ch1_derivation

November 25, 2020

1 Derivations For Matrix Derivative Rule of $L = \lambda(\sigma_{\text{apply}}(\nu(X, W)))$

[1]:
```
from sympy import Matrix, Symbol, derive_by_array, Lambda, Function,␣
 ↪MatrixSymbol, Identity, Derivative, symbols, diff
from sympy.abc import x, i, j, a, b
```

[2]:
```
from functools import reduce
import itertools

from typing import *
import sys
import os

PATH: str = '/development/projects/statisticallyfit/github/
 ↪learningmathstat/PythonNeuralNetNLP'

UTIL_DISPLAY_PATH: str = PATH + "/src/utils/GeneralUtil/"

NEURALNET_PATH: str = PATH + '/src/NeuralNetworkStudy/books/
 ↪SethWeidman_DeepLearningFromScratch'

#os.chdir(PATH)
#assert os.getcwd() == NEURALNET_PATH

sys.path.append(PATH)
#assert PATH in sys.path

sys.path.append(UTIL_DISPLAY_PATH)
```

```
#assert UTIL_DISPLAY_PATH in sys.path

sys.path.append(NEURALNET_PATH)#
#assert NEURALNET_PATH in sys.path
```

[3]:
```
from src.utils.GeneralUtil import *
from src.MatrixCalculusStudy.MatrixDerivLib.symbols import Deriv
from src.MatrixCalculusStudy.MatrixDerivLib.diff import diffMatrix
from src.MatrixCalculusStudy.MatrixDerivLib.printingLatex import
    ↪myLatexPrinter

from IPython.display import display, Math
from sympy.interactive import printing
printing.init_printing(use_latex='mathjax', latex_printer= lambda e,
    ↪**kw: myLatexPrinter.doprint(e))
```

[4]:
```
import itertools

from functools import reduce

from typing import *
```

[5]:
```
n,m,p = 3,3,2



xi  = Symbol('xi')
xi_1  = Symbol('xi_1')
beta = Symbol('beta')


X = Matrix(n, m, lambda i,j : var_ij('x', i, j))
W = Matrix(m, p, lambda i,j : var_ij('w', i, j))

A = MatrixSymbol('X',n,m)
B = MatrixSymbol('W',m,p)

# matrix variable for sympy Lambda function arguments
M = MatrixSymbol('M', i, j)# abstract shape
```

[6]:
```
#compose(sigmaApply)(N).replace(sigmaApply, sigmaApply_).diff(N).
 ↪subs({N : vN(A,B)}).doit()
```

[7]:
```
###N = MatrixSymbol("N", n, p)# shape of A*B### use Nelem below


showGroup([
    X, W, Matrix(A)
])
```

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix}$$

$$\begin{bmatrix} X_{0,0} & X_{0,1} & X_{0,2} \\ X_{1,0} & X_{1,1} & X_{1,2} \\ X_{2,0} & X_{2,1} & X_{2,2} \end{bmatrix}$$

[8]:
```
v = Function("nu",applyfunc=True)
v_ = lambda a,b: a*b
vL = Lambda((a,b), a*b)
VL = Lambda((A,B), MatrixSymbol('V', A.shape[0], B.shape[1]))
vN = lambda mat1, mat2: Matrix(mat1.shape[0], mat2.shape[1], lambda
 ↪i, j: Symbol("n_{}{}".format(i+1, j+1))); vN

Nelem = vN(X, W)
Nspec = v_(X,W)
N = v(A,B)


showGroup([
    Nelem, Nspec, N, VL
])
```

$$\begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \\ n_{31} & n_{32} \end{bmatrix}$$

$$\begin{bmatrix} w_{11}x_{11} + w_{21}x_{12} + w_{31}x_{13} & w_{12}x_{11} + w_{22}x_{12} + w_{32}x_{13} \\ w_{11}x_{21} + w_{21}x_{22} + w_{31}x_{23} & w_{12}x_{21} + w_{22}x_{22} + w_{32}x_{23} \\ w_{11}x_{31} + w_{21}x_{32} + w_{31}x_{33} & w_{12}x_{31} + w_{22}x_{32} + w_{32}x_{33} \end{bmatrix}$$

$\nu(X, W)$

$((X,\ W) \mapsto V)$

[9]:
```
sigma = Function('sigma')
sigmaApply = Function("sigma_apply") #lambda matrix:  matrix.
 ↪applyfunc(sigma)
sigmaApply_ = lambda matrix: matrix.applyfunc(sigma)
sigmaApply_L = Lambda(M, M.applyfunc(sigma))


S = sigmaApply(N)
Sspec = S.subs({A:X, B:W}).replace(v, v_).replace(sigmaApply,
 ↪sigmaApply_)
Selem = S.replace(v, vN).replace(sigmaApply, sigmaApply_)


showGroup([
    S, Sspec, Selem
])
```

$\sigma_{apply}(\nu(X, W))$

$$\begin{bmatrix} \sigma(w_{11}x_{11} + w_{21}x_{12} + w_{31}x_{13}) & \sigma(w_{12}x_{11} + w_{22}x_{12} + w_{32}x_{13}) \\ \sigma(w_{11}x_{21} + w_{21}x_{22} + w_{31}x_{23}) & \sigma(w_{12}x_{21} + w_{22}x_{22} + w_{32}x_{23}) \\ \sigma(w_{11}x_{31} + w_{21}x_{32} + w_{31}x_{33}) & \sigma(w_{12}x_{31} + w_{22}x_{32} + w_{32}x_{33}) \end{bmatrix}$$

$$\begin{bmatrix} \sigma(n_{11}) & \sigma(n_{12}) \\ \sigma(n_{21}) & \sigma(n_{22}) \\ \sigma(n_{31}) & \sigma(n_{32}) \end{bmatrix}$$

[10]:
```
lambd = Function("lambda")
lambd_ = lambda matrix : sum(matrix)
#lambda_L = Lambda(M, sum(M))

ABres = MatrixSymbol("R", A.shape[0], B.shape[1])
lambd_L = Lambda(ABres, sum(ABres))

#L = lambd(sigmaApply(v(A,B)))
```

```
L = compose(lambd, sigmaApply, v)(A, B)
L
```

[10]: $\lambda(\sigma_{apply}(\nu(X, W)))$

[11]:
```
elemToSpecD = dict(itertools.chain(*[[(Nelem[i, j], Nspec[i, j]) for
    j in range(p)] for i in range(n)]))
elemToSpec = list(elemToSpecD.items())

specToElemD = {val:key for key, val in elemToSpecD.items()}
specToElem = list(specToElemD.items())

Matrix(elemToSpec)
```

[11]:

$$\begin{bmatrix} n_{11} & w_{11}x_{11} + w_{21}x_{12} + w_{31}x_{13} \\ n_{12} & w_{12}x_{11} + w_{22}x_{12} + w_{32}x_{13} \\ n_{21} & w_{11}x_{21} + w_{21}x_{22} + w_{31}x_{23} \\ n_{22} & w_{12}x_{21} + w_{22}x_{22} + w_{32}x_{23} \\ n_{31} & w_{11}x_{31} + w_{21}x_{32} + w_{31}x_{33} \\ n_{32} & w_{12}x_{31} + w_{22}x_{32} + w_{32}x_{33} \end{bmatrix}$$

[12]:
```
elemToSpecFuncD = dict(itertools.chain(*[[(Nelem[i, j],
    Function("n_{}{}".format(i + 1, j + 1))(Nspec[i, j])) for j in
    range(p)] for i in range(n)]))

elemToSpecFunc = list(elemToSpecFuncD.items())

specFuncToElemD = {val : key for key , val in elemToSpecFuncD.items()}
specFuncToElem = list(specFuncToElemD.items())

Matrix(elemToSpecFunc)
```

[12]:

$$\begin{bmatrix} n_{11} & \mathrm{n}_{11}\left(w_{11}x_{11} + w_{21}x_{12} + w_{31}x_{13}\right) \\ n_{12} & \mathrm{n}_{12}\left(w_{12}x_{11} + w_{22}x_{12} + w_{32}x_{13}\right) \\ n_{21} & \mathrm{n}_{21}\left(w_{11}x_{21} + w_{21}x_{22} + w_{31}x_{23}\right) \\ n_{22} & \mathrm{n}_{22}\left(w_{12}x_{21} + w_{22}x_{22} + w_{32}x_{23}\right) \\ n_{31} & \mathrm{n}_{31}\left(w_{11}x_{31} + w_{21}x_{32} + w_{31}x_{33}\right) \\ n_{32} & \mathrm{n}_{32}\left(w_{12}x_{31} + w_{22}x_{32} + w_{32}x_{33}\right) \end{bmatrix}$$

[13]:
```python
elemToNFuncD = dict(itertools.chain(*[[(Nelem[i, j],
 ↪Function("n_{}{}".format(i + 1, j + 1))(*X,*W)) for j in range(p)]
 ↪for i in range(n)]))

elemToNFunc = list(elemToNFuncD.items())

nfuncToElemD = {val: key for key, val in elemToNFuncD.items()}
nfuncToElem = list(nfuncToElemD.items())

Matrix(elemToNFunc)
```

[13]:

$$\left[\begin{matrix}n_{11} & \operatorname{n_{11}}{\left(x_{11},x_{12},x_{13},x_{21},x_{22},x_{23},x_{31},x_{32},x_{33},w_{11},w_{12},w_{21},w_{22},w_{31},w_{32} \right)}\\n_{12} & \operatorname{n_{12}}{\left(x_{11},x_{12},x_{13},x_{21},x_{22},x_{23},x_{31},x_{32},x_{33},w_{11},w_{12},w_{21},w_{22},w_{31},w_{32} \right)}\\n_{21} & \operatorname{n_{21}}{\left(x_{11},x_{12},x_{13},x_{21},x_{22},x_{23},x_{31},x_{32},x_{33},w_{11},w_{12},w_{21},w_{22},w_{31},w_{32} \right)}\\n_{22} & \operatorname{n_{22}}{\left(x_{11},x_{12},x_{13},x_{21},x_{22},x_{23},x_{31},x_{32},x_{33},w_{11},w_{12},w_{21},w_{22},w_{31},w_{32} \right)}\\n_{31} & \operatorname{n_{31}}{\left(x_{11},x_{12},x_{13},x_{21},x_{22},x_{23},x_{31},x_{32},x_{33},w_{11},w_{12},w_{21},w_{22},w_{31},w_{32} \right)}\\n_{32} & \operatorname{n_{32}}{\left(x_{11},x_{12},x_{13},x_{21},x_{22},x_{23},x_{31},x_{32},x_{33},w_{11},w_{12},w_{21},w_{22},w_{31},w_{32} \right)}\end{matrix}\right]$$

[14]:
```python
elemToNmatfuncD = dict(itertools.chain(*[[(Nelem[i, j],
 ↪Function("n_{}{}".format(i+1,j+1))(A,B) ) for j in range(p)] for i
 ↪in range(n)]))

elemToNmatfunc = list(elemToNmatfuncD.items())

nmatfuncToElemD = {val: key for key, val in elemToNmatfuncD.items()}
nmatfuncToElem = list(nmatfuncToElemD.items())

Matrix(elemToNmatfunc)
```

[14]:

$$\left[\begin{matrix}n_{11} & \operatorname{n_{11}}{\left(X,W \right)}\\n_{12} & \operatorname{n_{12}}{\left(X,W \right)}\\n_{21} & \operatorname{n_{21}}{\left(X,W \right)}\\n_{22} & \operatorname{n_{22}}{\left(X,W \right)}\\n_{31} & \operatorname{n_{31}}{\left(X,W \right)}\\n_{32} & \operatorname{n_{32}}{\left(X,W \right)}\end{matrix}\right]$$

[15]:
```python
nmatfuncToSpecD = dict(zip(elemToNmatfuncD.values(), elemToSpecD.
 ↪values()))

nmatfuncToSpec = list(nmatfuncToSpecD.items())
```

```
Matrix(nmatfuncToSpec)
```

[15]:

$$\left[\begin{matrix}\mathfrak{n}_{11}(X,W) & w_{11}x_{11}+w_{21}x_{12}+w_{31}x_{13}\\ \mathfrak{n}_{12}(X,W) & w_{12}x_{11}+w_{22}x_{12}+w_{32}x_{13}\\ \mathfrak{n}_{21}(X,W) & w_{11}x_{21}+w_{21}x_{22}+w_{31}x_{23}\\ \mathfrak{n}_{22}(X,W) & w_{12}x_{21}+w_{22}x_{22}+w_{32}x_{23}\\ \mathfrak{n}_{31}(X,W) & w_{11}x_{31}+w_{21}x_{32}+w_{31}x_{33}\\ \mathfrak{n}_{32}(X,W) & w_{12}x_{31}+w_{22}x_{32}+w_{32}x_{33}\end{matrix}\right]$$

[16]:

```
# Overall abstract
dL_dX_overallAbstract = compose(lambd, sigmaApply)(VL).diff(A).
 ↪replace(VL, v(A, B))

dL_dW_overallAbstract = compose(lambd, sigmaApply)(VL).diff(B).
 ↪replace(VL, v(A, B))

showGroup([
    dL_dX_overallAbstract,
    dL_dW_overallAbstract
])
```

$$\left.\frac{d}{d\xi_1}\sigma_{apply}(\xi_1)\right|_{\xi_1=\nu(X,W)}\frac{\partial}{\partial\sigma_{apply}(\nu(X,W))}\lambda(\sigma_{apply}(\nu(X,W)))\frac{\partial}{\partial X}\nu(X,W)$$

$$\left.\frac{d}{d\xi_1}\sigma_{apply}(\xi_1)\right|_{\xi_1=\nu(X,W)}\frac{\partial}{\partial\sigma_{apply}(\nu(X,W))}\lambda(\sigma_{apply}(\nu(X,W)))\frac{\partial}{\partial W}\nu(X,W)$$

[17]:

```
dL_dW_abstract = compose(lambd, sigmaApply, v)(A, B).replace(v, v_).
 ↪replace(sigmaApply, sigmaApply_).diff(B)
#L.replace(v,v_).replace(sigmaApply, sigmaApply_).diff(B)

showGroup([
    dL_dW_abstract,
    dL_dW_abstract.subs({lambd : lambd_L})
])
```

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=(d\mapsto\sigma(d))_{\circ}(XW)}X^T\left(d\mapsto\frac{d}{dd}\sigma(d)\right)_{\circ}(XW)$$

$$\left.\frac{d}{d\xi_1}\left(\xi_{10,0}+\xi_{10,1}+\xi_{11,0}+\xi_{11,1}+\xi_{12,0}+\xi_{12,1}\right)\right|_{\xi_1=(d\mapsto\sigma(d))_{\circ}(XW)}X^T\left(d\mapsto\frac{d}{dd}\sigma(d)\right)_{\circ}(XW)$$

[18]:
```
dL_dX_abstract = compose(lambd, sigmaApply, v)(A, B).replace(v, v_).
 ↪replace(sigmaApply, sigmaApply_).diff(A)
#L.replace(v, v_).replace(sigmaApply, sigmaApply_).diff(A)

dL_dX_abstract
```

[18]:
$$\frac{d}{d\xi_1}\lambda(\xi_1)\Big|_{\xi_1=(d\mapsto\sigma(d))_\circ(XW)}\left(d\mapsto\frac{d}{dd}\sigma(d)\right)_\circ(XW)\,W^T$$

[19]:
```
dL_dW_direct = L.replace(v, vN).replace(sigmaApply, sigmaApply_).
 ↪replace(lambd, lambd_).subs(elemToSpecD).diff(W).subs(specToElemD)

dL_dW_direct = dL_dW_direct.doit()

dL_dW_direct
```

[19]:
$$\begin{bmatrix} x_{11}\frac{d}{dn_{11}}\sigma(n_{11})+x_{21}\frac{d}{dn_{21}}\sigma(n_{21})+x_{31}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{11}\frac{d}{dn_{12}}\sigma(n_{12})+x_{21}\frac{d}{dn_{22}}\sigma(n_{22})+x_{31}\frac{d}{dn_{32}}\sigma(n_{32}) \\ x_{12}\frac{d}{dn_{11}}\sigma(n_{11})+x_{22}\frac{d}{dn_{21}}\sigma(n_{21})+x_{32}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{12}\frac{d}{dn_{12}}\sigma(n_{12})+x_{22}\frac{d}{dn_{22}}\sigma(n_{22})+x_{32}\frac{d}{dn_{32}}\sigma(n_{32}) \\ x_{13}\frac{d}{dn_{11}}\sigma(n_{11})+x_{23}\frac{d}{dn_{21}}\sigma(n_{21})+x_{33}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{13}\frac{d}{dn_{12}}\sigma(n_{12})+x_{23}\frac{d}{dn_{22}}\sigma(n_{22})+x_{33}\frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

[20]:
```
dL_dX_direct = L.replace(v, vN).replace(sigmaApply, sigmaApply_).
 ↪replace(lambd, lambd_).subs(elemToSpecD).diff(X).subs(specToElemD)

dL_dX_direct = dL_dX_direct.doit()

dL_dX_direct
```

[20]:
$$\begin{bmatrix} w_{11}\frac{d}{dn_{11}}\sigma(n_{11})+w_{12}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{21}\frac{d}{dn_{11}}\sigma(n_{11})+w_{22}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{31}\frac{d}{dn_{11}}\sigma(n_{11})+w_{32}\frac{d}{dn_{12}}\sigma(n_{12}) \\ w_{11}\frac{d}{dn_{21}}\sigma(n_{21})+w_{12}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{21}\frac{d}{dn_{21}}\sigma(n_{21})+w_{22}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{31}\frac{d}{dn_{21}}\sigma(n_{21})+w_{32}\frac{d}{dn_{22}}\sigma(n_{22}) \\ w_{11}\frac{d}{dn_{31}}\sigma(n_{31})+w_{12}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{21}\frac{d}{dn_{31}}\sigma(n_{31})+w_{22}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{31}\frac{d}{dn_{31}}\sigma(n_{31})+w_{32}\frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

[21]:
```
unapplied = sigmaApply_L(vN(A,B))
unapplied
# Also works: same as above:
#compose(sigmaApply, v)(A,B).replace(v, vN).replace(sigmaApply ,
 ↪sigmaApply_L)
```

[21]:
$$(d\mapsto\sigma(d))_\circ\left(\begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \\ n_{31} & n_{32} \end{bmatrix}\right)$$

[22]:
```
applied = unapplied.doit()
applied
```

[22]:

$$\begin{bmatrix}\sigma(n_{11}) & \sigma(n_{12})\\ \sigma(n_{21}) & \sigma(n_{22})\\ \sigma(n_{31}) & \sigma(n_{32})\end{bmatrix}$$

[23]:
```
dL_dW_step = compose(lambd, sigmaApply, v)(A,B).replace(v, v_).
↪replace(sigmaApply, sigmaApply_).diff(B).subs({A*B : vN(A,B)}).
↪doit()

showGroup([
    dL_dW_step,
    dL_dW_step.replace(unapplied, applied),
    # Carrying out the multplication:
    dL_dW_step.subs({A:X}).doit(), # replace won't work here
    dL_dW_step.subs({A:X}).doit().replace(unapplied, applied)
])
```

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=(d\mapsto\sigma(d))_\circ\left(\begin{bmatrix}n_{11} & n_{12}\\ n_{21} & n_{22}\\ n_{31} & n_{32}\end{bmatrix}\right)} X^T \begin{bmatrix}\frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12})\\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22})\\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32})\end{bmatrix}$$

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=\begin{bmatrix}\sigma(n_{11}) & \sigma(n_{12})\\ \sigma(n_{21}) & \sigma(n_{22})\\ \sigma(n_{31}) & \sigma(n_{32})\end{bmatrix}} X^T \begin{bmatrix}\frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12})\\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22})\\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32})\end{bmatrix}$$

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=(d\mapsto\sigma(d))_\circ\left(\begin{bmatrix}n_{11} & n_{12}\\ n_{21} & n_{22}\\ n_{31} & n_{32}\end{bmatrix}\right)} \begin{bmatrix}x_{11}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{21}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{31}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{11}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{21}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{31}\frac{d}{dn_{32}}\sigma(n_{32})\\ x_{12}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{22}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{32}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{12}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{22}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{32}\frac{d}{dn_{32}}\sigma(n_{32})\\ x_{13}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{23}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{33}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{13}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{23}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{33}\frac{d}{dn_{32}}\sigma(n_{32})\end{bmatrix}$$

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=\begin{bmatrix}\sigma(n_{11}) & \sigma(n_{12})\\ \sigma(n_{21}) & \sigma(n_{22})\\ \sigma(n_{31}) & \sigma(n_{32})\end{bmatrix}} \begin{bmatrix}x_{11}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{21}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{31}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{11}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{21}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{31}\frac{d}{dn_{32}}\sigma(n_{32})\\ x_{12}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{22}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{32}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{12}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{22}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{32}\frac{d}{dn_{32}}\sigma(n_{32})\\ x_{13}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{23}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{33}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{13}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{23}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{33}\frac{d}{dn_{32}}\sigma(n_{32})\end{bmatrix}$$

[24]:
```
dL_dX_step = compose(lambd, sigmaApply, v)(A,B).replace(v, v_).
 ↪replace(sigmaApply, sigmaApply_).diff(A).subs({A*B : vN(A,B)}).
 ↪doit()


showGroup([
    dL_dX_step,
    dL_dX_step.replace(unapplied, applied),
    dL_dX_step.subs({B:W}).doit(),
    dL_dX_step.subs({B:W}).doit().replace(unapplied, applied)
])
```

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=(d\mapsto\sigma(d))_\circ\left(\begin{bmatrix}n_{11} & n_{12}\\ n_{21} & n_{22}\\ n_{31} & n_{32}\end{bmatrix}\right)}\begin{bmatrix}\frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12})\\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22})\\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32})\end{bmatrix}W^T$$

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=\begin{bmatrix}\sigma(n_{11}) & \sigma(n_{12})\\ \sigma(n_{21}) & \sigma(n_{22})\\ \sigma(n_{31}) & \sigma(n_{32})\end{bmatrix}}\begin{bmatrix}\frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12})\\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22})\\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32})\end{bmatrix}W^T$$

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=(d\mapsto\sigma(d))_\circ\left(\begin{bmatrix}n_{11} & n_{12}\\ n_{21} & n_{22}\\ n_{31} & n_{32}\end{bmatrix}\right)}\begin{bmatrix}w_{11}\frac{d}{dn_{11}}\sigma(n_{11})+w_{12}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{21}\frac{d}{dn_{11}}\sigma(n_{11})+w_{22}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{31}\frac{d}{dn_{11}}\sigma(n_{11})+w_{32}\frac{d}{dn_{12}}\sigma(n_{12})\\ w_{11}\frac{d}{dn_{21}}\sigma(n_{21})+w_{12}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{21}\frac{d}{dn_{21}}\sigma(n_{21})+w_{22}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{31}\frac{d}{dn_{21}}\sigma(n_{21})+w_{32}\frac{d}{dn_{22}}\sigma(n_{22})\\ w_{11}\frac{d}{dn_{31}}\sigma(n_{31})+w_{12}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{21}\frac{d}{dn_{31}}\sigma(n_{31})+w_{22}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{31}\frac{d}{dn_{31}}\sigma(n_{31})+w_{32}\frac{d}{dn_{32}}\sigma(n_{32})\end{bmatrix}$$

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=\begin{bmatrix}\sigma(n_{11}) & \sigma(n_{12})\\ \sigma(n_{21}) & \sigma(n_{22})\\ \sigma(n_{31}) & \sigma(n_{32})\end{bmatrix}}\begin{bmatrix}w_{11}\frac{d}{dn_{11}}\sigma(n_{11})+w_{12}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{21}\frac{d}{dn_{11}}\sigma(n_{11})+w_{22}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{31}\frac{d}{dn_{11}}\sigma(n_{11})+w_{32}\frac{d}{dn_{12}}\sigma(n_{12})\\ w_{11}\frac{d}{dn_{21}}\sigma(n_{21})+w_{12}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{21}\frac{d}{dn_{21}}\sigma(n_{21})+w_{22}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{31}\frac{d}{dn_{21}}\sigma(n_{21})+w_{32}\frac{d}{dn_{22}}\sigma(n_{22})\\ w_{11}\frac{d}{dn_{31}}\sigma(n_{31})+w_{12}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{21}\frac{d}{dn_{31}}\sigma(n_{31})+w_{22}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{31}\frac{d}{dn_{31}}\sigma(n_{31})+w_{32}\frac{d}{dn_{32}}\sigma(n_{32})\end{bmatrix}$$

Trying to replace further to get the ones matrix for the deriv of lambda expression, but doesn't work, see code below for why (hadamard is not present, just matrix multiplication. Chain rule in this form doesn't know there should be hadamard product between deriv of λ expression and $\frac{dS}{dX}$ expression)

[25]:
```
dle = lambd(xi).diff(xi)
```

```
dle_repl = lambd(xi).diff(xi).subs(xi, applied).replace(lambd,
 ↪lambd_L)

showGroup([
    dle,
    dle_repl
])
```

$$\frac{d}{d\xi}\lambda(\xi)$$

$$\frac{\partial}{\partial \begin{bmatrix} \sigma(n_{11}) & \sigma(n_{12}) \\ \sigma(n_{21}) & \sigma(n_{22}) \\ \sigma(n_{31}) & \sigma(n_{32}) \end{bmatrix}} \left(\sigma(n_{11}) + \sigma(n_{12}) + \sigma(n_{21}) + \sigma(n_{22}) + \sigma(n_{31}) + \sigma(n_{32})\right)$$

[26]:
```
showGroup([
    dL_dW_abstract.replace(sigmaApply_L(A*B), xi),
    dL_dW_abstract.replace(sigmaApply_L(A*B), xi).doit(),
    dL_dW_abstract.replace(sigmaApply_L(A*B), xi).doit().replace(dle,
 ↪dle_repl) #.doit())
])

# NOTE here it says the matrices are not aligned if we execute doit()
 ↪to reveal the ones matrix that is dL_dS. True since assumption
 ↪here is matrix multplication with dL_dS and right hand side, but
 ↪in fact it is hadamard multiplication.
```

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=\xi} X^T \left(d \mapsto \frac{d}{dd}\sigma(d)\right)_\circ (XW)$$

$$\left(\frac{d}{d\xi}\lambda(\xi)\right) X^T \left(d \mapsto \frac{d}{dd}\sigma(d)\right)_\circ (XW)$$

$$\left(\frac{\partial}{\partial \begin{bmatrix} \sigma(n_{11}) & \sigma(n_{12}) \\ \sigma(n_{21}) & \sigma(n_{22}) \\ \sigma(n_{31}) & \sigma(n_{32}) \end{bmatrix}} \left(\sigma(n_{11}) + \sigma(n_{12}) + \sigma(n_{21}) + \sigma(n_{22}) + \sigma(n_{31}) + \sigma(n_{32})\right)\right) X^T \left(d \mapsto \frac{d}{dd}\sigma(d)\right)_\circ (XW)$$

The first part: $\frac{dL}{dS}$

Direct substitution way:

[27]:
```
showGroup([
    lambd(xi).diff(xi).subs(xi, applied),
    lambd(xi).diff(xi).subs(xi, applied).replace(lambd, lambd_L),
    lambd(xi).diff(xi).subs(xi, applied).replace(lambd, lambd_L).
 ↪doit()
])
```

$$\frac{\partial}{\partial \begin{bmatrix} \sigma(n_{11}) & \sigma(n_{12}) \\ \sigma(n_{21}) & \sigma(n_{22}) \\ \sigma(n_{31}) & \sigma(n_{32}) \end{bmatrix}} \lambda\left(\begin{bmatrix} \sigma(n_{11}) & \sigma(n_{12}) \\ \sigma(n_{21}) & \sigma(n_{22}) \\ \sigma(n_{31}) & \sigma(n_{32}) \end{bmatrix}\right)$$

$$\frac{\partial}{\partial \begin{bmatrix} \sigma(n_{11}) & \sigma(n_{12}) \\ \sigma(n_{21}) & \sigma(n_{22}) \\ \sigma(n_{31}) & \sigma(n_{32}) \end{bmatrix}} (\sigma(n_{11}) + \sigma(n_{12}) + \sigma(n_{21}) + \sigma(n_{22}) + \sigma(n_{31}) + \sigma(n_{32}))$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

The substitute into derivative way:

[28]:
```
showGroup([
    lambd(xi).diff(xi).subs(xi, unapplied),
    lambd(xi).diff(xi).subs(xi, unapplied).replace(unapplied,
 ↪applied),
    # gives same expression as in dldx
    lambd(xi).diff(xi).subs(xi, unapplied).replace(unapplied,
 ↪applied).replace(lambd, lambd_L)
])
```

$$\left.\frac{d}{d\xi}\lambda(\xi)\right|_{\xi=(d\mapsto\sigma(d))_{\circ}\left(\begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \\ n_{31} & n_{32} \end{bmatrix}\right)}$$

$$\left.\frac{d}{d\xi}\lambda(\xi)\right|_{\xi=\begin{bmatrix} \sigma(n_{11}) & \sigma(n_{12}) \\ \sigma(n_{21}) & \sigma(n_{22}) \\ \sigma(n_{31}) & \sigma(n_{32}) \end{bmatrix}}$$

$$\frac{d}{d\xi}\left(\xi_{0,0}+\xi_{0,1}+\xi_{1,0}+\xi_{1,1}+\xi_{2,0}+\xi_{2,1}\right)\Big|_{\xi=\begin{bmatrix}\sigma(n_{11}) & \sigma(n_{12})\\ \sigma(n_{21}) & \sigma(n_{22})\\ \sigma(n_{31}) & \sigma(n_{32})\end{bmatrix}}$$

The second part: $\frac{\partial N}{\partial X}\times\frac{\partial S}{\partial N}$

[29]:
```
dN_dW_times_dS_dN = compose(sigmaApply, v)(A,B).replace(v, v_).
 ↪replace(sigmaApply, sigmaApply_).diff(B).subs({A*B : vN(A,B)}).
 ↪doit()

showGroup([
    dN_dW_times_dS_dN,
    dN_dW_times_dS_dN.subs({A:X}), # replace won't work here
    # Carrying out the multplication:
    dN_dW_times_dS_dN.subs({A:X}).doit() # replace won't work here
])
```

$$X^T\begin{bmatrix}\frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12})\\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22})\\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32})\end{bmatrix}$$

$$\left(\begin{bmatrix}x_{11} & x_{12} & x_{13}\\ x_{21} & x_{22} & x_{23}\\ x_{31} & x_{32} & x_{33}\end{bmatrix}\right)^T\begin{bmatrix}\frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12})\\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22})\\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32})\end{bmatrix}$$

$$\begin{bmatrix}x_{11}\frac{d}{dn_{11}}\sigma(n_{11})+x_{21}\frac{d}{dn_{21}}\sigma(n_{21})+x_{31}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{11}\frac{d}{dn_{12}}\sigma(n_{12})+x_{21}\frac{d}{dn_{22}}\sigma(n_{22})+x_{31}\frac{d}{dn_{32}}\sigma(n_{32})\\ x_{12}\frac{d}{dn_{11}}\sigma(n_{11})+x_{22}\frac{d}{dn_{21}}\sigma(n_{21})+x_{32}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{12}\frac{d}{dn_{12}}\sigma(n_{12})+x_{22}\frac{d}{dn_{22}}\sigma(n_{22})+x_{32}\frac{d}{dn_{32}}\sigma(n_{32})\\ x_{13}\frac{d}{dn_{11}}\sigma(n_{11})+x_{23}\frac{d}{dn_{21}}\sigma(n_{21})+x_{33}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{13}\frac{d}{dn_{12}}\sigma(n_{12})+x_{23}\frac{d}{dn_{22}}\sigma(n_{22})+x_{33}\frac{d}{dn_{32}}\sigma(n_{32})\end{bmatrix}$$

[30]:
```
dN_dX_times_dS_dN = compose(sigmaApply, v)(A,B).replace(v, v_).
 ↪replace(sigmaApply, sigmaApply_).diff(A).subs({A*B : vN(A,B)}).
 ↪doit()

showGroup([
    dN_dX_times_dS_dN,
    dN_dX_times_dS_dN.subs({B:W}), # replace won't work here
    # Carrying out the multplication:
    dN_dX_times_dS_dN.subs({B:W}).doit() # replace won't work here
])
```

$$\begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix} W^T$$

$$\begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix} \left(\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix} \right)^T$$

$$\begin{bmatrix} w_{11}\frac{d}{dn_{11}}\sigma(n_{11}) + w_{12}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{21}\frac{d}{dn_{11}}\sigma(n_{11}) + w_{22}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{31}\frac{d}{dn_{11}}\sigma(n_{11}) + w_{32}\frac{d}{dn_{12}}\sigma(n_{12}) \\ w_{11}\frac{d}{dn_{21}}\sigma(n_{21}) + w_{12}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{21}\frac{d}{dn_{21}}\sigma(n_{21}) + w_{22}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{31}\frac{d}{dn_{21}}\sigma(n_{21}) + w_{32}\frac{d}{dn_{22}}\sigma(n_{22}) \\ w_{11}\frac{d}{dn_{31}}\sigma(n_{31}) + w_{12}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{21}\frac{d}{dn_{31}}\sigma(n_{31}) + w_{22}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{31}\frac{d}{dn_{31}}\sigma(n_{31}) + w_{32}\frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

[31]:
```
# THis seems right:
dL_dS = lambd(Selem).replace(lambd, lambd_L).diff(Selem)
# ANOTHER WAY: lambd(xi).diff(xi).subs(xi, applied).replace(lambd,
 ↪lambd_L).doit()

# THIS SEEMS WRONG : ??? how to tell for sure?
#lambd(Selem).diff(Selem).replace(lambd, lambd_L).doit()

dL_dS
```

[31]:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

[32]:
```
dS_dN = compose(sigmaApply)(M).replace(sigmaApply, sigmaApply_).
 ↪diff(M).subs({M : vN(A,B)}).doit()

dS_dN_abstract = compose(sigmaApply)(M).replace(sigmaApply,
 ↪sigmaApply_).diff(M).subs(M, v_(A,B))
# ANOTHER WAY: sigmaApply_L(M).diff(M).subs({M : Nelem}).doit()
# WRONG:
#dS_dN = sigmaApply(Nelem).replace(sigmaApply, sigmaApply_).
 ↪diff(Matrix(Nelem))
showGroup([
    dS_dN,
    dS_dN_abstract
])
```

$$\begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

$$\left(d \mapsto \frac{d}{dd}\sigma(d)\right)_{\circ}(XW)$$

$$\begin{aligned} \frac{\partial L}{\partial W} &= \frac{\partial L}{\partial S} \odot \left(\frac{\partial N}{\partial W} \times \frac{\partial S}{\partial N}\right) \\ &= \frac{\partial L}{\partial S} \odot \left(X^T \times \frac{\partial S}{\partial N}\right) \end{aligned}$$

where $\odot$ signifies the Hadamard product and $\times$ is matrix multiplication.

[33]:
```
from sympy import HadamardProduct

dN_dW = A.transpose()

dS_dW = dN_dW * dS_dN
dS_dW_abstract = compose(sigmaApply, v)(A,B).replace(v, v_).
 ↪replace(sigmaApply, sigmaApply_).diff(B)

dL_dW = HadamardProduct(dL_dS, dS_dW)
dL_dW_hadamard = dL_dW.subs(A,X).doit()

assert dL_dW == HadamardProduct(dL_dS, dN_dW * dS_dN )

showGroup([
    dS_dW,
    dS_dW_abstract,
    dS_dW.subs(A, X).doit(),
    dL_dW,
    dL_dW_hadamard
])
```

$$X^T \begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

$$X^T\left(d \mapsto \frac{d}{dd}\sigma(d)\right)_{\circ}(XW)$$

$$\begin{bmatrix} x_{11}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{21}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{31}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{11}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{21}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{31}\frac{d}{dn_{32}}\sigma(n_{32}) \\ x_{12}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{22}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{32}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{12}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{22}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{32}\frac{d}{dn_{32}}\sigma(n_{32}) \\ x_{13}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{23}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{33}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{13}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{23}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{33}\frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \circ \left(X^T \begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix} \right)$$

$$\begin{bmatrix} x_{11}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{21}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{31}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{11}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{21}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{31}\frac{d}{dn_{32}}\sigma(n_{32}) \\ x_{12}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{22}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{32}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{12}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{22}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{32}\frac{d}{dn_{32}}\sigma(n_{32}) \\ x_{13}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{23}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{33}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{13}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{23}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{33}\frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

$$\begin{aligned} \frac{\partial L}{\partial X} &= \left(\frac{\partial L}{\partial S} \odot \frac{\partial S}{\partial N}\right) \times \frac{\partial N}{\partial X} \\ &= \left(\frac{\partial L}{\partial S} \odot \frac{\partial S}{\partial N}\right) \times W^T \end{aligned}$$

where $\odot$ signifies the Hadamard product and $\times$ is matrix multiplication.

[34]:
```
dN_dX = B.transpose()

dS_dX = dS_dN * dN_dX
dS_dX_abstract = compose(sigmaApply, v)(A,B).replace(v, v_).
 ↪replace(sigmaApply, sigmaApply_).diff(A)

dL_dN = HadamardProduct(dL_dS, dS_dN)

dL_dX = dL_dN * dN_dX #).subs(B, W).doit()
dL_dX_hadamard = dL_dX.subs(B, W).doit()

assert dL_dX == HadamardProduct(dL_dS, dS_dN) * dN_dX

showGroup([
    dS_dX,
    dS_dX.subs(B, W),
    dS_dX_abstract,
    dL_dN,
#    dS_dX.subs(B, W).doit(),
    dL_dX,
    dL_dX_hadamard
])
```

$$\begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix} W^T$$

$$\begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix} \left(\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix} \right)^T$$

$$\left(d \mapsto \frac{d}{dd}\sigma(d) \right)_\circ (XW)\, W^T$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \circ \begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

$$\begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix} W^T$$

$$\begin{bmatrix} w_{11}\frac{d}{dn_{11}}\sigma(n_{11}) + w_{12}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{21}\frac{d}{dn_{11}}\sigma(n_{11}) + w_{22}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{31}\frac{d}{dn_{11}}\sigma(n_{11}) + w_{32}\frac{d}{dn_{12}}\sigma(n_{12}) \\ w_{11}\frac{d}{dn_{21}}\sigma(n_{21}) + w_{12}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{21}\frac{d}{dn_{21}}\sigma(n_{21}) + w_{22}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{31}\frac{d}{dn_{21}}\sigma(n_{21}) + w_{32}\frac{d}{dn_{22}}\sigma(n_{22}) \\ w_{11}\frac{d}{dn_{31}}\sigma(n_{31}) + w_{12}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{21}\frac{d}{dn_{31}}\sigma(n_{31}) + w_{22}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{31}\frac{d}{dn_{31}}\sigma(n_{31}) + w_{32}\frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

[35]:
```
showGroup([
    dL_dX_abstract,
    dL_dX_step,
    dL_dX,
    dL_dX_hadamard
])
```

$$\frac{d}{d\xi_1}\lambda(\xi_1)\Big|_{\xi_1=(d\mapsto\sigma(d))_\circ(XW)} \left(d \mapsto \frac{d}{dd}\sigma(d) \right)_\circ (XW)\, W^T$$

$$\frac{d}{d\xi_1}\lambda(\xi_1)\Big|_{\xi_1=(d\mapsto\sigma(d))_\circ\left(\begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \\ n_{31} & n_{32} \end{bmatrix}\right)} \begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix} W^T$$

$$\begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix} W^T$$

$$\begin{bmatrix} w_{11}\frac{d}{dn_{11}}\sigma(n_{11}) + w_{12}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{21}\frac{d}{dn_{11}}\sigma(n_{11}) + w_{22}\frac{d}{dn_{12}}\sigma(n_{12}) & w_{31}\frac{d}{dn_{11}}\sigma(n_{11}) + w_{32}\frac{d}{dn_{12}}\sigma(n_{12}) \\ w_{11}\frac{d}{dn_{21}}\sigma(n_{21}) + w_{12}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{21}\frac{d}{dn_{21}}\sigma(n_{21}) + w_{22}\frac{d}{dn_{22}}\sigma(n_{22}) & w_{31}\frac{d}{dn_{21}}\sigma(n_{21}) + w_{32}\frac{d}{dn_{22}}\sigma(n_{22}) \\ w_{11}\frac{d}{dn_{31}}\sigma(n_{31}) + w_{12}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{21}\frac{d}{dn_{31}}\sigma(n_{31}) + w_{22}\frac{d}{dn_{32}}\sigma(n_{32}) & w_{31}\frac{d}{dn_{31}}\sigma(n_{31}) + w_{32}\frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

[36]:
```
showGroup([
    dL_dW_abstract,
    dL_dW_step,
    dL_dW,
    dL_dW_hadamard
])
```

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=(d\mapsto\sigma(d))_\circ(XW)} \quad X^T\left(d \mapsto \frac{d}{dd}\sigma(d)\right)_\circ(XW)$$

$$\left.\frac{d}{d\xi_1}\lambda(\xi_1)\right|_{\xi_1=(d\mapsto\sigma(d))_\circ\left(\begin{bmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \\ n_{31} & n_{32} \end{bmatrix}\right)} \quad X^T\begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \circ \left(X^T\begin{bmatrix} \frac{d}{dn_{11}}\sigma(n_{11}) & \frac{d}{dn_{12}}\sigma(n_{12}) \\ \frac{d}{dn_{21}}\sigma(n_{21}) & \frac{d}{dn_{22}}\sigma(n_{22}) \\ \frac{d}{dn_{31}}\sigma(n_{31}) & \frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}\right)$$

$$\begin{bmatrix} x_{11}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{21}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{31}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{11}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{21}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{31}\frac{d}{dn_{32}}\sigma(n_{32}) \\ x_{12}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{22}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{32}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{12}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{22}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{32}\frac{d}{dn_{32}}\sigma(n_{32}) \\ x_{13}\frac{d}{dn_{11}}\sigma(n_{11}) + x_{23}\frac{d}{dn_{21}}\sigma(n_{21}) + x_{33}\frac{d}{dn_{31}}\sigma(n_{31}) & x_{13}\frac{d}{dn_{12}}\sigma(n_{12}) + x_{23}\frac{d}{dn_{22}}\sigma(n_{22}) + x_{33}\frac{d}{dn_{32}}\sigma(n_{32}) \end{bmatrix}$$

[37]:
```
compose(lambd, sigmaApply, v)(A,B).replace(lambd, lambd_L)
```

[37]:

$$(\sigma_{apply}(\nu(X,W)))_{0,0} + (\sigma_{apply}(\nu(X,W)))_{0,1} + (\sigma_{apply}(\nu(X,W)))_{1,0} + (\sigma_{apply}(\nu(X,W)))_{1,1} + (\sigma_{apply}(\nu(X,W)))_{2,0} + (\sigma_{apply}(\nu(X,W)))_{2,1}$$

[38]:
```
compose(lambd, sigmaApply, v)(A,B).replace(v,v_).subs({lambd:
  ↪lambd_L})#.subs({sigmaApply : sigmaApply_L})
```

[38]:

$$(\sigma_{apply}(XW))_{0,0} + (\sigma_{apply}(XW))_{0,1} + (\sigma_{apply}(XW))_{1,0} + (\sigma_{apply}(XW))_{1,1} + (\sigma_{apply}(XW))_{2,0} + (\sigma_{apply}(XW))_{2,1}$$

[39]:
```
compose(lambd, sigmaApply, v)(A,B).replace(v,v_).replace(sigmaApply,
  ↪sigmaApply_).replace(lambd, lambd_L)
```

[39]: $(d \mapsto \sigma(d))_\circ\,(XW)_{0,0} + (d \mapsto \sigma(d))_\circ\,(XW)_{0,1} + (d \mapsto \sigma(d))_\circ\,(XW)_{1,0} + (d \mapsto \sigma(d))_\circ\,(XW)_{1,1} + (d \mapsto \sigma(d))_\circ\,(XW)_{2,0} + (d \mapsto \sigma(d))_\circ\,(XW)_{2,1}$

[40]:
```
compose(lambd, sigmaApply, v)(A,B).replace(lambd, lambd_L).replace(v,
  ↪v_).replace(sigmaApply, sigmaApply_)
```

[40]: $(d \mapsto \sigma(d))_\circ\,(XW)_{0,0} + (d \mapsto \sigma(d))_\circ\,(XW)_{0,1} + (d \mapsto \sigma(d))_\circ\,(XW)_{1,0} + (d \mapsto \sigma(d))_\circ\,(XW)_{1,1} + (d \mapsto \sigma(d))_\circ\,(XW)_{2,0} + (d \mapsto \sigma(d))_\circ\,(XW)_{2,1}$

[41]:
```
compose(lambd, sigmaApply, v)(A,B).replace(v,v_).replace(sigmaApply,
  ↪sigmaApply_).replace(lambd, lambd_L).doit()
```

[41]: $\sigma(W_{0,0}X_{0,0} + W_{1,0}X_{0,1} + W_{2,0}X_{0,2}) + \sigma(W_{0,0}X_{1,0} + W_{1,0}X_{1,1} + W_{2,0}X_{1,2}) + \sigma(W_{0,0}X_{2,0} + W_{1,0}X_{2,1} + W_{2,0}X_{2,2}) + \sigma(W_{0,1}X_{0,0} + W_{1,1}X_{0,1} + W_{2,1}X_{0,2}) + \sigma(W_{0,1}X_{1,0} + W_{1,1}X_{1,1} + W_{2,1}X_{1,2}) + \sigma(W_{0,1}X_{2,0} + W_{1,1}X_{2,1} + W_{2,1}X_{2,2})$

[42]:
```
# Alternative to the above: using the lower case matrix element names
  ↪rather than upper case (from MatrixSymbol)
compose(lambd, sigmaApply, v)(A, B).replace(v, vN).
  ↪replace(sigmaApply, sigmaApply_).replace(lambd, lambd_).
  ↪subs(elemToSpecD)
```

[42]: $\sigma(w_{11}x_{11} + w_{21}x_{12} + w_{31}x_{13}) + \sigma(w_{11}x_{21} + w_{21}x_{22} + w_{31}x_{23}) + \sigma(w_{11}x_{31} + w_{21}x_{32} + w_{31}x_{33}) + \sigma(w_{12}x_{11} + w_{22}x_{12} + w_{32}x_{13}) + \sigma(w_{12}x_{21} + w_{22}x_{22} + w_{32}x_{23}) + \sigma(w_{12}x_{31} + w_{22}x_{32} + w_{32}x_{33})$

[43]:
```
compose(lambd, sigmaApply, v)(A,B).replace(v,v_).diff(B).
  ↪doit()#replace(sigmaApply, sigmaApply_)#.replace(lambd, lambd_L).
  ↪diff(B)
```

[43]: $$\left.\frac{d}{d\xi_1}\sigma_{apply}(\xi_1)\right|_{\xi_1=XW}\frac{\partial}{\partial\sigma_{apply}(XW)}\lambda(\sigma_{apply}(XW))\frac{\partial}{\partial W}XW$$

[44]:
```
compose(lambd, sigmaApply, v)(A,B).replace(v,v_).diff(B).
  ↪replace(lambd, lambd_L)
```

[44]: $$\frac{\partial}{\partial\sigma_{apply}(XW)}\left((\sigma_{apply}(XW))_{0,0} + (\sigma_{apply}(XW))_{0,1} + (\sigma_{apply}(XW))_{1,0} + (\sigma_{apply}(XW))_{1,1} + (\sigma_{apply}(XW))_{2,0} + (\sigma_{apply}(XW))_{2,1}\right)\left.\frac{d}{d\xi_1}\sigma_{apply}(\xi_1)\right|_{\xi_1=XW}\frac{\partial}{\partial W}XW$$