

ECE49595NL Lecture 27: Automatic Differentiation—II

Jeffrey Mark Siskind

Elmore Family School of Electrical and Computer Engineering

Spring 2024



PURDUE
UNIVERSITY®

Elmore Family School of Electrical
and Computer Engineering

© 2024 Jeffrey Mark Siskind. All rights reserved.

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D}f\ c$:

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D}f\ c$:

- evaluate f

The Essence of Forward-Mode AD

Taylor expansion:

$$f(\textcolor{red}{c} + \epsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\epsilon + \frac{f''(c)}{2!}\epsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\epsilon^i + \cdots$$

To compute $\mathcal{D}f\ c$:

- evaluate f at the **term** $\textcolor{red}{c} + \epsilon$

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D}f\ c$:

- evaluate f at the **term** $c + \varepsilon$ to get a **power series**,

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D}f\ c$:

- ▶ evaluate f at the **term** $c + \varepsilon$ to get a **power series**,
- ▶ extract the coefficient of ε ,

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D} f c$:

- ▶ evaluate f at the **term** $c + \varepsilon$ to get a **power series**,
- ▶ extract the coefficient of ε ,

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D} f c$:

- ▶ evaluate f at the **term** $c + \varepsilon$ to get a **power series**,
- ▶ extract the coefficient of ε , and
- ▶ multiply by $1!$

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D} f c$:

- ▶ evaluate f at the **term** $c + \varepsilon$ to get a **power series**,
- ▶ extract the coefficient of ε , and
- ▶ multiply by $1!$ (noop).

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D} f c$:

- ▶ evaluate f at the **term** $c + \varepsilon$ to get a **power series**,
- ▶ extract the coefficient of ε , and
- ▶ multiply by $1!$ (noop).

Key idea: Only need output to be a **finite truncated** power series $a + b\varepsilon$.

The Essence of Forward-Mode AD

Taylor expansion:

$$f(\textcolor{red}{c} + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D}f\ c$:

- ▶ evaluate f at the **term** $c + \varepsilon$ to get a **power series**,
- ▶ extract the coefficient of ε , and
- ▶ multiply by $1!$ (noop).

Key idea: Only need output to be a **finite** truncated power series $a + b\varepsilon$.

The input $\textcolor{red}{c} + \varepsilon$ is also a truncated power series.

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D} f c$:

- ▶ evaluate f at the **term** $c + \varepsilon$ to get a **power series**,
- ▶ extract the coefficient of ε , and
- ▶ multiply by $1!$ (noop).

Key idea: Only need output to be a **finite** truncated power series $a + b\varepsilon$.

The input $c + \varepsilon$ is also a truncated power series.

Can do a *nonstandard interpretation* of f over **truncated power series**.

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D}f\ c$:

- ▶ evaluate f at the **term** $c + \varepsilon$ to get a **power series**,
- ▶ extract the coefficient of ε , and
- ▶ multiply by $1!$ (noop).

Key idea: Only need output to be a **finite** truncated power series $a + b\varepsilon$.

The input $c + \varepsilon$ is also a truncated power series.

Can do a *nonstandard interpretation* of f over truncated power series.

Preserves control flow: Augments **original values** with **derivatives**.

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D}f\ c$:

- ▶ evaluate f at the **term** $c + \varepsilon$ to get a **power series**,
- ▶ extract the coefficient of ε , and
- ▶ multiply by $1!$ (noop).

Key idea: Only need output to be a **finite** truncated power series $a + b\varepsilon$.

The input $c + \varepsilon$ is also a truncated power series.

Can do a *nonstandard interpretation* of f over truncated power series.

Preserves control flow: Augments original values with derivatives.

$(\mathcal{D}f)$ is $\mathcal{O}(1)$ relative to f (both space and time).

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D}f\ c$:

- ▶ evaluate f at the **term** $c + \varepsilon$ to get a **power series**,
- ▶ extract the coefficient of ε , and
- ▶ multiply by $1!$ (noop).

Key idea: Only need output to be a **finite** truncated power series $a + b\varepsilon$.

The input $c + \varepsilon$ is also a truncated power series.

Can do a *nonstandard interpretation* of f over truncated power series.

Preserves control flow: Augments original values with derivatives.

$(\mathcal{D}f)$ is $\mathcal{O}(1)$ relative to f (both space and time).

These $a + b\varepsilon$ are called *dual numbers* and can be represented as $\langle a, b \rangle$.

The Essence of Forward-Mode AD

Taylor expansion:

$$f(c + \varepsilon) = \frac{f(c)}{0!} + \frac{f'(c)}{1!}\varepsilon + \frac{f''(c)}{2!}\varepsilon^2 + \cdots + \frac{f^{(i)}(c)}{i!}\varepsilon^i + \cdots$$

To compute $\mathcal{D}f\ c$:

- ▶ evaluate f at the **term** $c + \varepsilon$ to get a **power series**,
- ▶ extract the coefficient of ε , and
- ▶ multiply by $1!$ (noop).

Key idea: Only need output to be a **finite** truncated power series $a + b\varepsilon$.

The input $c + \varepsilon$ is also a truncated power series.

Can do a *nonstandard interpretation* of f over truncated power series.

Preserves control flow: Augments original values with derivatives.

$(\mathcal{D}f)$ is $\mathcal{O}(1)$ relative to f (both space and time).

These $a + b\varepsilon$ are called *dual numbers* and can be represented as $\langle a, b \rangle$.

(Analogous to complex numbers $a + bi$ represented as $\langle a, b \rangle$.)

Complex Numbers

$$i^2 = -1$$

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

$$(a + bi)(c + di) = ac + (ad + bc)i + bdi^2 = (ac - bd) + (ad + bc)i$$

Dual Numbers

$$\varepsilon^2 = 0, \text{ but } \varepsilon \neq 0$$

$$(a + b\varepsilon) + (c + d\varepsilon) = (a + c) + (b + d)\varepsilon$$

$$(a + b\varepsilon)(c + d\varepsilon) = ac + (ad + bc)\varepsilon + bd\varepsilon^2 = ac + (ad + bc)\varepsilon$$

Arithmetic on Truncated Power Series (i.e. Dual Numbers)

$$(x_0 + x_1\varepsilon + \mathcal{O}(\varepsilon^2)) + (y_0 + y_1\varepsilon + \mathcal{O}(\varepsilon^2)) = (x_0 + y_0) + (x_1 + y_1)\varepsilon + \mathcal{O}(\varepsilon^2)$$

Arithmetic on Truncated Power Series (i.e. Dual Numbers)

$$(x_0 + x_1\varepsilon + \mathcal{O}(\varepsilon^2)) + (y_0 + y_1\varepsilon + \mathcal{O}(\varepsilon^2)) = (x_0 + y_0) + (x_1 + y_1)\varepsilon + \mathcal{O}(\varepsilon^2)$$

$$\begin{aligned}(x_0 + x_1\varepsilon + \mathcal{O}(\varepsilon^2)) \times (y_0 + y_1\varepsilon + \mathcal{O}(\varepsilon^2)) \\ = (x_0 \times y_0) + (x_0 \times y_1 + x_1 \times y_0)\varepsilon + \mathcal{O}(\varepsilon^2)\end{aligned}$$

Arithmetic on Truncated Power Series (i.e. Dual Numbers)

$$(x_0 + x_1\varepsilon + \mathcal{O}(\varepsilon^2)) + (y_0 + y_1\varepsilon + \mathcal{O}(\varepsilon^2)) = (x_0 + y_0) + (x_1 + y_1)\varepsilon + \mathcal{O}(\varepsilon^2)$$

$$\begin{aligned}(x_0 + x_1\varepsilon + \mathcal{O}(\varepsilon^2)) \times (y_0 + y_1\varepsilon + \mathcal{O}(\varepsilon^2)) \\ = (x_0 \times y_0) + (x_0 \times y_1 + x_1 \times y_0)\varepsilon + \mathcal{O}(\varepsilon^2)\end{aligned}$$

$$u(x_0 + x_1\varepsilon + \mathcal{O}(\varepsilon^2)) = (u(x_0) + (u'(x_0) \times x_1)\varepsilon + \mathcal{O}(\varepsilon^2))$$

Arithmetic on Truncated Power Series (i.e. Dual Numbers)

$$(x_0 + x_1\varepsilon + \mathcal{O}(\varepsilon^2)) + (y_0 + y_1\varepsilon + \mathcal{O}(\varepsilon^2)) = (x_0 + y_0) + (x_1 + y_1)\varepsilon + \mathcal{O}(\varepsilon^2)$$

$$\begin{aligned}(x_0 + x_1\varepsilon + \mathcal{O}(\varepsilon^2)) \times (y_0 + y_1\varepsilon + \mathcal{O}(\varepsilon^2)) \\ = (x_0 \times y_0) + (x_0 \times y_1 + x_1 \times y_0)\varepsilon + \mathcal{O}(\varepsilon^2)\end{aligned}$$

$$u(x_0 + x_1\varepsilon + \mathcal{O}(\varepsilon^2)) = (u x_0) + (x_1 \times (u' x_0))\varepsilon + \mathcal{O}(\varepsilon^2)$$

$$\begin{aligned}b((x_0 + x_1\varepsilon + \mathcal{O}(\varepsilon^2)), (y_0 + y_1\varepsilon + \mathcal{O}(\varepsilon^2))) \\ = (b(x_0, y_0)) + (x_1 \times (b^{(1,0)}(x_0, y_0)) + y_1 \times (b^{(0,1)}(x_0, y_0)))\varepsilon + \mathcal{O}(\varepsilon^2)\end{aligned}$$