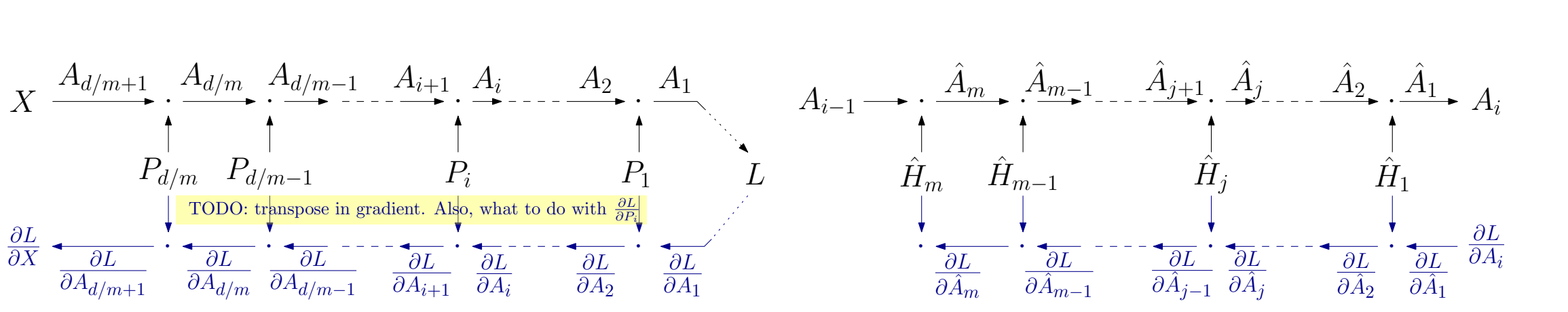
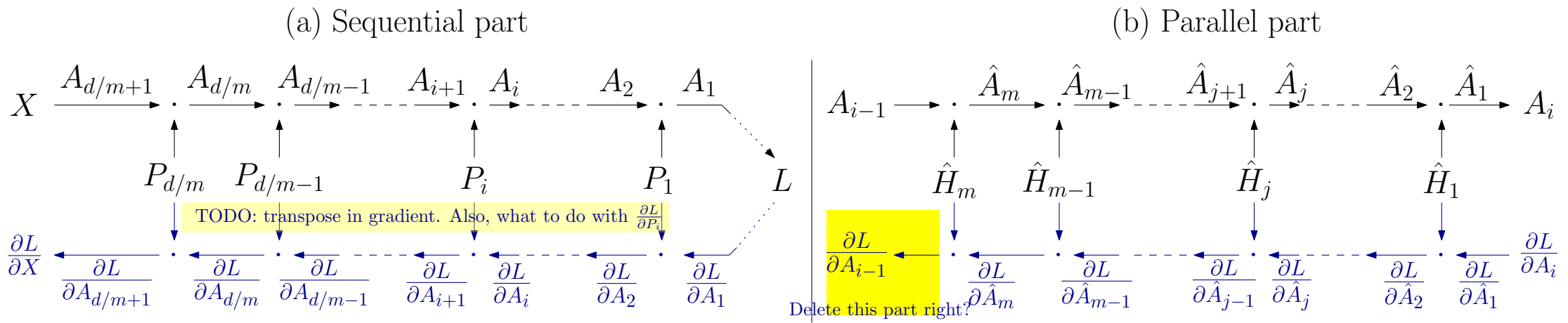
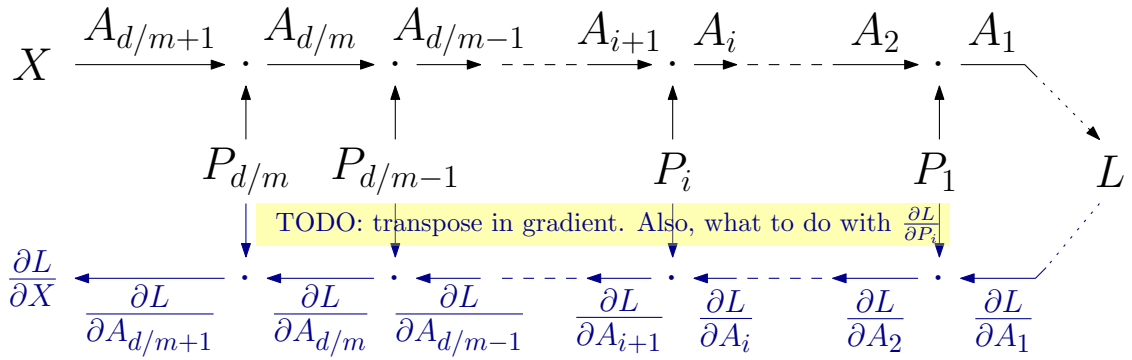
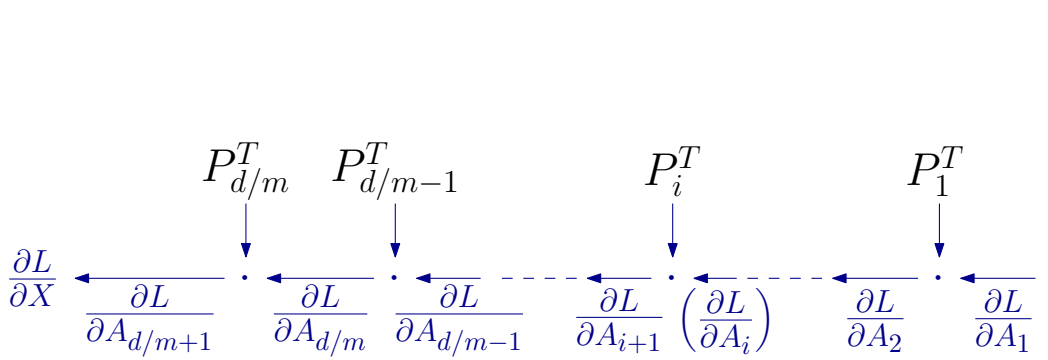
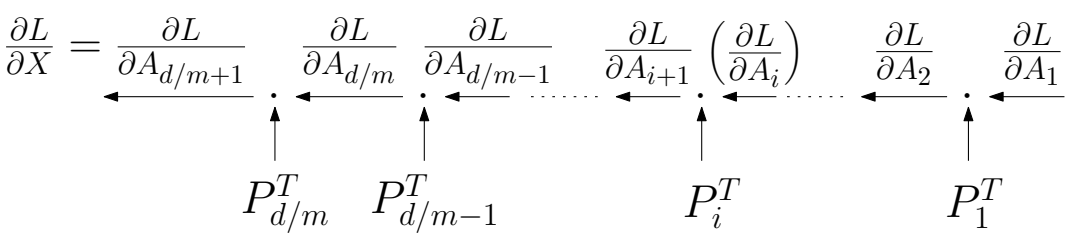
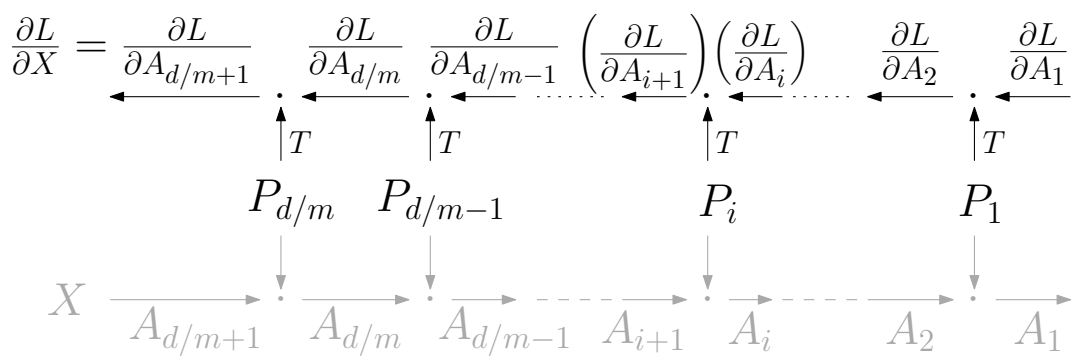

X
$A_{d/m+1}$
$A_{d/m}$
$A_{d/m-1}$
A_{i+1}
A_i
A_2
A_1
L
$P_{d/m}$
$P_{d/m-1}$
P_i
P_1
TODO: transpose in gradient. Also, what to do with $\frac{\partial L}{\partial P_i}$
$\frac{\partial L}{\partial X}$
$\frac{\partial L}{\partial A_{d/m+1}}$
$\frac{\partial L}{\partial A_{d/m}}$
$\frac{\partial L}{\partial A_{d/m-1}}$
$\frac{\partial L}{\partial A_{i+1}}$
$\frac{\partial L}{\partial A_i}$
$\frac{\partial L}{\partial A_2}$
$\frac{\partial L}{\partial A_1}$
A_{i-1}
$\hat{A}_m$
$\hat{A}_{m-1}$
$\hat{A}_{j+1}$
$\hat{A}_j$
$\hat{A}_2$
$\hat{A}_1$
A_i
$\hat{H}_m$
$\hat{H}_{m-1}$
$\hat{H}_j$
$\hat{H}_1$
$\frac{\partial L}{\partial A_i}$
$\frac{\partial L}{\partial \hat{A}_m}$
$\frac{\partial L}{\partial \hat{A}_{m-1}}$
$\frac{\partial L}{\partial \hat{A}_{j-1}}$
$\frac{\partial L}{\partial \hat{A}_j}$
$\frac{\partial L}{\partial \hat{A}_2}$
$\frac{\partial L}{\partial \hat{A}_1}$

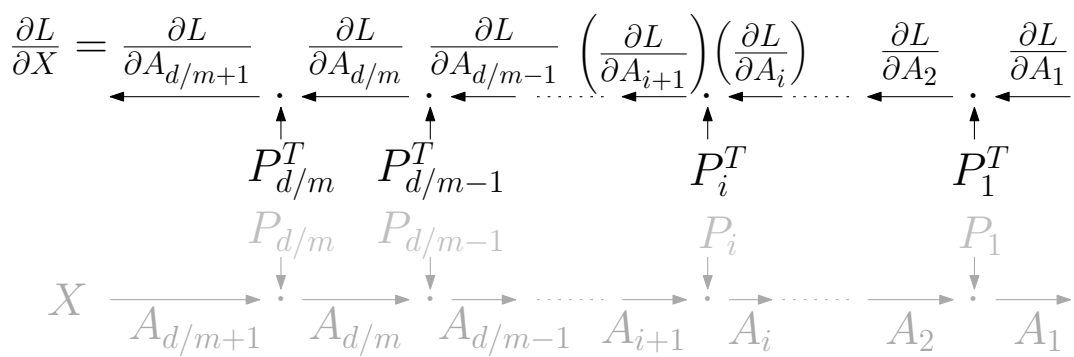

(a) Sequential part
(b) Parallel part
X
$A_{d/m+1}$
$A_{d/m}$
$A_{d/m-1}$
A_{i+1}
A_i
A_2
A_1
L
$P_{d/m}$
$P_{d/m-1}$
P_i
P_1
TODO: transpose in gradient. Also, what to do with $\frac{\partial L}{\partial P_i}$
$\frac{\partial L}{\partial X}$
$\frac{\partial L}{\partial A_{d/m+1}}$
$\frac{\partial L}{\partial A_{d/m}}$
$\frac{\partial L}{\partial A_{d/m-1}}$
$\frac{\partial L}{\partial A_{i+1}}$
$\frac{\partial L}{\partial A_i}$
$\frac{\partial L}{\partial A_2}$
$\frac{\partial L}{\partial A_1}$
A_{i-1}
$\hat{A}_m$
$\hat{A}_{m-1}$
$\hat{A}_{j+1}$
$\hat{A}_j$
$\hat{A}_2$
$\hat{A}_1$
A_i
$\hat{H}_m$
$\hat{H}_{m-1}$
$\hat{H}_j$
$\hat{H}_1$
$\frac{\partial L}{\partial A_{i-1}}$
Delete this part right?
$\frac{\partial L}{\partial \hat{A}_m}$
$\frac{\partial L}{\partial \hat{A}_{m-1}}$
$\frac{\partial L}{\partial \hat{A}_{j-1}}$
$\frac{\partial L}{\partial \hat{A}_j}$
$\frac{\partial L}{\partial \hat{A}_2}$
$\frac{\partial L}{\partial \hat{A}_1}$
$\frac{\partial L}{\partial A_i}$

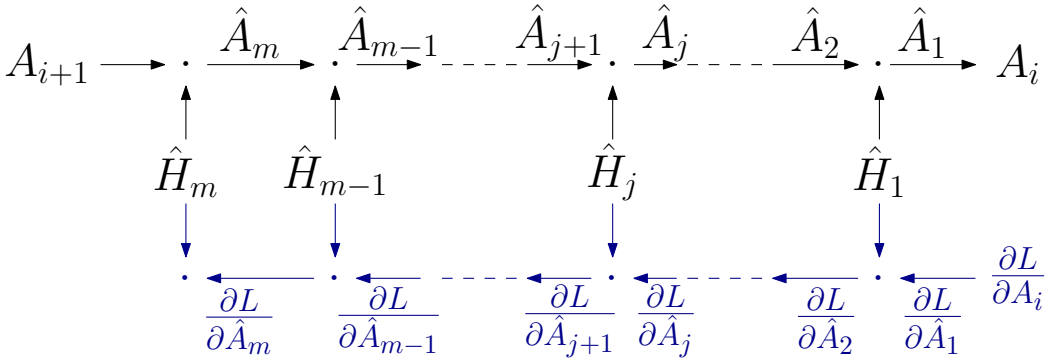

X
$A_{d/m+1}$
$A_{d/m}$
$A_{d/m-1}$
A_{i+1}
A_i
A_2
A_1
L
$P_{d/m}$
$P_{d/m-1}$
P_i
P_1
TODO: transpose in gradient. Also, what to do with $\frac{\partial L}{\partial P_i}$
$\frac{\partial L}{\partial X}$
$\frac{\partial L}{\partial A_{d/m+1}}$
$\frac{\partial L}{\partial A_{d/m}}$
$\frac{\partial L}{\partial A_{d/m-1}}$
$\frac{\partial L}{\partial A_{i+1}}$
$\frac{\partial L}{\partial A_i}$
$\frac{\partial L}{\partial A_2}$
$\frac{\partial L}{\partial A_1}$

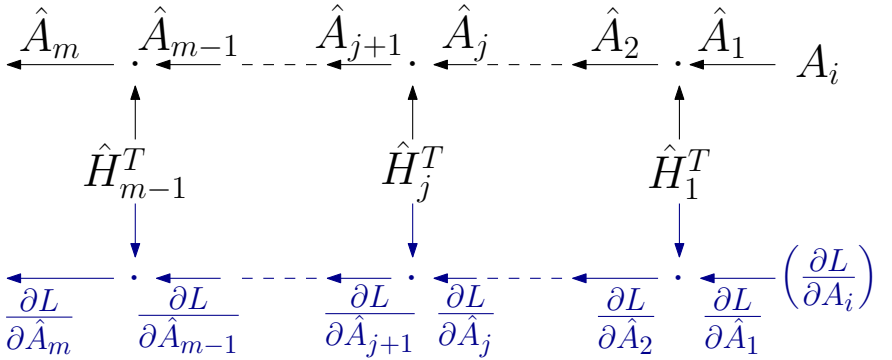

$P^T_{d/m}$
$P^T_{d/m-1}$
P^T_i
P^T_1
$\frac{\partial L}{\partial X}$
$\frac{\partial L}{\partial A_{d/m+1}}$
$\frac{\partial L}{\partial A_{d/m}}$
$\frac{\partial L}{\partial A_{d/m-1}}$
$\frac{\partial L}{\partial A_{i+1}}$
$\left(\frac{\partial L}{\partial A_i}\right)$
$\frac{\partial L}{\partial A_2}$
$\frac{\partial L}{\partial A_1}$

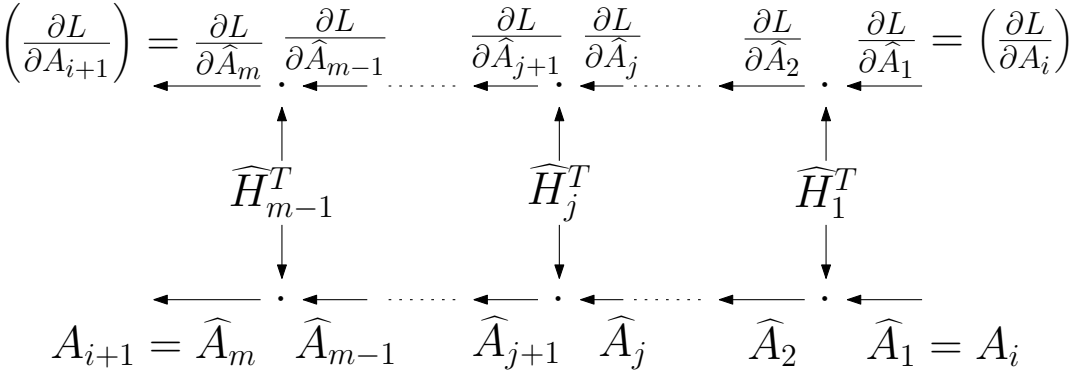

$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial A_{d/m+1}}$
$\frac{\partial L}{\partial A_{d/m}}$
$\frac{\partial L}{\partial A_{d/m-1}}$
$\frac{\partial L}{\partial A_{i+1}}$
$\left(\frac{\partial L}{\partial A_i}\right)$
$\frac{\partial L}{\partial A_2}$
$\frac{\partial L}{\partial A_1}$
$P^T_{d/m}$
$P^T_{d/m-1}$
P^T_i
P^T_1

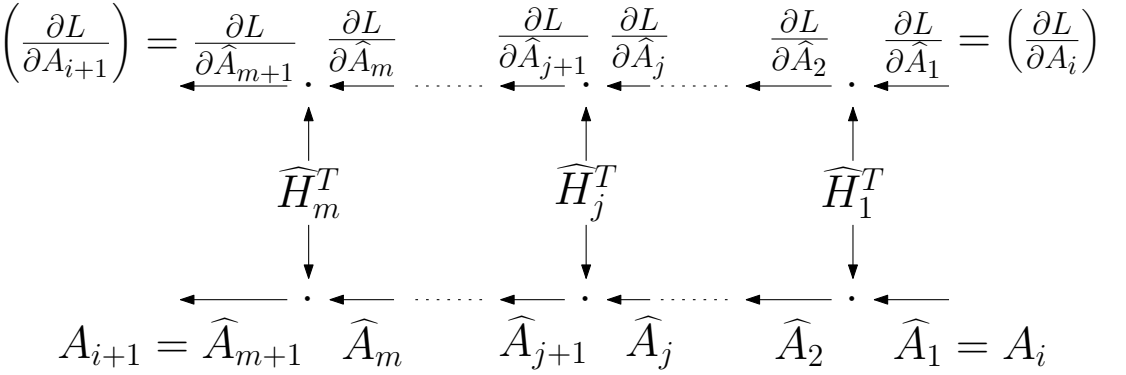

$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial A_{d/m+1}}$
$\frac{\partial L}{\partial A_{d/m}}$
$\frac{\partial L}{\partial A_{d/m-1}}$
$\left(\frac{\partial L}{\partial A_{i+1}}\right)\left(\frac{\partial L}{\partial A_i}\right)$
$\frac{\partial L}{\partial A_2}$
$\frac{\partial L}{\partial A_1}$
T
T
T
T
$P_{d/m}$
$P_{d/m-1}$
P_i
P_1
X
$A_{d/m+1}$
$A_{d/m}$
$A_{d/m-1}$
A_{i+1}
A_i
A_2
A_1


$\frac{\partial L}{\partial X} =$
$\frac{\partial L}{\partial A_{d/m+1}}$
$\frac{\partial L}{\partial A_{d/m}}$
$\frac{\partial L}{\partial A_{d/m-1}}$
$\left(\frac{\partial L}{\partial A_{i+1}}\right)\left(\frac{\partial L}{\partial A_i}\right)$
$\frac{\partial L}{\partial A_2}$
$\frac{\partial L}{\partial A_1}$
$P_{d/m}^T$
$P_{d/m-1}^T$
P_i^T
P_1^T
$P_{d/m}$
$P_{d/m-1}$
P_i
P_1
X
$A_{d/m+1}$
$A_{d/m}$
$A_{d/m-1}$
A_{i+1}
A_i
A_2
A_1


A_{i+1}
$\hat{A}_m$
$\hat{A}_{m-1}$
$\hat{A}_{j+1}$
$\hat{A}_j$
$\hat{A}_2$
$\hat{A}_1$
A_i
$\hat{H}_m$
$\hat{H}_{m-1}$
$\hat{H}_j$
$\hat{H}_1$
$\frac{\partial L}{\partial \hat{A}_m}$
$\frac{\partial L}{\partial \hat{A}_{m-1}}$
$\frac{\partial L}{\partial \hat{A}_{j+1}}$
$\frac{\partial L}{\partial \hat{A}_j}$
$\frac{\partial L}{\partial \hat{A}_2}$
$\frac{\partial L}{\partial \hat{A}_1}$
$\frac{\partial L}{\partial A_i}$


$\hat{A}_m$
$\hat{A}_{m-1}$
$\hat{A}_{j+1}$
$\hat{A}_j$
$\hat{A}_2$
$\hat{A}_1$
A_i
$\hat{H}^T_{m-1}$
$\hat{H}^T_j$
$\hat{H}^T_1$
$\frac{\partial L}{\partial \hat{A}_m}$
$\frac{\partial L}{\partial \hat{A}_{m-1}}$
$\frac{\partial L}{\partial \hat{A}_{j+1}}$
$\frac{\partial L}{\partial \hat{A}_j}$
$\frac{\partial L}{\partial \hat{A}_2}$
$\frac{\partial L}{\partial \hat{A}_1}$
$\left(\frac{\partial L}{\partial A_i}\right)$


$\left(\frac{\partial L}{\partial A_{i+1}}\right) = \frac{\partial L}{\partial \widehat{A}_m}$
$\frac{\partial L}{\partial \widehat{A}_{m-1}}$
$\frac{\partial L}{\partial \widehat{A}_{j+1}}$
$\frac{\partial L}{\partial \widehat{A}_j}$
$\frac{\partial L}{\partial \widehat{A}_2}$
$\frac{\partial L}{\partial \widehat{A}_1} = \left(\frac{\partial L}{\partial A_i}\right)$
$\widehat{H}_{m-1}^T$
$\widehat{H}_j^T$
$\widehat{H}_1^T$
$A_{i+1} = \widehat{A}_m$
$\widehat{A}_{m-1}$
$\widehat{A}_{j+1}$
$\widehat{A}_j$
$\widehat{A}_2$
$\widehat{A}_1 = A_i$


$\left(\frac{\partial L}{\partial A_{i+1}}\right) = \frac{\partial L}{\partial \widehat{A}_{m+1}}$
$\frac{\partial L}{\partial \widehat{A}_m}$
$\frac{\partial L}{\partial \widehat{A}_{j+1}}$
$\frac{\partial L}{\partial \widehat{A}_j}$
$\frac{\partial L}{\partial \widehat{A}_2}$
$\frac{\partial L}{\partial \widehat{A}_1} = \left(\frac{\partial L}{\partial A_i}\right)$
$\widehat{H}_m^T$
$\widehat{H}_j^T$
$\widehat{H}_1^T$
$A_{i+1} = \widehat{A}_{m+1}$
$\widehat{A}_m$
$\widehat{A}_{j+1}$
$\widehat{A}_j$
$\widehat{A}_2$
$\widehat{A}_1 = A_i$