

Distance matrices, Gram matrices and multidimensional scaling (MDS)

MATH 2740 – Mathematics of Data Science – Unit U07

Julien Arino

`julien.arino@umanitoba.ca`

Department of Mathematics @ University of Manitoba

Fall 2026

Outline



From coordinates to distances and back

In principal component analysis (Unit U05), we start with a data matrix $X \in \mathcal{M}_{np}(\mathbb{R})$ containing the coordinates of n observations $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ and project them onto a lower-dimensional subspace $\mathbb{R}^k$ ($k \leq p$)

From the coordinates $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$, it is straightforward to compute all pairwise Euclidean distances

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| = \sqrt{(\mathbf{x}_i - \mathbf{x}_j) \bullet (\mathbf{x}_i - \mathbf{x}_j)}, \quad i, j = 1, \dots, n$$

The inverse problem: what if we do *not* know the coordinates X , and only have a table of pairwise distances or dissimilarities $D = [d_{ij}] \in \mathcal{M}_n(\mathbb{R})$ between n objects?

Where do distance-only datasets arise?

In many scientific applications, objects do not come with explicit coordinates in $\mathbb{R}^p$, or only pairwise comparisons are measured:

- ▶ **Geography and transport:** road distances, flight distances or travel times between n cities
- ▶ **Genomics and evolutionary biology:** sequence edit distances or evolutionary divergence times between n species
- ▶ **Psychology and sensory science:** human judgements of dissimilarity between colours, sounds, faces or products
- ▶ **Cheminformatics and text analysis:** structural dissimilarity between molecules or cosine/Jaccard distances between documents

The goal of multidimensional scaling (MDS)

Given an $n \times n$ matrix $D = [d_{ij}]$ of pairwise distances between n objects:

1. **Exact Euclidean embedding:** do there exist points $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ (for some dimension p) such that

$$\|\mathbf{x}_i - \mathbf{x}_j\| = d_{ij} \quad \forall i, j = 1, \dots, n?$$

2. **Dimensionality reduction and visualisation:** for a chosen small dimension k (typically $k = 1, 2$ or 3), find coordinates $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^k$ whose pairwise Euclidean distances $\|\mathbf{x}_i - \mathbf{x}_j\|$ best approximate d_{ij}

This family of methods is called **multidimensional scaling (MDS)**. When D is treated as a Euclidean distance matrix and coordinates are recovered via spectral decomposition, it is called **classical MDS** (or **principal coordinate analysis, PCoA**)

Non-uniqueness: rigid motions preserve distances

Before trying to recover coordinates $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ from pairwise distances $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$, observe that coordinates can *never* be unique!

Proposition 95 (Invariance of Euclidean distances under rigid motions)

Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$, let $\mathbf{s} \in \mathbb{R}^p$ be a translation vector and let $Q \in \mathcal{M}_p(\mathbb{R})$ be an orthogonal matrix ($Q^T Q = \mathbb{I}_p$). Define transformed points

$$\mathbf{y}_i = Q\mathbf{x}_i + \mathbf{s}, \quad i = 1, \dots, n$$

Then all pairwise Euclidean distances are preserved:

$$\|\mathbf{y}_i - \mathbf{y}_j\| = \|\mathbf{x}_i - \mathbf{x}_j\| \quad \forall i, j = 1, \dots, n$$

Thus, distances only determine a point configuration up to **translation**, **rotation** and **reflection** (collectively called **rigid motions** or **Euclidean isometries**)

Proof of Proposition ??

Let $\mathbf{y}_i = Q\mathbf{x}_i + \mathbf{s}$ for $i = 1, \dots, n$, where $\mathbf{s} \in \mathbb{R}^p$ and $Q^T Q = \mathbb{I}_p$

First, taking the difference between $\mathbf{y}_i$ and $\mathbf{y}_j$ eliminates the translation vector $\mathbf{s}$:

$$\mathbf{y}_i - \mathbf{y}_j = (Q\mathbf{x}_i + \mathbf{s}) - (Q\mathbf{x}_j + \mathbf{s}) = Q\mathbf{x}_i - Q\mathbf{x}_j = Q(\mathbf{x}_i - \mathbf{x}_j)$$

Next, compute the squared Euclidean norm using the orthogonality of Q :

$$\begin{aligned}\|\mathbf{y}_i - \mathbf{y}_j\|^2 &= \|Q(\mathbf{x}_i - \mathbf{x}_j)\|^2 \\ &= (Q(\mathbf{x}_i - \mathbf{x}_j))^T (Q(\mathbf{x}_i - \mathbf{x}_j)) \\ &= (\mathbf{x}_i - \mathbf{x}_j)^T Q^T Q (\mathbf{x}_i - \mathbf{x}_j) \\ &= (\mathbf{x}_i - \mathbf{x}_j)^T \mathbb{I}_p (\mathbf{x}_i - \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|^2\end{aligned}$$

Taking the non-negative square root yields $\|\mathbf{y}_i - \mathbf{y}_j\| = \|\mathbf{x}_i - \mathbf{x}_j\|$



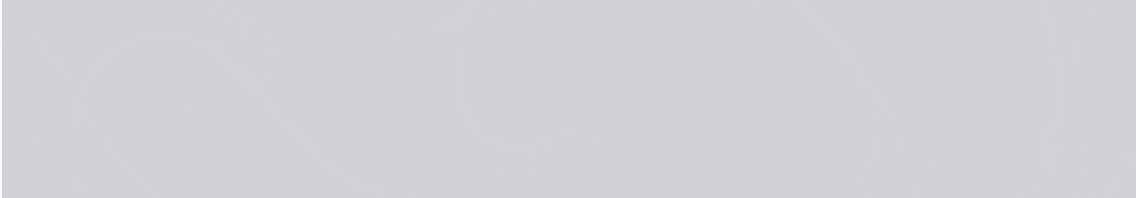
Removing translation ambiguity by centring at the origin

Because shifting all points by a constant vector $\mathbf{s} \in \mathbb{R}^p$ leaves every pairwise distance unchanged, we must fix a reference location to obtain a well-defined reconstruction

Convention in classical MDS: we require the reconstructed points $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ to have their **centroid** (sample mean) at the origin $\mathbf{0} \in \mathbb{R}^p$:

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \mathbf{0} \iff \sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$$

- ▶ Fixing $\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$ completely removes the translation ambiguity
- ▶ Any remaining ambiguity is purely an orthogonal transformation $Q \in \mathcal{M}_p(\mathbb{R})$ (rotation or reflection about the origin)
- ▶ Algebraically, why does fixing the origin help? Because once an origin is fixed, **distances determine inner products**, and inner products give us a **Gram matrix** that we can diagonalise!





Data matrix convention: observations in rows

Throughout this unit, we represent a configuration of n points in $\mathbb{R}^p$ by a data matrix $X \in \mathcal{M}_{np}(\mathbb{R})$ whose i th row is the transpose of the column vector $\mathbf{x}_i \in \mathbb{R}^p$:

$$X = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \in \mathcal{M}_{np}(\mathbb{R})$$

- ▶ Rows $i = 1, \dots, n$ are **observations** (objects, samples, cities or individuals)
- ▶ Columns $j = 1, \dots, p$ are **features** (also called variables or coordinate axes)

The Gram matrix of a set of vectors

Definition 96 (Gram matrix)

Let $X \in \mathcal{M}_{np}(\mathbb{R})$ have rows $\mathbf{x}_1^T, \dots, \mathbf{x}_n^T \in \mathbb{R}^p$. The **Gram matrix** (or **inner product matrix**) of $\mathbf{x}_1, \dots, \mathbf{x}_n$ is the $n \times n$ matrix

$$G = XX^T \in \mathcal{M}_n(\mathbb{R})$$

whose (i, j) entry is the standard inner product between $\mathbf{x}_i$ and $\mathbf{x}_j$:

$$g_{ij} = \mathbf{x}_i \bullet \mathbf{x}_j = \mathbf{x}_i^T \mathbf{x}_j = \sum_{m=1}^p x_{im}x_{jm}, \quad i, j = 1, \dots, n$$

Explicitly, in block row-times-column form:

$$G = XX^T = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} (\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n) = \begin{pmatrix} \mathbf{x}_1^T \mathbf{x}_1 & \mathbf{x}_1^T \mathbf{x}_2 & \cdots & \mathbf{x}_1^T \mathbf{x}_n \\ \mathbf{x}_2^T \mathbf{x}_1 & \mathbf{x}_2^T \mathbf{x}_2 & \cdots & \mathbf{x}_2^T \mathbf{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_n^T \mathbf{x}_1 & \mathbf{x}_n^T \mathbf{x}_2 & \cdots & \mathbf{x}_n^T \mathbf{x}_n \end{pmatrix}$$

Geometric meaning of the Gram matrix

Every entry of the Gram matrix $G = XX^T$ encodes geometric information relative to the origin $\mathbf{0}$:

- ▶ **Diagonal entries** give squared lengths (squared distances from the origin):

$$g_{ii} = \mathbf{x}_i^T \mathbf{x}_i = \|\mathbf{x}_i\|^2 \geq 0, \quad i = 1, \dots, n$$

- ▶ **Off-diagonal entries** encode lengths and angles θ_{ij} between vectors $\mathbf{x}_i$ and $\mathbf{x}_j$:

$$g_{ij} = \mathbf{x}_i^T \mathbf{x}_j = \|\mathbf{x}_i\| \|\mathbf{x}_j\| \cos \theta_{ij}, \quad i \neq j$$

- ▶ **Orthogonal invariance:** if we rotate or reflect all points by an orthogonal matrix $Q \in \mathcal{M}_p(\mathbb{R})$ (so $\mathbf{x}_i \mapsto Q\mathbf{x}_i$ and $X \mapsto XQ^T$), the Gram matrix is **unchanged**:

$$(XQ^T)(XQ^T)^T = XQ^T QX^T = X\mathbb{I}_p X^T = XX^T = G$$

Positive semi-definite and positive definite matrices

Definition 97 (Positive semi-definite and positive definite matrices)

Let $A \in \mathcal{M}_n(\mathbb{R})$ be a symmetric matrix ($A^T = A$).

- ▶ A is **positive semi-definite** (**PSD**) if

$$\mathbf{v}^T A \mathbf{v} \geq 0 \quad \forall \mathbf{v} \in \mathbb{R}^n$$

- ▶ A is **positive definite** (**PD**) if

$$\mathbf{v}^T A \mathbf{v} > 0 \quad \forall \mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$$

Recall from Theorem ?? (Unit U04) that all eigenvalues of a real symmetric matrix $A \in \mathcal{M}_n(\mathbb{R})$ are real. Equivalently:

- ▶ A is positive semi-definite $\iff$ all eigenvalues of A satisfy $\lambda_i \geq 0$
- ▶ A is positive definite $\iff$ all eigenvalues of A satisfy $\lambda_i > 0$ (in which case A is invertible)

You have already encountered these matrices!

Even without the name, we have already proved and used these properties for matrices of the form $A^T A$ and AA^T (with $A \in \mathcal{M}_{np}(\mathbb{R})$):

- ▶ **Unit U02 (Theorem ??):** we proved that for any $A \in \mathcal{M}_{np}(\mathbb{R})$, both $AA^T \in \mathcal{M}_n(\mathbb{R})$ and $A^T A \in \mathcal{M}_p(\mathbb{R})$ are **symmetric**
- ▶ **Unit U03 (Least squares, invertibility of $A^T A$):** for any $\mathbf{x} \in \mathbb{R}^p$, we wrote

$$\mathbf{x}^T (A^T A) \mathbf{x} = (A\mathbf{x})^T (A\mathbf{x}) = \|A\mathbf{x}\|^2 \geq 0$$

which says precisely that $A^T A$ is **positive semi-definite**! When A has linearly independent columns ($A\mathbf{x} = \mathbf{0} \iff \mathbf{x} = \mathbf{0}$), $\|A\mathbf{x}\|^2 > 0$ for all $\mathbf{x} \neq \mathbf{0}$, so $A^T A$ is **positive definite** (and hence invertible)

- ▶ **Unit U04 (SVD, Theorem ??):** we proved that all eigenvalues of $A^T A$ and AA^T are **real and non-negative** ($\lambda_i \geq 0$), which allowed us to define singular values $\sigma_i = \sqrt{\lambda_i}$
- ▶ **Unit U05 (PCA):** the sample covariance matrix $S_X = \frac{1}{n-1} \tilde{X}^T \tilde{X}$ is symmetric positive semi-definite, and its eigenvalues $\lambda_i \geq 0$ are the variances along the principal axes

Fundamental properties of Gram matrices

Notice that for a data matrix $X \in \mathcal{M}_{np}(\mathbb{R})$, its Gram matrix $G = XX^T \in \mathcal{M}_n(\mathbb{R})$ is simply the matrix AA^T from the SVD in Unit U04 (with $A = X$), whose eigenvectors are the left singular vectors of X

Theorem 98 (Properties of the Gram matrix)

Let $X \in \mathcal{M}_{np}(\mathbb{R})$ and let $G = XX^T \in \mathcal{M}_n(\mathbb{R})$ be its Gram matrix. Then:

- 1. G is symmetric ($G^T = G$) and positive semi-definite ($\mathbf{v}^T G \mathbf{v} \geq 0$ for all $\mathbf{v} \in \mathbb{R}^n$), with real non-negative eigenvalues $\lambda_1 \geq \dots \geq \lambda_n \geq 0$*
- 2. $\text{rank}(G) = \text{rank}(X) \leq \min(n, p)$*
- 3. $\text{tr}(G) = \sum_{i=1}^n \|\mathbf{x}_i\|^2 = \|X\|_F^2$*

Proof of Theorem ??

1. Symmetry and positive semi-definiteness: Setting $A = X$ in Theorem ?? (Unit U02), $G = XX^T$ is symmetric. By Theorem ?? (Unit U04), the eigenvalues of XX^T are real and non-negative, and as in the proof of Theorem ?? (applied to X^T), for any $\mathbf{v} \in \mathbb{R}^n$,

$$\mathbf{v}^T G \mathbf{v} = \mathbf{v}^T (XX^T) \mathbf{v} = (X^T \mathbf{v})^T (X^T \mathbf{v}) = \|X^T \mathbf{v}\|^2 \geq 0$$

2. Rank: By the same nullspace argument used for $A^T A$ in Unit U03 (applied to $A = X^T$), $XX^T \mathbf{v} = \mathbf{0} \iff \|X^T \mathbf{v}\|^2 = 0 \iff X^T \mathbf{v} = \mathbf{0}$, so $\text{null}(XX^T) = \text{null}(X^T)$. By the rank-nullity theorem (or because $\text{rank}(X)$ in Theorem ?? equals the number r of positive eigenvalues of XX^T), $\text{rank}(G) = \text{rank}(X^T) = \text{rank}(X) \leq \min(n, p)$

3. Trace: $\text{tr}(G) = \sum_{i=1}^n g_{ii} = \sum_{i=1}^n \mathbf{x}_i^T \mathbf{x}_i = \sum_{i=1}^n \|\mathbf{x}_i\|^2 = \sum_{i=1}^n \sum_{m=1}^p x_{im}^2 = \|X\|_F^2$



Characterisation of Gram matrices via spectral factorisation

In the SVD (Theorems ?? and ??, Unit U04), $X = U\Sigma V^T$ gives

$$XX^T = U\Sigma\Sigma^T U^T = \sum_{i=1}^r \sigma_i^2 \mathbf{u}_i \mathbf{u}_i^T$$

where $\lambda_i = \sigma_i^2 > 0$ are the positive eigenvalues of XX^T

Conversely, *any* symmetric PSD matrix B can be factored as a Gram matrix $B = XX^T$ by reversing this step, as shown by Theorem ?? later

Before stating that, we need to define the square root of a matrix...

Square root of a matrix

A square matrix S is called a **square root** of A if $S^2 = SS = A$, and we write $S = A^{1/2}$

For a diagonal matrix with non-negative diagonal entries $\lambda_i \geq 0$, its square root is obtained by simply taking the square root of each diagonal entry:

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) \implies \Lambda^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$$

Example (a simple 2×2 diagonal matrix):

$$\Lambda = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix} \implies \Lambda^{1/2} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

since

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 2^2 & 0 \\ 0 & 3^2 \end{pmatrix} = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}$$

Because a symmetric PSD matrix has eigenvalues $\lambda_i \geq 0$, the square roots $\sqrt{\lambda_i}$ are always real numbers!

Theorem 99 (Spectral factorisation of symmetric PSD matrices)

Let $B \in \mathcal{M}_n(\mathbb{R})$ be symmetric and positive semi-definite with $\text{rank}(B) = r \leq n$. Let its ordered eigenvalues be

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 \quad \text{and} \quad \lambda_{r+1} = \cdots = \lambda_n = 0$$

with corresponding orthonormal eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^n$. Define

$$V_r = (\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_r) \in \mathcal{M}_{nr}(\mathbb{R}), \quad \Lambda_r = \text{diag}(\lambda_1, \dots, \lambda_r) \in \mathcal{M}_r(\mathbb{R})$$

Then the matrix $X = V_r \Lambda_r^{1/2} \in \mathcal{M}_{nr}(\mathbb{R})$ satisfies $XX^T = B$

Proof of Theorem ??

Since $B \in \mathcal{M}_n(\mathbb{R})$ is real and symmetric, its eigenvalues are real (Theorem ??, Unit U04) and it has an orthonormal basis of eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^n$ (Theorem ??, Unit U04), giving $B = V\Lambda V^T$ with $V = [\mathbf{v}_1 \ \dots \ \mathbf{v}_n] \in \mathcal{M}_n(\mathbb{R})$ orthogonal and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$

Just as in the outer product form of the SVD (Theorem ??, Unit U04), only the $r = \text{rank}(B)$ positive eigenvalues $\lambda_1 \geq \dots \geq \lambda_r > 0$ contribute:

$$B = \sum_{i=1}^r \lambda_i \mathbf{v}_i \mathbf{v}_i^T = V_r \Lambda_r V_r^T$$

Exactly as we defined singular values $\sigma_i = \sqrt{\lambda_i}$ in Unit U04 (valid since B is PSD, so $\lambda_i > 0$ for $i = 1, \dots, r$), define $\Lambda_r^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r})$. Since $\Lambda_r^{1/2}$ is diagonal, $(\Lambda_r^{1/2})^T = \Lambda_r^{1/2}$ and $\Lambda_r^{1/2}(\Lambda_r^{1/2})^T = \Lambda_r$. Setting $X = V_r \Lambda_r^{1/2} \in \mathcal{M}_{nr}(\mathbb{R})$ gives

$$XX^T = (V_r \Lambda_r^{1/2})(V_r \Lambda_r^{1/2})^T = V_r \Lambda_r^{1/2} (\Lambda_r^{1/2})^T V_r^T = V_r \Lambda_r V_r^T = B \quad \square$$

Euclidean and squared Euclidean distance matrices

Definition 100 (Euclidean distance matrix and squared distance matrix)

Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ be the rows of $X \in \mathcal{M}_{np}(\mathbb{R})$

- The **Euclidean distance matrix** is $D = [d_{ij}] \in \mathcal{M}_n(\mathbb{R})$, where

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| = \sqrt{\sum_{m=1}^p (x_{im} - x_{jm})^2}, \quad i, j = 1, \dots, n$$

- The **squared Euclidean distance matrix** is $D^{(2)} = [d_{ij}^2] \in \mathcal{M}_n(\mathbb{R})$, obtained by squaring each entry of D elementwise (**Hadamard square** $D^{(2)} = D \odot D$):

$$[D^{(2)}]_{ij} = d_{ij}^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2, \quad i, j = 1, \dots, n$$

Warning: $D^{(2)}$ denotes the elementwise square $[d_{ij}^2]$, *not* the matrix product DD !

Basic properties of distance matrices

For any configuration of points $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$, the matrices $D = [d_{ij}]$ and $D^{(2)} = [d_{ij}^2]$ satisfy:

1. Zero diagonal (**hollow matrix**):

$$d_{ii} = \|\mathbf{x}_i - \mathbf{x}_i\| = 0 \implies \text{diag}(D) = \text{diag}(D^{(2)}) = \mathbf{0} \quad \text{and} \quad \text{tr}(D) = \text{tr}(D^{(2)}) = 0$$

2. Symmetry:

$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\| = \|\mathbf{x}_j - \mathbf{x}_i\| = d_{ji} \implies D^T = D \quad \text{and} \quad (D^{(2)})^T = D^{(2)}$$

3. Non-negativity: $d_{ij} \geq 0$ and $d_{ij}^2 \geq 0$ for all $i, j = 1, \dots, n$, with $d_{ij} = 0 \iff \mathbf{x}_i = \mathbf{x}_j$

4. Triangle inequality (for D , though *not* necessarily for $D^{(2)}$):

$$d_{ij} \leq d_{ik} + d_{kj} \quad \forall i, j, k = 1, \dots, n$$

Relating squared distances to inner products

Why do we work with the *squared* distance matrix $D^{(2)}$ rather than D ? Because squared Euclidean norms expand directly in terms of inner products!

For any two points $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^p$, using the bilinearity and symmetry of the inner product:

$$\begin{aligned}d_{ij}^2 &= \|\mathbf{x}_i - \mathbf{x}_j\|^2 = (\mathbf{x}_i - \mathbf{x}_j) \bullet (\mathbf{x}_i - \mathbf{x}_j) \\&= (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j) \\&= \mathbf{x}_i^T \mathbf{x}_i - \mathbf{x}_i^T \mathbf{x}_j - \mathbf{x}_j^T \mathbf{x}_i + \mathbf{x}_j^T \mathbf{x}_j \\&= \|\mathbf{x}_i\|^2 - 2\mathbf{x}_i^T \mathbf{x}_j + \|\mathbf{x}_j\|^2\end{aligned}$$

In terms of the entries $g_{ij} = \mathbf{x}_i^T \mathbf{x}_j$ of the Gram matrix $G = XX^T$:

$$\boxed{d_{ij}^2 = g_{ii} - 2g_{ij} + g_{jj}} \quad \forall i, j = 1, \dots, n$$

Notice that this is the law of cosines: $c^2 = a^2 + b^2 - 2ab \cos \theta$

Matrix formula for $D^{(2)}$ in terms of G

We can write the entrywise identity $d_{ij}^2 = g_{ii} - 2g_{ij} + g_{jj}$ as a single matrix equation

Theorem 101 (Squared distance matrix from the Gram matrix)

Let $X \in \mathcal{M}_{np}(\mathbb{R})$ have Gram matrix $G = XX^T \in \mathcal{M}_n(\mathbb{R})$. Let

$$\mathbf{g} = \text{diag}(G) = \begin{pmatrix} g_{11} \\ g_{22} \\ \vdots \\ g_{nn} \end{pmatrix} = \begin{pmatrix} \|\mathbf{x}_1\|^2 \\ \|\mathbf{x}_2\|^2 \\ \vdots \\ \|\mathbf{x}_n\|^2 \end{pmatrix} \in \mathbb{R}^n \quad \text{and} \quad \mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$$

Then the squared Euclidean distance matrix $D^{(2)} = [d_{ij}^2] \in \mathcal{M}_n(\mathbb{R})$ is given by

$$D^{(2)} = \mathbf{g}\mathbf{1}^T - 2G + \mathbf{1}\mathbf{g}^T$$

Proof of Theorem ??

Examine the (i, j) entry of the rank-one outer product matrices $\mathbf{g}\mathbf{1}^T$ and $\mathbf{1}\mathbf{g}^T$:

$$\mathbf{g}\mathbf{1}^T = \begin{pmatrix} g_{11} \\ g_{22} \\ \vdots \\ g_{nn} \end{pmatrix} \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} g_{11} & g_{11} & \cdots & g_{11} \\ g_{22} & g_{22} & \cdots & g_{22} \\ \vdots & \vdots & \ddots & \vdots \\ g_{nn} & g_{nn} & \cdots & g_{nn} \end{pmatrix} \implies [\mathbf{g}\mathbf{1}^T]_{ij} = g_{ii}$$

Similarly, taking the transpose:

$$\mathbf{1}\mathbf{g}^T = (\mathbf{g}\mathbf{1}^T)^T = \begin{pmatrix} g_{11} & g_{22} & \cdots & g_{nn} \\ g_{11} & g_{22} & \cdots & g_{nn} \\ \vdots & \vdots & \ddots & \vdots \\ g_{11} & g_{22} & \cdots & g_{nn} \end{pmatrix} \implies [\mathbf{1}\mathbf{g}^T]_{ij} = g_{jj}$$

Therefore, for every $(i, j) \in \{1, \dots, n\}^2$,

$$[\mathbf{g}\mathbf{1}^T - 2\mathbf{G} + \mathbf{1}\mathbf{g}^T]_{ij} = g_{ii} - 2g_{ij} + g_{jj} = \|\mathbf{x}_i\|^2 - 2\mathbf{x}_i^T \mathbf{x}_j + \|\mathbf{x}_j\|^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2 = d_{ij}^2 \quad \square$$

Example – Computing G and $D^{(2)}$ from X (1/2)

Consider $n = 3$ points in $\mathbb{R}^2$ given by the rows of

$$X = \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ -1 & 0 \end{pmatrix} \in \mathcal{M}_{32}(\mathbb{R}), \quad \mathbf{x}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

First, compute the Gram matrix $G = XX^T$:

$$G = \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 3 & -1 \\ 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 5 & -1 \\ 5 & 10 & -3 \\ -1 & -3 & 1 \end{pmatrix}$$

The vector of diagonal entries of G (squared norms of the points) is

$$\mathbf{g} = \text{diag}(G) = \begin{pmatrix} 5 \\ 10 \\ 1 \end{pmatrix}$$

Example – Computing G and $D^{(2)}$ from X (2/2)

Using Theorem ??, we form $\mathbf{g}\mathbf{1}^T$, $\mathbf{1}\mathbf{g}^T$ and $D^{(2)} = \mathbf{g}\mathbf{1}^T - 2G + \mathbf{1}\mathbf{g}^T$:

$$\mathbf{g}\mathbf{1}^T = \begin{pmatrix} 5 & 5 & 5 \\ 10 & 10 & 10 \\ 1 & 1 & 1 \end{pmatrix}, \quad \mathbf{1}\mathbf{g}^T = \begin{pmatrix} 5 & 10 & 1 \\ 5 & 10 & 1 \\ 5 & 10 & 1 \end{pmatrix}$$

Hence:

$$D^{(2)} = \begin{pmatrix} 5 & 5 & 5 \\ 10 & 10 & 10 \\ 1 & 1 & 1 \end{pmatrix} - 2 \begin{pmatrix} 5 & 5 & -1 \\ 5 & 10 & -3 \\ -1 & -3 & 1 \end{pmatrix} + \begin{pmatrix} 5 & 10 & 1 \\ 5 & 10 & 1 \\ 5 & 10 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 5 & 8 \\ 5 & 0 & 17 \\ 8 & 17 & 0 \end{pmatrix}$$

Check directly:

- ▶ $d_{12}^2 = \|(1, 2)^T - (3, 1)^T\|^2 = (-2)^2 + 1^2 = 5$
- ▶ $d_{13}^2 = \|(1, 2)^T - (-1, 0)^T\|^2 = 2^2 + 2^2 = 8$
- ▶ $d_{23}^2 = \|(3, 1)^T - (-1, 0)^T\|^2 = 4^2 + 1^2 = 17$



The all-ones vector $\mathbf{1}$ and the all-ones matrix $\mathbb{J}_n$

Two simple building blocks in $\mathbb{R}^n$ and $\mathcal{M}_n(\mathbb{R})$:

- ▶ The **all-ones vector** $\mathbf{1} \in \mathbb{R}^n$:

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n \implies \mathbf{1}^T \mathbf{1} = \sum_{i=1}^n 1^2 = n \quad \text{and} \quad \|\mathbf{1}\| = \sqrt{n}$$

- ▶ The **all-ones matrix** $\mathbb{J}_n = \mathbf{1}\mathbf{1}^T \in \mathcal{M}_n(\mathbb{R})$:

$$\mathbb{J}_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \in \mathcal{M}_n(\mathbb{R}) \implies \mathbb{J}_n^2 = (\mathbf{1}\mathbf{1}^T)(\mathbf{1}\mathbf{1}^T) = \mathbf{1}(\mathbf{1}^T \mathbf{1})\mathbf{1}^T = n\mathbb{J}_n$$

Definition of the centring matrix

Definition 102 (Centring matrix)

For $n \geq 1$, the **centring matrix** of order n is the $n \times n$ matrix

$$C_n = \mathbb{I}_n - \frac{1}{n} \mathbb{1} \mathbb{1}^T = \mathbb{I}_n - \frac{1}{n} \mathbb{J}_n \in \mathcal{M}_n(\mathbb{R})$$

where $\mathbb{I}_n$ is the $n \times n$ identity matrix, $\mathbb{1} = (1, \dots, 1)^T \in \mathbb{R}^n$ and $\mathbb{J}_n = \mathbb{1} \mathbb{1}^T$ is the $n \times n$ matrix of all ones

Entrywise, the (i, j) entry of C_n is

$$[C_n]_{ij} = \delta_{ij} - \frac{1}{n} = \begin{cases} 1 - \frac{1}{n} = \frac{n-1}{n} & \text{if } i = j \\ -\frac{1}{n} & \text{if } i \neq j \end{cases}$$

Explicit examples of C_n for small n

Writing out $C_n = \mathbb{I}_n - \frac{1}{n}\mathbb{J}_n$ explicitly for $n = 2, 3, 4$:

$$C_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

$$C_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$$

$$C_4 = \frac{1}{4} \begin{pmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix}$$

Notice that in every C_n , each row and each column sums to zero!

Action of C_n on a vector: subtracting the mean

Why is C_n called the *centring matrix*? Let $\mathbf{v} = (v_1, \dots, v_n)^T \in \mathbb{R}^n$ be a column vector of n observations of a feature

First note that the inner product $\mathbf{1}^T \mathbf{v}$ is the sum of the entries of $\mathbf{v}$:

$$\mathbf{1}^T \mathbf{v} = \sum_{i=1}^n v_i = n\bar{v}, \quad \text{where} \quad \bar{v} = \frac{1}{n} \sum_{i=1}^n v_i$$

Multiplying $\mathbf{v}$ on the left by C_n gives:

$$C_n \mathbf{v} = \left(\mathbb{I}_n - \frac{1}{n} \mathbf{1} \mathbf{1}^T \right) \mathbf{v} = \mathbf{v} - \frac{1}{n} \mathbf{1} (\mathbf{1}^T \mathbf{v}) = \mathbf{v} - \bar{v} \mathbf{1} = \begin{pmatrix} v_1 - \bar{v} \\ v_2 - \bar{v} \\ \vdots \\ v_n - \bar{v} \end{pmatrix}$$

Thus, $C_n \mathbf{v}$ subtracts the arithmetic mean $\bar{v}$ from every component of $\mathbf{v}$, producing a **zero-mean** (centred) vector!

Action of C_n on a data matrix and link to covariance

Now let $X \in \mathcal{M}_{np}(\mathbb{R})$ be a data matrix with columns $\mathbf{c}_1, \dots, \mathbf{c}_p \in \mathbb{R}^n$ (the p features) and rows $\mathbf{x}_1^T, \dots, \mathbf{x}_n^T \in \mathbb{R}^p$ (the n observations)

- **Column-wise view:** left-multiplying X by C_n centres every column of X simultaneously:

$$\tilde{X} = C_n X = (C_n \mathbf{c}_1 \quad C_n \mathbf{c}_2 \quad \cdots \quad C_n \mathbf{c}_p)$$

- **Row-wise view:** if $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \frac{1}{n} X^T \mathbf{1} \in \mathbb{R}^p$ is the centroid of the n points, then

$$\tilde{X} = C_n X = X - \frac{1}{n} \mathbf{1} \mathbf{1}^T X = X - \mathbf{1} \bar{\mathbf{x}}^T = \begin{pmatrix} (\mathbf{x}_1 - \bar{\mathbf{x}})^T \\ \vdots \\ (\mathbf{x}_n - \bar{\mathbf{x}})^T \end{pmatrix}$$

shifts the centroid of the n points to the origin $\mathbf{0} \in \mathbb{R}^p$

- **Link to Unit U05:** the sample covariance matrix $S_X \in \mathcal{M}_p(\mathbb{R})$ of X can be written compactly as

$$S_X = \frac{1}{n-1} \tilde{X}^T \tilde{X} = \frac{1}{n-1} (C_n X)^T (C_n X) = \frac{1}{n-1} X^T C_n^T C_n X = \frac{1}{n-1} X^T C_n X$$

Algebraic and geometric properties of the centring matrix

Theorem 103 (Properties of the centring matrix C_n)

Let $C_n = \mathbb{I}_n - \frac{1}{n}\mathbb{1}\mathbb{1}^T \in \mathcal{M}_n(\mathbb{R})$. Then:

1. **Symmetry:** $C_n^T = C_n$
2. **Idempotence:** $C_n^2 = C_n$ (and more generally $C_n^k = C_n$ for all integers $k \geq 1$)
3. **Null space:** $C_n\mathbb{1} = \mathbf{0}$ and $\mathbb{1}^T C_n = \mathbf{0}^T$; in fact, $\text{null}(C_n) = \text{span}(\mathbb{1})$
4. **Orthogonal projector:** C_n is the orthogonal projection matrix onto the $(n - 1)$ -dimensional subspace of zero-mean vectors

$$W = \text{span}(\mathbb{1})^\perp = \left\{ \mathbf{v} \in \mathbb{R}^n : \mathbb{1}^T \mathbf{v} = \sum_{i=1}^n v_i = 0 \right\}$$

5. **Spectrum, trace and rank:** the eigenvalues of C_n are $\lambda = 0$ (with multiplicity 1, eigenspace $\text{span}(\mathbb{1})$) and $\lambda = 1$ (with multiplicity $n - 1$, eigenspace W).
Consequently,

$$\text{rank}(C_n) = \text{tr}(C_n) = n - 1$$

Proof of Theorem ?? – Parts 1 and 2

1. C_n is symmetric ($C_n^T = C_n$): Using linearity of the transpose, $\mathbb{I}_n^T = \mathbb{I}_n$ and $(\mathbb{1}\mathbb{1}^T)^T = (\mathbb{1}^T)^T\mathbb{1}^T = \mathbb{1}\mathbb{1}^T$:

$$C_n^T = \left(\mathbb{I}_n - \frac{1}{n} \mathbb{1}\mathbb{1}^T \right)^T = \mathbb{I}_n^T - \frac{1}{n} (\mathbb{1}\mathbb{1}^T)^T = \mathbb{I}_n - \frac{1}{n} \mathbb{1}\mathbb{1}^T = C_n \quad \square$$

2. C_n is idempotent ($C_n^2 = C_n$): Expanding the square and using the associativity of matrix multiplication together with $\mathbb{1}^T\mathbb{1} = n$:

$$\begin{aligned} C_n^2 &= \left(\mathbb{I}_n - \frac{1}{n} \mathbb{1}\mathbb{1}^T \right) \left(\mathbb{I}_n - \frac{1}{n} \mathbb{1}\mathbb{1}^T \right) \\ &= \mathbb{I}_n^2 - \frac{1}{n} \mathbb{1}\mathbb{1}^T - \frac{1}{n} \mathbb{1}\mathbb{1}^T + \frac{1}{n^2} (\mathbb{1}\mathbb{1}^T)(\mathbb{1}\mathbb{1}^T) \\ &= \mathbb{I}_n - \frac{2}{n} \mathbb{1}\mathbb{1}^T + \frac{1}{n^2} \mathbb{1} \underbrace{(\mathbb{1}^T\mathbb{1})}_{=n} \mathbb{1}^T \\ &= \mathbb{I}_n - \frac{2}{n} \mathbb{1}\mathbb{1}^T + \frac{1}{n} \mathbb{1}\mathbb{1}^T = \mathbb{I}_n - \frac{1}{n} \mathbb{1}\mathbb{1}^T = C_n \quad \square \end{aligned}$$

Proof of Theorem ?? – Part 3

3. $C_n \mathbf{1} = \mathbf{0}$, $\mathbf{1}^T C_n = \mathbf{0}^T$ **and** $\text{null}(C_n) = \text{span}(\mathbf{1})$:

First, multiply C_n on the right by $\mathbf{1}$:

$$C_n \mathbf{1} = \left(\mathbb{I}_n - \frac{1}{n} \mathbf{1} \mathbf{1}^T \right) \mathbf{1} = \mathbf{1} - \frac{1}{n} \mathbf{1} (\mathbf{1}^T \mathbf{1}) = \mathbf{1} - \frac{1}{n} \mathbf{1} (n) = \mathbf{1} - \mathbf{1} = \mathbf{0}$$

Taking transposes and using $C_n^T = C_n$ immediately gives

$$\mathbf{1}^T C_n = \mathbf{1}^T C_n^T = (C_n \mathbf{1})^T = \mathbf{0}^T$$

To show $\text{null}(C_n) = \text{span}(\mathbf{1})$, let $\mathbf{v} \in \mathbb{R}^n$. Then

$$\mathbf{v} \in \text{null}(C_n) \iff C_n \mathbf{v} = \mathbf{0} \iff \mathbf{v} - \left(\frac{1}{n} \mathbf{1}^T \mathbf{v} \right) \mathbf{1} = \mathbf{0} \iff \mathbf{v} = \bar{v} \mathbf{1} \iff \mathbf{v} \in \text{span}(\mathbf{1})$$

Thus $\text{null}(C_n) = \text{span}(\mathbf{1})$, which is a 1-dimensional subspace of $\mathbb{R}^n$
($\text{nullity}(C_n) = 1$)



Proof of Theorem ?? – Part 4

4. C_n is the orthogonal projection matrix onto $W = \text{span}(\mathbf{1})^\perp$:

Let $L = \text{span}(\mathbf{1})$. Since $\{\mathbf{1}\}$ is an orthogonal basis of the 1-dimensional subspace L with $\|\mathbf{1}\|^2 = \mathbf{1}^T \mathbf{1} = n$, the orthogonal projection of any vector $\mathbf{v} \in \mathbb{R}^n$ onto L is

$$\text{proj}_L(\mathbf{v}) = \frac{\mathbf{1} \bullet \mathbf{v}}{\|\mathbf{1}\|^2} \mathbf{1} = \frac{\mathbf{1}^T \mathbf{v}}{n} \mathbf{1} = \left(\frac{1}{n} \mathbf{1} \mathbf{1}^T \right) \mathbf{v}$$

By definition of the orthogonal complement $W = L^\perp = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{1}^T \mathbf{v} = 0\}$, the orthogonal projection of $\mathbf{v}$ onto W is the component of $\mathbf{v}$ orthogonal to L :

$$\text{proj}_W(\mathbf{v}) = \text{perp}_L(\mathbf{v}) = \mathbf{v} - \text{proj}_L(\mathbf{v}) = \mathbf{v} - \left(\frac{1}{n} \mathbf{1} \mathbf{1}^T \right) \mathbf{v} = \left(\mathbb{I}_n - \frac{1}{n} \mathbf{1} \mathbf{1}^T \right) \mathbf{v} = C_n \mathbf{v}$$

Note also that for any $\mathbf{v} \in \mathbb{R}^n$, the centred vector $C_n \mathbf{v}$ and the mean vector $\bar{v} \mathbf{1} = (\mathbb{I}_n - C_n) \mathbf{v}$ are orthogonal:

$$(C_n \mathbf{v}) \bullet (\bar{v} \mathbf{1}) = \bar{v} \mathbf{v}^T C_n^T \mathbf{1} = \bar{v} \mathbf{v}^T (C_n \mathbf{1}) = 0 \quad \square$$

Proof of Theorem ?? – Part 5

5. Eigenvalues, trace and rank of C_n :

Let $(\lambda, \mathbf{v})$ be an eigenpair of C_n ($\mathbf{v} \neq \mathbf{0}$ and $C_n \mathbf{v} = \lambda \mathbf{v}$). Applying C_n to both sides and using idempotence $C_n^2 = C_n$:

$$C_n \mathbf{v} = C_n^2 \mathbf{v} = C_n(C_n \mathbf{v}) = C_n(\lambda \mathbf{v}) = \lambda(C_n \mathbf{v}) = \lambda^2 \mathbf{v}$$

Since $\mathbf{v} \neq \mathbf{0}$, $(\lambda^2 - \lambda)\mathbf{v} = \mathbf{0} \implies \lambda(\lambda - 1) = 0$, so every eigenvalue of C_n is either $\lambda = 0$ or $\lambda = 1$:

- ▶ For $\lambda = 0$: the eigenspace is $\text{null}(C_n) = \text{span}(\mathbf{1})$, which has dimension 1 (multiplicity 1)
- ▶ For $\lambda = 1$: $\mathbf{v} \in W = \text{span}(\mathbf{1})^\perp \iff \mathbf{1}^T \mathbf{v} = 0 \implies C_n \mathbf{v} = \mathbf{v} - \frac{1}{n} \mathbf{1}(\mathbf{1}^T \mathbf{v}) = \mathbf{v}$, so the eigenspace for $\lambda = 1$ is W , which has dimension $n - 1$ (multiplicity $n - 1$)

Finally, the trace equals the sum of the diagonal entries (and the sum of the eigenvalues):

$$\text{tr}(C_n) = \sum_{j=1}^n \left(1 - \frac{1}{n}\right) = n \left(\frac{n-1}{n}\right) = n - 1 = \text{rank}(C_n) \quad \square$$





From $D^{(2)}$ back to G : why double centring works

Recall Theorem ??: given the Gram matrix $G = XX^T$, the squared distance matrix is

$$D^{(2)} = \mathbf{g}\mathbf{1}^T - 2G + \mathbf{1}\mathbf{g}^T$$

We now want to **invert** this relationship: given only $D^{(2)}$, how can we isolate G ?

Look closely at the two nuisance terms $\mathbf{g}\mathbf{1}^T$ and $\mathbf{1}\mathbf{g}^T$:

► $\mathbf{g}\mathbf{1}^T$ has $\mathbf{1}^T$ on the **right**, and we know that $\mathbf{1}^T C_n = \mathbf{0}^T$!

► $\mathbf{1}\mathbf{g}^T$ has $\mathbf{1}$ on the **left**, and we know that $C_n \mathbf{1} = \mathbf{0}$!

Therefore, if we multiply $D^{(2)}$ on the **left** by C_n and on the **right** by C_n , both rank-one nuisance terms $\mathbf{g}\mathbf{1}^T$ and $\mathbf{1}\mathbf{g}^T$ are completely annihilated!

This operation $D^{(2)} \mapsto -\frac{1}{2}C_n D^{(2)} C_n$ is called **double centring**

The double centring theorem

Theorem 104 (Double centring formula)

Let $X \in \mathcal{M}_{np}(\mathbb{R})$ be a data matrix with squared Euclidean distance matrix $D^{(2)} = [d_{ij}^2] \in \mathcal{M}_n(\mathbb{R})$, and let $\tilde{X} = C_n X$ be the column-centred data matrix with centred Gram matrix $B = \tilde{X}\tilde{X}^T = C_n X X^T C_n$. Then

$$B = -\frac{1}{2}C_n D^{(2)} C_n = \tilde{X}\tilde{X}^T$$

In particular, if the points in X are already centred at the origin ($C_n X = X$, so $\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}$), then $B = -\frac{1}{2}C_n D^{(2)} C_n = G = X X^T$ recovers the exact Gram matrix of X

Proof of Theorem ??

By Theorem ??, $D^{(2)} = \mathbf{g}\mathbf{1}^T - 2G + \mathbf{1}\mathbf{g}^T$, where $G = XX^T$ and $\mathbf{g} = \text{diag}(G)$
Multiply $D^{(2)}$ on the left by C_n and on the right by C_n :

$$C_n D^{(2)} C_n = C_n (\mathbf{g}\mathbf{1}^T) C_n - 2C_n G C_n + C_n (\mathbf{1}\mathbf{g}^T) C_n$$

Using associativity and Part 3 of Theorem ?? ($C_n \mathbf{1} = \mathbf{0}$ and $\mathbf{1}^T C_n = \mathbf{0}^T$):

$$C_n (\mathbf{g}\mathbf{1}^T) C_n = (C_n \mathbf{g}) (\underbrace{\mathbf{1}^T C_n}_{=\mathbf{0}^T}) = (C_n \mathbf{g}) \mathbf{0}^T = \mathbf{0}_{n \times n}$$

$$C_n (\mathbf{1}\mathbf{g}^T) C_n = (\underbrace{C_n \mathbf{1}}_{=\mathbf{0}}) (\mathbf{g}^T C_n) = \mathbf{0} (\mathbf{g}^T C_n) = \mathbf{0}_{n \times n}$$

Thus both outer-product terms vanish, leaving

$$C_n D^{(2)} C_n = -2C_n G C_n = -2C_n (XX^T) C_n^T = -2(C_n X)(C_n X)^T = -2\tilde{X}\tilde{X}^T$$

Dividing both sides by -2 gives $-\frac{1}{2} C_n D^{(2)} C_n = C_n G C_n = \tilde{X}\tilde{X}^T$. If $C_n X = X$, then $\tilde{X} = X$ and $-\frac{1}{2} C_n D^{(2)} C_n = XX^T = G$



Scalar form of double centring

Why is $B = -\frac{1}{2}C_n D^{(2)} C_n$ called *double centring*?

- ▶ Left-multiplying $D^{(2)}$ by C_n subtracts the **column means** of $D^{(2)}$ from each column
- ▶ Right-multiplying by $C_n = C_n^T$ subtracts the **row means** from each row

Let $\bar{d}_{i\cdot}^2 = \frac{1}{n} \sum_{j=1}^n d_{ij}^2$ (mean of row i), $\bar{d}_{\cdot j}^2 = \frac{1}{n} \sum_{i=1}^n d_{ij}^2$ (mean of column j) and $\bar{d}_{\cdot\cdot}^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2$ (grand mean of $D^{(2)}$). Expanding $B = -\frac{1}{2} (\mathbb{I}_n - \frac{1}{n}\mathbb{J}_n) D^{(2)} (\mathbb{I}_n - \frac{1}{n}\mathbb{J}_n)$ gives the entrywise formula:

$$b_{ij} = -\frac{1}{2} \left(d_{ij}^2 - \bar{d}_{i\cdot}^2 - \bar{d}_{\cdot j}^2 + \bar{d}_{\cdot\cdot}^2 \right), \quad i, j = 1, \dots, n$$

Consequently, every row sum and every column sum of B is zero: $B\mathbb{1} = \mathbf{0}$ and $\mathbb{1}^T B = \mathbf{0}^T$, so $\lambda = 0$ is always an eigenvalue of B with eigenvector $\mathbb{1}$!

The classical MDS algorithm

Definition 105 (Classical MDS algorithm)

Given an $n \times n$ distance or dissimilarity matrix $D = [d_{ij}] \in \mathcal{M}_n(\mathbb{R})$ and a target embedding dimension $k \in \{1, \dots, n-1\}$:

1. **Square the distances:** form $D^{(2)} = [d_{ij}^2] \in \mathcal{M}_n(\mathbb{R})$
2. **Double-centre:** compute the inner product matrix $B = -\frac{1}{2} C_n D^{(2)} C_n$
3. **Spectral decomposition:** compute the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ and orthonormal eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in \mathbb{R}^n$ of B
4. **Construct k -dimensional coordinates:** assuming $\lambda_1 \geq \dots \geq \lambda_k > 0$, let $V_k = [\mathbf{v}_1 \ \dots \ \mathbf{v}_k] \in \mathcal{M}_{nk}(\mathbb{R})$ and $\Lambda_k = \text{diag}(\lambda_1, \dots, \lambda_k) \in \mathcal{M}_k(\mathbb{R})$. The k -dimensional MDS coordinate matrix is

$$X_k = V_k \Lambda_k^{1/2} = (\sqrt{\lambda_1} \mathbf{v}_1 \quad \sqrt{\lambda_2} \mathbf{v}_2 \quad \dots \quad \sqrt{\lambda_k} \mathbf{v}_k) \in \mathcal{M}_{nk}(\mathbb{R})$$

Schoenberg's theorem: when is a distance matrix Euclidean?

What if the input matrix $D = [d_{ij}]$ did *not* come from Euclidean points in $\mathbb{R}^p$, or came from geodesic/road distances on a curved surface?

Theorem 106 (Schoenberg, 1935; Young and Householder, 1938)

Let $D = [d_{ij}] \in \mathcal{M}_n(\mathbb{R})$ be a symmetric hollow matrix ($d_{ii} = 0$, $d_{ij} = d_{ji} \geq 0$) and let $B = -\frac{1}{2}C_n D^{(2)} C_n$. Then:

1. There exist points $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^r$ whose pairwise Euclidean distances equal d_{ij} if and only if B is **positive semi-definite** with $\text{rank}(B) \leq r$
2. The smallest dimension r in which the n points can be isometrically embedded in $\mathbb{R}^r$ is $r = \text{rank}(B)$ (the number of strictly positive eigenvalues of B), and $r \leq n - 1$

Dimensionality reduction, strain and goodness of fit

When $B = -\frac{1}{2}C_n D^{(2)} C_n$ has rank $r > k$ (or small negative eigenvalues due to non-Euclidean noise) and we keep only the top k positive eigenpairs:

- By the **Eckart-Young-Mirsky theorem** (Unit U04), $B_k = X_k X_k^T = \sum_{i=1}^k \lambda_i \mathbf{v}_i \mathbf{v}_i^T$ is the best rank- k positive semi-definite approximation to B in the Frobenius norm, minimising the **classical MDS strain**:

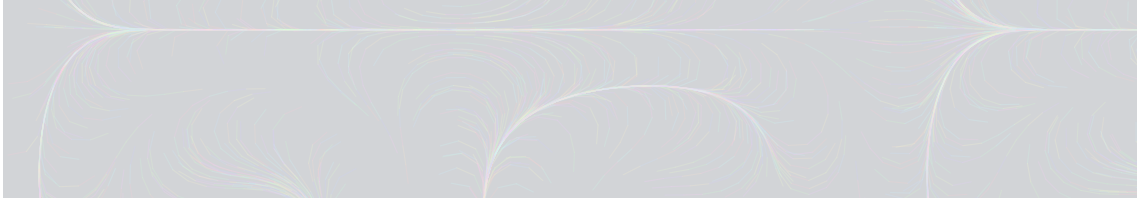
$$\text{Strain}(X_k) = \|B - X_k X_k^T\|_F^2 = \sum_{i=k+1}^n \lambda_i^2$$

- The **goodness of fit** (proportion of variance or dispersion explained) by the first k MDS dimensions:
 - When B is positive semi-definite ($\lambda_i \geq 0$ for all i):

$$\text{GOF}_1(k) = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} = \frac{\sum_{i=1}^k \lambda_i}{\text{tr}(B)}$$

- When B has some negative eigenvalues (two standard conventions in `cmdscale`):

$$\text{GOF}_1(k) = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} \quad \text{or} \quad \text{GOF}_2(k) = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n |\lambda_i|}$$



Example 1 – Recovering 1D coordinates from $D^{(2)}$ (1/4)

Suppose $n = 3$ objects have the squared Euclidean distance matrix

$$D^{(2)} = \begin{pmatrix} 0 & 4 & 16 \\ 4 & 0 & 4 \\ 16 & 4 & 0 \end{pmatrix}$$

(i.e. $d_{12} = 2$, $d_{23} = 2$ and $d_{13} = 4$)

Problem:

1. Compute the double-centred Gram matrix $B = -\frac{1}{2}C_3 D^{(2)} C_3$
2. Find the eigenvalues and orthonormal eigenvectors of B
3. Determine the minimal embedding dimension $r = \text{rank}(B)$ and recover the centred coordinates X_r

Example 1 – Computing $B = -\frac{1}{2}C_3D^{(2)}C_3$ (2/4)

Recall $C_3 = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$. First compute $C_3D^{(2)}$:

$$C_3D^{(2)} = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 0 & 4 & 16 \\ 4 & 0 & 4 \\ 16 & 4 & 0 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} -20 & 4 & 28 \\ -16 & -8 & -16 \\ 36 & 4 & -12 \end{pmatrix} = \frac{4}{3} \begin{pmatrix} -5 & 1 & 7 \\ -4 & -2 & -4 \\ 9 & 1 & -3 \end{pmatrix}$$

Now right-multiply by C_3 and multiply by $-\frac{1}{2}$:

$$C_3D^{(2)}C_3 = \frac{4}{9} \begin{pmatrix} -5 & 1 & 7 \\ -4 & -2 & -4 \\ 9 & 1 & -3 \end{pmatrix} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} = \frac{4}{9} \begin{pmatrix} -18 & 0 & 18 \\ 0 & 0 & 0 \\ 18 & 0 & -18 \end{pmatrix} = \begin{pmatrix} -8 & 0 & 8 \\ 0 & 0 & 0 \\ 8 & 0 & -8 \end{pmatrix}$$

$$\implies B = -\frac{1}{2}C_3D^{(2)}C_3 = \begin{pmatrix} 4 & 0 & -4 \\ 0 & 0 & 0 \\ -4 & 0 & 4 \end{pmatrix}$$

Example 1 – Eigenpairs and coordinate recovery (3/4)

We have

$$B = \begin{pmatrix} 4 & 0 & -4 \\ 0 & 0 & 0 \\ -4 & 0 & 4 \end{pmatrix}$$

- ▶ Notice that row 1 and row 3 are opposites and row 2 is zero, so $\text{rank}(B) = 1$
- ▶ Since $\text{tr}(B) = 4 + 0 + 4 = 8$ and $\text{rank}(B) = 1$, the eigenvalues of B are immediately

$$\lambda_1 = 8, \quad \lambda_2 = 0, \quad \lambda_3 = 0$$

- ▶ For $\lambda_1 = 8$, $(B - 8\mathbb{I}_3)\mathbf{v}_1 = \mathbf{0} \implies v_{11} + v_{13} = 0$ and $v_{12} = 0$. Normalising gives the unit eigenvector

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \quad (\text{or } -\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix})$$

Example 1 – Verification of recovered 1D coordinates (4/4)

Since B is positive semi-definite with $r = \text{rank}(B) = 1$, the 3 points lie on a 1-dimensional line ($\mathbb{R}^1$)!

Recovering the 1-dimensional coordinate vector $X_1 = \sqrt{\lambda_1} \mathbf{v}_1 \in \mathcal{M}_{31}(\mathbb{R})$:

$$X_1 = \sqrt{8} \left(\frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right) = \frac{2\sqrt{2}}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \\ 2 \end{pmatrix}$$

Check:

- ▶ Centred at the origin: $x_1 + x_2 + x_3 = -2 + 0 + 2 = 0$
- ▶ Pairwise squared distances:

$$(x_1 - x_2)^2 = (-2 - 0)^2 = 4, \quad (x_2 - x_3)^2 = (0 - 2)^2 = 4, \quad (x_1 - x_3)^2 = (-2 - 2)^2 = 16$$

which reproduces $D^{(2)}$ exactly! (Choosing $-\mathbf{v}_1$ gives $X_1 = (2, 0, -2)^T$, the reflection across the origin)

Example 2 – Recovering a 2D triangle from distances (1/4)

Now consider $n = 3$ points with squared Euclidean distance matrix

$$D^{(2)} = \begin{pmatrix} 0 & 6 & 6 \\ 6 & 0 & 12 \\ 6 & 12 & 0 \end{pmatrix}$$

(an isosceles right triangle with leg length $\sqrt{6}$ and hypotenuse $\sqrt{12}$, since $6 + 6 = 12$)

Let us compute $B = -\frac{1}{2} C_3 D^{(2)} C_3$ using the row/column mean formula

$$b_{ij} = -\frac{1}{2} (d_{ij}^2 - \bar{d}_{i.}^2 - \bar{d}_{.j}^2 + \bar{d}_{..}^2):$$

► Row (and column) means of $D^{(2)}$:

$$\bar{d}_{1.}^2 = \frac{0 + 6 + 6}{3} = 4, \quad \bar{d}_{2.}^2 = \frac{6 + 0 + 12}{3} = 6, \quad \bar{d}_{3.}^2 = \frac{6 + 12 + 0}{3} = 6$$

► Grand mean of $D^{(2)}$:

$$\bar{d}_{..}^2 = \frac{4 + 6 + 6}{3} = \frac{16}{3}$$

Example 2 – Forming B and finding its eigenpairs (2/4)

Computing each entry $b_{ij} = -\frac{1}{2}(d_{ij}^2 - \bar{d}_i^2 - \bar{d}_j^2 + \frac{16}{3})$ (or multiplying $-\frac{1}{2}C_3D^{(2)}C_3$ directly):

$$B = -\frac{1}{2}C_3D^{(2)}C_3 = \frac{1}{3} \begin{pmatrix} 4 & -2 & -2 \\ -2 & 10 & -8 \\ -2 & -8 & 10 \end{pmatrix} = \begin{pmatrix} 4/3 & -2/3 & -2/3 \\ -2/3 & 10/3 & -8/3 \\ -2/3 & -8/3 & 10/3 \end{pmatrix}$$

Let us find the eigenvalues and orthonormal eigenvectors of B :

- ▶ We always know $\mathbf{v}_3 = \frac{1}{\sqrt{3}}(1, 1, 1)^T$ is an eigenvector with eigenvalue $\lambda_3 = 0$
- ▶ By symmetry between points 2 and 3, try $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(0, 1, -1)^T$:

$$B\mathbf{v}_1 = \frac{1}{3\sqrt{2}} \begin{pmatrix} 4 & -2 & -2 \\ -2 & 10 & -8 \\ -2 & -8 & 10 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = \frac{1}{3\sqrt{2}} \begin{pmatrix} 0 \\ 18 \\ -18 \end{pmatrix} = 6\mathbf{v}_1 \implies \lambda_1 = 6$$

Example 2 – Second eigenpair and 2D coordinates (3/4)

- ▶ Since $\text{tr}(B) = \frac{4+10+10}{3} = 8 = \lambda_1 + \lambda_2 + \lambda_3$ and $\lambda_1 = 6, \lambda_3 = 0$, the remaining eigenvalue is

$$\lambda_2 = 8 - 6 - 0 = 2$$

- ▶ The unit eigenvector $\mathbf{v}_2$ must be orthogonal to both $\mathbf{v}_3 = \frac{1}{\sqrt{3}}(1, 1, 1)^T$ and $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(0, 1, -1)^T$, which gives

$$\mathbf{v}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \quad \left(\text{check: } B\mathbf{v}_2 = \frac{1}{3\sqrt{6}} \begin{pmatrix} 12 \\ -6 \\ -6 \end{pmatrix} = 2\mathbf{v}_2 \right)$$

Since $\lambda_1 = 6 > 0$, $\lambda_2 = 2 > 0$ and $\lambda_3 = 0$, $\text{rank}(B) = 2$, so the 3 points embed isometrically in $\mathbb{R}^2$!

Example 2 – Constructing X_2 and verifying distances (4/4)

Using $X_2 = V_2 \Lambda_2^{1/2} = (\sqrt{\lambda_1} \mathbf{v}_1 \quad \sqrt{\lambda_2} \mathbf{v}_2)$ with $\sqrt{\lambda_1} = \sqrt{6}$ and $\sqrt{\lambda_2} = \sqrt{2}$:

$$X_2 = \begin{pmatrix} \sqrt{6} \cdot 0 & \sqrt{2} \cdot \frac{2}{\sqrt{6}} \\ \sqrt{6} \cdot \frac{1}{\sqrt{2}} & \sqrt{2} \cdot \frac{-1}{\sqrt{6}} \\ \sqrt{6} \cdot \frac{-1}{\sqrt{2}} & \sqrt{2} \cdot \frac{-1}{\sqrt{6}} \end{pmatrix} = \begin{pmatrix} 0 & \frac{2}{\sqrt{3}} \\ \sqrt{3} & -\frac{1}{\sqrt{3}} \\ -\sqrt{3} & -\frac{1}{\sqrt{3}} \end{pmatrix} \in \mathcal{M}_{32}(\mathbb{R})$$

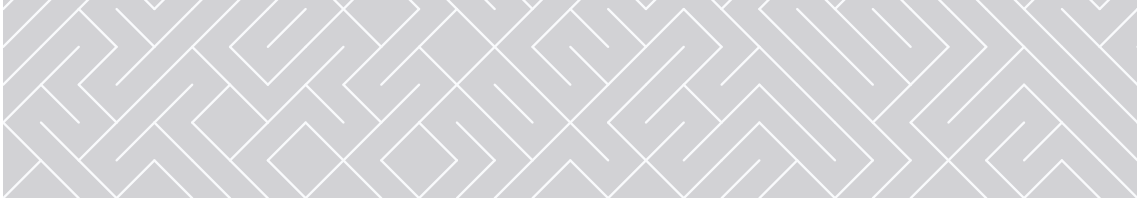
Verification:

- ▶ Column sums are zero: $0 + \sqrt{3} - \sqrt{3} = 0$ and $\frac{2}{\sqrt{3}} - \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{3}} = 0$
- ▶ Pairwise squared distances between rows $\mathbf{x}_1^T, \mathbf{x}_2^T, \mathbf{x}_3^T$ of X_2 :

$$\|\mathbf{x}_1 - \mathbf{x}_2\|^2 = (0 - \sqrt{3})^2 + \left(\frac{2}{\sqrt{3}} + \frac{1}{\sqrt{3}}\right)^2 = 3 + 3 = 6 = d_{12}^2$$

$$\|\mathbf{x}_1 - \mathbf{x}_3\|^2 = (0 + \sqrt{3})^2 + \left(\frac{2}{\sqrt{3}} + \frac{1}{\sqrt{3}}\right)^2 = 3 + 3 = 6 = d_{13}^2$$

$$\|\mathbf{x}_2 - \mathbf{x}_3\|^2 = (\sqrt{3} + \sqrt{3})^2 + 0^2 = (2\sqrt{3})^2 = 12 = d_{23}^2$$





Two $n \times p$ products: $\tilde{X}^T \tilde{X}$ versus $\tilde{X} \tilde{X}^T$

Suppose we have a data matrix $X \in \mathcal{M}_{np}(\mathbb{R})$ (n observations and p features) and its column-centred data matrix $\tilde{X} = C_n X \in \mathcal{M}_{np}(\mathbb{R})$. From $\tilde{X}$, we can form two symmetric positive semi-definite matrices:

Feature scatter / covariance (PCA, Unit U05)

$$\tilde{X}^T \tilde{X} \in \mathcal{M}_p(\mathbb{R}) \quad \left(\text{or } S_X = \frac{1}{n-1} \tilde{X}^T \tilde{X} \right)$$

- ▶ Size: $p \times p$ (features $\times$ features)
- ▶ (j, ℓ) entry: inner product between centred column j and centred column ℓ
- ▶ Diagonalising $\tilde{X}^T \tilde{X}$ gives the principal directions (loadings) in $\mathbb{R}^p$

Centred Gram matrix (Classical MDS, Unit U07)

$$B = \tilde{X} \tilde{X}^T = -\frac{1}{2} C_n D^{(2)} C_n \in \mathcal{M}_n(\mathbb{R})$$

- ▶ Size: $n \times n$ (observations $\times$ observations)
- ▶ (i, j) entry: inner product $\tilde{\mathbf{x}}_i^T \tilde{\mathbf{x}}_j$ between centred observation i and centred observation j
- ▶ Diagonalising $\tilde{X} \tilde{X}^T$ gives the principal coordinates in $\mathbb{R}^n$

Connecting PCA and classical MDS via the SVD

Let $\tilde{X} = C_n X \in \mathcal{M}_{np}(\mathbb{R})$ be the column-centred data matrix of rank $r \leq \min(n, p)$, and write its thin singular value decomposition (Unit U04):

$$\tilde{X} = U_r \Sigma_r V_r^T$$

where:

- ▶ $U_r = [\mathbf{u}_1 \cdots \mathbf{u}_r] \in \mathcal{M}_{nr}(\mathbb{R})$ has orthonormal columns ($U_r^T U_r = \mathbb{I}_r$)
- ▶ $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r) \in \mathcal{M}_r(\mathbb{R})$ contains the positive singular values $\sigma_1 \geq \dots \geq \sigma_r > 0$
- ▶ $V_r = [\mathbf{v}_1 \cdots \mathbf{v}_r] \in \mathcal{M}_{pr}(\mathbb{R})$ has orthonormal columns ($V_r^T V_r = \mathbb{I}_r$)

Now express both $\tilde{X}^T \tilde{X}$ and $B = \tilde{X} \tilde{X}^T$ in terms of the SVD of $\tilde{X}$:

$$\tilde{X}^T \tilde{X} = V_r \Sigma_r^2 V_r^T \in \mathcal{M}_p(\mathbb{R}) \quad \text{and} \quad B = \tilde{X} \tilde{X}^T = U_r \Sigma_r^2 U_r^T \in \mathcal{M}_n(\mathbb{R})$$

Duality theorem: PCA scores equal classical MDS coordinates

Theorem 107 (Equivalence of PCA and Euclidean classical MDS)

Let $X \in \mathcal{M}_{np}(\mathbb{R})$ be a data matrix with column-centred data matrix $\tilde{X} = C_n X$ having thin SVD $\tilde{X} = U_r \Sigma_r V_r^T$, and let $D = [\|\mathbf{x}_i - \mathbf{x}_j\|] \in \mathcal{M}_n(\mathbb{R})$ be the Euclidean distance matrix between the rows of X . Then for any $k \leq r$:

1. The non-zero eigenvalues of $\tilde{X}^T \tilde{X} \in \mathcal{M}_p(\mathbb{R})$ and $B = -\frac{1}{2} C_n D^{(2)} C_n = \tilde{X} \tilde{X}^T \in \mathcal{M}_n(\mathbb{R})$ are identical:

$$\lambda_i = \sigma_i^2 > 0, \quad i = 1, \dots, r$$

2. The k -dimensional principal component score matrix $Y_k = \tilde{X} V_k \in \mathcal{M}_{nk}(\mathbb{R})$ from PCA and the k -dimensional coordinate matrix $X_k = U_k \Lambda_k^{1/2} \in \mathcal{M}_{nk}(\mathbb{R})$ from classical MDS on D are **identical** (up to sign choices of the eigenvectors):

$$Y_k = \tilde{X} V_k = U_k \Sigma_k = U_k \Lambda_k^{1/2} = X_k$$

Proof of Theorem ??

Let $\tilde{X} = U_r \Sigma_r V_r^T$, where $U_r = [\mathbf{u}_1 \cdots \mathbf{u}_r]$, $V_r = [\mathbf{v}_1 \cdots \mathbf{v}_r]$ and $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r)$

1. PCA projection onto top k right singular vectors: The columns of $V_k = [\mathbf{v}_1 \cdots \mathbf{v}_k] \in \mathcal{M}_{pk}(\mathbb{R})$ are the top k orthonormal eigenvectors of $\tilde{X}^T \tilde{X}$ (and of $S_X = \frac{1}{n-1} \tilde{X}^T \tilde{X}$). Using $V_r^T V_k = \begin{pmatrix} \mathbb{I}_k \\ \mathbf{0}_{(r-k) \times k} \end{pmatrix}$, the PCA score matrix is

$$Y_k = \tilde{X} V_k = (U_r \Sigma_r V_r^T) V_k = U_k \Sigma_k$$

2. Classical MDS on Euclidean distances of X : By Theorem ??, double centring $D^{(2)}$ yields $B = -\frac{1}{2} C_n D^{(2)} C_n = \tilde{X} \tilde{X}^T = U_r \Sigma_r^2 U_r^T$. Thus the top k eigenvalues of B are $\lambda_i = \sigma_i^2$ ($\Lambda_k^{1/2} = \Sigma_k$) with orthonormal eigenvectors $U_k = [\mathbf{u}_1 \cdots \mathbf{u}_k] \in \mathcal{M}_{nk}(\mathbb{R})$. The classical MDS coordinate matrix is therefore

$$X_k = U_k \Lambda_k^{1/2} = U_k \Sigma_k = Y_k \quad \square$$

Computational and practical payoff of PCA–MDS duality

Why is Theorem ?? so useful in practice?

1. Tall-and-skinny data ($n \gg p$, many observations, few features):

- ▶ $\tilde{X}^T \tilde{X}$ is a small $p \times p$ matrix, whereas $\tilde{X} \tilde{X}^T$ is a huge $n \times n$ matrix
- ▶ Computing PCA via $\tilde{X}^T \tilde{X}$ costs $O(np^2 + p^3)$, which is much faster than $O(n^3)$ for $\tilde{X} \tilde{X}^T$

2. Short-and-wide data ($p \gg n$, few observations, huge number of features):

- ▶ E.g. gene expression ($n = 50$ patients, $p = 20,000$ genes) or high-resolution images ($n = 100$ faces, $p = 10^6$ pixels)
- ▶ $\tilde{X}^T \tilde{X}$ is $20,000 \times 20,000$, whereas the Gram matrix $B = \tilde{X} \tilde{X}^T$ is only 50×50 !
- ▶ Diagonalising the $n \times n$ Gram matrix $B = \tilde{X} \tilde{X}^T$ (classical MDS) gives the exact PCA scores in $O(n^2 p + n^3)$ time!

3. Coordinate-free and kernel extensions:

- ▶ Classical MDS works when X is unknown and only distances D are given
- ▶ Replacing $\mathbf{x}_i^T \mathbf{x}_j$ by a general similarity kernel $k(\mathbf{x}_i, \mathbf{x}_j)$ leads directly to **kernel PCA**!

Verifying PCA and classical MDS equivalence in R

Let us first verify Theorem ?? numerically in R on a synthetic 5×3 centred data matrix $\tilde{X}$:

```
set.seed(2740)
X_tilde = scale(matrix(rnorm(15), nrow = 5, ncol = 3), center = TRUE, scale
pca_scores = prcomp(X_tilde, center = FALSE)$x[, 1:2]
mds_coords = cmdscale(dist(X_tilde), k = 2)
# Compare absolute values (eigenvectors are defined up to a sign +/- 1)
max(abs(abs(pca_scores) - abs(mds_coords)))

## [1] 2.442491e-15
```

Up to machine precision ($\approx 10^{-16}$) and column signs, the PCA scores `prcomp(X_tilde)$x[,1:2]` and the classical MDS coordinates `cmdscale(dist(X_tilde), k=2)` are identical!

Geographical example: Canadian cities distance matrix

Now suppose we are only given approximate great-circle distances (in kilometres) between $n = 8$ Canadian cities, with **no latitude or longitude coordinates**:

```
cities = c("Victoria", "Vancouver", "Calgary", "Regina",  
           "Winnipeg", "Toronto", "Montreal", "Halifax")  
D_can = matrix(c(  
    0,    90,   725, 1215, 1690, 3390, 3760, 4490,  
    90,    0,   675, 1165, 1640, 3355, 3685, 4430,  
    725,  675,    0,   505,   990, 2705, 3020, 3770,  
    1215, 1165,   505,    0,   500, 2030, 2360, 3125,  
    1690, 1640,   990,   500,    0, 1515, 1825, 2590,  
    3390, 3355, 2705, 2030, 1515,    0,   505, 1265,  
    3760, 3685, 3020, 2360, 1825,   505,    0,   790,  
    4490, 4430, 3770, 3125, 2590, 1265,   790,    0  
), nrow = 8, byrow = TRUE, dimnames = list(cities, cities))
```

Step-by-step manual classical MDS vs cmdscale

Let us implement the classical MDS algorithm directly from $B = -\frac{1}{2}C_n D^{(2)} C_n$ and compare with R's built-in `cmdscale` function:

```
n = nrow(D_can)
C_n = diag(n) - matrix(1, n, n) / n
B = -0.5 * C_n %*% (D_can^2) %*% C_n
ev = eigen(B, symmetric = TRUE)
round(ev$values, 1)

## [1] 20985217.2    224941.6    67722.5    17544.1    3459.2    0.0
## [8] -482729.4

X_manual = ev$vectors[, 1:2] %*% diag(sqrt(ev$values[1:2]))
rownames(X_manual) = cities
mds_builtin = cmdscale(as.dist(D_can), k = 2, eig = TRUE)
max(abs(abs(X_manual) - abs(mds_builtin$points)))

## [1] 0
```

Eigenvalue spectrum and goodness of fit

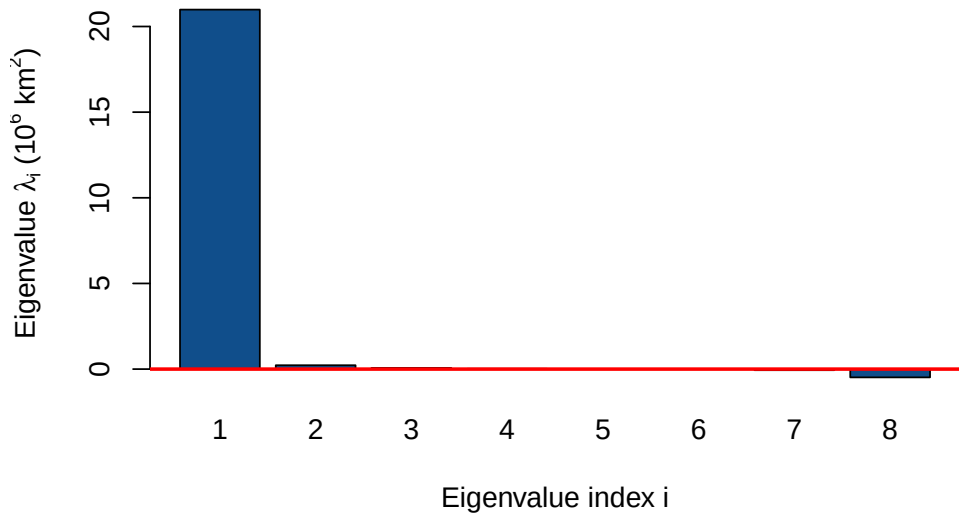
Notice how the first two eigenvalues dominate the spectrum of B :

```
gof_top2 = sum(ev$values[1:2]) / sum(ev$values[ev$values > 0])
gof_1d    = ev$values[1] / sum(ev$values[ev$values > 0])
c(GOF_1D = round(gof_1d, 4), GOF_2D = round(gof_top2, 4))

## GOF_1D GOF_2D
## 0.9853 0.9958
```

The first dimension alone (East–West axis across Canada) explains 98.5% of the positive trace, and 2 dimensions explain 99.58%!

Eigenvalues of double-centred Gram matrix B (Canadian cities)

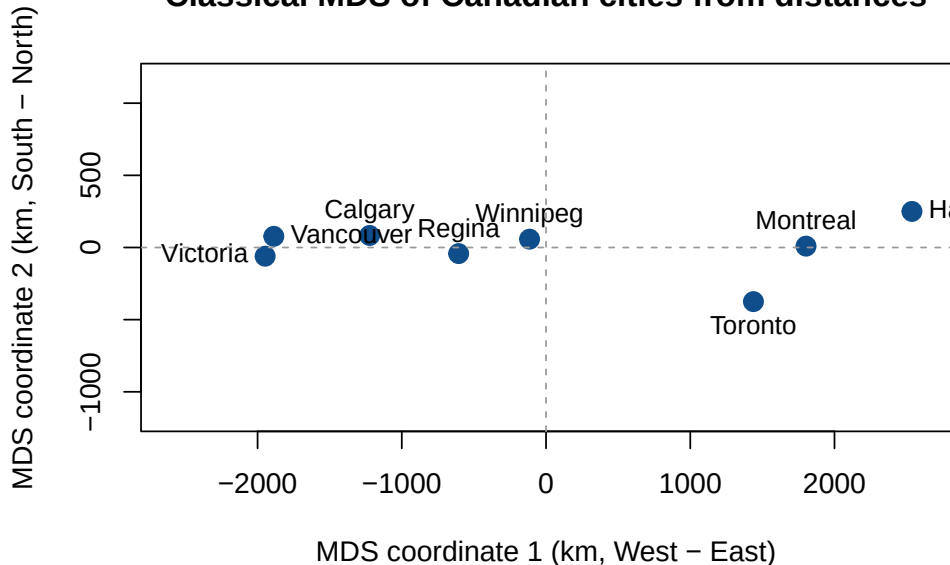


Plotting the reconstructed 2D map of Canadian cities

Recall that classical MDS only determines coordinates up to orthogonal transformations (rotations and reflections). We plot the raw MDS coordinates and then flip signs if needed so West is left and North is up:

```
coords = X_manual
# Ensure West < 0 and Winnipeg is North of Toronto
if (coords["Vancouver", 1] > 0) coords[, 1] = -coords[, 1]
if (coords["Winnipeg", 2] < coords["Toronto", 2]) coords[, 2] = -coords[, 2]
plot(coords[, 1], coords[, 2], pch = 19, cex = 1.5, col = "dodgerblue4",
      asp = 1, xlim = c(-2600, 2600), ylim = c(-900, 900),
      xlab = "MDS coordinate 1 (km, West - East)",
      ylab = "MDS coordinate 2 (km, South - North)",
      main = "Classical MDS of Canadian cities from distances")
abline(h = 0, v = 0, lty = 2, col = "grey60")
```

Classical MDS of Canadian cities from distances



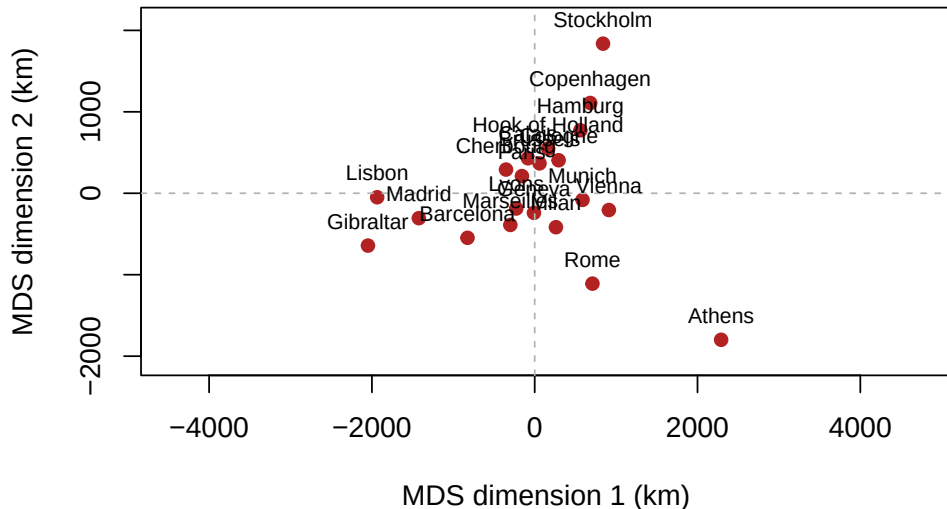
Classical example: European road distances (eurodist)

R's built-in datasets package includes eurodist, a table of road distances (in km) between 21 European cities:

```
D_euro = as.matrix(eurodist)
mds_euro = cmdscale(eurodist, k = 2, eig = TRUE)
x_eu = mds_euro$points[, 1]
y_eu = -mds_euro$points[, 2] # Flip North-South axis
plot(x_eu, y_eu, pch = 19, col = "firebrick", asp = 1,
     xlim = range(x_eu) * 1.15, ylim = range(y_eu) * 1.15,
     xlab = "MDS dimension 1 (km)", ylab = "MDS dimension 2 (km)",
     main = "Classical MDS of 21 European cities (eurodist)")
abline(h = 0, v = 0, lty = 2, col = "grey70")
```

Even though road distances include detours around mountains and seas (so B has small negative eigenvalues), the top 2 eigenvectors of $-\frac{1}{2}C_{21}D^{(2)}C_{21}$ faithfully reconstruct the map of Europe!

Classical MDS of 21 European cities (eurodist)



Summary of Unit U07

- ▶ **Gram matrix:** for $X \in \mathcal{M}_{np}(\mathbb{R})$, $G = XX^T \in \mathcal{M}_n(\mathbb{R})$ has entries $g_{ij} = \mathbf{x}_i^T \mathbf{x}_j$, is symmetric positive semi-definite and satisfies $\text{rank}(G) = \text{rank}(X)$
- ▶ **From Gram to squared distances:** with $\mathbf{g} = \text{diag}(G)$,

$$D^{(2)} = \mathbf{g}\mathbf{1}^T - 2G + \mathbf{1}\mathbf{g}^T$$

- ▶ **Centring matrix:** $C_n = \mathbb{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is the symmetric, idempotent orthogonal projection matrix onto $\text{span}(\mathbf{1})^\perp$ with $\text{rank}(C_n) = \text{tr}(C_n) = n - 1$ and $C_n\mathbf{1} = \mathbf{0}$
- ▶ **Double centring and classical MDS:** $B = -\frac{1}{2}C_nD^{(2)}C_n = \tilde{X}\tilde{X}^T$ recovers the centred Gram matrix from distances alone, and $X_k = V_k\Lambda_k^{1/2}$ recovers k -dimensional coordinates
- ▶ **PCA–MDS duality:** via the SVD $\tilde{X} = U\Sigma V^T$, classical MDS on Euclidean distances of X produces the exact same low-dimensional coordinates $U_k\Sigma_k$ as PCA on X