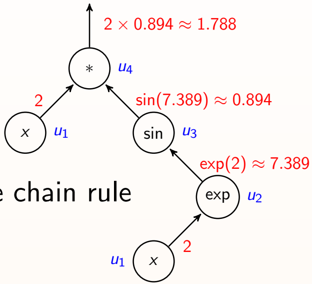
$$u_1 = x = 2$$
$$u_2 = \exp(u_1) = \exp(2) \approx 7.389$$
$$u_3 = \sin(u_2) = \sin(7.389) \approx 0.894$$
$$u_4 = u_1 \times u_3 = 2 \times 0.894 \approx 1.788$$



Automatic Differentiation: Application of Chain Rule

The evaluation of the expression tree can be represented by the following chain of calculations

$$
\begin{aligned}
u_1 &= x = 2 \\
u_2 &= \exp(u_1) = \exp(2) \approx 7.389 \\
u_3 &= \sin(u_2) = \sin(7.389) \approx 0.894 \\
u_4 &= u_1 \times u_3 = 2 \times 0.894 \approx 1.788
\end{aligned}
$$



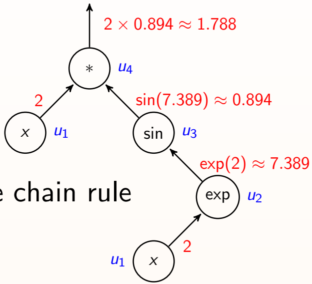
The derivatives can be calculated by the chain rule

$$
\begin{aligned}
\frac{\mathrm{d}u_4}{\mathrm{d}x} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x} \\
\frac{\mathrm{d}u_2}{\mathrm{d}x} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} \\
\frac{\mathrm{d}u_1}{\mathrm{d}x} &= \frac{\mathrm{d}x}{\mathrm{d}x} = 1
\end{aligned}
$$

Automatic Differentiation: Application of Chain Rule

The evaluation of the expression tree can be represented by the following chain of calculations

$$
\begin{aligned}
u_1 &= x = 2 \\
u_2 &= \exp(u_1) = \exp(2) \approx 7.389 \\
u_3 &= \sin(u_2) = \sin(7.389) \approx 0.894 \\
u_4 &= u_1 \times u_3 = 2 \times 0.894 \approx 1.788
\end{aligned}
$$



The derivatives can be calculated by the chain rule

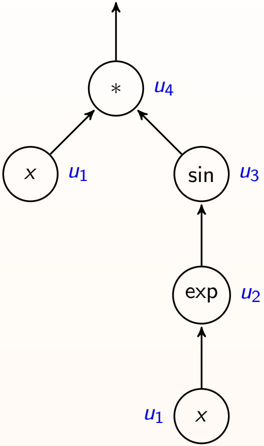
$$
\begin{aligned}
\frac{\mathrm{d}u_4}{\mathrm{d}x} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x} \\
\frac{\mathrm{d}u_2}{\mathrm{d}x} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} \\
\frac{\mathrm{d}u_1}{\mathrm{d}x} &= \frac{\mathrm{d}x}{\mathrm{d}x} = 1
\end{aligned}
$$

$\Rightarrow$ The AD process propagates these calculation up through the expression tree.

Automatic Differentiation: Application of Chain Rule

$$
\begin{aligned}
u_1 &= x = 2 \\
u_2 &= \exp(u_1) \\
&= \exp(2) \approx 7.389 \\
u_3 &= \sin(u_2) \\
&= \sin(7.389) \approx 0.894 \\
u_4 &= u_1 \times u_3 \\
&= 2 \times 0.894 \approx 1.788
\end{aligned}
$$

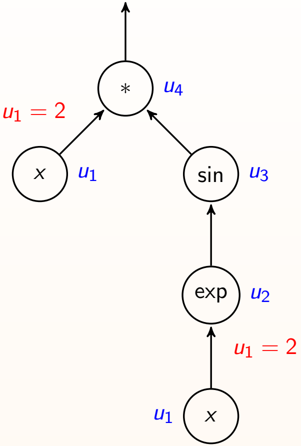
$$
\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}x} &= \frac{\mathrm{d}x}{\mathrm{d}x} = 1 \\
\frac{\mathrm{d}u_2}{\mathrm{d}x} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x} \\
\frac{\mathrm{d}u_4}{\mathrm{d}x} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x}
\end{aligned}
$$



Automatic Differentiation: Application of Chain Rule

$$
\begin{aligned}
u_1 &= x = 2 \\
u_2 &= \exp(u_1) \\
&= \exp(2) \approx 7.389 \\
u_3 &= \sin(u_2) \\
&= \sin(7.389) \approx 0.894 \\
u_4 &= u_1 \times u_3 \\
&= 2 \times 0.894 \approx 1.788
\end{aligned}
$$

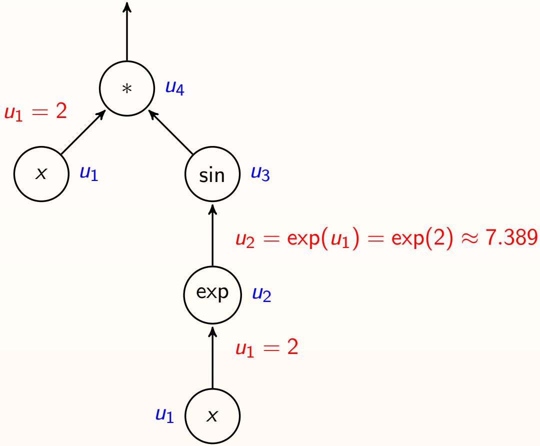
$$
\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}x} &= \frac{\mathrm{d}x}{\mathrm{d}x} = 1 \\
\frac{\mathrm{d}u_2}{\mathrm{d}x} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x} \\
\frac{\mathrm{d}u_4}{\mathrm{d}x} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x}
\end{aligned}
$$



Automatic Differentiation: Application of Chain Rule

$$u_1 = x = 2$$
$$u_2 = \exp(u_1)$$
$$= \exp(2) \approx 7.389$$
$$u_3 = \sin(u_2)$$
$$= \sin(7.389) \approx 0.894$$
$$u_4 = u_1 \times u_3$$
$$= 2 \times 0.894 \approx 1.788$$

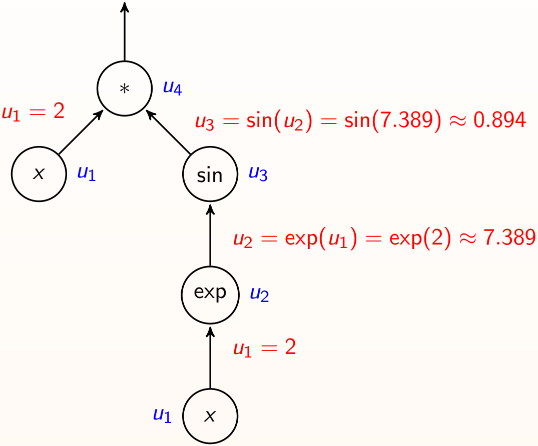
$$\frac{du_1}{dx} = \frac{dx}{dx} = 1$$
$$\frac{du_2}{dx} = \frac{du_2}{du_1}\frac{du_1}{dx}$$
$$\frac{du_3}{dx} = \frac{du_3}{du_2}\frac{du_2}{dx}$$
$$\frac{du_4}{dx} = \frac{du_4}{du_1}\frac{du_1}{dx} + \frac{du_4}{du_3}\frac{du_3}{dx}$$



Automatic Differentiation: Application of Chain Rule

$$
\begin{aligned}
u_1 &= x = 2 \\
u_2 &= \exp(u_1) \\
&= \exp(2) \approx 7.389 \\
u_3 &= \sin(u_2) \\
&= \sin(7.389) \approx 0.894 \\
u_4 &= u_1 \times u_3 \\
&= 2 \times 0.894 \approx 1.788
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}x} &= \frac{\mathrm{d}x}{\mathrm{d}x} = 1 \\
\frac{\mathrm{d}u_2}{\mathrm{d}x} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x} \\
\frac{\mathrm{d}u_4}{\mathrm{d}x} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x}
\end{aligned}
$$



Automatic Differentiation: Application of Chain Rule

$$u_1 = x = 2$$

$$u_2 = \exp(u_1) = \exp(2) \approx 7.389$$

$$u_3 = \sin(u_2) = \sin(7.389) \approx 0.894$$

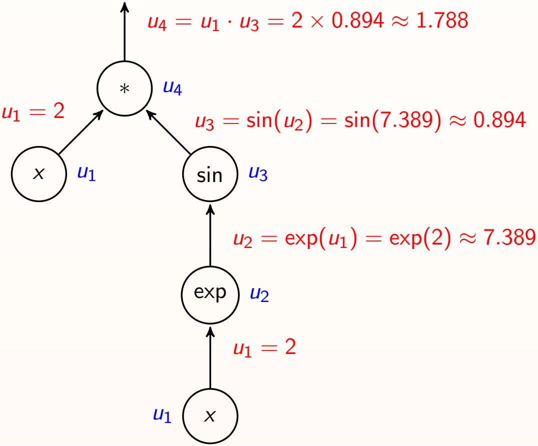
$$u_4 = u_1 \times u_3 = 2 \times 0.894 \approx 1.788$$

$$\frac{du_1}{dx} = \frac{dx}{dx} = 1$$

$$\frac{du_2}{dx} = \frac{du_2}{du_1}\frac{du_1}{dx}$$

$$\frac{du_3}{dx} = \frac{du_3}{du_2}\frac{du_2}{dx}$$

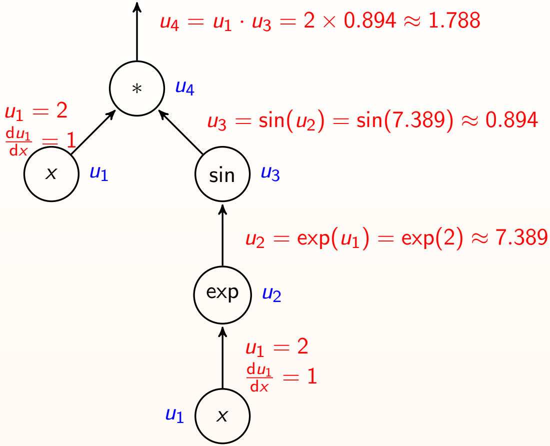
$$\frac{du_4}{dx} = \frac{du_4}{du_1}\frac{du_1}{dx} + \frac{du_4}{du_3}\frac{du_3}{dx}$$



Automatic Differentiation: Application of Chain Rule

$$
\begin{aligned}
u_1 &= x = 2 \\
u_2 &= \exp(u_1) \\
&= \exp(2) \approx 7.389 \\
u_3 &= \sin(u_2) \\
&= \sin(7.389) \approx 0.894 \\
u_4 &= u_1 \times u_3 \\
&= 2 \times 0.894 \approx 1.788
\end{aligned}
$$

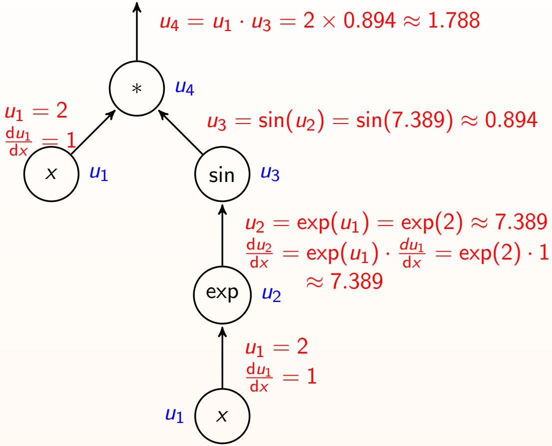
$$
\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}x} &= \frac{\mathrm{d}x}{\mathrm{d}x} = 1 \\
\frac{\mathrm{d}u_2}{\mathrm{d}x} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x} \\
\frac{\mathrm{d}u_4}{\mathrm{d}x} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x}
\end{aligned}
$$



Automatic Differentiation: Application of Chain Rule

$$
\begin{aligned}
u_1 &= x = 2 \\
u_2 &= \exp(u_1) \\
&= \exp(2) \approx 7.389 \\
u_3 &= \sin(u_2) \\
&= \sin(7.389) \approx 0.894 \\
u_4 &= u_1 \times u_3 \\
&= 2 \times 0.894 \approx 1.788
\end{aligned}
$$

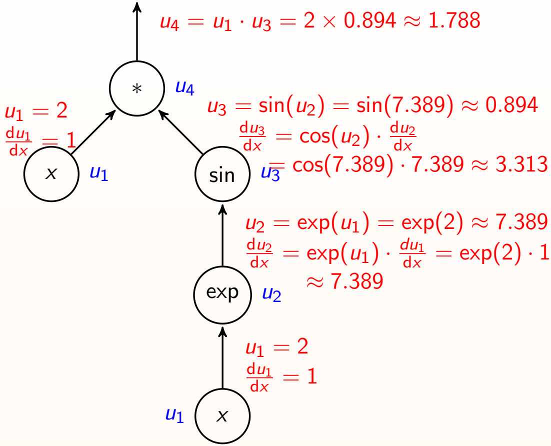
$$
\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}x} &= \frac{\mathrm{d}x}{\mathrm{d}x} = 1 \\
\frac{\mathrm{d}u_2}{\mathrm{d}x} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x} \\
\frac{\mathrm{d}u_4}{\mathrm{d}x} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x}
\end{aligned}
$$



Automatic Differentiation: Application of Chain Rule

$$
\begin{aligned}
u_1 &= x = 2 \\
u_2 &= \exp(u_1) \\
&= \exp(2) \approx 7.389 \\
u_3 &= \sin(u_2) \\
&= \sin(7.389) \approx 0.894 \\
u_4 &= u_1 \times u_3 \\
&= 2 \times 0.894 \approx 1.788
\end{aligned}
$$

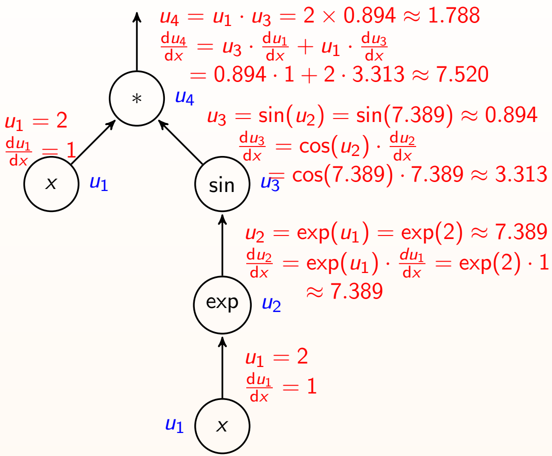
$$
\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}x} &= \frac{\mathrm{d}x}{\mathrm{d}x} = 1 \\
\frac{\mathrm{d}u_2}{\mathrm{d}x} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x} \\
\frac{\mathrm{d}u_4}{\mathrm{d}x} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x}
\end{aligned}
$$



Automatic Differentiation: Application of Chain Rule

$$
\begin{aligned}
u_1 &= x = 2 \\
u_2 &= \exp(u_1) \\
&= \exp(2) \approx 7.389 \\
u_3 &= \sin(u_2) \\
&= \sin(7.389) \approx 0.894 \\
u_4 &= u_1 \times u_3 \\
&= 2 \times 0.894 \approx 1.788
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}x} &= \frac{\mathrm{d}x}{\mathrm{d}x} = 1 \\
\frac{\mathrm{d}u_2}{\mathrm{d}x} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x} \\
\frac{\mathrm{d}u_4}{\mathrm{d}x} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x}
\end{aligned}
$$



Forward Mode Automatic Differentiation

- The forward mode of automatic differentiation accumulates derivatives by evaluating (from the bottom of the tree up)

$$w_i = \frac{\mathrm{d}u_i}{\mathrm{d}x} = \sum_{j \in \{\text{children of } i\}} \frac{\mathrm{d}u_i}{\mathrm{d}u_j}\frac{\mathrm{d}u_j}{\mathrm{d}x} = \sum_{j \in \{\text{children of } i\}} \frac{\mathrm{d}u_i}{\mathrm{d}u_j} w_j$$

- It does so by one upward (=forward) pass through the tree evaluating function value u_i and derivative $w_i = \frac{\mathrm{d}u_i}{\mathrm{d}x}$ for each node at the same time.
- Work required is *a bit more than twice* (actually $1 + \omega$ times)[1] that of just evaluating the function

[1] The factor ω has been stated as 1.5, 3 or 5

Forward Mode Automatic Differentiation

- The forward mode of automatic differentiation accumulates derivatives by evaluating (from the bottom of the tree up)

$$w_i = \frac{\mathrm{d}u_i}{} = \sum \quad \frac{\mathrm{d}u_i\,\mathrm{d}u_j}{} = \sum \quad \frac{\mathrm{d}u_i}{} w$$

Can use the same process for functions of several variables: $f(x, y)$. In that case we evaluate

$$u_i, \qquad w_i^x = \frac{\mathrm{d}u_i}{\mathrm{d}x}, \qquad w_i^y = \frac{\mathrm{d}u_i}{\mathrm{d}y}$$

in the same pass

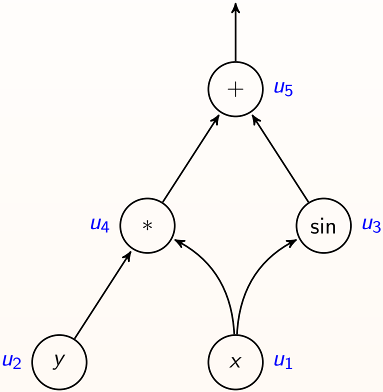
- It does … evaluat … :h node a …
- Work r … es)[1] that of just evaluating the function

[1] The factor ω has been stated as 1.5, 3 or 5

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

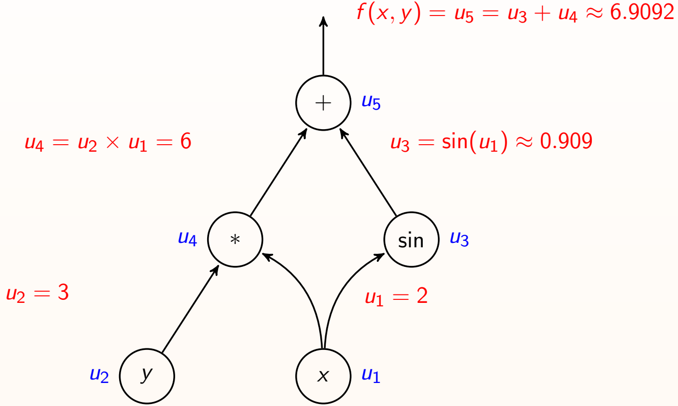


- Forward AD passes up

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

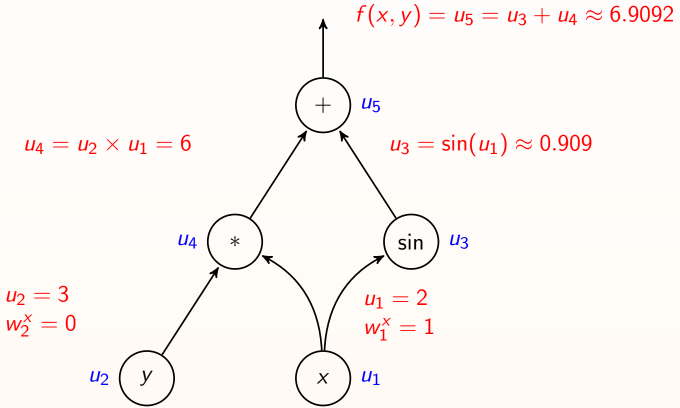


- Forward AD passes up values u_i

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x,y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

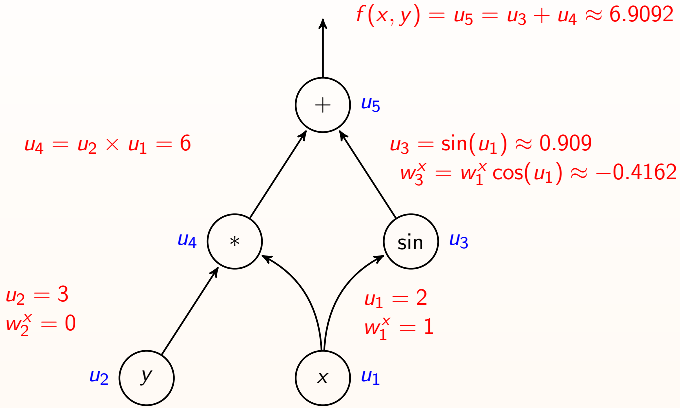


- Forward AD passes up values u_i
- and derivaties $w_i^x = \frac{\mathrm{d}u_i}{\mathrm{d}x}$

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

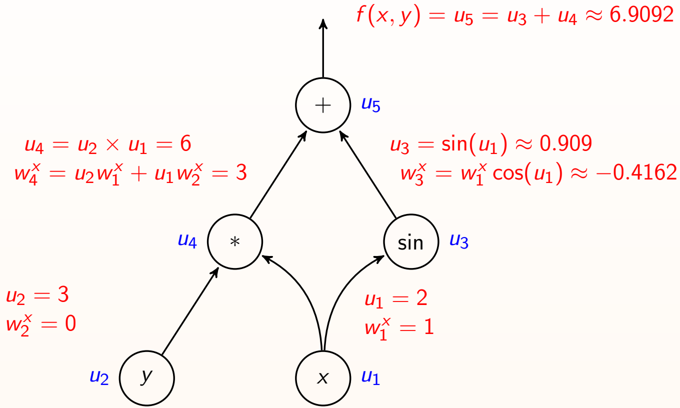


- Forward AD passes up values u_i
- and derivaties $w_i^x = \frac{\mathrm{d}u_i}{\mathrm{d}x}$

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

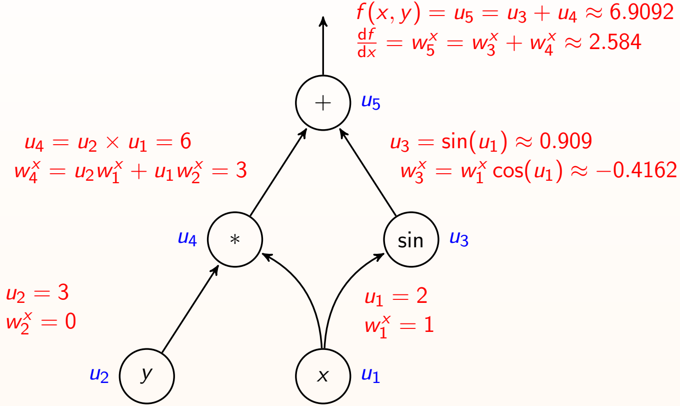


- Forward AD passes up values u_i
- and derivaties $w_i^x = \frac{\mathrm{d}u_i}{\mathrm{d}x}$

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

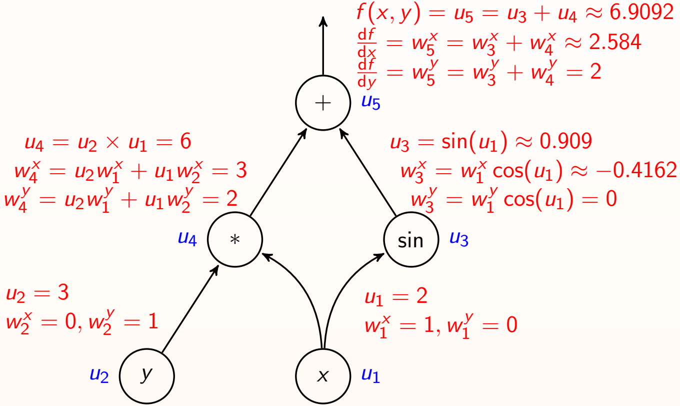


- Forward AD passes up values u_i
- and derivaties $w_i^x = \frac{du_i}{dx}$

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

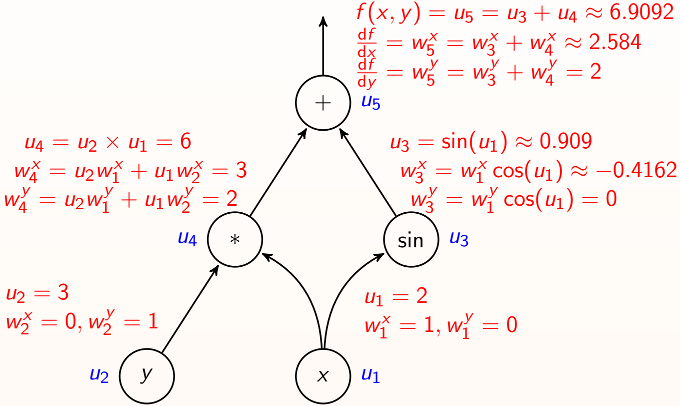


- Forward AD passes up values u_i
- and derivaties $w_i^x = \frac{du_i}{dx}$, $w_i^y = \frac{du_i}{dy}$

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$



- Forward AD passes up values u_i
- and derivaties $w_i^x = \frac{\mathrm{d}u_i}{\mathrm{d}x}, w_i^y = \frac{\mathrm{d}u_i}{\mathrm{d}y}$

$\Big\}$ evaluated in one forward (upwards) pass

Automatic Differentiation: Functions of several variables

- Evaluating function and (complete) gradient requires one pass of the expression tree
- At each node evaluate the function and chain rule propagation for n components of the gradient
- Total work is $(1 + \omega n)\times$ "cost of function evaluation"

Automatic Differentiation: Functions of several variables

- Evaluating function and (complete) gradient requires one pass of the expression tree
- At each node evaluate the function and chain rule propagation for n components of the gradient
- Total work is $(1 + \omega n)\times$ "cost of function evaluation"

It turns out we can do better!

Reverse Mode Automatic Differentiation

The Forward AD process accumulates the following (by traversing the tree):

$$
\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}x_k} &= \frac{\mathrm{d}x_k}{\mathrm{d}x_k} = 1 \\
\frac{\mathrm{d}u_2}{\mathrm{d}x_k} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x_k} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x_k} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x_k} \\
\frac{\mathrm{d}u_4}{\mathrm{d}x_k} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x_k} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x_k}
\end{aligned}
$$

Reverse Mode Automatic Differentiation

The Forward AD process accumulates the following (by traversing the tree):

$$\frac{\mathrm{d}u_1}{\mathrm{d}x_k} = \frac{\mathrm{d}x_k}{\mathrm{d}x_k} = 1$$
$$\frac{\mathrm{d}u_2}{\mathrm{d}x_k} = \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x_k}$$
$$\frac{\mathrm{d}u_3}{\mathrm{d}x_k} = \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x_k}$$
$$\frac{\mathrm{d}u_4}{\mathrm{d}x_k} = \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x_k} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x_k}$$

Resulting in the evaluation

$$\frac{\mathrm{d}u_4}{\mathrm{d}x_k} = \left(\frac{\mathrm{d}u_4}{\mathrm{d}u_1} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}u_1}\right)\frac{\mathrm{d}u_1}{\mathrm{d}x_k}$$

Reverse Mode Automatic Differentiation

The Forward AD process accumulates the following (by traversing the tree):

$$
\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}x_k} &= \frac{\mathrm{d}x_k}{\mathrm{d}x_k} = 1 \\
\frac{\mathrm{d}u_2}{\mathrm{d}x_k} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x_k} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x_k} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x_k} \\
\frac{\mathrm{d}u_4}{\mathrm{d}x_k} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x_k} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x_k}
\end{aligned}
$$

Resulting in the evaluation

$$
\frac{\mathrm{d}u_4}{\mathrm{d}x_k} = \left(\frac{\mathrm{d}u_4}{\mathrm{d}u_1} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}u_1}\right)\frac{\mathrm{d}u_1}{\mathrm{d}x_k}
$$

- The tree traversal process accumulates this from right to left
- Only the final (right most) term depends on x_k (the variable differentiated w.r.t)

Reverse Mode Automatic Differentiation

The Forward AD process accumulates the following (by traversing the tree):

$$
\begin{aligned}
\frac{\mathrm{d}u_1}{\mathrm{d}x_k} &= \frac{\mathrm{d}x_k}{\mathrm{d}x_k} = 1 \\
\frac{\mathrm{d}u_2}{\mathrm{d}x_k} &= \frac{\mathrm{d}u_2}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x_k} \\
\frac{\mathrm{d}u_3}{\mathrm{d}x_k} &= \frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}x_k} \\
\frac{\mathrm{d}u_4}{\mathrm{d}x_k} &= \frac{\mathrm{d}u_4}{\mathrm{d}u_1}\frac{\mathrm{d}u_1}{\mathrm{d}x_k} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}x_k}
\end{aligned}
$$

Resulting in the evaluation

$$
\frac{\mathrm{d}u_4}{\mathrm{d}x_k} = \left(\frac{\mathrm{d}u_4}{\mathrm{d}u_1} + \frac{\mathrm{d}u_4}{\mathrm{d}u_3}\frac{\mathrm{d}u_3}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}u_1}\right)\frac{\mathrm{d}u_1}{\mathrm{d}x_k}
$$

- The tree traversal process accumulates this from right to left
- Only the final (right most) term depends on x_k (the variable differentiated w.r.t)
- To get derivatives w.r.t. other variables could reverse the process

$\rightarrow$ Accumulate downwards on tree!

Reverse Mode Automatic Differentiation

Reverse mode AD accumulates the derivatives by traversing the tree from top to bottom to evaluate

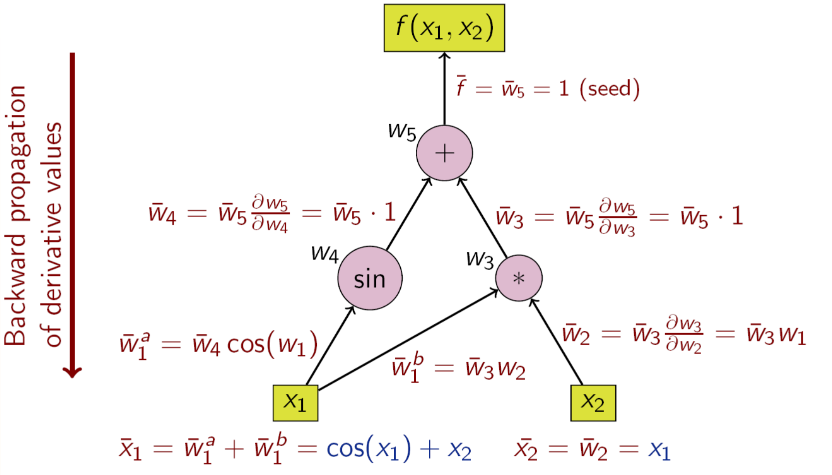
$$\bar{w}_i = \frac{\mathrm{d}f}{\mathrm{d}u_i}$$

Chain rule

$$f(u_1) = f(u_2(u_1)) \Rightarrow \frac{\mathrm{d}f}{\mathrm{d}u_1} = \frac{\mathrm{d}f}{\mathrm{d}u_2}\frac{\mathrm{d}u_2}{\mathrm{d}u_1}$$

$$\Rightarrow \bar{w}_1 = \bar{w}_2 \frac{\mathrm{d}u_2}{\mathrm{d}u_1}$$

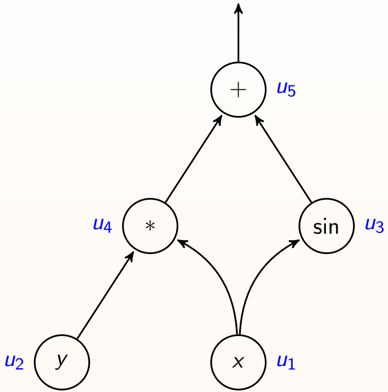
Reverse Mode Automatic Differentiation



Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

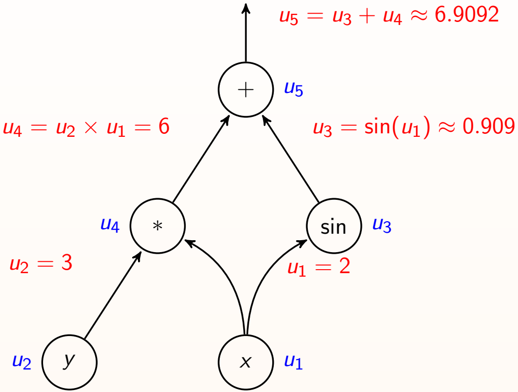


- Reverse AD passes up as before

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

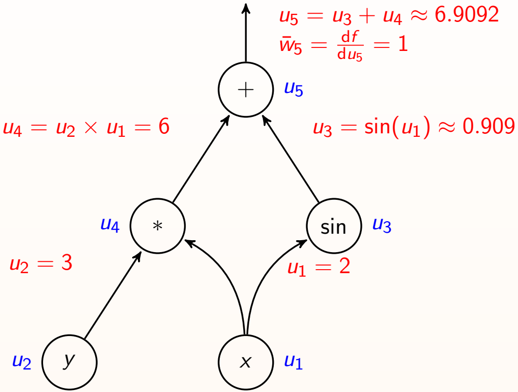


- Reverse AD passes up values u_i as before

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

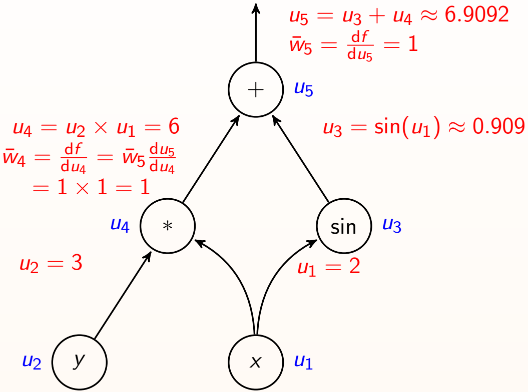


- Reverse AD passes up values u_i as before
- Adjoints $\bar{w}_i = \frac{\mathrm{d}f}{\mathrm{d}u_i}$ are passed top to bottom (reverse).

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

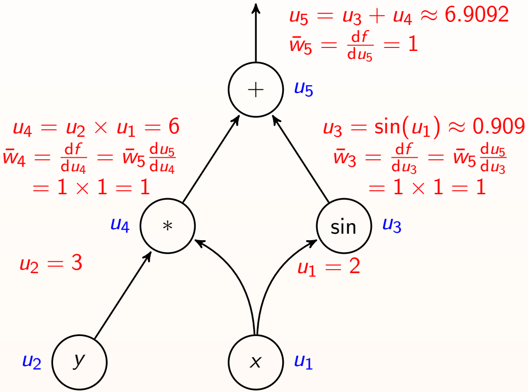


- Reverse AD passes up values u_i as before
- Adjoints $\bar{w}_i = \frac{df}{du_i}$ are passed top to bottom (reverse).

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

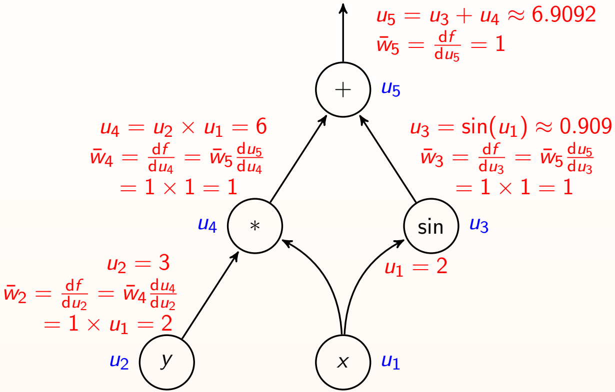


- Reverse AD passes up values u_i as before
- Adjoints $\bar{w}_i = \frac{\mathrm{d}f}{\mathrm{d}u_i}$ are passed top to bottom (reverse).

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

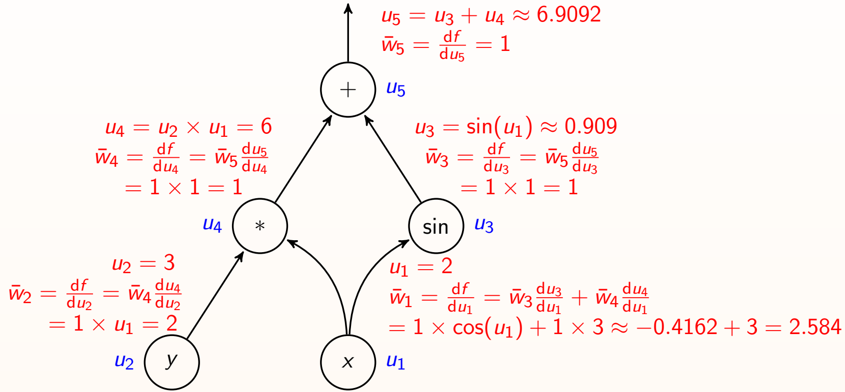


- Reverse AD passes up values u_i as before
- Adjoints $\bar{w}_i = \frac{\mathrm{d}f}{\mathrm{d}u_i}$ are passed top to bottom (reverse).

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$

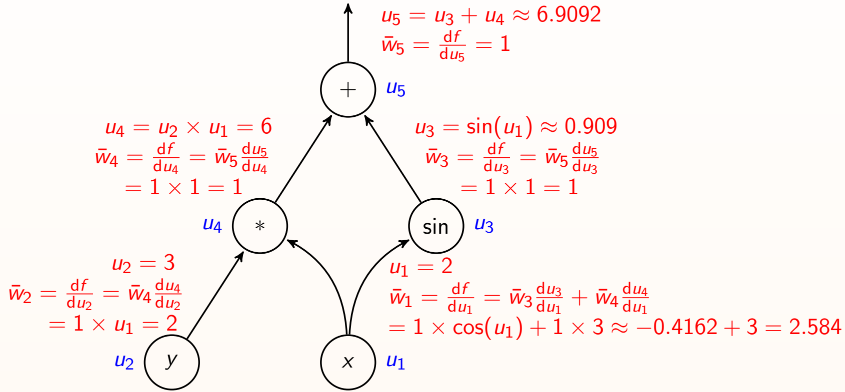


- Reverse AD passes up values u_i as before
- Adjoints $\bar{w}_i = \frac{\mathrm{d}f}{\mathrm{d}u_i}$ are passed top to bottom (reverse).

Automatic Differentiation: Functions of several variables

How do you evaluate

$$f(x, y) = xy + \sin(x) \text{ for } x = 2, y = 3?$$



- Reverse AD passes up values u_i as before
- Adjoints $\bar{w}_i = \frac{\mathrm{d}f}{\mathrm{d}u_i}$ are passed top to bottom (reverse).
- Derivaties $\frac{\mathrm{d}f}{\mathrm{d}x} = \bar{w}_2$, $\frac{\mathrm{d}f}{\mathrm{d}y} = \bar{w}_1$ obtained in same pass

Reverse Mode Automatic Differentiation: Complexity

The reverse mode of AD obtains

- function value $f(x, y)$ in one upward (forward) pass
- The complete gradient in one downward (reverse) pass
- Total work is $(1 + \omega)\times$ "cost of function evaluation"

Reverse Mode Automatic Differentiation: Complexity

The reverse mode of AD obtains

- function value $f(x, y)$ in one upward (forward) pass
- The complete gradient in one downward (reverse) pass
- Total work is $(1 + \omega)\times$ "cost of function evaluation"

Theorem (Complexity of Reverse Mode AD)

The reverse mode of AD is able to obtain the complete exact gradient of a $\mathbb{R}^n \to \mathbb{R}$ function at a cost of less than $5\times$ the cost of a function evaluation.
[Independent of the dimension of the gradient (n)!]

Automatic Differentiation: other issue

- AD works on an evaluation tree!
  - typically originating from a symbolic formula
  - could use (almost any) sequence of calculations (algorithm!)
- Calculating Hessians (second derivatives):
  - Can use forward mode as before. Need tree traversal for every element of the Hessian!(expensive)
  - Can use backward mode to get gradient and then forward mode for Hessian (or vice versa)
  - Not possible to do pure backward mode
- Implementation
  - AD is built into all modelling systems (XPressMP, AMPL, etc)
  - Also many libraries/modules available

NEOS

NEOS: Network-Enabled Optimization System

https://neos-server.org

NLP Solvers on NEOS

- Filter: SQP
- CONOPT: Generalised Reduced Gradients
- IPOPT: Interior Point
- SNOPT: SQP
- LOQO: Interior Point
- MINOS: projected Augmented Lagrangian
- Mosek: Interior Point
- Knitro: Interior Point or SQP
- LANCELOT: Augmented Lagrangian