

$$\sigma(x) = \frac{1}{1+e^x}$$

$$\begin{aligned} \frac{\partial \sigma}{\partial x}(x) &= \frac{\partial}{\partial x} \left( \frac{e^x}{1+e^x} \right) = \frac{(1+e^x)e^x - e^x(e^x)}{(1+e^x)^2} = \left( \frac{e^x}{1+e^x} \right) \left( \frac{1-e^x+e^x}{(1+e^x)} \right) \\ &= \sigma(x) \left( \frac{1+e^x}{1+e^x} - \frac{e^x}{(1+e^x)} \right) = \sigma(x)(1-\sigma(x)) \end{aligned}$$

$$\text{Thus } \frac{\partial \sigma(x)}{\partial x}(x) = \sigma(x)(1-\sigma(x)).$$

Objective function:

$$\arg \max_{V_c, V_w} = \Sigma_{(w,c) \in D} \log(\sigma(\langle v_c, v_w \rangle)) + \Sigma_{(w,c) v \in D'} \log(\sigma(-\langle v_c, v_w \rangle))$$

$$\nabla_{v_c} (\langle v_c, v_w \rangle)_k = \frac{\partial}{\partial (v_c)_k} \sum_i (v_c)_i (v_w)_i = \frac{\partial}{\partial (v_c)_k} \left( \sum_{i \neq k} (v_c)_i (v_w)_i + (v_c)_k (v_w)_k \right) = (v_w)_k$$

by symmetry, the gradient  $\nabla_{v_w} \langle v_c, v_w \rangle$  is  $v_c$ . Let  $B \in \mathbb{R}^{n \times m}$ .

$$\nabla_B (\langle v_c, v_w \rangle_B)_{kl} = \frac{\partial}{\partial B_{kl}} \sum_{ij} B_{ij} (v_c)_i (v_w)_j =$$

$$\frac{\partial}{\partial B_{kl}} \left( \sum_{i,j \neq k} (v_c)_i (v_w)_j + B_{kl} (v_c)_k (v_w)_l \right) = (v_c)_l (v_w)_k = (v_c \otimes v_w)_{kl}$$

Then the gradient of  $\log(\sigma(\langle v_c, v_w \rangle))$  with respect to  $v_c$  is

$$\begin{aligned} \frac{\partial}{\partial (v_c)_k} (\log(\sigma(\langle v_c, v_w \rangle))) &= \left( \frac{1}{\sigma(\langle v_c, v_w \rangle)} \right) (\sigma(\langle v_c, v_w \rangle) (1 - \sigma(\langle v_c, v_w \rangle))) (v_w)_k \\ &= (1 - \sigma(\langle v_c, v_w \rangle)) (v_w)_k \end{aligned}$$

again by symmetry we can show that the gradient with respect to  $v_w$  is similarly  $\nabla_{v_w} (\log(\sigma(\langle v_c, v_w \rangle))) = (1 - \sigma(\langle v_c, v_w \rangle)) v_c$

Let  $P \in \mathbb{R}^{n \times n}$  and  $U, V \in \mathbb{R}^{n \times d}$  and let  $f = \|P - UV^T\|_F^2$

$$\begin{aligned}
(\nabla_U f)_{ij} &= \frac{\partial}{\partial U_{ij}} \|P - UV^T\|_F^2 = \frac{\partial}{\partial U_{ij}} \left( \sum_{i_1, i_2=1}^n (P_{i_1, i_2} - \sum_{i_3=1}^d U_{i_1, i_3} V_{i_3, i_2}^T)^2 \right) \\
&= \frac{\partial}{\partial U_{ij}} \left( \left[ \sum_{i_1 \neq i, i_2=1}^n (P_{i_1, i_2} - \sum_{i_3=1}^d U_{i_1, i_3} V_{i_3, i_2}^T)^2 \right] + (P_{i, i_2} - \sum_{i_3=1}^d U_{i, i_3} V_{i_3, i_2}^T)^2 \right) \\
&= \frac{\partial}{\partial U_{ij}} \sum_{i_2=1}^n (P_{i, i_2} - \sum_{i_3=1}^d U_{i, i_3} V_{i_3, i_2}^T)^2 \\
&= \sum_{i_2=1}^n 2(P_{i, i_2} - \sum_{i_3=1}^d U_{i, i_3} V_{i_3, i_2}^T) \frac{\partial}{\partial U_{ij}} (P_{i, i_2} - \sum_{i_3=1}^d U_{i, i_3} V_{i_3, i_2}^T) \\
&= \sum_{i_2=1}^n 2(P_{i, i_2} - \sum_{i_3=1}^d U_{i, i_3} V_{i_3, i_2}^T) (V_{j, i_2}^T) \\
&= \sum_{i_2=1}^n 2(P_{i, i_2} - (UV^T)_{i, i_2}) (V_{i_2, j}) = 2(PV - UV^T V)_{ij} \\
(\nabla_V f)_{ij} &= \frac{\partial}{\partial V_{ij}} \|P - UV^T\|_F^2 = \frac{\partial}{\partial V_{ij}} \|P^T - VU^T\|_F^2
\end{aligned}$$

Now note that by the generality of the last gradient computation we can show that  $\nabla_V f = 2(P^T U - VU^T U)$ .

Consider the function  $g = \lambda \|U\|_F^2$  for  $\lambda \in \mathbb{R}$ , then the gradient with respect to  $U$  is

$$(\nabla_U g)_{ij} = \lambda \frac{\partial}{\partial U_{ij}} \left( \sum_{i_1, i_2=1}^{n, d} U_{i_1, i_2}^2 \right) = \lambda \frac{\partial}{\partial U_{ij}} \left( \left[ \sum_{i_1 \neq i, i_2 \neq j}^{n, d} U_{i_1, i_2}^2 \right] + U_{ij}^2 \right) = 2\lambda U_{ij}$$

Let  $P \in \mathbb{R}^{n \times n}$ ,  $U \in \mathbb{R}^{n \times d}$ , and  $B \in \mathbb{R}^{d \times d}$ . Consider the function  $f : \mathbb{R}^{n \times n} \times \mathbb{R}^{n \times d} \times \mathbb{R}^{n \times n} \mapsto \mathbb{R}$  such that

$$f(P, U, B) = \|P - UBB^T U^T\|_F^2$$

Letting  $C = BB^T$

$$\begin{aligned} (\nabla_U f)_{ij} &= \frac{\partial}{\partial U_{ij}} \|P - UBB^T U^T\|_F^2 = \frac{\partial}{\partial U_{ij}} \|P - UCU^T\|_F^2 = \frac{\partial}{\partial U_{ij}} \sum_{i_1, i_2}^{n, n} (P_{i_1, i_2} - (UCU^T)_{i_1, i_2})^2 \\ &= \frac{\partial}{\partial U_{ij}} \sum_{i_1, i_2}^{n, n} (P_{i_1, i_2} - \sum_{i_3}^d (U_{i_1, i_3} (CU^T)_{i_3, i_2})^2) = \frac{\partial}{\partial U_{ij}} \sum_{i_1, i_2}^{n, n} (P_{i_1, i_2} - \sum_{i_3, i_4}^{d, d} U_{i_1, i_3} C_{i_3, i_4} U_{i_4, i_2}^T)^2 \\ &= \frac{\partial}{\partial U_{ij}} ([ \sum_{\substack{i_1, i_2 \\ (i_1, i_2) \neq (i, j)}}^{n, n} (P_{i_1, i_2} - \sum_{i_3, i_4}^{d, d} U_{i_1, i_3} C_{i_3, i_4} U_{i_4, i_2}^T)^2 ] + (P_{i, j} - \sum_{i_3, i_4}^{d, d} U_{i, i_3} C_{i_3, i_4} U_{i_4, j}^T)^2) \\ &= \frac{\partial}{\partial U_{ij}} ((P_{i, j} - \sum_{i_3, i_4}^{d, d} U_{i, i_3} C_{i_3, i_4} U_{i_4, j}^T)^2) \\ &= 2(P_{i, j} - \sum_{i_3, i_4}^{d, d} U_{i, i_3} C_{i_3, i_4} U_{i_4, j}^T) \frac{\partial}{\partial U_{ij}} (P_{i, j} - \sum_{i_3, i_4}^{d, d} U_{i, i_3} C_{i_3, i_4} U_{i_4, j}^T) \\ &= 2(P_{i, j} - \sum_{i_3, i_4}^{d, d} U_{i, i_3} C_{i_3, i_4} U_{i_4, j}^T) \frac{\partial}{\partial U_{ij}} (- \sum_{i_3, i_4}^{d, d} U_{i, i_3} C_{i_3, i_4} U_{i_4, j}^T) \\ &= 2(P_{i, j} - \sum_{i_3, i_4}^{d, d} U_{i, i_3} C_{i_3, i_4} U_{i_4, j}^T) \frac{-\partial}{\partial U_{ij}} ([ \sum_{\substack{i_3, i_4 \\ i_3 \neq j}}^{d, d} U_{i, i_3} C_{i_3, i_4} U_{i_4, j}^T ] + U_{i, j} \sum_{i_4}^d C_{j, i_4} U_{i_4, j}^T) \\ &= 2(P_{i, j} - \sum_{i_3, i_4}^{d, d} U_{i, i_3} C_{i_3, i_4} U_{i_4, j}^T) \frac{-\partial}{\partial U_{ij}} (U_{i, j} \sum_{i_4}^d C_{j, i_4} U_{i_4, j}^T) \\ &= -2(P_{i, j} - \sum_{i_3, i_4}^{d, d} U_{i, i_3} C_{i_3, i_4} U_{i_4, j}^T) (\sum_{i_4}^d C_{j, i_4} U_{i_4, j}^T) \\ &= -2(P_{i, j} - (UCU^T)_{i, j}) (CU^T)_{j, j} = -2(P_{i, j} - (UBB^T U^T)_{i, j}) (BB^T U^T)_{j, j} \end{aligned}$$