1 Landmark MDS:

Landmark MDS is based on the MDS algorithm. We can first have a quick look at how MDS works. For a given distance matrix $D^{(x)}$ we need to find $k = -\frac{1}{2}HD^XH$ where $H = I - \frac{1}{n}ee^T$. From this we can find that $K = X^TX = V\Lambda V^T$. The low-dimensional map of $Y$ will be:

$$Y = \Lambda_d^{1/2}V_d$$

The problem with the MDS algorithm is that the matrices $D^X$ and $K$ are not sparse. It is therefore expensive to compute eigen-decompositions. To reduce the computational work required we can use Landmark MDS which is equivalent to the Nyström approximation.

2 Nyström Approximation

Suppose we have $n$ data points from which we can choose $m$ data points randomly from the sets $D^X$ and $K$. Without loss of generality we can permute these points so that they represent the first $m$ points in $D^X$ and $K$. 

Consider the matrices:

\[ K = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \]

\[ D^X = \begin{pmatrix} E & F \\ F^T & G \end{pmatrix} \]

Where \( A \) is a known \( m \) by \( m \) matrix and \( B \) is a known \( m \) by \( n - m \) matrix. The idea is to estimate the unknown \( n - m \) by \( n - m \) matrix \( C \). If \( K \) is a positive semi-definite matrix then it is a Gram matrix. It can then be expressed as an inner product:

\[ K = X^T X = V \Lambda V^T \]

Initially \( A = R^T R \). After we apply MDS we get \( R = \Gamma^{1/2} U^T \). Also, \( B = R^T S \). After we apply MDS we get \( S = R^{-T} B \). We can rewrite the equation for \( R \) as:

\[ R^T = U \Gamma^{1/2} \]

And then:

\[ R^{-T} = \Gamma^{-1/2} U^T \]

Then we can substitute that back into the earlier equation for:

\[ S = \Gamma^{-1/2} U^T B \]

To estimate \( C \) we need to recognize that \( C = S^T S \). So from the above equation for \( S \) we
get an expression for an estimate for $C$:

$$
C = S^T S
= B^T U \Gamma^{-1/2} \Gamma^{-1/2} U^T B
= B^T R^{-1} R^{-T} B
= B^T A^{-1} B
$$

So then we can estimate $C$ by first finding $A$ and $B$ and then we can complete the matrix in the following way. Nyström approximation approximate $K$ as:

$$
\hat{K} = \begin{pmatrix}
A & B \\
B^T & B^T A^{-1} B
\end{pmatrix}
$$