Q 1) (PCA) $f(z,w) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{d}(w_i^T z_j - x_{ij})^2$

(Robust PCA)

$$f(z,w) = \sum_{i=1}^{n}\sum_{j=1}^{d}|w_i^T z_j - x_{ij}|$$

Use the approximation to use gradient descent for robust PCA

$\forall x: |x| \approx \sqrt{x^2+\varepsilon}$ (smooth)

if $r_{ij} = w_i^T z_j - x_{ij}$

$$\Rightarrow f(z,w) = \sum_{i=1}^{n}\sum_{j=1}^{d}\sqrt{r_{ij}^2+\varepsilon}$$

Need to modify PCAObjW, PCAObjZ to output the correct function value & gradient

for gradient

$$\frac{\partial f}{\partial z_i} = \sum_{i=1}^{n}\sum_{j=1}^{d}\left[\frac{\partial\left[\sqrt{r_{ij}^2+\varepsilon}\right]}{\partial(z_i)}\right]$$

$$= \sum_{j=1}^{d}\left(\frac{\partial\sqrt{r_{ij}^2+\varepsilon}}{\partial(r_{ij})}\right)\cdot\left(\frac{\partial(r_{ij})}{\partial z_i}\right) \quad \text{[Chain Rule]}$$

Same as PCA → implemented in the current code in PCA gradient

Same thing wrt w_j

[$f(z,w)$ is symmetric in w, z]

Q2) Use geodesic distance instead of L_2 distance

① $\forall i, j$: compute $D(i,j) = \|x_i - x_j\|_2$
[same as before]

② Construct a graph s.t $A \rightarrow$ adjacency matrix
$A \rightarrow$ Edge from $i \rightarrow j$ if j is a k-nearest neighbour of i wrt the distance in D

③ $\rightarrow$ Need a distance metric G for Dijkstra code
$\forall i, j$: compute the shortest distance d_{ij} from $i \rightarrow j$ on the graph.
Input to dijkstra code $\rightarrow$ matrix G.

$G_{ij} = \infty$ if i is not connected to j

else $G_{ij} = D_{ij}$

④ Use distance d_{ij} [output from Dijkstra] instead of D_{ij} in the MDS code.

Subtleties

$\rightarrow$ G needs to be symmetric (graph needs to be undirected)

$\rightarrow$ Don't include the point i in its nearest neighbours $\Rightarrow D(i,i) = \infty$

Q2.2 if graph is disconnected there might be (i,j) s.t $d_{ij} = \infty \xrightarrow[\text{case}]{\text{in this}}$ just $d_{ij} \rightarrow$ maximum finite distance across (i,j)