Proof. Let us consider the partial derivative w.r.t $w_{i,j}^{(l)}$. For this let $I_{i,j}^{(l)}$, $A_j^{(l+1)}$ and $B_i^{(l-1)}$ be 3 matrices of size $W^{(l)}, W^{(l+1)}$ and $W^{(l-1)}$ respectively such that $I_{i,j}^{(l)}[i,i] = 1$ and reset all entries are 0, $A_j^{(l+1)}[k,j] = W^{(l+1)}[k,j], \forall k$ and rest all entries are 0 and $B_i^{(l-1)}[i,k] = W^{(l-1)}[i,k], \forall k$ and rest all entries are one. Using these 3 matrices and the weight matrices we can compute the gradient as

$$\frac{\partial \phi(\mathbf{w}, \mathbf{x})}{\partial w_{i,j}^{(l)}} = W^{(H)} \cdots W^{(l+2)} \cdot A_j^{(l+1)} \cdot I_{i,j}^{(l)} \cdot B_i^{(l-1)} \cdot W^{l-2} \cdots W^1 \cdot \|x\| \tag{7}$$

Let M'(l,i,j) be a matrix such that

$$M'_{l,i,j} = A_j^{(l+1)} \cdot I_{i,j}^{(l)} \cdot B_i^{(l-1)}$$

Although we have scalar values taking spectral norm on both the sides of eq 7 we get

$$\left| \frac{\partial \phi(\mathbf{w}, \mathbf{x})}{\partial w_{i,j}^{(l)}} \right| = \prod_{k=1}^{H} \|W^{(k)}\|_\sigma \frac{\|M'_{l,i,j}\|_\sigma}{\|W^{(l+1)}\|_\sigma \cdot \|W^{(l)}\|_\sigma \cdot \|W^{(l-1)}\|_\sigma} \|x\|$$

Now lets define another matrix M_l such that $(p,q)^{th}$ element of matrix $M_l[p,q] = \|M'_{l,i,j}\|_\sigma$. Now the expression for 2 norm of the gradient vector directly gives us the required expression $\square$