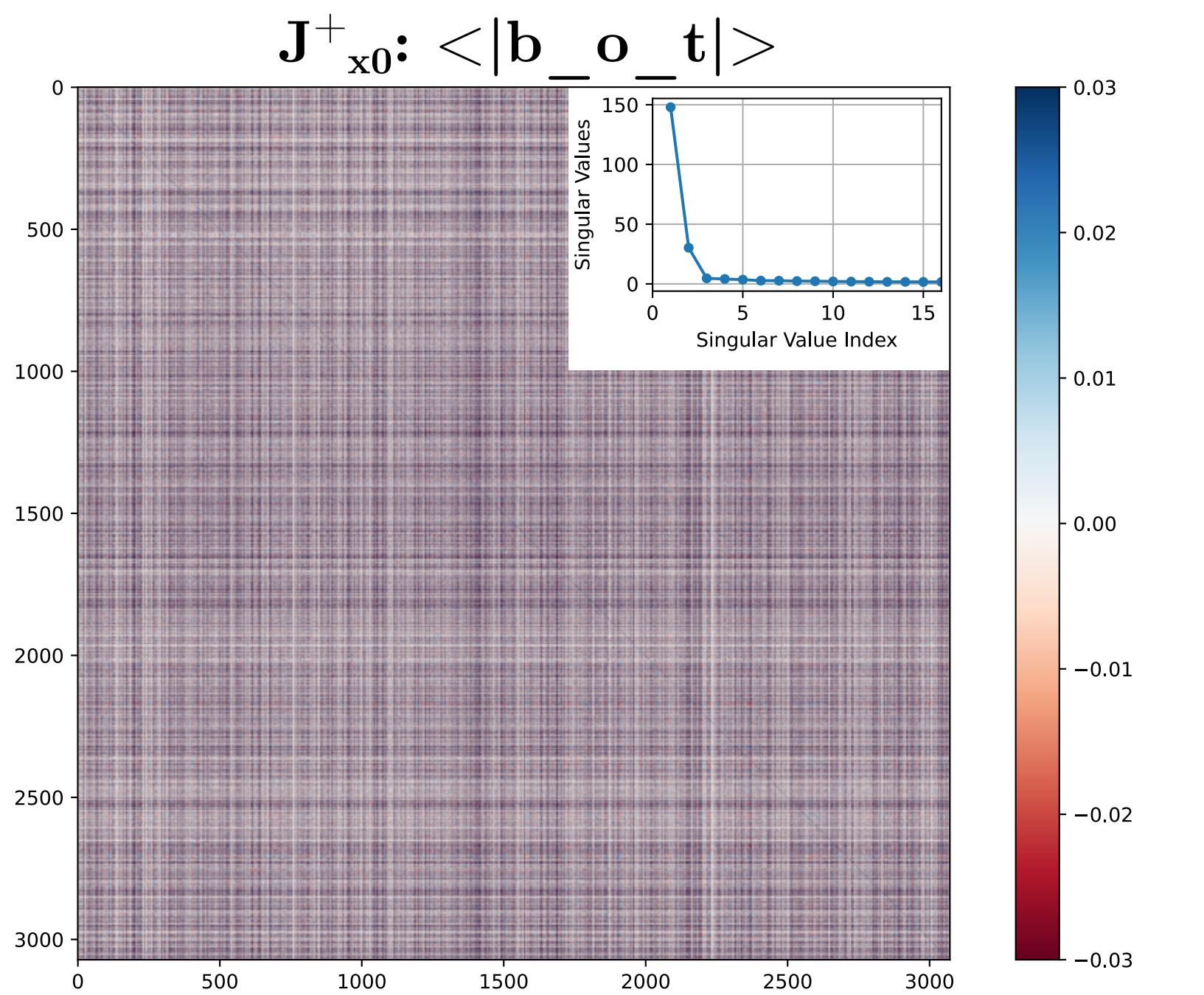


**A**

$$t_0 = <|\text{begin\_of\_text}|>, x_0 = \text{embed}(t_0), y_{T1} = \text{The}$$

$$y_1(x) = \text{model}(x_0) = J^+_{x0}(x)*x_0$$



One-token input: <|b\_o\_t|>  
Predicted token: The

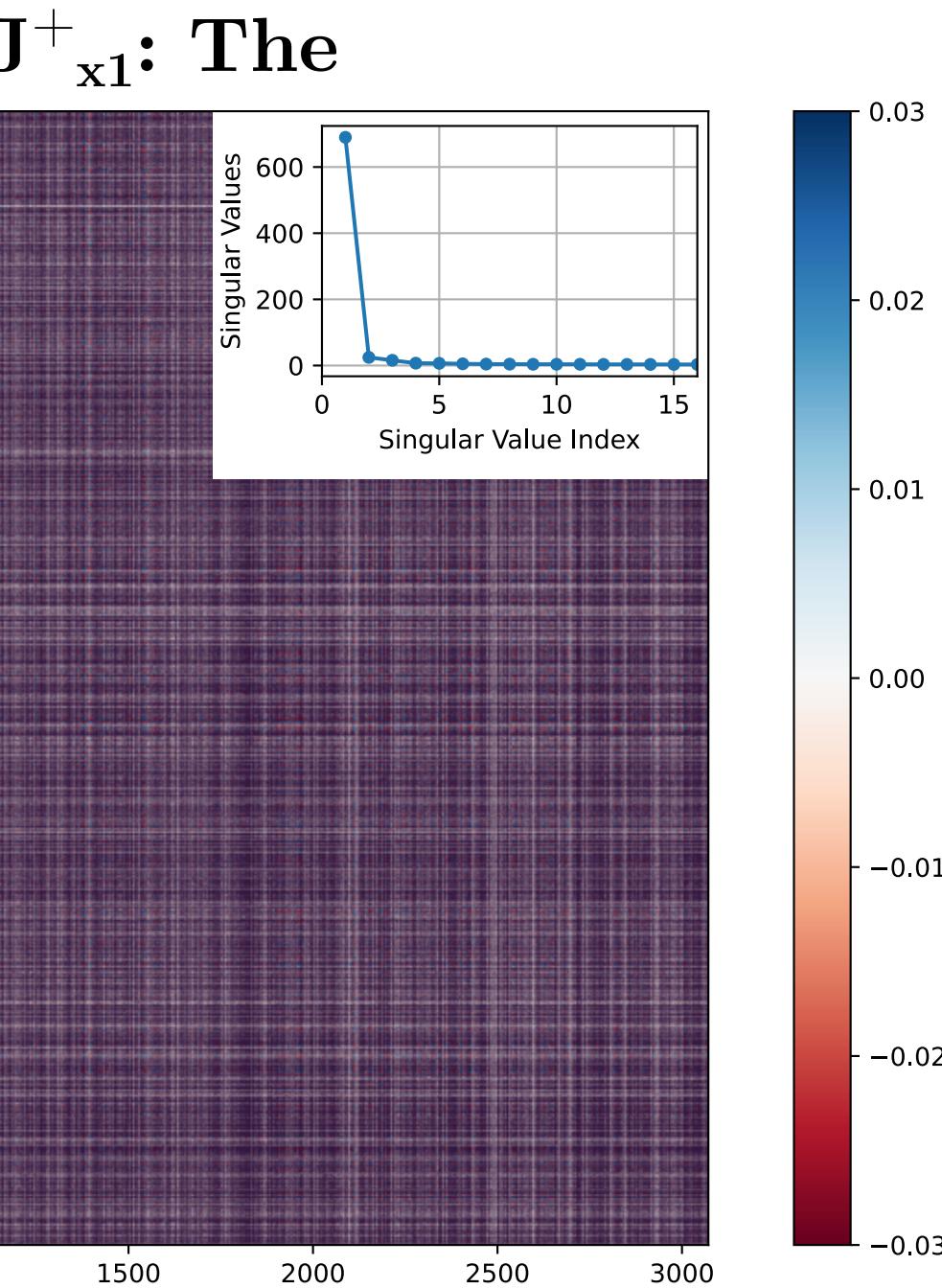
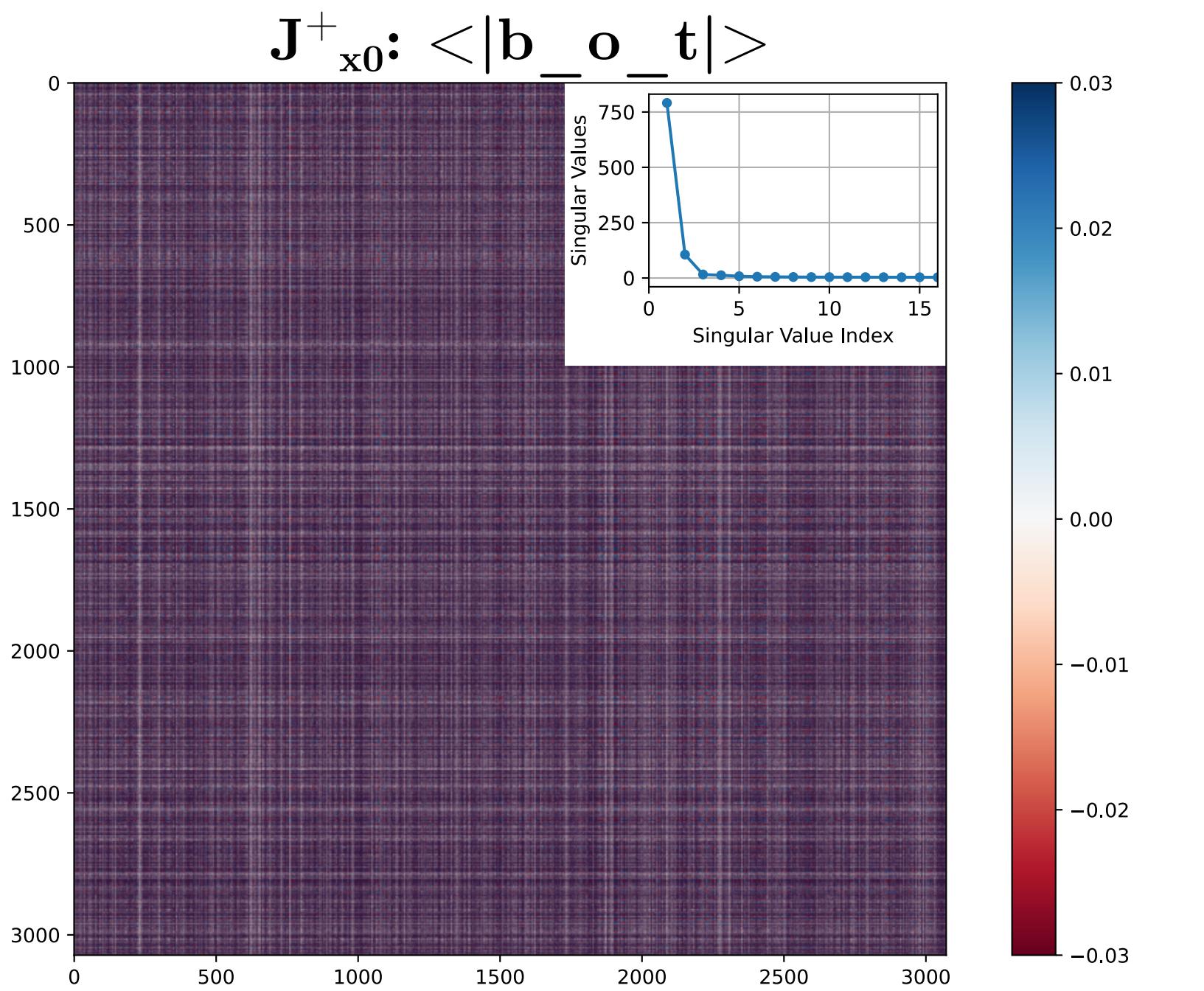
$J^+_{xi}$ : psuedo-Jacobian matrix of the i-th token

**B**

$$t_0 = <|\text{begin\_of\_text}|>, t_1 = \text{The}, (x_0, x_1) = \text{embed}(t_0, t_1), y_{T2} = \text{' '}$$

$$y_2(x) = \text{model}(x_0, x_1) = J^+_{x0}(x)*x_0 + J^+_{x1}(x)*x_1$$

Two-token input: :<|b\_o\_t|>, The  
Predicted token: : ' '

**C**

$$t_0 = <|\text{begin\_of\_text}|>, t_1 = \text{The}, t_2 = :\backslash n, (x_0, x_1, x_2) = \text{embed}(t_0, t_1, t_2), y_{T3} = 201$$

$$y_3(x) = \text{model}(x_0, x_1, x_2) = J^+_{x0}(x)*x_0 + J^+_{x1}(x)*x_1 + J^+_{x2}(x)*x_2$$

Three-token input: <|b\_o\_t|>, The, ' '  
Predicted token: 201

