12.20

P.302の下の方にあるとおり

$$p(\mathbf{x}|\boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\Psi}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C}), \quad \mathbf{C} = \mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi}$$

である。これより尤度関数は

$x_1 \cdots x_D$は独立なので

$$p(\mathbf{X}|\boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\Psi}) = \prod_{n=1}^{N} p(\mathbf{x}_n|\boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\Psi}) = \prod_{n=1}^{N} \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}, \mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi})$$

となる。対数をとると

$$\ln p(\mathbf{X}|\boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\Psi}) = \sum_{n=1}^{N} \ln \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}, \mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi})$$

$$= \sum_{n=1}^{N} \ln \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}_n - \boldsymbol{\mu})^T (\mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi})^{-1} (\mathbf{x}_n - \boldsymbol{\mu})\right\}$$

$$= \sum_{n=1}^{N} \left\{-\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi}| - \frac{1}{2}(\mathbf{x}_n - \boldsymbol{\mu})^T (\mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi})^{-1} (\mathbf{x}_n - \boldsymbol{\mu})\right\}$$

となる。$\boldsymbol{\mu}$について微分すると

$$\frac{\partial \ln p(\mathbf{X}|\boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\Psi})}{\partial \boldsymbol{\mu}} = \sum_{n=1}^{N} -\frac{1}{2}(-1)\,2(\mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi})^{-1}(\mathbf{x}_n - \boldsymbol{\mu})$$

$$= \sum_{n=1}^{N} (\mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi})^{-1}(\mathbf{x}_n - \boldsymbol{\mu}) \quad \cdots ①$$

- $\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x} = 2\mathbf{A}\mathbf{x}$ ← denominator layout

  ↑$\mathbf{A}$が対称の場合
- $\mathbf{W}\mathbf{W}^T = (\mathbf{u}_1 \cdots \mathbf{u}_n)\begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{pmatrix}$

  $= \mathbf{u}_1\mathbf{u}_1^T + \cdots + \mathbf{u}_n\mathbf{u}_n^T$

  $\mathbf{u}_i\mathbf{u}_i^T$は対称なので$\mathbf{W}\mathbf{W}^T$は対称

  $\boldsymbol{\Psi} = \begin{pmatrix} \phi_1^2 & & 0 \\ & \ddots & \\ 0 & & \phi_D^2 \end{pmatrix}$は対称　よって$\mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi}$は対称
- ベクトル微分のchain rule

  $g(\mathbf{u}), \mathbf{u}(\mathbf{x})$とすると

  $\frac{\partial g(\mathbf{u})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial g}{\partial \mathbf{u}}$ (denominator layout)

さらに$\boldsymbol{\mu}$について微分すると

$$\frac{\partial^2 \ln p(\mathbf{X}|\boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\Psi})}{\partial \boldsymbol{\mu}^2} = \frac{\partial}{\partial \boldsymbol{\mu}} \sum_{n=1}^{N} (\mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi})^{-1}(\mathbf{x}_n - \boldsymbol{\mu})$$

(C.19) $\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{a} = \frac{\partial}{\partial \mathbf{x}} \mathbf{a}^T \mathbf{x} = \mathbf{a}$ (denominator layout)

$$= \sum_{n=1}^{N} (-1)(\mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi})^{-1}$$

$$= -N(\mathbf{W}\mathbf{W}^T + \boldsymbol{\Psi})^{-1} \quad \cdots ②$$

となる。ここで、任意の$\mathbf{a} \neq 0$に対して

$$\mathbf{a}^T \mathbf{W}\mathbf{W}^T \mathbf{a} = (\mathbf{W}^T\mathbf{a})^T(\mathbf{W}^T\mathbf{a}) = (\mathbf{W}^T\mathbf{a})^2$$

少なくとも1つは主成分があると仮定してるので
少なくとも1つの$\mathbf{u}_i \neq 0$であるので

$$= \left(\begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_M^T \end{pmatrix} \mathbf{a}\right)^2 = \begin{pmatrix} \mathbf{u}_1^T \mathbf{a} \\ \vdots \\ \mathbf{u}_M^T \mathbf{a} \end{pmatrix}^2 = (\mathbf{u}_1^T \mathbf{a})^2 + \cdots + (\mathbf{u}_M^T \mathbf{a})^2 > 0$$

なので$\mathbf{W}\mathbf{W}^T$は正定値である。

また

$a^T \Phi a = a^T \begin{pmatrix} \phi_1^2 & & 0 \\ & \ddots & \\ 0 & & \phi_D^2 \end{pmatrix} a = (\phi_1 a_1)^2 + \cdots + (\phi_D a_D)^2 \geq 0$，等号成立は $\phi_1 = \cdots = \phi_D = 0$ のとき

なので，Φ は半正定値である。

なので $WW^T + \Phi$ は正定値である。

$a^T (WW^T + \Phi) a = a^T WW^T a + a^T \Phi a > 0$

正定値と半正定値の和は正定値である。

なので $(WW^T + \Phi)^{-1}$ は正定値である。

正定値行列の固有値 $\lambda_i > 0$

これの逆行列の固有値 $\frac{1}{\lambda_i} > 0$ なので

正定値の逆行列も正定値である

なので，② より

対数尤度関数の μ に関する2階微分は，負定値である。

よって対数尤度関数は，$\sqcap$ 関数である。なので停留点が最大値となる。

停留点は①を 0 とすることで得られる

$$0 = \sum_{n=1}^{N} (WW^T + \Phi)^{-1} (x_n - \mu)$$

$$= (WW^T + \Phi)^{-1} \left\{ \left(\sum_{n=1}^{N} x_n \right) - N\mu \right\}$$

$$\therefore \left(\sum_{n=1}^{N} x_n \right) - N\mu = 0$$

$$\therefore \mu = \frac{1}{N} \sum_{n=1}^{N} x_n$$

$$\therefore \mu = \bar{x}$$

を得る。