

1 Einops

Reference: <https://github.com/arogozhnikov/einops>

1.1 Features

- self-documenting notation for layouts of input and output arrays
- low number of backend functions to implement
- focus on data rearrangements and simple transformations (axes reordering, decomposition, composition, reduction, repeats)
- focus on 1 tensor/array transformations
- notation uses strings
- supported notations: named axis, anonymous axis, unitary axis, ellipsis, (de)compose parenthesis
- supports a list of arrays as input with implied additional outer dimension corresponding to the list
- inferrable dimension sizes, given partial info as parameters
- hide backend framework inconsistency of notations for common array rearrangement operations
- use of proxy classes for specific backends
- caching of tensor type map to backend type for performance
- caching of patterns and axes
- caching of patterns, axes, and input shape: compute unknown axis sizes and shape verification on first time, otherwise reuse sequence of commands previously generated
- inverse transformations are easy to read off by switching patterns for input and output

1.2 Approaches

- evidence based for API design, via real world use cases
- explicit separation of a few functions over 1 function, for better error messages
- consideration of adoption friction and ease of use

1.3 Known Issues

- does not enforce axes alignment between operations
- no means of integrated analysis/tracking of shapes

2 Tensor Indexing

Index notation (for a binary operation):

$*(s_1, s_2, s_3)$

where

s_1 : input index set

s_2 : input index set

s_3 : output index set

2.1 Properties

- associative

let

$$s_3 \subseteq s_1 \cup s_2$$

$$s_4 \cap (s_1 \cup s_2) = \emptyset$$

then,

$$\begin{aligned} &*(s_3 s_4, s_4, s_3) (*(s_1, s_2 s_4, s_3 s_4)(A, B), C) \\ &= *(s_1, s_2, s_3)(A, *(s_2 s_4, s_4, s_2)(B, C)) \end{aligned}$$

order of evaluations:

$$(s_1 \rightarrow s_2 s_4) \rightarrow s_4 \rightarrow s_3$$

vs

$$s_1 \rightarrow (s_2 s_4 \rightarrow s_4) \rightarrow s_3$$

- commutative

$$(s_1, s_2, s_3)(A, B) = *(s_2, s_1, s_3)(B, A)$$

- distributive

$$\begin{aligned} &*(s_1, s_2, s_3)(A, B) + *(s_1, s_2, s_3)(A, C) \\ &= *(s_1, s_2, s_3)(A, B + C) \end{aligned}$$

where $s_3 \subseteq s_1 \cup s_2$

3 Derivative Definition

$$f : \mathbb{R}^{n_1 \times \dots \times n_k} \rightarrow \mathbb{R}^{m_1 \times \dots \times m_l}$$

$$D \in \mathbb{R}^{m_1 \times \dots \times m_l \times n_1 \times \dots \times n_k}$$

$$\lim_{h \rightarrow 0} \frac{\|f(x+h) - f(x) - D \circ h\|}{\|h\|} = 0$$

$$\iff D \text{ is a derivative of } f \text{ at } x$$

where inner tensor product is:

$$D \circ h = *(s_1 s_2, s_2, s_1)(D, h)$$

4 Forward Mode

$$\sum_i \frac{\partial v_i}{\partial x_j} \frac{\partial f}{\partial x_j} = \frac{\partial f}{\partial x_j}$$

where x_j are leaf input variables and

where pushforwards of predecessor nodes v_i are computed and cached by the time $\frac{\partial f}{\partial x_j}$ is computed

notation: let $\dot{v} = \frac{\partial v}{\partial x_j}$ be the pushforward of v

Generalized cases of local node connections:

- general unary function
- elementwise unary function
- binary addition
- binary multiplication

We seek to compute pushforwards for the above types of ops.

4.1 Pushforward of General Unary Function

let:

f be a unary function with:

- domain index set s_1
- range index set s_2

x be an input variable with index set s_3

$C = f(A)$, where A and C are nodes in expression DAG,

then, the pushforward \dot{C} is:

$$\dot{C} = *(s_2 s_1, s_1 s_3, s_2 s_3)(f'(A), \dot{A})$$

where f' is the derivative of f

4.1.1 Proof

$$f' \text{ is derivative of } f \iff \lim_{\tilde{h} \rightarrow 0} \frac{\|f(A+\tilde{h}) - f(A) - f'(A) \circ \tilde{h}\|}{\|\tilde{h}\|} = 0$$

$$\text{let } \tilde{h} = A(x+h) - A(x)$$

$$\tilde{h} \rightarrow 0 \text{ as } h \rightarrow 0$$

$$\lim_{\tilde{h} \rightarrow 0} \frac{\|f(A+\tilde{h}) - f(A) - f'(A) \circ \tilde{h}\|}{\|\tilde{h}\|} = 0$$

$$\lim_{h \rightarrow 0} \frac{\|f(A+A(x+h)-A(x)) - f(A) - f'(A) \circ (A(x+h)-A(x))\|}{\|A(x+h)-A(x)\|} = 0$$

$$\lim_{h \rightarrow 0} \frac{\|f(A(x+h)) - f(A) - f'(A) \circ (A(x+h)-A(x))\|}{\|A(x+h)-A(x)\|} = 0$$

$$\text{let } \dot{A} \text{ be derivative of } A \iff \lim_{h \rightarrow 0} \frac{\|A(x+h) - A(x) - \dot{A} \circ h\|}{\|h\|} = 0$$

triangular inequality:

$$\lim_{h \rightarrow 0} \frac{\|A(x+h) - A(x)\|}{\|h\|} - \frac{\|\dot{A} \circ h\|}{\|h\|}$$

$$\leq \lim_{h \rightarrow 0} \frac{\|A(x+h) - A(x) - \dot{A} \circ h\|}{\|h\|} = 0$$

$$\lim_{h \rightarrow 0} \frac{\|A(x+h) - A(x)\|}{\|h\|} = \frac{\|\dot{A} \circ h\|}{\|h\|}$$

substitute:

$$\lim_{h \rightarrow 0} \frac{\|f(A(x+h)) - f(A) - f'(A) \circ (A(x+h)-A(x))\|}{\|A(x+h)-A(x)\|} = 0$$

$$\lim_{h \rightarrow 0} \frac{\|f(A(x+h)) - f(A) - f'(A) \circ (\dot{A} \circ h)\|}{\|A(x+h)-A(x)\|} = 0$$

using definition of tensor inner product:

$$\dot{A} \circ h = *(s_1 s_3, s_3, s_1)(\dot{A}, h)$$

$$f'(A) \circ (\dot{A} \circ h) = f'(A) \circ (*(s_1 s_3, s_3, s_1)(\dot{A}, h))$$

$$f'(A) \circ (\dot{A} \circ h) = *(s_2 s_1, s_1, s_2)(f'(A), *(s_1 s_3, s_3, s_1)(\dot{A}, h))$$

using associativity:

$$f'(A) \circ (\dot{A} \circ h) = *(s_2 s_3, s_3, s_2)(*(s_2 s_1, s_1 s_3, s_2 s_3)(f'(A), \dot{A}), h)$$

$$f'(A) \circ (\dot{A} \circ h) = *(s_2 s_1, s_1 s_3, s_2 s_3)(f'(A), \dot{A}) \circ h$$

$$\lim_{h \rightarrow 0} \frac{\|f(A(x+h)) - f(A) - *(s_2 s_1, s_1 s_3, s_2 s_3)(f'(A), \dot{A}) \circ h\|}{\|A(x+h)-A(x)\|} = 0$$

$$\tilde{h} \rightarrow 0 \text{ as } h \rightarrow 0$$

$$(\exists k) \|A(x+h) - A(x)\| \leq \frac{1}{k} \|h\|$$

$$(\exists k) \frac{k}{\|h\|} \leq \frac{1}{\|A(x+h)-A(x)\|}$$

then

$$\lim_{h \rightarrow 0} \frac{k \|f(A(x+h)) - f(A) - *(s_2 s_1, s_1 s_3, s_2 s_3)(f'(A), \dot{A}) \circ h\|}{\|h\|} \leq$$

$$\lim_{h \rightarrow 0} \frac{\|f(A(x+h)) - f(A) - *(s_2 s_1, s_1 s_3, s_2 s_3)(f'(A), \dot{A}) \circ h\|}{\|A(x+h)-A(x)\|} = 0$$

$$\lim_{h \rightarrow 0} \frac{\|f(A(x+h)) - f(A) - *(s_2 s_1, s_1 s_3, s_2 s_3)(f'(A), \dot{A}) \circ h\|}{\|h\|} = 0$$

$$\iff *(s_2 s_1, s_1 s_3, s_2 s_3)(f'(A), \dot{A}) = \dot{C}$$

is derivative of $f(A(x))$ wrt. x , then \dot{C} is a pushforward of C

4.2 Pushforward of Elementwise Unary Function

let:

x be an input variable with index set s_2

f is an elementwise function

A is a node in the expression DAG with index set s_1

$$C = f(A)$$

then:

pushforward of node C , \dot{C} , is:

$$\dot{C} = *(s_1, s_1 s_2, s_1 s_2)(f'(A), \dot{A})$$

where f' is derivative of f

4.2.1 Proof

using definition of derivative:

$$\lim_{h \rightarrow 0} \frac{1}{\|h\|} \|A(x+h) - A(x) - \dot{A} \circ h\| = 0$$

$$\iff \dot{A} \text{ is derivative of } A$$

per scalar tensor entry indexed by multi-indices s :

$$\lim_{h \rightarrow 0} \frac{1}{\|h\|} |A(x+h)_s - A(x)_s - (\dot{A} \circ h)_s| = 0$$

using chain rule:

$$\text{let } g(x) = A(x)_s$$

$$f(g(x)) = f(A(x)_s)$$

$$(f(g(x)))' = f'(g(x))g'(x)$$

$$g'(x) = \dot{A}_s$$

$$(f(g(x)))' = f'(A(x)_s)\dot{A}_s$$

using definition of derivative:

$$(f(g(x)))' \text{ is derivative of } f(g(x))$$

$$\iff \lim_{h \rightarrow 0} \frac{1}{\|h\|} |f(g(x+h)) - f(g(x)) - (f(g(x)))' \circ h_s| = 0$$

$$\lim_{h \rightarrow 0} \frac{1}{\|h\|} |f(A(x+h)_s) - f(A(x)_s) - f'(A(x)_s)\dot{A}_s \circ h_s| = 0$$

$$\lim_{h \rightarrow 0} \frac{1}{\|h\|} |f(A(x+h)_s) - f(A(x)_s) - f'(A(x)_s)(\dot{A} \circ h)_s| = 0$$

above holds for all multi-indices $s \implies$

$$\lim_{h \rightarrow 0} \frac{\|f(A(x+h)) - f(A(x)) - *(s_1, s_1 s_2, s_1 s_2)(f'(A(x)), \dot{A} \circ h)\|}{\|h\|} = 0$$

using inner tensor product definition:

$$*(s_1, s_1, s_1)(f'(A(x)), \dot{A} \circ h) = *(s_1, s_1, s_1)(f'(A(x)), *(s_1 s_2, s_2, s_1)(\dot{A}, h))$$

using associativity rule:

$$*(s_1, s_1, s_1)(f'(A(x)), \dot{A} \circ h) = *(s_1 s_2, s_2, s_1)(*(s_1, s_1 s_2, s_1 s_2)(f'(A(x)), \dot{A}), h)$$

$$*(s_1, s_1, s_1)(f'(A(x)), \dot{A} \circ h) = *(s_1, s_1 s_2, s_1 s_2)(f'(A(x)), \dot{A}) \circ h$$

substitute:

$$\lim_{h \rightarrow 0} \frac{\|f(A(x+h)) - f(A(x)) - *(s_1, s_1 s_2, s_1 s_2)(f'(A(x)), \dot{A}) \circ h\|}{\|h\|} = 0$$

$$\iff *(s_1, s_1 s_2, s_1 s_2)(f'(A(x)), \dot{A}) \text{ is derivative of } f(A(x))$$

wrt. x

$$\iff \dot{C}, \text{ pushforward of } C, \text{ is } *(s_1, s_1 s_2, s_1 s_2)(f'(A(x)), \dot{A})$$

4.3 Pushforward of Binary Addition

let $C = f(A, B)$ where f is addition

then, $\dot{C} = \dot{A} + \dot{B}$ (sum of pushforwards of summands)

4.4 Pushforward of Binary Multiplication

let $C = *(s_1, s_2, s_3)(A, B)$, then

$\dot{C} = *(s_1 s_4, s_2, s_3 s_4)(\dot{A}, B) + *(s_1, s_2 s_4, s_3 s_4)(A, \dot{B})$

4.4.1 Proof

use definition of derivative: $\dot{C} = \frac{\partial C}{\partial A} \dot{A} + \frac{\partial C}{\partial B} \dot{B}$

consider the case of $\frac{\partial C}{\partial B} \dot{B}$ (contribution from node B):

$$\begin{aligned} \frac{\partial C}{\partial B} \dot{B} &= C(x+h) - C(x) - \dot{C} \circ h \\ &= *(s_1, s_2, s_3)(A, B(x+h)) \\ &\quad - *(s_1, s_2, s_3)(A, B(x)) \\ &\quad - *(s_1, s_2 s_4, s_3 s_4)(A, \dot{B}) \circ h \end{aligned}$$

where we let $\dot{C} = *(s_1, s_2 s_4, s_3 s_4)(A, \dot{B})$

where s_4 is the index set of input x to B

using inner tensor definition:

$$x \circ y = *(s_1 s_2, s_2, s_1)(x, y)$$

$$\begin{aligned} C(x+h) - C(x) - \dot{C} \circ h &= \\ &= *(s_1, s_2, s_3)(A, B(x+h)) \\ &\quad - *(s_1, s_2, s_3)(A, B(x)) \\ &\quad - *(s_3 s_4, s_4, s_3)(*(s_1, s_2 s_4, s_3 s_4)(A, \dot{B}), h) \end{aligned}$$

using associativity:

$$\begin{aligned} &*(s_3 s_4, s_4, s_3)(*(s_1, s_2 s_4, s_3 s_4)(A, \dot{B}), h) \\ &= *(s_1, s_2, s_3)(A, *(s_2 s_4, s_4, s_2)(\dot{B}, h)) \end{aligned}$$

$$\begin{aligned} C(x+h) - C(x) - \dot{C} \circ h &= \\ &= *(s_1, s_2, s_3)(A, B(x+h)) \\ &\quad - *(s_1, s_2, s_3)(A, B(x)) \\ &\quad - *(s_1, s_2, s_3)(A, *(s_2 s_4, s_4, s_2)(\dot{B}, h)) \end{aligned}$$

using distributivity:

$$\begin{aligned} C(x+h) - C(x) - \dot{C} \circ h &= \\ &= *(s_1, s_2, s_3)(A, B(x+h) - B(x) - *(s_2 s_4, s_4, s_2)(\dot{B}, h)) \end{aligned}$$

using definition of derivative:

$$\lim_{h \rightarrow 0} \frac{\|B(x+h) - B(x) - \dot{B} \circ h\|}{\|h\|} = 0 \iff$$

\dot{B} is a derivative of B at x

$$\begin{aligned} &\lim_{h \rightarrow 0} \frac{\|C(x+h) - C(x) - \dot{C} \circ h\|}{\|h\|} \\ &\leq \|A\| \lim_{h \rightarrow 0} \frac{\|B(x+h) - B(x) - \dot{B} \circ h\|}{\|h\|} = 0 \end{aligned}$$

$\iff *(s_1, s_2 s_4, s_3 s_4)(A, \dot{B})$ is the pushforward contribution to \dot{C} from B

similar logic follows for contribution to \dot{C} from A , then the proof is complete

5 Reverse Mode

5.1 Pullback of General Unary Function

node v_i stores $\frac{\partial y_j}{\partial v_i} = \bar{v}_i$

todo

computing for node z :

$\bar{z} = \frac{\partial y_j}{\partial z} = \sum_f \frac{\partial y_j}{\partial f} \frac{\partial f}{\partial z}$ where f is successor/output node of z ,
eg: there is an outgoing edge from z to f

$\bar{z} = \sum_f \bar{f} \frac{\partial f}{\partial z}$, where \bar{f} is the pullback of f and is computed
and cached at node f by the time \bar{z} needs to be updated

$O(m)$ rounds for m outputs y_j

Generalized cases of local node connections:

- general unary function
- elementwise unary function
- binary addition
- binary multiplication

5.2 Pullback of Elementwise Unary Function

todo

5.3 Pullback of Binary Addition

let:

$$C = +(A, B)$$

then,

$$\bar{A} = \bar{C}$$

$$\bar{B} = \bar{C}$$

where \bar{C} is the pullback of C

5.4 Pullback of Binary Multiplication

let: Y be an output node with index set s_4

$C = *(s_1, s_2, s_3)(A, B)$ be a multiplication node of expression

DAG

then, contribution of node C to pullback \bar{B} is:

$$*(s_4 s_3, s_1, s_4 s_2)(\bar{C}, A)$$

contribution of node C to pullback \bar{A} is:

$$*(s_4 s_3, s_2, s_4 s_1)(\bar{C}, B)$$

5.4.1 Proof

focus on derivation of pullback contribution to \bar{B}

using definition of derivative:

$$\lim_{\tilde{h} \rightarrow 0} \frac{1}{\|\tilde{h}\|} \|Y(C + \tilde{h}) - Y(C) - \frac{\partial Y}{\partial C} \circ \tilde{h}\| = 0$$

$$\iff \bar{C} = \frac{\partial Y}{\partial C} \text{ is derivative of } Y \text{ evaluated at } C$$

let $\tilde{h} = *(s_1, s_2, s_3)(A, h)$, then

$$\begin{aligned} & Y(C + \tilde{h}) - Y(C) - \bar{C} \circ \tilde{h} \\ &= Y(C + *(s_1, s_2, s_3)(A, h)) - Y(C) - \bar{C} \circ *(s_1, s_2, s_3)(A, h) \\ &= Y(*(s_1, s_2, s_3)(A, B) + *(s_1, s_2, s_3)(A, h)) - Y(*(s_1, s_2, s_3)(A, B)) - \\ & \quad \bar{C} \circ *(s_1, s_2, s_3)(A, h) \\ &= Y(*(s_1, s_2, s_3)(A, B + h)) - Y(*(s_1, s_2, s_3)(A, B)) - \bar{C} \circ \\ & \quad *(s_1, s_2, s_3)(A, h) \end{aligned}$$

using definition of tensor inner product:

$$\bar{C} \circ *(s_1, s_2, s_3)(A, h) = *(s_4 s_3, s_3, s_4)(\bar{C}, *(s_1, s_2, s_3)(A, h))$$

using associativity property:

$$\begin{aligned} \bar{C} \circ *(s_1, s_2, s_3)(A, h) &= *(s_4 s_2, s_2, s_4)(*(s_4 s_3, s_1, s_4 s_2)(\bar{C}, A), h) \\ \bar{C} \circ *(s_1, s_2, s_3)(A, h) &= *(s_4 s_3, s_1, s_4 s_2)(\bar{C}, A) \circ h \end{aligned}$$

$$\begin{aligned} & Y(C + \tilde{h}) - Y(C) - \bar{C} \circ \tilde{h} \\ &= Y(*(s_1, s_2, s_3)(A, B + h)) - Y(*(s_1, s_2, s_3)(A, B)) \\ & \quad - *(s_4 s_3, s_1, s_4 s_2)(\bar{C}, A) \circ h \end{aligned}$$

substitute:

$$0 = \lim_{h \rightarrow 0} \frac{1}{\|*(s_1, s_2, s_3)(A, h)\|} \|Y(*(s_1, s_2, s_3)(A, B + h)) - Y(*(s_1, s_2, s_3)(A, B)) - *(s_4 s_3, s_1, s_4 s_2)(\bar{C}, A) \circ h\|$$

$$\iff *(s_4 s_3, s_1, s_4 s_2)(\bar{C}, A) \text{ is derivative of}$$

$$Y(*(s_1, s_2, s_3)(A, B)) \text{ wrt. } B$$

then, contribution of C to \bar{B} , eg $(\frac{\partial Y}{\partial B})_{\text{via } C}$ is:

$$*(s_4 s_3, s_1, s_4 s_2)(\bar{C}, A)$$

use similar logic for derivation of pullback contribution to \bar{A} :

$$\text{contribution of } C \text{ to } \bar{A} \text{ is: } *(s_4 s_3, s_2, s_4 s_1)(\bar{C}, B)$$