Steps for Batch Normalization with Backpropagation

Soon-Hyung Yook

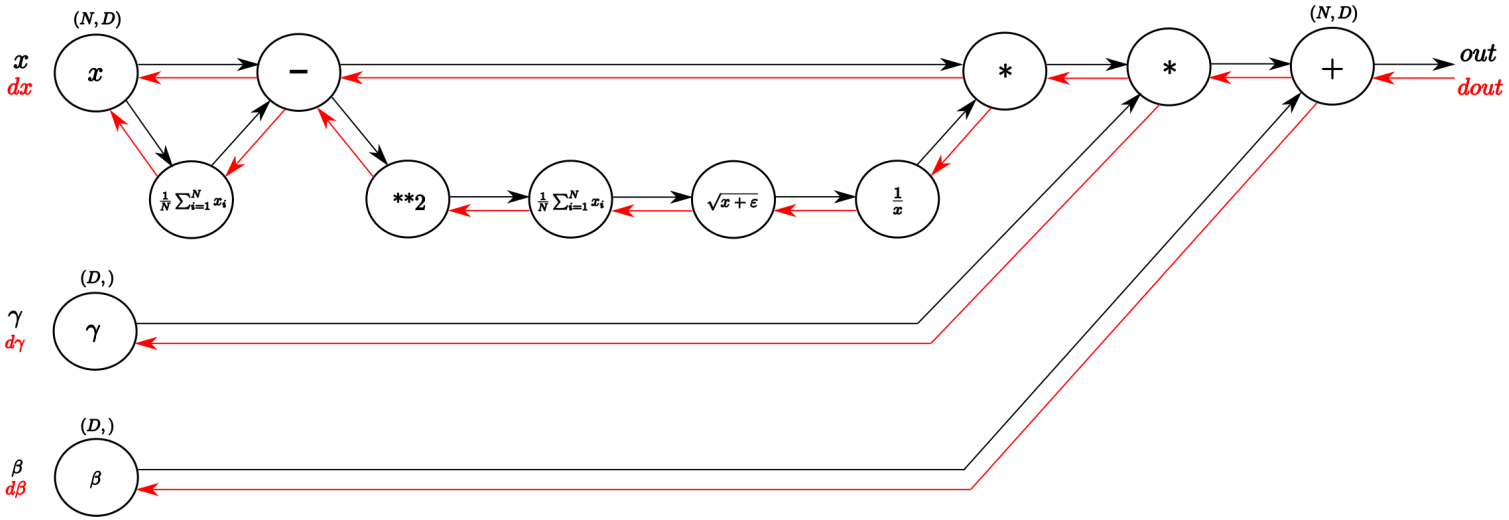
Department of Physics, Kyung Hee University

Table of Contents

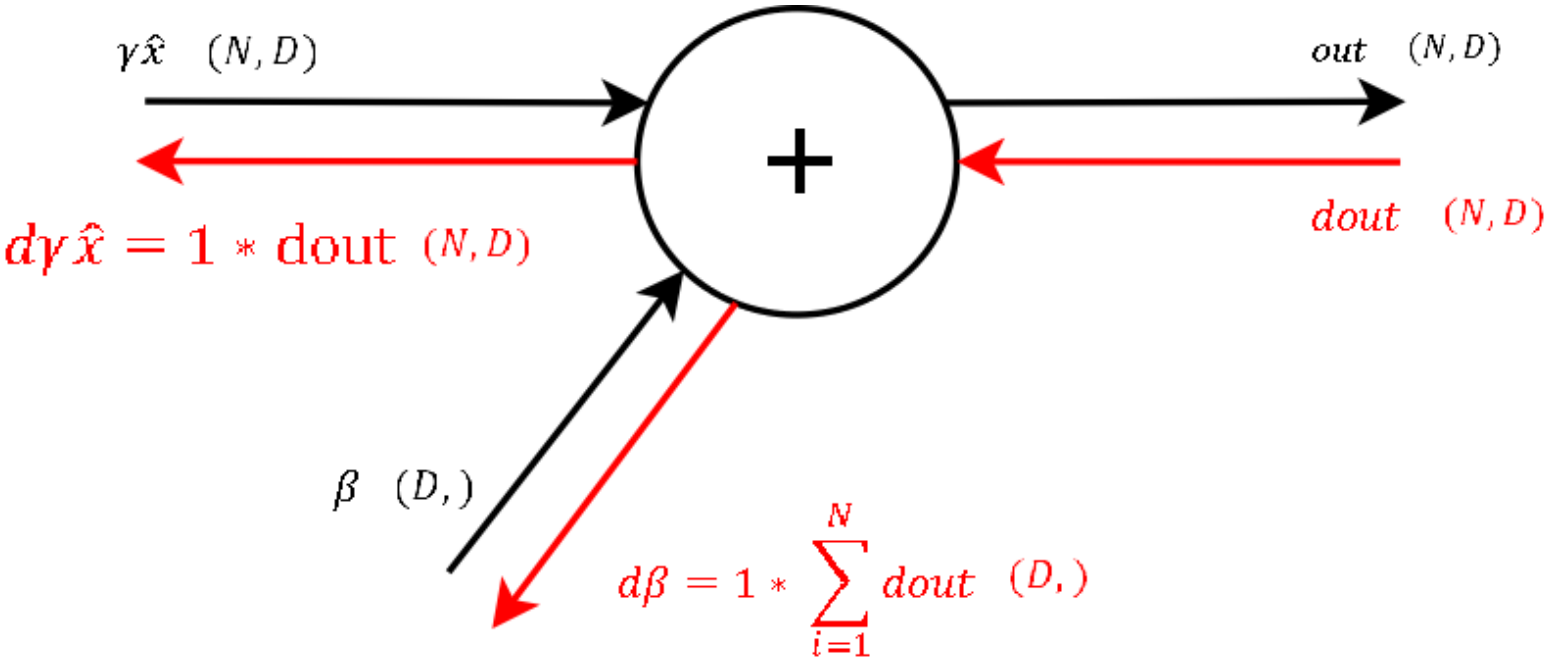
1. Circuit

2. Steps

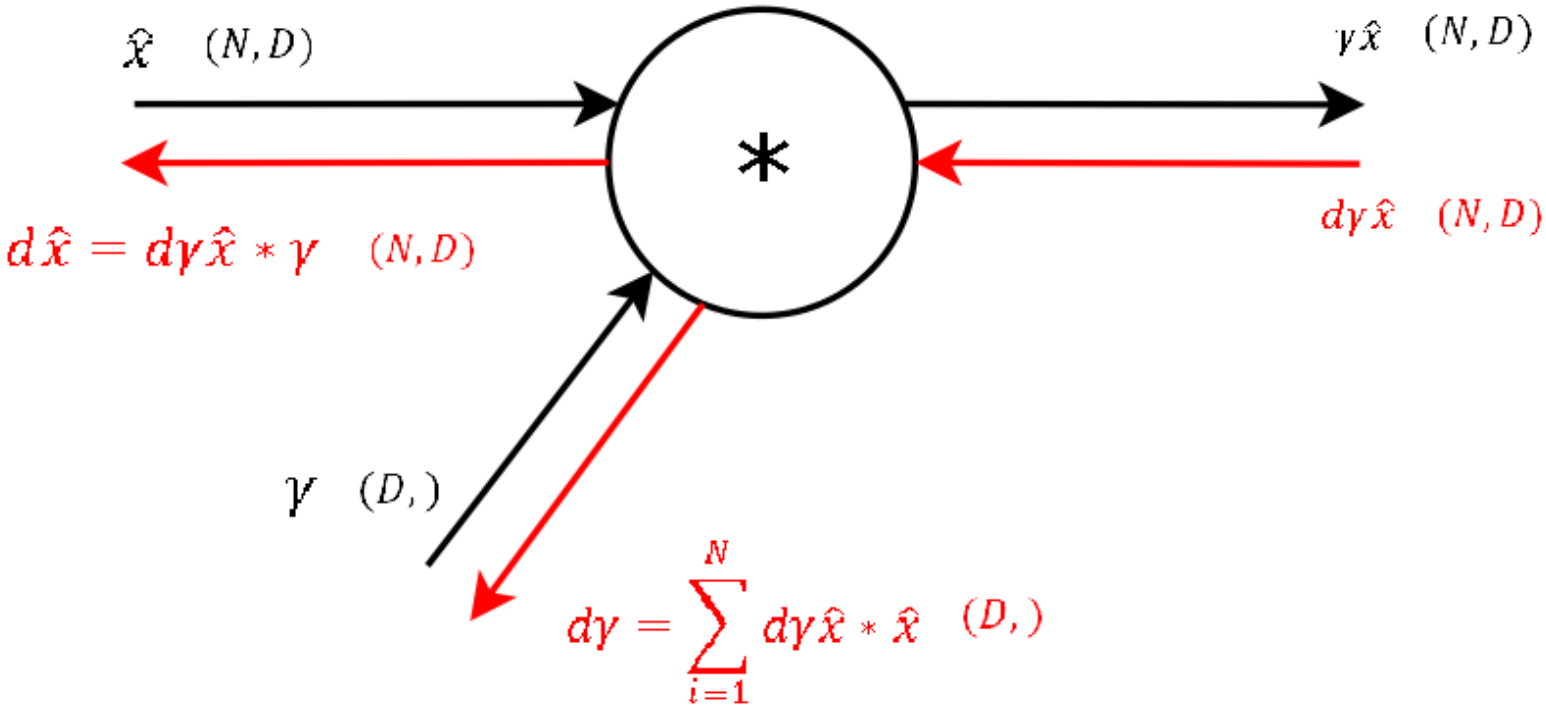
Batch Normal Circuit



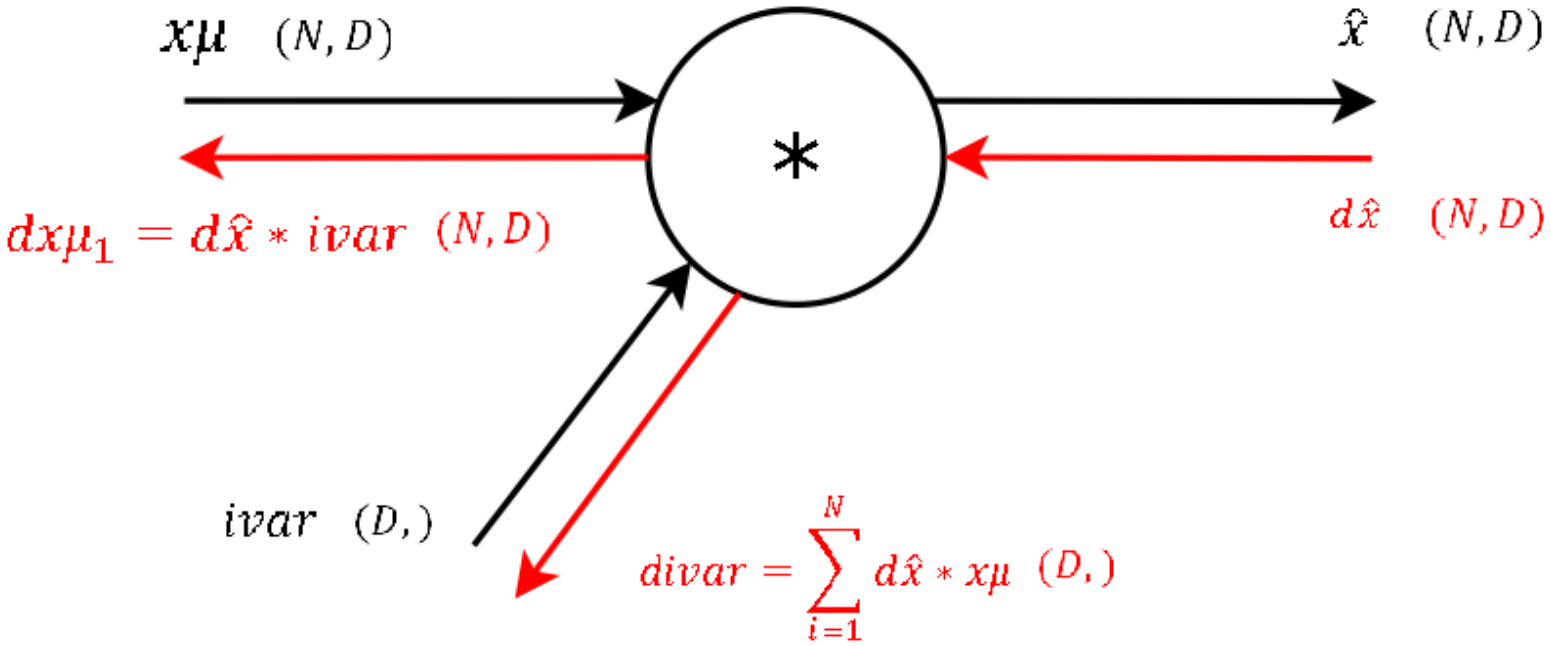
Step 9


$\gamma\hat{x}$ (N, D)
out (N, D)
$d\gamma\hat{x} = 1 * \text{dout}$ (N, D)
dout (N, D)
β $(D,)$
$d\beta = 1 * \sum_{i=1}^{N} dout$ $(D,)$

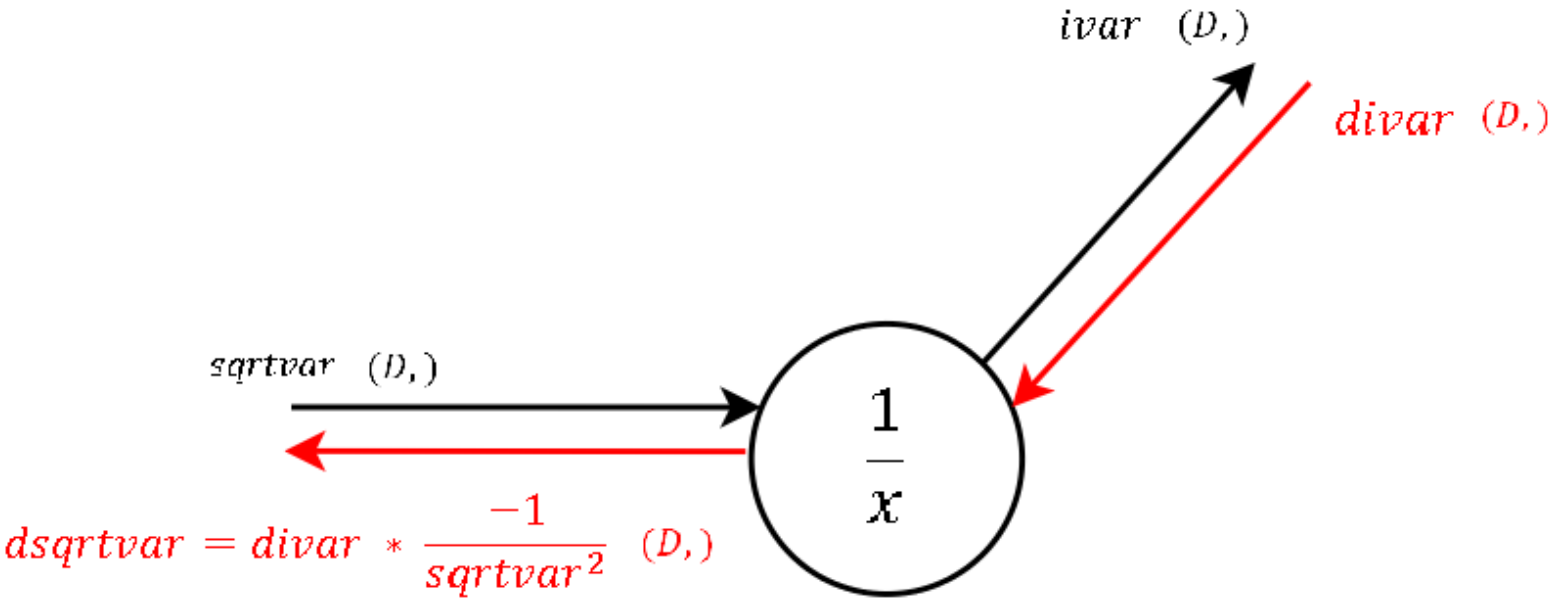
Step 8


$\hat{x}$ (N, D)
$\gamma\hat{x}$ (N, D)
$*$
$d\hat{x} = d\gamma\hat{x} * \gamma$ (N, D)
$d\gamma\hat{x}$ (N, D)
γ $(D,)$
$d\gamma = \sum_{i=1}^{N} d\gamma\hat{x} * \hat{x}$ $(D,)$

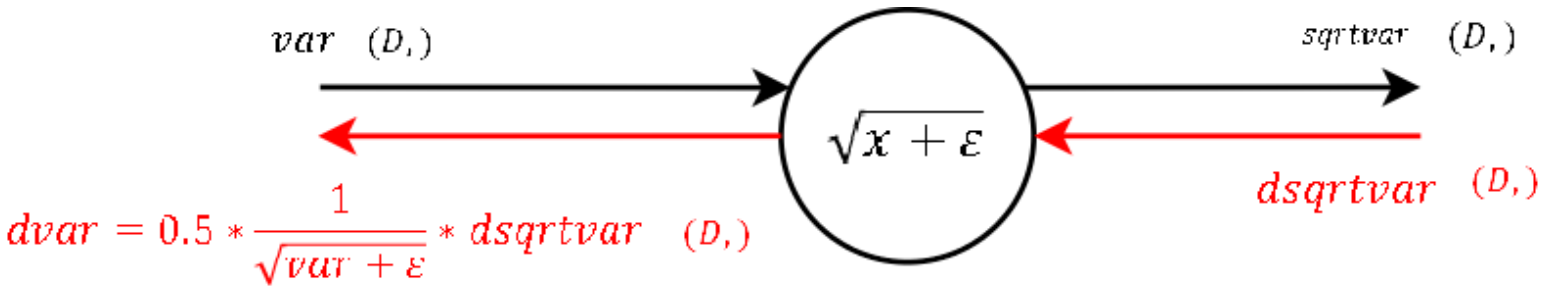
Step 7


$x\mu \quad (N, D)$
$\hat{x} \quad (N, D)$
$*$
$dx\mu_1 = d\hat{x} * ivar \quad (N, D)$
$d\hat{x} \quad (N, D)$
$ivar \quad (D,)$
$divar = \sum_{i=1}^{N} d\hat{x} * x\mu \quad (D,)$

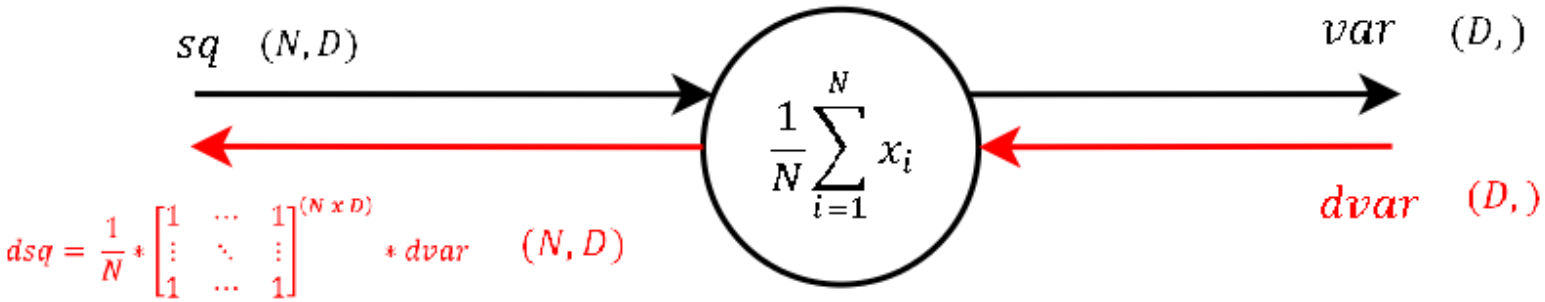
Step 6



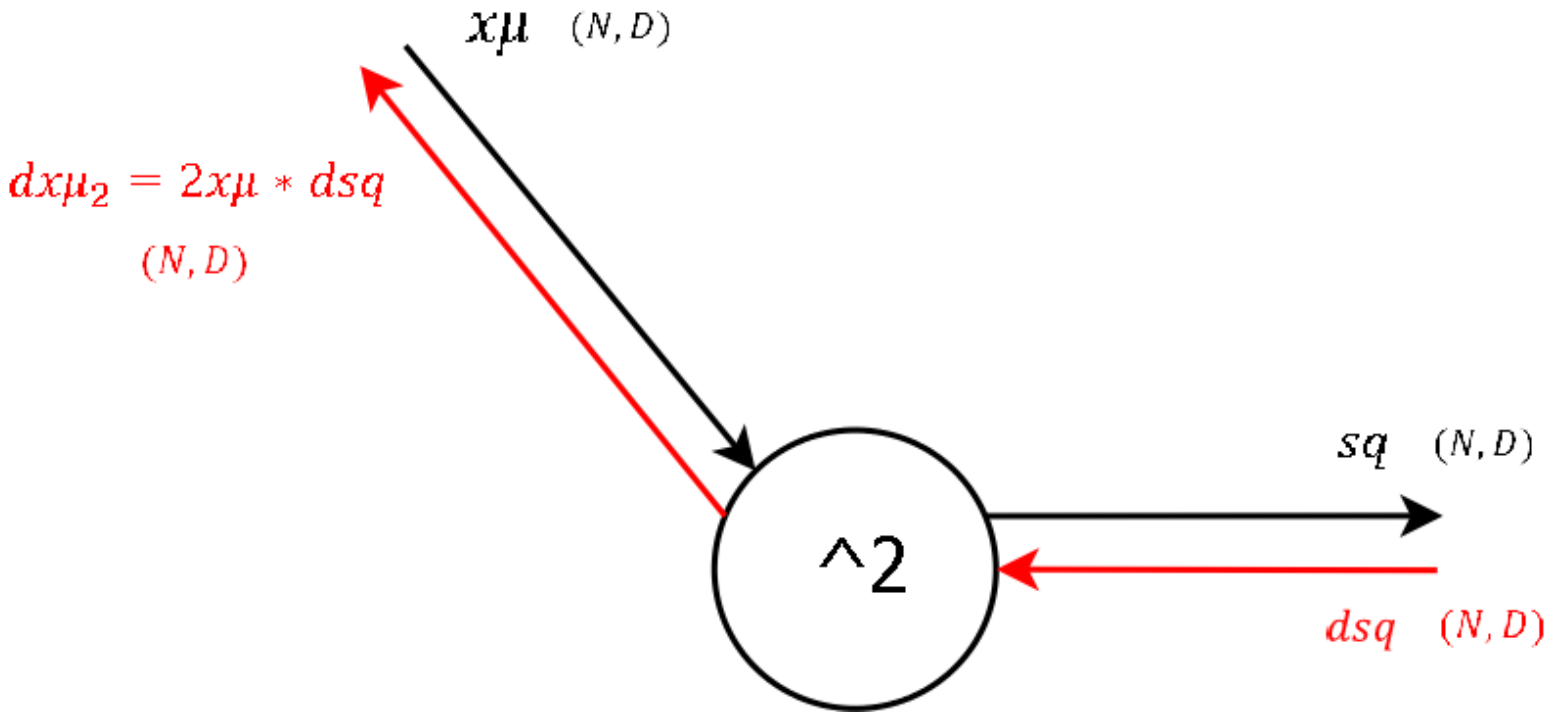
Step 5


$var \quad (D,)$
$\sqrt{x+\varepsilon}$
$sqrtvar \quad (D,)$
$dsqrtvar \quad (D,)$
$dvar = 0.5 * \frac{1}{\sqrt{var+\varepsilon}} * dsqrtvar \quad (D,)$

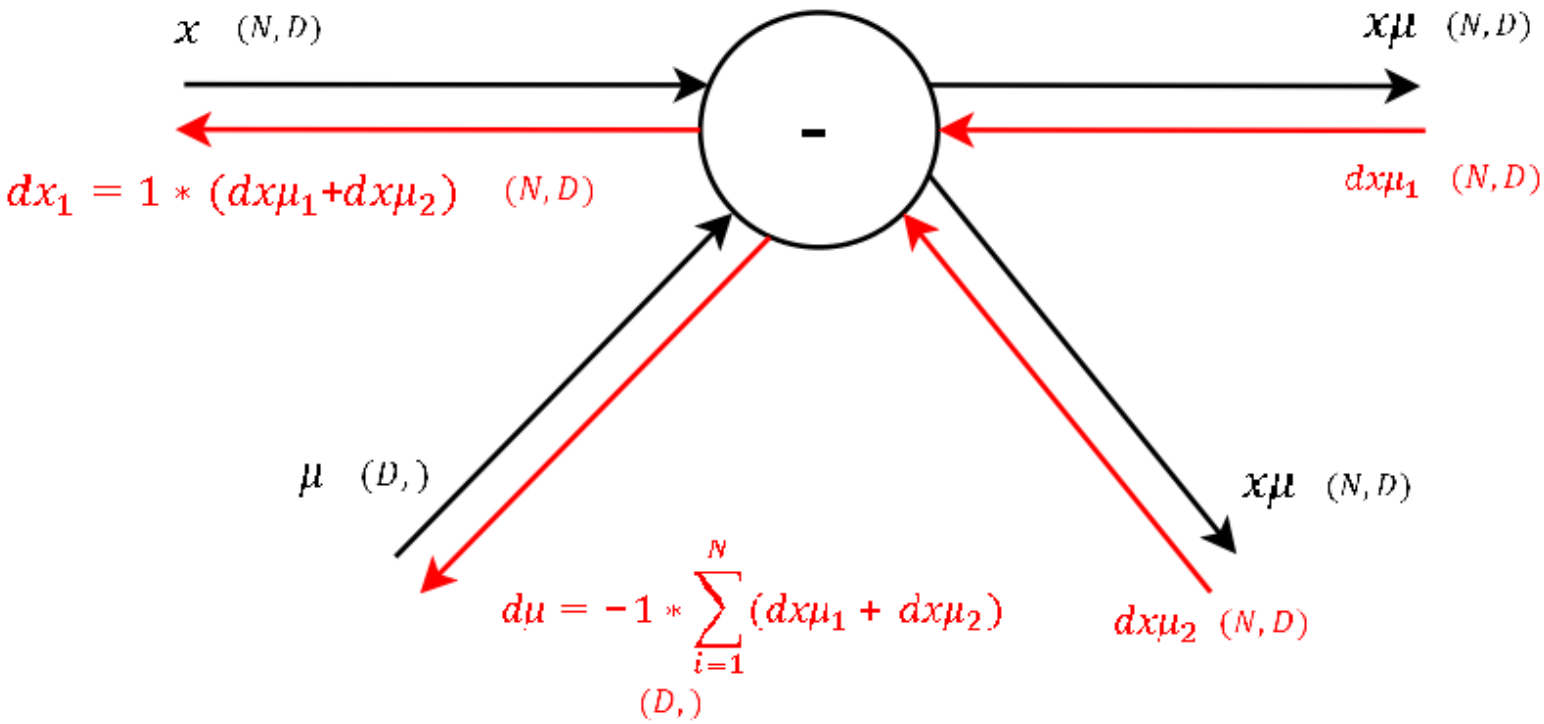
Step 4


sq (N, D)
var (D,)
$\frac{1}{N}\sum_{i=1}^{N} x_i$
dvar (D,)
$dsq = \frac{1}{N} * \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}^{(N \times D)} * dvar$ (N, D)

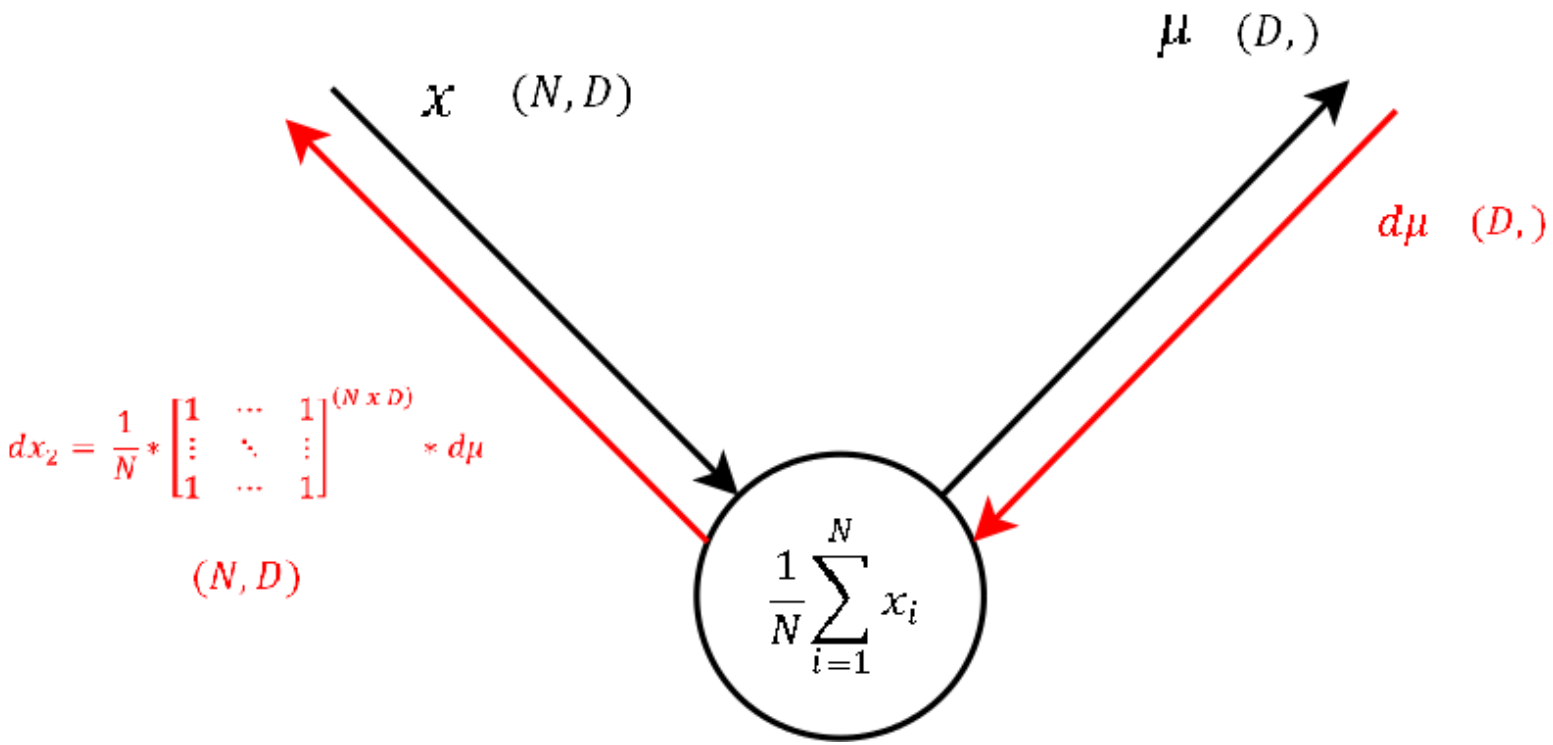
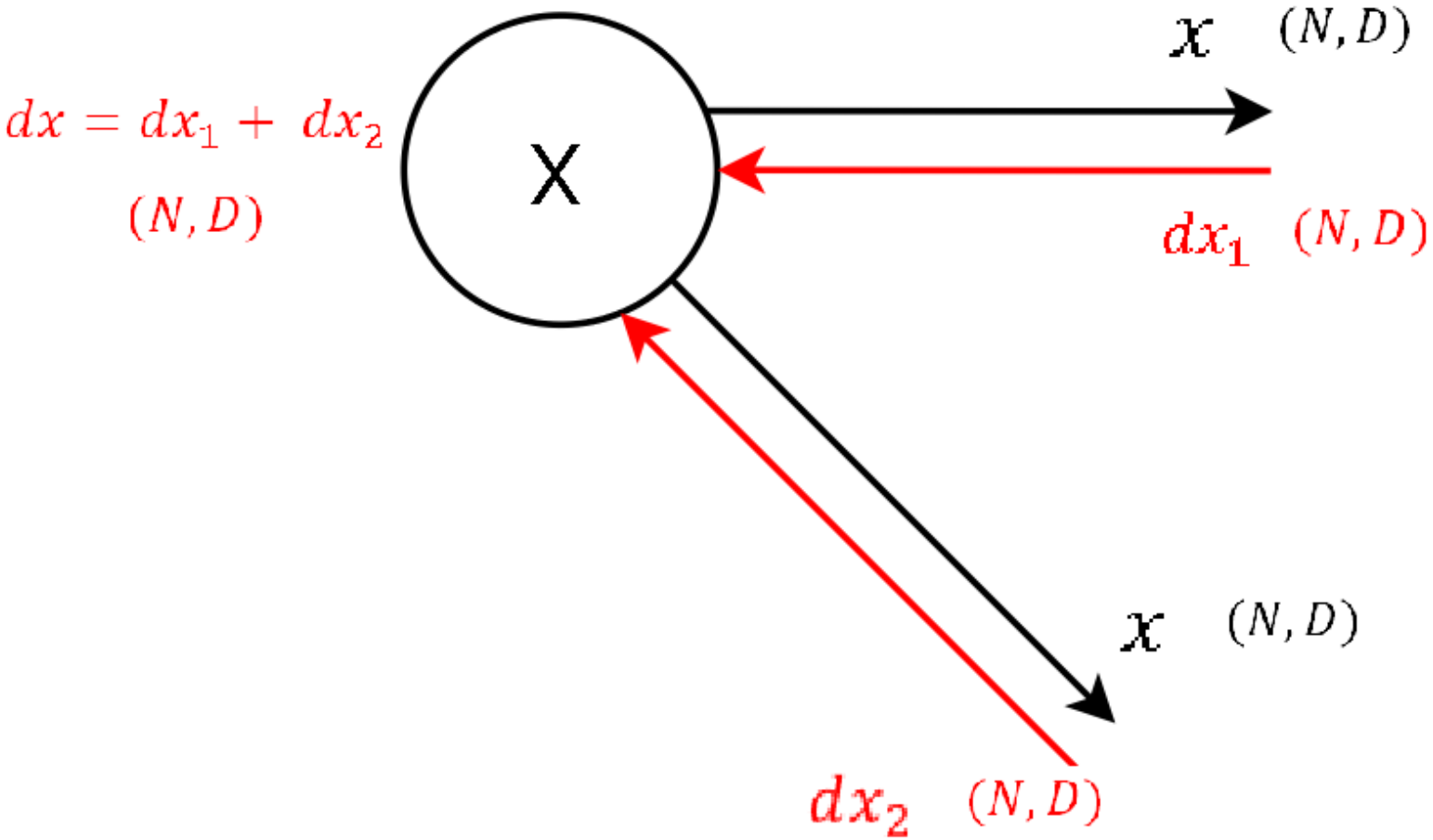
Step 3



Step 2


x (N,D)
$x\mu$ (N,D)
$-$
$dx_1 = 1 * (dx\mu_1 + dx\mu_2)$ (N,D)
$dx\mu_1$ (N,D)
μ $(D,)$
$x\mu$ (N,D)
$d\mu = -1 * \sum_{i=1}^{N} (dx\mu_1 + dx\mu_2)$
$(D,)$
$dx\mu_2$ (N,D)

Step 1


$x \quad (N, D)$
$\mu \quad (D,)$
$d\mu \quad (D,)$
$dx_2 = \frac{1}{N} * \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}^{(N \times D)} * d\mu$
(N, D)
$\frac{1}{N}\sum_{i=1}^{N} x_i$

Step 0