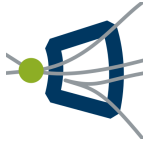


Data Analysis Block Course: Statistics and Machine learning



Exercise: Forward and Reverse Automatic Differentiation by Hand - Solutions

Prof. Dr. Mikhail Mikhasenko

Assistants: Hendric Jonas, Christoph Tönnis



25 September 2026

[Indico schedule](#)

[RUB Moodle](#)

Part I - Forward-mode AD with dual numbers

Solution 1.1 - Scalar input, scalar output

We evaluate $f(x) = \log(xe^x + 3)$ at $x = 1$. Start with the dual number $\hat{x} = 1 + \varepsilon$, so the input tangent is $\dot{x} = 1$. The forward-mode rule is

$$f(x + \dot{x}\varepsilon) = f(x) + f'(x)\dot{x}\varepsilon \quad (1)$$

Apply the elementary operations in order, dropping all terms containing ε^2 :

1. **Exponential:**

$$a = e^{\hat{x}} = e^{1+\varepsilon} = e + e\varepsilon$$

2. **Multiplication:**

$$b = \hat{x}a = (1 + \varepsilon)(e + e\varepsilon) = e + 2e\varepsilon$$

3. **Addition:**

$$c = b + 3 = (e + 3) + 2e\varepsilon$$

4. **Logarithm:**

$$f(\hat{x}) = \log(e + 3) + \frac{2e}{e + 3}\varepsilon$$

The primal value and tangent therefore give

$$f(1) = \log(e + 3) \quad f'(1) = \frac{2e}{e + 3}$$

Solution 1.2 - Vector input, scalar output

For $g(x, y, z) = xy + \sin z + y^2$, the ordinary function value is

$$g(1, 2, 0) = 1 \cdot 2 + \sin 0 + 2^2 = 6$$

The three basis seeds

Each sweep starts again at $(1, 2, 0)$. Set the tangent in the selected direction to 1 and all other input tangents to 0.

1. **Seed in the x-direction:** Use

$$\hat{x} = 1 + \varepsilon \quad \hat{y} = 2 \quad \hat{z} = 0$$

Forward propagation gives

$$\begin{aligned}\hat{x}\hat{y} &= 2 + 2\varepsilon \\ \sin \hat{z} &= 0 \\ \hat{y}^2 &= 4 \\ g(\hat{x}, \hat{y}, \hat{z}) &= 6 + 2\varepsilon\end{aligned}$$

Hence $\partial g / \partial x = 2$ at the chosen point.

2. **Seed in the y-direction:** Use

$$\hat{x} = 1 \quad \hat{y} = 2 + \varepsilon \quad \hat{z} = 0$$

Forward propagation gives

$$\begin{aligned}\hat{x}\hat{y} &= 2 + \varepsilon \\ \sin \hat{z} &= 0 \\ \hat{y}^2 &= (2 + \varepsilon)^2 = 4 + 4\varepsilon \\ g(\hat{x}, \hat{y}, \hat{z}) &= 6 + 5\varepsilon\end{aligned}$$

Hence $\partial g / \partial y = 5$.

3. **Seed in the z-direction:** Use

$$\hat{x} = 1 \quad \hat{y} = 2 \quad \hat{z} = \varepsilon$$

Since $\sin(\varepsilon) = \sin 0 + \cos 0 \varepsilon = \varepsilon$, we obtain

$$g(\hat{x}, \hat{y}, \hat{z}) = 2 + \varepsilon + 4 = 6 + \varepsilon$$

Hence $\partial g / \partial z = 1$.

The full gradient is therefore

$$\nabla g(1, 2, 0) = \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix}$$

Directional derivative

For $\mathbf{v} = (1, -1, 2)^T$, seed all inputs at once:

$$\hat{x} = 1 + \varepsilon \quad \hat{y} = 2 - \varepsilon \quad \hat{z} = 2\varepsilon$$

One forward sweep gives

$$\begin{aligned}\hat{x}\hat{y} &= 2 + \varepsilon \\ \sin \hat{z} &= 2\varepsilon \\ \hat{y}^2 &= 4 - 4\varepsilon \\ g(\hat{x}, \hat{y}, \hat{z}) &= 6 - \varepsilon\end{aligned}$$

Thus $D_{\mathbf{v}}g = -1$, in agreement with the check

$$\nabla g^T \mathbf{v} = 2(1) + 5(-1) + 1(2) = -1$$

Solution 1.3 - Vector input, vector output

For $\mathbf{h}(x, y, z) = (xy + z, x^2 + \sin y - z^2)^T$, the ordinary output is

$$\mathbf{h}(1, 0, 1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

The three basis seeds

Each sweep starts again at $(1, 0, 1)$ with all unselected input tangents set to zero.

1. **Seed in the x-direction:** Use

$$\hat{x} = 1 + \varepsilon \quad \hat{y} = 0 \quad \hat{z} = 1$$

The two output components are

$$\begin{aligned}\hat{h}_1 &= (1 + \varepsilon)0 + 1 = 1 + 0\varepsilon \\ \hat{h}_2 &= (1 + \varepsilon)^2 + \sin 0 - 1 = 2\varepsilon\end{aligned}$$

The first Jacobian column is $(0, 2)^T$.

2. **Seed in the y-direction:** Use

$$\hat{x} = 1 \quad \hat{y} = \varepsilon \quad \hat{z} = 1$$

The output components are

$$\begin{aligned}\hat{h}_1 &= 1 + \varepsilon \\ \hat{h}_2 &= \sin(\varepsilon) = \varepsilon\end{aligned}$$

The second Jacobian column is $(1, 1)^T$.

3. **Seed in the z-direction:** Use

$$\hat{x} = 1 \quad \hat{y} = 0 \quad \hat{z} = 1 + \varepsilon$$

The output components are

$$\begin{aligned}\hat{h}_1 &= 1 + \varepsilon \\ \hat{h}_2 &= 1 - (1 + \varepsilon)^2 = -2\varepsilon\end{aligned}$$

The third Jacobian column is $(1, -2)^T$.

Putting the columns together gives

$$J_{\mathbf{h}}(1, 0, 1) = \begin{pmatrix} 0 & 1 & 1 \\ 2 & 1 & -2 \end{pmatrix}$$

One sweep in the specified direction

For $\mathbf{v} = (1, 2, -1)^T$, use

$$\hat{x} = 1 + \varepsilon \quad \hat{y} = 2\varepsilon \quad \hat{z} = 1 - \varepsilon$$

One forward sweep gives

$$\begin{aligned}\hat{h}_1 &= (1 + \varepsilon)2\varepsilon + (1 - \varepsilon) = 1 + \varepsilon \\ \hat{h}_2 &= (1 + \varepsilon)^2 + \sin(2\varepsilon) - (1 - \varepsilon)^2 = 6\varepsilon\end{aligned}$$

The tangent vector is therefore

$$J_{\mathbf{h}}\mathbf{v} = \begin{pmatrix} 1 \\ 6 \end{pmatrix}$$

This agrees with multiplication of the Jacobian by $\mathbf{v}$.

Part II - Reverse-mode AD / Backward AD

Solution 2.1 - Scalar input, scalar output

Write $f(x) = \log(xe^x + 3)$ as elementary operations:

$$\begin{aligned}v_1 &= e^x \\ v_2 &= xv_1 \\ v_3 &= v_2 + 3 \\ v_4 &= \log v_3 \\ f &= v_4\end{aligned}$$

Forward pass

At $x = 1$, the intermediate values are

$$v_1 = e \quad v_2 = e \quad v_3 = e + 3 \quad v_4 = \log(e + 3)$$

Backward pass

Initialize all adjoints to zero, then set $\bar{v}_4 = 1$. Propagate backwards through the logarithm and addition:

$$\begin{aligned}\bar{v}_3 &= \bar{v}_4 \frac{1}{v_3} = \frac{1}{e + 3} \\ \bar{v}_2 &= \bar{v}_3 = \frac{1}{e + 3}\end{aligned}$$

The multiplication $v_2 = xv_1$ gives a direct contribution to x and an adjoint for v_1 :

$$\begin{aligned}\bar{x}_{\text{direct}} &= \bar{v}_2 v_1 = \frac{e}{e + 3} \\ \bar{v}_1 &= \bar{v}_2 x = \frac{1}{e + 3}\end{aligned}$$

Finally, the exponential $v_1 = e^x$ contributes

$$\bar{x}_{\text{via } v_1} = \bar{v}_1 e^x = \frac{e}{e + 3}$$

Add both paths to obtain

$$f'(1) = \bar{x} = \frac{e}{e+3} + \frac{e}{e+3} = \frac{2e}{e+3}$$

This is the same derivative obtained with dual numbers in Exercise 1.1.

Solution 2.2 - Three inputs, scalar output

Use $a = xy$, $b = \sin z$, $c = y^2$ and $d = a + b + c$, with scalar output $g = d$.

Forward pass

At $(x, y, z) = (1, 2, 0)$, the intermediate values are

$$a = 2 \quad b = 0 \quad c = 4 \quad d = 6$$

Backward pass

Initialize all adjoints to zero, then set $\bar{d} = 1$. Since $d = a + b + c$, this gives

$$\bar{a} = 1 \quad \bar{b} = 1 \quad \bar{c} = 1$$

Apply the local backward rules:

- **From** $a = xy$:

$$\bar{x} += \bar{a}y = 2 \quad \bar{y} += \bar{a}x = 1$$

- **From** $b = \sin z$:

$$\bar{z} += \bar{b} \cos z = 1$$

- **From** $c = y^2$:

$$\bar{y} += \bar{c}2y = 4$$

The adjoint of y receives contributions through both xy and y^2 , so $\bar{y} = 1 + 4 = 5$. Thus

$$\nabla g(1, 2, 0) = \begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix}$$

This agrees with the gradient in Exercise 1.2.

Solution 2.3 - Three inputs, two outputs

Use the computation graph defined by

$$a = xy$$

$$b = x^2$$

$$c = \sin y$$

$$d = z^2$$

$$h_1 = a + z$$

$$h_2 = b + c - d$$

At $(x, y, z) = (1, 0, 1)$, the forward pass gives

$$a = 0 \quad b = 1 \quad c = 0 \quad d = 1 \quad h_1 = 1 \quad h_2 = 0$$

Each reverse sweep starts with **all adjoints reset to zero**. Then assign the two output adjoints from the chosen seed.

First output seed

For $\mathbf{w}_1 = (1, 0)^T$, set $\bar{h}_1 = 1$ and $\bar{h}_2 = 0$. The operation $h_1 = a + z$ gives

$$\bar{a} += \bar{h}_1 = 1 \quad \bar{z} += \bar{h}_1 = 1$$

The second output contributes nothing. Propagating through $a = xy$ gives

$$\bar{x} += \bar{a} y = 0 \quad \bar{y} += \bar{a} x = 1$$

Hence the input adjoints are

$$J_{\mathbf{h}}^T \mathbf{w}_1 = \begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

This is the first Jacobian row, written as a column vector.

Second output seed

Reset all adjoints, then use $\mathbf{w}_2 = (0, 1)^T$, so $\bar{h}_1 = 0$ and $\bar{h}_2 = 1$. From $h_2 = b + c - d$ we obtain

$$\bar{b} = 1 \quad \bar{c} = 1 \quad \bar{d} = -1$$

The first output contributes nothing. Propagate through the three elementary operations:

$$\begin{aligned} \bar{x} += \bar{b} 2x &= 2 \\ \bar{y} += \bar{c} \cos y &= 1 \\ \bar{z} += \bar{d} 2z &= -2 \end{aligned}$$

Thus

$$J_{\mathbf{h}}^T \mathbf{w}_2 = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}$$

Putting the two rows together gives the complete Jacobian

$$J_{\mathbf{h}}(1, 0, 1) = \begin{pmatrix} 0 & 1 & 1 \\ 2 & 1 & -2 \end{pmatrix}$$

General output seed

For $\mathbf{w} = (3, -1)^T$, combine the two previously computed results:

$$J_{\mathbf{h}}^T \mathbf{w} = 3 \nabla h_1 - \nabla h_2 = 3 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \\ 5 \end{pmatrix}$$

Solutions - Short conceptual questions

1. With $\varepsilon^2 = 0$, all terms of second and higher order in ε vanish. The first-order term remains:

$$f(x + \dot{x}\varepsilon) = f(x) + f'(x)\dot{x}\varepsilon$$

The coefficient of ε therefore follows the derivative rules, including the chain rule for compositions.

2. For $f : \mathbb{R}^{100} \rightarrow \mathbb{R}$, forward mode needs **100 basis-direction sweeps**. Reverse mode needs **one reverse sweep** after a forward pass, because the output is scalar.
3. For $f : \mathbb{R} \rightarrow \mathbb{R}^{100}$, **one forward sweep** gives the full single Jacobian column, containing all 100 derivatives with respect to the one input. Forward mode is the natural choice.
4. The adjoint of an intermediate variable v is

$$\bar{v} = \frac{\partial L}{\partial v} \tag{2}$$

It measures the sensitivity of the chosen scalar output L to that variable.

5. A variable can influence the output through several paths. Each path contributes to the total derivative, so reverse mode must **accumulate all contributions** to its adjoint.

Forward mode versus reverse mode

For $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $J \in \mathbb{R}^{m \times n}$, the two modes efficiently compute different Jacobian products:

Mode	Seed	One sweep	Full Jacobian
Forward	Input direction $\mathbf{v} \in \mathbb{R}^n$	$J\mathbf{v}$	n basis sweeps
Reverse	Output direction $\mathbf{w} \in \mathbb{R}^m$	$J^T\mathbf{w}$	m basis sweeps

- **Few inputs, many outputs:** forward mode is often attractive.
- **Many inputs, few outputs:** reverse mode is often attractive.
- **Many parameters, one scalar loss:** reverse mode is especially attractive, as in neural-network training.

A reverse sweep requires a preceding forward pass to obtain the intermediate values. The counts above refer to one seed direction per sweep.