ECE59500CV Lecture 6: Automatic Differentiation—III

Jeffrey Mark Siskind

School of Electrical and Computer Engineering

Fall 2020

PURDUE
UNIVERSITY

Matrix multiplication and transposition

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

Programs as function composition

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

$$\begin{aligned} \mathbf{x}_1 &= f_1\ \mathbf{x}_0 \\ &\vdots \\ \mathbf{x}_n &= f_n\ \mathbf{x}_{n-1} \end{aligned}$$

The chain rule applied to programs

$$\mathcal{J}\;f\;\mathbf{x}_0=(\mathcal{J}\;f_n\;\mathbf{x}_{n-1})\times\cdots\times(\mathcal{J}\;f_1\;\mathbf{x}_0)$$

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

Computing the Jacobian

$$
\begin{aligned}
\overset{\rightharpoondown}{\mathbf{X}_1} &= (\mathcal{J}\, f_1\ \mathbf{x}_0) \\
\overset{\rightharpoondown}{\mathbf{X}_2} &= (\mathcal{J}\, f_2\ \mathbf{x}_1) \times \overset{\rightharpoondown}{\mathbf{X}_1} \\
&\ \vdots \\
\overset{\rightharpoondown}{\mathbf{X}_n} &= (\mathcal{J}\, f_n\ \mathbf{x}_{n-1}) \times \overset{\rightharpoondown}{\mathbf{X}_{n-1}}
\end{aligned}
$$

Computing the transpose of the Jacobian

$$\begin{aligned}
\overleftharpoon{\mathbf{X}_{n-1}} &= (\mathcal{J}\, f_n\ \mathbf{x}_{n-1})^{\top} \\
\overleftharpoon{\mathbf{X}_{n-2}} &= (\mathcal{J}\, f_{n-1}\ \mathbf{x}_{n-2})^{\top} \times \overleftharpoon{\mathbf{X}_{n-1}} \\
&\vdots \\
\overleftharpoon{\mathbf{X}_{0}} &= (\mathcal{J}\, f_1\ \mathbf{x}_{0})^{\top} \times \overleftharpoon{\mathbf{X}_{1}}
\end{aligned}$$

Unary machine-state transition functions

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

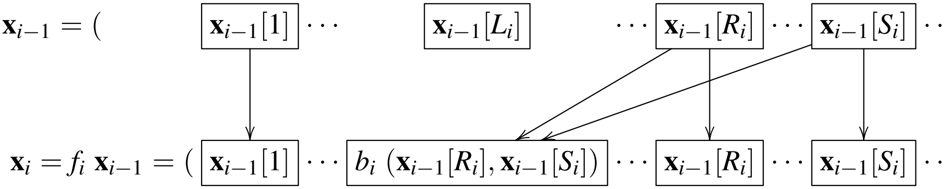
Binary machine-state transition functions

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

What a unary machine-state transition function does

$$\begin{tikzcd}
\mathbf{x}_{i-1} = (& \boxed{\mathbf{x}_{i-1}[1]} \arrow[d] & \cdots & \boxed{\mathbf{x}_{i-1}[L_i]} & \cdots & \boxed{\mathbf{x}_{i-1}[R_i]} \arrow[dll] \arrow[d] & \cdots & \boxed{\mathbf{x}_{i-1}[m]} \arrow[d] &) \\
\mathbf{x}_i = f_i\ \mathbf{x}_{i-1} = (& \boxed{\mathbf{x}_{i-1}[1]} & \cdots & \boxed{u_i\ \mathbf{x}_{i-1}[R_i]} & \cdots & \boxed{\mathbf{x}_{i-1}[R_i]} & \cdots & \boxed{\mathbf{x}_{i-1}[m]} &)
\end{tikzcd}$$

What a binary machine-state transition function does



$$\mathbf{x}_{i-1} = (\boxed{\mathbf{x}_{i-1}[1]} \cdots \boxed{\mathbf{x}_{i-1}[L_i]} \cdots \boxed{\mathbf{x}_{i-1}[R_i]} \cdots \boxed{\mathbf{x}_{i-1}[S_i]} \cdot\cdot$$

$$\mathbf{x}_i = f_i\ \mathbf{x}_{i-1} = (\boxed{\mathbf{x}_{i-1}[1]} \cdots \boxed{b_i\left(\mathbf{x}_{i-1}[R_i], \mathbf{x}_{i-1}[S_i]\right)} \cdots \boxed{\mathbf{x}_{i-1}[R_i]} \cdots \boxed{\mathbf{x}_{i-1}[S_i]} \cdot\cdot$$

Computing a Jacobian-vector product

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

$$\begin{aligned}
\overrightharpoon{\mathbf{x}_1} &= (\mathcal{J}\ f_1\ \mathbf{x}_0) \times \overrightharpoon{\mathbf{x}_0} \\
&\vdots \\
\overrightharpoon{\mathbf{x}_n} &= (\mathcal{J}\ f_n\ \mathbf{x}_{n-1}) \times \overrightharpoon{\mathbf{x}_{n-1}}
\end{aligned}$$

Computing a vector-Jacobian product

$$\overleftharpoondown{\mathbf{x}_0} = (\mathcal{J}\, f\ \mathbf{x}_0)^\top \times \overleftharpoondown{\mathbf{x}_n}$$

$$\begin{aligned}
\overleftharpoondown{\mathbf{x}_{n-1}} &= (\mathcal{J}\, f_n\ \mathbf{x}_{n-1})^\top \times \overleftharpoondown{\mathbf{x}_n}\\
&\vdots\\
\overleftharpoondown{\mathbf{x}_0} &= (\mathcal{J}\, f_1\ \mathbf{x}_0)^\top \times \overleftharpoondown{\mathbf{x}_1}
\end{aligned}$$

Jacobian of a unary primitive

$$\begin{array}{r|ccccccccc|}
 & & & & \overset{L_i}{\downarrow} & & & \overset{R_i}{\downarrow} & & \\ \hline
 & 1 & & & & & & & & \\
 & & \ddots & & & & & & & \\
 & & & 1 & & & & & & \\
L_i \rightarrow & & & & 0 & & & u' & & \\
 & & & & & 1 & & & & \\
 & & & & & & \ddots & & & \\
R_i \rightarrow & & & & & & & 1 & & \\
 & & & & & & & & \ddots & \\
 & & & & & & & & & 1 \\ \hline
\end{array}$$

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

Jacobian of a binary primitive

$$
\begin{array}{c|ccccccccccc|}
 & & & & L_i & & & R_i & & S_i & & \\
 & & & & \downarrow & & & \downarrow & & \downarrow & & \\
\hline
 & 1 & & & & & & & & & & \\
 & & \ddots & & & & & & & & & \\
 & & & 1 & & & & & & & & \\
L_i \rightarrow & & & & 0 & & & b'_1 & & b'_2 & & \\
 & & & & & 1 & & & & & & \\
 & & & & & & \ddots & & & & & \\
R_i \rightarrow & & & & & & & 1 & & & & \\
 & & & & & & & & \ddots & & & \\
S_i \rightarrow & & & & & & & & & 1 & & \\
 & & & & & & & & & & \ddots & \\
 & & & & & & & & & & & 1 \\
\hline
\end{array}
$$

$$b'_1 = \mathcal{D}_1\ b_i\ (\mathbf{x}_{i-1}[R_i], \mathbf{x}_{i-1}[S_i])$$
$$b'_2 = \mathcal{D}_2\ b_i\ (\mathbf{x}_{i-1}[R_i], \mathbf{x}_{i-1}[S_i])$$

Unary single-step Jacobian-vector product

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

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

Binary single-step Jacobian-vector product

$$\begin{pmatrix}
\overrightarrow{\mathbf{x}_{i-1}}[1]\\
\vdots\\
\overrightarrow{\mathbf{x}_{i-1}}[L_i-1]\\
b_1'\times\overrightarrow{\mathbf{x}_{i-1}}[R_i]+b_2'\times\overrightarrow{\mathbf{x}_{i-1}}[S_i]\\
\overrightarrow{\mathbf{x}_{i-1}}[L_i+1]\\
\vdots\\
\overrightarrow{\mathbf{x}_{i-1}}[R_i]\\
\vdots\\
\overrightarrow{\mathbf{x}_{i-1}}[S_i]\\
\vdots\\
\overrightarrow{\mathbf{x}_{i-1}}[m]
\end{pmatrix}
=
\begin{pmatrix}
1 &&&&&&&&&&\\
&\ddots&&&&&&&&&\\
&&1&&&&&&&&\\
&&&0&&&b_1'&&b_2'&&\\
&&&&1&&&&&&\\
&&&&&\ddots&&&&&\\
&&&&&&1&&&&\\
&&&&&&&\ddots&&&\\
&&&&&&&&1&&\\
&&&&&&&&&\ddots&\\
&&&&&&&&&&1
\end{pmatrix}
\begin{pmatrix}
\overrightarrow{\mathbf{x}_{i-1}}[1]\\
\vdots\\
\overrightarrow{\mathbf{x}_{i-1}}[L_i-1]\\
\overrightarrow{\mathbf{x}_{i-1}}[L_i]\\
\overrightarrow{\mathbf{x}_{i-1}}[L_i+1]\\
\vdots\\
\overrightarrow{\mathbf{x}_{i-1}}[R_i]\\
\vdots\\
\overrightarrow{\mathbf{x}_{i-1}}[S_i]\\
\vdots\\
\overrightarrow{\mathbf{x}_{i-1}}[m]
\end{pmatrix}$$

$$\begin{aligned}
b_1' &= \mathcal{D}_1\ b_i\ (\mathbf{x}_{i-1}[R_i],\mathbf{x}_{i-1}[S_i])\\
b_2' &= \mathcal{D}_2\ b_i\ (\mathbf{x}_{i-1}[R_i],\mathbf{x}_{i-1}[S_i])
\end{aligned}$$

Unary single-step vector-Jacobian product

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

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

Binary single-step vector-Jacobian product

$$\begin{pmatrix}\overleftharpoon{\mathbf{x}}_i[1]\\ \vdots\\ \overleftharpoon{\mathbf{x}}_i[L_i-1]\\ 0\\ \overleftharpoon{\mathbf{x}}_i[L_i+1]\\ \vdots\\ b'_1\times\overleftharpoon{\mathbf{x}}_i[L_i]+\overleftharpoon{\mathbf{x}}_i[R_i]\\ \vdots\\ b'_2\times\overleftharpoon{\mathbf{x}}_i[L_i]+\overleftharpoon{\mathbf{x}}_i[S_i]\\ \vdots\\ \overleftharpoon{\mathbf{x}}_i[m]\end{pmatrix}=\begin{pmatrix}1&&&&&&&&&&\\ &\ddots&&&&&&&&&\\ &&1&&&&&&&&\\ &&&0&&&&&&&\\ &&&&1&&&&&&\\ &&&&&\ddots&&&&&\\ &&&b'_1&&&1&&&&\\ &&&&&&&\ddots&&&\\ &&&b'_2&&&&&1&&\\ &&&&&&&&&\ddots&\\ &&&&&&&&&&1\end{pmatrix}\begin{pmatrix}\overleftharpoon{\mathbf{x}}_i[1]\\ \vdots\\ \overleftharpoon{\mathbf{x}}_i[L_i-1]\\ \overleftharpoon{\mathbf{x}}_i[L_i]\\ \overleftharpoon{\mathbf{x}}_i[L_i+1]\\ \vdots\\ \overleftharpoon{\mathbf{x}}_i[R_i]\\ \vdots\\ \overleftharpoon{\mathbf{x}}_i[S_i]\\ \vdots\\ \overleftharpoon{\mathbf{x}}_i[m]\end{pmatrix}$$

$$b'_1=\mathcal{D}_1\;b_i\;(\mathbf{x}_{i-1}[R_i],\mathbf{x}_{i-1}[S_i])$$
$$b'_2=\mathcal{D}_2\;b_i\;(\mathbf{x}_{i-1}[R_i],\mathbf{x}_{i-1}[S_i])$$

Forward Mode

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

Forward Mode

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

Forward Mode

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

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

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

Reverse Mode

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

Reverse Mode

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

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

Forward Mode by Overloading

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

Forward Mode by Overloading

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

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

Forward Mode by Overloading

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

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

Forward Mode by Overloading

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

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

Reverse Mode

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