- classical multidimensional scaling
  - assume that the *observed* $n \times n$ proximity matrix **D** (a random matrix) is a matrix of Euclidean distances derived from a raw $n \times q$ data matrix, **X**, which is *not observed*. (a random matrix)
  - define an $n \times n$ matrix **B**

$$P\Lambda P' = \mathbf{B} = \mathbf{X}\mathbf{X}' (= (\underbrace{\mathbf{XM}}_{X^*})(\underbrace{\mathbf{XM}}_{X^*})') = XMM'X' = XX'$$

M: an orthogonal matrix — similar situation encounted in factor analysis → $\boxed{\text{factor rotation}}$

- the elements of **B** are given by

$$b_{ij} = \sum_{k=1}^{q} x_{ik} x_{jk}$$

- the squared Euclidean distances between the rows of **X** can be written in terms of the elements of **B** as $\sum_k (x_{ik} - x_{jk})^2 = \sum_k x_{ik}^2 + \sum_k x_{jk}^2 - 2 \sum_k x_{ik} x_{jk}$

$$d_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}$$

- idea: If the b_{ij}'s could be found in terms of the d_{ij}'s in the equation above, then we can derive **X** from **B** by factoring **B**.
- to obtain **B** from **D**, no unique solution exists unless a location constraint is introduced. Usually, the center of the columns of **X** are set at origin, i.e.,

$$\sum_{i=1}^{n} x_{ik} = 0, \quad \text{for all } k$$

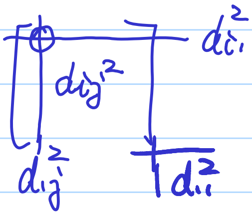
mean of each variable in X (column)

- these constraints imply that sum of the terms in any row of **B** must be 0, i.e.,

$$\sum_{j=1}^{n} b_{ij} = \sum_{j=1}^{n} \sum_{k=1}^{q} x_{ik} x_{jk} = \sum_{k=1}^{q} x_{ik} \left(\sum_{j=1}^{n} x_{jk} \right) = 0$$

- Let T be the trace of **B**. To obtain **B** from **D**, notice that
  - $\sum_{i=1}^{n} d_{ij}^2 = T + nb_{jj}$ $= \sum_i b_{ii} + \sum_i b_{jj} - 2\sum_i b_{ij}$.
  - $\sum_{j=1} d_{ij}^2 = nb_{ii} + T$
  - $\sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^2 = 2nT$
- the elements of **B** can be found from **D** as

(c.f.) main-effect model in regression



$$b_{ij} = -\frac{1}{2} \left[d_{ij}^2 - d_{i\cdot}^2 - d_{\cdot j}^2 + d_{\cdot\cdot}^2 \right]$$

where $d_{i\cdot}^2 = (\sum_{j=1}^{n} d_{ij}^2)/n$, $d_{\cdot j}^2 = (\sum_{i=1}^{n} d_{ij}^2)/n$, $d_{\cdot\cdot}^2 = (\sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^2)/n^2$

$= (T+nb_{jj})/n = \frac{T}{n} + b_{jj}$ $\quad = \frac{T}{n} + b_{jj}$ $\quad = \frac{2nT}{n^2} = \frac{2T}{n}$

- **B** (symmetric matrix) can be written as

$$\mathbf{B} = \mathbf{V \Lambda V}', \leftarrow \text{spectral decomposition}$$

where $\mathbf{\Lambda} = diag[\lambda_1, \cdots, \lambda_n]$ $(\lambda_1 \geq \cdots \geq \lambda_n)$ is the diagonal matrix of eigenvalues of **B** and $\mathbf{V} = [\mathbf{V_1}, \cdots, \mathbf{V_n}]$ is the corresponding matrix of normalized eigenvectors (i.e., $\mathbf{V_i'V_i} = 1$)

PCA $\Sigma = P\Lambda P'$

FA $\Sigma = (P\Lambda^{1/2})(P\Lambda^{1/2})'$

- Note: when **D** arises from an $n \times q$ data matrix, the rank of **B** is q (i.e, the last $n-q$ eigenvalues should be zero)
- So, **B** can be chosen as

$$V\Lambda V' = \left[V^* \,\vdots\, \cdots \right] \left[\begin{array}{c|c} \Lambda^* & 0 \\ \hline 0 & 0 \end{array} \right] \left[\begin{array}{c} V^{*\prime} \\ \hline \vdots \end{array} \right]$$

$$\mathbf{B} = \mathbf{V}^* \mathbf{\Lambda}^* \mathbf{V}^{*\prime}, = V^* \Lambda^{*1/2} \Lambda^{*1/2} V^{*\prime}$$

where $\mathbf{V}^*$ contains the first q eigenvectors and $\mathbf{\Lambda}^*$ the first q eigenvalues

- Thus, a solution of **X** is $\mathbf{X} = \mathbf{V}^* \mathbf{\Lambda}^{*1/2}$

- The adequacy of the q-dimensional representation can be judged by the size of the criterion

$$\left(\sum_{i=1}^{q} \lambda_i\right) / \left(\sum_{i=1}^{n} \lambda_i\right)$$ ← must choose a q.

- When the observed proximity matrix is not Euclidean, the matrix **B** is not positive-definite. In such case, some of the eigenvalues of **B** will be negative; correspondingly, some coordinate values will be complex numbers.
  - If **B** has only a small number of small negative eigenvalues, it's still possible to use the eigenvectors associated with the q largest positive eigenvalues
  - adequacy of the resulting solution might be assessed using
    - $\left(\sum_{i=1}^{q} |\lambda_i|\right) / \left(\sum_{i=1}^{n} |\lambda_i|\right) \equiv P_q^{(1)}$
    - $\left(\sum_{i=1}^{q} \lambda_i^2\right) / \left(\sum_{i=1}^{n} \lambda_i^2\right) \equiv P_q^{(2)}$.

• Some other issues

- metric scaling (define loss function for **D** and the distance matrix based on **X**, called *stress*, and find **X** to minimize stress)
- non-metric scaling (applied when the actual values of **D** is not reliable, but their orders can be trusted)
- 3-way multidimensional scaling (proximity matrix result from individual assessments of dissimilarity and more than one individual is sampled)
- asymmetric proximity matrix ($d_{ij} \neq d_{ji}$)

❖ **Reading**: Reference, 5.1, 5.2