The $j$-th i-vector of speaker $i$, denoted by $x_{ij} \in R^d$, is decomposed as:

$$x_{ij} = \mu_i + e_{ij}$$

Within-speaker variability

- Two independent Gaussians: $\mu_i \sim N(0, S_\mu)$, $e_{ij} \sim N(0, S_e)$
- Training (EM algorithm)
  $$\min_{\Theta} \sum_i E_{p(h_i|x_i; \Theta)} \left[ \log p(h_i; \Theta^{t+1}) \right]$$
- Testing
  $$r(x_1, x_2) = \log \frac{p(x_1, x_2 | H_1)}{p(x_1, x_2 | H_2)} = \log p(x_1) - \log p(x_2)$$

The calculation of $p(x_i)$ could be accelerated, which only involves inversion of diagonal matrices. **Complexity: $O(d^3) \rightarrow O(d)$**

Connection with PLDAS

<table>
<thead>
<tr>
<th>Method</th>
<th>JB</th>
<th>two-covariance</th>
<th>SPLDA</th>
<th>Kaldi PLDA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observation</strong></td>
<td>$x_i \equiv {x_{ij}, j = 1, \ldots, m_i}$</td>
<td>$x_i = \frac{1}{m_i} \sum_{j=1}^{m_i} x_{ij}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td>$h_i \equiv {\mu_i, {e_{ij}}}$</td>
<td>$x_{ij} = F z_{ij} + e_{ij}$</td>
<td>$\tilde{x}<em>i = \mu_i + e</em>{i1}$</td>
<td></td>
</tr>
<tr>
<td><strong>EM objective function</strong></td>
<td>$E_{p(b_i</td>
<td>x_i; \Theta^{t+1})} \left[ \log p(h_i) \right]$</td>
<td>$E_{p(b_i</td>
<td>x_i; \Theta)} \left[ \log p(x, h_i) \right]$</td>
</tr>
<tr>
<td><strong>Subspace dimensionality setting</strong></td>
<td>loose</td>
<td>strict</td>
<td>loose</td>
<td></td>
</tr>
<tr>
<td><strong>EM convergence</strong></td>
<td>fast</td>
<td>slow</td>
<td>fast</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The summary of the similarities and difference between JB, SPLDA, Kaldi PLDA and the two-covariance model, $x_{ij}$ denotes the $j$-th i-vector of speaker $i$, $\mu_i \sim N(0, S_\mu)$ is the identity variable for speaker $i$, modeled by the between-class covariance $S_\mu$, $e_{ij} \sim N(0, S_e)$ is the interresidual, modeled by the within-class covariance $S_e$. For SPLDA, $x_i \sim N(0, I)$ stands for the identity variable.

- **EM algorithm for SPLDA:**
  $$\max_{\Theta} E_{p(b,x;\Theta)} \left[ \log p(x, b; \Theta) \right]$$
  $$E_{p(b_i|x_i; \Theta)} \left[ \log p(h_i) \right] = E_{p(b_i|x_i; \Theta)} \left[ \log p(x, h_i) \right]$$

When $A_t$ is small and $F_{t+1} \approx F_t$,

$$x_{ij} \approx F_{t-1} + \epsilon_{ij}$$

The EM update for SPLDA could easily be stuck into non-local minima with small $A_t$.

The EM update for JB does not have such problem.

- **JB calculates the joint likelihood**
  $$p(x_i) = N(0, \Sigma_{x_i})$$

- **Kaldi calculates the likelihood of the single average i-vector $x_{\tilde{i}}$.**
  $$p(x_{\tilde{i}}) = N \left( 0, F T + \frac{1}{n_i} \Lambda \right)$$

**Experiments**

- **Speaker Verification Performance**
  ![](Image 3862x2022.png)

  (a) SRE10 MALE (b) SRE10 FEMALE

<table>
<thead>
<tr>
<th>System</th>
<th>SRE10 MALE</th>
<th>DCF10</th>
<th>DCF08</th>
<th>SRE10 FEMALE</th>
<th>DCF10</th>
<th>DCF08</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA+Cos</td>
<td>1.905</td>
<td>0.293</td>
<td>0.091</td>
<td>2.619</td>
<td>0.390</td>
<td>0.126</td>
</tr>
<tr>
<td>Kaldi PLDA</td>
<td>1.290</td>
<td>0.284</td>
<td>0.079</td>
<td>1.944</td>
<td>0.345</td>
<td>0.102</td>
</tr>
<tr>
<td>SPLDA</td>
<td>1.010</td>
<td>0.217</td>
<td>0.055</td>
<td>1.621</td>
<td>0.287</td>
<td>0.079</td>
</tr>
<tr>
<td>JB</td>
<td>0.894</td>
<td>0.188</td>
<td>0.048</td>
<td>1.485</td>
<td>0.245</td>
<td>0.069</td>
</tr>
</tbody>
</table>

- **Subspace Dimensionality**
  ![](Image 2678x1488.png)

  ![](Image 2881x151.png)

  Fig. 3. (a) The negative log-likelihood of JB (EM with exact or approximated statistics) and SPLDA during training. (b) The zoom-in of negative log-likelihood convergence curves for JB with exact and approximated EM statistics.

- **Convergence Rate**
  ![](Image 2692x1157.png)

  ![](Image 3108x117.png)

  Fig. 2. The influence of subspace dimensionality on JB and SPLDA using NIST SRE10 core condition male test data.