Tensor Operations

Tensor Addition and Scalar Tensor Multiplication

These two operations will be defined elementwise, that is to say

Given **A**, **B** tensors, c a scalar,

$$(\mathrm{A}+\mathrm{B})_{\mathrm{i}_1\ \mathrm{i}_2\ \mathrm{i}_3\ \ldots \mathrm{i}_\mathrm{n}}=\mathrm{A}_{\mathrm{i}_1\ \mathrm{i}_2\ \mathrm{i}_3\ \ldots \mathrm{i}_\mathrm{n}}+\mathrm{B}_{\mathrm{i}_1\ \mathrm{i}_2\ \mathrm{i}_3\ \ldots \mathrm{i}_\mathrm{n}}$$

$$(\mathrm{cA})_{\mathrm{i}_1\ \mathrm{i}_2\ \mathrm{i}_3\ \ldots \mathrm{i}_\mathrm{n}}=\mathrm{c}*\mathrm{A}_{\mathrm{i}_1\ \mathrm{i}_2\ \mathrm{i}_3\ \ldots \mathrm{i}_\mathrm{n}}$$

Or extending this a bit, given **A**, **B** tensors, $c,\ d$ scalars,

$$(\mathrm{cA}+\mathrm{dB})_{\mathrm{i}_1\ \mathrm{i}_2\ \mathrm{i}_3\ \ldots \mathrm{i}_\mathrm{n}}=\mathrm{cA}_{\mathrm{i}_1\ \mathrm{i}_2\ \mathrm{i}_3\ \ldots \mathrm{i}_\mathrm{n}}+\mathrm{dB}_{\mathrm{i}_1\ \mathrm{i}_2\ \mathrm{i}_3\ \ldots \mathrm{i}_\mathrm{n}}$$

This has a few important ramifications, namely

- Tensors can only be added if all of their dimensions are aligned
- Tensor scalar multiplication and tensor addition do not change the overall dimension of the tensor
- Tensor scalar multiplication distributes over tensor addition
- Tensor scalar multiplication and tensor addition can be used to form vector spaces of tensors
- All linear operators (e.g. derivatives) will distribute and act normally with the scalar multiplications

Tensor Transposition

This operations is the analog of the matrix transpose in that it changes the way indexing works in a tensor

Given a tensor **A** with axes $i_1,\ i_2,\ \ldots\ i_\mathrm{N}$ we define the tensor transpose $\mathbf{A}_{\mathrm{T}(\{\mathrm{k}_\mathrm{m}\})}$ such that

$$\mathrm{A}_{\mathrm{T}(k_1,\ k_2,\ldots,k_N)} \text{ has axes } i_{\mathrm{k}_1},\ i_{k_2},\ \ldots\ i_{k_N}$$

$$(\mathrm{A}_{\mathrm{T}(k_1,\ k_2,\ldots,k_N)})_{j_1\ j_2\ \ldots,j_N}=\mathrm{A}_{k_{\mathrm{j}_1}\mathrm{k}_{\mathrm{j}_2}\ \ldots\ \mathrm{k}_{\mathrm{j}_\mathrm{N}}}$$

Note that for the case where $\mathrm{N} = 2$ the traditional transpose is

$$\mathrm{A}^{\mathrm{T}}=\mathrm{A}_{\mathrm{T}(2,\ 1)}$$

Axes Shift

If we are simply shift one axis to a new position, we define an alternate notation for convenience

Given a tensor **A** with axes $i_1,\ i_2,\ \ldots\ i_\mathrm{N}$, assuming WLOG that n>m

$$\mathrm{A}^{\mathrm{n}\to\mathrm{m}}=\mathrm{A}_{\mathrm{T}(1,\ 2,\ \ldots,\ \mathrm{m}-1,\ \mathrm{n},\ \mathrm{m},\ \mathrm{m}+1,\ \ldots,\ \mathrm{n}-1,\ \mathrm{n}+1,\ \ldots,\ \mathrm{N})}$$

Axes Swap

If we are simply swap two axes, we define an alternate notation for convenience

Given a tensor $\mathbf{A}$ with axes i_1, i_2, ... i_{N}, assuming WLOG that n>m

$$\mathrm{A}^{\mathrm{n}\leftrightarrow\mathrm{m}}=\mathrm{A}_{\mathrm{T}(1,\ 2,\ \ldots,\ \mathrm{m}-1,\ \mathrm{n},\ \mathrm{m}+1,\ \ldots,\ \mathrm{n}-1,\ \mathrm{m},\ \mathrm{n}+1,\ \ldots,\ \mathrm{N})}$$

Property 1 (Derivative Relations)

WLOG we will state this only for swaps, but the same ideas hold for standard transpositions

Scalar Derivatives

Given a swap $\mathbf{A}^{\mathrm{n}\to\mathrm{m}}$ and a single cordinate x

$$\frac{\partial}{\partial x}\mathbf{A}^{\mathrm{n}\to\mathrm{m}}=\left(\frac{\partial}{\partial x}\mathbf{A}\right)^{\mathrm{n}\to\mathrm{m}}$$

Proof

$\frac{\partial}{\partial x}$ is an elementwise operator

Vector Derivatives

Given a swap $\mathbf{A}^{\mathrm{n}\to\mathrm{m}}$ and a vector of cordinates X

$$\nabla_{\mathrm{X}^{\mathrm{k}}}\mathrm{A}^{\mathrm{n}\to\mathrm{m}}=(\nabla_{\mathrm{X}}\mathbf{A})^{\mathrm{n}+\mathrm{k}\to\mathrm{m}+\mathrm{k}}$$

Proof

$\nabla_{\mathrm{X}^{\mathrm{k}}}$ increases the rank of the tensor by k but operates elementwise

Property 2 (Composition)

Given a tensor $\mathbf{A}$ and two transpositions k_1, k_2,…,k_N and j_1, j_2,…,j_N

$$(\mathrm{A}_{\mathrm{T}(k_1,\ k_2,\ldots,k_N)})_{\mathrm{T}(j_1,\ j_2,\ldots,j_N)}=\mathrm{A}_{\mathrm{T}(k_{j_1},\ k_{j_2},\ldots,k_{j_N})}$$

Proof

By definition?

Property 3 (Indexing)

Given a transpositon $\mathbf{A}_{\mathrm{T}(k_1,\ k_2,\ldots,k_N)}$ and an ordering of the k_n, $\{o_n\}$

$$(A_{T(k_1,\ k_2,\ldots,k_N)})_{i_1\ i_2\ldots\ i_N} = A_{i_{o_1}\ i_{o_2}\ \ldots i_{o_N}}$$

(By ordering we mean a sequence such that $\forall_n\ k_{o_n} < k_{o_{n+1}}$)

Proof

We won't do a full proof, which would basically have to be done by induction by showing that it works for a single axis swap and then showing that if it works for n swaps it works for $n+1$

Assume that $\mathbf{A}$ has axes $\{a_n\} \Rightarrow A_{T(k_1,\ k_2,\ldots,k_N)}$ has axes $\{a_{k_n}\}$

When we ask for element $i_1\ i_2\ldots\ i_N$ we are asking, therefore, for element i_m of axis a_{k_m}

E.g. suppose $m = 5$, i_5 will be taken from a_{k_5} then suppose $k_5 = 2$

$\Rightarrow$ we want element i_5 from axis 2 in the original

$\Rightarrow$ we need to find an *ordering* of the $i_1\ i_2\ldots\ i_N$ such that $o_2 = 5$

$\therefore$ we use the ordering of the $\{k_n\}$

Corr (Shift Indexing)

Given a shift $\mathbf{A}^{n\to m}$ WLOG assume $n < m$

$$(A^{n\to m})_{i_1,\ldots,\ i_n,\ldots,\ i_m,\ \ldots\ i_N} = A_{i_1,\ldots,\ i_m,\ i_n,\ \ldots,\ i_{m-1},\ i_{m+1},\ \ldots,\ i_N}$$

or in more prosaic terms, the indices shift in the opposite direction that the axes shift

Property 4 (Shift Compositions)

Given a double shift $(\mathbf{A}^{n\to m})^{l\to k}$

if m=l, $(\mathbf{A}^{n\to m})^{l\to k} = \mathbf{A}^{n\to k}$, otherwise there is no reduction to be had

Table of N=3 cases

Proof

WLOG we will assume $\mathbf{A}$ has dimension $N = \max\{n, m, l, k\} - \min\{n, m, l, k\}$

Case 1 : n<m<l< k

If n<m<l<k, we have two distinct uncomposable shift operations, n→m and l→k

Note that in this case and only in this case (I think) the operations commute

Case 2: n<m=l<k

WLOG suppose n=1 and k=N

$$\begin{aligned}\left(A^{1\to m}\right)^{m\to N} &= (A_{T(2,\ \ldots,\ m-1,\ 1,\ m,\ \ldots,\ N)})_{T(1,\ \ldots,\ m-1,\ m+1,\ \ldots,\ N,\ m)}\\ &= A_{T(2,\ \ldots,\ N,\ 1)}\end{aligned}$$

$$=A^{1\to N}$$

Case 3: n<m & l<k & n=l

WLOG suppose n=1 and k=N

$$\left(A^{1\to m}\right)^{1\to N}=(A_{T(2,\ \ldots,\ m-1,\ 1,\ m,\ \ldots,\ N)})_{T(2,\ \ldots,\ N,\ 1)}$$
$$=A_{T(3,\ \ldots,\ m-1,\ 1,\ m,\ \ldots,\ N,\ 2)}$$

$\therefore$ Uncomposable in general

Case 4: n<m & l<k & m=k

WLOG suppose n=1 and k=N

$$\left(A^{1\to N}\right)^{l\to N}=(A_{T(2,\ \ldots,\ N,\ 1)})_{T(1,\ \ldots,\ k-1,\ k+1,\ \ldots,\ N,\ k)}$$
$$=A_{T(2,\ \ldots,\ N,\ 1,\ k)}$$

$\therefore$ Uncomposable in general

Case 5: n>m & l<k & n=k

WLOG suppose m=1 and k=N

$$\left(A^{n\to 1}\right)^{l\to N}=A_{T(N,\ 1,\ \ldots,\ l-1,\ l+1,\ \ldots,\ N-1,\ l)}$$

$\therefore$ Uncomposable in general

Tensor Derivatives

Given a tensor $\mathbf{A}$, and some variable x upon which the elements of $\mathbf{A}$ may depend we'll define the derivative element-wise such that

$$\left(\frac{\partial}{\partial x}A\right)_{i_1\, i_2 \ldots\, i_n}=\frac{\partial}{\partial x}A_{i_1\, i_2 \ldots\, i_n}$$

We will also define derivative tensors such that given a vector of coordinates $\mathbf{x}$ and a scalar function f,

$$\nabla_x f=\left(\frac{\partial}{\partial x_1}f \quad \frac{\partial}{\partial x_2}f \quad \cdots \quad \frac{\partial}{\partial x_n}f\right)$$

if instead of scalar f we have a tensor $\mathbf{A}$ everything is basically the same but we get the tensor of derivatives

Higher dimensional tensor derivative operators may also be defined, e.g. given another vector of coordinates $\mathbf{y}$ we can have

$$\nabla_{x,y} f = \begin{pmatrix} \frac{\partial^2}{\partial x_1 \, \partial y_1} f & \frac{\partial^2}{\partial x_1 \, \partial y_2} f & \cdots & \frac{\partial^2}{\partial x_1 \, \partial y_m} f \\ \frac{\partial^2}{\partial x_2 \, \partial y_1} f & & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial^2}{\partial x_n \, \partial y_1} f & & \cdots & \frac{\partial^2}{\partial x_n \, \partial y_m} f \end{pmatrix}$$

Chain Rule

We'll define the chain rule over a tensor of derivatives in the standard kind of way by supplying a Jacobian for the transformation, e.g. given a vector of coordinates $\mathbf{x}$ and a function f and a set of coordinates upon which f depends, $\mathbf{y}$,

$$\nabla_x f = \nabla_x y \, \nabla_y f$$

The same basic form holds for higher dimensional derivative tensors, e.g. given a vector valued function $\mathbf{g}$

$$\nabla_x g = \nabla_x y \, \nabla_y g$$

And for a general derivative tensor for a tensor valued function $\mathbf{A}$

$$\nabla_x A = \nabla_x y \odot \nabla_y A$$

where $\odot$ is the tensor dot operation talked about later

We'll also note that this still works in a slightly different form with univariate derivatives

$$\frac{\partial}{\partial x_i} f = (\nabla_x y)_i \, \nabla_y f$$

Proof

Case 1:

$$\begin{aligned} (\nabla_x f)_i &= \frac{\partial}{\partial x_i} f \\ &= \sum_{j=1}^{N} \frac{\partial y_j}{\partial x_i} \frac{\partial}{\partial y_j} f \\ &= (\nabla_x y)_i \cdot \nabla_y f \end{aligned}$$

Case 2:

$$\begin{aligned} (\nabla_x g)_{ij} &= \left(\frac{\partial}{\partial x_i} g \right)_j \\ &= \frac{\partial}{\partial x_i} g_j \end{aligned}$$

$$
\begin{aligned}
&=\sum_k \frac{\partial y_k}{\partial x_i} \frac{\partial}{\partial y_k} g_j \\
&=(\nabla_x y)_i \cdot (\nabla_y g_j)
\end{aligned}
$$

Case 3:

Left to reader

Tensor Dot

Definition

We'll define the tensor dot $\odot$ to be a binary tensor operation that generalizes the dot product. It will operate recursively such that:

- Given a vector **a** of length N and tensor **B** with primary dimension N we'll define this such that

$$a \odot B = \sum_{i=1}^{N} a_i B_i$$

- Given a tensor **A** with primary dimension N and innermost dimension M and a tensor **B** with primary dimension M we'll define this such that

$$A \odot B = (A_1 \odot B \quad A_2 \odot B \quad \cdots \quad A_N \odot B)$$

Property 1 (associativity)

Given **A**, **B**, **C** tensors,

$$(A \odot B) \odot C = A \odot (B \odot C)$$

Proof

LTR

Property 2 (dot product)

Given vectors **a**, **b** of length N

$$a \odot b = a \cdot b$$

where $\cdot$ is the classic dot product

Proof

Basically just by definition

$$a \odot b = \sum_{i=1}^{N} a_i\, b_i = a \cdot b$$

Property 3 (vector matrix)

Given a vector **a** of length N and matrix $\mathbf{B}_{NxM}$

$$a \odot B = aB$$

where **aB** is interpreted in the normal sense of a matrix product

Proof

We start with

$$a \odot B = \sum_{i=1}^{N} a_i\, B_i$$

Given $1 \le i \le N$

$$a_i\, B_i = \begin{pmatrix} a_i\, B_{i1} & a_i\, B_{i2} & \cdots & a_i\, B_{iM} \end{pmatrix}$$

then

$$\begin{aligned} \sum_{i=1}^{N} a_i\, B_i &= \begin{pmatrix} \sum_{i=1}^{N} a_i\, B_{i1} & \sum_{i=1}^{N} a_i\, B_{i2} & \cdots & \sum_{i=1}^{N} a_i\, B_{iM} \end{pmatrix} \\ &= \begin{pmatrix} a \cdot B_{:1} & a \cdot B_{:2} & \cdots & a \cdot B_{:M} \end{pmatrix} \end{aligned}$$

which is of course the classic vector matrix product

Property 4 (matrix product)

Given matrices $\mathbf{A}_{NxM}$ and $\mathbf{B}_{MxL}$

$$A \odot B = AB$$

Proof

$$A \odot B = \begin{pmatrix} A_1 \odot B & A_2 \odot B & \cdots & A_N \odot B \end{pmatrix}$$

Given $1 \le i \le M$, then, by Property 2,

$$\begin{aligned} \begin{pmatrix} A_1 \odot B & A_2 \odot B & \cdots & A_N \odot B \end{pmatrix} &= \begin{pmatrix} A_1\, B & A_2\, B & \cdots & A_N\, B \end{pmatrix} \\ &= \begin{pmatrix} A_1 \cdot B_{:1} & A_1 \cdot B_{:2} & \cdots & A_1 \cdot B_{:L} \\ A_2 \cdot B_{:1} & & & \\ \vdots & & \ddots & \vdots \\ A_N \cdot B_{:1} & & \cdots & A_N \cdot B_{:L} \end{pmatrix} \end{aligned}$$

which is how **AB** is defined

Property 5 (indexing)

Given $\mathbf{A}$, $\mathbf{B}$ tensors of dimension N and M, suppose $k>0$, we have

$$(A\odot B)_{i_1\, i_2 \ldots\, i_{N-1}\, j_2\, \ldots j_M} = A_{i_1\, i_2 \ldots\, i_{N-1}} \cdot B_{:j_2\, \ldots j_M} \quad (1)$$

$$(A\odot B)_{i_1\, i_2 \ldots\, i_{N-k}} = A_{i_1\, i_2 \ldots\, i_{N-1}} \odot B \quad (2)$$

$$(A\odot B)_{i_1\, i_2 \ldots\, i_{N-1}\, j_2\, \ldots j_{M-k}} = A_{i_1\, i_2 \ldots\, i_{N-1}} \odot B_{:j_2\, \ldots j_{M-k}:\ldots} \quad (3)$$

where : means to index every element along that axis

Intuitively this means that we simply distribute the indices from left to right until we have run out

Proof

(I proved this for an older statement first so the proof doesn't really directly correspond to what I stated, but is equivalent)

(1)

LTR

(2)

We will show this first for the case that k = N-1

$$\begin{aligned}(A\odot B)_i &= (A_1\odot B \quad A_2\odot B \quad \cdots \quad A_N\odot B)_i \\ &= A_i\odot B\end{aligned}$$

Then we'll do this by induction, à la Property 6

Base Case

Suppose N=2, this reduces to (1), therefore it holds

Induction on N:

Suppose this holds for all $\mathbf{A'}$ with dimension $N-1$. Next as we have

$$A\odot B = (A_1\odot B \quad A_2\odot B \quad \cdots \quad A_N\odot B)$$

and $\forall_i$ $\dim(A_i) = N-1$

$$\begin{aligned}(A\odot B)_{i_1\, i_2 \ldots\, i_{N-1}} &= (A_{i_1}\odot B)_{i_2 \ldots\, i_{N-1}} \\ &= (A_{i_1})_{i_2 \ldots\, i_{N-1}}\odot B \\ &= A_{i_1\, i_2 \ldots\, i_{N-1}}\odot B\end{aligned}$$

(3)

By (1) we have:

$$(A\odot B)_{i_1\, i_2 \dots\, i_{N-1}\, j_2\, \dots j_M} = (A_{i_1\, i_2 \dots\, i_{N-1}} \odot B)_{j_2\, \dots j_M}$$

Next note that $\dim(A_{i_1\, i_2 \dots\, i_{N-1}}) = 1$, so

$$\begin{aligned}
(A_{i_1\, i_2 \dots\, i_{N-1}} \odot B)_{j_2\, \dots j_M} &= \left(\sum_{k=1}^{N} (A_{i_1\, i_2 \dots\, i_{N-1}})_k\, B_k\right)_{j_2\, \dots j_M} \\
&= \sum_{k=1}^{N} (A_{i_1\, i_2 \dots\, i_{N-1} k}\, B_k)_{j_2\, \dots j_M} \text{ (by distribution of indexing)} \\
&= \sum_{k=1}^{N} A_{i_1\, i_2 \dots\, i_{N-1} k}\, B_{k\, j_2\, \dots j_M} \text{(by commutation of indexing and scalar mul.)} \\
&= A_{i_1\, i_2 \dots\, i_{N-1}} \cdot B_{:j_2\, \dots j_M}
\end{aligned}$$

Property 6 (dimension)

Given **A** and **B** tensors,

$$\dim(A\odot B) = \dim(A) + \dim(B) - 2$$

Proof

Lemma:

If $\dim(A) = 1$,

$$\dim(A\odot B) = \dim(B) - 1$$

Lemma Proof:

$$A \odot B = \sum_{i=1}^{N} A_i\, B_i$$

Given **A'**, **B'** tensors with the same dimensions, we'll recall that

$$\dim(A' + B') = \dim(A') = \dim(B')$$

therefore

$$\begin{aligned}
\dim\left(A \odot B\right) &= \dim\left(\sum_{i=1}^{N} A_i\, B_i\right) \\
&= \dim(B_i) \\
&= \dim(B) - 1
\end{aligned}$$

QED

Returning to the main problem, we'll use induction on $\dim(A)$, taking the Lemma as our base case

Suppose that $\forall A'$ with $\dim(A') = n-1$, $\dim(A'\odot B) = n-1+\dim(B)-2$

Given A with dim(A) = n,

$$A \odot B = (A_1 \odot B \quad A_2 \odot B \quad \cdots \quad A_N \odot B)$$

Given $1 \le i \le N$,

$$\begin{aligned}\dim(A_i \odot B) &= n-1+\dim(B)-2\\ &= n+\dim(B)-3\\ &= \dim(A)+\dim(B)-3\end{aligned}$$

Then as

$$A \odot B = (A_1 \odot B \quad A_2 \odot B \quad \cdots \quad A_N \odot B)$$

since we have added a new outer dimension, we have

$$\begin{aligned}\forall_i \ \dim(A \odot B) &= \dim(A_i \odot B) + 1\\ &= \dim(A)+\dim(B)-3+1\\ &= \dim(A)+\dim(B)-2\end{aligned}$$

And then by induction we have this for all n

Property 7 (scalar multiplication)

Given **A**, **B**, tensors, c a scalar,

$$A \odot (cB) = c(A \odot B)$$

Proof

The general strategy for this proof would be the same as for Property 6, where we show it holds for tensors of dimension one and then generalize upwards via induction on the dimension of **A** or **C**. In the interest of time, we will simply show that it operates for dim(A)=1

Suppose dim(A) = 1,

$$\begin{aligned}A \odot (cB) &= \sum_{i=1}^{N} A_i(cB_i)\\ &= c\sum_{i=1}^{N} A_i B_i\\ &= c(A \odot B)\end{aligned}$$

Property 8 (distributivity)

Given **A**, **B**, **C** tensors,

$$A \odot (B+C) = A \odot B + A \odot C$$

and

$$(A+B)\odot C = A\odot C + B\odot C$$

Proof

As in Property 7, we will simply show that it operates for tensors of dimension one.

Suppose dim(A) = 1,

$$\begin{aligned}
A\odot(B+C) &= \sum_{i=1}^{N} A_i * (B+C)_i \\
&= \sum_{i=1}^{N} A_i * (B_i + C_i) \\
&= \sum_{i=1}^{N} A_i\, B_i + A_i\, C_i \\
&= \sum_{i=1}^{N} A_i\, B_i + \sum_{i=1}^{N} A_i\, C_i \\
&= A\odot B + A\odot C
\end{aligned}$$

Further, now suppose dim(B)=1,

$$\begin{aligned}
(A+B)\odot C &= \sum_{i=1}^{N} (A+B)_i\, C_i \\
&= \sum_{i=1}^{N} (A_i + B_i)\, C_i \\
&= \sum_{i=1}^{N} A_i\, C_i + \sum_{i=1}^{N} B_i\, C_i \\
&= A\odot C + B\odot C
\end{aligned}$$

Property 9 (product rule)

9.a: scalar derivatives

Given **A**, **B** tensors, x some variable upon which the elements of **A** and **B** may depend

$$\frac{\partial}{\partial x} A\odot B = \left(\frac{\partial}{\partial x} A\right)\odot B + A\odot\left(\frac{\partial}{\partial x} B\right)$$

Proof

As in Property 7, we will simply show that it operates for tensors of dimension one.

Suppose dim(A) = 1,

$$\frac{\partial}{\partial \mathrm{x}} A\odot B = \frac{\partial}{\partial \mathrm{x}} \sum_{i=1}^{N} A_i\, B_i$$

$$=\sum_{i=1}^{N}\frac{\partial}{\partial x}(A_i\, B_i)$$

then as the derivative will operate element-wise on the tensor—

is this obvious?—we will have

$$\frac{\partial}{\partial x}(A_i\, B_i)_{j_1\, j_2\, \ldots j_n}=\frac{\partial}{\partial x}A_i(B_i)_{j_1\, j_2\, \ldots j_n}$$
$$=\left(\frac{\partial}{\partial x}A_i\right)(B_i)_{j_1\, j_2\, \ldots j_n}+A_i\left(\frac{\partial}{\partial x}(B_i)_{j_1\, j_2\, \ldots j_n}\right)$$

therefore

$$\frac{\partial}{\partial x}A\odot B=\sum_{i=1}^{N}\frac{\partial}{\partial x}(A_i\, B_i)$$
$$=\sum_{i=1}^{N}\left(\frac{\partial}{\partial x}A_i\right)B_i+A_i\left(\frac{\partial}{\partial x}B_i\right)$$
$$=\sum_{i=1}^{N}\left(\frac{\partial}{\partial x}A_i\right)B_i+\sum_{i=1}^{N}A_i\left(\frac{\partial}{\partial x}B_i\right)$$
$$=\left(\frac{\partial}{\partial x}A\right)\odot B+A\odot\left(\frac{\partial}{\partial x}B\right)$$

9.b: vector derivatives

For vector derivatives this has a slightly different form

$$\nabla_X(A\odot B)=(\nabla_X A)\odot B\;+\;(A\odot(\nabla_{X_i}B))_{i=1\ldots}$$

or, using transposition notation, if A is of rank a and B is of rank b

$$\nabla_X(A\odot B)=(\nabla_X A)\odot B\;+\;\left(A\odot(\nabla_X B)^{2\to 1}\right)^{a\to 1}$$

Proof

Property 10 (transposition)

Given **A** with axes $\{a_n\}_{n=1}^{N+1}$ **B** with asxes $\{b_m\}_{m=1}^{M+1}$ and some transposition $\{k_l\}$,

if $\forall_i\; i\le N \Rightarrow k_i \le N$ and $\forall_j\; j>N \Rightarrow k_j>N$,

$$(A\odot B)_{T(\{k_l\})}=A_{T(\{k_l|l\le N\})}\odot B_{T(\{k_l-N|l>N\})}$$

if $\forall_i\; i>M \Rightarrow k_i \le M$ and $\forall_j\; j\le M \Rightarrow k_j>M$,

$$(A\odot B)_{T(\{k_l\})}=B_{T(\{k_l|l\le M\})}\odot A_{T(\{k_l-M|l>M\})}$$

otherwise $\nexists$ $\{i_l\}$, $\{j_l\}$ such that

$$(A \odot B)_{T(\{k_l\})} = A_{T(\{i_l\})} \odot B_{T(\{j_l\})} \text{ or } (A \odot B)_{T(\{k_l\})} = B_{T(\{j_l\})} \odot A_{T(\{i_l\})}$$

Relation to Einstein Summation Notation (einsum)

Basically a tensor dot is a simpler version of einsum where we only contract along the inner dimensions, e.g.

Given tensors $\mathbf{A}_{ijkl}$ and $\mathbf{B}_{l\alpha\beta\gamma}$ we have

$$A \odot B = A_{ijkl} \, B_{l\alpha\beta\gamma} = (AB)_{ijk\alpha\beta\gamma}$$

Note that when written like this the validity of Properties 2-6 are evident "by notation"

Notations

Since these derivatives involve a large number of $\nabla . \cdot$, $x^{\cdot \to \cdot}$, and $\cdot \odot \cdot$ operations we'll introduce a convenient shorthand where

$$\nabla_x A \equiv A_x$$
$$A^{a \to b} \equiv A^{a:b}$$
$$A \odot B \equiv AB$$

This means the product rule for tensor derivatives can be expressed as

$$\nabla_x (A \odot B) = A_x \, B + \left(A (B_x)^{2:1}\right)^{a:1}$$

Code