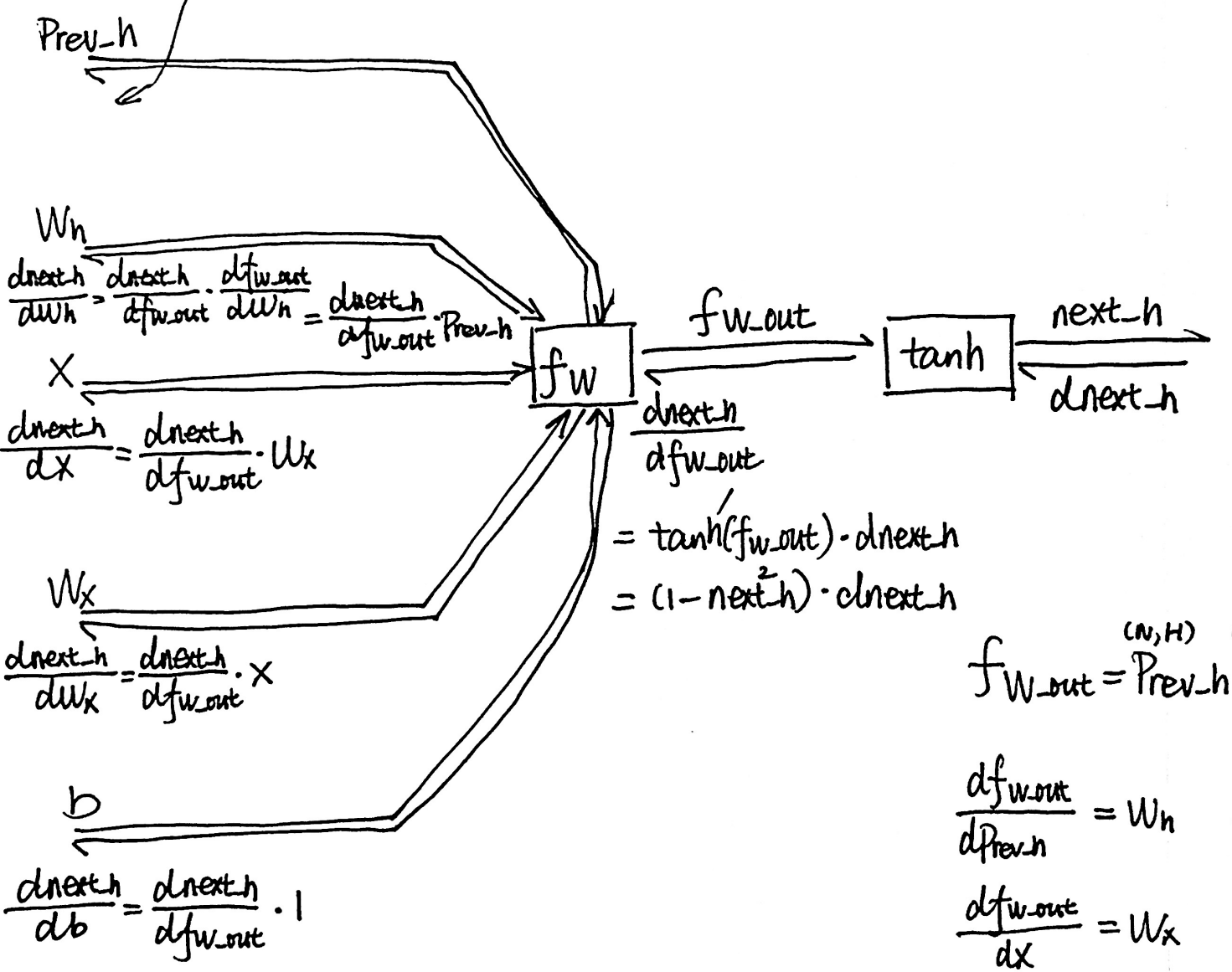
Vanilla recurrent neural networks.

$next\_h = \tanh(fw\_out)$

$\frac{d}{dx}(\tanh(x)) = 1 - \tanh^2(x)$

$\frac{dnext\_h}{dprev\_h} = \frac{dnext\_h}{dfw\_out} \cdot \frac{dfw\_out}{dprev\_h} = \frac{dnext\_h}{dfw\_out} \cdot W_h$



$\frac{dnext\_h}{dWh} = \frac{dnext\_h}{dfw\_out} \cdot \frac{dfw\_out}{dWh} = \frac{dnext\_h}{dfw\_out} \cdot Prev\_h$

$\frac{dnext\_h}{dX} = \frac{dnext\_h}{dfw\_out} \cdot W_x$

$\frac{dnext\_h}{dfw\_out} = \tanh'(fw\_out) \cdot dnext\_h$

$= (1 - next\_h^2) \cdot dnext\_h$

$\frac{dnext\_h}{dWx} = \frac{dnext\_h}{dfw\_out} \cdot X$

$\frac{dnext\_h}{db} = \frac{dnext\_h}{dfw\_out} \cdot 1$

$fw\_out = \underset{(N,H)}{Prev\_h} \cdot \underset{(H,H)}{W_h} + \underset{(N,D)}{X} \cdot \underset{(D,H)}{W_x} + \underset{(H)}{b}$

$\frac{dfw\_out}{dprev\_h} = W_h \qquad \frac{dfw\_out}{dWh} = Prev\_h$

$\frac{dfw\_out}{dX} = W_x \qquad \frac{dfw\_out}{dWx} = X$

$\frac{dfw\_out}{db} = 1$