Probabilistic latent variable models for shape correspondence analysis

Hernán F. García and Mauricio A. Álvarez

October 6, 2017

Abstract

This report provides a detailed analysis of the mathematical aspects related to our model for shape correspondence analysis.

1 The model

Based on the Iwatta's paper (see [1, 2]), we develop a non linear model for shape correspondence analysis using probabilistic latent variable models.

Suppose that we are given objects in D domains $\mathcal{X} = \{\mathbf{X}_d\}_{d=1}^{D}$ mapped to a Hilbert space $\mathcal{H}$, where $\mathbf{X}_d = \{\mathbf{x}_{dn}\}_{n=1}^{N_d}$ is a set of objects in the dth domain, and $\mathbf{x}_{dn} \in \mathbb{R}^{M_d}$ is the feature vector of the nth object in the dth domain. By introducing a function $k : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ called the *kernel*, that performs a given mapping over the objects, $\phi : \mathcal{X} \to \mathcal{H}$ such that $\forall x, x' \in \mathcal{X}$,

$$k(x, x') := \langle \phi(x), \phi(x') \rangle_{\mathcal{H}}, \tag{1}$$

we can cluster groups of correspondences by using a non-linear function that represents the shape descriptors in the Hilbert space.

Our notation is summarized in Table 1. Note that we are unaware of any correspondence between objects in different domains. The number of objects N_d and the dimensionality L_d for each domain in $\mathcal{H}$ can be different from those of other domains. Therefore, our task is to match clusters of descriptors (groupwise correspondences) across multiple brain structures in an unsupervised manner [2].

As in infinite Gaussian mixture models, our approach assumes that there are an infinite number of clusters related to each correspondence, and each cluster j has a latent vector $\boldsymbol{\zeta}_j \in \mathbb{R}^P$ in a latent space of dimension P. Descriptors that have the same cluster assignments s_{dn} are related by the same latent vector and considered to match (establish a groupwise correspondence).

Each object in $\boldsymbol{\phi}(\mathbf{x}_{dn}) \in \mathcal{H}$ in the dth domain is generated depending on the domain-specific projection matrix $\boldsymbol{\mathcal{W}}_d = \left[\boldsymbol{\varphi}(\mathbf{w}_1^d), \boldsymbol{\varphi}(\mathbf{w}_2^d), \ldots, \boldsymbol{\varphi}(\mathbf{w}_P^d)\right], \quad \boldsymbol{\mathcal{W}}_d \in \mathbb{R}^{L_d \times P}$ and latent vector $\boldsymbol{\zeta}_{s_{dn}}$ that is selected from a set of latent vectors $\mathbf{Z} = \{\boldsymbol{\zeta}_j\}_{j=1}^{\infty}$. Here, $s_{dn} = \{1, \ldots, \infty\}$ is the latent cluster assignment of object $\boldsymbol{\phi}(\mathbf{x}_{dn})$.

The proposed model is based on an infinite mixture model, where the probability of descriptor mapped in a Hilbert space $\boldsymbol{\phi}(\mathbf{x}_{dn})$ is given by

Table 1: Notation.

Symbol in $\mathcal{I}$	Symbol in $\mathcal{H}$	Description
D		Number of shapes
N_d		Number of objects ($3D$-landmarks) in the dth shape
M_d	L_d	Dimensionality of the observed features in the dth shape
K	P	Dimensionality of a latent vector
J	Q	Number of correspondences (latent vectors) to which objects are assigned
$\mathbf{x}_{dn}$	$\boldsymbol{\phi}(\mathbf{x}_{dn})$	Observation of the nth object in the dth shape, $\mathbf{x}_{dn} \in \mathbb{R}^{M_d}$
$\mathbf{z}_j$	$\boldsymbol{\zeta}_j$	Latent vector for the jth correspondence, $\boldsymbol{\zeta}_j \in \mathbb{R}^K$
$\mathbf{W}_d$	$\boldsymbol{\mathcal{W}}_d$	Projection matrix for the dth domain, $\mathbf{W}_d \in \mathbb{R}^{M_d \times K}$
θ_j		Mixture weight for the jth cluster, $\theta_j \geq 0$, $\sum_{j=1}^{\infty} \theta_j = 1$

$$p\left(\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)|\mathbf{Z},\boldsymbol{\mathcal{W}},\boldsymbol{\theta}\right)=\sum_{j=1}^{\infty}\theta_j\mathcal{N}\left(\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)|\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_j,\alpha^{-1}\mathbf{I}\right), \tag{2}$$

where $\boldsymbol{\mathcal{W}}=\{\boldsymbol{\mathcal{W}}_d\}_{d=1}^{D}$ is a set of projections matrices, $\boldsymbol{\theta}=(\theta_j)_{j=1}^{\infty}$ are the mixture weights, θ_j represents the probability that the jth cluster is chosen, and $\mathcal{N}\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$ denotes a normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. One important contribution derived in [1], is that we can analyze multiples structures with different properties and dimensionalities, by employing projection matrices for each brain structure (domain-specific). Figure 1 shows the scheme of the proposed model, in which the relationship between shape descriptors, and latent vectors is described.

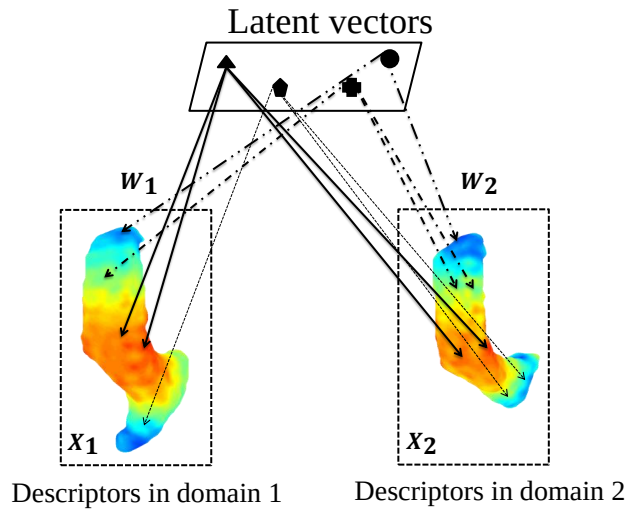

Figure 1: Scheme for the groupwise correspondence method. The figure shows an example of establishing clusters of correspondences in two domains (left ventrals).

In order to draw the cluster proportions, we use a stick-breaking process to generate mixture weights for a Dirichlet process with concentration parameter γ [1] (a,b and r are the hyperparameters). The joint probability of the data $\mathcal{X}$, and the cluster assignments $\mathbf{S}=\left\{\{s_{dn}\}_{n=1}^{N_d}\right\}_{d=1}^{D}$ are given by

$$p\left(\boldsymbol{\Phi},\mathbf{S}|\boldsymbol{\mathcal{W}},a,b,r,\gamma\right)=p\left(\mathbf{S}|\gamma\right)p\left(\boldsymbol{\Phi}|\mathbf{S},\boldsymbol{\mathcal{W}},a,b,r\right). \tag{3}$$

By marginalizing out the mixture weights $\boldsymbol{\theta}$, $p\left(\mathbf{S}|\gamma\right)$ becomes in

$$p\left(\mathbf{S}|\gamma\right)=\frac{\gamma^{J}\prod_{j=1}^{J}\left(N_{\cdot j}-1\right)!}{\gamma\left(\gamma+1\right)\ldots\left(\gamma+N-1\right)}, \tag{4}$$

where $N=\sum_{d=1}^{D}N_d$ is the total number of shape descriptors, $N_{\cdot j}$ represents the number of descriptors assigned to the cluster j, and J is the number of clusters that satisfies $N_{\cdot j}>0$.

1.1 The linear model (likelihood)

By marginalizing out latent vectors $\mathbf{Z}$ and the precision parameter α, the second factor of (**??**) is computed by

$$
\begin{aligned}
p\left(\mathbf{X}|\mathbf{S},\mathbf{W},a,b,r\right) &= \int\int \prod_{d=1}^{D}\prod_{n=1}^{N_d} \mathcal{N}\left(\mathbf{x}_{dn}|\mathbf{W}_d\mathbf{z}_{s_{dn}},\alpha^{-1}\mathbf{I}\right)\mathcal{G}\left(\alpha|a,b\right)\times\prod_{j=1}^{J}\mathcal{N}\left(\mathbf{z}_j|\mathbf{0},\left(\alpha r\right)^{-1}\mathbf{I}\right)d\mathbf{Z}d\alpha \\
&= \int\int \prod_{d=1}^{D}\prod_{n=1}^{N_d}\left(\frac{\alpha}{2\pi}\right)^{M_d/2}\exp\left(-\frac{\alpha}{2}||\mathbf{x}_{dn}-\mathbf{W}_d\mathbf{z}_{s_{dn}}||^2\right)\prod_{j=1}^{J}\left(\frac{\alpha r}{2\pi}\right)^{K/2} \\
&\times\exp\left(-\frac{\alpha r}{2}||\mathbf{z}_j||^2\right)\frac{b^a\alpha^{a-1}}{\Gamma\left(a\right)}\exp\left(-b\alpha\right)d\mathbf{Z}d\alpha \\
&= \frac{b^a}{\Gamma\left(a\right)}\int\int\left(\frac{\alpha}{2\pi}\right)^{\sum_d M_d N_d/2}\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}||\mathbf{x}_{dn}-\mathbf{W}_d\mathbf{z}_{s_{dn}}||^2\right)\left(\frac{\alpha r}{2\pi}\right)^{KJ/2} \\
&\times\exp\left(-\frac{\alpha r}{2}\sum_{j=1}^{J}||\mathbf{z}_j||^2\right)\exp\left(-b\alpha\right)\alpha^{a-1}d\mathbf{Z}d\alpha
\end{aligned}
\tag{5}
$$

Solving for the first exponential term

$$
\begin{aligned}
\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}||\mathbf{x}_{dn}-\mathbf{W}_d\mathbf{z}_{s_{dn}}||^2\right) = \exp\Big(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\big[&\mathbf{x}_{dn}^\top\mathbf{x}_{dn} \\
&-2\mathbf{z}_{s_{dn}}^\top\mathbf{W}_d^\top\mathbf{x}_{dn}+\mathbf{z}_{s_{dn}}^\top\mathbf{W}_d^\top\mathbf{W}_d\mathbf{z}_{s_{dn}}\big]\Big)
\end{aligned}
\tag{6}
$$

The equation in (5) becomes

$$
\begin{aligned}
p\left(\mathbf{X}|\mathbf{S},\mathbf{W},a,b,r\right) = \frac{b^a}{\Gamma\left(a\right)}\int\int\left(\frac{\alpha}{2\pi}\right)^{\sum_d M_d N_d/2}&\exp\Big(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\big[\mathbf{x}_{dn}^\top\mathbf{x}_{dn} \\
&-2\mathbf{z}_{s_{dn}}^\top\mathbf{W}_d^\top\mathbf{x}_{dn}+\mathbf{z}_{s_{dn}}^\top\mathbf{W}_d^\top\mathbf{W}_d\mathbf{z}_{s_{dn}}\big]\Big)\left(\frac{\alpha r}{2\pi}\right)^{KJ/2} \\
&\times\exp\left(-\frac{\alpha r}{2}\sum_{j=1}^{J}\mathbf{z}_j^\top\mathbf{z}_j\right)\exp\left(-b\alpha\right)\alpha^{a-1}d\mathbf{Z}d\alpha.
\end{aligned}
\tag{7}
$$

The exponential terms in (7) becomes

$$
\begin{aligned}
&\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\left[\mathbf{x}_{dn}^\top\mathbf{x}_{dn}-2\mathbf{z}_{s_{dn}}^\top\mathbf{W}_d^\top\mathbf{x}_{dn}+\mathbf{z}_{s_{dn}}^\top\mathbf{W}_d^\top\mathbf{W}_d\mathbf{z}_{s_{dn}}\right]-\frac{\alpha r}{2}\sum_{j=1}^{J}\mathbf{z}_j^\top\mathbf{z}_j-b\alpha\right) = \\
&\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\left[\mathbf{x}_{dn}^\top\mathbf{x}_{dn}-b\alpha\right)\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\left[-2\mathbf{z}_{s_{dn}}^\top\mathbf{W}_d^\top\mathbf{x}_{dn}+\mathbf{z}_{s_{dn}}^\top\mathbf{W}_d^\top\mathbf{W}_d\mathbf{z}_{s_{dn}}\right]\right. \\
&\left.-\frac{\alpha r}{2}\sum_{j=1}^{J}\mathbf{z}_j^\top\mathbf{z}_j\right)
\end{aligned}
\tag{8}
$$

By analyzing the nth objects that has the cluster assignment j ($n : s_{dn} = j$), the second factor in (8) becomes

$$\exp\Bigg(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\big[-2\mathbf{z}_{s_{dn}}^{\top}\mathbf{W}_d^{\top}\mathbf{x}_{dn}+\mathbf{z}_{s_{dn}}^{\top}\mathbf{W}_d^{\top}\mathbf{W}_d\mathbf{z}_{s_{dn}}\big]-\frac{\alpha r}{2}\sum_{j=1}^{J}\mathbf{z}_j^{\top}\mathbf{z}_j\Bigg)=\exp\Bigg(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n:s_{dn}\neq j}\big[-2\mathbf{z}_{s_{dn}}^{\top}\mathbf{W}_d^{\top}\mathbf{x}_{dn}$$

$$+\mathbf{z}_{s_{dn}}^{\top}\mathbf{W}_d^{\top}\mathbf{W}_d\mathbf{z}_{s_{dn}}\big]\Bigg)\times\underbrace{\exp\Bigg(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n:s_{dn}=j}\big[-2\mathbf{z}_{s_{dn}}^{\top}\mathbf{W}_d^{\top}\mathbf{x}_{dn}+\mathbf{z}_{s_{dn}}^{\top}\mathbf{W}_d^{\top}\mathbf{W}_d\mathbf{z}_{s_{dn}}\big]-\frac{\alpha r}{2}\sum_{j=1}^{J}\mathbf{z}_j^{\top}\mathbf{z}_j\Bigg)}_{C}$$

$$=\exp\Bigg(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{j=1}^{J}N_{dj}\big[-2\mathbf{z}_j^{\top}\mathbf{W}_d^{\top}\mathbf{x}_{dn}+\mathbf{z}_j^{\top}\mathbf{W}_d^{\top}\mathbf{W}_d\mathbf{z}_j\big]-\frac{\alpha r}{2}\sum_{j=1}^{J}\mathbf{z}_j^{\top}\mathbf{z}_j\Bigg)$$

$$=\exp\Bigg(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{j=1}^{J}\big[-2\mathbf{z}_j^{\top}\mathbf{W}_d^{\top}\sum_{n:s_{dn}=j}\mathbf{x}_{dn}+\mathbf{z}_j^{\top}N_{dj}\mathbf{W}_d^{\top}\mathbf{W}_d\mathbf{z}_j\big]-\frac{\alpha r}{2}\sum_{j=1}^{J}\mathbf{z}_j^{\top}\mathbf{z}_j\Bigg)$$

$$=\exp\Bigg(-\frac{\alpha}{2}\sum_{j=1}^{J}\big[-2\mathbf{z}_j^{\top}\sum_{d=1}^{D}\mathbf{W}_d^{\top}\sum_{n:s_{dn}=j}\mathbf{x}_{dn}+\mathbf{z}_j^{\top}\sum_{d=1}^{D}N_{dj}\mathbf{W}_d^{\top}\mathbf{W}_d\mathbf{z}_j\big]-\frac{\alpha r}{2}\sum_{j=1}^{J}\mathbf{z}_j^{\top}\mathbf{z}_j\Bigg)$$

$$=\exp\Bigg(-\frac{\alpha}{2}\sum_{j=1}^{J}\big[-2\mathbf{z}_j^{\top}\sum_{d=1}^{D}\mathbf{W}_d^{\top}\sum_{n:s_{dn}=j}\mathbf{x}_{dn}+\mathbf{z}_j^{\top}\sum_{d=1}^{D}N_{dj}\mathbf{W}_d^{\top}\mathbf{W}_d\mathbf{z}_j+r\mathbf{z}_j^{\top}\mathbf{z}_j\big]\Bigg)$$

$$=\exp\Bigg(-\frac{\alpha}{2}\sum_{j=1}^{J}\big[-2\mathbf{z}_j^{\top}\sum_{d=1}^{D}\mathbf{W}_d^{\top}\sum_{n:s_{dn}=j}\mathbf{x}_{dn}+\mathbf{z}_j^{\top}\left(\sum_{d=1}^{D}N_{dj}\mathbf{W}_d^{\top}\mathbf{W}_d+r\mathbf{I}\right)\mathbf{z}_j\big]\Bigg). \tag{9}$$

By using the quadratic property

$$-\frac{1}{2}\left(\mathbf{z}-\boldsymbol{\mu}\right)^{\top}\boldsymbol{C}^{-1}\left(\mathbf{z}-\boldsymbol{\mu}\right)=-\frac{1}{2}\left[\mathbf{z}^{\top}\boldsymbol{C}^{-1}\mathbf{z}-2\mathbf{z}^{\top}\boldsymbol{C}^{-1}\boldsymbol{\mu}+\boldsymbol{\mu}^{\top}\boldsymbol{C}^{-1}\boldsymbol{\mu}\right], \tag{10}$$

where

$$\mathbf{C}_j^{-1}=\sum_{d=1}^{D}N_{dj}\mathbf{W}_d^{\top}\mathbf{W}_d+r\mathbf{I}, \tag{11}$$

and

$$-2\mathbf{z}^{\top}\boldsymbol{C}^{-1}\boldsymbol{\mu}=-2\mathbf{z}_j^{\top}\sum_{d=1}^{D}\mathbf{W}_d^{\top}\sum_{n:s_{dn}=j}\mathbf{x}_{dn}$$

$$\boldsymbol{\mu}_j=\mathbf{C}_j\sum_{d=1}^{D}\mathbf{W}_d^{\top}\sum_{n:s_{dn}=j}\mathbf{x}_{dn}. \tag{12}$$

By completing the square as: $\arg=\arg+\frac{1}{2}\boldsymbol{\mu}^{\top}\boldsymbol{C}^{-1}\boldsymbol{\mu}-\frac{1}{2}\boldsymbol{\mu}^{\top}\boldsymbol{C}^{-1}\boldsymbol{\mu}$, the argument in (9) becomes

$$\exp\Bigg(-\frac{\alpha}{2}\sum_{j=1}^{J}\big[-2\mathbf{z}_j^{\top}\sum_{d=1}^{D}\mathbf{W}_d^{\top}\sum_{n:s_{dn}=j}\mathbf{x}_{dn}+\mathbf{z}_j^{\top}\left(\sum_{d=1}^{D}N_{dj}\mathbf{W}_d^{\top}\mathbf{W}_d+r\mathbf{I}\right)\mathbf{z}_j\big]\Bigg)$$

$$=\exp\Bigg(-\frac{\alpha}{2}\Big[\sum_{j=1}^{J}\left(\mathbf{z}_j-\boldsymbol{\mu}_j\right)^{\top}\mathbf{C}_j^{-1}\left(\mathbf{z}_j-\boldsymbol{\mu}_j\right)\Big]\Bigg)\exp\Bigg(-\frac{\alpha}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_j^{\top}\mathbf{C}_j^{-1}\boldsymbol{\mu}_j\Bigg) \tag{13}$$

Substituting (13) in (7) give us

$$p\left(\mathbf{X}|\mathbf{S},\mathbf{W},a,b,r\right)=\frac{b^a}{\Gamma\left(a\right)}\iint\left(\frac{\alpha}{2\pi}\right)^{\sum_d M_d N_d/2}\left(\frac{\alpha r}{2\pi}\right)^{KJ/2}\exp\Bigg(-\frac{\alpha}{2}\Big[\sum_{j=1}^{J}\left(\mathbf{z}_j-\boldsymbol{\mu}_j\right)^{\top}\mathbf{C}_j^{-1}\left(\mathbf{z}_j-\boldsymbol{\mu}_j\right)\Big]\Bigg)d\mathbf{Z}$$

$$\exp\Bigg(-\alpha\left[\frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\mathbf{x}_{dn}^{\top}\mathbf{x}_{dn}-\frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_j^{\top}\mathbf{C}_j^{-1}\boldsymbol{\mu}_j+b\right]\Bigg)\alpha^{a-1}d\alpha \tag{14}$$

In equation (14), factors related to $\mathbf{Z}$ are grouped together. We integrated out $\mathbf{Z}$ using

$$\int \exp\left(-\frac{1}{2}\left(\mathbf{z}_j-\boldsymbol{\mu}_j\right)^\top\left[\alpha^{-1}\mathbf{C}_j\right]^{-1}\left(\mathbf{z}_j-\boldsymbol{\mu}_j\right)\right)d\mathbf{z}_j=(2\pi)^{K/2}\left|\alpha^{-1}\mathbf{C}_j\right|^{1/2}=(2\pi)^{K/2}\,\alpha^{-K/2}|\mathbf{C}_j|^{1/2}, \tag{15}$$

which is the normalization constant of P-dimensional Gaussian distribution. Since we have the sum over the number of correspondences (latent vectors), K, the above equation ranges for all of these clusters. The equation (14), becomes

$$\begin{aligned} p\left(\mathbf{X}|\mathbf{S},\mathbf{W},a,b,r\right) &= \frac{b^a}{\Gamma\left(a\right)}\int\left(\frac{\alpha}{2\pi}\right)^{\sum_d M_d N_d/2}\left(\frac{\alpha r}{2\pi}\right)^{KJ/2}\prod_{j=1}^{J}\left[(2\pi)^{K/2}\,\alpha^{-K/2}|\mathbf{C}_j|^{1/2}\right] \\ &\exp\left(-\alpha\left[\frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\mathbf{x}_{dn}^\top\mathbf{x}_{dn}-\frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_j^\top\mathbf{C}_j^{-1}\boldsymbol{\mu}_j+b\right]\right)\alpha^{a-1}d\alpha \qquad (16) \\ &= \frac{b^a}{\Gamma\left(a\right)}\int\left(\frac{\alpha}{2\pi}\right)^{\sum_d M_d N_d/2}\left(\frac{\alpha r}{2\pi}\right)^{KJ/2}(2\pi)^{KJ/2}\,\alpha^{-KJ/2}\prod_{j=1}^{J}|\mathbf{C}_j|^{1/2} \\ &\exp\left(-\alpha\left[\frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\mathbf{x}_{dn}^\top\mathbf{x}_{dn}-\frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_j^\top\mathbf{C}_j^{-1}\boldsymbol{\mu}_j+b\right]\right)\alpha^{a-1}d\alpha \qquad (17) \end{aligned}$$

The α parameter is integrated out by using the following normalization constant of a Gamma distribution

$$\int \alpha^{a'-1}\exp\left(-b\alpha\right)d\alpha=\frac{\Gamma\left(a'\right)}{b'^{a'}}. \tag{18}$$

Finally the likelihood is given by

$$p\left(\mathbf{X}|\mathbf{S},\mathbf{W},a,b,r\right)=(2\pi)^{-\frac{\sum_d M_d N_d}{2}}\,r^{\frac{KJ}{2}}\,\frac{b^a}{b'^{a'}}\,\frac{\Gamma\left(a'\right)}{\Gamma\left(a\right)}\prod_{j=1}^{J}|\mathbf{C}_j|^{1/2}, \tag{19}$$

Here,

$$a'=a+\frac{\Sigma_d M_d N_d}{2},\qquad b'=b+\frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\mathbf{x}_{dn}^\top\mathbf{x}_{dn}-\frac{1}{2}\sum_{j=1}^{J}\boldsymbol{\mu}_j^\top\mathbf{C}_j^{-1}\boldsymbol{\mu}_j, \tag{20}$$

$$\boldsymbol{\mu}_j=\mathbf{C}_j\sum_{d=1}^{D}\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn},\qquad \mathbf{C}_j^{-1}=\sum_{d=1}^{D}N_{dj}\mathbf{W}_d^\top\mathbf{W}_d+r\mathbf{I}, \tag{21}$$

where N_{dj} is the number of descriptors assigned to cluster j in the shape d (domain).

1.2 Inference for the Linear Model

1.2.1 M-Step Linear

In the M-step, the projection matrices $\mathbf{W}$ are estimated by maximizing the logarithm of the joint likelihood (3). The gradient of the joint likelihood is computed by

$$\frac{\partial\log p\left(\mathbf{X},\mathbf{S}|\mathbf{W},a,b,r,\gamma\right)}{\partial\mathbf{W}_d}=\frac{\partial\log p\left(\mathbf{S}|\gamma\right)}{\partial\mathbf{W}_d}+\frac{\partial\log p\left(\mathbf{X}|\mathbf{S},\mathbf{W},a,b,r\right)}{\partial\mathbf{W}_d}, \tag{22}$$

Since the derivative of the first term in the above expression is zero, the expression becomes

$$\frac{\partial \log p\left(\mathbf{X},\mathbf{S}|\mathbf{W},a,b,r,\gamma\right)}{\partial \mathbf{W}_d} = \frac{\partial \log p\left(\mathbf{X}|\mathbf{S},\mathbf{W},a,b,r\right)}{\partial \mathbf{W}_d} = \frac{\partial \log \Big[(2\pi)^{-\frac{\sum_d M_d N_d}{2}} r^{\frac{JK}{2}} \frac{b^a}{b'^{a'}} \frac{\Gamma\left(a'\right)}{\Gamma(a)} \prod_{j=1}^{J} |\mathbf{C}_j|^{1/2} \Big]}{\partial \boldsymbol{\mathcal{W}}_d},$$

$$= \frac{\partial \log \left[\frac{cte}{b'^{a'}} \prod_{j=1}^{J} |\mathbf{C}_j|^{1/2} \right]}{\partial \mathbf{W}_d},$$

$$= -\frac{a'}{b'}\frac{\partial b'}{\partial \mathbf{W}_d} + \frac{1}{2}\sum_{j=1}^{J} \frac{\partial \log |\mathbf{C}_j|}{\partial \mathbf{W}_d} + 0,$$

$$= -\frac{a'}{b'}\frac{\partial b'}{\partial \mathbf{W}_d} + \frac{1}{2}\sum_{j=1}^{J} \mathrm{tr}\left(\mathbf{C}_j^{-1}\frac{\partial \mathbf{C}_j}{\partial \mathbf{W}_d}\right) \tag{23}$$

where

$$\frac{\partial b'}{\partial \mathbf{W}_d} = \frac{\partial}{\partial \mathbf{W}_d}\Big[\frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d} \mathbf{x}_{dn}^\top \mathbf{x}_{dn} - \frac{1}{2}\sum_{j=1}^{J} \boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\boldsymbol{\mu}_j + b\Big], \tag{24}$$

Here the second factor of the argument is the only which depends on $\boldsymbol{\mathcal{W}}_d$.

$$\begin{aligned}
\frac{\partial b'}{\partial \mathbf{W}_d} &= \frac{\partial}{\partial \mathbf{W}_d}\Big[-\frac{1}{2}\sum_{j=1}^{J} \boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\boldsymbol{\mu}_j \Big] \\
&= -\frac{1}{2}\sum_{j=1}^{J} \frac{\partial}{\partial \mathbf{W}_d}\Big[\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\boldsymbol{\mu}_j\Big] = -\frac{1}{2}\sum_{j=1}^{J} \mathrm{tr}\left(\frac{\partial}{\partial \mathbf{W}_d}\Big[\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\boldsymbol{\mu}_j\Big]\right) \\
&= -\frac{1}{2}\sum_{j=1}^{J} \mathrm{tr}\left(\frac{\partial \boldsymbol{\mu}_j^\top}{\partial \mathbf{W}_d}\mathbf{C}_j^{-1}\boldsymbol{\mu}_j + \boldsymbol{\mu}_j^\top\left(\frac{\partial \mathbf{C}_j^{-1}}{\partial \mathbf{W}_d}\boldsymbol{\mu}_j + \mathbf{C}_j^{-1}\frac{\partial \boldsymbol{\mu}_j}{\partial \mathbf{W}_d}\right)\right) \\
&= -\frac{1}{2}\sum_{j=1}^{J} \mathrm{tr}\left(\partial\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\boldsymbol{\mu}_j\right) + \mathrm{tr}\left(\boldsymbol{\mu}_j^\top \partial\mathbf{C}_j^{-1}\boldsymbol{\mu}_j\right) + \mathrm{tr}\left(\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\partial\boldsymbol{\mu}_j\right),
\end{aligned} \tag{25}$$

by applying trace properties (transpose elements)

$$\begin{aligned}
\frac{\partial b'}{\partial \mathbf{W}_d} &= -\frac{1}{2}\sum_{j=1}^{J} \mathrm{tr}\left(\partial\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\boldsymbol{\mu}_j\right) + \mathrm{tr}\left(\boldsymbol{\mu}_j^\top \partial\mathbf{C}_j^{-1}\boldsymbol{\mu}_j\right) + \mathrm{tr}\left(\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\partial\boldsymbol{\mu}_j\right) \\
&= -\frac{1}{2}\sum_{j=1}^{J} \mathrm{tr}\left(\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\partial\boldsymbol{\mu}_j\right) + \mathrm{tr}\left(\boldsymbol{\mu}_j^\top \partial\mathbf{C}_j^{-1}\boldsymbol{\mu}_j\right) + \mathrm{tr}\left(\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\partial\boldsymbol{\mu}_j\right) \\
&= -\frac{1}{2}\sum_{j=1}^{J} \mathrm{tr}\left(\boldsymbol{\mu}_j^\top \partial\mathbf{C}_j^{-1}\boldsymbol{\mu}_j\right) + 2\,\mathrm{tr}\left(\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\partial\boldsymbol{\mu}_j\right) \\
&= \underbrace{-\sum_{j=1}^{J} \mathrm{tr}\left(\boldsymbol{\mu}_j^\top \mathbf{C}_j^{-1}\partial\boldsymbol{\mu}_j\right)}_{A} \underbrace{-\frac{1}{2}\sum_{j=1}^{J} \mathrm{tr}\left(\boldsymbol{\mu}_j^\top \partial\mathbf{C}_j^{-1}\boldsymbol{\mu}_j\right)}_{B}
\end{aligned} \tag{26}$$

First, for the B part we have:

$$
\begin{aligned}
-\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top \partial \mathbf{C}_j^{-1}\boldsymbol{\mu}_j\right) &= -\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top \partial\left[\sum_{d=1}^{D} N_{dj}\mathbf{W}_d^\top\mathbf{W}_d + r\mathbf{I}\right]\boldsymbol{\mu}_j\right)\\
&= -\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\left[N_{dj}\left(\partial\mathbf{W}_d^\top\mathbf{W}_d + \mathbf{W}_d^\top\partial\mathbf{W}_d\right)\right]\boldsymbol{\mu}_j\right)\\
&= -\frac{1}{2}\sum_{j=1}^{J}\left[N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\partial\mathbf{W}_d^\top\mathbf{W}_d\boldsymbol{\mu}_j\right) + N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\boldsymbol{\mu}_j\right)\right]\\
&= -\frac{1}{2}\sum_{j=1}^{J}\left[N_{dj}\operatorname{tr}\left(\left(\boldsymbol{\mu}_j^\top\partial\mathbf{W}_d^\top\mathbf{W}_d\boldsymbol{\mu}_j\right)^\top\right) + N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\right)\right]\\
&= -\frac{1}{2}\sum_{j=1}^{J}\left[N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\boldsymbol{\mu}_j\right) + N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\right)\right]\\
&= -\sum_{j=1}^{J} N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\right). \qquad (27)
\end{aligned}
$$

By using derivatives properties for trace forms as

$$
\frac{\partial\operatorname{tr}\left[F\left(\mathbf{X}\right)\right]}{\partial\mathbf{X}} = f\left(\mathbf{X}\right)^\top, \qquad (28)
$$

where $f\left(\cdot\right)$ is the scalar derivative of $F\left(\cdot\right)$, the equation (27) becomes

$$
-\frac{1}{2}\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\partial\mathbf{C}_j^{-1}\boldsymbol{\mu}_j\right) = -\sum_{j=1}^{J} N_{dj}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\right)^\top = -\sum_{j=1}^{J} N_{dj}\mathbf{W}_d\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top. \qquad (29)
$$

Besides, for the A part we have:

$$
-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\mathbf{C}_j^{-1}\partial\boldsymbol{\mu}_j\right) \rightarrow \frac{\partial\boldsymbol{\mu}_j}{\partial\mathbf{W}_d} = \frac{\partial}{\partial\mathbf{W}_d}\left[\mathbf{C}_j\sum_{d=1}^{D}\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\right]. \qquad (30)
$$

The derivative for $\boldsymbol{\mu}_j$ is given by

$$
\frac{\partial\boldsymbol{\mu}_j}{\partial\mathbf{W}_d} = \underbrace{\partial\mathbf{C}_j\left(\sum_{d=1}^{D}\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\right)}_{C} + \underbrace{\mathbf{C}_j\partial\left(\sum_{d=1}^{D}\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\right)}_{D} \qquad (31)
$$

For the C part, we have

$$
\begin{aligned}
\partial\mathbf{C}_j\left(\sum_{d=1}^{D}\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\right) &= \frac{\partial}{\partial\mathbf{W}_d}\left[\left(\sum_{d=1}^{D}N_{dj}\mathbf{W}_d^\top\mathbf{W}_d + r\mathbf{I}\right)^{-1}\right]\left(\sum_{d=1}^{D}\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\right)\\
&= -\mathbf{C}_j\frac{\partial}{\partial\mathbf{W}_d}\left[\sum_{d=1}^{D}N_{dj}\mathbf{W}_d^\top\mathbf{W}_d + r\mathbf{I}\right]\mathbf{C}_j\sum_{d=1}^{D}\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\\
&= -\mathbf{C}_j\left(N_{dj}\left[\partial\mathbf{W}_d^\top\mathbf{W}_d + \mathbf{W}_d^\top\partial\mathbf{W}_d\right]\right)\boldsymbol{\mu}_j. \qquad (32)
\end{aligned}
$$

The D part is computed as

$$
\mathbf{C}_j\partial\left(\sum_{d=1}^{D}\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\right) = \mathbf{C}_j\partial\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}. \qquad (33)
$$

Then the A part becomes,

$$
\begin{aligned}
-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\mathbf{C}_j^{-1}\partial\boldsymbol{\mu}_j\right) &= -\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\mathbf{C}_j^{-1}\left[-\mathbf{C}_j\left(N_{dj}\left[\partial\mathbf{W}_d^\top\mathbf{W}_d+\mathbf{W}_d^\top\partial\mathbf{W}_d\right]\right)\boldsymbol{\mu}_j+\mathbf{C}_j\partial\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\right]\right)\\
&=\sum_{j=1}^{J}N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\mathbf{C}_j^{-1}\mathbf{C}_j\left(\partial\mathbf{W}_d^\top\mathbf{W}_d\right)\boldsymbol{\mu}_j\right)+\sum_{j=1}^{J}N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\mathbf{C}_j^{-1}\mathbf{C}_j\left(\mathbf{W}_d^\top\partial\mathbf{W}_d\right)\boldsymbol{\mu}_j\right)\\
&\quad-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\mathbf{C}_j^{-1}\mathbf{C}_j\partial\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\right)\\
&=\sum_{j=1}^{J}N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\partial\mathbf{W}_d^\top\mathbf{W}_d\boldsymbol{\mu}_j\right)+\sum_{j=1}^{J}N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\boldsymbol{\mu}_j\right)\\
&\quad-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\partial\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\right)\\
&=\sum_{j=1}^{J}N_{dj}\operatorname{tr}\left(\left(\boldsymbol{\mu}_j^\top\partial\mathbf{W}_d^\top\mathbf{W}_d\boldsymbol{\mu}_j\right)^\top\right)+\sum_{j=1}^{J}N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\right)\\
&\quad-\sum_{j=1}^{J}\operatorname{tr}\left(\left(\boldsymbol{\mu}_j^\top\partial\mathbf{W}_d^\top\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\right)^\top\right)\\
&=\sum_{j=1}^{J}N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\boldsymbol{\mu}_j\right)+\sum_{j=1}^{J}N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\right)\\
&\quad-\sum_{j=1}^{J}\operatorname{tr}\left(\sum_{n:s_{dn}=j}\mathbf{x}_{dn}^\top\partial\mathbf{W}_d\boldsymbol{\mu}_j\right)\\
&=\sum_{j=1}^{J}N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\right)+\sum_{j=1}^{J}N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\right)\\
&\quad-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j\sum_{n:s_{dn}=j}\mathbf{x}_{dn}^\top\partial\mathbf{W}_d\right)\\
&=2\sum_{j=1}^{J}N_{dj}\operatorname{tr}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\partial\mathbf{W}_d\right)-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j\sum_{n:s_{dn}=j}\mathbf{x}_{dn}^\top\partial\mathbf{W}_d\right). \qquad (34)
\end{aligned}
$$

By using the derivatives properties for trace forms described above, we have

$$
\begin{aligned}
-\sum_{j=1}^{J}\operatorname{tr}\left(\boldsymbol{\mu}_j^\top\mathbf{C}_j^{-1}\partial\boldsymbol{\mu}_j\right) &= 2\sum_{j=1}^{J}N_{dj}\left(\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\mathbf{W}_d^\top\right)^\top-\sum_{j=1}^{J}\left(\boldsymbol{\mu}_j\sum_{n:s_{dn}=j}\mathbf{x}_{dn}^\top\right)^\top\\
&=2\sum_{j=1}^{J}N_{dj}\mathbf{W}_d\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top-\sum_{j=1}^{J}\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\boldsymbol{\mu}_j^\top. \qquad (35)
\end{aligned}
$$

Finally

$$\begin{aligned}\frac{\partial b'}{\partial \mathbf{W}_d} &= 2\sum_{j=1}^{J} N_{dj}\mathbf{W}_d\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top - \sum_{j=1}^{J}\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\boldsymbol{\mu}_j^\top - -\sum_{j=1}^{J} N_{dj}\mathbf{W}_d\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top \\ &= \sum_{j=1}^{J} N_{dj}\mathbf{W}_d\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top - \sum_{j=1}^{J}\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\boldsymbol{\mu}_j^\top \\ &= \sum_{j=1}^{J}\left\{ N_{dj}\mathbf{W}_d\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top - \sum_{n:s_{dn}=j}\mathbf{x}_{dn}\boldsymbol{\mu}_j^\top \right\}. \end{aligned} \tag{36}$$

For the part B,

$$\begin{aligned}\frac{1}{2}\sum_{j=1}^{J}\mathrm{Tr}\left(\mathbf{C}_j^{-1}\frac{\partial \mathbf{C}_j}{\partial \mathbf{W}_d}\right) &= \frac{1}{2}\sum_{j=1}^{J}\mathrm{tr}\left(\mathbf{C}_j^{-1}\left(-\mathbf{C}_j\left(N_{dj}\left[\partial\mathbf{W}_d^\top\mathbf{W}_d + \mathbf{W}_d^\top\partial\mathbf{W}_d\right]\right)\mathbf{C}_j\right)\right) \\ &= -\frac{1}{2}\sum_{j=1}^{J} N_{dj}\,\mathrm{tr}\left(\mathbf{C}_j^{-1}\mathbf{C}_j\left[\partial\mathbf{W}_d^\top\mathbf{W}_d + \mathbf{W}_d^\top\partial\mathbf{W}_d\right]\mathbf{C}_j\right) \\ &= -\frac{1}{2}\sum_{j=1}^{J} N_{dj}\,\mathrm{tr}\left(\mathbf{C}_j^{-1}\mathbf{C}_j\left[\partial\mathbf{W}_d^\top\mathbf{W}_d + \mathbf{W}_d^\top\partial\mathbf{W}_d\right]\mathbf{C}_j\right) \\ &= -\frac{1}{2}\sum_{j=1}^{J} N_{dj}\,\mathrm{tr}\left(\partial\mathbf{W}_d^\top\mathbf{W}_d\mathbf{C}_j\right) - \frac{1}{2}\sum_{j=1}^{J} N_{dj}\,\mathrm{tr}\left(\mathbf{W}_d^\top\partial\mathbf{W}_d\mathbf{C}_j\right) \\ &= -\frac{1}{2}\sum_{j=1}^{J} N_{dj}\,\mathrm{tr}\left(\left(\partial\mathbf{W}_d^\top\mathbf{W}_d\mathbf{C}_j\right)^\top\right) - \frac{1}{2}\sum_{j=1}^{J} N_{dj}\,\mathrm{tr}\left(\mathbf{C}_j\mathbf{W}_d^\top\partial\mathbf{W}_d\right) \\ &= -\frac{1}{2}\sum_{j=1}^{J} N_{dj}\,\mathrm{tr}\left(\mathbf{C}_j\mathbf{W}_d^\top\partial\mathbf{W}_d\right) - \frac{1}{2}\sum_{j=1}^{J} N_{dj}\,\mathrm{tr}\left(\mathbf{C}_j\mathbf{W}_d^\top\partial\mathbf{W}_d\right) \\ &= -\sum_{j=1}^{J} N_{dj}\,\mathrm{tr}\left(\mathbf{C}_j\mathbf{W}_d^\top\partial\mathbf{W}_d\right) = -\sum_{j=1}^{J} N_{dj}\mathbf{W}_d\mathbf{C}_j. \end{aligned} \tag{37}$$

Finally the derivative of the log-likelihood is computed as

$$\begin{aligned}\frac{\partial \log p\left(\mathbf{X},\mathbf{S}|\mathbf{W},a,b,r,\gamma\right)}{\partial \mathbf{W}_d} &= -\frac{a'}{b'}\frac{\partial b'}{\partial \mathbf{W}_d} + \frac{1}{2}\sum_{j=1}^{J}\mathrm{tr}\left(\mathbf{C}_j^{-1}\frac{\partial \mathbf{C}_j}{\partial \mathbf{W}_d}\right) \\ &= -\frac{a'}{b'}\left[\sum_{j=1}^{J}\left\{ N_{dj}\mathbf{W}_d\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top - \sum_{n:s_{dn}=j}\mathbf{x}_{dn}\boldsymbol{\mu}_j^\top \right\}\right] - \sum_{j=1}^{J} N_{dj}\mathbf{W}_d\mathbf{C}_j. \end{aligned} \tag{38}$$

We can obtain the projection matrices that maximize the joint likelihood analytically as follows,

$$\mathbf{W}_d = -\frac{a'}{b'}\left(\sum_{j=1}^{J}\sum_{n:s_{dn}=j}\mathbf{x}_{dn}\boldsymbol{\mu}_j^\top\right)\left(\sum_{j=1}^{J} N_{dj}\mathbf{C}_j + \frac{a'}{b'}N_{dj}\boldsymbol{\mu}_j\boldsymbol{\mu}_j^\top\right)^{-1}. \tag{39}$$

1.3 Likelihood for the non-linear model

For the non-linear model, we give the derivation of the likelihood in (2), in which latent vectors $\mathbf{Z}$ and precision parameter α are analytically integrated out. To this end, we use two mappings functions $\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)$ and $\boldsymbol{\varphi}\left(\mathbf{w}_j^d\right)$, in order to represent our observations in the Hilbert space. First let us define the following expressions:

Let us define the projection matrix $\boldsymbol{\mathcal{W}}_d$ in $\mathcal{H}$ as

$$\boldsymbol{\mathcal{W}}_d = \begin{pmatrix} \varphi_1\left(\mathbf{w}_1^d\right) & \varphi_1\left(\mathbf{w}_2^d\right) & \cdots & \varphi_1\left(\mathbf{w}_P^d\right) \\ \varphi_2\left(\mathbf{w}_1^d\right) & \varphi_2\left(\mathbf{w}_2^d\right) & \cdots & \varphi_2\left(\mathbf{w}_P^d\right) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{L_d}\left(\mathbf{w}_1^d\right) & \varphi_{L_d}\left(\mathbf{w}_2^d\right) & \cdots & \varphi_{L_d}\left(\mathbf{w}_P^d\right) \end{pmatrix} \tag{40}$$

$$\begin{aligned} p\left(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right) &= \int\int \prod_{d=1}^{D}\prod_{n=1}^{N_d} \mathcal{N}\left(\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)|\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}, \alpha^{-1}\mathbf{I}\right)\mathcal{G}\left(\alpha|a,b\right) \times \prod_{j=1}^{Q}\mathcal{N}\left(\boldsymbol{\zeta}_j|\mathbf{0}, \left(\alpha r\right)^{-1}\mathbf{I}\right) d\mathbf{Z}d\alpha \\ &= \int\int \prod_{d=1}^{D}\prod_{n=1}^{N_d}\left(\frac{\alpha}{2\pi}\right)^{L_d/2}\exp\left(-\frac{\alpha}{2}||\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) - \boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}||^2\right)\prod_{j=1}^{Q}\left(\frac{\alpha r}{2\pi}\right)^{P/2} \\ &\times \exp\left(-\frac{\alpha r}{2}||\boldsymbol{\zeta}_j||^2\right)\frac{b^a\alpha^{a-1}}{\Gamma\left(a\right)}\exp\left(-b\alpha\right)d\mathbf{Z}d\alpha \\ &= \frac{b^a}{\Gamma\left(a\right)}\int\int\left(\frac{\alpha}{2\pi}\right)^{\sum_d L_d N_d/2}\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}||\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) - \boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}||^2\right)\left(\frac{\alpha r}{2\pi}\right)^{PQ/2} \\ &\times \exp\left(-\frac{\alpha r}{2}\sum_{j=1}^{Q}||\boldsymbol{\zeta}_j||^2\right)\exp\left(-b\alpha\right)\alpha^{a-1}d\mathbf{Z}d\alpha \end{aligned} \tag{41}$$

Solving for the first exponential term

$$\begin{aligned} \exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}||\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) - \boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}||^2\right) = \exp\Big(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\big[\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)^\top\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) \\ - 2\boldsymbol{\zeta}_{s_{dn}}^\top\boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) + \boldsymbol{\zeta}_{s_{dn}}^\top\boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}\big]\Big) \end{aligned} \tag{42}$$

The equation in (41) becomes

$$\begin{aligned} p\left(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right) = \frac{b^a}{\Gamma\left(a\right)}\int\int\left(\frac{\alpha}{2\pi}\right)^{\sum_d L_d N_d/2}\exp\Big(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\big[\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)^\top\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) \\ - 2\boldsymbol{\zeta}_{s_{dn}}^\top\boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) + \boldsymbol{\zeta}_{s_{dn}}^\top\boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}\big]\Big)\left(\frac{\alpha r}{2\pi}\right)^{PQ/2} \\ \times \exp\left(-\frac{\alpha r}{2}\sum_{j=1}^{Q}\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j\right)\exp\left(-b\alpha\right)\alpha^{a-1}d\mathbf{Z}d\alpha. \end{aligned} \tag{43}$$

The exponential terms in (7) becomes

$$\begin{aligned} &\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\left[\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)^\top\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) - 2\boldsymbol{\zeta}_{s_{dn}}^\top\boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) + \boldsymbol{\zeta}_{s_{dn}}^\top\boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}\right] - \frac{\alpha r}{2}\sum_{j=1}^{Q}\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j - b\alpha\right) = \\ &\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\left[\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)^\top\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) - b\alpha\right)\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\left[-2\boldsymbol{\zeta}_{s_{dn}}^\top\boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) + \boldsymbol{\zeta}_{s_{dn}}^\top\boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}\right]\right. \\ &\left.-\frac{\alpha r}{2}\sum_{j=1}^{Q}\boldsymbol{\zeta}_j^\top\boldsymbol{\zeta}_j\right) \end{aligned} \tag{44}$$

By analyzing the nth objects that has the cluster assignment j ($n : s_{dn} = j$), the second factor in (44) becomes

$$
\begin{aligned}
&\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\big[-2\boldsymbol{\zeta}_{s_{dn}}^{\top}\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)+\boldsymbol{\zeta}_{s_{dn}}^{\top}\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}\big]-\frac{\alpha r}{2}\sum_{j=1}^{Q}\boldsymbol{\zeta}_j^{\top}\boldsymbol{\zeta}_j\right)\\
&=\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n:s_{dn}\neq j}\big[-2\boldsymbol{\zeta}_{s_{dn}}^{\top}\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)+\boldsymbol{\zeta}_{s_{dn}}^{\top}\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}\big]\right)\\
&\times\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{n:s_{dn}=j}\big[-2\boldsymbol{\zeta}_{s_{dn}}^{\top}\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)+\boldsymbol{\zeta}_{s_{dn}}^{\top}\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}\big]-\frac{\alpha r}{2}\sum_{j=1}^{Q}\boldsymbol{\zeta}_j^{\top}\boldsymbol{\zeta}_j\right)\\
&=\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{j=1}^{Q}N_{dj}\big[-2\boldsymbol{\zeta}_j^{\top}\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)+\boldsymbol{\zeta}_j^{\top}\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_j\big]-\frac{\alpha r}{2}\sum_{j=1}^{J}\boldsymbol{\zeta}_j^{\top}\boldsymbol{\zeta}_j\right)\\
&=\exp\left(-\frac{\alpha}{2}\sum_{d=1}^{D}\sum_{j=1}^{Q}\big[-2\boldsymbol{\zeta}_j^{\top}\boldsymbol{\mathcal{W}}_d^{\top}\sum_{n:s_{dn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)+\boldsymbol{\zeta}_j^{\top}N_{dj}\mathbf{W}_d^{\top}\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_j\big]-\frac{\alpha r}{2}\sum_{j=1}^{Q}\boldsymbol{\zeta}_j^{\top}\boldsymbol{\zeta}_j\right)\\
&=\exp\left(-\frac{\alpha}{2}\sum_{j=1}^{Q}\big[-2\boldsymbol{\zeta}_j^{\top}\sum_{d=1}^{D}\boldsymbol{\mathcal{W}}_d^{\top}\sum_{n:s_{dn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)+\boldsymbol{\zeta}_j^{\top}\sum_{d=1}^{D}N_{dj}\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_j\big]-\frac{\alpha r}{2}\sum_{j=1}^{Q}\boldsymbol{\zeta}_j^{\top}\boldsymbol{\zeta}_j\right)\\
&=\exp\left(-\frac{\alpha}{2}\sum_{j=1}^{Q}\big[-2\boldsymbol{\zeta}_j^{\top}\sum_{d=1}^{D}\boldsymbol{\mathcal{W}}_d^{\top}\sum_{n:s_{dn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)+\boldsymbol{\zeta}_j^{\top}\sum_{d=1}^{D}N_{dj}\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_j+r\boldsymbol{\zeta}_j^{\top}\boldsymbol{\zeta}_j\big]\right)\\
&=\exp\left(-\frac{\alpha}{2}\sum_{j=1}^{Q}\big[-2\boldsymbol{\zeta}_j^{\top}\sum_{d=1}^{D}\boldsymbol{\mathcal{W}}_d^{\top}\sum_{n:s_{dn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)+\boldsymbol{\zeta}_j^{\top}\left(\sum_{d=1}^{D}N_{dj}\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\mathcal{W}}_d+r\mathbf{I}\right)\boldsymbol{\zeta}_j\big]\right),
\end{aligned}
\tag{45}
$$

where N_{dj} is the number of objects assigned to cluster j in de domain d. By using the quadratic property

$$
-\frac{1}{2}\left(\mathbf{z}-\boldsymbol{\mu}\right)^{\top}\boldsymbol{\Lambda}^{-1}\left(\mathbf{z}-\boldsymbol{\mu}\right)=-\frac{1}{2}\left[\mathbf{z}^{\top}\boldsymbol{\Lambda}^{-1}\mathbf{z}-2\mathbf{z}^{\top}\boldsymbol{\Lambda}^{-1}\boldsymbol{\mu}+\boldsymbol{\mu}^{\top}\boldsymbol{\Lambda}^{-1}\boldsymbol{\mu}\right],
\tag{46}
$$

where

$$
\boldsymbol{\Lambda}_j^{-1}=\sum_{d=1}^{D}N_{dj}\underbrace{\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\mathcal{W}}_d}_{A}+r\mathbf{I},
\tag{47}
$$

and

$$
\begin{aligned}
-2\mathbf{z}^{\top}\boldsymbol{\Lambda}^{-1}\boldsymbol{\mu}&=-2\boldsymbol{\zeta}_j^{\top}\sum_{d=1}^{D}\boldsymbol{\mathcal{W}}_d^{\top}\sum_{n:s_{dn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)\\
\boldsymbol{\mu}_j&=\boldsymbol{\Lambda}_j\sum_{d=1}^{D}\boldsymbol{\mathcal{W}}_d^{\top}\sum_{n:s_{dn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)\\
\boldsymbol{\mu}_j&=\boldsymbol{\Lambda}_j\sum_{d=1}^{D}\sum_{n:s_{dn}=j}\underbrace{\boldsymbol{\mathcal{W}}_d^{\top}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)}_{B}.
\end{aligned}
\tag{48}
$$

We can rewrite the expression in A by placing a given kernel for $\boldsymbol{\mathcal{W}}_d$

$$\boldsymbol{\mathcal{W}}_d^\top \boldsymbol{\mathcal{W}}_d = \begin{bmatrix} \boldsymbol{\varphi}\left(\mathbf{w}_1^d\right)^\top \boldsymbol{\varphi}\left(\mathbf{w}_1^d\right) & \boldsymbol{\varphi}\left(\mathbf{w}_1^d\right)^\top \boldsymbol{\varphi}\left(\mathbf{w}_2^d\right) & \cdots & \boldsymbol{\varphi}\left(\mathbf{w}_1^d\right)^\top \boldsymbol{\varphi}\left(\mathbf{w}_P^d\right) \\ \boldsymbol{\varphi}\left(\mathbf{w}_2^d\right)^\top \boldsymbol{\varphi}\left(\mathbf{w}_1^d\right) & \boldsymbol{\varphi}\left(\mathbf{w}_2^d\right)^\top \boldsymbol{\varphi}\left(\mathbf{w}_2^d\right) & \cdots & \boldsymbol{\varphi}\left(\mathbf{w}_2^d\right)^\top \boldsymbol{\varphi}\left(\mathbf{w}_P^d\right) \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\varphi}\left(\mathbf{w}_P^d\right)^\top \boldsymbol{\varphi}\left(\mathbf{w}_1^d\right) & \boldsymbol{\varphi}\left(\mathbf{w}_P^d\right)^\top \boldsymbol{\varphi}\left(\mathbf{w}_2^d\right) & \cdots & \boldsymbol{\varphi}\left(\mathbf{w}_P^d\right)^\top \boldsymbol{\varphi}\left(\mathbf{w}_P^d\right) \end{bmatrix}$$
$$= \begin{bmatrix} k\left(\mathbf{w}_1^d, \mathbf{w}_1^d\right) & k\left(\mathbf{w}_1^d, \mathbf{w}_2^d\right) & \cdots & k\left(\mathbf{w}_1^d, \mathbf{w}_P^d\right) \\ k\left(\mathbf{w}_2^d, \mathbf{w}_1^d\right) & k\left(\mathbf{w}_2^d, \mathbf{w}_2^d\right) & \cdots & k\left(\mathbf{w}_2^d, \mathbf{w}_P^d\right) \\ \vdots & \vdots & \ddots & \vdots \\ k\left(\mathbf{w}_P^d, \mathbf{w}_1^d\right) & k\left(\mathbf{w}_P^d, \mathbf{w}_2^d\right) & \cdots & k\left(\mathbf{w}_P^d, \mathbf{w}_P^d\right) \end{bmatrix} = \hat{\mathbf{K}}_d. \tag{49}$$

We can rewrite the expression in B by placing a given kernel of the form

$$\boldsymbol{\mathcal{W}}_d^\top \boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) = \begin{bmatrix} \boldsymbol{\varphi}\left(\mathbf{w}_1^d\right)^\top \boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) \\ \boldsymbol{\varphi}\left(\mathbf{w}_2^d\right)^\top \boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) \\ \vdots \\ \boldsymbol{\varphi}\left(\mathbf{w}_P^d\right)^\top \boldsymbol{\phi}\left(\mathbf{x}_{dn}\right) \end{bmatrix} = \begin{bmatrix} k\left(\mathbf{w}_1^d, \mathbf{x}_{dn}\right) \\ k\left(\mathbf{w}_2^d, \mathbf{x}_{dn}\right) \\ \vdots \\ k\left(\mathbf{w}_P^d, \mathbf{x}_{dn}\right) \end{bmatrix} = \hat{\mathbf{k}}_d, \tag{50}$$

where $\hat{\mathbf{k}}_d$ represents the kernel evaluated in the $\mathbf{x}_{dn}$ objects that has the cluster assignment j.

By completing the square as: $\arg = \arg + \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{\Lambda}_{\boldsymbol{j}}^{-\mathbf{1}} \boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{\Lambda}_{\boldsymbol{j}}^{-\mathbf{1}} \boldsymbol{\mu}$, the argument in (9) becomes

$$\exp\left(-\frac{\alpha}{2} \sum_{j=1}^{Q} \big[-2\boldsymbol{\zeta}_j^\top \sum_{d=1}^{D} \sum_{n:s_{dn}=j} \hat{\mathbf{k}}_d + \boldsymbol{\zeta}_j^\top \left(\sum_{d=1}^{D} N_{dj} \hat{\mathbf{K}}_d + r\mathbf{I} \right) \boldsymbol{\zeta}_j \big] \right)$$
$$= \exp\left(-\frac{\alpha}{2} \Big[\sum_{j=1}^{Q} \left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right)^\top \boldsymbol{\Lambda}_j^{-1} \left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right) \Big] \right) \exp\left(-\frac{\alpha}{2} \sum_{j=1}^{Q} \boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1} \boldsymbol{\mu}_j \right) \tag{51}$$

Substituting (51) in (43) give us

$$p\left(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right) = \frac{b^a}{\Gamma\left(a\right)} \iint \left(\frac{\alpha}{2\pi}\right)^{\sum_d L_d N_d/2} \left(\frac{\alpha r}{2\pi}\right)^{PQ/2} \exp\left(-\frac{\alpha}{2} \Big[\sum_{j=1}^{Q} \left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right)^\top \boldsymbol{\Lambda}_j^{-1} \left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right) \Big] \right) d\mathbf{Z}$$
$$\exp\left(-\alpha \left[\frac{1}{2} \sum_{d=1}^{D} \sum_{n=1}^{N_d} k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right) - \frac{1}{2} \sum_{j=1}^{Q} \boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1} \boldsymbol{\mu}_j + b \right] \right) \alpha^{a-1} d\alpha \tag{52}$$

In equation (52), factors related to $\mathbf{Z}$ are grouped together. We integrated out $\mathbf{Z}$ using

$$\int \exp\left(-\frac{1}{2} \left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right)^\top \left[\alpha^{-1} \boldsymbol{\Lambda}_j\right]^{-1} \left(\boldsymbol{\zeta}_j - \boldsymbol{\mu}_j\right) \right) d\boldsymbol{\zeta}_j = \left(2\pi\right)^{P/2} \left|\alpha^{-1} \boldsymbol{\Lambda}_j\right|^{1/2} = \left(2\pi\right)^{P/2} \alpha^{-P/2} |\boldsymbol{\Lambda}_j|^{1/2}, \tag{53}$$

which is the normalization constant of P-dimensional Gaussian distribution. Since we have the sum over the number of correspondences (latent vectors), P, the above equation ranges for all of these clusters. The equation (52), becomes

$$p\left(\mathbf{\Phi}|\mathbf{S},\boldsymbol{\mathcal{W}},a,b,r\right)=\frac{b^{a}}{\Gamma\left(a\right)}\int\left(\frac{\alpha}{2\pi}\right)^{\sum_{d}L_{d}N_{d}/2}\left(\frac{\alpha r}{2\pi}\right)^{PQ/2}\prod_{j=1}^{Q}\left[\left(2\pi\right)^{P/2}\alpha^{-P/2}|\mathbf{\Lambda}_{j}|^{1/2}\right]$$
$$\exp\left(-\alpha\left[\frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_{d}}k\left(\mathbf{x}_{dn},\mathbf{x}_{dn}\right)-\frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_{j}^{\top}\mathbf{\Lambda}_{j}^{-1}\boldsymbol{\mu}_{j}+b\right]\right)\alpha^{a-1}d\alpha \tag{54}$$
$$=\frac{b^{a}}{\Gamma\left(a\right)}\int\left(\frac{\alpha}{2\pi}\right)^{\sum_{d}L_{d}N_{d}/2}\left(\frac{\alpha r}{2\pi}\right)^{PQ/2}\left(2\pi\right)^{PQ/2}\alpha^{-PQ/2}\prod_{j=1}^{Q}|\mathbf{\Lambda}_{j}|^{1/2}$$
$$\exp\left(-\alpha\left[\frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_{d}}k\left(\mathbf{x}_{dn},\mathbf{x}_{dn}\right)-\frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_{j}^{\top}\mathbf{\Lambda}_{j}^{-1}\boldsymbol{\mu}_{j}+b\right]\right)\alpha^{a-1}d\alpha \tag{55}$$

The α parameter is integrated out by using the following normalization constant of a Gamma distribution

$$\int\alpha^{a'-1}\exp\left(-b\alpha\right)d\alpha=\frac{\Gamma\left(a'\right)}{b'^{a'}}. \tag{56}$$

Finally the likelihood of the nonlinear model is given by

$$p\left(\mathbf{\Phi}|\mathbf{S},\boldsymbol{\mathcal{W}},a,b,r\right)=\left(2\pi\right)^{-\frac{\sum_{d}L_{d}N_{d}}{2}}r^{\frac{PQ}{2}}\frac{b^{a}}{b'^{a'}}\frac{\Gamma\left(a'\right)}{\Gamma\left(a\right)}\prod_{j=1}^{Q}|\mathbf{\Lambda}_{j}|^{1/2}, \tag{57}$$

Here,

$$a'=a+\frac{\Sigma_{d}L_{d}N_{d}}{2},\qquad b'=b+\frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_{d}}k\left(\mathbf{x}_{dn},\mathbf{x}_{dn}\right)-\frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_{j}^{\top}\mathbf{\Lambda}_{j}^{-1}\boldsymbol{\mu}_{j}, \tag{58}$$

and

$$\boldsymbol{\mu}_{j}=\mathbf{\Lambda}_{j}\sum_{d=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_{d},\qquad\mathbf{\Lambda}_{j}^{-1}=\sum_{d=1}^{D}N_{dj}\hat{\mathbf{K}}_{d}+r\mathbf{I}. \tag{59}$$

1.4 Posterior

The posterior for the precision parameter α is given by

$$p\left(\alpha|\mathbf{\Phi},\mathbf{S},\boldsymbol{\mathcal{W}},a,b\right)=\mathcal{G}\left(a',b'\right), \tag{60}$$

and the posterior for the latent vector $\boldsymbol{\zeta}_{j}$ is given by

$$p\left(\boldsymbol{\zeta}_{j}|\alpha,\mathbf{\Phi},\mathbf{S},\boldsymbol{\mathcal{W}},r\right)=\mathcal{N}\left(\boldsymbol{\mu}_{j},\alpha^{-1}\mathbf{\Lambda}_{j}\right) \tag{61}$$

The derivation for these posteriors is given by

$$
\begin{aligned}
&p\left(\alpha|\mathbf{\Phi},\mathbf{S},\boldsymbol{\mathcal{W}},a,b\right)\prod_{j=1}^{Q}p\left(\boldsymbol{\zeta}_j|\alpha,\mathbf{\Phi},\mathbf{S},\boldsymbol{\mathcal{W}},r\right)\propto p\left(\mathbf{\Phi}|\alpha,\mathbf{Z},\mathbf{S},\boldsymbol{\mathcal{W}},a,b,r\right)p\left(\alpha|a,b\right)\prod_{j=1}^{Q}p\left(\boldsymbol{\zeta}_j|\alpha,r\right)\\
&=\prod_{d=1}^{D}\prod_{n=1}^{N_d}\mathcal{N}\left(\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)|\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}},\alpha^{-1}\mathbf{I}\right)\mathcal{G}\left(\alpha|a,b\right)\times\prod_{j=1}^{Q}\mathcal{N}\left(\boldsymbol{\zeta}_j|\mathbf{0},\left(\alpha r\right)^{-1}\mathbf{I}\right)\\
&=\prod_{d=1}^{D}\prod_{n=1}^{N_d}\left(\frac{\alpha}{2\pi}\right)^{L_d/2}\exp\left(-\frac{\alpha}{2}||\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)-\boldsymbol{\mathcal{W}}_d\boldsymbol{\zeta}_{s_{dn}}||^2\right)\prod_{j=1}^{Q}\left(\frac{\alpha r}{2\pi}\right)^{P/2}\\
&\times\exp\left(-\frac{\alpha r}{2}||\boldsymbol{\zeta}_j||^2\right)\frac{b^a\alpha^{a-1}}{\Gamma\left(a\right)}\exp\left(-b\alpha\right)\\
&\propto\alpha^{a'-1}\exp\left(-b'\alpha\right)\prod_{j=1}^{Q}|\mathbf{\Lambda}_j|^{-1/2}\exp\left(-\frac{\alpha}{2}\left(\boldsymbol{\zeta}_j-\boldsymbol{\mu}_j\right)^{\top}\mathbf{\Lambda}_j^{-1}\left(\boldsymbol{\zeta}_j-\boldsymbol{\mu}_j\right)\right)\\
&\propto\mathcal{G}\left(a',b'\right)\prod_{j=1}^{Q}\mathcal{N}\left(\boldsymbol{\mu}_j,\alpha^{-1}\mathbf{\Lambda}_j\right).
\end{aligned}
\tag{62}
$$

2 Inference for the non-linear model

2.1 M-step

In the M-step, the projection matrices $\boldsymbol{\mathcal{W}}$ are estimated by maximizing the logarithm of the joint likelihood (3). The gradient of the joint likelihood is computed by

$$
\frac{\partial\log p\left(\mathbf{\Phi},\mathbf{S}|\boldsymbol{\mathcal{W}},a,b,r,\gamma\right)}{\partial\boldsymbol{\mathcal{W}}_d}=\frac{\partial\log p\left(\mathbf{S}|\gamma\right)}{\partial\boldsymbol{\mathcal{W}}_d}+\frac{\partial\log p\left(\mathbf{\Phi}|\mathbf{S},\boldsymbol{\mathcal{W}},a,b,r\right)}{\partial\boldsymbol{\mathcal{W}}_d},
\tag{63}
$$

Since the derivative of the first term in the above expression is zero, the expression becomes

$$
\begin{aligned}
\frac{\partial\log p\left(\mathbf{\Phi},\mathbf{S}|\boldsymbol{\mathcal{W}},a,b,r,\gamma\right)}{\partial\boldsymbol{\mathcal{W}}_d}&=\frac{\partial\log p\left(\mathbf{\Phi}|\mathbf{S},\boldsymbol{\mathcal{W}},a,b,r\right)}{\partial\boldsymbol{\mathcal{W}}_d}=\frac{\partial\log\left[\left(2\pi\right)^{-\frac{\sum_d L_d N_d}{2}}r^{\frac{PQ}{2}}\frac{b^a}{b'^{a'}}\frac{\Gamma\left(a'\right)}{\Gamma(a)}\prod_{j=1}^{Q}|\mathbf{\Lambda}_j|^{1/2}\right]}{\partial\boldsymbol{\mathcal{W}}_d},\\
&=\frac{\partial\log\left[\frac{cte}{b'^{a'}}\prod_{j=1}^{Q}|\mathbf{\Lambda}_j|^{1/2}\right]}{\partial\boldsymbol{\mathcal{W}}_d},\\
&=-\frac{a'}{b'}\frac{\partial b'}{\partial\boldsymbol{\mathcal{W}}_d}+\frac{1}{2}\sum_{j=1}^{Q}\frac{\partial\log|\mathbf{\Lambda}_j|}{\partial\boldsymbol{\mathcal{W}}_d}+0,\\
&=-\frac{a'}{b'}\frac{\partial b'}{\partial\boldsymbol{\mathcal{W}}_d}+\frac{1}{2}\sum_{j=1}^{Q}\mathrm{Tr}\left(\mathbf{\Lambda}_j^{-1}\frac{\partial\mathbf{\Lambda}_j}{\partial\boldsymbol{\mathcal{W}}_d}\right)
\end{aligned}
\tag{64}
$$

where

$$
\frac{\partial b'}{\partial\boldsymbol{\mathcal{W}}_d}=\frac{\partial}{\partial\boldsymbol{\mathcal{W}}_d}\Big[\frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}k\left(\mathbf{x}_{dn},\mathbf{x}_{dn}\right)-\frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^{\top}\mathbf{\Lambda}_j^{-1}\boldsymbol{\mu}_j+b\Big],
\tag{65}
$$

Here the second factor of the argument is the only which depends on $\boldsymbol{\mathcal{W}}_d$.

$$\begin{aligned}
\frac{\partial b'}{\partial \boldsymbol{\mathcal{W}}_d} &= \frac{\partial}{\partial \boldsymbol{\mathcal{W}}_d}\Big[-\frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\Big] \\
&= -\frac{1}{2}\sum_{j=1}^{Q}\frac{\partial}{\partial \boldsymbol{\mathcal{W}}_d}\Big[\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\Big] \\
&= -\frac{1}{2}\sum_{j=1}^{Q}\frac{\partial}{\partial \boldsymbol{\mathcal{W}}_d}\Big[\left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d\right)^\top \boldsymbol{\Lambda}_j^{-1}\left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d\right)\Big] \\
&= -\frac{1}{2}\sum_{j=1}^{Q}\frac{\partial}{\partial \boldsymbol{\mathcal{W}}_d}\Big[\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d^\top \boldsymbol{\Lambda}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d\Big] \\
&= -\frac{1}{2}\sum_{j=1}^{Q}\frac{\partial}{\partial \boldsymbol{\mathcal{W}}_d}\Big[\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d^\top \boldsymbol{\Lambda}_j^\top \sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d\Big]
\end{aligned} \tag{66}$$

Since the derivative with respect to $\boldsymbol{\mathcal{W}}_d$ takes the *dth*-projection matrix, the sum over $e = 1, \ldots, d, \ldots, D$ only affect the *dth*-projection matrix. Besides, by using product derivatives properties, we must to compute the derivative of

$$\frac{\partial b'}{\partial \boldsymbol{\mathcal{W}}_d} = -\frac{1}{2}\Big[\sum_{n:s_{dn}=j}\Big(\partial\hat{\mathbf{k}}_d\Big)^\top \boldsymbol{\Lambda}_j^\top \sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d + \sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d^\top\Big[\Big(\partial\boldsymbol{\Lambda}_j^\top\Big)\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d + \boldsymbol{\Lambda}_j^\top\sum_{n:s_{dn}=j}\partial\hat{\mathbf{k}}_d\Big]\Big] \tag{67}$$

Since $\hat{\mathbf{k}}_d = \boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\phi}(\mathbf{x}_{dn})$ the above derivative $\partial\hat{\mathbf{k}}_d$ becomes,

$$\frac{\partial\hat{\mathbf{k}}_d}{\partial \boldsymbol{\mathcal{W}}_d} = \frac{\partial}{\partial \boldsymbol{\mathcal{W}}_d}\Big[\boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\phi}(\mathbf{x}_{dn})\Big] = \boldsymbol{\phi}(\mathbf{x}_{dn}). \tag{68}$$

For the $\boldsymbol{\Lambda}_j$, the derivative is computed as

$$\begin{aligned}
\frac{\partial\boldsymbol{\Lambda}_j}{\partial \boldsymbol{\mathcal{W}}_d} &= \frac{\partial}{\partial \boldsymbol{\mathcal{W}}_d}\left(\left[\sum_{e=1}^{D}N_{dj}\hat{\mathbf{K}}_d + r\mathbf{I}\right]^{-1}\right), \\
&= -\boldsymbol{\Lambda}_j\frac{\partial}{\partial \boldsymbol{\mathcal{W}}_d}\left(\sum_{e=1}^{D}N_{dj}\hat{\mathbf{K}}_d + r\mathbf{I}\right)\boldsymbol{\Lambda}_j, \\
&= -N_{dj}\boldsymbol{\Lambda}_j\frac{\partial\hat{\mathbf{K}}_d}{\partial \boldsymbol{\mathcal{W}}_d}\boldsymbol{\Lambda}_j,
\end{aligned} \tag{69}$$

where $\hat{\mathbf{K}}_d = \boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\mathcal{W}}_d$. The derivative $\frac{\partial\hat{\mathbf{K}}_d}{\partial \boldsymbol{\mathcal{W}}_d}$ is given by

$$\frac{\partial\hat{\mathbf{K}}_d}{\partial \boldsymbol{\mathcal{W}}_d} = \frac{\partial}{\partial \boldsymbol{\mathcal{W}}_d}\Big[\boldsymbol{\mathcal{W}}_d^\top\boldsymbol{\mathcal{W}}_d\Big] = 2\boldsymbol{\mathcal{W}}_d \tag{70}$$

By replacing (67) and (69) in (64), the gradient for $\boldsymbol{\mathcal{W}}_d$ becomes

$$
\begin{aligned}
\frac{\partial \log p\left(\boldsymbol{\Phi}, \mathbf{S} | \boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \boldsymbol{\mathcal{W}}_d} &= -\frac{a'}{b'}\frac{\partial b'}{\partial \boldsymbol{\mathcal{W}}_d} + \frac{1}{2}\sum_{j=1}^{Q} \operatorname{Tr}\left(\boldsymbol{\Lambda}_j^{-1}\frac{\partial \boldsymbol{\Lambda}_j}{\partial \boldsymbol{\mathcal{W}}_d}\right), \\
&= -\frac{a'}{b'}\frac{\partial b'}{\partial \boldsymbol{\mathcal{W}}_d} - \frac{1}{2}\sum_{j=1}^{Q} \operatorname{Tr}\left(\boldsymbol{\Lambda}_j^{-1} N_{dj} \boldsymbol{\Lambda}_j \frac{\partial \hat{\mathbf{K}}_d}{\partial \boldsymbol{\mathcal{W}}_d}\boldsymbol{\Lambda}_j\right), \\
&= -\frac{a'}{b'}\frac{\partial b'}{\partial \boldsymbol{\mathcal{W}}_d} - \frac{1}{2}\sum_{j=1}^{Q} N_{dj} \operatorname{Tr}\left(\frac{\partial \hat{\mathbf{K}}_d}{\partial \boldsymbol{\mathcal{W}}_d}\boldsymbol{\Lambda}_j\right), \\
&= \frac{1}{2}\frac{a'}{b'}\Bigg[\sum_{n:s_{dn}=j}\left(\partial \hat{\mathbf{k}}_d\right)^{\top}\boldsymbol{\Lambda}_j^{\top}\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d \\
&+ \sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d^{\top}\Big[\left(\partial \boldsymbol{\Lambda}_j^{\top}\right)\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d + \boldsymbol{\Lambda}_j^{\top}\sum_{n:s_{dn}=j}\partial \hat{\mathbf{k}}_d\Big]\Bigg] \\
&- \frac{1}{2}\sum_{j=1}^{Q} N_{dj} \operatorname{Tr}\left(\frac{\partial \hat{\mathbf{K}}_d}{\partial \boldsymbol{\mathcal{W}}_d}\boldsymbol{\Lambda}_j\right), \qquad (71) \\
&= \frac{1}{2}\frac{a'}{b'}\Bigg[\sum_{n:s_{dn}=j}\left(\partial \hat{\mathbf{k}}_d\right)^{\top}\boldsymbol{\mu}_j + \sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d^{\top} \\
&\times \Big[\left(\partial \boldsymbol{\Lambda}_j^{\top}\right)\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d + \boldsymbol{\Lambda}_j^{\top}\sum_{n:s_{dn}=j}\partial \hat{\mathbf{k}}_d\Big]\Bigg] \\
&- \frac{1}{2}\sum_{j=1}^{Q} N_{dj} \operatorname{Tr}\left(\frac{\partial \hat{\mathbf{K}}_d}{\partial \boldsymbol{\mathcal{W}}_d}\boldsymbol{\Lambda}_j\right), \\
&= \frac{1}{2}\frac{a'}{b'}\Bigg[\sum_{n:s_{dn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)^{\top}\boldsymbol{\mu}_j + \sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d^{\top} \\
&\times \Big[\left(-N_{dj}\boldsymbol{\Lambda}_j\frac{\partial \hat{\mathbf{K}}_d}{\partial \boldsymbol{\mathcal{W}}_d}\boldsymbol{\Lambda}_j\right)^{\top}\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d + \boldsymbol{\Lambda}_j^{\top}\sum_{n:s_{dn}=j}\partial \hat{\mathbf{k}}_d\Big]\Bigg] \\
&- \frac{1}{2}\sum_{j=1}^{Q} N_{dj} \operatorname{Tr}\left(\frac{\partial \hat{\mathbf{K}}_d}{\partial \boldsymbol{\mathcal{W}}_d}\boldsymbol{\Lambda}_j\right), \\
&= \frac{1}{2}\frac{a'}{b'}\Bigg[\sum_{n:s_{dn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)^{\top}\boldsymbol{\mu}_j - 2N_{dj}\boldsymbol{\mu}_j^{\top}\boldsymbol{\mathcal{W}}_d\boldsymbol{\mu}_j + \boldsymbol{\mu}_j^{\top}\sum_{n:s_{dn}=j}\boldsymbol{\phi}\left(\mathbf{x}_{dn}\right)\Bigg]\Bigg] \\
&- \frac{1}{2}\sum_{j=1}^{Q} 2N_{dj} \operatorname{Tr}\left(\boldsymbol{\mathcal{W}}_d\boldsymbol{\Lambda}_j\right), \qquad (72)
\end{aligned}
$$

2.2 Kernel derivatives

Since $\hat{\mathbf{k}}_d$ depends on $\boldsymbol{\mathcal{W}}_d$,

$$
\frac{\partial \hat{\mathbf{k}}_d}{\partial \boldsymbol{\mathcal{W}}_d} = \frac{\partial}{\partial \boldsymbol{\mathcal{W}}_d}\begin{bmatrix} k\left(\mathbf{w}_1^d, \mathbf{x}_{dn}\right) \\ k\left(\mathbf{w}_2^d, \mathbf{x}_{dn}\right) \\ \vdots \\ k\left(\mathbf{w}_P^d, \mathbf{x}_{dn}\right)\end{bmatrix} = \frac{\partial \hat{\mathbf{k}}_d}{\partial \boldsymbol{\Theta}} \qquad (73)
$$

where $\boldsymbol{\Theta} = \left\{l, \sigma^2, \{\mathbf{w}_i\}_{i=1}^{P}\right\}$ are the kernel hyperparameters of a squared exponential kernel of the form $k\left(x, x'\right) = \sigma^2 \exp\left(-\frac{\left(x-x'\right)^2}{2l^2}\right)$.

$$\frac{\partial \hat{\mathbf{k}}_d}{\partial \boldsymbol{\Theta}} = \begin{bmatrix} \frac{\partial k\left(\mathbf{w}_1^d, \mathbf{x}_{dn}\right)}{\partial \sigma^2} + \frac{\partial k\left(\mathbf{w}_2^d, \mathbf{x}_{dn}\right)}{\partial \sigma^2} + \cdots + \frac{\partial k\left(\mathbf{w}_P^d, \mathbf{x}_{dn}\right)}{\partial \sigma^2} \\ \frac{\partial k\left(\mathbf{w}_1^d, \mathbf{x}_{dn}\right)}{\partial l^2} + \frac{\partial k\left(\mathbf{w}_2^d, \mathbf{x}_{dn}\right)}{\partial l^2} + \cdots + \frac{\partial k\left(\mathbf{w}_P^d, \mathbf{x}_{dn}\right)}{\partial l^2} \\ \frac{\partial k\left(\mathbf{w}_1^d, \mathbf{x}_{dn}\right)}{\partial \mathbf{w}_1^d} + \frac{\partial k\left(\mathbf{w}_2^d, \mathbf{x}_{dn}\right)}{\partial \mathbf{w}_1^d} + \cdots + \frac{\partial k\left(\mathbf{w}_P^d, \mathbf{x}_{dn}\right)}{\partial \mathbf{w}_1^d} \\ \frac{\partial k\left(\mathbf{w}_1^d, \mathbf{x}_{dn}\right)}{\partial \mathbf{w}_P^d} + \frac{\partial k\left(\mathbf{w}_2^d, \mathbf{x}_{dn}\right)}{\partial \mathbf{w}_P^d} + \cdots + \frac{\partial k\left(\mathbf{w}_P^d, \mathbf{x}_{dn}\right)}{\partial \mathbf{w}_P^d} \end{bmatrix}^\top \tag{74}$$

Let us define the following derivatives with respect to the kernel hyperparameters in the equation (74), $k\left(\mathbf{w}_i^d, \mathbf{x}_{dn}\right) = \sigma^2 \exp\left(-\frac{\beta}{2}\left(\mathbf{w}_i^d - \mathbf{x}_{dn}\right)^\top \left(\mathbf{w}_i^d - \mathbf{x}_{dn}\right)\right)$, where β is the inverwidth $\beta = 1/l^2$

$$\frac{\partial k\left(\mathbf{w}_i^d, \mathbf{x}_{dn}\right)}{\partial \sigma^2} = \frac{1}{\sigma^2} k\left(\mathbf{w}_i^d, \mathbf{x}_{dn}\right), \tag{75}$$

$$\frac{\partial k\left(\mathbf{w}_1^d, \mathbf{x}_{dn}\right)}{\partial \beta} = -(1/2)\left(\mathbf{w}_i^d - \mathbf{x}_{dn}\right)^\top \left(\mathbf{w}_i^d - \mathbf{x}_{dn}\right) k\left(\mathbf{w}_i^d, \mathbf{x}_{dn}\right), \tag{76}$$

$$\frac{\partial k\left(\mathbf{w}_i^d, \mathbf{x}_{dn}\right)}{\partial \mathbf{w}_i^d} = -\beta\left(\mathbf{w}_i^d - \mathbf{x}_{dn}\right) k\left(\mathbf{w}_i^d, \mathbf{x}_{dn}\right) \tag{77}$$

and

$$\frac{\partial k\left(\mathbf{w}_i^d, \mathbf{x}_{dn}\right)}{\partial \mathbf{w}_j^d} = 0, \qquad \forall i \neq j. \tag{78}$$

Besides for the derivative of $\frac{\partial \hat{\mathbf{K}}_d}{\partial \boldsymbol{\mathcal{W}}_d}$, we have,

$$\frac{\partial \hat{\mathbf{K}}_d}{\partial \boldsymbol{\mathcal{W}}_d} = \begin{bmatrix} \frac{\partial k\left(\mathbf{w}_1^d, \mathbf{w}_1^d\right)}{\partial \boldsymbol{\Theta}} & \frac{\partial k\left(\mathbf{w}_1^d, \mathbf{w}_2^d\right)}{\partial \boldsymbol{\Theta}} & \cdots & \frac{\partial k\left(\mathbf{w}_1^d, \mathbf{w}_P^d\right)}{\partial \boldsymbol{\Theta}} \\ \frac{\partial k\left(\mathbf{w}_2^d, \mathbf{w}_1^d\right)}{\partial \boldsymbol{\Theta}} & \frac{\partial k\left(\mathbf{w}_2^d, \mathbf{w}_2^d\right)}{\partial \boldsymbol{\Theta}} & \cdots & \frac{\partial k\left(\mathbf{w}_2^d, \mathbf{w}_P^d\right)}{\partial \boldsymbol{\Theta}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial k\left(\mathbf{w}_P^d, \mathbf{w}_1^d\right)}{\partial \boldsymbol{\Theta}} & \frac{\partial k\left(\mathbf{w}_P^d, \mathbf{w}_2^d\right)}{\partial \boldsymbol{\Theta}} & \cdots & \frac{\partial k\left(\mathbf{w}_P^d, \mathbf{w}_P^d\right)}{\partial \boldsymbol{\Theta}} \end{bmatrix} \tag{79}$$

3 Derivatives for the optimization problem

In the M-step, the projection matrices $\boldsymbol{\mathcal{W}}$ are estimated by maximizing the logarithm of the joint likelihood (3). The gradient of the joint likelihood is computed by

$$\frac{\partial \log p\left(\boldsymbol{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \boldsymbol{\mathcal{W}}_d} = \frac{\partial \log p\left(\mathbf{S}|\gamma\right)}{\partial \boldsymbol{\mathcal{W}}_d} + \frac{\partial \log p\left(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right)}{\partial \boldsymbol{\mathcal{W}}_d}, \tag{80}$$

Since the derivative of the first term in the above expression is zero, the expression becomes

$$\begin{aligned} \frac{\partial \log p\left(\boldsymbol{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \boldsymbol{\mathcal{W}}_d} = \frac{\partial \log p\left(\boldsymbol{\Phi}|\mathbf{S}, \boldsymbol{\mathcal{W}}, a, b, r\right)}{\partial \boldsymbol{\mathcal{W}}_d} &= \frac{\partial \log\left[(2\pi)^{-\frac{\sum_d L_d N_d}{2}} r^{\frac{PQ}{2}} \frac{b^a}{b'^{a'}} \frac{\Gamma\left(a'\right)}{\Gamma(a)} \prod_{j=1}^Q |\boldsymbol{\Lambda}_j|^{1/2}\right]}{\partial \boldsymbol{\mathcal{W}}_d}, \\ &= \frac{\partial \log\left[\frac{cte}{b'^{a'}} \prod_{j=1}^Q |\boldsymbol{\Lambda}_j|^{1/2}\right]}{\partial \boldsymbol{\mathcal{W}}_d}, \\ &= -\frac{a'}{b'} \frac{\partial b'}{\partial \boldsymbol{\mathcal{W}}_d} + \frac{1}{2} \sum_{j=1}^Q \frac{\partial \log |\boldsymbol{\Lambda}_j|}{\partial \boldsymbol{\mathcal{W}}_d} + 0, \\ &= -\frac{a'}{b'} \frac{\partial b'}{\partial \boldsymbol{\mathcal{W}}_d} + \frac{1}{2} \sum_{j=1}^Q \mathrm{Tr}\left(\boldsymbol{\Lambda}_j^{-1} \frac{\partial \boldsymbol{\Lambda}_j}{\partial \boldsymbol{\mathcal{W}}_d}\right) \end{aligned} \tag{81}$$

We set three different kernels to analyze $k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right)$, $k\left(\mathbf{w}_i^d, \mathbf{x}_{dn}\right)$ and $k\left(\mathbf{w}_i^d, \mathbf{w}_j^d\right)$. Then, by using a rbf kernel for each one, the hyperparameters are set as:

$$\boldsymbol{\theta} = \left\{\beta_{xx}, \quad \sigma_{xx}^2, \quad , \beta_{xw}, \quad \sigma_{xw}^2, \quad \beta_{ww}, \quad \sigma_{ww}^2, \quad \{\mathbf{w}_i\}_{i=1}^P\right\} \tag{82}$$

First, for $k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right)$ we have the following derivatives,

$$\begin{aligned}
\frac{\partial \log p\left(\boldsymbol{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \beta_{xx}} &= -\frac{a'}{b'}\frac{\partial b'}{\partial \beta_{xx}} && (83)\\
&= -\frac{a'}{b'}\frac{\partial}{\partial \beta_{xx}}\left\{b + \frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d} k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right) - \frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right\} && (84)\\
&= -\frac{a'}{b'}\left\{\frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\frac{\partial}{\partial \beta_{xx}} k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right)\right\} && (85)\\
&= \frac{1}{4}\frac{a'}{b'}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\left(\mathbf{x}_{dn} - \mathbf{x}_{dn}\right)^\top\left(\mathbf{x}_{dn} - \mathbf{x}_{dn}\right) k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right) && (86)\\
&= 0, && (87)
\end{aligned}$$

for the kernel inversewidth β_{xx},

$$\begin{aligned}
\frac{\partial \log p\left(\boldsymbol{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \sigma_{xx}^2} &= -\frac{a'}{b'}\frac{\partial b'}{\partial \sigma_{xx}^2} && (88)\\
&= -\frac{a'}{b'}\frac{\partial}{\partial \sigma_{xx}^2}\left\{b + \frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d} k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right) - \frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right\} && (89)\\
&= -\frac{a'}{b'}\left\{\frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\frac{\partial}{\partial \sigma_{xx}^2} k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right)\right\} && (90)\\
&= -\frac{a'}{2b'}\sum_{d=1}^{D}\sum_{n=1}^{N_d}\frac{1}{\sigma_{xx}^2} k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right) && (91)
\end{aligned}$$

for the kernel variance σ_{xx}^2.

Besides the kernel derivatives for the $k\left(\mathbf{w}_i^d, \mathbf{x}_{dn}\right)$ are

$$\frac{\partial \log p\left(\mathbf{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \beta_{wx}} = -\frac{a'}{b'}\frac{\partial b'}{\partial \beta_{wx}}, \tag{92}$$

$$= -\frac{a'}{b'}\frac{\partial}{\partial \beta_{wx}}\left\{b + \frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d} k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right) - \frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right\}, \tag{93}$$

$$= -\frac{a'}{b'}\frac{\partial}{\partial \beta_{wx}}\left\{-\frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right\}, \tag{94}$$

$$= -\frac{a'}{b'}\frac{\partial}{\partial \beta_{wx}}\left\{\left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d\right)^\top \boldsymbol{\Lambda}_j^{-1}\left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d\right)\right\}, \tag{95}$$

$$= \frac{a'}{2b'}\left\{\left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\partial\hat{\mathbf{k}}_d\right)^\top \boldsymbol{\Lambda}_j^{-1}\left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d\right)\right. \tag{96}$$

$$\left. + \left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\hat{\mathbf{k}}_d\right)^\top \boldsymbol{\Lambda}_j^{-1}\left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\partial\hat{\mathbf{k}}_d\right)\right\}, \tag{97}$$

$$= \frac{a'}{2b'}\left\{\left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j} -\frac{1}{2}\left(\mathbf{w}_i^d - \mathbf{x}_{dn}\right)^\top\left(\mathbf{w}_i^d - \mathbf{x}_{dn}\right)\hat{\mathbf{k}}_d\right)^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right. \tag{98}$$

$$\left. + \boldsymbol{\mu}_j^\top\boldsymbol{\Lambda}_j^{-1}\left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j} -\frac{1}{2}\left(\mathbf{w}_i^d - \mathbf{x}_{dn}\right)^\top\left(\mathbf{w}_i^d - \mathbf{x}_{dn}\right)\hat{\mathbf{k}}_d\right)\right\}, \tag{99}$$

for the β_{wx} parameter,

$$\frac{\partial \log p\left(\mathbf{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \sigma^2_{wx}} = -\frac{a'}{b'}\frac{\partial b'}{\partial \sigma^2_{wx}}, \tag{100}$$

$$= -\frac{a'}{b'}\frac{\partial}{\partial \sigma^2_{wx}}\left\{b + \frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d} k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right) - \frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right\}, \tag{101}$$

$$= -\frac{a'}{b'}\frac{\partial}{\partial \sigma^2_{wx}}\left\{-\frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right\}, \tag{102}$$

$$= \frac{a'}{2b'}\left\{\left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\frac{1}{\sigma^2_{wx}}\hat{\mathbf{k}}_d\right)^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right. \tag{103}$$

$$\left. + \boldsymbol{\mu}_j^\top\boldsymbol{\Lambda}_j^{-1}\left(\boldsymbol{\Lambda}_j\sum_{e=1}^{D}\sum_{n:s_{dn}=j}\frac{1}{\sigma^2_{wx}}\hat{\mathbf{k}}_d\right)\right\}, \tag{104}$$

for the $\sigma 2_{wx}$ parameter,

$$\frac{\partial \log p\left(\mathbf{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \mathbf{w}_i^d} = -\frac{a'}{b'}\frac{\partial b'}{\partial \mathbf{w}_i^d}, \tag{105}$$

$$= -\frac{a'}{b'}\frac{\partial}{\partial \mathbf{w}_i^d}\left\{b + \frac{1}{2}\sum_{d=1}^{D}\sum_{n=1}^{N_d} k\left(\mathbf{x}_{dn}, \mathbf{x}_{dn}\right) - \frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right\}, \tag{106}$$

$$= -\frac{a'}{b'}\frac{\partial}{\partial \mathbf{w}_i^d}\left\{-\frac{1}{2}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right\}, \tag{107}$$

$$= \frac{a'}{2b'}\left\{\left(\boldsymbol{\Lambda}_j \sum_{n:s_{dn}=j} -\beta_{wx}\left(\mathbf{w}_i^d - \mathbf{x}_{dn}\right)\hat{\mathbf{k}}_d\right)^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j \tag{108}$$

$$+\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\left(\boldsymbol{\Lambda}_j \sum_{n:s_{dn}=j} -\beta_{wx}\left(\mathbf{w}_i^d - \mathbf{x}_{dn}\right)\hat{\mathbf{k}}_d\right)\right\}, \tag{109}$$

for the $\mathbf{w}_i^d$ parameter.

The last kernel is defined as $k\left(\mathbf{w}_i^d, \mathbf{w}_{i'}^d\right)$, with hyperparmeters σ_{ww}^2, β_{ww} and $\mathbf{w}_i^d$. These derivatives are computed as follows

$$\frac{\partial \log p\left(\mathbf{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \beta_{ww}} = \left[-\frac{a'}{b'}\frac{\partial b'}{\partial \beta_{ww}} + \frac{1}{2}\sum_{j=1}^{Q}\mathrm{Tr}\left(\boldsymbol{\Lambda}_j^{-1}\frac{\partial \boldsymbol{\Lambda}_j}{\partial \beta_{ww}}\right)\right], \tag{110}$$

$$= \frac{\partial}{\partial \beta_{ww}}\left\{\frac{a'}{2b'}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top \boldsymbol{\Lambda}_j^{-1}\boldsymbol{\mu}_j\right\} + \frac{1}{2}\sum_{j=1}^{Q}\mathrm{Tr}\left(\boldsymbol{\Lambda}_j^{-1}\frac{\partial \boldsymbol{\Lambda}_j}{\partial \beta_{ww}}\right), \tag{111}$$

$$= \frac{\partial}{\partial \beta_{ww}}\left\{\frac{a'}{2b'}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top\left(\sum_{d=1}^{D}N_{dj}\hat{\mathbf{K}}_d + r\mathbf{I}\right)\boldsymbol{\mu}_j\right\} \tag{112}$$

$$+\frac{1}{2}\sum_{j=1}^{Q}\mathrm{Tr}\left(\boldsymbol{\Lambda}_j^{-1}\frac{\partial}{\partial \beta_{ww}}\left(\sum_{d=1}^{D}N_{dj}\hat{\mathbf{K}}_d + r\mathbf{I}\right)^{-1}\right), \tag{113}$$

$$= \frac{\partial}{\partial \beta_{ww}}\left\{\frac{a'}{2b'}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top\left(\sum_{d=1}^{D}N_{dj}\hat{\mathbf{K}}_d + r\mathbf{I}\right)\boldsymbol{\mu}_j\right\} \tag{114}$$

$$+\frac{1}{2}\sum_{j=1}^{Q}\mathrm{Tr}\left(\boldsymbol{\Lambda}_j^{-1}\frac{\partial}{\partial \beta_{ww}}\left(\sum_{d=1}^{D}N_{dj}\hat{\mathbf{K}}_d + r\mathbf{I}\right)^{-1}\right), \tag{115}$$

$$= \left\{\frac{a'}{2b'}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top\frac{\partial}{\partial \beta_{ww}}\left(\sum_{d=1}^{D}N_{dj}\hat{\mathbf{K}}_d\right)\boldsymbol{\mu}_j\right\} \tag{116}$$

$$-\frac{1}{2}\sum_{j=1}^{Q}\mathrm{Tr}\left(\boldsymbol{\Lambda}_j^{-1}\boldsymbol{\Lambda}_j\frac{\partial}{\partial \beta_{ww}}\left(\sum_{d=1}^{D}N_{dj}\hat{\mathbf{K}}_d + r\mathbf{I}\right)\boldsymbol{\Lambda}_j\right), \tag{117}$$

$$= \left\{\frac{a'}{2b'}\sum_{j=1}^{Q}\boldsymbol{\mu}_j^\top\left(\sum_{d=1}^{D}N_{dj}\left(-(1/2)\left(\mathbf{P}_2\right)\hat{\mathbf{K}}_d\right)\right)\boldsymbol{\mu}_j\right\} \tag{118}$$

$$-\frac{1}{2}\sum_{j=1}^{Q}\mathrm{Tr}\left(\left(\sum_{d=1}^{D}N_{dj}\left(-(1/2)\left(\mathbf{P}_2\right)\right)\hat{\mathbf{K}}_d\right)\boldsymbol{\Lambda}_j\right), \tag{119}$$

where $\mathbf{P}_2$ is the pairwise distance of $\mathbf{W}_d$

$$\frac{\partial \log p\left(\mathbf{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \sigma_{ww}^2} = \left[-\frac{a'}{b'}\frac{\partial b'}{\partial \sigma_{ww}^2} + \frac{1}{2}\sum_{j=1}^{Q} \mathrm{Tr}\left(\mathbf{\Lambda}_j^{-1}\frac{\partial \mathbf{\Lambda}_j}{\partial \sigma_{ww}^2}\right)\right], \tag{120}$$

$$= \left\{ \frac{a'}{2b'}\sum_{j=1}^{Q} \boldsymbol{\mu}_j^{\top}\frac{\partial}{\partial \sigma_{ww}^2}\left(\sum_{d=1}^{D} N_{dj}\hat{\mathbf{K}}_d\right)\boldsymbol{\mu}_j \right\} \tag{121}$$

$$-\frac{1}{2}\sum_{j=1}^{Q}\mathrm{Tr}\left(\mathbf{\Lambda}_j^{-1}\mathbf{\Lambda}_j\frac{\partial}{\partial \sigma_{ww}^2}\left(\sum_{d=1}^{D} N_{dj}\hat{\mathbf{K}}_d + r\mathbf{I}\right)\mathbf{\Lambda}_j\right), \tag{122}$$

$$= \left\{ \frac{a'}{2b'}\sum_{j=1}^{Q} \boldsymbol{\mu}_j^{\top}\left(\sum_{d=1}^{D} N_{dj}\left(\frac{1}{\sigma_{ww}^2}\hat{\mathbf{K}}_d\right)\right)\boldsymbol{\mu}_j \right\} \tag{123}$$

$$-\frac{1}{2}\sum_{j=1}^{Q}\mathrm{Tr}\left(\left(\sum_{d=1}^{D} N_{dj}\frac{1}{\sigma_{ww}^2}\hat{\mathbf{K}}_d\right)\mathbf{\Lambda}_j\right). \tag{124}$$

Finally for the $\mathbf{w}_i^d$ parameter we have

$$\frac{\partial \log p\left(\mathbf{\Phi}, \mathbf{S}|\boldsymbol{\mathcal{W}}, a, b, r, \gamma\right)}{\partial \mathbf{w}_i^d} = \left[-\frac{a'}{b'}\frac{\partial b'}{\partial \mathbf{w}_i^d} + \frac{1}{2}\sum_{j=1}^{Q} \mathrm{Tr}\left(\mathbf{\Lambda}_j^{-1}\frac{\partial \mathbf{\Lambda}_j}{\partial \mathbf{w}_i^d}\right)\right], \tag{125}$$

$$= \left\{ \frac{a'}{2b'}\sum_{j=1}^{Q} \boldsymbol{\mu}_j^{\top}\frac{\partial}{\partial \mathbf{w}_i^d}\left(\sum_{d=1}^{D} N_{dj}\hat{\mathbf{K}}_d\right)\boldsymbol{\mu}_j \right\} \tag{126}$$

$$-\frac{1}{2}\sum_{j=1}^{Q}\mathrm{Tr}\left(\mathbf{\Lambda}_j^{-1}\mathbf{\Lambda}_j\frac{\partial}{\partial \mathbf{w}_i^d}\left(\sum_{d=1}^{D} N_{dj}\hat{\mathbf{K}}_d + r\mathbf{I}\right)\mathbf{\Lambda}_j\right), \tag{127}$$

$$= \left\{ \frac{a'}{2b'}\sum_{j=1}^{Q} \boldsymbol{\mu}_j^{\top}\left(\sum_{d=1}^{D} N_{dj}\left(-\beta_{ww}\mathbf{P}_1\hat{\mathbf{K}}_d\right)\right)\boldsymbol{\mu}_j \right\} \tag{128}$$

$$-\frac{1}{2}\sum_{j=1}^{Q}\mathrm{Tr}\left(\left(\sum_{d=1}^{D} N_{dj} - \beta_{ww}\mathbf{P}_1\hat{\mathbf{K}}_d\right)\mathbf{\Lambda}_j\right), \tag{129}$$

where $\mathbf{P}_1$ define mathematical notation

References

[1] Tomoharu Iwata, Tsutomu Hirao, and Naonori Ueda. Unsupervised cluster matching via probabilistic latent variable models, 2013.

[2] Tomoharu Iwata, Tsutomu Hirao, and Naonori Ueda. Probabilistic latent variable models for unsupervised many-to-many object matching. *Information Processing and Management*, 52(4):682 – 697, 2016.