GMM Appendix B: Covariance Matrix Structure in Multivariate Gaussians

The Multivariate Gaussian[1]

For a d-dimensional Gaussian distribution:

$$p(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

The covariance matrix $\boldsymbol{\Sigma}$ is a $d \times d$ positive definite matrix that completely determines the shape, orientation, and spread of the distribution.

1. Courtesy of Claude.ai

Different Forms of Covariance Matrices

1. Full Covariance Matrix (Unrestricted)

$$\Sigma_{\text{full}} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_d^2 \end{pmatrix}$$

Properties:

- Symmetric: $\sigma_{ij} = \sigma_{ji}$
- Number of free parameters: $\frac{d(d+1)}{2}$
- Can represent any orientation and shape

2. Diagonal Covariance Matrix

$$\Sigma_{\text{diag}} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d^2 \end{pmatrix}$$

Properties:

- All off-diagonal elements are zero: $\sigma_{ij} = 0$ for $i \neq j$
- Number of free parameters: d
- Dimensions are **independent** (uncorrelated)

Geometric Interpretation:

- Ellipsoids are **axis-aligned** (principal axes parallel to coordinate axes)
- Each dimension can have different spread

3. Spherical (Isotropic) Covariance Matrix

$$\Sigma_{\text{spherical}} = \sigma^2 \mathbf{I} = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}$$

Properties:

- All dimensions have the **same variance:** $\sigma_i^2 = \sigma^2$ for all i
- Number of free parameters: 1
- Special case of diagonal covariance

Geometric Interpretation: - Contours are **hyperspheres** (circles in 2D, spheres in 3D) - Equal spread in all directions - No preferred direction

Example (2D):

4. Tied (Shared) Covariance Matrix

All mixture components share the **same** covariance matrix:

$$\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_K = \mathbf{\Sigma}_{\text{shared}}$$

Properties:

- Can be full, diagonal, or spherical
- Number of parameters doesn't scale with K
- All clusters have the same shape and orientation

Geometric Interpretation:

- All ellipsoids have the same shape, size, and orientation
- Only the centers (means) differ between components
- This is equivalent to **Linear Discriminant Analysis (LDA)** for classification

Usage in GMMs:

Visual Comparison of All Types

▶ Code

GMM Example with Different Covariance Types

▶ Code

```
BIC Scores (lower is better):
  Full: 3538.70
  Diag: 3782.72
  Spherical: 3846.10
  Tied: 3815.55
```

Number of Parameters

For a GMM with K components in d dimensions:

Covariance Type	Parameters per Component	Total Covariance Parameters
Full	$\frac{d(d+1)}{2}$	$K \cdot \frac{d(d+1)}{2}$
Diagonal	d	$K \cdot d$
Spherical	1	K
Tied Full	$\frac{d(d+1)}{2}$	$\frac{d(d+1)}{2}$
Tied Diagonal	d	d
Tied Spherical	1	1

▶ Code

```
Parameter counts for K=5 components in d=10 dimensions:

  Full:            275 parameters
  Diagonal:        50 parameters
  Spherical:       5 parameters
  Tied Full:       55 parameters
```

Choosing the Right Covariance Structure

Use Full Covariance when:

- You have lots of data relative to dimensionality
- Clusters have different shapes and orientations
- Features are correlated within clusters
- Maximum flexibility is needed

Use Diagonal Covariance when:

- Features are approximately independent
- You want computational efficiency
- Data is moderately sized
- A good default choice for many applications