Full-Posterior PLDA in Speaker Recognition — Technical Literature Review —

Andreas Nautsch, Alicia Lozano Diez, Daniel Ramos Castro
da/sec h_da and ATVS-UAM

2017-12-12 (3rd Version of 2016-03-02)

1 Introduction

PLDA estimates the similarity of two biometric features by performing a likelihood ratio test within a linear subspace assuming Gaussian subject and noise distributions [1]. In [1, chapter 18], a PLDA family is depicted:

- the subspace identity model,
- the original PLDA,
- non-linear models,
- asymmetric bilinear models,
- symmetric bilinear and multi-linear models.

Inspired by the original PLDA and the subspace identity model, the speaker recognition community derived further PLDA variants [2, 3, 4, 5, 6], examining different flavors of model compositions and underlying distributions:

- Heavy-Tailed PLDA (HT-PLDA) [2, 4, 5],
- Gaussian PLDA (G-PLDA) [4, 5, 6],
- Two-Covariance model (2-Cov) [3, 5, 6],
- Two-Gaussian model (2-Gau) [5],
- Pairwise SVM (PSVM) [5].

Research in speaker recognition regarding unconstrained conditions addresses effects of different capture/transmission channels, noise and duration variation in reference and/or probe samples. Mismatches in capture channel, i.e. tel. vs. mic., motivated HT-PLDA [2]. Sample duration analysis [4] motivated the application of length-normalization in G-PLDA as an alternative to HT-PLDA due to Radial Gaussianization (RG) effects. Thus, RD and G-PLDA became the defacto state-of-the-art for comparing i-vector features. In theory, i-vectors represent the charactersitic offset within a high-dimensional universal cluster, which is represented as an intermediate-sized Gaussian-distributed random variable consisting of a single-point estimate, the i-vector itself, and the posterior i-vector covariance.

Full-Posterior PLDA (FP-PLDA) was motivated [7, 8] in order to address arbitrary voice sample durations in references and probes by taking i-vector posterior covariances into account, i.e. fully considering the i-vector random variable.

This report states FP-PLDA in context to the established PLDA variants, and depicts the latest research progress of FP-PLDA flavors and optimization approaches.

2 Theory on the PLDA family

In [1], PLDA is motivated by the factor analyses model:

$$\begin{aligned} \boldsymbol{x_i} &= \boldsymbol{\mu} + \boldsymbol{\Phi}\,\boldsymbol{h_i} + \boldsymbol{\epsilon_i}, \\ \Pr(\boldsymbol{x_i}\,|\,\boldsymbol{h_i}) &= \mathcal{N}\left(\boldsymbol{\mu} + \boldsymbol{\Phi}\,\boldsymbol{h_i}, \boldsymbol{\Sigma}\right), \\ \boldsymbol{h_i} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}\right), \\ \boldsymbol{\epsilon_i} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}\right), \textit{with: } \operatorname{diag}[\boldsymbol{\Sigma}], \\ \boldsymbol{x_i} &\sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Phi}\,\boldsymbol{\Phi}' + \boldsymbol{\Sigma}\right), \end{aligned} \tag{1}$$

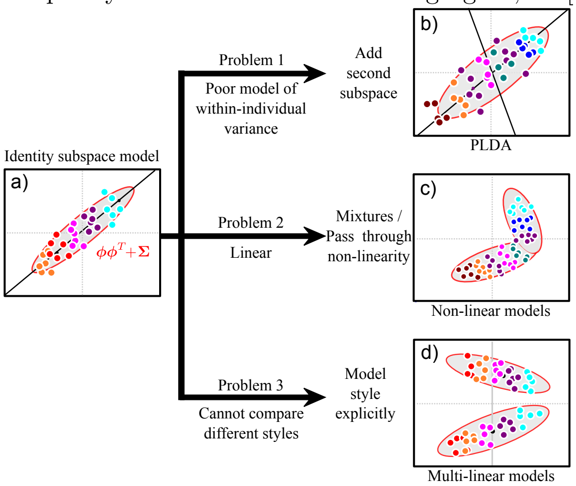
where $\boldsymbol{x_i}$ is a data example, $\boldsymbol{\mu}$ the overall data mean, $\boldsymbol{\Sigma}$ the residual covariance, $\boldsymbol{\Phi} = [\phi_1, \dots, \phi_K]$ contains K factors in its columns, modeling the subspace of the hidden variable $\boldsymbol{h_i}$, and residuals $\boldsymbol{\epsilon_i}$. Factor analysis models are trained by the expectation maximization (EM) algoirthm using training data $\{\boldsymbol{x_i}\}_{i=1}^N$. In the E-step, the posterior distribution $\Pr(\boldsymbol{h_i}\,|\,\boldsymbol{x_i})$ is estimated:

$$\begin{aligned} \Pr(\boldsymbol{h_i}\,|\,\boldsymbol{x_i}) &= \mathcal{N}\left(\boldsymbol{L^{-1}}\,\boldsymbol{\Phi}'\,\boldsymbol{\Sigma}^{-1}\,(\boldsymbol{x_i} - \boldsymbol{\mu}), \boldsymbol{L^{-1}}\right), \\ \boldsymbol{L} &= \boldsymbol{I} + \boldsymbol{\Phi}'\,\boldsymbol{\Sigma}^{-1}\,\boldsymbol{\Phi}. \end{aligned} \tag{2}$$

In the M-step, parameter updates $\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Sigma}}$ are computed based on zero, first and second order moments $N, \mathrm{E}[\boldsymbol{h_i}], \mathrm{E}[\boldsymbol{h_i}\,\boldsymbol{h_i'}]$ of the estimated posterior distribution in order to maximize the overall model fit:

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x_i}, \\ \bar{\boldsymbol{x}}_{\boldsymbol{i}} &= \boldsymbol{x_i} - \hat{\boldsymbol{\mu}}, \textit{ as centralized sample}, \\ \hat{\boldsymbol{\Phi}} &= \left(\sum_{i=1}^{N}\bar{\boldsymbol{x}}_{\boldsymbol{i}}\;\mathrm{E}[\boldsymbol{h_i}]'\right)\left(\sum_{i=1}^{N}\mathrm{E}[\boldsymbol{h_i}\,\boldsymbol{h_i'}]\right)^{-1}, \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{N}\sum_{i=1}^{N}\operatorname{diag}\left[\bar{\boldsymbol{x}}_{\boldsymbol{i}}\,\bar{\boldsymbol{x}}_{\boldsymbol{i}}{}' - \boldsymbol{\Phi}\;\mathrm{E}[\boldsymbol{h_i}]\,\bar{\boldsymbol{x}}_{\boldsymbol{i}}{}'\right]. \end{aligned} \tag{3}$$

In the following, the subspace identity and the original PLDA model are discussed. Non- and multi-linear models, depicted in [1], will only be briefy addressed, since they assume either distribution mixtures (non-linear) or too large style variations (multi-linear). The speaker recognition community addresses so far purely *Problem 1* of the following figure, cf. [1].



2.1 Subspace Identity Model

The subspace identity model [1] extends the factor analysis by addressing within-subject variabilities among samples J_i among subjects I, such that $N = \sum_{i=1}^{I} J_i$. The generative model is depicted regarding $\boldsymbol{x_{ij}}$:

$$\begin{aligned} \boldsymbol{x_{ij}} &= \boldsymbol{\mu} + \boldsymbol{\Phi}\,\boldsymbol{h_i} + \boldsymbol{\epsilon_{ij}}, \\ \Pr(\boldsymbol{x_{ij}}\,|\,\boldsymbol{h_i}) &= \mathcal{N}\left(\boldsymbol{\mu} + \boldsymbol{\Phi}\,\boldsymbol{h_i}, \boldsymbol{\Sigma}\right), \\ \boldsymbol{h_i} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}\right), \\ \boldsymbol{\epsilon_{ij}} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}\right), \textit{with: } \operatorname{diag}[\boldsymbol{\Sigma}], \\ \boldsymbol{x_{ij}} &\sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Phi}\,\boldsymbol{\Phi}' + \boldsymbol{\Sigma}\right) \end{aligned} \tag{4}$$

where $\boldsymbol{\Phi}\,\boldsymbol{\Phi}'$ corresponds to the between-subject variance, and $\boldsymbol{\Sigma}$ corresponds to the within-subject variance. Posterior distributions are estimated by:

$$\begin{aligned} \boldsymbol{L_i} &= \boldsymbol{I} + J_i\,\boldsymbol{\Phi}'\,\boldsymbol{\Sigma}^{-1}\,\boldsymbol{\Phi}, \\ \mathrm{E}[\boldsymbol{h_i}] &= \boldsymbol{L}^{-1}\,\boldsymbol{\Phi}'\,\boldsymbol{\Sigma}^{-1} \sum_{j=1}^{J_i} (\boldsymbol{x_{ij}} - \boldsymbol{\mu}), \\ \Pr(\boldsymbol{h_i}\,|\,\boldsymbol{x_{i;1,\ldots,J_i}}) &= \mathcal{N}\left(\mathrm{E}[\boldsymbol{h_i}], \boldsymbol{L}^{-1}\right), \\ \mathrm{E}[\boldsymbol{h_i}\,\boldsymbol{h_i'}] &= \boldsymbol{L}^{-1} + \mathrm{E}[\boldsymbol{h_i}]\ \mathrm{E}[\boldsymbol{h_i}]', \end{aligned} \tag{5}$$

and models are updated regarding:

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{J_i} \boldsymbol{x_{ij}}, \\ \boldsymbol{\bar{x}_{ij}} &= \boldsymbol{x_{ij}} - \hat{\boldsymbol{\mu}}, \textit{ as centralized sample}, \\ \boldsymbol{\bar{x}_{i\bullet}} &= [\boldsymbol{\bar{x}_{i1}}', \boldsymbol{\bar{x}_{i2}}', \ldots, \boldsymbol{\bar{x}_{iJ_i}}']', \\ \hat{\boldsymbol{\Phi}} &= \left(\sum_{i=1}^{I} \sum_{j=1}^{J_i} \boldsymbol{\bar{x}_{ij}}\ \mathrm{E}[\boldsymbol{h_i}]' \right) \left(\sum_{i=1}^{N} J_i\ \mathrm{E}[\boldsymbol{h_i}\,\boldsymbol{h_i'}] \right)^{-1}, \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{J_i} \operatorname{diag}\left[\boldsymbol{\bar{x}_{ij}}\,\boldsymbol{\bar{x}_{ij}}' - \hat{\boldsymbol{\Phi}}\ \mathrm{E}[\boldsymbol{h_i}]\,\boldsymbol{\bar{x}_{ij}}' \right]. \end{aligned} \tag{6}$$

Note: the $\hat{\boldsymbol{\Sigma}}$ update can be seen as (a) summing up i.i.d. Gaussians of within-subject variabilities, which (b) are modeled by the difference of the total variability $\boldsymbol{\bar{x}_{ij}}\,\boldsymbol{\bar{x}_{ij}}$ to the updated empirical between-subject variability $\hat{\boldsymbol{\Phi}}\ \mathrm{E}[\boldsymbol{h_i}]\,\boldsymbol{\bar{x}_{ij}}'$, where $\mathrm{E}[\boldsymbol{h_i}]\,\boldsymbol{\bar{x}_{ij}}'$ denotes the empirical between variability for estimating $\hat{\boldsymbol{\Phi}}\,\hat{\boldsymbol{\Phi}}'$ given the data, and (c) averaging all within-subject variabilities in order to achieve one i.i.d. $\hat{\boldsymbol{\Sigma}}$, that is representative for all subjects.

LLR scores S of reference and probe data $\boldsymbol{x_r}, \boldsymbol{x_p}$ are computed by Bayesian inference:

$$\begin{aligned} S(\boldsymbol{x_r}, \boldsymbol{x_p}) &= \log \mathcal{N}\left(\begin{bmatrix} \boldsymbol{x_r} \\ \boldsymbol{x_p} \end{bmatrix} \middle| \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\text{tot}} & \boldsymbol{\Phi}\,\boldsymbol{\Phi}' \\ \boldsymbol{\Phi}\,\boldsymbol{\Phi}' & \boldsymbol{\Sigma}_{\text{tot}} \end{bmatrix} \right) \\ &\quad - \log \mathcal{N}\left(\begin{bmatrix} \boldsymbol{x_r} \\ \boldsymbol{x_p} \end{bmatrix} \middle| \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\text{tot}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{\text{tot}} \end{bmatrix} \right), \\ \boldsymbol{\Sigma}_{\text{tot}} &= \boldsymbol{\Phi}\,\boldsymbol{\Phi}' + \boldsymbol{\Sigma}. \end{aligned} \tag{7}$$

2.2 Original PLDA

The original PLDA [1] assumes further style (channel) influences, which are interpreted as additive Gaussian random variables $\boldsymbol{s_{ij}}$ that smoothly result in an additive manner to the generative model:

$$\begin{aligned} \boldsymbol{x_{ij}} &= \boldsymbol{\mu} + \boldsymbol{\Phi}\,\boldsymbol{h_i} + \boldsymbol{\Psi}\,\boldsymbol{s_{ij}} + \boldsymbol{\epsilon_{ij}}, \\ \Pr(\boldsymbol{x_{ij}}\,|\,\boldsymbol{h_i}, \boldsymbol{s_{ij}}) &= \mathcal{N}\left(\boldsymbol{\mu} + \boldsymbol{\Phi}\,\boldsymbol{h_i} + \boldsymbol{\Psi}\,\boldsymbol{s_{ij}}, \boldsymbol{\Sigma}\right), \\ \boldsymbol{h_i} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}\right), \\ \boldsymbol{s_{ij}} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}\right), \\ \boldsymbol{\epsilon_{ij}} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}\right), \textit{with: } \operatorname{diag}[\boldsymbol{\Sigma}], \\ \boldsymbol{x_{ij}} &\sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Phi}\,\boldsymbol{\Phi}' + \boldsymbol{\Sigma}\right). \end{aligned} \tag{8}$$

For the sake of tractable EM equations, PLDA is examined in terms of compound systems, which can easily be substituted in the subspace identity model, cf. 2.1. The compound generative system for the E-step can be expressed as:

$$\begin{aligned} \begin{bmatrix} \boldsymbol{x_{i1}} \\ \boldsymbol{x_{i2}} \\ \vdots \\ \boldsymbol{x_{iJ_i}} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{\Psi} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{\Phi} & \boldsymbol{0} & \boldsymbol{\Psi} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \boldsymbol{0} \\ \boldsymbol{\Phi} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Psi} \end{bmatrix} \begin{bmatrix} \boldsymbol{h_i} \\ \boldsymbol{s_{i1}} \\ \vdots \\ \boldsymbol{s_{iJ_i}} \end{bmatrix} \\ &\quad + \begin{bmatrix} \boldsymbol{\epsilon_{i1}} \\ \boldsymbol{\epsilon_{i2}} \\ \vdots \\ \boldsymbol{\epsilon_{iJ_i}} \end{bmatrix}, \\ \boldsymbol{x_{i\bullet}} &= \boldsymbol{\mu}^* + \boldsymbol{\Phi}^*\,\boldsymbol{h}_i^* + \boldsymbol{\epsilon_{i\bullet}}, \\ \boldsymbol{\Sigma}^* &= \operatorname{diag}[\boldsymbol{\Sigma}, \boldsymbol{\Sigma}, \ldots, \boldsymbol{\Sigma}], \\ \boldsymbol{L}_i^* &= \boldsymbol{I} + \boldsymbol{\Phi}^{*\prime}\,\boldsymbol{\Sigma}^{*-1}\,\boldsymbol{\Phi}^*, \boldsymbol{L}_i^* \textit{ contains factor } J_i, \\ \mathrm{E}[\boldsymbol{h}_i^*] &= \boldsymbol{L}_i^{-1}\,\boldsymbol{\Phi}^{*\prime}\,\boldsymbol{\Sigma}^{*-1}\,\boldsymbol{\bar{x}_{i\bullet}}, \\ \mathrm{E}[\boldsymbol{h}_i^*\,\boldsymbol{h}_i^{*\prime}] &= \boldsymbol{L}_i^{-1} + \mathrm{E}[\boldsymbol{h}_i^*]\ \mathrm{E}[\boldsymbol{h}_i^*]', \end{aligned} \tag{9}$$

where single $\mathrm{E}[\boldsymbol{h}_{ij}^*], \mathrm{E}[\boldsymbol{h}_{ij}^*\,\boldsymbol{h}_{ij}^{*\prime}]$ estimates can be obtained either by sub-indexing, or by single computation. Note: the complete data of a subject is utilized in order to estimate the depending posterior distribution $\boldsymbol{h}_i^*$, then sample-depending posterior estimates are derived.

In the M-step, the compound is formulated as [1]:

$$\begin{aligned}
\boldsymbol{x}_{ij} &= \boldsymbol{\mu} + [\boldsymbol{\Phi} \quad \boldsymbol{\Psi}] \begin{bmatrix} \boldsymbol{h}_i \\ \boldsymbol{s}_{ij} \end{bmatrix} + \boldsymbol{\epsilon}_{ij}, \qquad (10)\\
\boldsymbol{x}_{ij} &= \boldsymbol{\mu} + \boldsymbol{\Phi}^{**}\, \boldsymbol{h}_i^{**} + \boldsymbol{\epsilon}_{ij},\\
\begin{bmatrix} \hat{\boldsymbol{\Phi}} & \boldsymbol{Q} \\ \boldsymbol{Q}' & \hat{\boldsymbol{\Psi}} \end{bmatrix} &= \hat{\boldsymbol{\Phi}}^{**}.\\
\hat{\boldsymbol{\Sigma}} &= \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{J_i} \mathrm{diag}[\bar{\boldsymbol{x}}_{ij}\, \bar{\boldsymbol{x}}_{ij}{}' - [\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}]\ \mathrm{E}[\boldsymbol{h}_{ij}^{**}]\, \bar{\boldsymbol{x}}_{ij}{}'],
\end{aligned}$$

where the E-step is utilized e.g., $\mathrm{E}[\boldsymbol{h}_{ij}^{**}] = \mathrm{E}[\boldsymbol{h}_{ij}^{*}]$, where $\boldsymbol{\Phi}^{**}, \hat{\boldsymbol{\Phi}}^{**}$ represent compound representation suitable for the M-step, and $\boldsymbol{Q}$ as the covariance of $\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{\Psi}}$ can be omitted.

Finally, a compund generative model can be established for LLR scoring as well [1]:

$$\begin{aligned}
\begin{bmatrix} \boldsymbol{x}_r \\ \boldsymbol{x}_p \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{\Psi} & \boldsymbol{0} \\ \boldsymbol{\Phi} & \boldsymbol{0} & \boldsymbol{\Psi} \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_{r,p} \\ \boldsymbol{s}_r \\ \boldsymbol{s}_p \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_r \\ \boldsymbol{\epsilon}_p \end{bmatrix},\\
\boldsymbol{x}_i^{***} &= \boldsymbol{\mu}^{***} + \boldsymbol{\Phi}^{***}\, \boldsymbol{h}_i^{***} + \boldsymbol{\epsilon}_{ij}^{***}. \qquad (11)
\end{aligned}$$

2.3 G-PLDA

In speaker recognition [2, 4], eq. (8) is revisited w.r.t. (a) the subject-specific term $\boldsymbol{B}_i = \boldsymbol{\mu} + \boldsymbol{\Phi}\, \boldsymbol{h}_i$ does not depend on a particular sample, and (b) the sample-depending term $\boldsymbol{W}_{ij} = \boldsymbol{\Psi}\, \boldsymbol{s}_{ij} + \boldsymbol{\epsilon}_{ij}$ describes the within-subject variability, where all latent variables $\boldsymbol{h}, \boldsymbol{s}, \boldsymbol{\epsilon}$ are assumed to be statistically independent. Eq. (8) can be re-formulated w.r.t. $\boldsymbol{B}_i, \boldsymbol{W}_{ij}$:

$$\begin{aligned}
\boldsymbol{x}_{ij} &= \boldsymbol{B}_i + \boldsymbol{W}_{ij}, \qquad (12)\\
\boldsymbol{B}_i &\sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Phi}\, \boldsymbol{\Phi}'\right),\\
\boldsymbol{W}_{ij} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Psi}\, \boldsymbol{\Psi}' + \boldsymbol{\Sigma}\right), \textit{with: } \mathrm{diag}[\boldsymbol{\Sigma}].
\end{aligned}$$

In G-PLDA [4], eq. (8) is simplified to:

$$\begin{aligned}
\boldsymbol{x}_{ij} &= \boldsymbol{\mu} + \boldsymbol{\Phi}\, \boldsymbol{h}_i + \boldsymbol{\epsilon}_{ij}, \qquad (13)\\
\boldsymbol{h}_i &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}\right),\\
\boldsymbol{\epsilon}_{ij} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}\right), \textit{with: } \mathrm{full}[\boldsymbol{\Sigma}],\\
\boldsymbol{\Sigma} &= \boldsymbol{\Psi}\, \boldsymbol{\Psi}' + \boldsymbol{\varphi}^{-1}, \textit{with: } \mathrm{diag}[\boldsymbol{\varphi}^{-1}],
\end{aligned}$$

where $\boldsymbol{\varphi}^{-1}$ represents the covariance [1] of $\boldsymbol{\epsilon}_{ij}$ in eq. (8). Regarding to the equations of the subspace identity model, effectively only the M-step update $\hat{\boldsymbol{\Sigma}}$ in eq. (6) needs to be adapted from diagonal to full covaraince matrix estimation [4]:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{J_i} \bar{\boldsymbol{x}}_{ij}\, \bar{\boldsymbol{x}}_{ij}{}' - \hat{\boldsymbol{\Phi}}\ \mathrm{E}[\boldsymbol{h}_i]\, \bar{\boldsymbol{x}}_{ij}{}'. \qquad (14)$$

Further, in [6, 4, 9], a *minimum-divergence* (MD) step is applied in order to avoid saddle points. MD assumes $\boldsymbol{h}_i, \boldsymbol{s}_{ij}$ priors *could* be in a non-standard Gaussian form, maximizes w.r.t. its parameters and finds standard prior equivalent representations. In [4], the MD-step is conducted as:

$$\begin{aligned}
\boldsymbol{R} &= \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{J_i} \mathrm{E}[\boldsymbol{h}_{ij}\, \boldsymbol{h}_{ij}] \qquad (15)\\
\hat{\boldsymbol{\Phi}}_{\mathrm{MD}} &= \hat{\boldsymbol{\Phi}}\ \mathrm{chol}(\boldsymbol{R}).
\end{aligned}$$

2.4 2-Cov

In 2-Cov [3], a simple linear-Gaussian generative model is adopted with between- and within-subject covariance matrices $\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{W}}$, which have the same dimensionality as the feature vector $\boldsymbol{x}_{ij}$ [6]:

$$\begin{aligned}
\boldsymbol{x}_{ij} &= \boldsymbol{\mu} + \boldsymbol{h}_i + \boldsymbol{\epsilon}_t, \qquad (16)\\
\boldsymbol{h}_i &\sim \mathcal{N}\left(\boldsymbol{h}_i \,|\, \boldsymbol{\mu}, \boldsymbol{\mathcal{B}}\right), \textit{with: } \mathrm{full}[\boldsymbol{B}],\\
\boldsymbol{x}_{ij} \,|\, \boldsymbol{h}_i &= \mathcal{N}\left(\boldsymbol{x}_{ij} \,|\, \boldsymbol{h}_i, \boldsymbol{\mathcal{W}}\right), \textit{with: } \mathrm{full}[\boldsymbol{W}],\\
\boldsymbol{\mathcal{B}} &= \sum_{i=1}^{I} \frac{J_i}{N}\ (\boldsymbol{h}_i - \boldsymbol{\mu})\ (\boldsymbol{h}_i - \boldsymbol{\mu})',\\
\boldsymbol{\mathcal{W}} &= \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{J_i} (x_{ij} - \boldsymbol{h}_i)\ (x_{ij} - \boldsymbol{h}_i)'.
\end{aligned}$$

LLRs are computed by [3, 5, 10]:

$$\begin{aligned}
S(\boldsymbol{x}_r, \boldsymbol{x}_p) &= \bar{\boldsymbol{x}}_r'\, \boldsymbol{\mathcal{P}}\, \bar{\boldsymbol{x}}_p + \frac{1}{2}\left(\bar{\boldsymbol{x}}_r'\, \boldsymbol{\mathcal{Q}}\, \bar{\boldsymbol{x}}_r + \bar{\boldsymbol{x}}_p'\, \boldsymbol{\mathcal{Q}}\, \bar{\boldsymbol{x}}_p\right),\\
\boldsymbol{\mathcal{P}} &= \boldsymbol{\mathcal{W}}^{-1}\left(2\,\boldsymbol{\mathcal{W}}^{-1} + \boldsymbol{\mathcal{B}}^{-1}\right)\boldsymbol{\mathcal{W}}^{-1}, \qquad (17)\\
\boldsymbol{\mathcal{Q}} &= \boldsymbol{\mathcal{W}}^{-1}\left(2\,\boldsymbol{\mathcal{W}}^{-1} + \boldsymbol{\mathcal{B}}^{-1}\right)\boldsymbol{\mathcal{W}}^{-1}\\
&\quad - \boldsymbol{\mathcal{W}}^{-1}\left(\boldsymbol{\mathcal{W}}^{-1} + \boldsymbol{\mathcal{B}}^{-1}\right)\boldsymbol{\mathcal{W}}^{-1}.
\end{aligned}$$

As depicted in [3, 5], the 2-Cov and the original PLDA models share the same structure, they only differ in their covariance matrices:

$$\begin{aligned}
\boldsymbol{B}_{\textit{orignal PLDA}} &= \left(\boldsymbol{\Phi}\, \boldsymbol{\Phi}'\right)^{-1}, \qquad (18)\\
\boldsymbol{W}_{\textit{orignal PLDA}} &= \left(\boldsymbol{\Psi}\, \boldsymbol{\Psi}' + \boldsymbol{\Sigma}\right)^{-1}, \textit{with: } \mathrm{diag}[\Sigma],\\
\boldsymbol{B}_{\textit{G-PLDA}} &= \left(\boldsymbol{\Phi}\, \boldsymbol{\Phi}'\right)^{-1},\\
\boldsymbol{W}_{\textit{G-PLDA}} &= \boldsymbol{\Sigma}^{-1}, \textit{with: } \mathrm{full}[\Sigma],
\end{aligned}$$

A straight-forward parameter estimation with source code are provided by [6] for the original PLDA, G-PLDA and 2-Cov. However, the 2-Cov scoring seems odd comparing it to [10], and furthermore, the G-PLDA results differ from the implementation in [4].

2.5 Other PLDA flavors

This section briefly depicts generative model ideas behind non-linear models, multi-linear models, 2-Gau and PSVM.

A. Non-linear models [1] assume categories $c_k \in [1, \ldots, C]$ as hidden latent variables that define clusters, where each cluster has different parameters:

$$\begin{aligned}
\Pr(\boldsymbol{x}_{ij} \,|\, c_k, h_i, s_{ij}) &= \mathcal{N}\left(\boldsymbol{m}_{c_k}, \boldsymbol{\Sigma}_{c_k}\right), \qquad (19)\\
\boldsymbol{m}_{c_k} &= \boldsymbol{\mu}_{c_k} + \boldsymbol{\Phi}_{c_k}\, \boldsymbol{h}_i + \boldsymbol{\Psi}_{c_k}\, \boldsymbol{s}_{sj}.
\end{aligned}$$

[1] $\boldsymbol{\varphi}^{-1}$ is a covariance matrix, and $\boldsymbol{\varphi}$ is a precision matrix.

In [11], a non-linear model, referred to as *mixture PLDA*, was examined in noisy environments, conducting mixtures SNR-dependingly.

B. Multi-linear models [1] are referred to be not marginalizable for all hidden random variables e.g., a generative equation for three latent variables $\boldsymbol{h}, \boldsymbol{s}, \boldsymbol{t}$:

$$\boldsymbol{x_{ijkl}} = \boldsymbol{\mu} + \boldsymbol{\Phi} \times_2 \boldsymbol{h_i} \times_3 \boldsymbol{s_k} \times_4 \boldsymbol{t_l} + \boldsymbol{\epsilon_{ijkl}}, \tag{20}$$

where the dot product notations $\times_{2,3,4}$ take the dot product of the second, third, fourth dimensions of $\boldsymbol{\Phi}$ w.r.t. $\boldsymbol{h}, \boldsymbol{s}, \boldsymbol{t}$. However, *bi-linear* models are referred to obtain closed-form solutions [1].

C. HT-PLDA [2] are motivated due non-Gaussian effects observed on microphone data. Therefore, the original PLDA priors $\boldsymbol{h_i}, \boldsymbol{s_{ij}}, \boldsymbol{\epsilon_{ij}}$ were assumed to be t-distributed rather than standard normal:

$$\begin{aligned} \boldsymbol{h_i} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{u^{-1}}\,\boldsymbol{I}\right), \textit{where } \boldsymbol{u} \sim \Gamma\left(\frac{n_1}{2}, \frac{n_1}{2}\right), \\ \boldsymbol{s_{ij}} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{v^{-1}}\,\boldsymbol{I}\right), \textit{where } \boldsymbol{v} \sim \Gamma\left(\frac{n_2}{2}, \frac{n_2}{2}\right), \\ \boldsymbol{\epsilon_{ij}} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{w^{-1}}\,\boldsymbol{I}\right), \textit{where } \boldsymbol{w} \sim \Gamma\left(\frac{n_3}{2}, \frac{n_3}{2}\right), \end{aligned} \tag{21}$$

with degrees-of-freedom n_1, n_2, n_3. However, RG (length-normalization) [4] forced i-vector features to be Gaussian distributed, no matter the channel, motivating G-PLDA as state-of-the-art speaker recognition.

D. 2-Gau [5] aims at modeling subject and channel distribution components of i-vector features directly. Thereby, one model represents the null hypothesis H_0 : {same subject} and the other model represents the alternative hypothesis H_A : {different subject}, which one may also interprete as between- and within-subject variability models:

$$\begin{aligned} \begin{bmatrix} \boldsymbol{x_r} \\ \boldsymbol{x_p} \end{bmatrix}_{H_0} &\sim \mathcal{N}\left(\boldsymbol{\mu_{H_0}}, \boldsymbol{\mathcal{B}}\right), \\ \begin{bmatrix} \boldsymbol{x_r} \\ \boldsymbol{x_p} \end{bmatrix}_{H_A} &\sim \mathcal{N}\left(\boldsymbol{\mu_{H_A}}, \boldsymbol{\mathcal{W}}\right). \end{aligned} \tag{22}$$

E. PSVM [5] is trained to discriminate between H_0, H_A (reference, probe) pairs, which is in contrast to the conventional one-vs-all SVM framework. PSVM belongs to the PLDA family in so far as its score of a second order Taylor expansion of an i-vector pair can be formulated in a way leading to the 2-Cov scoring equation [5].

3 FP-PLDA and i-vectors

State-of-the-art speaker recognition systems comprise i-vector features [12], which are latent posterior variables, being also extracted by factor analysis techniques [12, 13]. In order to account for uncertainties during feature extraction, i.e. the estimation of a sample-depending i-vector posterior distribution, FP-PLDA is motivated [7, 8].

3.1 Estimation of i-vectors (brief)

An i-vector $\vec{\boldsymbol{i}}$ is a compact representation of a voice sample, which depends on an underlying statistical cluster of an acoustical space[2]. By observing acoustic features $\boldsymbol{\mathcal{X}}$ on each cluster c, zero and centered first order moments $\boldsymbol{N_{\mathcal{X}}^{(c)}}, \boldsymbol{F_{\mathcal{X}}^{(c)}}$ known as Baum-Welch statistics can be estimated. The i-vector extractor is a factor analysis model $(\boldsymbol{T}, \boldsymbol{\Sigma})$ for Baum-Welch statistics [12, 13, 14, 15, 16] modeling the i-vector posterior distribution $\boldsymbol{X_{\mathcal{X}}}$:

$$\begin{aligned} \boldsymbol{X_{\mathcal{X}}} &\sim \mathcal{N}\left(\vec{\boldsymbol{i}}_{\boldsymbol{\mathcal{X}}}, \boldsymbol{\Gamma_{\mathcal{X}}}\right), \\ \vec{\boldsymbol{i}}_{\boldsymbol{\mathcal{X}}} &= \boldsymbol{\Gamma_{\mathcal{X}}}\,\boldsymbol{T'}\,\boldsymbol{\Sigma^{-1}}\,\boldsymbol{F_{\mathcal{X}}}, \\ \boldsymbol{\Gamma_{\mathcal{X}}^{-1}} &= \boldsymbol{I} + \sum_{c=1}^{C} \boldsymbol{N_{\mathcal{X}}^{(c)}}\,\boldsymbol{T^{(c)\prime}}\,\boldsymbol{\Sigma^{(c)-1}}\,\boldsymbol{T^{(c)}}. \end{aligned} \tag{23}$$

The conventional PLDA models subspaces of the single-point i-vector estimates $\vec{\boldsymbol{i}}_{\boldsymbol{\mathcal{X}}}$. FP-PLDA models uncertainty estimates $\boldsymbol{\Gamma_{\mathcal{X}}}$ as additive latent variables to the sample-depending term $\boldsymbol{W_{i\mathcal{X}}}$ [15, 16].

However, conventionally, i-vectors are processed in order to achieve a more discriminative feature space by applying LDA and WCCN. Furthermore by RG, i-vectors are projected onto a unit sphere. In order to account for this transformations, $\boldsymbol{X_{\mathcal{X}}}$ needs to be projected as well [17]:

- LDA with transformation matrix[3] $\boldsymbol{L}$:

  $$\boldsymbol{X_{\mathcal{X}}} \sim \mathcal{N}(\boldsymbol{L}\,\vec{\boldsymbol{i}}_{\boldsymbol{\mathcal{X}}}, \boldsymbol{L}\,\boldsymbol{\Gamma_{\mathcal{X}}}\,\boldsymbol{L'}), \tag{24}$$

  Note: in [18], an LDA variation is proposed that also takes the i-vector uncertainty $\boldsymbol{\Gamma_{\mathcal{X}}}$ into account (ULDA),

- WCCN with mean centering $\boldsymbol{m}$ and whitening matrix $\boldsymbol{W}$:

  $$\boldsymbol{X_{\mathcal{X}}} \sim \mathcal{N}(\boldsymbol{B}\,(\boldsymbol{L}\,\vec{\boldsymbol{i}}_{\boldsymbol{\mathcal{X}}} - \boldsymbol{m}), \boldsymbol{B}\,\boldsymbol{L}\,\boldsymbol{\Gamma_{\mathcal{X}}}\,\boldsymbol{L'}\,\boldsymbol{B'}), \tag{25}$$

- RG is a first-order Taylor series expansion [7], passing the uncertainties through the processing, which can be achieved either by the simple length-normalization (LN):

  $$\begin{aligned} \boldsymbol{X_{\mathcal{X}}} &\sim \mathcal{N}(\boldsymbol{x_{ij}}, \boldsymbol{\Gamma_{ij}}), \\ l_{\mathcal{X}} &= \|\boldsymbol{B}\,(\boldsymbol{L}\,\vec{\boldsymbol{i}}_{\boldsymbol{\mathcal{X}}} - \boldsymbol{m})\|, \\ \boldsymbol{x_{ij}} &= \frac{\boldsymbol{B}\,(\boldsymbol{L}\,\vec{\boldsymbol{i}}_{\boldsymbol{\mathcal{X}}} - \boldsymbol{m})}{l_{\mathcal{X}}}, \\ \boldsymbol{\Gamma_{ij}} &= \frac{\boldsymbol{B}\,\boldsymbol{L}\,\boldsymbol{\Gamma_{\mathcal{X}}}\,\boldsymbol{L'}\,\boldsymbol{B'}}{l_{\mathcal{X}}^2}, \end{aligned} \tag{26}$$

[2] Conventionally, a GMM which is referred to as UBM.

[3] Usually, LDA performs a mean subtraction as well. However, since the i-vector extractor models i-vectors to be 0-centered in its factor analysis model, mean-subtraction is omitted in this framework.

or by the first-order approximation, referred to as projected LN (PLN):

$$\boldsymbol{\Gamma}_{ij} = \frac{(\boldsymbol{I} - \boldsymbol{x}_{ij}\,\boldsymbol{x}'_{ij})\,\boldsymbol{B}\,\boldsymbol{L}\,\boldsymbol{\Gamma}_{\boldsymbol{\mathcal{X}}}\,\boldsymbol{L}'\,\boldsymbol{B}'\,(\boldsymbol{I} - \boldsymbol{x}_{ij}\,\boldsymbol{x}'_{ij})}{l^2_{\mathcal{X}}}. \tag{27}$$

3.2 FP-PLDA model

In FP-PLDA [7, 8], uncertainty is propagated in terms of additive noise to the prior residual distribution[4], reformulating eq. (13):

$$\begin{aligned} \boldsymbol{x}_{ij} &= \boldsymbol{\mu} + \boldsymbol{\Phi}\,\boldsymbol{h}_i + \bar{\boldsymbol{\epsilon}}_{ij}, \\ \bar{\boldsymbol{\epsilon}}_{ij} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma} + \boldsymbol{\Gamma}_{ij}\right) \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Lambda}_{ij}\right). \end{aligned} \tag{28}$$

Note that in this section, the indexing ij purely represents an indexing of a certain sample j in the sample set i, without the intent to denote any subject-dependencies, i.e. this notation is kept for comparison tractability to the conventional PLDA model. Indexed subject-dependencies are only addressed by $\boldsymbol{h}_i$.

3.2.1 Adapting the EM Training

As for the original PLDA and G-PLDA, compound systems are formulated regarding EM-steps. In the E-step, the Cholesky decomposite of the posterior covariance $\boldsymbol{\Gamma}_{ij} = \boldsymbol{C}_{ij}\,\boldsymbol{C}'_{ij}$ is utilized as an additional channel component $\boldsymbol{s}^*_{ij}$ with $\boldsymbol{0}$-mean [8]:

$$\begin{aligned} \boldsymbol{x}_{ij} &= \boldsymbol{\mu} + \boldsymbol{\Phi}\,\boldsymbol{h}_i + \boldsymbol{C}_{ij}\,\boldsymbol{s}^*_{ij} + \bar{\boldsymbol{\epsilon}}^*_{ij}, \\ \boldsymbol{s}^*_{ij} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Gamma}_{ij}\right), \\ \bar{\boldsymbol{\epsilon}}^*_{ij} &\sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}\right), \textit{with: } \text{full}[\boldsymbol{\Sigma}], \end{aligned} \tag{29}$$

which appears to match the formulation of the original PLDA with sample-dependent channel factors. Model consistency is contained due to:

$$\begin{aligned} \boldsymbol{x}_{ij} &\sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Phi}\,\boldsymbol{\Phi}' + \boldsymbol{C}_{ij}\,\boldsymbol{C}'_{ij} + \boldsymbol{\Sigma}\right), \\ &\sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Phi}\,\boldsymbol{\Phi}' + \boldsymbol{\Gamma}_{ij} + \boldsymbol{\Sigma}\right), \\ &\sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Phi}\,\boldsymbol{\Phi}' + \boldsymbol{\Lambda}_{ij}\right). \end{aligned} \tag{30}$$

In analogue to eq. (9), the E-step compund system is expressed by [8]:

$$\begin{bmatrix} \boldsymbol{x}_{i1} \\ \boldsymbol{x}_{i2} \\ \vdots \\ \boldsymbol{x}_{iJ_i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{C}_{i1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{\Phi} & \boldsymbol{0} & \boldsymbol{C}_{i2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \boldsymbol{0} \\ \boldsymbol{\Phi} & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{C}_{iJ_i} \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_i \\ \boldsymbol{s}^*_{i1} \\ \vdots \\ \boldsymbol{s}^*_{iJ_i} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\Sigma} \end{bmatrix}, \textit{with: } \text{full}[\boldsymbol{\Sigma}]. \tag{31}$$

The compound system for the M-step can be also expressed in terms of the original PLDA formulations in eq. (10):

$$\boldsymbol{x}_{ij} = \boldsymbol{\mu} + [\boldsymbol{\Phi} \quad \boldsymbol{C}_{ij}] \begin{bmatrix} \boldsymbol{h}_i \\ \boldsymbol{s}_{ij} \end{bmatrix} + \boldsymbol{\epsilon}_{ij}. \tag{32}$$

Compared to the original PLDA and G-PLDA, FP-PLDA requires to process the E-steps sample-wise, i.e.:

- $\boldsymbol{L}_i^{-1}$ is conducted from all posterior covariances,
- $\mathrm{E}[\boldsymbol{h}^*_{ij}], \mathrm{E}[\boldsymbol{h}^*_i]$ are estimated,
- $\mathrm{E}[\boldsymbol{h}^*_{ij}\,\boldsymbol{h}^{*\prime}_{ij}], \mathrm{E}[\boldsymbol{h}^*_i\,\boldsymbol{h}^{*\prime}_i]$.

3.2.2 Updating the FP-PLDA model

Model updates are derived with MD-step using the equations depicted in section 2.3. Note: the terms are stated w.r.t. sample-wise processing, but usually they are optimized.

3.2.3 Sample-wise LLR estimations

Contrary to G-PLDA, FP-PLDA LLR scoring cannot be completely optimized, i.e. scores are computed individually w.r.t. to depending posterior distributions (covariances). In [16], the LLR scoring for FP-PLDA is referred to regarding a set of reference samples $\mathcal{R}$ and a set of probe samples $\mathcal{P}$ with $\mathcal{G} = \mathcal{R} \cup \mathcal{P}$, and ϕ, γ denoting the dimensionalities of the PLDA sub-space $\boldsymbol{\Phi}$ and the posterior covariance $\boldsymbol{\Gamma}$, respectively:

$$\begin{aligned} S(\mathcal{R}, \mathcal{P}) &= \sigma\left(\mathcal{R}, \mathcal{P}\right) - \sigma\left(\mathcal{G}\right) + \frac{\phi}{2}\,\log 2\,\pi, \\ \sigma\left(\mathcal{G}\right) &= -\frac{1}{2}\,\log|\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{G}}}| + \frac{1}{2}\,\boldsymbol{\mu}'_{\boldsymbol{\mathcal{G}}}\,\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{G}}}\,\boldsymbol{\mu}_{\boldsymbol{\mathcal{G}}}, \\ \boldsymbol{\Lambda}_{\boldsymbol{\mathcal{S}} \in \{\boldsymbol{\mathcal{G}}, \boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{P}}\}} &= \boldsymbol{I} + \boldsymbol{\Phi}' \left(\sum_{s \in \mathcal{S}} \boldsymbol{\Lambda}_s^{-1}\right) \boldsymbol{\Phi}, \\ \boldsymbol{\mu}_{\boldsymbol{\mathcal{G}}} &= \boldsymbol{\Lambda}_{\boldsymbol{\mathcal{G}}}^{-1}\,\boldsymbol{\Phi}' \sum_{g \in \mathcal{G}} \boldsymbol{\Lambda}_g^{-1}\,\bar{\boldsymbol{x}}_g, \\ \sigma\left(\mathcal{R}, \mathcal{P}\right) &= -\frac{1}{2}\,\log|\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{R}},\boldsymbol{\mathcal{P}}}| + \boldsymbol{F}'_{\boldsymbol{\mathcal{R}}}\,\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{R}},\boldsymbol{\mathcal{P}}}\,\boldsymbol{F}_{\boldsymbol{\mathcal{P}}} \\ &\quad + \frac{1}{2}\left(\boldsymbol{F}'_{\boldsymbol{\mathcal{R}}}\,\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{R}},\boldsymbol{\mathcal{P}}}\,\boldsymbol{F}_{\boldsymbol{\mathcal{R}}} + \boldsymbol{F}'_{\boldsymbol{\mathcal{P}}}\,\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{R}},\boldsymbol{\mathcal{P}}}\,\boldsymbol{F}_{\boldsymbol{\mathcal{P}}}\right), \\ \boldsymbol{\Lambda}_{\boldsymbol{\mathcal{R}},\boldsymbol{\mathcal{P}}} &= \left(\boldsymbol{I} + \boldsymbol{\Lambda}_{\boldsymbol{\mathcal{R}}} + \boldsymbol{\Lambda}_{\boldsymbol{\mathcal{P}}}\right)^{-1}, \\ \boldsymbol{F}_{\boldsymbol{\mathcal{G}}} &= \boldsymbol{\Phi}' \sum_{g \in \mathcal{G}} \boldsymbol{\Lambda}_g\,\bar{\boldsymbol{x}}_g. \end{aligned} \tag{33}$$

Since samples cannot be processed at once as in (optimized) G-PLDA, one may consider system complexity as a performance constraint for optimization. In [16], complexities of G-PLDA and FP-PLDA are compared:

Comparator	Complexity per sample	probe	comparison
G-PLDA	γ	$\gamma\,\phi$	ϕ
FP-PLDA	γ^3	$\gamma^2\,\phi$	ϕ^3

[4]In literature, usually precision matrices are referred to rather than covariance matrices. For the sake of easier comprehension, this report refers to covariance matrices.

In order to achieve lower FP-PLDA complexity, *approximated FP-PLDA* is motivated [16]. Note: in original and G-PLDA optimizations were possible regarding how the sample sets are processed, which due to the inclusion of the posterior covariances is not applicable in FP-PLDA, such that other complexity optimization strategies are sought. However, by denoting reference distributions as certain, uncertainty propagation needs only to be addressed regarding probe samples. This method is referred to as *asymmetric FP-PLDA* [15] or *uncertainty decoding* (UD) [17].

3.3 Approximating FP-PLDA

In [16], three diagonalizations are proposed: i-vector posterior $\hat{\boldsymbol{\Gamma}}_{\boldsymbol{ij}}$, residual covariance $\hat{\boldsymbol{\Lambda}}_{\boldsymbol{ij}}$, speaker identity posteriors $\hat{\boldsymbol{D}}_{\boldsymbol{\mathcal{R}}}, \hat{\boldsymbol{D}}_{\boldsymbol{\mathcal{P}}}$. By combining all of them, the FP-PLDA complexity can be reduced to $\left(\gamma^2, \gamma\,\phi, \phi\right)$.

3.3.1 Diagonalizing i-vector posterior

The i-vector posterior can be approximated in three different ways [16]:

Diagonalizing the posterior covariance matrix

$$\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{X}}} = \boldsymbol{\Gamma}_{\boldsymbol{\mathcal{X}}} \circ \boldsymbol{I}. \tag{34}$$

Diagonalization during posterior composition
Zero-order statistics $\boldsymbol{N}^{(\boldsymbol{c})}_{\boldsymbol{\mathcal{X}}}$ are replaced by the weight of the c-th component w_c, and further lets denote the eigen-decomposition of $\hat{\boldsymbol{\Gamma}}^{-\boldsymbol{1}}_{\boldsymbol{\mathcal{X}}}$ as $\boldsymbol{U}\,\boldsymbol{\Sigma}^{-\boldsymbol{1}}\,\boldsymbol{U}'$:

$$\begin{aligned}
\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{X}}} &= \left(\sum_c w_c\,\boldsymbol{T}^{(\boldsymbol{c})\,\prime}\,\boldsymbol{\Sigma}^{(\boldsymbol{c})\,-\boldsymbol{1}}\,\boldsymbol{T}^{(\boldsymbol{c})}\right)^{-1}, s.t. \qquad (35)\\
\hat{\boldsymbol{w}} &= \boldsymbol{U}'\,\boldsymbol{w}, \quad \hat{\boldsymbol{T}} = \boldsymbol{T}\,\boldsymbol{U}, s.t.\\
\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\mathcal{X}}} &= \left(\boldsymbol{U}'\,\boldsymbol{\Gamma}^{-\boldsymbol{1}}_{\boldsymbol{\mathcal{X}}}\,\boldsymbol{U}\right)^{-1},\\
\hat{\vec{\boldsymbol{i}}}_{\boldsymbol{\mathcal{X}}} &= \boldsymbol{U}'\,\vec{\boldsymbol{i}}_{\boldsymbol{\mathcal{X}}}.
\end{aligned}$$

Diagonalization by Heteroscedastic LDA
Conventionally, LDA targets an optimal discriminativity by seeking a minimal scatter ratio in order to estimate a transformation matrix. Furthermore, Heteroscedastic LDA (HLDA) aims at a maximum likelihood of the feature space, while estimating transformation matrices. Thereby, HLDA performs a diagonalization.

3.3.2 Diagonalizing the Residual Covariance

By eigen-decomposing the G-PLDA precision matrix as $\boldsymbol{\Sigma}^{-\boldsymbol{1}} = \boldsymbol{\Lambda} = \boldsymbol{V}_{\boldsymbol{\Lambda}}\,\boldsymbol{D}_{\boldsymbol{\Lambda}}\,\boldsymbol{V}'_{\boldsymbol{\Lambda}}$, where $\boldsymbol{D}_{\boldsymbol{\Lambda}}$ is diagonal, the residual covariance $\boldsymbol{\Lambda}_{\boldsymbol{ij}}$ can be re-written and approximated by $\hat{\boldsymbol{\Lambda}}_{\boldsymbol{ij}}$ as [16]:

$$\begin{aligned}
\boldsymbol{\Lambda}^{-\boldsymbol{1}}_{\boldsymbol{ij}} &= \left(\boldsymbol{\Lambda}^{-\boldsymbol{1}} + \boldsymbol{\Gamma}_{\boldsymbol{ij}}\right)^{-1} \qquad (36)\\
&= \left(\boldsymbol{V}_{\boldsymbol{\Lambda}}\,\boldsymbol{D}^{-\boldsymbol{1}}_{\boldsymbol{\Lambda}}\,\boldsymbol{V}'_{\boldsymbol{\Lambda}} + \boldsymbol{\Gamma}_{\boldsymbol{ij}}\right)^{-1}\\
&= \boldsymbol{V}_{\boldsymbol{\Lambda}}\left(\boldsymbol{D}^{-\boldsymbol{1}}_{\boldsymbol{\Lambda}} + \boldsymbol{V}'_{\boldsymbol{\Lambda}}\,\boldsymbol{\Gamma}_{\boldsymbol{ij}}\,\boldsymbol{V}_{\boldsymbol{\Lambda}}\right)^{-1}\boldsymbol{V}'_{\boldsymbol{\Lambda}},\\
\boldsymbol{\Lambda}^{\boldsymbol{D}}_{\boldsymbol{ij}} &= \left(\boldsymbol{D}^{-\boldsymbol{1}}_{\boldsymbol{\Lambda}} + \boldsymbol{V}'_{\boldsymbol{\Lambda}}\,\boldsymbol{\Gamma}_{\boldsymbol{ij}}\,\boldsymbol{V}_{\boldsymbol{\Lambda}} \circ \boldsymbol{I}\right)^{-1},\\
\hat{\boldsymbol{\Lambda}}_{\boldsymbol{ij}} &= \boldsymbol{V}_{\boldsymbol{\Lambda}}\,\boldsymbol{\Lambda}^{\boldsymbol{D}}_{\boldsymbol{ij}}\,\boldsymbol{V}'_{\boldsymbol{\Lambda}}, s.t.\\
\hat{\boldsymbol{\Lambda}}^{-\boldsymbol{1}}_{\boldsymbol{\mathcal{R}},\boldsymbol{\mathcal{P}}} &= \boldsymbol{I} + \boldsymbol{\Phi}'\,\boldsymbol{V}_{\boldsymbol{\Lambda}}\left(\sum_{r\in\mathcal{R}}\boldsymbol{\Lambda}^{\boldsymbol{D}}_{\boldsymbol{r}} + \sum_{p\in\mathcal{P}}\boldsymbol{\Lambda}^{\boldsymbol{D}}_{\boldsymbol{p}}\right)\boldsymbol{V}'_{\boldsymbol{\Lambda}}\,\boldsymbol{\Phi}.
\end{aligned}$$

Note: the approximation effects become negligible for longer-duration samples, since the i-vector posterior covariance becomes smaller, recovering the exact PLDA solution [16].

3.3.3 Diagonalizing Speaker Identity Posterior

The speaker identity posterior $\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{R}},\boldsymbol{\mathcal{P}}}$ can be diagonalized in a similar manner by eigen-decomposing the precision term $\boldsymbol{\Phi}'\,\boldsymbol{\Lambda}\,\boldsymbol{\Phi}$ as $\boldsymbol{V}_{\boldsymbol{Y}}\,\boldsymbol{D}_{\boldsymbol{Y}}\,\boldsymbol{V}'_{\boldsymbol{Y}}$, such that the approximation terms $\hat{\boldsymbol{D}}_{\boldsymbol{\mathcal{R}}}, \hat{\boldsymbol{D}}_{\boldsymbol{\mathcal{P}}}$ can be derived [16]:

$$\begin{aligned}
\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{R}},\boldsymbol{\mathcal{P}}} &= \left(\boldsymbol{I} + \boldsymbol{V}_{\boldsymbol{Y}}\,\left(\boldsymbol{D}_{\boldsymbol{\mathcal{R}}} + \boldsymbol{D}_{\boldsymbol{\mathcal{P}}}\right)\,\boldsymbol{V}'_{\boldsymbol{Y}}\right)^{-1}, \qquad (37)\\
&= \boldsymbol{V}_{\boldsymbol{Y}}\,\left(\boldsymbol{I} + \boldsymbol{D}_{\boldsymbol{\mathcal{R}}} + \boldsymbol{D}_{\boldsymbol{\mathcal{P}}}\right)^{-1}\boldsymbol{V}'_{\boldsymbol{Y}},\\
\boldsymbol{D}_{\boldsymbol{\mathcal{R}}} &= \boldsymbol{V}'_{\boldsymbol{Y}}\,\boldsymbol{\Phi}'\,\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{R}}}\,\boldsymbol{\Phi}\,\boldsymbol{V}_{\boldsymbol{Y}},\\
\boldsymbol{D}_{\boldsymbol{\mathcal{P}}} &= \boldsymbol{V}'_{\boldsymbol{Y}}\,\boldsymbol{\Phi}'\,\boldsymbol{\Lambda}_{\boldsymbol{\mathcal{P}}}\,\boldsymbol{\Phi}\,\boldsymbol{V}_{\boldsymbol{Y}},\\
\hat{\boldsymbol{D}}_{\boldsymbol{\mathcal{R}}} &= \boldsymbol{D}_{\boldsymbol{\mathcal{R}}} \circ \boldsymbol{I},\\
\hat{\boldsymbol{D}}_{\boldsymbol{\mathcal{P}}} &= \boldsymbol{D}_{\boldsymbol{\mathcal{P}}} \circ \boldsymbol{I}.
\end{aligned}$$

3.4 Full-Posterior 2-Cov

In [10], UD is introduced to the 2-Cov model, the LLR scoring asymmetrically accounts for the probe uncertainty:

$$\begin{aligned}
S(\boldsymbol{x_r}, \boldsymbol{x_p}) &= \bar{\boldsymbol{x}}'_{\boldsymbol{r}}\,\boldsymbol{\mathcal{P}}\,\boldsymbol{\mu}_{\bar{\boldsymbol{x}}_{\boldsymbol{p}}} + \frac{1}{2}\,\bar{\boldsymbol{x}}'_{\boldsymbol{r}}\,\boldsymbol{\mathcal{Q}}\,\bar{\boldsymbol{x}}_{\boldsymbol{r}} \qquad (38)\\
&\quad + \frac{1}{2}\left(\mathrm{Tr}(\boldsymbol{\Gamma}_{\boldsymbol{p}}\,\boldsymbol{\mathcal{Q}}) + \boldsymbol{\mu}'_{\bar{\boldsymbol{x}}_{\boldsymbol{p}}}\,\boldsymbol{\mathcal{Q}}\,\boldsymbol{\mu}_{\bar{\boldsymbol{x}}_{\boldsymbol{p}}}\right),\\
\boldsymbol{L} &= \boldsymbol{\Sigma}^{-\boldsymbol{1}}_{\boldsymbol{N}} + \boldsymbol{\Sigma}^{-\boldsymbol{1}}_{\boldsymbol{X}}\\
\boldsymbol{\mu}_{\bar{\boldsymbol{x}}_{\boldsymbol{p}}} &= \boldsymbol{L}^{-\boldsymbol{1}}\left(\boldsymbol{\Sigma}^{-\boldsymbol{1}}_{\boldsymbol{N}}\,\left(\boldsymbol{x_p} - \boldsymbol{\mu_N}\right) + \boldsymbol{\Sigma}^{-\boldsymbol{1}}_{\boldsymbol{X}}\,\boldsymbol{\mu_X}\right),
\end{aligned}$$

with total i-vector noise distribution $\mathcal{N}\left(\boldsymbol{\mu_X}, \boldsymbol{\Sigma_X}\right)$ and random noise distribution $\mathcal{N}\left(\boldsymbol{\mu_N}, \boldsymbol{\Sigma_N}\right)$, respectively.

4 Acknowledgements

This work was partially supported by the German Federal Ministry of Education and Research (BMBF) within the Center for Research in Security and Privacy (CRISP), and the Hesse government (project no. 467/15-09, BioMobile).

References

[1] S.J.D. Prince, *Computer Vision: Models, Learning, and Inference*, Cambridge University Press, 2012.

[2] P. Kenny, "Bayesian speaker verification with heavy-tailed priors," in *Proc. Odyssey 2010: The Speaker and Language Recognition Workshop*, 2010.

[3] N. Bümmer and E. de Villiers, "The speaker partitioning problem," in *Proc. Odyssey 2010: The Speaker and Language Recognition Workshop*, 2010.

[4] D. Garcia-Romero and C.Y. Espy-Wilson, "Analysis of i-vector Length Normalization in Speaker Recognition Systems," in *Proc. Interspeech*, 2011, pp. 249–252.

[5] S. Cumani and P. Laface, "Generative pairwise models for speaker recognition," in *Proc. Odyssey 2014: The Speaker and Language Recognition Workshop*, 2014, pp. 273–279.

[6] A. Sizov, K.A. Lee, and T. Kinnunen, *Proc. Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshop, S+SSPR 2014, Joensuu, Finland, August 20-22, 2014.*, chapter Unifying Probabilistic Linear Discriminant Analysis Variants in Biometric Authentication, pp. 464 – 475, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

[7] S. Cumani, O. Plchot, and P. Laface, "Probabilistic linear discriminant analysis of i-vector posterior distributions," in *Int. Conf. on Audio, Speech and Signal Processing (ICASSP 2013)*, 2013, pp. 7644–7648.

[8] P. Kenny, T. Stafylakis, P. Ouellet, Md.J. Alam, and P. Dumouchel, "PLDA for speaker verification with utterances of arbitrary duration," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2013)*, 2013.

[9] N. Brümmer, "EM for probabilistic LDA," Tech. Rep., Agnitio Research, Cape Town, 2010.

[10] D. Matrouf, W.B. Kheder, P.-M. Bousquet, M. Ajili, and J.-F. Bonastre, "Dealing with additive noise in speaker recognition systems based on i-vector approach," in *Proc. 23rd European Signal Processing Conference (EUSIPCO 2015)*, 2015.

[11] M.-W. Mak, X. Pang, and J.-T. Chien, "Mixture of PLDA for Noise Robust I-vector Speaker Verification," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 24, no. 1, pp. 130–142, 2016.

[12] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-End Factor Analysis For Speaker Verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 788–798, 2011.

[13] P. Kenny, "Joint Factor Analysis of Speaker and Session Variability: Theory and Algorithms," Tech. Rep., Centre de recherche informatique de Montréal (CRIM), 2005.

[14] P. Kenny, "A Small Footprint i-Vector Extractor," in *Proc. Odyssey 2012: The Speaker and Language Recognition Workshop*, 2012.

[15] S. Cumani, P. Laface, and O. Plchot, "On the use of i-vector posterior distributions in Probabilistic Linear Discriminant Analysis," *IEEE/ACM Transactions On Audio, Speech and Language Processing*, vol. 22, no. 4, pp. 846 – 857, 2014.

[16] S. Cumani, "Fast Scoring of Full Posterior PLDA models," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 2036–2045, 2015.

[17] R. Saeidi and P. Alku, "Accounting For Uncertainty of i-vectors in Speaker Recognition Using Uncertainty Propagation and Modified Imputation," in *Proc. Interspeech*, 2015, pp. 3546–3550.

[18] R. Saeidi, R. Fernandez Astudillo, and D. Kolossa, "Uncertain LDA: Including observation uncertainties in discriminative transforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press.