Algorithmic Differentiation

Manuel Alberto Parra Díaz
MSc. Mathematics

Faculty of Economics
Universidad del Rosario

August 25, 2022

Presentation Outline

Matrix Formulation

Definition 1.1

If A is a matrix, the *adjoint matrix of A* is defined by the conjugated transpose of A.

If A is a real-valued matrix, the *adjoint matrix of A* is defined by the transpose of A.

Let $Z = (X_1, X_2, ..., X_N)$ be a vector of input parameters X_j, F be the main function (which is elementary) of a model and Y be the output of the model.

Let $Z = (X_1, X_2, ..., X_N)$ be a vector of input parameters X_j, F be the main function (which is elementary) of a model and Y be the output of the model.

Then, by chain-rule

$$\dot{F} = \frac{\partial F}{\partial Z} = \frac{\partial F}{\partial X_1}\dot{X}_1 + \cdots + \frac{\partial F}{\partial X_N}\dot{X}_N$$

Let $Z = (X_1, X_2, ..., X_N)$ be a vector of input parameters X_j, F be the main function (which is elementary) of a model and Y be the output of the model.

Then, by chain-rule

$$\dot{F} = \frac{\partial F}{\partial Z} = \frac{\partial F}{\partial X_1}\dot{X}_1 + \cdots + \frac{\partial F}{\partial X_N}\dot{X}_N$$

So,

$$\dot{F} = \nabla F \cdot (\dot{X}_1, ..., \dot{X}_N) = \nabla F \cdot \dot{Z}$$

Thus, we can write

$$\begin{bmatrix} \dot{X}_1 \\ \vdots \\ \dot{X}_N \\ \dot{F} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \frac{\partial F}{\partial X_1} & \frac{\partial F}{\partial X_2} & \cdots & \frac{\partial F}{\partial X_N} \end{bmatrix} \begin{bmatrix} \dot{X}_1 \\ \vdots \\ \dot{X}_N \end{bmatrix} \tag{1}$$

Thus, we can write

$$\begin{bmatrix} \dot{X}_1 \\ \vdots \\ \dot{X}_N \\ \dot{F} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \frac{\partial F}{\partial X_1} & \frac{\partial F}{\partial X_2} & \cdots & \frac{\partial F}{\partial X_N} \end{bmatrix} \begin{bmatrix} \dot{X}_1 \\ \vdots \\ \dot{X}_N \end{bmatrix} \tag{1}$$

Since this equation is the consequence of the chain-rule over the function F, it is the matrix representation of the forward mode.

Thus, we can write

$$\begin{bmatrix} \dot{X}_1 \\ \vdots \\ \dot{X}_N \\ \dot{F} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \dfrac{\partial F}{\partial X_1} & \dfrac{\partial F}{\partial X_2} & \cdots & \dfrac{\partial F}{\partial X_N} \end{bmatrix} \begin{bmatrix} \dot{X}_1 \\ \vdots \\ \dot{X}_N \end{bmatrix} \tag{1}$$

Since this equation is the consequence of the chain-rule over the function F, it is the matrix representation of the forward mode.

Moreover, if we want to compute the derivative with respect to X_j, the $i-th$ component of the vector $\dot{Z}$ will be $\delta_{i,j}$ in the equation (1).

Example:

If $F(a, b, c) = abc$, then

$$\begin{bmatrix} \dot{a} \\ \dot{b} \\ \dot{c} \\ \dot{F} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ bc & ac & ab \end{bmatrix} \begin{bmatrix} \dot{a} \\ \dot{b} \\ \dot{c} \end{bmatrix} \Longrightarrow \dot{F} = bc\dot{a} + ac\dot{b} + ab\dot{c}$$

Example:

If $F(a,b,c) = abc$, then

$$\begin{bmatrix} \dot{a} \\ \dot{b} \\ \dot{c} \\ \dot{F} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ bc & ac & ab \end{bmatrix} \begin{bmatrix} \dot{a} \\ \dot{b} \\ \dot{c} \end{bmatrix} \implies \dot{F} = bc\dot{a} + ac\dot{b} + ab\dot{c}$$

And, we use $\dot{Z} = (0,1,0)$ to compute $\dfrac{\partial F}{\partial b}$:

$$\begin{bmatrix} \dot{a} \\ \dot{b} \\ \dot{c} \\ \dot{F} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ bc & ac & ab \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \implies \dot{F} = ac = \frac{\partial F}{\partial b}.$$

For the adjoint mode, we defined the vector of *adjoints* by

$$\begin{bmatrix} \overline{X}_1 \\ \vdots \\ \overline{X}_N \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & \dfrac{\partial F}{\partial X_1} \\ 0 & 1 & \cdots & 0 & \dfrac{\partial F}{\partial X_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \dfrac{\partial F}{\partial X_N} \end{bmatrix} \begin{bmatrix} \overline{X}_1 \\ \vdots \\ \overline{X}_N \\ \overline{F} \end{bmatrix} \tag{2}$$

Note that, in equation (2), the first matrix in the right side is the adjoint matrix of the first matrix in the right side of the equation (1).

Therefore we get the following system

$$\overline{X}_1 = \overline{X}_1 + \frac{\partial F}{\partial X_1}\overline{F}$$

$$\overline{X}_2 = \overline{X}_2 + \frac{\partial F}{\partial X_2}\overline{F}$$

$$\vdots$$

$$\overline{X}_N = \overline{X}_N + \frac{\partial F}{\partial X_N}\overline{F}$$

Therefore we get the following system

$$\overline{X}_1 = \overline{X}_1 + \frac{\partial F}{\partial X_1}\overline{F}$$
$$\overline{X}_2 = \overline{X}_2 + \frac{\partial F}{\partial X_2}\overline{F}$$
$$\vdots$$
$$\overline{X}_N = \overline{X}_N + \frac{\partial F}{\partial X_N}\overline{F}$$

Now, we can obtain $\dot{F}$ with a single run of these equations initializing $\overline{F} = 1$ and $\overline{X}_1 = \overline{X}_2 = ... = \overline{X}_N = 0$.

Example:

If $F(a,b,c) = abc$ the system of equations is

$$\bar{a} = \bar{a} + \frac{\partial F}{\partial a}\overline{F}$$
$$\bar{b} = \bar{b} + \frac{\partial F}{\partial b}\overline{F}$$
$$\bar{c} = \bar{c} + \frac{\partial F}{\partial c}\overline{F}$$

Example:

If $F(a,b,c) = abc$ the system of equations is

$$\bar{a} = \bar{a} + \frac{\partial F}{\partial a}\bar{F}$$
$$\bar{b} = \bar{b} + \frac{\partial F}{\partial b}\bar{F}$$
$$\bar{c} = \bar{c} + \frac{\partial F}{\partial c}\bar{F}$$

And, if $\bar{F} = 1$ and $\bar{a} = \bar{b} = \bar{c} = 0$

$$\bar{a} = \frac{\partial F}{\partial a}, \bar{b} = \frac{\partial F}{\partial b}, \bar{c} = \frac{\partial F}{\partial c}$$

What about non-elementary functions?

What about non-elementary functions?

Let $Z = (X_1, X_2, ..., X_N)$ be a vector of input parameters X_j and Y be the output of the model.

If

$$F(Z) = F_p \circ F_{p-1} \circ ... \circ F_2 \circ F_1(Z)$$

we generalize the matrix representation of the algorithmic differentiation modes with matrix products

What about non-elementary functions?

Let $Z = (X_1, X_2, ..., X_N)$ be a vector of input parameters X_j and Y be the output of the model.

If

$$F(Z) = F_p \circ F_{p-1} \circ ... \circ F_2 \circ F_1(Z)$$

we generalize the matrix representation of the algorithmic differentiation modes with matrix products

Forward Mode

$$M = M_1 M_2 \cdots M_{p-1} M_p$$

What about non-elementary functions?

Let $Z = (X_1, X_2, ..., X_N)$ be a vector of input parameters X_j and Y be the output of the model.

If

$$F(Z) = F_p \circ F_{p-1} \circ ... \circ F_2 \circ F_1(Z)$$

we generalize the matrix representation of the algorithmic differentiation modes with matrix products

Forward Mode

$$M = M_1 M_2 \cdots M_{p-1} M_p$$

Adjoint Mode

$$M^T = M_p^T M_{p-1}^T \cdots M_2^T M_1^T$$

Example:

If $F(Z) = F_2(F_1(Z))$, then

$$Z \xrightarrow{F_1} A \xrightarrow{F_2} B \longrightarrow Y$$

Example:

If $F(Z) = F_2(F_1(Z))$, then

$$Z \xrightarrow{F_1} A \xrightarrow{F_2} B \longrightarrow Y$$

$$\dot{A} = \frac{\partial A}{\partial Z}\dot{Z} \qquad \dot{B} = \frac{\partial B}{\partial A}\dot{A} \qquad \dot{Y} = \frac{\partial Y}{\partial B}\dot{B}$$

Example:

If $F(Z) = F_2(F_1(Z))$, then

$$Z \xrightarrow{F_1} A \xrightarrow{F_2} B \longrightarrow Y$$

$$\dot{A} = \frac{\partial A}{\partial Z}\dot{Z} \qquad \dot{B} = \frac{\partial B}{\partial A}\dot{A} \qquad \dot{Y} = \frac{\partial Y}{\partial B}\dot{B}$$

$$\dot{Y} = \underbrace{\frac{\partial Y}{\partial B}\frac{\partial B}{\partial A}\frac{\partial A}{\partial Z}}_{M}\dot{Z}$$

Example:

If $F(Z) = F_2(F_1(Z))$, then

$$Z \xrightarrow{F_1} A \xrightarrow{F_2} B \longrightarrow Y$$

$$\dot{A} = \frac{\partial A}{\partial Z}\dot{Z} \qquad \dot{B} = \frac{\partial B}{\partial A}\dot{A} \qquad \dot{Y} = \frac{\partial Y}{\partial B}\dot{B}$$

$$\dot{Y} = \underbrace{\frac{\partial Y}{\partial B}\frac{\partial B}{\partial A}\frac{\partial A}{\partial Z}}_{M}\dot{Z}$$

$$\dot{Z} \longrightarrow \dot{A} \longrightarrow \dot{B} \longrightarrow \dot{Y}$$

If $\overline{Y} = 1$,

$$\overline{Z} = \left(\frac{\partial Y}{\partial Z}\right)^T = \left(\frac{\partial Y}{\partial A}\frac{\partial A}{\partial Z}\right)^T = \left(\frac{\partial A}{\partial Z}\right)^T \overline{A}$$

$$\overline{A} = \left(\frac{\partial Y}{\partial A}\right)^T = \left(\frac{\partial Y}{\partial B}\frac{\partial B}{\partial A}\right)^T = \left(\frac{\partial B}{\partial A}\right)^T \overline{B}$$

$$\overline{B} = \left(\frac{\partial Y}{\partial B}\right)^T = \left(\frac{\partial Y}{\partial B}\right)^T \cdot 1 = \left(\frac{\partial Y}{\partial B}\right)^T \overline{Y}$$

If $\overline{Y} = 1$,

$$\overline{Z} = \left(\frac{\partial Y}{\partial Z}\right)^T = \left(\frac{\partial Y}{\partial A}\frac{\partial A}{\partial Z}\right)^T = \left(\frac{\partial A}{\partial Z}\right)^T \overline{A}$$

$$\overline{A} = \left(\frac{\partial Y}{\partial A}\right)^T = \left(\frac{\partial Y}{\partial B}\frac{\partial B}{\partial A}\right)^T = \left(\frac{\partial B}{\partial A}\right)^T \overline{B}$$

$$\overline{B} = \left(\frac{\partial Y}{\partial B}\right)^T = \left(\frac{\partial Y}{\partial B}\right)^T \cdot 1 = \left(\frac{\partial Y}{\partial B}\right)^T \overline{Y}$$

$$\overline{Z} = \underbrace{\left(\frac{\partial A}{\partial Z}\right)^T \left(\frac{\partial B}{\partial A}\right)^T \left(\frac{\partial Y}{\partial B}\right)^T}_{M^T} \overline{Y}$$

If $\overline{Y} = 1$,

$$\overline{Z} = \left(\frac{\partial Y}{\partial Z}\right)^T = \left(\frac{\partial Y}{\partial A}\frac{\partial A}{\partial Z}\right)^T = \left(\frac{\partial A}{\partial Z}\right)^T \overline{A}$$

$$\overline{A} = \left(\frac{\partial Y}{\partial A}\right)^T = \left(\frac{\partial Y}{\partial B}\frac{\partial B}{\partial A}\right)^T = \left(\frac{\partial B}{\partial A}\right)^T \overline{B}$$

$$\overline{B} = \left(\frac{\partial Y}{\partial B}\right)^T = \left(\frac{\partial Y}{\partial B}\right)^T \cdot 1 = \left(\frac{\partial Y}{\partial B}\right)^T \overline{Y}$$

$$\overline{Z} = \underbrace{\left(\frac{\partial A}{\partial Z}\right)^T \left(\frac{\partial B}{\partial A}\right)^T \left(\frac{\partial Y}{\partial B}\right)^T}_{M^T} \overline{Y}$$

$$\overline{Y} \longrightarrow \overline{B} \longrightarrow \overline{A} \longrightarrow \overline{Z}$$

Algorithms

Main Assumption:

- The programs use a finite set of elementary functions.
- All other values are compositions of the elementary functions.
- The derivatives of the elementary functions are known or the derivatives can be computed by chain-rule.

Main Assumption:

- The programs use a finite set of elementary functions.
- All other values are compositions of the elementary functions.
- The derivatives of the elementary functions are known or the derivatives can be computed by chain-rule.
- The program that represents the algorithm to compute F is a simple linear program (SAC).

We start with the computation of a function

$$F : \mathbb{R}^{p_z} \longrightarrow \mathbb{R}$$
$$z \longrightarrow y = f(z)$$

where $z = (z_0, z_1, ..., z_{p_z-1}) = z[0 : p_z - 1]$ are the inputs of the function.

$a[i : j]$ represents the vector of dimension $j - i + 1$ and components a_k for $i \leq k \leq j$.

- The algorithm starts with the inputs z and goes to y through a lot of intermediary variables.

- The algorithm starts with the inputs z and goes to y through a lot of intermediary variables.
- All the intermediary values used in the program will be a component of a vector b.

- The algorithm starts with the inputs z and goes to y through a lot of intermediary variables.
- All the intermediary values used in the program will be a component of a vector b.
- There are p_b intermediary variables in the program.

- The algorithm starts with the inputs z and goes to y through a lot of intermediary variables.
- All the intermediary values used in the program will be a component of a vector b.
- There are p_b intermediary variables in the program.
- $b = b[1 : p_b]$ and an intermediary variable is denoted by $b[j]$ for $1 \leq j \leq p_b$.

- The algorithm starts with the inputs z and goes to y through a lot of intermediary variables.
- All the intermediary values used in the program will be a component of a vector b.
- There are p_b intermediary variables in the program.
- $b = b[1 : p_b]$ and an intermediary variable is denoted by $b[j]$ for $1 \leq j \leq p_b$.
- At each line of code, the variable $b[j]$ is defined depending potentially on all the previous intermediary variables and inputs.

- The algorithm starts with the inputs z and goes to y through a lot of intermediary variables.
- All the intermediary values used in the program will be a component of a vector b.
- There are p_b intermediary variables in the program.
- $b = b[1 : p_b]$ and an intermediary variable is denoted by $b[j]$ for $1 \leq j \leq p_b$.
- At each line of code, the variable $b[j]$ is defined depending potentially on all the previous intermediary variables and inputs.
- We initialize $b[j]$ for $-p_z + 1 \leq j \leq 0$ with the input variables.

- The algorithm starts with the inputs z and goes to y through a lot of intermediary variables.
- All the intermediary values used in the program will be a component of a vector b.
- There are p_b intermediary variables in the program.
- $b = b[1 : p_b]$ and an intermediary variable is denoted by $b[j]$ for $1 \leq j \leq p_b$.
- At each line of code, the variable $b[j]$ is defined depending potentially on all the previous intermediary variables and inputs.
- We initialize $b[j]$ for $-p_z + 1 \leq j \leq 0$ with the input variables.
- Each line of code perform the computation of a function g_j to define $b[j]$.

Initialization

$$[j = -p_z + 1 : 0] \qquad b[j] = z[j + p_z - 1]$$

Algorithm

$$[j = 1 : p_b] \qquad b[j] = g_j(b[-p_z + 1 : j - 1])$$

Output

$$z = b[p_b]$$

Example: Let be $f : \mathbb{R}^4 \longrightarrow \mathbb{R}$ defined by

$$f(z_0, z_1, z_2, z_3) = \cos\left(z_0 + e^{z_1}\right)\left[\sin\left(z_2\right) + \cos\left(z_3\right)\right] + z_1^{3/2} + z_3$$

Example: Let be $f : \mathbb{R}^4 \longrightarrow \mathbb{R}$ defined by

$$f(z_0, z_1, z_2, z_3) = \cos\left(z_0 + e^{z_1}\right)\left[\sin\left(z_2\right) + \cos\left(z_3\right)\right] + z_1^{3/2} + z_3$$

then, $p_z = 4$, $p_b = 4$, $z = z[0:3] = (z_0, z_1, z_2, z_3)$ and $y = f(z)$.

Example: Let be $f : \mathbb{R}^4 \longrightarrow \mathbb{R}$ defined by

$$f(z_0, z_1, z_2, z_3) = \cos\left(z_0 + e^{z_1}\right)\left[\sin\left(z_2\right) + \cos\left(z_3\right)\right] + z_1^{3/2} + z_3$$

then, $p_z = 4$, $p_b = 4$, $z = z[0:3] = (z_0, z_1, z_2, z_3)$ and $y = f(z)$.

Initialization

$$[j = -3 : 0] \qquad b[j] = z[j+3]$$

Example: Let be $f : \mathbb{R}^4 \longrightarrow \mathbb{R}$ defined by

$$f(z_0, z_1, z_2, z_3) = \cos\left(z_0 + e^{z_1}\right)\left[\sin\left(z_2\right) + \cos\left(z_3\right)\right] + z_1^{3/2} + z_3$$

then, $p_z = 4$, $p_b = 4$, $z = z[0:3] = (z_0, z_1, z_2, z_3)$ and $y = f(z)$.

Initialization

$$[j = -3 : 0] \qquad b[j] = z[j+3]$$

$$\begin{aligned} b[-3] &= z[0] = z_0 \\ b[-2] &= z[1] = z_1 \\ b[-1] &= z[2] = z_2 \\ b[0] &= z[3] = z_3 \end{aligned}$$

$$f(z_0, z_1, z_2, z_3) = \cos\left(z_0 + e^{z_1}\right)\left[\sin\left(z_2\right) + \cos\left(z_3\right)\right] + z_1^{3/2} + z_3$$

Algorithm

$$[j = 1 : 4] \qquad b[j] = g_j(b[-3 : j-1])$$

$$f(z_0, z_1, z_2, z_3) = \cos(z_0 + e^{z_1})[\sin(z_2) + \cos(z_3)] + z_1^{3/2} + z_3$$

Algorithm

$$[j = 1:4] \qquad b[j] = g_j(b[-3:j-1])$$

$$b[1] = g_1(b[-3:0]) = b[-3] + math.exp(b[-2])$$
$$b[2] = g_2(b[-3:1]) = math.sin(b[-1]) + math.cos(b[0])$$
$$b[3] = g_3(b[-3:2]) = math.pow(b[-2], 1.5d) + b[0]$$
$$b[4] = g_4(b[-3:3]) = math.cos(b[1]) * b[2] + b[3]$$

$$f(z_0, z_1, z_2, z_3) = \cos(z_0 + e^{z_1})[\sin(z_2) + \cos(z_3)] + z_1^{3/2} + z_3$$

Algorithm

$$[j = 1 : 4] \qquad b[j] = g_j(b[-3 : j-1])$$

$$b[1] = g_1(b[-3:0]) = b[-3] + math.exp(b[-2])$$
$$b[2] = g_2(b[-3:1]) = math.sin(b[-1]) + math.cos(b[0])$$
$$b[3] = g_3(b[-3:2]) = math.pow(b[-2], 1.5d) + b[0]$$
$$b[4] = g_4(b[-3:3]) = math.cos(b[1]) * b[2] + b[3]$$

Output

$$z = b[4]$$

Forward Mode

- The goal is to compute $\dfrac{\partial Y}{\partial z_i}$.

- The goal is to compute $\dfrac{\partial Y}{\partial z_i}$.
- We need to compute for each $-p_z + 1 \leq j \leq p_b$ the value

$$\begin{aligned}
\dot{b}[j:i] &= \frac{\partial}{\partial z[i]} b[j] \\
&= \sum_{k=-p_z+1}^{j-1} \frac{\partial}{\partial b[k]} b[j] \cdot \frac{\partial}{\partial z[i]} b[k] \\
&= \sum_{k=-p_z+1}^{j-1} \frac{\partial}{\partial b[k]} b[j] \cdot \dot{b}[k:i] \\
\dot{b}[j:i] &= \sum_{k=-p_z+1}^{j-1} \frac{\partial}{\partial b[k]} g_j(b[-p_z+1:j-1]) \cdot \dot{b}[k:i]
\end{aligned}$$

- Note that for $-p_z + 1 \leq j \leq 0$, $\dot{b}[j : i] = \delta_{i,-j}$.

- Note that for $-p_z + 1 \leq j \leq 0$, $\dot{b}[j:i] = \delta_{i,-j}$.
- In the step j, the values $\dot{b}[k:i]$ for $k < j$ are known and the derivatives of g_j can be computed by the program (Main Assumption).

Initialization

$$[j = -p_z + 1 : 0][i = 0 : p_z - 1] \qquad \dot{b}[j, i] = \delta_{i,-j}$$

Initialization

$$[j = -p_z + 1 : 0][i = 0 : p_z - 1] \qquad \dot{b}[j, i] = \delta_{i,-j}$$

Algorithm

$$[j = 1 : p_b][i = 0 : p_z - 1] \qquad \dot{b}[j, i] = \sum_{k=-p_z+1}^{j-1} \frac{\partial}{\partial b[k]} g_j \cdot \dot{b}[k : i]$$

Initialization

$$[j = -p_z + 1 : 0][i = 0 : p_z - 1] \qquad \dot{b}[j, i] = \delta_{i,-j}$$

Algorithm

$$[j = 1 : p_b][i = 0 : p_z - 1] \qquad \dot{b}[j, i] = \sum_{k=-p_z+1}^{j-1} \frac{\partial}{\partial b[k]} g_j \cdot \dot{b}[k : i]$$

Output

$$[i = 0 : p_z - 1] \qquad \frac{\partial Y}{\partial z_i} = \dot{b}[p_b, i]$$

Example: Let be $f : \mathbb{R}^4 \longrightarrow \mathbb{R}$ defined by

$$f(z_0, z_1, z_2, z_3) = \cos\left(z_0 + e^{z_1}\right)\left[\sin\left(z_2\right) + \cos\left(z_3\right)\right] + z_1^{3/2} + z_3$$

then, $p_z = 4$, $p_b = 4$, $z = z[0:3] = (z_0, z_1, z_2, z_3)$ and $y = f(z)$.

Example: Let be $f : \mathbb{R}^4 \longrightarrow \mathbb{R}$ defined by

$$f(z_0, z_1, z_2, z_3) = \cos\left(z_0 + e^{z_1}\right)\left[\sin\left(z_2\right) + \cos\left(z_3\right)\right] + z_1^{3/2} + z_3$$

then, $p_z = 4$, $p_b = 4$, $z = z[0:3] = (z_0, z_1, z_2, z_3)$ and $y = f(z)$.

Initialization

$$[j = -3 : 0][i = 0 : 3] \qquad \dot{b}[j, i] = \delta_{i,-j}$$

Example: Let be $f : \mathbb{R}^4 \longrightarrow \mathbb{R}$ defined by

$$f(z_0, z_1, z_2, z_3) = \cos\left(z_0 + e^{z_1}\right)\left[\sin\left(z_2\right) + \cos\left(z_3\right)\right] + z_1^{3/2} + z_3$$

then, $p_z = 4$, $p_b = 4$, $z = z[0:3] = (z_0, z_1, z_2, z_3)$ and $y = f(z)$.

Initialization

$$[j = -3 : 0][i = 0 : 3] \qquad \dot{b}[j, i] = \delta_{i,-j}$$

$$\dot{b}[j, i] = \begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} -3 \\ -2 \\ -1 \\ 0 \end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Algorithm

$$[j = 1:4][i = 0:3] \qquad \dot{b}[j,i] = \sum_{k=-3}^{j-1} \frac{\partial}{\partial b[k]} g_j \cdot \dot{b}[k:i]$$

Algorithm

$$[j = 1:4][i = 0:3] \qquad \dot{b}[j,i] = \sum_{k=-3}^{j-1} \frac{\partial}{\partial b[k]} g_j \cdot \dot{b}[k:i]$$

$$\dot{b}[1,i] = \sum_{k=-3}^{0} \frac{\partial}{\partial b[k]} g_1 \cdot \dot{b}[k:i]$$

Algorithm

$$[j = 1:4][i = 0:3] \qquad \dot{b}[j,i] = \sum_{k=-3}^{j-1} \frac{\partial}{\partial b[k]} g_j \cdot \dot{b}[k:i]$$

$$\dot{b}[1,i] = \sum_{k=-3}^{0} \frac{\partial}{\partial b[k]} g_1 \cdot \dot{b}[k:i]$$

$$\dot{b}[2,i] = \sum_{k=-3}^{1} \frac{\partial}{\partial b[k]} g_2 \cdot \dot{b}[k:i]$$

Algorithm

$$[j = 1:4][i = 0:3] \qquad \dot{b}[j,i] = \sum_{k=-3}^{j-1} \frac{\partial}{\partial b[k]} g_j \cdot \dot{b}[k:i]$$

$$\dot{b}[1,i] = \sum_{k=-3}^{0} \frac{\partial}{\partial b[k]} g_1 \cdot \dot{b}[k:i]$$

$$\dot{b}[2,i] = \sum_{k=-3}^{1} \frac{\partial}{\partial b[k]} g_2 \cdot \dot{b}[k:i]$$

$$\dot{b}[3,i] = \sum_{k=-3}^{2} \frac{\partial}{\partial b[k]} g_3 \cdot \dot{b}[k:i]$$

Algorithm

$$[j = 1:4][i = 0:3] \qquad \dot{b}[j,i] = \sum_{k=-3}^{j-1} \frac{\partial}{\partial b[k]} g_j \cdot \dot{b}[k:i]$$

$$\dot{b}[1,i] = \sum_{k=-3}^{0} \frac{\partial}{\partial b[k]} g_1 \cdot \dot{b}[k:i]$$

$$\dot{b}[2,i] = \sum_{k=-3}^{1} \frac{\partial}{\partial b[k]} g_2 \cdot \dot{b}[k:i]$$

$$\dot{b}[3,i] = \sum_{k=-3}^{2} \frac{\partial}{\partial b[k]} g_3 \cdot \dot{b}[k:i]$$

$$\dot{b}[4,i] = \sum_{k=-3}^{3} \frac{\partial}{\partial b[k]} g_4 \cdot \dot{b}[k:i]$$

- $g_1 = b[-3] + math.exp(b[-2])$

$$\begin{aligned}\dot{b}[1,i] &= \frac{\partial g_1}{\partial b[-3]}\dot{b}[-3:i] + \frac{\partial g_1}{\partial b[-2]}\dot{b}[-2:i]\\ &= 1\dot{b}[-3:i] + math.exp(b[-2])\dot{b}[-2:i]\end{aligned}$$

- $g_1 = b[-3] + math.exp(b[-2])$

$$\begin{aligned}\dot{b}[1,i] =& \frac{\partial g_1}{\partial b[-3]}\dot{b}[-3:i] + \frac{\partial g_1}{\partial b[-2]}\dot{b}[-2:i]\\ =& 1\dot{b}[-3:i] + math.exp(b[-2])\dot{b}[-2:i]\end{aligned}$$

- $g_2 = math.sin(b[-1]) + math.cos(b[0])$

$$\begin{aligned}\dot{b}[2,i] =& \frac{\partial g_2}{\partial b[-1]}\dot{b}[-1:i] + \frac{\partial g_2}{\partial b[0]}\dot{b}[0:i]\\ =& math.cos(b[-1])\dot{b}[-1:i] - math.sin(b[0])\dot{b}[0:i]\end{aligned}$$

- $g_3 = math.pow(b[-2], 1.5d) + b[0]$

$$\begin{aligned}\dot{b}[3,i] =& \frac{\partial g_3}{\partial b[-2]}\dot{b}[-2:i] + \frac{\partial g_3}{\partial b[0]}\dot{b}[0:i] \\ =& 1.5d * math.sqrt(b[-2])\dot{b}[-2:i] + 1\dot{b}[0:i]\end{aligned}$$

- $g_3 = math.pow(b[-2], 1.5d) + b[0]$

$$\dot{b}[3,i] = \frac{\partial g_3}{\partial b[-2]}\dot{b}[-2:i] + \frac{\partial g_3}{\partial b[0]}\dot{b}[0:i]$$
$$= 1.5d * math.sqrt(b[-2])\dot{b}[-2:i] + 1\dot{b}[0:i]$$

- $g_4 = math.cos(b[1]) * b[2] + b[3]$

$$\dot{b}[4,i] = \frac{\partial g_4}{\partial b[1]}\dot{b}[1:i] + \frac{\partial g_4}{\partial b[2]}\dot{b}[2:i] + \frac{\partial g_4}{\partial b[3]}\dot{b}[3:i]$$
$$= -math.sin(b[1]) * b[2]\dot{b}[1:i]$$
$$+ math.cos(b[1])\dot{b}[2:i] + 1\dot{b}[3:i]$$

$$
\begin{aligned}
&\dot{b}[4, i] \\
=& - math.sin(b[1]) * b[2]\dot{b}[1:i] + math.cos(b[1])\dot{b}[2:i] + 1\dot{b}[3:i] \\
=& - math.sin(b[1]) * b[2]\left(1\dot{b}[-3:i] + math.exp(b[-2])\dot{b}[-2:i]\right) \\
&+ math.cos(b[1])\left(math.cos(b[-1])\dot{b}[-1:i] - math.sin(b[0])\dot{b}[0:i]\right) \\
&+ 1\left(1.5d * math.sqrt(b[-2])\dot{b}[-2:i] + 1\dot{b}[0:i]\right) \\
=& \frac{\partial y}{\partial z_i}
\end{aligned}
$$

Adjoint Mode

- The goal is to compute $\dfrac{\partial y}{\partial z_i}$.

- The goal is to compute $\dfrac{\partial y}{\partial z_i}$.
- We need to compute the value

$$\bar{b}[j] = \frac{\partial y}{\partial b[j]} = \sum_{k=j+1}^{p_b} \frac{\partial y}{\partial b[k]} \cdot \frac{\partial b[k]}{\partial b[j]} = \sum_{k=j+1}^{p_b} \bar{b}[k] \frac{\partial g_k}{\partial b[j]}$$

for each intermediary variable with $-p_z + 1 \leq j \leq p_b$

- The goal is to compute $\dfrac{\partial y}{\partial z_i}$.
- We need to compute the value

$$\bar{b}[j] = \frac{\partial y}{\partial b[j]} = \sum_{k=j+1}^{p_b} \frac{\partial y}{\partial b[k]} \cdot \frac{\partial b[k]}{\partial b[j]} = \sum_{k=j+1}^{p_b} \bar{b}[k] \frac{\partial g_k}{\partial b[j]}$$

  for each intermediary variable with $-p_z + 1 \leq j \leq p_b$
- For $j = p_b$,

$$\bar{b}[p_b] = \frac{\partial y}{\partial b[p_b]} = 1$$

- The goal is to compute $\frac{\partial y}{\partial z_i}$.
- We need to compute the value

$$\bar{b}[j] = \frac{\partial y}{\partial b[j]} = \sum_{k=j+1}^{p_b} \frac{\partial y}{\partial b[k]} \cdot \frac{\partial b[k]}{\partial b[j]} = \sum_{k=j+1}^{p_b} \bar{b}[k] \frac{\partial g_k}{\partial b[j]}$$

  for each intermediary variable with $-p_z + 1 \leq j \leq p_b$
- For $j = p_b$,

$$\bar{b}[p_b] = \frac{\partial y}{\partial b[p_b]} = 1$$

- $\bar{y} = 1.0$ from the matrix representation

Initialization

$$\bar{b}[p_b] = 1.0$$

$$\bar{y} = 1.0$$

Initialization

$$\overline{b}[p_b] = 1.0$$

$$\overline{y} = 1.0$$

Algorithm

$$[j = p_b - 1 : -1 : -p_z + 1] \qquad \overline{b}[j] = \sum_{k=j+1}^{p_b} \overline{b}[k] \frac{\partial g_k}{\partial b[j]}$$

Initialization

$$\overline{b}[p_b] = 1.0$$

$$\overline{y} = 1.0$$

Algorithm

$$[j = p_b - 1 : -1 : -p_z + 1] \qquad \overline{b}[j] = \sum_{k=j+1}^{p_b} \overline{b}[k] \frac{\partial g_k}{\partial b[j]}$$

Output

$$[i = 0 : p_z - 1] \qquad \frac{\partial y}{\partial z_i} = \overline{b}[i - p_z + 1]$$

Example: Let be $f : \mathbb{R}^4 \longrightarrow \mathbb{R}$ defined by

$$f(z_0, z_1, z_2, z_3) = \cos\left(z_0 + e^{z_1}\right)\left[\sin\left(z_2\right) + \cos\left(z_3\right)\right] + z_1^{3/2} + z_3$$

then, $p_z = 4$, $p_b = 4$, $z = z[0:3] = (z_0, z_1, z_2, z_3)$ and $y = f(z)$.

Example: Let be $f : \mathbb{R}^4 \longrightarrow \mathbb{R}$ defined by

$$f(z_0, z_1, z_2, z_3) = \cos\left(z_0 + e^{z_1}\right)\left[\sin\left(z_2\right) + \cos\left(z_3\right)\right] + z_1^{3/2} + z_3$$

then, $p_z = 4$, $p_b = 4$, $z = z[0:3] = (z_0, z_1, z_2, z_3)$ and $y = f(z)$.

Initialization

$$\bar{b}[4] = 1.0$$

$$\bar{y} = 1.0$$

Example: Let be $f : \mathbb{R}^4 \longrightarrow \mathbb{R}$ defined by

$$f(z_0, z_1, z_2, z_3) = \cos\left(z_0 + e^{z_1}\right)\left[\sin\left(z_2\right) + \cos\left(z_3\right)\right] + z_1^{3/2} + z_3$$

then, $p_z = 4$, $p_b = 4$, $z = z[0:3] = (z_0, z_1, z_2, z_3)$ and $y = f(z)$.

Initialization

$$\overline{b}[4] = 1.0$$

$$\overline{y} = 1.0$$

Algorithm

$$[j = 3 : -1 : -3] \qquad \overline{b}[j] = \sum_{k=j+1}^{4} \overline{b}[k] \frac{\partial g_k}{\partial b[j]}$$

$$
\begin{aligned}
g_1 &= b[-3] + math.exp(b[-2]) \\
g_2 &= math.sin(b[-1]) + math.cos(b[0]) \\
g_3 &= math.pow(b[-2], 1.5d) + b[0] \\
g_4 &= math.cos(b[1]) * b[2] + \bar{b}[3] \\
\bar{b}[4] &= 1.0
\end{aligned}
$$

- $j = 3$

$$\bar{b}[3] = \bar{b}[4] \frac{\partial g_4}{\partial \bar{b}[3]} = 1$$

$$
\begin{aligned}
g_1 &= b[-3] + math.exp(b[-2]) \\
g_2 &= math.sin(b[-1]) + math.cos(b[0]) \\
g_3 &= math.pow(b[-2], 1.5d) + b[0] \\
g_4 &= math.cos(b[1]) * \bar{b}[2] + b[3] \\
\bar{b}[4] &= 1.0 \\
\bar{b}[3] &= 1.0
\end{aligned}
$$

- $j = 2$

$$
\bar{b}[2] = \bar{b}[3]\frac{\partial g_3}{\partial b[2]} + \bar{b}[4]\frac{\partial g_4}{\partial \bar{b}[2]} = math.cos(b[1])
$$

$$
\begin{aligned}
g_1 &= b[-3] + math.exp(b[-2]) \\
g_2 &= math.sin(b[-1]) + math.cos(b[0]) \\
g_3 &= math.pow(b[-2], 1.5d) + b[0] \\
g_4 &= math.cos(\overline{b}[1]) * b[2] + b[3] \\
\overline{b}[4] &= 1.0 \\
\overline{b}[3] &= 1.0 \\
\overline{b}[2] &= math.cos(b[1])
\end{aligned}
$$

- $j = 1$

$$
\begin{aligned}
\overline{b}[1] &= \overline{b}[2]\frac{\partial g_2}{\partial b[1]} + \overline{b}[3]\frac{\partial g_3}{\partial b[1]} + \overline{b}[4]\frac{\partial g_4}{\partial \overline{b}[1]} \\
&= -b[2] \cdot math.sin(b[1])
\end{aligned}
$$

$$
\begin{aligned}
g_1 &= b[-3] + math.exp(b[-2]) \\
g_2 &= math.sin(b[-1]) + math.cos(b[0]) \\
g_3 &= math.pow(b[-2], 1.5d) + b[0] \\
g_4 &= math.cos(b[1]) * b[2] + b[3] \\
\overline{b}[4] &= 1.0 \\
\overline{b}[3] &= 1.0 \\
\overline{b}[2] &= math.cos(b[1]) \\
\overline{b}[1] &= -b[2] \cdot math.sin(b[1])
\end{aligned}
$$

$$\bar{z}[3] = 1 - math.sin(b[0]) \cdot math.cos(b[1])$$
$$\bar{z}[2] = math.cos(b[-1]) \cdot math.cos(b[1])$$
$$\bar{z}[1] = 1.5d \cdot math.sqrt(b[-2]) - b[2] \cdot math.sin(b[1]) \cdot math.exp(b[-2])$$
$$\bar{z}[0] = -b[2] \cdot math.sin(b[1])$$

References

[1] M. Henrard. Algorithmic differentiation in finance explained. 2017.

[2] C. Homescu. "Adjoints and automatic (algorithmic) differentiation in computational finance". In: (2011).

Python Code Examples

[1] Forward Mode

[2] Adjoint Mode