Lecture Notes on Uncertainty Propagation for I-Vector/PLDA Speaker Verification

Man-Wai MAK

Dept. of Electronic and Information Engineering,
The Hong Kong Polytechnic University
enmwmak@polyu.edu.hk

Abstract

This document provides the detailed formulations and implementation of uncertainty propagation. The formulations are derived based on the PLDA model with Eigen-channels.

Please cite this document as: M.W. Mak, "Lecture Notes on Uncertainty Propagation for I-Vector/PLDA Speaker Verification", *Technical Report and Lecture Note Series, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University*, April 2015.

Keywords: Speaker verification, PLDA, I-vectors, Uncertainty propagation

1. I-vectors

1.1. Generative Model

The i-vectors are assumed to be generated by the following factor analysis model:[1]

$$\boldsymbol{\mu}_s = \boldsymbol{\mu} + \boldsymbol{T}\mathbf{w}_s \tag{1}$$

where $\boldsymbol{\mu}_s$ is the GMM-supervector corresponding to speaker s, $\boldsymbol{\mu}$ is obtained by stacking the mean vectors of a UBM, $\boldsymbol{T}$ is a low-rank total variability matrix modeling the speaker and channel variability, $\mathbf{w}_s$ is the latent factor. Given an utterance with acoustic vectors $\mathcal{X}_s$, the posterior mean of $\mathbf{w}_s$ is the i-vector representing the utterance.

[1] In this document, matrices and vectors in GMM-supervector space are represented by italic bold and low-dimensional vectors such as i-vectors are represented by boldface.

2. I-vector Extraction

Denote the acoustic vectors of an utterance as:

$$\mathcal{O} = \{\mathbf{o}_1, \ldots, \mathbf{o}_T\}$$

where T is the number of acoustic vectors (frames). Given a UBM $\mathbf{\Lambda}^{(b)} = \{\lambda_k^{(b)}, \boldsymbol{\mu}_k^{(b)}, \mathbf{\Sigma}_k^{(b)}\}_{k=1}^{M}$ with M mixture components and acoustic vectors $\mathcal{O}$ from speaker s, the zero-order and centered first-order Baum-Welch statistics are computed as follows [4, 1, 3]:

$$n_{i,k} = \sum_{t=1}^{T} \Pr(C_t = k|\mathbf{o}_t) \quad \text{and} \quad \tilde{\mathbf{f}}_{i,k} = \sum_{t=1}^{T} \Pr(C_t = k|\mathbf{o}_t)(\mathbf{o}_t - \boldsymbol{\mu}_k^{(b)}) \tag{2}$$

where $C_t \in \{1, \ldots, M\}$ is a discrete random variable indicating which of the M mixtures is responsible for generating $\mathbf{o}_t$ and

$$\Pr(C_t = k|\mathbf{o}_t) = \frac{\lambda_k^{(b)} \mathcal{N}(\mathbf{o}_t; \boldsymbol{\mu}_k^{(b)}, \mathbf{\Sigma}_k^{(b)})}{\sum_{j=1}^{M} \lambda_j^{(b)} \mathcal{N}(\mathbf{o}_t; \boldsymbol{\mu}_j^{(b)}, \mathbf{\Sigma}_j^{(b)})}, \quad k = 1, \ldots, M$$

is the posterior probability of mixture component k given $\mathbf{o}_t$. The posterior covariance and posterior mean associated with the i-vector of speaker s are given by:

$$\text{Cov}(\mathbf{w}_s, \mathbf{w}_s|\mathcal{O}) = \mathbf{L}_s^{-1} \tag{3}$$

$$\langle \mathbf{w}_s|\mathcal{O}\rangle = \mathbf{L}_s^{-1} \boldsymbol{T}^{\mathsf{T}} {\mathbf{\Sigma}^{(b)}}^{-1} \tilde{\boldsymbol{f}}_s \tag{4}$$

where

$$\mathbf{L}_s = \mathbf{I} + \boldsymbol{T}^{\mathsf{T}} {\mathbf{\Sigma}^{(b)}}^{-1} \boldsymbol{N}_s \boldsymbol{T} \tag{5}$$

is a precision matrix and $\mathbf{I}$ is the identity matrix. $\boldsymbol{N}_s$ is an $MD \times MD$ diagonal matrix whose diagonal blocks are $n_{i,k}\mathbf{I}$. $\tilde{\boldsymbol{f}}_s$ is an $MD \times 1$ supervector formed by concatenating the centered first-order Baum-Welch statistics $\tilde{\mathbf{f}}_{i,k}$. $\mathbf{\Sigma}^{(b)}$ is a covariance matrix modeling the residual variability not captured by the $MD \times R$ total variability matrix $\boldsymbol{T}$. In practice, we substitute this matrix by the covariance matrices of the UBM, i.e., $\mathbf{\Sigma}^{(b)} = \text{diag}\{\mathbf{\Sigma}_1^{(b)}, \ldots, \mathbf{\Sigma}_M^{(b)}\}$. The posterior mean (Eq. 4) is the i-vector representing the speaker s.

3. PLDA Models with Uncertainty Propagation

3.1. PLDA Modeling

In standard i-vector/PLDA systems, the i-vectors $\mathbf{w}_s$ needed to be *pre-processed* by whitening, length normalization and optionally LDA followed by WCCN. The pre-processing procedure is an important step for effective Gaussian PLDA modelling. If whitening and length normalization are not performed, this pre-processing step can be represented by a linear transformation. Specifically, given the r-th session of speaker s, the pre-processed i-vector is given by

$$\mathbf{i}_{s,r} = \mathbf{P}\mathbf{w}_{s,r} \tag{6}$$

where $\mathbf{P}$ is a transformation matrix combining both LDA and WCCN. The i-vector $\mathbf{i}_{s,r}$ is then directly plugged into the PLDA model:

$$\mathbf{i}_{s,r} = \mathbf{m} + \mathbf{V}\mathbf{y}_s + \boldsymbol{\epsilon}_{s,r}, \tag{7}$$

where $\mathbf{m}$ is the global mean of pre-processed i-vectors, $\mathbf{V}$ is the speaker loading matrix and $\boldsymbol{\epsilon}_{s,r} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ represents the residue that cannot be modelled by the linear model. Eq. 6 and Eq. 7 suggest that i-vector extraction and PLDA modelling are de-coupled. The two processes can be de-coupled because only point estimated i-vectors are used for PLDA modeling.

3.2. Uncertainty Propagation

In [5], Kenny et al. argue that i-vector extraction and PLDA modelling should be tightly coupled when the utterances are short. The main argument is that short utterances lead to large posterior covariance $\mathbf{L}_s^{-1}$ (Eq. 5) in the estimated i-vectors. To take the covariance of i-vectors into account during PLDA modelling, Kenny et al. [5] proposed a method called uncertainty propagation. The method uses the posterior distribution of i-vectors rather than their point estimates. To propagate this information to the PLDA model, an additional session- and speaker-dependent factor is added to the PLDA model. More specifically, the pre-processed i-vector ($\mathbf{i}_{s,r}$) of the r-th session of speaker s is assumed to be generated by the following model:

$$\mathbf{i}_{s,r} = \mathbf{m} + \mathbf{V}\mathbf{y}_s + \mathbf{U}_{s,r}\mathbf{x}_{s,r} + \boldsymbol{\epsilon}_{s,r} \qquad \boldsymbol{\epsilon}_{s,r} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{8}$$

where $\mathbf{U}_{s,r}$ models the i-vector variability and $\mathbf{x}_{s,r}$ is the speaker- and session-dependent factor. If length-normalization is not applied to i-vectors,

$\mathbf{U}_{s,r}$ can be computed from the Cholesky decomposition of

$$\mathbf{U}_{s,r}\mathbf{U}_{s,r}^{\mathsf{T}} = \mathrm{cov}(\mathbf{i}_{s,r}, \mathbf{i}_{s,r}) = \mathbf{P}\mathbf{L}_{s,r}^{-1}\mathbf{P}^{\mathsf{T}} \tag{9}$$

where

$$\mathbf{L}_{s,r}^{-1} = \mathbf{I} + \boldsymbol{T}^{\mathsf{T}}{\boldsymbol{\Sigma}^{(b)}}^{-1}\boldsymbol{N}_{s,r}\boldsymbol{T}.$$

When length normalization is applied, the situation becomes more complicated because length normalization is a non-linear transformation so that the pre-processing procedure mentioned earlier cannot be represented by a linear transformation. Nevertheless, Kenny et al. suggested the following expedient method:

$$\begin{aligned} \mathbf{U}_{s,r}\mathbf{U}_{s,r}^{\mathsf{T}} &= \mathrm{cov}(\mathbf{w}_{s,r}, \mathbf{w}_{s,r}) = \mathbf{P}\mathbf{L}_{s,r}^{-1}\mathbf{P}^{\mathsf{T}} \\ \mathbf{U}_{s,r} &\leftarrow \frac{\mathbf{U}_{s,r}}{\|\mathbf{w}_{s,r}\|} \end{aligned} \tag{10}$$

where in the first step, $\mathbf{U}_{s,r}$ is computed from the Cholesky decomposition of $\mathbf{L}_{s,r}^{-1}$.

4. EM Formulations for PLDA with Uncertainty Propagation

Assume that we are given a set of length-normalized [2] i-vectors $\mathcal{I} = \{\mathbf{i}_{ij}; i = 1, \ldots, N; j = 1, \ldots, H_i\}$ obtained from N training speakers where speaker i has H_i sessions. Eq. 8 can be written as:

$$\mathbf{i}_{i,j} = \mathbf{m} + [\mathbf{V} \quad \mathbf{U}_{i,j}] \begin{bmatrix} \mathbf{y}_i \\ \mathbf{x}_{i,j} \end{bmatrix} + \boldsymbol{\epsilon}_{i,j} = \mathbf{m} + \mathbf{B}_{i,j}\hat{\mathbf{z}}_{i,j} + \boldsymbol{\epsilon}_{i,j},$$

where $\mathbf{B}_{i,j} = [\mathbf{V} \quad \mathbf{U}_{i,j}]$ and $\hat{\mathbf{z}}_{i,j} = [\mathbf{y}_i^{\mathsf{T}} \quad \mathbf{x}_{i,j}^{\mathsf{T}}]^{\mathsf{T}}$. The model parameters $\boldsymbol{\theta} = \{\mathbf{m}, \mathbf{V}, \boldsymbol{\Sigma}\}$ are estimated by an EM algorithm, and the loading matrix $\mathbf{U}_{i,j}$ can be estimated from the posterior covariance of the i-vector $\mathbf{i}_{i,j}$ as in Eq. 9 or Eq. 10.

4.1. E-Step

In the E-step, we compute the posterior expectation of $\mathbf{y}_i$ by marginalizing over $\mathbf{x}_{i,j}$. Thus, the posterior density of $\mathbf{y}_i$ is written as:

$$\begin{aligned} p(\mathbf{y}_i|\mathbf{i}_{i,j}, \boldsymbol{\theta}) &\propto p(\mathbf{i}_{i,j}|\mathbf{y}_i, \boldsymbol{\theta})p(\mathbf{y}_i) \\ &= \int p(\mathbf{i}_{i,j}, \mathbf{x}_{i,j}|\mathbf{y}_i, \boldsymbol{\theta})p(\mathbf{y}_i)\mathrm{d}\mathbf{x}_{i,j} \end{aligned}$$

$$
\begin{aligned}
&= \int p(\mathbf{i}_{i,j}|\mathbf{y}_i, \mathbf{x}_{i,j}, \boldsymbol{\theta}) p(\mathbf{x}_{i,j}) p(\mathbf{y}_i) \mathrm{d}\mathbf{x}_{i,j} \\
&= \int \mathcal{N}(\mathbf{i}_{i,j}|\mathbf{m} + \mathbf{V}\mathbf{y}_i + \mathbf{U}_{i,j}\mathbf{x}_{i,j}, \boldsymbol{\Sigma}) \mathcal{N}(\mathbf{x}_{i,j}|\mathbf{0}, \mathbf{I}) \mathcal{N}(\mathbf{y}_i|\mathbf{0}, \mathbf{I}) \mathrm{d}\mathbf{x}_{i,j} \\
&= \mathcal{N}(\mathbf{i}_{i,j}|\mathbf{m} + \mathbf{V}\mathbf{y}_i, \boldsymbol{\Phi}_{i,j}) \mathcal{N}(\mathbf{y}_i|\mathbf{0}, \mathbf{I}) \\
&\propto \exp\left\{ \mathbf{y}_i^{\mathsf{T}} \mathbf{V}^{\mathsf{T}} \boldsymbol{\Phi}_{i,j}^{-1} (\mathbf{i}_{i,j} - \mathbf{m}) - \frac{1}{2} \mathbf{y}_i^{\mathsf{T}} (\mathbf{I} + \mathbf{V}^{\mathsf{T}} \boldsymbol{\Phi}_{i,j}^{-1} \mathbf{V}) \mathbf{y}_i \right\} \qquad (11)
\end{aligned}
$$

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

$$
\begin{aligned}
\langle \mathbf{y}_i | \mathbf{i}_{i,j} \rangle &= \left(\mathbf{I} + \mathbf{V}^{\mathsf{T}} \boldsymbol{\Phi}_{i,j}^{-1} \mathbf{V} \right)^{-1} \mathbf{V}^{\mathsf{T}} \boldsymbol{\Phi}_{i,j}^{-1} (\mathbf{i}_{i,j} - \mathbf{m}) \\
\left\langle \mathbf{y}_i \mathbf{y}_i^{\mathsf{T}} | \mathbf{i}_{i,j} \right\rangle &= \left(\mathbf{I} + \mathbf{V}^{\mathsf{T}} \boldsymbol{\Phi}_{i,j}^{-1} \mathbf{V} \right)^{-1} + \langle \mathbf{y}_i | \mathbf{i}_{i,j} \rangle \langle \mathbf{y}_i | \mathbf{i}_{i,j} \rangle^{\mathsf{T}}.
\end{aligned} \qquad (12)
$$

If all of the i-vectors of speaker i are given, the posterior expectations become:

$$
\begin{aligned}
\langle \mathbf{y}_i | \tilde{\mathbf{i}}_i \rangle &= \left(\mathbf{I} + \sum\nolimits_{j=1}^{H_i} \mathbf{V}^{\mathsf{T}} \boldsymbol{\Phi}_{i,j}^{-1} \mathbf{V} \right)^{-1} \mathbf{V}^{\mathsf{T}} \sum\nolimits_{j=1}^{H_i} \boldsymbol{\Phi}_{i,j}^{-1} (\mathbf{i}_{i,j} - \mathbf{m}) \\
\left\langle \mathbf{y}_i \mathbf{y}_i^{\mathsf{T}} | \tilde{\mathbf{i}}_i \right\rangle &= \left(\mathbf{I} + \sum\nolimits_{j=1}^{H_i} \mathbf{V}^{\mathsf{T}} \boldsymbol{\Phi}_{i,j}^{-1} \mathbf{V} \right)^{-1} + \langle \mathbf{y}_i | \tilde{\mathbf{i}}_i \rangle \langle \mathbf{y}_i | \tilde{\mathbf{i}}_i \rangle^{\mathsf{T}}.
\end{aligned} \qquad (13)
$$

where $\tilde{\mathbf{i}}_i$ represents the stacking of all i-vectors of speaker i.

To compute the posterior expectation of $\mathbf{x}_{i,j}$, we marginalize over $\mathbf{y}_i$ by expressing the posterior density of $\mathbf{x}_{i,j}$ as follows:

$$
\begin{aligned}
p(\mathbf{x}_{i,j}|\mathbf{i}_{i,j}, \boldsymbol{\theta}) &\propto p(\mathbf{i}_{i,j}|\mathbf{x}_{i,j}, \boldsymbol{\theta}) p(\mathbf{x}_{i,j}) \\
&= \int p(\mathbf{i}_{i,j}, \mathbf{y}_i|\mathbf{x}_{i,j}, \boldsymbol{\theta}) p(\mathbf{x}_{i,j}) \mathrm{d}\mathbf{y}_i \\
&= \int p(\mathbf{i}_{i,j}|\mathbf{y}_i, \mathbf{x}_{i,j}, \boldsymbol{\theta}) p(\mathbf{x}_{i,j}) p(\mathbf{y}_i) \mathrm{d}\mathbf{y}_i \\
&= \int \mathcal{N}(\mathbf{i}_{i,j}|\mathbf{m} + \mathbf{V}\mathbf{y}_i + \mathbf{U}_{i,j}\mathbf{x}_{i,j}, \boldsymbol{\Sigma}) \mathcal{N}(\mathbf{x}_{i,j}|\mathbf{0}, \mathbf{I}) \mathcal{N}(\mathbf{y}_i|\mathbf{0}, \mathbf{I}) \mathrm{d}\mathbf{y}_i \\
&= \mathcal{N}(\mathbf{i}_{i,j}|\mathbf{m} + \mathbf{U}_{i,j}\mathbf{x}_{i,j}, \boldsymbol{\Psi}) \mathcal{N}(\mathbf{x}_{i,j}|\mathbf{0}, \mathbf{I}) \\
&\propto \exp\left\{ \mathbf{x}_{i,j}^{\mathsf{T}} \mathbf{U}_{i,j}^{\mathsf{T}} \boldsymbol{\Psi}^{-1} (\mathbf{i}_{i,j} - \mathbf{m}) - \frac{1}{2} \mathbf{x}_{i,j}^{\mathsf{T}} (\mathbf{I} + \mathbf{U}_{i,j}^{\mathsf{T}} \boldsymbol{\Psi}^{-1} \mathbf{U}_{i,j}) \mathbf{x}_{i,j} \right\}
\end{aligned} \qquad (14)
$$

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

$$\begin{aligned}\langle \mathbf{x}_{i,j}|\mathbf{i}_{i,j}\rangle &= \left(\mathbf{I} + \mathbf{U}_{i,j}^{\mathsf{T}}\boldsymbol{\Psi}^{-1}\mathbf{U}_{i,j}\right)^{-1}\mathbf{U}_{i,j}^{\mathsf{T}}\boldsymbol{\Psi}^{-1}(\mathbf{i}_{i,j}-\mathbf{m}) \\ \left\langle \mathbf{x}_{i,j}\mathbf{x}_{i,j}^{\mathsf{T}}|\mathbf{i}_{i,j}\right\rangle &= \left(\mathbf{I} + \mathbf{U}_{i,j}^{\mathsf{T}}\boldsymbol{\Psi}^{-1}\mathbf{U}_{i,j}\right)^{-1} + \langle \mathbf{x}_{i,j}|\mathbf{i}_{i,j}\rangle\langle \mathbf{x}_{i,j}|\mathbf{i}_{i,j}\rangle^{\mathsf{T}}.\end{aligned} \tag{15}$$

4.2. M-Step

In the M-step, we assume that the latent factors $\mathbf{y}_i$ and $\mathbf{x}_{i,j}$ are independent. We maximize the following auxiliary function:

$$\begin{aligned}Q(\boldsymbol{\theta}) &= \mathbb{E}_{\mathcal{Y},\mathcal{X}}\left\{\sum\nolimits_{i,j}\ln\mathcal{N}\left(\mathbf{i}_{i,j}\middle|\mathbf{m}+\mathbf{V}\mathbf{y}_i+\mathbf{U}_{i,j}\mathbf{x}_{i,j},\boldsymbol{\Sigma}\right)\mathcal{N}(\mathbf{y}_i|\mathbf{0},\mathbf{I})\mathcal{N}(\mathbf{x}_{i,j}|\mathbf{0},\mathbf{I})\middle|\mathcal{I},\boldsymbol{\theta}\right\} \\ &= -\frac{1}{2}\sum_{i,j}\mathbb{E}_{\mathcal{Y},\mathcal{X}}\Big\{\log|\boldsymbol{\Sigma}| + (\mathbf{i}_{i,j}-\mathbf{m}-\mathbf{V}\mathbf{y}_i-\mathbf{U}_{i,j}\mathbf{x}_{i,j})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1} \\ &\qquad \times(\mathbf{i}_{i,j}-\mathbf{m}-\mathbf{V}\mathbf{y}_i-\mathbf{U}_{i,j}\mathbf{x}_{i,j}) + \mathbf{y}_i^{\mathsf{T}}\mathbf{y}_i + \mathbf{x}_{i,j}^{\mathsf{T}}\mathbf{x}_{i,j}\big|\mathcal{I},\boldsymbol{\theta}\Big\}\end{aligned} \tag{16}$$

where $\mathcal{Y} = \{\mathbf{y}_i; i = 1,\ldots,N\}$ and $\mathcal{X} = \{\mathbf{x}_{i,j}; i = 1,\ldots,N; j = 1,\ldots,H_i\}$. Differentiating Eq. 16 with respect to $\mathbf{V}$ and $\boldsymbol{\Sigma}$, we obtain

$$\begin{aligned}\mathbf{V} &= \left[\sum\nolimits_{i,j}\left(\mathbf{i}_{i,j}-\mathbf{m}-\mathbf{U}_{i,j}\langle\mathbf{x}_{i,j}|\mathbf{i}_{i,j}\rangle\right)\langle\mathbf{y}_i|\tilde{\mathbf{i}}_i\rangle^{\mathsf{T}}\right]\left[\sum\nolimits_{i,j}\langle\mathbf{y}_i\mathbf{y}_i^{\mathsf{T}}|\tilde{\mathbf{i}}_i\rangle\right]^{-1} \\ \boldsymbol{\Sigma} &= \frac{1}{\sum_i H_i}\left\{\sum\nolimits_{i,j}\left[(\mathbf{i}_{i,j}-\mathbf{m})(\mathbf{i}_{i,j}-\mathbf{m})^{\mathsf{T}} - \left(\mathbf{V}\langle\mathbf{y}_i|\tilde{\mathbf{i}}_i\rangle + \mathbf{U}_{i,j}\langle\mathbf{x}_{i,j}|\mathbf{i}_{i,j}\rangle\right)(\mathbf{i}_{i,j}-\mathbf{m})^{\mathsf{T}}\right]\right\}\end{aligned}$$

5. PLDA Scoring with UP

5.1. Exact Scoring

Given a test i-vector $\mathbf{i}_t$ and target-speaker's i-vector $\mathbf{i}_s$, the likelihood ratio score is

$$\begin{aligned}S_{\mathrm{LR}}(\mathbf{i}_s,\mathbf{i}_t) &= \frac{p(\mathbf{i}_s,\mathbf{i}_t|\text{same-speaker})}{p(\mathbf{i}_s,\mathbf{i}_t|\text{different-speakers})} \\ &= \frac{\int\int\int p(\mathbf{i}_s,\mathbf{i}_t,\mathbf{y},\mathbf{x}_s,\mathbf{x}_t)|\boldsymbol{\theta})\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x}_s\mathrm{d}\mathbf{x}_t}{\int\int p(\mathbf{i}_s,\mathbf{y}_s,\mathbf{x}_s|\boldsymbol{\theta})\mathrm{d}\mathbf{y}_s\mathrm{d}\mathbf{x}_s\int\int p(\mathbf{x}_t,\mathbf{y}_t,\mathbf{x}_t|\boldsymbol{\theta})\mathrm{d}\mathbf{y}_t\mathrm{d}\mathbf{x}_t} \\ &= \frac{\int\int\int p(\mathbf{i}_s,\mathbf{i}_t|\mathbf{y},\mathbf{x}_s,\mathbf{x}_t,\boldsymbol{\theta})p(\mathbf{y})p(\mathbf{x}_s)p(\mathbf{x}_t)\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x}_s\mathrm{d}\mathbf{x}_t}{\int\int p(\mathbf{i}_s|\mathbf{y}_s,\mathbf{x}_s,\boldsymbol{\theta})p(\mathbf{y}_s)p(\mathbf{x}_s)\mathrm{d}\mathbf{y}_s\mathrm{d}\mathbf{x}_s\int\int p(\mathbf{i}_t|\mathbf{y}_t,\mathbf{x}_t,\boldsymbol{\theta})p(\mathbf{y}_t)p(\mathbf{x}_t)\mathrm{d}\mathbf{y}_t\mathrm{d}\mathbf{x}_t} \\ &= \frac{\mathcal{N}(\left[\mathbf{i}_s^{\mathsf{T}}\ \ \mathbf{i}_t^{\mathsf{T}}\right]^{\mathsf{T}}|\hat{\mathbf{m}},\hat{\boldsymbol{\Sigma}}_{st})}{\mathcal{N}\left(\left[\mathbf{i}_s^{\mathsf{T}}\ \ \mathbf{i}_t^{\mathsf{T}}\right]^{\mathsf{T}}|\hat{\mathbf{m}},\mathrm{diag}\{\boldsymbol{\Sigma}_s,\boldsymbol{\Sigma}_t\}\right)}\end{aligned} \tag{17}$$

where

$$\hat{\mathbf{m}} = \begin{bmatrix} \mathbf{m}^\mathsf{T} & \mathbf{m}^\mathsf{T} \end{bmatrix}^\mathsf{T} \tag{18}$$

$$\hat{\mathbf{\Sigma}}_{st} = \begin{bmatrix} \mathbf{V}\mathbf{V}^\mathsf{T} + \mathbf{U}_s\mathbf{U}_s^\mathsf{T} + \mathbf{\Sigma} & \mathbf{V}\mathbf{V}^\mathsf{T} \\ \mathbf{V}\mathbf{V}^\mathsf{T} & \mathbf{V}\mathbf{V}^\mathsf{T} + \mathbf{U}_t\mathbf{U}_t^\mathsf{T} + \mathbf{\Sigma} \end{bmatrix} \tag{19}$$

$$\mathbf{\Sigma}_s = \mathbf{V}\mathbf{V}^\mathsf{T} + \mathbf{U}_s\mathbf{U}_s^\mathsf{T} + \mathbf{\Sigma} \tag{20}$$

$$\mathbf{\Sigma}_t = \mathbf{V}\mathbf{V}^\mathsf{T} + \mathbf{U}_t\mathbf{U}_t^\mathsf{T} + \mathbf{\Sigma} \tag{21}$$

Define $\mathbf{\Sigma}_{ac} = \mathbf{V}\mathbf{V}^\mathsf{T}$. We may simplify the logarithm of Eq. 17 as follows:

$$\begin{aligned}
&\log S_{LR}(\mathbf{i}_s, \mathbf{i}_t) \\
&= -\frac{1}{2}\begin{bmatrix}\mathbf{i}_s \\ \mathbf{i}_t\end{bmatrix}^\mathsf{T}\begin{bmatrix}\mathbf{\Sigma}_s & \mathbf{\Sigma}_{ac} \\ \mathbf{\Sigma}_{ac} & \mathbf{\Sigma}_t\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{i}_s \\ \mathbf{i}_t\end{bmatrix} + \frac{1}{2}\begin{bmatrix}\mathbf{i}_s \\ \mathbf{i}_t\end{bmatrix}^\mathsf{T}\begin{bmatrix}\mathbf{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_t\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{i}_s \\ \mathbf{i}_t\end{bmatrix} \\
&\quad - \frac{1}{2}\log\begin{vmatrix}\mathbf{\Sigma}_s & \mathbf{\Sigma}_{ac} \\ \mathbf{\Sigma}_{ac} & \mathbf{\Sigma}_t\end{vmatrix} + \frac{1}{2}\log\begin{vmatrix}\mathbf{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_t\end{vmatrix} \\
&= -\frac{1}{2}\begin{bmatrix}\mathbf{i}_s \\ \mathbf{i}_t\end{bmatrix}^\mathsf{T}\begin{bmatrix}(\mathbf{\Sigma}_s - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_t^{-1}\mathbf{\Sigma}_{ac})^{-1} & -\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma}_{ac}(\mathbf{\Sigma}_t - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma})^{-1}] \\ -(\mathbf{\Sigma}_t - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma}_{ac})^{-1}\mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1} & (\mathbf{\Sigma}_t - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma}_{ac})^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{i}_s \\ \mathbf{i}_t\end{bmatrix} \\
&\quad + \frac{1}{2}\begin{bmatrix}\mathbf{i}_s \\ \mathbf{i}_t\end{bmatrix}^\mathsf{T}\begin{bmatrix}\mathbf{\Sigma}_s^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_t^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{i}_s \\ \mathbf{i}_t\end{bmatrix} - \frac{1}{2}\log\begin{vmatrix}\mathbf{\Sigma}_s & \mathbf{\Sigma}_{ac} \\ \mathbf{\Sigma}_{ac} & \mathbf{\Sigma}_t\end{vmatrix} + \frac{1}{2}\log\begin{vmatrix}\mathbf{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_t\end{vmatrix} \\
&= \frac{1}{2}\begin{bmatrix}\mathbf{i}_s \\ \mathbf{i}_t\end{bmatrix}^\mathsf{T}\begin{bmatrix}\mathbf{\Sigma}_s^{-1} - (\mathbf{\Sigma}_s - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_t^{-1}\mathbf{\Sigma}_{ac})^{-1} & -\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma}_{ac}(\mathbf{\Sigma}_t - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma})^{-1}] \\ -(\mathbf{\Sigma}_t - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma}_{ac})^{-1}\mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1} & \mathbf{\Sigma}_t^{-1} - (\mathbf{\Sigma}_t - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma}_{ac})^{-1}\end{bmatrix} \\
&\quad - \frac{1}{2}\log\begin{vmatrix}\mathbf{\Sigma}_s & \mathbf{\Sigma}_{ac} \\ \mathbf{\Sigma}_{ac} & \mathbf{\Sigma}_t\end{vmatrix} + \frac{1}{2}\log\begin{vmatrix}\mathbf{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_t\end{vmatrix} \\
&= \frac{1}{2}\begin{bmatrix}\mathbf{i}_s^\mathsf{T} & \mathbf{i}_t^\mathsf{T}\end{bmatrix}\begin{bmatrix}\mathbf{A}_{s,t} & \mathbf{B}_{s,t} \\ \mathbf{B}_{s,t} & \mathbf{C}_{s,t}\end{bmatrix}\begin{bmatrix}\mathbf{i}_s \\ \mathbf{i}_t\end{bmatrix} + D_{s,t} \\
&= \frac{1}{2}\left[\mathbf{i}_s^\mathsf{T}\mathbf{A}_{s,t}\mathbf{i}_s + \mathbf{i}_s^\mathsf{T}\mathbf{B}_{s,t}\mathbf{i}_t + \mathbf{i}_t^\mathsf{T}\mathbf{B}_{s,t}\mathbf{i}_s + \mathbf{i}_t^\mathsf{T}\mathbf{C}_{s,t}\mathbf{i}_t\right] + D_{s,t} \\
&= \frac{1}{2}\left[\mathbf{i}_s^\mathsf{T}\mathbf{A}_{s,t}\mathbf{i}_s + 2\mathbf{i}_s^\mathsf{T}\mathbf{B}_{s,t}\mathbf{i}_t + \mathbf{i}_t^\mathsf{T}\mathbf{C}_{s,t}\mathbf{i}_t\right] + D_{s,t}
\end{aligned} \tag{22}$$

where

$$\mathbf{A}_{s,t} = \mathbf{\Sigma}_s^{-1} - (\mathbf{\Sigma}_s - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_t^{-1}\mathbf{\Sigma}_{ac})^{-1} \tag{23}$$

$$\mathbf{B}_{s,t} = -\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma}_{ac}(\mathbf{\Sigma}_t - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma})^{-1} \tag{24}$$

$$\mathbf{C}_{s,t} = \mathbf{\Sigma}_t^{-1} - (\mathbf{\Sigma}_t - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma}_{ac})^{-1} \tag{25}$$

$$D_{s,t} = -\frac{1}{2}\log\begin{vmatrix}\boldsymbol{\Sigma}_s & \boldsymbol{\Sigma}_{ac}\\ \boldsymbol{\Sigma}_{ac} & \boldsymbol{\Sigma}_t\end{vmatrix} + \frac{1}{2}\log\begin{vmatrix}\boldsymbol{\Sigma}_s & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_t\end{vmatrix} \tag{26}$$

Given a test i-vector $\mathbf{i}_t$ and a group of target-speaker's i-vectors $\mathcal{I}_s = \{\mathbf{i}_1, \ldots, \mathbf{i}_{H_s}\}$, the log-likelihood ratio score is

$$\log S_{LR}(\mathcal{I}_s, \mathbf{i}_t) = \frac{1}{H_s}\sum_{j=1}^{H_s}\left(\frac{1}{2}\mathbf{i}_{s,j}^{\mathsf{T}}\mathbf{A}_{s,j,t}\mathbf{i}_{s,j} + \mathbf{i}_{s,j}^{\mathsf{T}}\mathbf{B}_{s,j,t}\mathbf{i}_t + \frac{1}{2}\mathbf{i}_t^{\mathsf{T}}\mathbf{C}_{s,j,t}\mathbf{i}_t + D_{s,j,t}\right), \tag{27}$$

where we extend the subscript s to (s, j).

References

[1] Dehak, N., Kenny, P., Dehak, R., Dumouchel, P., Ouellet, P., 2011. Front-end factor analysis for speaker verification. IEEE Trans. on Audio, Speech, and Language Processing 19, 788–798.

[2] Garcia-Romero, D., Espy-Wilson, C., 2011. Analysis of i-vector length normalization in speaker recognition systems, in: Interspeech'2011, pp. 249–252.

[3] Kenny, P., 2012. A small footprint i-vector extractor, in: Proc. Odyssey 2012, pp. 1–6.

[4] Kenny, P., Ouellet, P., Dehak, N., Gupta, V., Dumouchel, P., 2008. A study of inter-speaker variability in speaker verification. IEEE Trans. on Audio, Speech and Language Processing 16, 980–988.

[5] Kenny, P., Stafylakis, T., Ouellet, P., Alam, M.J., Dumouchel, P., 2013. PLDA for speaker verification with utterances of arbitrary duration, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 7649–7653.