

ECE59500CV Lecture 4: Automatic Differentiation—III

Jeffrey Mark Siskind

Elmore Family School of Electrical and Computer Engineering

Fall 2021



PURDUE
UNIVERSITY®

Elmore Family School of Electrical
and Computer Engineering

© 2021 Jeffrey Mark Siskind. All rights reserved.

Matrix multiplication and transposition

$$(\mathbf{X}_1 \times \mathbf{X}_2)^T = \mathbf{X}_2^T \times \mathbf{X}_1^T$$

Programs as function composition

$$f = f_1 \circ \cdots \circ f_n$$

$$\mathbf{x}_1 = f_1 \mathbf{x}_0$$

$$\vdots$$

$$\mathbf{x}_n = f_n \mathbf{x}_{n-1}$$

The chain rule applied to programs

$$\begin{aligned}\mathcal{J} f \mathbf{x}_0 &= (\mathcal{J} f_n \mathbf{x}_{n-1}) \times \cdots \times (\mathcal{J} f_1 \mathbf{x}_0) \\ (\mathcal{J} f \mathbf{x}_0)^\top &= (\mathcal{J} f_1 \mathbf{x}_0)^\top \times \cdots \times (\mathcal{J} f_n \mathbf{x}_{n-1})^\top\end{aligned}$$

Computing the Jacobian

$$\overline{\mathbf{X}}_1' = (\mathcal{J} f_1 \mathbf{x}_0)$$

$$\overline{\mathbf{X}}_2' = (\mathcal{J} f_2 \mathbf{x}_1) \times \overline{\mathbf{X}}_1'$$

⋮

$$\overline{\mathbf{X}}_n' = (\mathcal{J} f_n \mathbf{x}_{n-1}) \times \overline{\mathbf{X}}_{n-1}'$$

Computing the transpose of the Jacobian

$$\begin{aligned}\overline{\mathbf{X}}_{n-1} &= (\mathcal{J} f_n \mathbf{x}_{n-1})^\top \\ \overline{\mathbf{X}}_{n-2} &= (\mathcal{J} f_{n-1} \mathbf{x}_{n-2})^\top \times \overline{\mathbf{X}}_{n-1} \\ &\vdots \\ \overline{\mathbf{X}}_0 &= (\mathcal{J} f_1 \mathbf{x}_0)^\top \times \overline{\mathbf{X}}_1\end{aligned}$$

Unary machine-state transition functions

$$\mathbf{x}[L_i] := u_i \mathbf{x}[R_i]$$

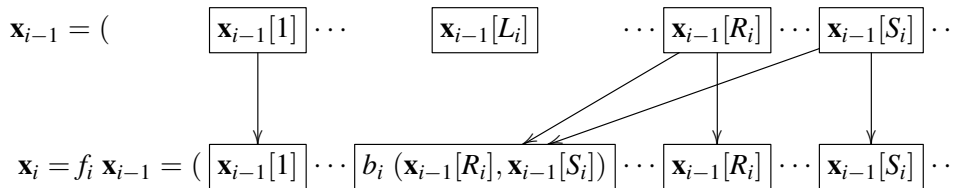
Binary machine-state transition functions

$$\mathbf{x}[L_i] := b(\mathbf{x}[R_i], \mathbf{x}[S_i])$$

What a unary machine-state transition function does

$$\mathbf{x}_{i-1} = (\boxed{\mathbf{x}_{i-1}[1]} \cdots \boxed{\mathbf{x}_{i-1}[L_i]} \cdots \boxed{\mathbf{x}_{i-1}[R_i]} \cdots \boxed{\mathbf{x}_{i-1}[m]})$$
$$\mathbf{x}_i = f_i \mathbf{x}_{i-1} = (\boxed{\mathbf{x}_{i-1}[1]} \cdots \boxed{u_i \mathbf{x}_{i-1}[R_i]} \cdots \boxed{\mathbf{x}_{i-1}[R_i]} \cdots \boxed{\mathbf{x}_{i-1}[m]})$$

What a binary machine-state transition function does



Computing a Jacobian-vector product

$$\overline{\mathbf{x}}_n = (\mathcal{J} f \mathbf{x}_0) \times \overline{\mathbf{x}}_0$$

$$\overline{\mathbf{x}}_1 = (\mathcal{J} f_1 \mathbf{x}_0) \times \overline{\mathbf{x}}_0$$

⋮

$$\overline{\mathbf{x}}_n = (\mathcal{J} f_n \mathbf{x}_{n-1}) \times \overline{\mathbf{x}}_{n-1}$$

Computing a vector-Jacobian product

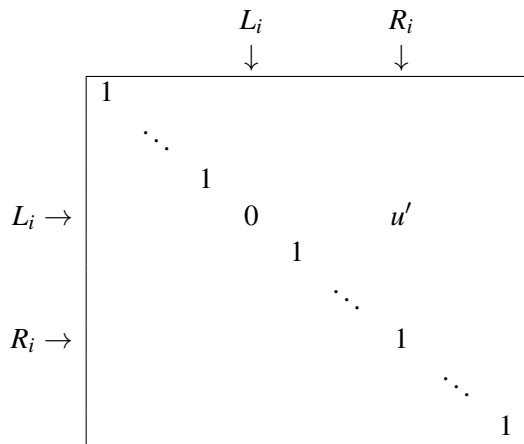
$$\overline{\mathbf{x}}_0 = (\mathcal{J} f_{\mathbf{x}_0})^\top \times \overline{\mathbf{x}}_n$$

$$\overline{\mathbf{x}}_{n-1} = (\mathcal{J} f_n \mathbf{x}_{n-1})^\top \times \overline{\mathbf{x}}_n$$

$$\vdots$$

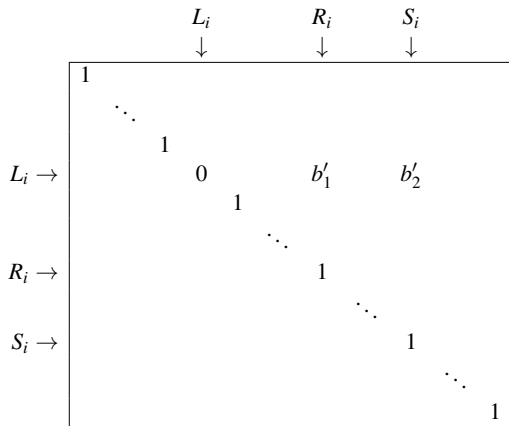
$$\overline{\mathbf{x}}_0 = (\mathcal{J} f_1 \mathbf{x}_0)^\top \times \overline{\mathbf{x}}_1$$

Jacobian of a unary primitive



$$u' = \mathcal{D} u_i \mathbf{x}_{i-1}[R_i]$$

Jacobian of a binary primitive



$$b'_1 = \mathcal{D}_1 b_i (\mathbf{x}_{i-1}[R_i], \mathbf{x}_{i-1}[S_i])$$

$$b'_2 = \mathcal{D}_2 b_i (\mathbf{x}_{i-1}[R_i], \mathbf{x}_{i-1}[S_i])$$

Unary single-step vector-Jacobian product

$$\begin{pmatrix} \bar{\mathbf{x}}_i[1] \\ \vdots \\ \bar{\mathbf{x}}_i[L_i - 1] \\ 0 \\ \bar{\mathbf{x}}_i[L_i + 1] \\ \vdots \\ u' \times \bar{\mathbf{x}}_i[L_i] + \bar{\mathbf{x}}_i[R_i] \\ \vdots \\ \bar{\mathbf{x}}_i[m] \end{pmatrix} = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 0 & & & & & \\ & & & & 1 & & & & \\ & & & & & \ddots & & & \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix} \begin{pmatrix} \bar{\mathbf{x}}_i[1] \\ \vdots \\ \bar{\mathbf{x}}_i[L_i - 1] \\ \bar{\mathbf{x}}_i[L_i] \\ \bar{\mathbf{x}}_i[L_i + 1] \\ \vdots \\ \bar{\mathbf{x}}_i[R_i] \\ \vdots \\ \bar{\mathbf{x}}_i[m] \end{pmatrix}$$

$$u' = \mathcal{D} u_i \mathbf{x}_{i-1}[R_i]$$

$$f = f_1 \circ \cdots \circ f_n$$

Forward Mode

$$f = f_1 \circ \cdots \circ f_n$$
$$\mathcal{J}(f)(x_0) = \mathcal{J}(f_n)(x_{n-1}) \times \cdots \times \mathcal{J}(f_1)(x_0)$$

Forward Mode

$$\begin{aligned}f &= f_1 \circ \cdots \circ f_n \\ \mathcal{J}(f)(x_0) &= \mathcal{J}(f_n)(x_{n-1}) \times \cdots \times \mathcal{J}(f_1)(x_0) \\ \acute{x}_n &= \mathcal{J}(f)(x_0) \times \acute{x}_0\end{aligned}$$

Forward Mode

$$\begin{aligned}f &= f_1 \circ \cdots \circ f_n \\ \mathcal{J}(f)(x_0) &= \mathcal{J}(f_n)(x_{n-1}) \times \cdots \times \mathcal{J}(f_1)(x_0) \\ \acute{x}_n &= \mathcal{J}(f)(x_0) \times \acute{x}_0\end{aligned}$$

$$\begin{aligned}x_1 &= f_1(x_0) \\ \acute{x}_1 &= \mathcal{J}(f_1)(x_0) \times \acute{x}_0 \\ &\vdots \\ x_n &= f_n(x_{n-1}) \\ \acute{x}_n &= \mathcal{J}(f_n)(x_{n-1}) \times \acute{x}_{n-1}\end{aligned}$$

Reverse Mode

$$f = f_1 \circ \cdots \circ f_n$$

Reverse Mode

$$f = f_1 \circ \cdots \circ f_n$$
$$\mathcal{J}(f)(x_0)^\top = \mathcal{J}(f_1)(x_0)^\top \times \cdots \times \mathcal{J}(f_n)(x_{n-1})^\top$$

Reverse Mode

$$\begin{aligned}f &= f_1 \circ \cdots \circ f_n \\ \mathcal{J}(f)(x_0)^\top &= \mathcal{J}(f_1)(x_0)^\top \times \cdots \times \mathcal{J}(f_n)(x_{n-1})^\top \\ \dot{x}_0 &= \mathcal{J}(f)(x_0)^\top \times \dot{x}_n\end{aligned}$$

Reverse Mode

$$\begin{aligned}f &= f_1 \circ \cdots \circ f_n \\ \mathcal{J}(f)(x_0)^\top &= \mathcal{J}(f_1)(x_0)^\top \times \cdots \times \mathcal{J}(f_n)(x_{n-1})^\top \\ \dot{x}_0 &= \mathcal{J}(f)(x_0)^\top \times \dot{x}_n\end{aligned}$$

$$x_1 = f_1(x_0)$$

$$\vdots$$

$$x_n = f_n(x_{n-1})$$

$$\dot{x}_{n-1} = \mathcal{J}(f_n)(x_{n-1}) \times \dot{x}_n$$

$$\vdots$$

$$\dot{x}_0 = \mathcal{J}(f_1)(x_0) \times \dot{x}_1$$

Forward Mode by Overloading

$$x_1 = f_1(x_0)$$

$$\dot{x}_1 = \mathcal{J}(f_1)(x_0) \times \dot{x}_0$$

$$\vdots$$

$$x_n = f_n(x_{n-1})$$

$$\dot{x}_n = \mathcal{J}(f_n)(x_{n-1}) \times \dot{x}_{n-1}$$

Forward Mode by Overloading

$$x_1 = f_1(x_0)$$

$$\acute{x}_1 = \mathcal{J}(f_1)(x_0) \times \acute{x}_0$$

$$\vdots$$

$$x_n = f_n(x_{n-1})$$

$$\acute{x}_n = \mathcal{J}(f_n)(x_{n-1}) \times \acute{x}_{n-1}$$

$$x_i = f_i(x_{i-1})$$

Forward Mode by Overloading

$$x_1 = f_1(x_0)$$

$$\acute{x}_1 = \mathcal{J}(f_1)(x_0) \times \acute{x}_0$$

$$\vdots$$

$$x_n = f_n(x_{n-1})$$

$$\acute{x}_n = \mathcal{J}(f_n)(x_{n-1}) \times \acute{x}_{n-1}$$

$$x_i = f_i(x_{i-1})$$

$$\langle x_i, \acute{x}_i \rangle = \langle f_i(x_{i-1}), \mathcal{J}(f_i)(x_{i-1}) \times \acute{x}_{i-1} \rangle$$

Forward Mode by Overloading

$$x_1 = f_1(x_0)$$

$$\acute{x}_1 = \mathcal{J}(f_1)(x_0) \times \acute{x}_0$$

$$\vdots$$

$$x_n = f_n(x_{n-1})$$

$$\acute{x}_n = \mathcal{J}(f_n)(x_{n-1}) \times \acute{x}_{n-1}$$

$$x_i = f_i(x_{i-1})$$

$$\langle x_i, \acute{x}_i \rangle = \langle f_i(x_{i-1}), \mathcal{J}(f_i)(x_{i-1}) \times \acute{x}_{i-1} \rangle$$

$$\overrightarrow{x_i} = \overrightarrow{f_i}(\overrightarrow{x_{i-1}})$$

Reverse Mode

$$x_1 = f_1(x_0)$$

$$\vdots$$

$$x_n = f_n(x_{n-1})$$

$$\dot{x}_{n-1} = \mathcal{J}(f_n)(x_{n-1}) \times \dot{x}_n$$

$$\vdots$$

$$\dot{x}_0 = \mathcal{J}(f_1)(x_0) \times \dot{x}_1$$