Lecture Notes 7: EM algorithm and Multidimensional Scaling

Professor: Zhihua Zhang *Scribe:Yao Cheng, Yuchen Sha*

6 Expectation-Maximization(EM algorithm)

6.5 Convergence of EM algorithm

Suppose $\theta^{(t)}$ and $\theta^{(t+1)}$ are the parameters from two successive iterations of EM. We will now prove that $L(\theta^{(t)}) \leq L(\theta^{(t+1)})$,that is,$log\, p(x|\theta^{(t)}) \leq log\, p(x|\theta^{(t+1)})$.We know

$$Q(\theta|\theta^{(t)}) = \textstyle\int (log\, p(x, z|\theta))(p(z|x, \theta^{(t)}))dz.$$

Let $H(\theta|\theta^{(t)}) = log\, p(x|\theta) - Q(\theta|\theta^{(t)})$,then

$$\begin{aligned}
H(\theta|\theta^{(t)}) &= log\, p(x|\theta) - Q(\theta|\theta^{(t)}) \\
&= log\, p(x|\theta) \int p(z|x, \theta^{(t)})dz - \int (log\, p(x, z|\theta))(p(z|x, \theta^{(t)}))dz \\
&(p(z|x, \theta^{(t)}) \text{ is a distribution of z}) \\
&= \int log\, p(x|\theta)p(z|x, \theta^{(t)})dz - \int (log\, p(x, z|\theta))(p(z|x, \theta^{(t)}))dz \\
&= \int log\, \frac{p(x|\theta)}{p(x, z|\theta)} p(z|x, \theta(t))dz \\
&= -\int log\, p(z|x, \theta)p(z|x, \theta^{(t)})dz
\end{aligned}$$

Thus

$$log\, p(x|\theta^{(t+1)}) - log\, p(x|\theta^{(t)}) = Q(\theta|\theta^{(t+1)}) + H(\theta|\theta^{(t+1)}) - (Q(\theta|\theta^{(t)}) + H(\theta|\theta^{(t)}))$$

Since

$$\theta^{(t+1)} = \arg\max_{\theta} Q(\theta|\theta^{(t)})$$

We have

$$Q(\theta^{(t+1)}|\theta^{(t)}) \geq Q(\theta^{(t)}|\theta^{(t)})$$

Then we prove

$$H(\theta^{(t+1)}|\theta^{(t)}) \geq H(\theta^{(t)}|\theta^{(t)})$$

$$
\begin{aligned}
H(\theta^{(t+1)}|\theta^{(t)}) - H(\theta^{(t)}|\theta^{(t)}) &= -\int log\, p(z|x,\theta^{(t+1)})p(z|x,\theta^{(t)})dz + \int log\, p(z|x,\theta^{(t)})p(z|x,\theta^{(t)}) \\
&= \int log\, \frac{p(z|x,\theta^{(t)})}{p(z|x,\theta^{(t+1)})} p(z|x,\theta^{t})dz \\
&= -\int log\, \frac{p(z|x,\theta^{(t+1)})}{p(z|x,\theta^{(t)})} p(z|x,\theta^{t})dz \\
&\geq -log \int \frac{p(z|x,\theta^{(t+1)})}{p(z|x,\theta^{(t)})} p(z|x,\theta^{t})dz \\
&= 0.
\end{aligned}
$$

Hence,

$$log\, p(x|\theta^{(t+1)}) \geq log\, p(x|\theta^{(t)})$$

EM causes the likelihood to converge monotonically.

6.6 Equivalence of EM algorithm and MLE

To prove the correctness of EM algorithm, we now show the equivalence of EM algorithm and MLE when solving probabilistic PCA.
Based on the previous notes, we have the following results. Please refer to *Lecture Notes 6* for detailed proof if necessary.
1. The estimation of $\mathbf{W}$ and τ derived by maximum likelihood estimation(MLE) is:

$$\tau = \frac{1}{p-q} \sum_{j=q+1}^{p} \mathbf{\Gamma}_j \tag{1}$$

$$\mathbf{W} = \mathbf{\Phi}_q(\mathbf{\Gamma}_q - \tau \mathbf{I}_q)^{\frac{1}{2}} \mathbf{V}^T \tag{2}$$

2.The iterative formulas for $\mathbf{W}$ and τ derived by EM algorithm is:

$$\mathbf{W}^{(t+1)} = (\sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})\langle \mathbf{z}_i^T \rangle)(\sum_{i=1}^{n} \langle \mathbf{z}_i, \mathbf{z}_i^T \rangle)^{-1} \tag{3}$$

$$\tau^{(t+1)} = \frac{1}{p}[tr(\mathbf{S}) - \frac{1}{n} \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})^T \mathbf{W}^{(t+1)} \langle \mathbf{z}_i \rangle] \tag{4}$$

where

$$\langle \mathbf{z}_i \rangle = \mathbf{M}_{(t)}^{-1} \mathbf{W}_{(t)}^T (\mathbf{x}_i - \boldsymbol{\mu}) \tag{5}$$

$$
\begin{aligned}
\langle \mathbf{z}_i, \mathbf{z}_i^T \rangle &= \tau_{(t)} \mathbf{M}_{(t)}^{-1} + \langle \mathbf{z}_i \rangle \langle \mathbf{z}_i \rangle^T \\
&= \tau_{(t)} \mathbf{M}_{(t)}^{-1} + \mathbf{M}_{(t)}^{-1} \mathbf{W}_{(t)}^T (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \mathbf{W}(t) \mathbf{M}_{(t)}^{-1}
\end{aligned} \tag{6}
$$

Let $\mathbf{W}$ converges to $\hat{\mathbf{W}}$, τ converges to $\hat{\tau}$. Substitute $\langle \mathbf{z}_i \rangle$ and $\langle \mathbf{z}_i, \mathbf{z}_i^T \rangle$ with the above formulas and we get

$$\hat{\mathbf{W}} = \mathbf{S}\hat{\mathbf{W}}(\tau \mathbf{I}_q + \mathbf{M}^{-1} \hat{\mathbf{W}}^T \mathbf{S} \hat{\mathbf{W}})^{-1} \tag{7}$$

$$\hat{\tau} = \frac{1}{p}(tr(\mathbf{S}) - tr(\mathbf{S}\hat{\mathbf{W}}\mathbf{M}^{-1}\hat{\mathbf{W}}^T)) \tag{8}$$

Here we just omit the hat on $\mathbf{W}$ and τ.Then according to equation(7),

$$
\begin{aligned}
\mathbf{W}(\tau\mathbf{I}_q + \mathbf{M}^{-1}\mathbf{W}^T\mathbf{S}\mathbf{W}) &= \mathbf{S}\mathbf{W} \\
\tau\mathbf{W} + \mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T\mathbf{S}\mathbf{W} &= \mathbf{S}\mathbf{W} \\
\mathbf{W}\mathbf{M}^{-1}\mathbf{W}^T\mathbf{S}\mathbf{W} &= (\mathbf{S} - \tau\mathbf{I}_p)\mathbf{W} \\
\mathbf{W}(\tau\mathbf{I}_q + \mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{S}\mathbf{W} &= (\mathbf{S} - \tau\mathbf{I}_p)\mathbf{W} \\
(\tau\mathbf{I}_p + \mathbf{W}\mathbf{W}^T)^{-1}\mathbf{W}\mathbf{W}^T\mathbf{S}\mathbf{W} &= (\mathbf{S} - \tau\mathbf{I}_p)\mathbf{W} \\
(\tau\mathbf{I}_p + \mathbf{W}\mathbf{W}^T)(\mathbf{S} - \tau\mathbf{I}_p)\mathbf{W} &= \mathbf{W}\mathbf{W}^T\mathbf{S}\mathbf{W} \\
(\tau\mathbf{S} - \tau^2\mathbf{I}_p + \mathbf{W}\mathbf{W}^T\mathbf{S} - \tau\mathbf{W}\mathbf{W}^T)\mathbf{W} &= \mathbf{W}\mathbf{W}^T\mathbf{S}\mathbf{W} \\
\tau\mathbf{S}\mathbf{W} - \tau^2\mathbf{W} + \mathbf{W}\mathbf{W}^T\mathbf{S}\mathbf{W} - \tau\mathbf{W}\mathbf{W}^T\mathbf{W} &= \mathbf{W}\mathbf{W}^T\mathbf{S}\mathbf{W} \\
\tau\mathbf{S}\mathbf{W} &= \tau\mathbf{W}\mathbf{W}^T\mathbf{W} + \tau^2\mathbf{W} \\
\mathbf{S}\mathbf{W} &= \mathbf{W}\mathbf{W}^T\mathbf{W} + \tau\mathbf{W} \\
\mathbf{S}\mathbf{W} &= \mathbf{W}(\mathbf{W}^T\mathbf{W} + \tau\mathbf{I}_q) \\
\mathbf{S}\mathbf{W} &= \mathbf{W}(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T + \tau\mathbf{V}\mathbf{V}^T) \\
&\quad (\text{Note: } \mathbf{W}^T\mathbf{W} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T) \\
\mathbf{S}\mathbf{W} &= \mathbf{W}\mathbf{V}(\mathbf{\Lambda} + \tau\mathbf{I}_q)\mathbf{V}^T \\
\mathbf{S}\mathbf{W}\mathbf{V} &= \mathbf{W}\mathbf{V}(\mathbf{\Lambda} + \tau\mathbf{I}_q) \\
\mathbf{S}\mathbf{W}\mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}} &= \mathbf{W}\mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}}(\mathbf{\Lambda} + \tau\mathbf{I}_q)
\end{aligned}
$$

Moreover,

$$\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{V}^T\mathbf{W}^T\mathbf{W}\mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}} = \mathbf{I}$$

Thus, $\mathbf{\Lambda} + \tau\mathbf{I}_q$ is composed of eigenvalues of S and $\mathbf{W}\mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}}$ is composed of a set of orthonormal eigenvectors.
Let $\mathbf{\Phi}_q = \mathbf{W}\mathbf{V}\mathbf{\Lambda}^{-\frac{1}{2}}, \mathbf{\Gamma} = \mathbf{\Lambda} + \tau\mathbf{I}_q$. Together we have

$$
\begin{aligned}
\mathbf{S}\mathbf{\Phi}_q &= \mathbf{\Phi}_q\mathbf{\Gamma}_q \\
\mathbf{W} &= \mathbf{\Phi}_q(\mathbf{\Gamma}_q - \tau\mathbf{I}_q)^{\frac{1}{2}}\mathbf{V}^T
\end{aligned}
$$

τ can be derived similarly. It can be seen that the formulas are the same as those derived using maximum likelihood estimation.

7 Multidimensional Scaling

Definition 7.1 *A distance matrix* $\mathbf{D}$ *is called Euclidean if there exists a configuration of points in some Euclidean space where interpoint distances are given by* $\mathbf{D}$*, that is, if for some p,there exist points* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{R}^p$ *such that* $d_{rs}^2 = ||\mathbf{x}_r - \mathbf{x}_s||^2 = (\mathbf{x}_r - \mathbf{x}_s)^T(\mathbf{x}_r - \mathbf{x}_s)$.

Theorem 7.1 *Let* $\mathbf{D}$ *be a distance matrix and define* $\mathbf{B} = \mathbf{HAH}$*, where* $\mathbf{A} = [a_{rs}]$ *where* $a_{rs} = -\frac{1}{2}d_{rs}^2$*.Then D is Euclidean iff B is p.s.d.*

Proof:
(a)Firstly, we prove that if $\mathbf{D}$ is a distance matrix, then $\mathbf{B} = \mathbf{HAH}$ as defined is *p.s.d.*
Suppose $\mathbf{D}$ is the matrix of Euclidean interpoint distances for a configuration $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)^T$.
Then $\bar{\mathbf{z}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i$.

$$\begin{aligned}
\mathbf{B} &= \mathbf{HAH} \\
&= (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n I_n^T) A (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T) \\
&= \mathbf{A} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\mathbf{A} - \frac{1}{n}\mathbf{A}\mathbf{1}_n\mathbf{1}_n^T + \frac{1}{n^2}\mathbf{1}_n\mathbf{1}_n^T A \mathbf{1}_n\mathbf{1}_n^T
\end{aligned}$$

Hence, we have,

$$b_{rs} = a_{rs} - \bar{a}_{.s} - \bar{a}_{r.} + \bar{a}$$

Here $\bar{a}_{.s}$ represent the mean of sth column of $\mathbf{A}$, $\bar{a}_{r.}$ represents the mean of rth row of $\mathbf{A}$, $\bar{a}$ represents the mean of $\mathbf{A}$.
Thus,

$$b_{rs} = -\frac{1}{2}||\mathbf{z}_r - \mathbf{z}_s||^2 + \frac{1}{2n}\sum_{i=1}^{n} d_{ri}^2 + \frac{1}{2n}\sum_{i=1}^{n} d_{is}^2 - \frac{1}{n^2}\sum_{i,j} d_{i,j}^2$$

We calculate the formula part by part:
(1)

$$-\frac{1}{2}||\mathbf{z}_r - \mathbf{z}_s||^2 = -\frac{1}{2}\mathbf{z}_r^T\mathbf{z}_r - \frac{1}{2}\mathbf{z}_s\mathbf{z}_s^T + \mathbf{z}_r^T\mathbf{z}_s;$$

(2)

$$\frac{1}{2n}\sum_{i=1}^{n} d_{ri}^2 = \frac{1}{2n}\sum_{i=1}^{n}(\mathbf{z}_r - \mathbf{z}_i)^T(\mathbf{z}_r - \mathbf{z}_i) = \frac{1}{2}\mathbf{z}_r^T\mathbf{z}_r + \frac{1}{2n}\sum_{i}^{n}\mathbf{z}_i^T\mathbf{z}_i - \mathbf{z}_r^T\bar{\mathbf{z}};$$

(3)

$$\frac{1}{2n}\sum_{i=1}^{n} d_{is}^2 = \frac{1}{2n}\sum_{i=1}^{n}(\mathbf{z}_i - \mathbf{z}_s)^T(\mathbf{z}_i - \mathbf{z}_s) = \frac{1}{2}\mathbf{z}_s^T\mathbf{z}_s + \frac{1}{2n}\sum_{i}^{n}\mathbf{z}_i^T\mathbf{z}_i - \mathbf{z}_s^T\bar{\mathbf{z}};$$

(4)

$$\begin{aligned}
\frac{1}{2n^2}\sum_{i,j} d_{i,j}^2 &= \frac{1}{2n^2}\sum_{i,j}^{n}(\mathbf{z}_i^T\mathbf{z}_i + \mathbf{z}_j^T\mathbf{z}_j - 2\mathbf{z}_i\mathbf{z}_j) \\
&= \frac{1}{2n}\sum_{i=1}^{n}\mathbf{z}_i^T\mathbf{z}_i + \frac{1}{2n}\sum_{j=1}^{n}(\mathbf{z}_j^T\mathbf{z}_j) - (\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i)^T(\frac{1}{n}\sum_{j=1}^{n}\mathbf{z}_j) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i^T\mathbf{z}_i - \bar{\mathbf{z}}^T\bar{\mathbf{z}};
\end{aligned}$$

Hence,

$$\begin{aligned}
b_{rs} &= \bar{\mathbf{z}}^T\mathbf{z} + \bar{\mathbf{z}}_s^T\mathbf{z}_r - \bar{\mathbf{z}}_r^T\bar{\mathbf{z}} - \mathbf{z}_s^T\bar{\mathbf{z}} \\
&= \mathbf{z}_s^T(\mathbf{z}_r - \bar{\mathbf{z}}) + \bar{\mathbf{z}}^T(\bar{\mathbf{z}} - \mathbf{z}_r) \\
&= (\mathbf{z}_r - \bar{\mathbf{z}})^T(\mathbf{z}_s - \bar{\mathbf{z}});
\end{aligned}$$

Let $f(\mathbf{z}) = \mathbf{z} - \bar{\mathbf{z}}$, then $b_{rs} = f(\mathbf{z}_r)^T f(\mathbf{z}_s)$. Thus B is *p.s.d.*
(b) Now we prove the opposite direction. That is, if $\mathbf{B}$ is *p.s.d.* of rank p, then a configuration corresponds to $\mathbf{B}$ can be constructed as follows:
Let $\lambda_1 > \lambda_2 > ... > \lambda_p$ denote the positive eigenvalues of $\mathbf{B}$ with corresponding eigenvectors $\mathbf{Z} = (\mathbf{z}_{(1)}, \mathbf{z}_{(2)}, ..., \mathbf{z}_{(p)})$ normalized by $\mathbf{z}_{(i)}^T \mathbf{z}_{(i)} = \lambda_i (i = 1, 2...p)$, then points p_r in $\mathbf{R}^p$ with coordinates $\mathbf{z}_r = (\mathbf{z}_{(r_1)}), \mathbf{z}_{(r_2)}, ..., \mathbf{z}_{(r_p)})^T$ have interpoint distance given by D. Further, this configuration has center of $\bar{\mathbf{z}} = 0$.
Now we show the construction is correct.
Let $\mathbf{\Lambda} = diag(\lambda_1, ...\lambda_p), \mathbf{U} = [\mathbf{U}(1), \mathbf{U}(2), ..., \mathbf{U}(p)]$ where $\mathbf{U}(i) = \mathbf{z}_{(i)} \lambda_i^{-\frac{1}{2}}$.
Thus,

$$\mathbf{U} = \mathbf{Z}\mathbf{\Lambda}^{-\frac{1}{2}}$$

$$\mathbf{U}^T\mathbf{U} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Z}\mathbf{Z}^T\mathbf{\Lambda}^{-\frac{1}{2}} = \mathbf{I}_p$$

We can know $\mathbf{U}$ is composed of a set of orthonormal eigenvectors with $\mathbf{U}(i)$ corresponds to λ_i. The diagonal of $\mathbf{\Lambda}$ is composed of eigenvalues of $\mathbf{B}$.
Hence, the spectral decomposition of $\mathbf{B}$ is $\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^T = \mathbf{Z}\mathbf{Z}^T$, which means $b_{rs} = \mathbf{z}_r^T \mathbf{z}_s$.
According to the construction of $\mathbf{D}$, if $\mathbf{D}$ is the Euclidean distance matrix of $p_1, p_2, ...p_n$,

$$\mathbf{D} = (d_{rs}^2)$$

where $d_{rs}^2 = (\mathbf{z}_r - \mathbf{z}_s)^T(\mathbf{z}_r - \mathbf{z}_s) = \mathbf{z}_r^T\mathbf{z}_r - 2\mathbf{z}_r^T\mathbf{z}_s + \mathbf{z}_s^T\mathbf{z}_s = b_{rr} - 2b_{rs} + b_{ss}$.
According to (a),

$$b_{rs} = a_{rs} - \bar{a}_{.s} - \bar{a}_{r.} + \bar{a}$$

Thus we have

$$\begin{aligned}
d_{rs}^2 &= b_{rr} - 2b_{rs} + b_{ss} \\
&= a_{rr} - \bar{a}_{r.} - \bar{a}_{.r} + \bar{a} + a_{ss} - \bar{a}_{s.} - \bar{a}_{.s} + \bar{a} - 2a_{rs} + 2\bar{a}_{r.} + 2\bar{a}_{.s} - 2\bar{a} \\
&= \bar{a}_{r.} + \bar{a}_{.s} - \bar{a}_{.r} - \bar{a}_{s.} - 2a_{rs} \\
&= -2a_{rs}
\end{aligned}$$

As we can see, the result conforms to the definition of A, which means D is really the Euclidean distance matrix of $p_1, p_2, ...p_n$.
Remarks: Here every row of Z corresponds to a point. Moreover, since $\mathbf{H}\mathbf{1}_n = \mathbf{0} = 0 \cdot \mathbf{1}_n$, we can get $\mathbf{B}\mathbf{1}_n = \mathbf{H}\mathbf{A}\mathbf{H}\mathbf{1}_n = 0 \cdot \mathbf{1}_n$. Thus $\mathbf{1}_n$ is an eigenvector of B corresponds to eigenvalue 0. And eigenvectors corresponds to different eigenvalues are orthogonal. Hence, $\mathbf{Z}^T\mathbf{1}_n = \mathbf{0}$,which is equivalent to $\sum_{r=1}^n \mathbf{z}_r = 0$. Therefore, the mean of all points pass the origin, which guarantees the uniqueness of $\mathbf{Z}$.

Add a point:
Now we consider when a new point comes, how to find the coordinates of that point. We formalize the problem as follows:
Given a distance matrix $\mathbf{D}_{n\times n}$ of n points $p_1, p_2, ..., p_n$. The distance between a new point p_{n+1} with the n points is $(d_{1,n+1}, d_{2,n+1}, ..., d_{n,n+1})$,find the coordinates of p_{n+1}.

Here we suppose the coordinates of p_{n+1} is $\mathbf{z}_{n+1} = (\mathbf{z}_{n+1,1}, \mathbf{z}_{n+1,2}, ...\mathbf{z}_{n+1,p})^T$.
Firstly, we can use $\mathbf{D}$ to get $\mathbf{B}$ and then get $\mathbf{Z}_{n\times p}$.
Then

$$\begin{aligned} d_{i,n+1}^2 &= \sum_{k=1}^{p} (\mathbf{z}_{n+1,k} - \mathbf{z}_{i,k})^2 \\ &= \sum_{k=1}^{p} \mathbf{z}_{n+1,k}^2 + \sum_{k=1}^{p} \mathbf{z}_{i,k}^2 - 2\sum_{k=1}^{p} \mathbf{z}_{n+1,k}\mathbf{z}_{i,k} \\ &= d_{n+1}^2 + d_i^2 - 2\sum_{k=1}^{p} \mathbf{z}_{n+1,k}\mathbf{z}_{i,k} \end{aligned}$$

Thus

$$\begin{aligned} \sum_{i=1}^{n} d_{i,n+1}^2 &= nd_{n+1}^2 + \sum_{i=1}^{n} d_i^2 - 2\sum_{k=1}^{p}(\sum_{i=1}^{n} \mathbf{z}_{i,k})\mathbf{z}_{n+1,k} \\ &= nd_{n+1}^2 + \sum_{i=1}^{n} d_i^2 \\ &(\text{Note:} \textstyle\sum_{i=1}^{n} \mathbf{z}_{i,k} = 0) \\ d_{n+1}^2 &= \frac{1}{n}\sum_{i=1}^{n}(d_{i,n+1}^2 - d_i^2) \end{aligned}$$

Substitute d_{n+1}^2 with the formula above, we can get

$$\begin{aligned} d_{i,n+1}^2 &= \frac{1}{n}\sum_{i=1}^{n}(d_{i,n+1}^2 - d_i^2) + d_i^2 - 2\sum_{k=1}^{p} \mathbf{z}_{n+1,k}\mathbf{z}_{i,k} \\ 2\sum_{k=1}^{p} \mathbf{z}_{n+1,k}\mathbf{z}_{i,k} &= d_i^2 - d_{i,n+1}^2 - \frac{1}{n}\sum_{i=1}^{n}(d_i^2 - d_{i,n+1}^2) \end{aligned}$$

Let $\alpha_i = d_i^2 - d_{i,n+1}^2$, then

$$\begin{aligned} 2\sum_{k=1}^{p} \mathbf{z}_{n+1,k}\mathbf{z}_{i,k} &= \alpha_i - \frac{1}{n}\sum_{i=j} n\alpha_j \\ 2\mathbf{Z}\mathbf{z}_{n+1} &= \boldsymbol{\alpha} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\boldsymbol{\alpha} \end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ...\alpha_n)^T$,thus

$$
\begin{aligned}
(\mathbf{Z}^T\mathbf{Z})\mathbf{z}_{n+1} &= \frac{1}{2}\mathbf{Z}^T(\boldsymbol{\alpha} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\boldsymbol{\alpha}) \\
\boldsymbol{\Lambda}\mathbf{z}_{n+1} &= \frac{1}{2}\mathbf{Z}^T(\boldsymbol{\alpha} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\boldsymbol{\alpha}) \\
\mathbf{z}_{n+1} &= \frac{1}{2}\boldsymbol{\Lambda}^{-1}\mathbf{Z}^T(\boldsymbol{\alpha} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\boldsymbol{\alpha}) \\
\mathbf{z}_{n+1} &= \frac{1}{2}\boldsymbol{\Lambda}^{-1}\mathbf{Z}^T\mathbf{H}\boldsymbol{\alpha}
\end{aligned}
$$

Summary:
We can see MDS is based on distance matrix, and PCO is a special instance of MD-S.Compare PCO with PCA, we can find PCA is based on kernel matrix, which is *p.s.d*, while PCO is based on distance matrix, which is *n.d*. However, the results of these two methods are in fact the same. While dealing with a real problem, you can choose the one that of lower complexity.