PS 3 Yang Tian

1.1 $Y_{4\times1} = A_{4\times9}\, X_{9\times1}$

$$= \begin{bmatrix} W_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & W_{10} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & W_{01} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & W_{00} \end{bmatrix} \begin{bmatrix} x_{00} \\ x_{01} \\ x_{02} \\ x_{10} \\ x_{11} \\ x_{12} \\ x_{20} \\ x_{21} \\ x_{22} \end{bmatrix}$$

1.2 $Y_{16\times1} = A_{16\times4}\, X_{4\times1}$

$$= \begin{bmatrix} w_{00} & 0 & 0 & 0 \\ w_{01} & 0 & 0 & 0 \\ 0 & w_{00} & 0 & 0 \\ 0 & w_{01} & 0 & 0 \\ w_{10} & 0 & 0 & 0 \\ w_{11} & 0 & 0 & 0 \\ 0 & w_{10} & 0 & 0 \\ 0 & w_{11} & 0 & 0 \\ 0 & 0 & w_{00} & 0 \\ 0 & 0 & w_{01} & 0 \\ 0 & 0 & 0 & w_{00} \\ 0 & 0 & 0 & w_{01} \\ 0 & 0 & w_{10} & 0 \\ 0 & 0 & w_{11} & 0 \\ 0 & 0 & 0 & w_{10} \\ 0 & 0 & 0 & w_{11} \end{bmatrix} \begin{bmatrix} x_{00} \\ x_{01} \\ x_{10} \\ x_{11} \end{bmatrix}$$

1.3 Affine transformation for a conv. layer with (4,1,1,1)

$$A = \begin{bmatrix} w_0 & w_0 & 0 & \\ 0 & & w_0 & w_0 \\ w_1 & w_1 & 0 & \\ 0 & & w_1 & w_1 \\ w_2 & w_2 & 0 & \\ 0 & & w_2 & w_2 \\ w_3 & w_3 & 0 & \\ 0 & & w_3 & w_3 \end{bmatrix}$$

Affine transformation for a deconv layer with (1, 1, 2, 2)

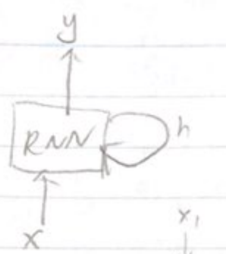
$$A = \begin{bmatrix} w_0 & & & \\ w_1 & & & \\ & w_0 & & 0 \\ & w_1 & & \\ w_2 & & & \\ w_3 & & & \\ & w_2 & & \\ & w_3 & & \\ & & w_0 & \\ & & w_1 & \\ & & & w_0 \\ & & & w_1 \\ 0 & & w_2 & \\ & & w_3 & \\ & & & w_2 \\ & & & w_3 \end{bmatrix}$$

Clearly after permuting rows, two As are the same. In other words, they are the identical operation

2.1 If G is a DAG, there exists ~~and only~~ one node with no incoming edges. We mark this node as v_1 and remove all edges out of it. Then there will be more nodes with no incoming edges. Recursively mark these incoming degree 0 node and remove all edges out from it, we will find a valid topological sort order.

2.2 If G has a topological order, it means along the order, for any node, other nodes can be either its ancestors or its childen (grandchilden), but cannot be both. this is equalent to the definition of DAG

3.1



$h_0 = 0$

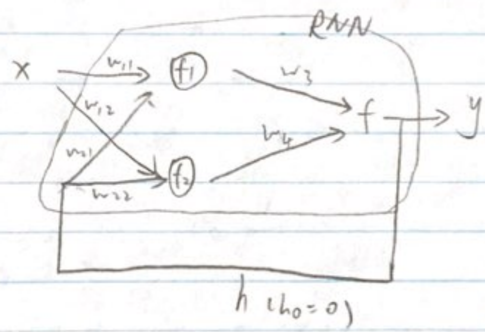
$h_t = h_{t-1}$ XOR x_t

$y_t = h_t$

XOR (exclusively or)

	x_1						
Input		0	1	0	1	1	0
h_t / y_t	0	0	1	1	0	1	1
	h_0	h_1/y_1					

more specifically.



$w_{11} = 1$ $\quad b_1 = -1$ $\quad f_1 = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$

$w_{12} = -1$

$w_{21} = -1$ $\quad b_2 = -1$ $\quad f_2 = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$

$w_{22} = 1$

$w_3 = 1$

$w_4 = 1$ $\quad b_3 = -0.1$ $\quad f = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$

3.2 $\quad h_t = W^T h_{t-1} = (W^T)^2 h_{t-2} = \dots = (W^T)^t h_0$

$$\leq \rho^t(W)\, h_0$$

If $|\rho(W)| > 1$ then h_t will exponetially grow to undefined large field, where gradient explodes

If $|\rho(W)| < 1$ then h_t will exponetially diminish to undefinell small field where gradient vanishes.