Lecture Notes on SNR-Invariant PLDA

Man-Wai MAK

Aug. 2015

Abstract

This document provides the derivations of the equations in the paper: Na Li and M.W. Mak, "SNR-Invariant PLDA Modeling in Nonparametric Subspace for Robust Speaker Verification", *IEEE/ACM Trans. on Audio Speech and Language Processing*, vol. 23, no. 10, pp. 1648-1659, Oct. 2015.

Please cite this document as: M.W. Mak, Lecture Notes on SNR-Invariant PLDA, Technical Report and Lecture Note Series, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, April 2015.

1 SNR-Invariant PLDA

1.1 Generative Model

Denote $\mathbf{x}_{ij}^k$ as the j-th D-dimensional i-vector from speaker i, where the i-vector is obtained from an utterance with SNR falling into the k-th SNR group. Then, we have

$$\mathbf{x}_{ij}^k = \mathbf{m} + \mathbf{V}\mathbf{h}_i + \mathbf{U}\mathbf{w}_k + \boldsymbol{\epsilon}_{ij}^k \tag{1}$$

where $\mathbf{m}$ is a $D \times 1$ vector representing the global mean of i-vectors, $\mathbf{h}_i$ is a $P \times 1$ vector denoting the speaker factor with the prior distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{w}_k$ is a $Q \times 1$ vector denoting the latent SNR factor with prior distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, $\boldsymbol{\epsilon}_{ij}^k$ is a $D \times 1$ vector denoting the residual with distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, $\mathbf{V}$ is a $D \times P$ matrix whose columns span the speaker subspace, and $\mathbf{U}$ is a $D \times Q$ matrix whose columns span the SNR subspace.

1.2 EM Formulation

Denote all of the training i-vectors as $\mathcal{X} = \{\mathbf{x}_{ij}^k | i = 1, \ldots, S; j = 1, \ldots, H_i(k); k = 1, \ldots, K\}$, where S is the number of training speakers, $H_i(k)$ is the number of utterances from speaker i at the k-th SNR group, and K is the number of SNR groups. Eq. 1 can be written as:

$$\mathbf{x}_{ij}^k = \mathbf{m} + [\mathbf{V} \;\; \mathbf{U}] \begin{bmatrix} \mathbf{h}_i \\ \mathbf{w}_k \end{bmatrix} + \boldsymbol{\epsilon}_{ij}^k = \mathbf{m} + \mathbf{B}\hat{\mathbf{z}}_{ik} + \boldsymbol{\epsilon}_{ij}^k,$$

where $\mathbf{B} = [\mathbf{V} \;\; \mathbf{U}]$ and $\hat{\mathbf{z}}_{ik} = [\mathbf{h}_i^\mathsf{T} \;\; \mathbf{w}_k^\mathsf{T}]^\mathsf{T}$. The model parameters $\boldsymbol{\theta} = \{\mathbf{m}, \mathbf{V}, \mathbf{U}, \boldsymbol{\Sigma}\}$ are estimated by an EM algorithm, in which the loading matrices $\mathbf{V}$ and $\mathbf{U}$ can be estimated either separately or jointly.

1.2.1 Decoupled Estimation of Loading Matrices

In this approach, we focus on one latent factor at a time and marginalize over the other latent factors. For example, to estimate $\mathbf{V}$, we compute the posterior expectation of $\mathbf{h}_i$ by marginalizing over $\mathbf{w}_k$. Thus, the posterior density of $\mathbf{h}_i$ is written as:

$$\begin{aligned} p(\mathbf{h}_i | \mathbf{x}_{ij}^k, \boldsymbol{\theta}) &\propto p(\mathbf{x}_{ij}^k | \mathbf{h}_i, \boldsymbol{\theta}) p(\mathbf{h}_i) \\ &= \int p(\mathbf{x}_{ij}^k, \mathbf{w}_k | \mathbf{h}_i, \boldsymbol{\theta}) p(\mathbf{h}_i) \mathrm{d}\mathbf{w}_k \\ &= \int p(\mathbf{x}_{ij}^k | \mathbf{h}_i, \mathbf{w}_k, \boldsymbol{\theta}) p(\mathbf{w}_k) p(\mathbf{h}_i) \mathrm{d}\mathbf{w}_k \end{aligned}$$

$$
\begin{aligned}
&= \int \mathcal{N}(\mathbf{x}_{ij}^k|\mathbf{m}+\mathbf{V}\mathbf{h}_i+\mathbf{U}\mathbf{w}_k,\boldsymbol{\Sigma})\mathcal{N}(\mathbf{w}_k|\mathbf{0},\mathbf{I})\mathcal{N}(\mathbf{h}_i|\mathbf{0},\mathbf{I})\mathrm{d}\mathbf{w}_k \\
&= \mathcal{N}(\mathbf{x}_{ij}^k|\mathbf{m}+\mathbf{V}\mathbf{h}_i,\boldsymbol{\Phi})\mathcal{N}(\mathbf{h}_i|\mathbf{0},\mathbf{I}) \\
&\propto \exp\left\{\mathbf{h}_i^\mathsf{T}\mathbf{V}^\mathsf{T}\boldsymbol{\Phi}^{-1}(\mathbf{x}_{ij}^k-\mathbf{m})-\frac{1}{2}\mathbf{h}_i^\mathsf{T}(\mathbf{I}+\mathbf{V}^\mathsf{T}\boldsymbol{\Phi}^{-1}\mathbf{V})\mathbf{h}_i\right\} \qquad (2)
\end{aligned}
$$

where $\boldsymbol{\Phi}=\mathbf{U}\mathbf{U}^\mathsf{T}+\boldsymbol{\Sigma}$. Comparing this posterior density with a standard Gaussian, we have

$$
\begin{aligned}
\langle\mathbf{h}_i|\mathbf{x}_{ij}^k\rangle &= \left(\mathbf{I}+\mathbf{V}^\mathsf{T}\boldsymbol{\Phi}^{-1}\mathbf{V}\right)^{-1}\mathbf{V}^\mathsf{T}\boldsymbol{\Phi}^{-1}(\mathbf{x}_{ij}^k-\mathbf{m}) \\
\langle\mathbf{h}_i\mathbf{h}_i^\mathsf{T}|\mathbf{x}_{ij}^k\rangle &= \left(\mathbf{I}+\mathbf{V}^\mathsf{T}\boldsymbol{\Phi}^{-1}\mathbf{V}\right)^{-1}+\langle\mathbf{h}_i|\mathbf{x}_{ij}^k\rangle\langle\mathbf{h}_i|\mathbf{x}_{ij}^k\rangle^\mathsf{T}.
\end{aligned} \qquad (3)
$$

If all of the i-vectors of speaker i are given, we evaluate the joint posterior

$$
\begin{aligned}
p(\mathbf{h}_i|\mathbf{x}_{ij}^k\ \forall j \text{ and } k,\boldsymbol{\theta}) &\propto \prod_{k=1}^{K}\prod_{j=1}^{H_i(k)} p(\mathbf{x}_{ij}^k|\mathbf{h}_i,\boldsymbol{\theta})p(\mathbf{h}_i) \\
&\propto \exp\left\{\mathbf{h}_i^\mathsf{T}\mathbf{V}^\mathsf{T}\boldsymbol{\Phi}^{-1}\sum_{k=1}^{K}\sum_{j=1}^{H_i(k)}(\mathbf{x}_{ij}^k-\mathbf{m})-\frac{1}{2}\mathbf{h}_i^\mathsf{T}\left(\mathbf{I}+\sum_{k=1}^{K}H_i(k)\mathbf{V}^\mathsf{T}\boldsymbol{\Phi}^{-1}\mathbf{V}\right)\mathbf{h}_i\right\}
\end{aligned} \qquad (4)
$$

Then, the posterior expectations become:

$$
\begin{aligned}
\langle\mathbf{h}_i|\mathcal{X}_i\rangle &= \left(\mathbf{I}+\sum_{k=1}^{K}H_i(k)\mathbf{V}^\mathsf{T}\boldsymbol{\Phi}^{-1}\mathbf{V}\right)^{-1}\mathbf{V}^\mathsf{T}\boldsymbol{\Phi}^{-1}\sum_{k=1}^{K}\sum_{j=1}^{H_i(k)}(\mathbf{x}_{ij}^k-\mathbf{m}) \\
\langle\mathbf{h}_i\mathbf{h}_i^\mathsf{T}|\mathcal{X}_i\rangle &= \left(\mathbf{I}+\sum_{k=1}^{K}H_i(k)\mathbf{V}^\mathsf{T}\boldsymbol{\Phi}^{-1}\mathbf{V}\right)^{-1}+\langle\mathbf{h}_i|\mathcal{X}_i\rangle\langle\mathbf{h}_i|\mathcal{X}_i\rangle^\mathsf{T},
\end{aligned} \qquad (5)
$$

where $\mathcal{X}_i$ represents the set of i-vectors from speaker i in all of the K SNR groups.

Similarly, to compute the posterior expectations of $\mathbf{w}_k$, we marginalize over $\mathbf{h}_i$'s. Thus, the posterior density of $\mathbf{w}_k$ is

$$
\begin{aligned}
p(\mathbf{w}_k|\mathbf{x}_{ij}^k,\boldsymbol{\theta}) &\propto \int p(\mathbf{x}_{ij}^k|\mathbf{h}_i,\mathbf{w}_k,\boldsymbol{\theta})p(\mathbf{h}_i)p(\mathbf{w}_k)\mathrm{d}\mathbf{h}_i \\
&= \int \mathcal{N}(\mathbf{x}_{ij}^k|\mathbf{m}+\mathbf{V}\mathbf{h}_i+\mathbf{U}\mathbf{w}_k,\boldsymbol{\Sigma})\mathcal{N}(\mathbf{h}_i|\mathbf{0},\mathbf{I})\mathcal{N}(\mathbf{w}_k|\mathbf{0},\mathbf{I})\mathrm{d}\mathbf{h}_i \\
&= \mathcal{N}(\mathbf{x}_{ij}^k|\mathbf{m}+\mathbf{U}\mathbf{w}_k,\boldsymbol{\Psi})\mathcal{N}(\mathbf{w}_k|\mathbf{0},\mathbf{I}) \\
&\propto \exp\left\{\mathbf{w}_k^\mathsf{T}\mathbf{U}^\mathsf{T}\boldsymbol{\Psi}^{-1}(\mathbf{x}_{ij}^k-\mathbf{m})-\frac{1}{2}\mathbf{w}_k^\mathsf{T}(\mathbf{I}+\mathbf{U}^\mathsf{T}\boldsymbol{\Psi}^{-1}\mathbf{U})\mathbf{w}_k\right\}
\end{aligned} \qquad (6)
$$

which results in

$$
\begin{aligned}
\langle\mathbf{w}_k|\mathbf{x}_{ij}^k\rangle &= \left(\mathbf{I}+\mathbf{U}^\mathsf{T}\boldsymbol{\Psi}^{-1}\mathbf{U}\right)^{-1}\mathbf{U}^\mathsf{T}\boldsymbol{\Psi}^{-1}(\mathbf{x}_{ij}^k-\mathbf{m}) \\
\langle\mathbf{w}_k\mathbf{w}_k^\mathsf{T}|\mathbf{x}_{ij}^k\rangle &= \left(\mathbf{I}+\mathbf{U}^\mathsf{T}\boldsymbol{\Psi}^{-1}\mathbf{U}\right)^{-1}+\langle\mathbf{w}_k|\mathbf{x}_{ij}^k\rangle\langle\mathbf{w}_k|\mathbf{x}_{ij}^k\rangle^\mathsf{T}
\end{aligned} \qquad (7)
$$

where $\boldsymbol{\Psi}=\mathbf{V}\mathbf{V}^\mathsf{T}+\boldsymbol{\Sigma}$. Given all of the i-vectors from the k-th SNR group, we can compute

the posterior expectations as follows:

$$\begin{aligned}\langle \mathbf{w}_k|\mathcal{X}^k\rangle &= \left(\mathbf{I}+\sum_{i=1}^{S} H_i(k)\mathbf{U}^\mathsf{T}\boldsymbol{\Psi}^{-1}\mathbf{U}\right)^{-1}\mathbf{U}^\mathsf{T}\boldsymbol{\Psi}^{-1}\sum_{i=1}^{S}\sum_{j=1}^{H_i(k)}(\mathbf{x}_{ij}^k-\mathbf{m})\\ \langle \mathbf{w}_k\mathbf{w}_k^\mathsf{T}|\mathcal{X}^k\rangle &= \left(\mathbf{I}+\sum_{i=1}^{S} H_i(k)\mathbf{U}^\mathsf{T}\boldsymbol{\Psi}^{-1}\mathbf{U}\right)^{-1}+\langle \mathbf{w}_k|\mathcal{X}^k\rangle\langle \mathbf{w}_k|\mathcal{X}^k\rangle^\mathsf{T}.\end{aligned} \tag{8}$$

where $\mathcal{X}^k$ is the set of i-vectors from the k-th SNR group. Eq. 5 and Eq. 8 constitute the E-step.

In the M-step, we assume that the latent factors $\mathbf{h}_i$ and $\mathbf{w}_k$ are independent and maximize the following auxiliary function:

$$\begin{aligned}Q(\boldsymbol{\theta}) &= \mathbb{E}_{\mathcal{H},\mathcal{W}}\left\{\sum\nolimits_{ijk}\ln\mathcal{N}\left(\mathbf{x}_{ij}^k|\mathbf{m}+\mathbf{V}\mathbf{h}_i+\mathbf{U}\mathbf{w}_k,\boldsymbol{\Sigma}\right)\mathcal{N}(\mathbf{h}_i|\mathbf{0},\mathbf{I})\mathcal{N}(\mathbf{w}_k|\mathbf{0},\mathbf{I})\middle|\mathcal{X},\boldsymbol{\theta}\right\}\\ &= -\frac{1}{2}\sum_{ijk}\mathbb{E}_{\mathcal{H},\mathcal{W}}\left\{\log|\boldsymbol{\Sigma}|+\left(\mathbf{x}_{ij}^k-\mathbf{m}-\mathbf{V}\mathbf{h}_i-\mathbf{U}\mathbf{w}_k\right)^\mathsf{T}\boldsymbol{\Sigma}^{-1}\right.\\ &\qquad\left.\left(\mathbf{x}_{ij}^k-\mathbf{m}-\mathbf{V}\mathbf{h}_i-\mathbf{U}\mathbf{w}_k\right)+\mathbf{h}_i^\mathsf{T}\mathbf{h}_i+\mathbf{w}_k^\mathsf{T}\mathbf{w}_k|\mathcal{X},\boldsymbol{\theta}\right\}\end{aligned} \tag{9}$$

where $\mathcal{H}=\{\mathbf{h}_i; i=1,\ldots,H_i\}$ and $\mathcal{W}=\{\mathbf{w}_k; k=1,\ldots,K\}$. Differentiating Eq. 9 with respect to $\mathbf{V}$, $\mathbf{U}$, and $\boldsymbol{\Sigma}$, we obtain

$$\begin{aligned}\mathbf{V} &= \left[\sum\nolimits_{ijk}(\mathbf{x}_{ij}^k-\mathbf{m}-\mathbf{W}\langle \mathbf{w}_k|\mathcal{X}^k\rangle)\langle \mathbf{h}_i|\mathcal{X}^k\rangle^\mathsf{T}\right]\left[\sum\nolimits_{ijk}\langle \mathbf{h}_i\mathbf{h}_i^\mathsf{T}|\mathcal{X}_i\rangle\right]^{-1}\\ \mathbf{W} &= \left[\sum\nolimits_{ijk}(\mathbf{x}_{ij}^k-\mathbf{m}-\mathbf{V}\langle \mathbf{h}_i|\mathcal{X}_i\rangle)\langle \mathbf{w}_k|\mathcal{X}^k\rangle^\mathsf{T}\right]\left[\sum\nolimits_{ijk}\langle \mathbf{w}_k\mathbf{w}_k^\mathsf{T}|\mathcal{X}^k\rangle\right]^{-1}\\ \boldsymbol{\Sigma} &= \frac{1}{\sum_{ik}H_i(k)}\left\{\sum\nolimits_{ijk}\left[(\mathbf{x}_{ij}^k-\mathbf{m})(\mathbf{x}_{ij}^k-\mathbf{m})^\mathsf{T}-(\mathbf{V}\langle \mathbf{h}_i|\mathcal{X}_i\rangle+\mathbf{U}\langle \mathbf{w}_k|\mathcal{X}^k\rangle)(\mathbf{x}_{ij}^k-\mathbf{m})^\mathsf{T}\right]\right\}\end{aligned} \tag{10}$$

where the posterior expectations have been computed in the E-step (Eq. 5 and Eq. 8).

1.2.2 Joint Estimation

In this approach, we assume that the latent factors $\mathbf{h}_i$ and $\mathbf{w}_k$ are dependent so that their joint posterior should be estimated. Using Eq. 1, we can express the i-vectors from speaker i as

$$\begin{bmatrix}\mathbf{x}_{i1}^1\\ \vdots\\ \mathbf{x}_{iH_i(1)}^1\\ \mathbf{x}_{i1}^2\\ \vdots\\ \mathbf{x}_{iH_i(2)}^2\\ \vdots\\ \mathbf{x}_{i1}^K\\ \vdots\\ \mathbf{x}_{iH_i(K)}^K\end{bmatrix}=\begin{bmatrix}\mathbf{m}\\ \vdots\\ \mathbf{m}\\ \mathbf{m}\\ \vdots\\ \mathbf{m}\\ \vdots\\ \mathbf{m}\\ \vdots\\ \mathbf{m}\end{bmatrix}+\begin{bmatrix}\mathbf{V}&\mathbf{U}&\mathbf{0}&\cdots&\mathbf{0}\\ \vdots&\vdots&\vdots&\ddots&\vdots\\ \mathbf{V}&\mathbf{U}&\mathbf{0}&\cdots&\mathbf{0}\\ \mathbf{V}&\mathbf{0}&\mathbf{U}&\cdots&\mathbf{0}\\ \vdots&\vdots&\vdots&\ddots&\vdots\\ \mathbf{V}&\mathbf{0}&\mathbf{U}&\cdots&\mathbf{0}\\ \vdots&\vdots&\vdots&\vdots&\vdots\\ \mathbf{V}&\mathbf{0}&\mathbf{0}&\cdots&\mathbf{U}\\ \vdots&\vdots&\vdots&\ddots&\vdots\\ \mathbf{V}&\mathbf{0}&\mathbf{0}&\cdots&\mathbf{U}\end{bmatrix}\begin{bmatrix}\mathbf{h}_i\\ \mathbf{w}_1\\ \vdots\\ \mathbf{w}_K\end{bmatrix}+\begin{bmatrix}\boldsymbol{\epsilon}_{i1}^1\\ \vdots\\ \boldsymbol{\epsilon}_{iH_i(1)}^1\\ \boldsymbol{\epsilon}_{i1}^2\\ \vdots\\ \boldsymbol{\epsilon}_{iH_i(2)}^2\\ \vdots\\ \boldsymbol{\epsilon}_{i1}^K\\ \vdots\\ \boldsymbol{\epsilon}_{iH_i(K)}^K\end{bmatrix} \tag{11}$$

Eq. 11 can be written in a compact form:

$$\tilde{\mathbf{x}}_i = \tilde{\mathbf{m}} + \mathbf{A}\tilde{\mathbf{z}}_i + \tilde{\boldsymbol{\epsilon}}_i, \tag{12}$$

where the correspondence between the terms in Eq. 11 and Eq. 12 is obvious.

Given an initial value $\boldsymbol{\theta}$, we aim to find a new estimate $\boldsymbol{\theta}'$ that maximizes the auxiliary function:

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}') &= \mathbb{E}_{\mathcal{Z}}\left\{\sum\nolimits_i \ln p(\tilde{\mathbf{x}}_i|\tilde{\mathbf{z}}_i,\boldsymbol{\theta}')p(\tilde{\mathbf{z}}_i)\middle|\mathcal{X},\boldsymbol{\theta}\right\} \\ &= \mathbb{E}_{\mathcal{Z}}\left\{\sum\nolimits_i \ln \mathcal{N}(\tilde{\mathbf{x}}_i|\tilde{\mathbf{m}}' + \mathbf{A}'\tilde{\mathbf{z}}_i, \tilde{\boldsymbol{\Sigma}}')\mathcal{N}(\tilde{\mathbf{z}}_i|\mathbf{0},\mathbf{I})\middle|\mathcal{X},\boldsymbol{\theta}\right\}. \end{aligned} \tag{13}$$

To simplify notations, we drop the symbol ($'$) in Eq. 13 and ignore the constant terms independent on the model parameters, which results in

$$Q(\boldsymbol{\theta}) = \sum_i \left[-\frac{1}{2}\log|\tilde{\boldsymbol{\Sigma}}| - \frac{1}{2}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{m}} - \mathbf{A}\langle\tilde{\mathbf{z}}_i|\mathcal{X}\rangle)^{\mathsf{T}}\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\mathbf{x}}_{ij} - \tilde{\mathbf{m}} - \mathbf{A}\langle\tilde{\mathbf{z}}_i|\mathcal{X}\rangle)\right]. \tag{14}$$

Evaluating $Q(\boldsymbol{\theta})$ requires computing the posterior expectations:

$$\langle\tilde{\mathbf{z}}_i|\mathcal{X}_i\rangle \quad \text{and} \quad \langle\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^{\mathsf{T}}|\mathcal{X}_i\rangle. \tag{15}$$

These expectations can be obtained by comparing the posterior distribution

$$\begin{aligned} p(\tilde{\mathbf{z}}_i|\tilde{\mathbf{x}}_i,\boldsymbol{\theta}) &\propto p(\tilde{\mathbf{x}}_i|\tilde{\mathbf{z}}_i,\boldsymbol{\theta})p(\tilde{\mathbf{z}}_i) \\ &= \mathcal{N}(\tilde{\mathbf{x}}_i|\tilde{\mathbf{m}} + \mathbf{A}\tilde{\mathbf{z}}_i, \tilde{\boldsymbol{\Sigma}})\mathcal{N}(\tilde{\mathbf{z}}_{ik}|\mathbf{0},\mathbf{I}) \end{aligned}$$

with a Gaussian distribution, which result in the following formulae for the E-step:

$$\begin{aligned} \langle\tilde{\mathbf{z}}_i|\mathcal{X}\rangle &= \mathbf{L}^{-1}\mathbf{A}^{\mathsf{T}}\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\mathbf{x}}_i - \tilde{\mathbf{m}}) \\ \langle\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^{\mathsf{T}}|\mathcal{X}\rangle &= \mathbf{L}^{-1} + \langle\tilde{\mathbf{z}}_i|\mathcal{X}\rangle\langle\tilde{\mathbf{z}}_i|\mathcal{X}\rangle^{\mathsf{T}} \\ \mathbf{L} &= \mathbf{I} + \mathbf{A}^{\mathsf{T}}\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{A}. \end{aligned} \tag{16}$$

The posterior expectation of $\mathbf{h}_i$ can now be extracted from Eq. 16 as follows:

$$\langle\mathbf{h}_i|\mathcal{X}_i\rangle = \langle\tilde{\mathbf{z}}_i|\mathcal{X}_i\rangle_{1:P} \quad \text{and} \quad \left\langle\mathbf{h}_i\mathbf{h}_i^{\mathsf{T}}|\mathcal{X}_i\right\rangle = \langle\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^{\mathsf{T}}|\mathcal{X}_i\rangle_{1:P}^{1:P}$$

where $\langle\mathbf{Z}\rangle_{p:q}^{r:s}$ means extracting a sub-matrix of $\mathbf{Z}$ from rows p to q and columns r to s. Similarly, the posterior expectation of $\mathbf{h}_i\mathbf{w}_k$ can be extracted from Eq. 16 as follows:

$$\left\langle\mathbf{h}_i\mathbf{w}_k^{\mathsf{T}}|\mathcal{X}_i^k\right\rangle = \langle\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^{\mathsf{T}}|\mathcal{X}_i^k\rangle_{1:P}^{P+(k-1)Q+1:P+kQ}.$$

To compute the posterior expectation of $\mathbf{w}^k$, we may express the i-vectors from the

k-th SNR group as follows:

$$\begin{bmatrix} \mathbf{x}_{11}^k \\ \vdots \\ \mathbf{x}_{1H_1(k)}^1 \\ \mathbf{x}_{21}^k \\ \vdots \\ \mathbf{x}_{2H_2(k)}^k \\ \vdots \\ \mathbf{x}_{N1}^k \\ \vdots \\ \mathbf{x}_{NH_S(k)}^k \end{bmatrix} = \begin{bmatrix} \mathbf{m} \\ \vdots \\ \mathbf{m} \\ \mathbf{m} \\ \vdots \\ \mathbf{m} \\ \vdots \\ \mathbf{m} \\ \vdots \\ \mathbf{m} \end{bmatrix} + \begin{bmatrix} \mathbf{U} & \mathbf{V} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{U} & \mathbf{V} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{U} & \mathbf{0} & \mathbf{V} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{U} & \mathbf{0} & \mathbf{V} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{U} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{U} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{w}_k \\ \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{11}^k \\ \vdots \\ \boldsymbol{\epsilon}_{1H_1(k)}^k \\ \boldsymbol{\epsilon}_{21}^k \\ \vdots \\ \boldsymbol{\epsilon}_{2H_2(k)}^k \\ \vdots \\ \boldsymbol{\epsilon}_{N1}^k \\ \vdots \\ \boldsymbol{\epsilon}_{NH_S(k)}^k \end{bmatrix}$$

which can be written in a compact form:

$$\tilde{\mathbf{x}}^k = \tilde{\mathbf{m}} + \mathbf{C}\tilde{\mathbf{z}}^k + \tilde{\boldsymbol{\epsilon}}^k. \tag{17}$$

Similar to Eq. 16, in the E-step, we need to compute

$$\begin{aligned} \langle \tilde{\mathbf{z}}^k | \mathcal{X}^k \rangle &= \mathbf{L}^{-1}\mathbf{C}^{\mathsf{T}}\tilde{\boldsymbol{\Sigma}}^{-1}(\tilde{\mathbf{x}}^k - \tilde{\mathbf{m}}) \\ \langle \tilde{\mathbf{z}}^k (\tilde{\mathbf{z}}^k)^{\mathsf{T}} | \mathcal{X}^k \rangle &= \mathbf{L}^{-1} + \langle \tilde{\mathbf{z}}^k | \mathcal{X}^k \rangle \langle \tilde{\mathbf{z}}^k | \mathcal{X}^k \rangle^{\mathsf{T}} \end{aligned} \tag{18}$$

where

$$\mathbf{L} = \mathbf{I} + \mathbf{C}^{\mathsf{T}}\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{C}.$$

The posterior expectation of $\mathbf{w}^k$ can now be extracted from Eq. 18 as follows:

$$\left\langle \mathbf{w}^k | \mathcal{X}^k \right\rangle = \left\langle \tilde{\mathbf{z}}^k | \mathcal{X}^k \right\rangle_{1:Q} \quad \text{and} \quad \left\langle \mathbf{w}^k (\mathbf{w}^k)^{\mathsf{T}} | \mathcal{X}^k \right\rangle = \left\langle \tilde{\mathbf{z}}^k (\tilde{\mathbf{z}}^k)^{\mathsf{T}} | \mathcal{X}^k \right\rangle_{1:Q}^{1:Q}$$

where $\langle \mathbf{Z} \rangle_{p:q}^{r:s}$ means extracting a sub-matrix of $\mathbf{Z}$ from rows p to q and columns r to s.

In the M step, we consider one i-vector at a time and express it in the following form:

$$\begin{aligned} \mathbf{x}_{ij}^k &= \mathbf{m} + [\mathbf{V} \;\; \mathbf{W}] \begin{bmatrix} \mathbf{h}_i \\ \mathbf{w}_k \end{bmatrix} + \boldsymbol{\epsilon}_{ij}^k \\ &= \mathbf{m} + \mathbf{B}\hat{\mathbf{z}}_{ik} + \boldsymbol{\epsilon}_{ij}^k \qquad\qquad j = 1, \ldots, H_i. \end{aligned}$$

where $\mathbf{B} = [\mathbf{V} \;\; \mathbf{W}]^{\mathsf{T}}$ and $\hat{\mathbf{z}}_{ik} = [\mathbf{h}_i^{\mathsf{T}} \;\; \mathbf{w}_k^{\mathsf{T}}]^{\mathsf{T}}$.

The auxiliary function becomes

$$Q(\boldsymbol{\theta}'|\boldsymbol{\theta}) = \mathbb{E}_{\mathcal{Z}} \left\{ \sum\nolimits_{ijk} \ln \mathcal{N}\left(\mathbf{x}_{ij}^k | \mathbf{m}' + \mathbf{B}'\hat{\mathbf{z}}_{ik}, \boldsymbol{\Sigma}'\right) \mathcal{N}(\hat{\mathbf{z}}_{ik} | \mathbf{0}, \mathbf{I}) \middle| \mathcal{X}, \boldsymbol{\theta} \right\}. \tag{19}$$

Differentiate Eq. 19 with respect to $\mathbf{B}'$ and $\boldsymbol{\Sigma}'$ and set the results to zero, we obtain the update formulae:

$$\begin{aligned} \mathbf{m}' &= \frac{\sum_{ijk} \mathbf{x}_{ij}^k}{\sum_{ik} H_i} \\ \mathbf{B}' &= \left[\sum\nolimits_{ijk} (\mathbf{x}_{ij}^k - \mathbf{m}') \langle \hat{\mathbf{z}}_{ik} | \mathcal{X} \rangle^{\mathsf{T}}\right] \left[\sum\nolimits_{ijk} \langle \hat{\mathbf{z}}_{ik} \hat{\mathbf{z}}_{ik}^{\mathsf{T}} | \mathcal{X} \rangle\right]^{-1} \\ \boldsymbol{\Sigma}' &= \frac{1}{\sum_{ik} H_i} \left\{ \sum\nolimits_{ijk} \left[(\mathbf{x}_{ij}^k - \mathbf{m}')(\mathbf{x}_{ij}^k - \mathbf{m}')^{\mathsf{T}} - \mathbf{B}' \langle \hat{\mathbf{z}}_{ik} | \mathcal{X} \rangle (\mathbf{x}_{ij}^k - \mathbf{m}')^{\mathsf{T}} \right] \right\} \end{aligned}$$

where

$$\langle \hat{\mathbf{z}}_{ik}|\mathcal{X}\rangle = \begin{bmatrix} \langle \mathbf{h}_i|\mathcal{X}_i\rangle \\ \\ \langle \mathbf{w}_k|\mathcal{X}^k\rangle \end{bmatrix}$$

$$\left\langle \hat{\mathbf{z}}_{ik}\hat{\mathbf{z}}_{ik}^{\mathsf{T}}|\mathcal{X}\right\rangle = \begin{bmatrix} \langle \mathbf{h}_i\mathbf{h}_i^{\mathsf{T}}|\mathcal{X}_i\rangle & \langle \mathbf{h}_i\mathbf{w}_k^{\mathsf{T}}|\mathcal{X}_i^k\rangle \\ \\ \langle \mathbf{h}_i\mathbf{w}_k^{\mathsf{T}}|\mathcal{X}_i^k\rangle^{\mathsf{T}} & \langle \mathbf{w}_k\mathbf{w}_k^{\mathsf{T}}|\mathcal{X}^k\rangle \end{bmatrix}$$

2 Likelihood-Ratio Scores

Denote $\mathbf{x}_s$ and $\mathbf{x}_t$ as the length-normalized i-vectors of a target-speaker and a test speaker (claimant), respectively. If $\mathbf{x}_s$ and $\mathbf{x}_t$ are from the same speaker, then we have

$$\begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mathbf{m} \\ \mathbf{m} \end{bmatrix} + \begin{bmatrix} \mathbf{V} & \mathbf{U} & \mathbf{0} \\ \mathbf{V} & \mathbf{0} & \mathbf{U} \end{bmatrix} \begin{bmatrix} \mathbf{h} \\ \mathbf{w}_s \\ \mathbf{w}_t \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_s \\ \boldsymbol{\epsilon}_t \end{bmatrix} \tag{20}$$

where $\mathbf{h}$ represents the speaker factors shared by both i-vectors and $\mathbf{w}_s$ and $\mathbf{w}_t$ represent the SNR factors of the two utterances, respectively. Eq. 20 can be written in a compact form:

$$\hat{\mathbf{x}}_{st} = \hat{\mathbf{m}} + \hat{\mathbf{A}}\hat{\mathbf{z}}_{st} + \hat{\boldsymbol{\epsilon}}_{st}.$$

Assuming that the length-normalized i-vectors follow a Gaussian distribution, the distribution of $\hat{\mathbf{x}}_{st}$ can be obtained by marginalizing over all possible latent factors as follows:

$$\begin{aligned} p(\hat{\mathbf{x}}_{st}|\text{same-speaker}) &= \int p(\hat{\mathbf{x}}_{st}|\hat{\mathbf{z}}_{st})p(\hat{\mathbf{z}}_{st})\mathrm{d}\hat{\mathbf{z}}_{st} \\ &= \int \mathcal{N}(\hat{\mathbf{x}}_{st}|\hat{\mathbf{m}} + \hat{\mathbf{A}}\hat{\mathbf{z}}_{st}, \hat{\boldsymbol{\Sigma}})\mathcal{N}(\hat{\mathbf{z}}_{st}|\mathbf{0}, \mathbf{I}) \\ &= \mathcal{N}(\hat{\mathbf{x}}_{st}|\hat{\mathbf{m}}, \hat{\mathbf{A}}\hat{\mathbf{A}}^{\mathsf{T}} + \hat{\boldsymbol{\Sigma}}) \\ &= \mathcal{N}\left(\begin{bmatrix}\mathbf{x}_s \\ \mathbf{x}_t\end{bmatrix} \middle| \begin{bmatrix}\mathbf{m} \\ \mathbf{m}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{tot} & \boldsymbol{\Sigma}_{ac} \\ \boldsymbol{\Sigma}_{ac} & \boldsymbol{\Sigma}_{tot}\end{bmatrix}\right) \end{aligned} \tag{21}$$

where $\hat{\boldsymbol{\Sigma}} = \text{diag}\{\boldsymbol{\Sigma}, \boldsymbol{\Sigma}\}$, $\boldsymbol{\Sigma}_{tot} = \mathbf{V}\mathbf{V}^{\mathsf{T}} + \mathbf{U}\mathbf{U}^{\mathsf{T}} + \boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{ac} = \mathbf{V}\mathbf{V}^{\mathsf{T}}$. If $\mathbf{x}_s$ and $\mathbf{x}_t$ are from the utterances of two different speakers, we have

$$\begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mathbf{m} \\ \mathbf{m} \end{bmatrix} + \begin{bmatrix} \mathbf{V} & \mathbf{0} & \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} & \mathbf{0} & \mathbf{U} \end{bmatrix} \begin{bmatrix} \mathbf{h}_s \\ \mathbf{h}_t \\ \mathbf{w}_s \\ \mathbf{w}_t \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_s \\ \boldsymbol{\epsilon}_t \end{bmatrix} \tag{22}$$

which can be compactly written as

$$\hat{\mathbf{x}}_{st} = \hat{\mathbf{m}} + \bar{\mathbf{A}}\bar{\mathbf{z}}_{st} + \hat{\boldsymbol{\epsilon}}_{st}$$

The distribution of $\hat{\mathbf{x}}_{st}$ can be obtained by marginalizing over $\bar{\mathbf{z}}_{st}$:

$$\begin{aligned} p(\hat{\mathbf{x}}_{st}|\text{diff-speaker}) &= \int p(\hat{\mathbf{x}}_{st}|\bar{\mathbf{z}}_{st})p(\bar{\mathbf{z}}_{st})\mathrm{d}\bar{\mathbf{z}}_{st} \\ &= \int \mathcal{N}(\hat{\mathbf{x}}_{st}|\hat{\mathbf{m}} + \bar{\mathbf{A}}\bar{\mathbf{z}}_{st}, \hat{\boldsymbol{\Sigma}})\mathcal{N}(\bar{\mathbf{z}}_{st}|\mathbf{0}, \mathbf{I}) \\ &= \mathcal{N}(\hat{\mathbf{x}}_{st}|\hat{\mathbf{m}}, \bar{\mathbf{A}}\bar{\mathbf{A}}^{\mathsf{T}} + \hat{\boldsymbol{\Sigma}}) \end{aligned}$$

$$= \mathcal{N}\left(\begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix} \middle| \begin{bmatrix} \mathbf{m} \\ \mathbf{m} \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{tot} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{tot} \end{bmatrix} \right) \tag{23}$$

Combining Eq. 21 and Eq. 23 and assuming all i-vectors have been mean subtracted ($\mathbf{x} \leftarrow \mathbf{x} - \mathbf{m}$), we have the log-likelihood ratio score:

$$\begin{aligned}
S_{\text{LR}}(\mathbf{x}_s, \mathbf{x}_t) &= \log \frac{\mathcal{N}\left(\begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix} \middle| \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{tot} & \mathbf{\Sigma}_{ac} \\ \mathbf{\Sigma}_{ac} & \mathbf{\Sigma}_{tot} \end{bmatrix} \right)}{\mathcal{N}\left(\begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix} \middle| \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_{tot} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{tot} \end{bmatrix} \right)} \\
&= -\frac{1}{2} \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mathbf{\Sigma}_{tot} & \mathbf{\Sigma}_{ac} \\ \mathbf{\Sigma}_{ac} & \mathbf{\Sigma}_{tot} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mathbf{\Sigma}_{tot} & 0 \\ 0 & \mathbf{\Sigma}_{tot} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix} + \text{const} \\
&= -\frac{1}{2} \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} (\mathbf{\Sigma}_{tot} - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac})^{-1} & -\mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac}(\mathbf{\Sigma}_{tot} - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac})^{-1} \\ -(\mathbf{\Sigma}_{tot} - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac})^{-1}\mathbf{\Sigma}_{ac}\mathbf{\Sigma}_{tot}^{-1} & (\mathbf{\Sigma}_{tot} - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix} \\
&\quad + \frac{1}{2} \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mathbf{\Sigma}_{tot}^{-1} & 0 \\ 0 & \mathbf{\Sigma}_{tot}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix} + \text{const} \\
&= \frac{1}{2} \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \mathbf{\Sigma}_{tot}^{-1} - (\mathbf{\Sigma}_{tot} - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac})^{-1} & \mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac}(\mathbf{\Sigma}_{tot} - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac})^{-1} \\ \mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac}(\mathbf{\Sigma}_{tot} - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac})^{-1} & \mathbf{\Sigma}_{tot}^{-1} - (\mathbf{\Sigma}_{tot} - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix} + \text{const} \\
&= \frac{1}{2} \begin{bmatrix} \mathbf{x}_s^{\mathrm{T}} & \mathbf{x}_t^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{Q} & \mathbf{P} \\ \mathbf{P} & \mathbf{Q} \end{bmatrix} \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_t \end{bmatrix} + \text{const} \\
&= \frac{1}{2}[\mathbf{x}_s^{\mathrm{T}}\mathbf{Q}\mathbf{x}_s + \mathbf{x}_s^{\mathrm{T}}\mathbf{P}\mathbf{x}_t + \mathbf{x}_t^{\mathrm{T}}\mathbf{P}\mathbf{x}_s + \mathbf{x}_t^{\mathrm{T}}\mathbf{Q}\mathbf{x}_t] + \text{const} \\
&= \frac{1}{2}[\mathbf{x}_s^{\mathrm{T}}\mathbf{Q}\mathbf{x}_s + 2\mathbf{x}_s^{\mathrm{T}}\mathbf{P}\mathbf{x}_t + \mathbf{x}_t^{\mathrm{T}}\mathbf{Q}\mathbf{x}_t] + \text{const}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{Q} &= \mathbf{\Sigma}_{tot}^{-1} - (\mathbf{\Sigma}_{tot} - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac})^{-1} \\
\mathbf{P} &= \mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac}(\mathbf{\Sigma}_{tot} - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_{tot}^{-1}\mathbf{\Sigma}_{ac})^{-1}.
\end{aligned} \tag{24}$$