The weight update equations proposed in the paper are dimensionally inconsistent; the shapes of the matrices make some subtraction and multiplication operations invalid as written.

$$\underbrace{W_1}_{m_1\times d} \leftarrow \underbrace{W_1}_{m_1\times d} + \mu_1\Big(\underbrace{b_1}_{d\times m_1} - \underbrace{A_1}_{d\times d}\underbrace{W_1}_{m_1\times d}\Big)$$

$$\underbrace{W_2}_{m_2\times m_1} \leftarrow \underbrace{W_2}_{m_2\times m_1} + \mu_2\Big(\underbrace{b_2}_{m_1\times m_2} - \underbrace{A_2}_{m_1\times m_1}\underbrace{W_2}_{m_2\times m_1}\Big)$$

The following alternative updates are dimensionally consistent, although I am not sure about their correctness with respect to the underlying algorithm:

$$\underbrace{W_1}_{m_1\times d} \leftarrow \underbrace{W_1}_{m_1\times d} + \mu_1\Big(\underbrace{b_1^T}_{m_1\times d} - \underbrace{W_1}_{m_1\times d}\underbrace{A_1}_{d\times d}\Big)$$

$$\underbrace{W_2}_{m_2\times m_1} \leftarrow \underbrace{W_2}_{m_2\times m_1} + \mu_2\Big(\underbrace{b_2^T}_{m_2\times m_1} - \underbrace{W_2}_{m_2\times m_1}\underbrace{A_2}_{m_1\times m_1}\Big)$$

The matrix dimensions for the neural network are summarized in the following table to provide a clear reference for the shapes of inputs, weights, biases, and layer outputs.

Symbol	**Dimensions**	**Meaning**
X	$\mathbb{R}^{d\times n}$	Input data matrix, where d is the number of features per sample and n is the number of samples
W_1	$\mathbb{R}^{m_1\times d}$	First layer weight matrix, where m_1 is the number of neurons in the first layer
β_1	$\mathbb{R}^{m_1\times 1}$	First layer bias vector, added to each of the n columns of the layer output
Z_1	$\mathbb{R}^{m_1\times n}$	First layer pre-activation output $(W_1X + b_1)$
H_1	$\mathbb{R}^{m_1\times n}$	First layer post-activation output after applying nonlinearity $\sigma_1(Z_1)$
W_2	$\mathbb{R}^{m_2\times m_1}$	Second layer weight matrix, where m_2 is the number of neurons in the second layer
β_2	$\mathbb{R}^{m_2\times 1}$	Second layer bias vector, added to each of the n columns of the layer output
Z_2	$\mathbb{R}^{m_2\times n}$	Second layer pre-activation output $(W_2H_1 + b_2)$
H_2	$\mathbb{R}^{m_2\times n}$	Second layer post-activation output after applying nonlinearity $\sigma_2(Z_2)$

The update equations for the matrices, as given in the paper, are as follows:

$$
\begin{aligned}
\underbrace{A_1}_{d\times d} &\leftarrow \underbrace{\frac{r}{r+N}A_1}_{d\times d} + \underbrace{\frac{1}{r+N}\overset{d\times n}{X}\overset{n\times d}{X^T}}_{d\times d} \\
\underbrace{b_1}_{d\times m_1} &\leftarrow \underbrace{\frac{r}{r+N}b_1}_{d\times m_1} + \underbrace{\frac{1}{r+N}\overset{d\times n}{X}\overset{n\times m_1}{Z_1^T}}_{d\times m_1} \\
\underbrace{A_2}_{m_1\times m_1} &\leftarrow \underbrace{\frac{r}{r+N}A_2}_{m_1\times m_1} + \underbrace{\frac{1}{r+N}\overset{m_1\times n}{H_1}\overset{n\times m_1}{H_1^T}}_{m_1\times m_1} \\
\underbrace{b_2}_{m_1\times m_2} &\leftarrow \underbrace{\frac{r}{r+N}b_2}_{m_1\times m_2} + \underbrace{\frac{1}{r+N}\overset{m_1\times n}{H_1}\overset{n\times m_2}{Z_2^T}}_{m_1\times m_2}
\end{aligned}
$$

The dimensions of the matrices used in the update equations can be seen in the next table.

Symbol	**Dimensions**	**Meaning**
r	scalar	Forgetting factor used in the update equations
N	scalar	Number of samples in the current batch
A_1	$\mathbb{R}^{d\times d}$	Updated matrix estimate for first-layer correlation using XX^T
b_1	$\mathbb{R}^{d\times m_1}$	Updated bias-like matrix estimate for first layer using XZ_1^T
XX^T	$\mathbb{R}^{d\times d}$	Covariance-like term used in A_1 update
XZ_1^T	$\mathbb{R}^{d\times m_1}$	Outer-product term used in b_1 update
A_2	$\mathbb{R}^{m_1\times m_1}$	Updated matrix estimate for second-layer correlation using $H_1H_1^T$
b_2	$\mathbb{R}^{m_1\times m_2}$	Updated bias-like matrix estimate for second layer using $H_1Z_2^T$
$H_1H_1^T$	$\mathbb{R}^{m_1\times m_1}$	Outer-product term used in A_2 update
$H_1Z_2^T$	$\mathbb{R}^{m_1\times m_2}$	Outer-product term used in b_2 update

Even if the bias is absorbed into the weight matrices, the original operations remain invalid:

$$
\begin{aligned}
\underbrace{\widetilde{W}_1}_{m_1\times(d+1)} &\leftarrow \underbrace{\widetilde{W}_1}_{m_1\times(d+1)} + \mu_1\Big(\underbrace{\tilde{b}_1}_{(d+1)\times m_1} - \underbrace{\widetilde{A}_1}_{(d+1)\times(d+1)}\underbrace{\widetilde{W}_1}_{m_1\times(d+1)}\Big)\\
\underbrace{\widetilde{W}_2}_{m_2\times(m_1+1)} &\leftarrow \underbrace{\widetilde{W}_2}_{m_2\times(m_1+1)} + \mu_2\Big(\underbrace{\tilde{b}_2}_{(m_1+1)\times m_2} - \underbrace{\widetilde{A}_2}_{(m_1+1)\times(m_1+1)}\underbrace{\widetilde{W}_2}_{m_2\times(m_1+1)}\Big)
\end{aligned}
$$

The following updates are dimensionally consistent in the augmented form, but again, I am not certain about their correctness:

$$
\begin{aligned}
\underbrace{\widetilde{W}_1}_{m_1\times(d+1)} &\leftarrow \underbrace{\widetilde{W}_1}_{m_1\times(d+1)} + \mu_1\Big(\underbrace{b_1^T}_{m_1\times(d+1)} - \underbrace{\widetilde{W}_1}_{m_1\times(d+1)}\underbrace{A_1}_{(d+1)\times(d+1)}\Big)\\
\underbrace{\widetilde{W}_2}_{m_2\times(m_1+1)} &\leftarrow \underbrace{\widetilde{W}_2}_{m_2\times(m_1+1)} + \mu_2\Big(\underbrace{b_2^T}_{m_2\times(m_1+1)} - \underbrace{\widetilde{W}_2}_{m_2\times(m_1+1)}\underbrace{A_2}_{(m_1+1)\times(m_1+1)}\Big)
\end{aligned}
$$

The following table summarizes the new dimensions for the augmented matrices.

Symbol	**Dimensions**	**Meaning**
$\widetilde{X}$	$\mathbb{R}^{(d+1)\times n}$	Augmented input with a row of ones for bias absorption
$\widetilde{W}_1$	$\mathbb{R}^{m_1\times(d+1)}$	First layer augmented weight matrix including bias
$\widetilde{H}_1$	$\mathbb{R}^{(m_1+1)\times n}$	Augmented first layer output H_1 with a row of ones for bias absorption in the second layer
$\widetilde{W}_2$	$\mathbb{R}^{m_2\times(m_1+1)}$	Second layer augmented weight matrix including bias

The update equations for the augmented matrices then become

$$
\begin{aligned}
\underbrace{\widetilde{A}_1}_{(d+1)\times(d+1)} &\leftarrow \underbrace{\frac{r}{r+N}\widetilde{A}_1}_{(d+1)\times(d+1)} + \underbrace{\frac{1}{r+N}\overset{(d+1)\times n}{\widetilde{X}}\;\overset{n\times(d+1)}{\widetilde{X}^T}}_{(d+1)\times(d+1)}\\
\underbrace{\tilde{b}_1}_{(d+1)\times m_1} &\leftarrow \underbrace{\frac{r}{r+N}\tilde{b}_1}_{(d+1)\times m_1} + \underbrace{\frac{1}{r+N}\overset{(d+1)\times n}{\widetilde{X}}\;\overset{n\times m_1}{Z_1^T}}_{(d+1)\times m_1}\\
\underbrace{\widetilde{A}_2}_{(m_1+1)\times(m_1+1)} &\leftarrow \underbrace{\frac{r}{r+N}\widetilde{A}_2}_{(m_1+1)\times(m_1+1)} + \underbrace{\frac{1}{r+N}\overset{(m_1+1)\times n}{\widetilde{H}_1}\;\overset{n\times(m_1+1)}{\widetilde{H}_1^T}}_{(m_1+1)\times(m_1+1)}\\
\underbrace{\tilde{b}_2}_{(m_1+1)\times m_2} &\leftarrow \underbrace{\frac{r}{r+N}\tilde{b}_2}_{(m_1+1)\times m_2} + \underbrace{\frac{1}{r+N}\overset{(m_1+1)\times n}{\widetilde{H}_1}\;\overset{n\times m_2}{Z_2^T}}_{(m_1+1)\times m_2}
\end{aligned}
$$

The following table summarizes the dimensions of the augmented matrices used in the new update equations.

Symbol	**Dimensions**	**Meaning**
$\widetilde{A}_1$	$\mathbb{R}^{(d+1)\times(d+1)}$	Augmented first-layer correlation matrix estimate using $\widetilde{X}\widetilde{X}^T$
$\widetilde{b}_1$	$\mathbb{R}^{(d+1)\times m_1}$	Augmented first-layer bias-like matrix estimate using $\widetilde{X}Z_1^T$
$\widetilde{X}\widetilde{X}^T$	$\mathbb{R}^{(d+1)\times(d+1)}$	Outer-product of augmented input for first-layer correlation
$\widetilde{X}Z_1^T$	$\mathbb{R}^{(d+1)\times m_1}$	Outer-product of augmented input with first-layer pre-activation for bias update
$\widetilde{A}_2$	$\mathbb{R}^{(m_1+1)\times(m_1+1)}$	Augmented second-layer correlation matrix estimate using $\widetilde{H}_1\widetilde{H}_1^T$
$\widetilde{b}_2$	$\mathbb{R}^{(m_1+1)\times m_2}$	Augmented second-layer bias-like matrix estimate using $\widetilde{H}_1 Z_2^T$
$\widetilde{H}_1\widetilde{H}_1^T$	$\mathbb{R}^{(m_1+1)\times(m_1+1)}$	Outer-product of augmented first-layer output for second-layer correlation
$\widetilde{H}_1 Z_2^T$	$\mathbb{R}^{(m_1+1)\times m_2}$	Outer-product of augmented first-layer output with second-layer pre-activation for bias update