Problem 1.a. Replace all $\vec{X}_i$ with $\alpha\vec{X}_i$, for an $\alpha \in \mathbb{R}^+/\{0\}$.

Solution:

$$\begin{aligned}
E(W) &= \sum_i \left|\alpha X_i - \sum_j W_{ij}\alpha\vec{X}_j\right|^2 \\
&= \sum_i \left|\alpha\left(\vec{X}_i - \sum_j W_{ij}\vec{X}_j\right)\right|^2 \\
&= \sum_i \alpha^2 \left|\vec{X}_i - \sum_j W_{ij}\vec{X}_j\right|^2 \\
&= \alpha^2 \sum_i \left|\vec{X}_i - \sum_j W_{ij}\vec{X}_j\right|^2
\end{aligned}$$

, which is the original problem scaled by some constant, thus it is minimized by the same W as original problem. □

Problem 1.b. Replace all $\vec{X}_i$ with $\vec{X}_i + \vec{v}$, where $\vec{v} \in \mathbb{R}^D$.

Solution:

$$\begin{aligned}
E(W) &= \sum_i \left|\vec{X}_i + \vec{v} - \sum_j W_{ij}\left(\vec{X}_j + \vec{v}\right)\right|^2 \\
&= \sum_i \left|\vec{X}_i + \vec{v} - \sum_j W_{ij}\vec{X}_j + W_{ij}\vec{v}\right|^2 \\
&= \sum_i \left|\vec{X}_i + \vec{v} - \sum_j W_{ij}\vec{X}_j - \vec{v}\sum_j W_{ij}\right|^2 \\
&= \sum_i \left|\vec{X}_i + \vec{v} - \sum_j W_{ij}\vec{X}_j - \vec{v}\right|^2 \\
&= \sum_i \left|X_i - \sum_j W_{ij}\vec{X}_j\right|^2
\end{aligned}$$

, which is the original problem.

□

Problem 1.c. Replace all $\vec{X}_i$ with $U \cdot \vec{X}_i$, where U is an orthogonal $D \times D$ matrix (this additionally includes mirror symmetries

Solution:

$$\begin{aligned}
E\left(W\right) &= \sum_i \left| U \cdot \vec{X}_i - \sum_j W_{ij} U \cdot \vec{X}_j \right|^2 \\
&= \sum_i \left| U \cdot \vec{X}_i - U \cdot \sum_j W_{ij} \vec{X}_j \right|^2 \\
&= \sum_i \left| U \cdot \left(\vec{X}_i - \sum_j W_{ij} \vec{X}_j \right) \right|^2 \\
&= \sum_i ||U||^2 \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2 \\
&= \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2
\end{aligned}$$

, where we have made use of the associative property of matrix multiplication and the fact that for orthogonal matrices $||U||^2 = U \cdot U^T = I$, with I the identity matrix. □

Problem 2.a.

Solution: Rewrite the optimization problem as:

$$\begin{aligned}
E &= \left| \mathbf{1} \vec{X}^\top - \sum_j W_j \vec{\eta}_j \right|^2 \\
&= \left| \sum_j W_j \mathbf{1} \vec{X}^\top - \sum_j W_j \vec{\eta}_j \right|^2 \\
&= \left(W^\top \mathbf{1} \vec{X}^\top - W^\top \eta \right)^2 \\
&= \left(W^\top \left(\mathbf{1} \vec{X}^\top - \eta \right) \right)^2 \\
&= W^\top \left(\mathbf{1} \vec{X}^\top - \vec{\eta} \right) \left(\mathbf{1} \vec{X}^\top - \vec{\eta} \right)^\top W = W^\top C W
\end{aligned}$$

, $W^\top \mathbf{1} = 1$ by definition of w. □

Problem 2.b.

Solution:
$\mathcal{L}(w,\lambda) = w^\top C w - \lambda\left(w^\top \mathbf{1} - 1\right)$
Where does $\nabla\mathcal{L}(w,\lambda) = \left[\frac{\partial}{\partial w}\mathcal{L}(w,\lambda), \frac{\partial}{\partial \lambda}\mathcal{L}(w,\lambda)\right] = [0,0]$?

$$\begin{cases} \frac{\partial}{\partial w}\mathcal{L}(w,\lambda) = 2w^\top C - \lambda\mathbf{1} = 0 \rightarrow w^\top = \frac{\lambda}{2}\mathbf{1}C^{-1} \rightarrow w = \frac{\lambda}{2}C^{-1}\mathbf{1} \\ \frac{\partial}{\partial \lambda}\mathcal{L}(w,\lambda) = -w^\top\mathbf{1} + 1 = 0 \end{cases}$$

$$\begin{cases} w^\top = \frac{\lambda}{2}\mathbf{1}C^{-1} \\ -w^\top\mathbf{1} + 1 = 0 \end{cases} \rightarrow \lambda = 2\mathbf{1}C\mathbf{1}^\top$$

$$\begin{cases} w = \frac{\lambda}{2}C^{-1}\mathbf{1} \\ \lambda = 2\mathbf{1}C\mathbf{1}^\top \end{cases}$$

Finally,

$$w = \mathbf{1}C\mathbf{1}^\top C^{-1}\mathbf{1} = \frac{C^{-1}\mathbf{1}}{\left(\mathbf{1}C\mathbf{1}^\top\right)^{-1}} = \frac{C^{-1}\mathbf{1}}{\mathbf{1}^\top C^{-1}\mathbf{1}}$$

□

Problem 2.c. Show that the minimum w can be equivalently found by solving the equation

$$Cw = 1,$$

and then rescaling w such that $w^T\mathbf{1} = 1$.

Solution: Solving $Cw = \mathbf{1}$ gives $w = C^{-1}\mathbf{1}$. For any minimal w we have

$$w^T\mathbf{1} = \mathbf{1}^T w = \mathbf{1}^T C^{-1}\mathbf{1}.$$

Dividing by this constant gives the desired constrained minimum. □

Problem 3.a.

Solution: we have:

$$C = D_{KL}(p \parallel q) = \sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}\log(\frac{p_{ij}}{q_{ij}})$$
$$= \sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}\log(p_{ij}) - p_{ij}\log(q_{ij})$$

The partial derivative of C with respect to q_{ij} is given by:

$$\frac{\partial C}{\partial q_{ij}} = -p_{ij}.\frac{\partial}{\partial q_{ij}}(\log q_{ij}) = -\frac{p_{ij}}{q_{ij}}$$

□

Problem 3.b. The probability matrix q is now reparameterized as:

$$q_{ij} = \frac{exp(z_{ij})}{\sum_{k=1}^{N}\sum_{l=1}^{N} exp(z_{kl})}$$

let:

$$S = \sum_{k=1}^{N}\sum_{l=1}^{N} exp(z_{kl}) = \sum_{k=1}^{N}\sum_{l=1}^{N} w_{kl} \qquad and \qquad w_{ij} = exp(z_{ij})$$

leads to:

$$q_{ij} = \frac{w_{ij}}{S}$$

The partial derivative of q_{ij} with respect to w_{ij} is given by:

$$\frac{\partial q_{ij}}{\partial w_{ij}} = \frac{S - w_{ij}}{S^2} = \frac{1}{S} - \frac{q_{ij}}{S}$$

we apply the chain rule for partial derivatives:

$$\begin{aligned}
\frac{\partial C}{\partial z_{ij}} &= \frac{\partial C}{\partial q_{ij}} \times \frac{\partial q_{ij}}{\partial w_{ij}} \times \frac{\partial w_{ij}}{\partial z_{ij}} \\
&= -\frac{p_{ij}}{q_{ij}} \times \left(\frac{S - w_{ij}}{S^2}\right) \times w_{ij} \\
&= -\frac{p_{ij}}{q_{ij}} \times \left(\frac{S - w_{ij}}{S^2}\right) \times (S \times q_{ij}) \\
&= -p_{ij} \times \left(\frac{S - w_{ij}}{S}\right) \\
&= -p_{ij}\left(1 - \frac{w_{ij}}{S}\right) \\
&= -p_{ij}\left(1 - \frac{q_{ij}}{p_{ij}}\right) \\
&= -p_{ij} + q_{ij}
\end{aligned}$$

Since the $\sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}$=1 which gives: $\frac{w_{ij}}{S} = \frac{q_{ij}}{p_{ij}}$

Problem 3.d. The scores z_{ij} is given by:

$$z_{ij} = -\|y_i - y_j\|^2 \qquad (i.e \quad z_{ji} = -\|y_j - y_i\|^2)$$

Now we use the chain rule for derivatives which gives:

$$\frac{\partial C}{\partial y_i} = \sum_j \frac{\partial C}{\partial z_{ij}} \times \frac{\partial z_{ij}}{\partial y_i} + \sum_j \frac{\partial C}{\partial z_{ji}} \times \frac{\partial z_{ji}}{\partial y_i}$$

We start with the first term:

$$\begin{aligned}
\frac{\partial C}{\partial z_{ij}} &= \sum_j \frac{\partial C}{\partial z_{ij}}\Big(-2(y_i - y_j)\Big) + \sum_j \frac{\partial C}{\partial z_{ji}}\Big(+2(y_j - y_i)\Big) \\
&= -2\Bigg(\sum_j \frac{\partial C}{\partial z_{ij}}(y_i - y_j) + \sum_j \frac{\partial C}{\partial z_{ji}}(y_i - y_j)\Bigg) \\
&= -2(y_i - y_j)\Bigg(\sum_j \frac{\partial C}{\partial z_{ij}} + \sum_j \frac{\partial C}{\partial z_{ji}}\Bigg)
\end{aligned}$$

and using(3.b) which gives:

$$\frac{\partial C}{\partial z_{ij}} = -p_{ij} + q_{ij} \qquad and \qquad \frac{\partial C}{\partial z_{ji}} = -p_{ji} + q_{ji}$$

Now we replace in our equation:

$$\frac{\partial C}{\partial z_{ij}} = -2(y_i - y_j)\Big(-p_{ij} + q_{ij} - p_{ji} + q_{ji}\Big)$$

For SSNE, Both the P and Q matrices are symmetric, so $p_{ij} = p_{ji}$ and $q_{ij} = q_{ji}$ leading to:

$$\begin{aligned}
\frac{\partial C}{\partial z_{ij}} &= -2\sum_j (y_i - y_j)\Big(-2p_{ij} + 2q_{ij}\Big) \\
&= -4\sum_j (y_i - y_j)\Big(-p_{ij} + q_{ij}\Big) \\
&= 4\sum_j (y_i - y_j)\Big(p_{ij} - q_{ij}\Big)
\end{aligned}$$

Solution:

$$\frac{\partial C}{\partial z_{ij}} = 4\sum_j (y_i - y_j)\Big(p_{ij} - q_{ij}\Big)$$

□