Fast Scoring for PLDA with Uncertainty Propagation

Wei-wei LIN and Man-Wai Mak

June 2016

Department of Electronic and Information Engineering
The Hong Kong Polytechnic University

Contents

I-vector/PLDA

- State-of-the-art method
- I-vector extraction can be described as:

$$\boldsymbol{\beta} = \mathbf{m} + \mathbf{T}\boldsymbol{\eta} \qquad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Speaker supervector (61440x1)

GMM supervector

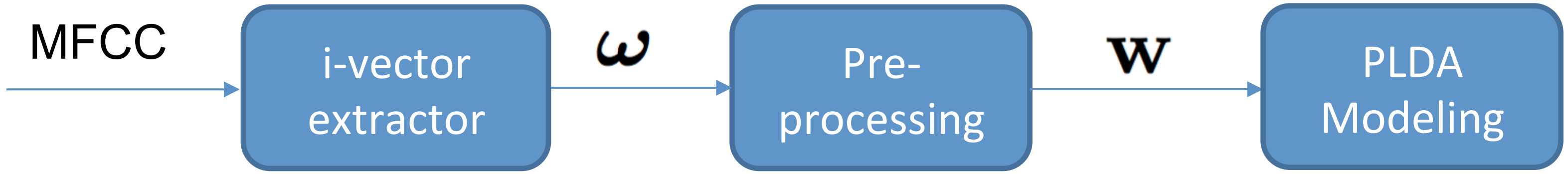
Total variability matrix (61440x500)

Total variability factor (500x1)

  - I-vector $\boldsymbol{\omega} = \langle \boldsymbol{\eta} | \mathcal{O} \rangle$ is the maximum-a-posteriori (MAP) estimate of $\boldsymbol{\eta}$
  - Instead of using the high-dimensional supervector $\boldsymbol{\beta}$ to represent speaker, we use more compact (low-dimension) i-vector $\boldsymbol{\omega}$ to represent speaker.
  - $\mathbf{T}$ represents the subspace where i-vectors can vary.

I-vector/PLDA

- Procedure of i-vector/PLDA



- In Gaussian PLDA, the preprocessed i-vector $\mathbf{w}_{i,j}$ from the *j*-th session of the *i*-th speaker is assumed to be generated from a factor analysis model:

$$\underbrace{\mathbf{w}_{i,j}}_{\text{Pre-processed i-vector}} = \underbrace{\boldsymbol{\mu}}_{\text{mean of i-vectors in training set}} + \underbrace{\mathbf{V}}_{\text{speaker subspace}}\underbrace{\mathbf{h}_i}_{\text{speaker factor}} + \underbrace{\boldsymbol{\epsilon}_{i,j}}_{\text{residual}}$$

I-vector/PLDA

- Given a test i-vector $\mathbf{w}_t$ and target-speaker's i-vectors $\mathbf{w}_s$, verification score is the log-likelihood ratio between two hypotheses:

$$\text{score} = \log\left[\frac{p(\mathbf{w}_s, \mathbf{w}_t|\text{same-speaker})}{p(\mathbf{w}_s, \mathbf{w}_t|\text{different-speakers})}\right]$$
$$= \frac{1}{2}\mathbf{w}_s^{\mathsf{T}}\boldsymbol{\Phi}\mathbf{w}_s + \mathbf{w}_s^{\mathsf{T}}\boldsymbol{\Psi}\mathbf{w}_t + \frac{1}{2}\mathbf{w}_t^{\mathsf{T}}\boldsymbol{\Phi}\mathbf{w}_t + \text{const}$$

where

$$\boldsymbol{\Phi} = \boldsymbol{\Sigma}_{tot}^{-1} - (\boldsymbol{\Sigma}_{tot} - \boldsymbol{\Sigma}_{ac}\boldsymbol{\Sigma}_{tot}^{-1}\boldsymbol{\Sigma}_{ac})^{-1}$$
$$\boldsymbol{\Psi} = \boldsymbol{\Sigma}_{tot}^{-1}\boldsymbol{\Sigma}_{ac}(\boldsymbol{\Sigma}_{tot} - \boldsymbol{\Sigma}_{ac}\boldsymbol{\Sigma}_{tot}^{-1}\boldsymbol{\Sigma}_{ac})^{-1}$$
$$\boldsymbol{\Sigma}_{ac} = \mathbf{V}\mathbf{V}^{\mathsf{T}} \qquad \boldsymbol{\Sigma}_{tot} = \mathbf{V}\mathbf{V}^{\mathsf{T}} + \boldsymbol{\Sigma}.$$

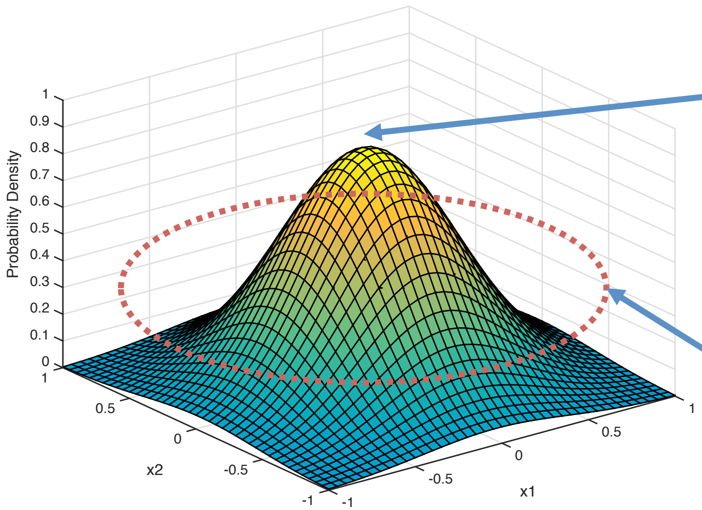
These matrices are independent of the test utterance. So, they can be pre-computed.

Problems with i-vector/PLDA

- Conventional i-vector/PLDA system has no ability to represent the **reliability** of i-vectors.
- This poses a severe problem for **short-utterance speaker verification**, because short utterances do not have enough data for MAP estimation. In such case, the prior dominates the MAP estimate.
- As a result, PLDA scores will favor same-speaker hypothesis for short utterances even if the test utterance is given by an impostor.

PLDA with Uncertainty Propagation

- In i-vector extraction, besides the posterior mean of the latent variable (i-vector) , we also have the posterior covariance matrix, which reflects the uncertainty of the i-vector estimate.



$$\boldsymbol{\omega} = \mathrm{cov}(\boldsymbol{\eta}, \boldsymbol{\eta}) \sum_{c=1}^{C} \mathbf{T}_c^{\mathsf{T}} \boldsymbol{\Sigma}_c^{-1} \tilde{\mathbf{f}}_c$$

$$\mathrm{cov}(\boldsymbol{\eta}, \boldsymbol{\eta}) = \mathbf{L}^{-1} = \left(\mathbf{I} + \sum_{c=1}^{C} N_c \mathbf{T}_c^{\mathsf{T}} \boldsymbol{\Sigma}_c^{-1} \mathbf{T}_c \right)^{-1}$$

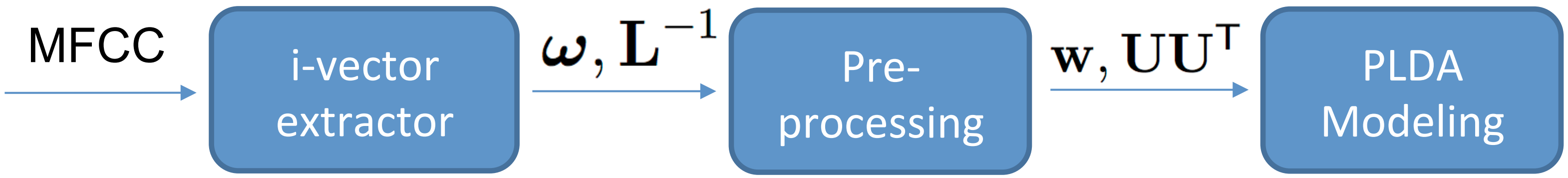
$\mathbf{L}$ is the precision matrix of the posterior density

N_c is zero-order sufficient statistics with respect to UBM

$\tilde{\mathbf{f}}_c$ is first-order sufficient statistics with respect to UBM

PLDA with Uncertainty Propagation

- Procedure of PLDA-UP (Kenny *et al*. 2013)



- Generative model

$$\mathbf{w}_{i,j} = \boldsymbol{\mu} + \mathbf{V}\mathbf{h}_i + \mathbf{U}_{i,j}\mathbf{z}_{i,j} + \boldsymbol{\epsilon}_{i,j}$$

  - $\mathbf{U}_{i,j}$ is the Cholesky decomposition of the posterior covariance matrix of the *j*-th utterance by the *i*-th speaker
  - The intra-speaker covariance matrix becomes:

$$\mathrm{cov}(\mathbf{w}_{i,j}, \mathbf{w}_{i,j}|\mathbf{h}_i) = \mathbf{U}_{i,j}\mathbf{U}_{i,j}^{\mathsf{T}} + \boldsymbol{\Sigma}$$

where $\mathbf{U}_{i,j}\mathbf{U}_{i,j}^{\mathsf{T}}$ changes from utterance to utterance, thus reflecting the reliability of the i-vector $\mathbf{w}_{i,j}$.

THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學

PLDA-UP

- The log-likelihood ratio score is:

$$\text{score} = \frac{1}{2}\mathbf{w}_s\mathbf{A}_{s,t}\mathbf{w}_s + \mathbf{w}_s^{\mathsf{T}}\mathbf{B}_{s,t}\mathbf{w}_t + \frac{1}{2}\mathbf{w}_t^{\mathsf{T}}\mathbf{C}_{s,t}\mathbf{w}_t + D_{s,t}$$

where

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

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

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

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

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

Terms that depend on test utterances must be evaluated during verification

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

$$\mathbf{\Sigma}_{ac} = \mathbf{V}\mathbf{V}^{\mathsf{T}}$$

Terms independent of test utterances can be pre-computed

PLDA vs PLDA with UP

Conventional PLDA Scoring Equation	Other terms needed to be evaluated during verification
$\text{score} = \frac{1}{2}\mathbf{w}_s^{\mathsf{T}}\mathbf{\Phi}\mathbf{w}_s + \mathbf{w}_s^{\mathsf{T}}\mathbf{\Phi}\mathbf{w}_t + \frac{1}{2}\mathbf{w}_t^{\mathsf{T}}\mathbf{\Phi}\mathbf{w}_t + \text{const}$	None
PLDA with UP Scoring Equation	**Other terms needed to be evaluated during verification**
$\text{score} = \frac{1}{2}\mathbf{w}_s\mathbf{A}_{s,t}\mathbf{w}_s + \mathbf{w}_s^{\mathsf{T}}\mathbf{B}_{s,t}\mathbf{w}_t + \frac{1}{2}\mathbf{w}_t^{\mathsf{T}}\mathbf{C}_{s,t}\mathbf{w}_t + D_{s,t}$	$\mathbf{A}_{s,t} = \mathbf{\Sigma}_s^{-1} - (\mathbf{\Sigma}_s - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_t^{-1}\mathbf{\Sigma}_{ac})^{-1}$ $\mathbf{B}_{s,t} = \mathbf{\Sigma}_s^{-1}\mathbf{\Sigma}_{ac}(\mathbf{\Sigma}_t - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma}_{ac})^{-1}$ $\mathbf{C}_{s,t} = \mathbf{\Sigma}_t^{-1} - (\mathbf{\Sigma}_t - \mathbf{\Sigma}_{ac}\mathbf{\Sigma}_s^{-1}\mathbf{\Sigma}_{ac})^{-1}$ $D_{s,t} = -\frac{1}{2}\log\begin{vmatrix}\mathbf{\Sigma}_s & \mathbf{\Sigma}_{ac}\\ \mathbf{\Sigma}_{ac} & \mathbf{\Sigma}_t\end{vmatrix} + \frac{1}{2}\log\begin{vmatrix}\mathbf{\Sigma}_s & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_t\end{vmatrix}$ $\mathbf{\Sigma}_t = \mathbf{V}\mathbf{V}^{\mathsf{T}} + \mathbf{U}_t\mathbf{U}_t^{\mathsf{T}} + \mathbf{\Sigma}$

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Contents

Motivation

- Posterior covariance of latent factors:

$$\mathrm{cov}(\boldsymbol{\eta}, \boldsymbol{\eta}) = \mathbf{L}^{-1} = \left(\mathbf{I} + \sum_{c=1}^{C} N_c \mathbf{T}_c^{\mathsf{T}} \boldsymbol{\Sigma}_c^{-1} \mathbf{T}_c \right)^{-1}$$

- N_c is proportional to the number of frames in an utterance, which suggests that the posterior covariance matrix $\mathbf{L}^{-1}$ quantifies the uncertainty through utterance duration.
- If two utterances are of approximately the same duration, their posterior covariance matrices should be similar.

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Fast Scoring for PLDA-UP

- We proposed grouping i-vectors according to their reliability.
- For each group, i-vectors' reliability is model by a posterior covariance matrix obtained from development data.
- The new PLDA model can be written as:

$$\mathbf{w}_{i,j}^{(k)} = \boldsymbol{\mu} + \mathbf{V}\mathbf{h}_i + \mathbf{U}_k\mathbf{z}_{i,j} + \boldsymbol{\epsilon}_{i,j}$$

  – k is the group identity to which $\mathbf{w}_{i,j}$ belongs
  – I-vectors within the same group share the same loading matrix $\mathbf{U}_k$.
  – The loading matrices $\{\mathbf{U}_k | k = 1, 2, \ldots, K\}$ are obtained from development data.
- Compared with the original PLDA-UP:

$$\mathbf{w}_{i,j} = \boldsymbol{\mu} + \mathbf{V}\mathbf{h}_i + \mathbf{U}_{i,j}\mathbf{z}_{i,j} + \boldsymbol{\epsilon}_{i,j}$$

Fast Scoring for PLDA-UP

- We proposed grouping i-vectors according to their reliability.
- For each group, i-vectors' reliability is model by a posterior covariance matrix obtained from development data.
- The new PLDA model can be written as:

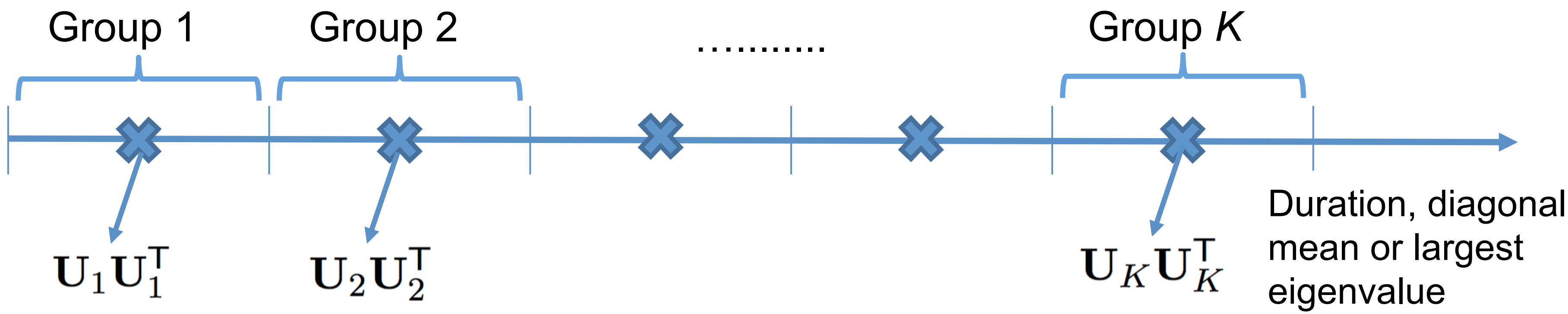
$$\mathbf{w}_{i,j}^{(k)} = \boldsymbol{\mu} + \mathbf{V}\mathbf{h}_i + \mathbf{U}_k \mathbf{z}_{i,j} + \boldsymbol{\epsilon}_{i,j}$$

  - k is the group identity to which $\mathbf{w}_{i,j}$ belongs
  - I-vectors within the same group share the same loading matrix $\mathbf{U}_k$.
  - The loading matrices $\{\mathbf{U}_k | k = 1, 2, \ldots, K\}$ are obtained from development data.
- Compared with the original PLDA-UP:

$$\mathbf{w}_{i,j} = \boldsymbol{\mu} + \mathbf{V}\mathbf{h}_i + \mathbf{U}_{i,j} \mathbf{z}_{i,j} + \boldsymbol{\epsilon}_{i,j}$$

Fast Scoring for PLDA-UP

- Three grouping schemes based on:
  1) Utterance duration
  2) Mean of diagonal elements of posterior covariance matrix
  3) Largest eigenvalue of posterior covariance matrix
- Basic procedures:
  1. Compute the posterior covariance matrices from development data
  2. For the *k*-th group, select the representative $\mathbf{U}_k\mathbf{U}_k^{\mathsf{T}}$



Fast Scoring for PLDA-UP

- During scoring, we find the group identities *m* and *n* of the target-speaker i-vector $\mathbf{w}_s$ and the test i-vector $\mathbf{w}_t$.
- Then, we retrieve pre-computed matrices $\{\mathbf{A}_{m,n}, \mathbf{B}_{m,n}, \mathbf{C}_{m,n}, D_{m,n}\}$ from the repository to compute the score

$$\text{score} = \frac{1}{2}\mathbf{w}_s^{\mathsf{T}}\mathbf{A}_{m,n}\mathbf{w}_s + \mathbf{w}_s^{\mathsf{T}}\mathbf{B}_{m,n}\mathbf{w}_t + \frac{1}{2}\mathbf{w}_t^{\mathsf{T}}\mathbf{C}_{m,n}\mathbf{w}_t + D_{m,n}$$

- Compared with the original PLDA-UP

$$\text{score} = \frac{1}{2}\mathbf{w}_s\mathbf{A}_{s,t}\mathbf{w}_s + \mathbf{w}_s^{\mathsf{T}}\mathbf{B}_{s,t}\mathbf{w}_t + \frac{1}{2}\mathbf{w}_t^{\mathsf{T}}\mathbf{C}_{s,t}\mathbf{w}_t + D_{s,t}$$

Fast Scoring for PLDA-UP

- During scoring, we find the group identities *m* and *n* of the target-speaker i-vector $\mathbf{w}_s$ and the test i-vector $\mathbf{w}_t$.
- Then, we retrieve pre-computed matrices $\{\mathbf{A}_{m,n}, \mathbf{B}_{m,n}, \mathbf{C}_{m,n}, D_{m,n}\}$ from the repository to compute the score

$$\text{score} = \frac{1}{2}\mathbf{w}_s^{\mathsf{T}}\mathbf{A}_{m,n}\mathbf{w}_s + \mathbf{w}_s^{\mathsf{T}}\mathbf{B}_{m,n}\mathbf{w}_t + \frac{1}{2}\mathbf{w}_t^{\mathsf{T}}\mathbf{C}_{m,n}\mathbf{w}_t + D_{m,n}$$

- Compared with the original PLDA-UP

$$\text{score} = \frac{1}{2}\mathbf{w}_s\mathbf{A}_{s,t}\mathbf{w}_s + \mathbf{w}_s^{\mathsf{T}}\mathbf{B}_{s,t}\mathbf{w}_t + \frac{1}{2}\mathbf{w}_t^{\mathsf{T}}\mathbf{C}_{s,t}\mathbf{w}_t + D_{s,t}$$

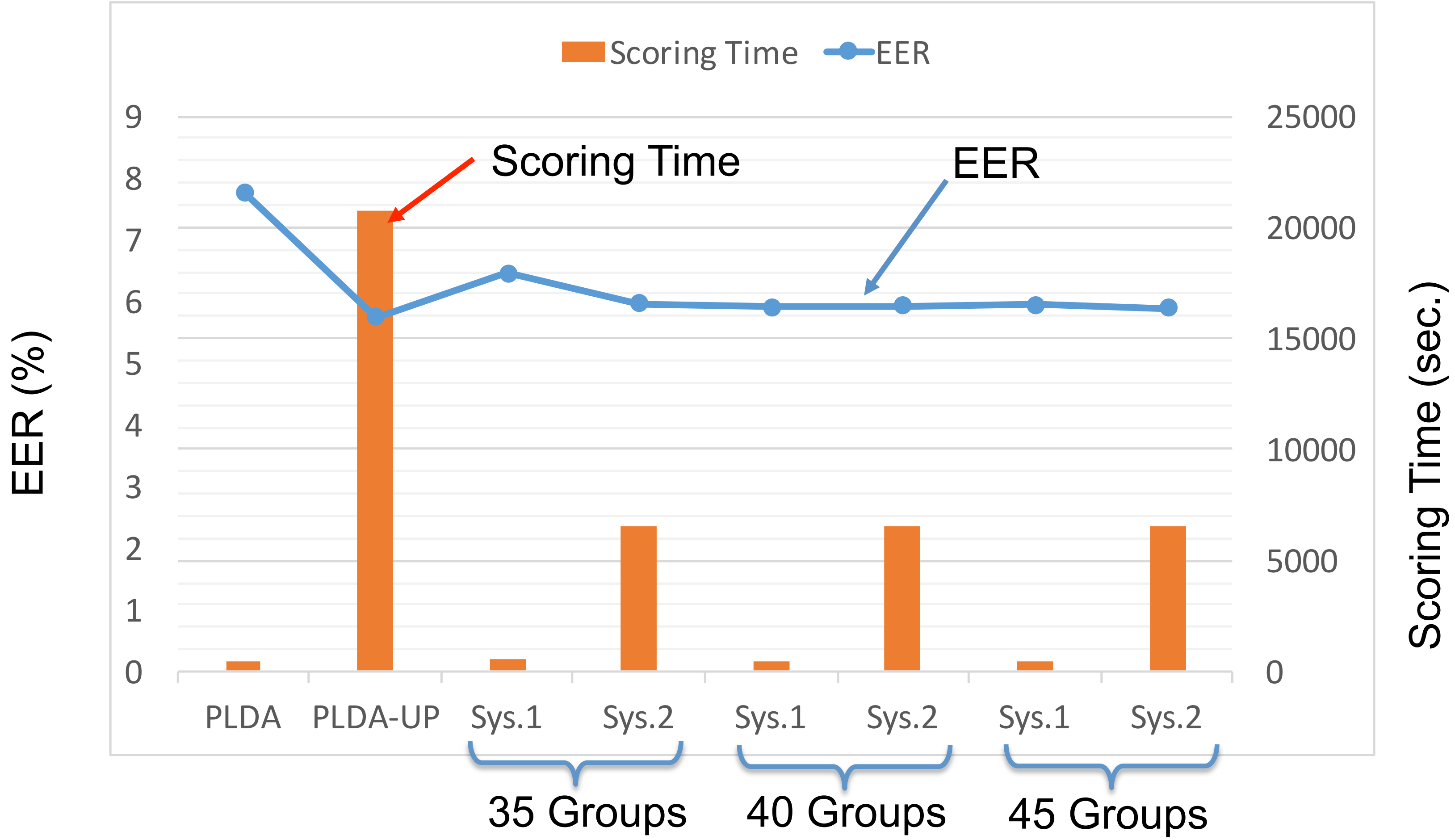
UP vs UP with Fast Scoring

PLDA with UP using fast scoring	Other Terms needed to be evaluated during verification
$\text{score} = \frac{1}{2}\mathbf{w}_s^{\mathsf{T}}\mathbf{A}_{m,n}\mathbf{w}_s + \mathbf{w}_s^{\mathsf{T}}\mathbf{B}_{m,n}\mathbf{w}_t + \frac{1}{2}\mathbf{w}_t^{\mathsf{T}}\mathbf{C}_{m,n}\mathbf{w}_t + D_{m,n}$	Determine the group index of test utterance
PLDA with UP using exact scoring	**Terms needed to be evaluated during verification**
$\text{score} = \frac{1}{2}\mathbf{w}_s\mathbf{A}_{s,t}\mathbf{w}_s + \mathbf{w}_s^{\mathsf{T}}\mathbf{B}_{s,t}\mathbf{w}_t + \frac{1}{2}\mathbf{w}_t^{\mathsf{T}}\mathbf{C}_{s,t}\mathbf{w}_t + D_{s,t}$	$\mathbf{A}_{s,t} = \boldsymbol{\Sigma}_s^{-1} - (\boldsymbol{\Sigma}_s - \boldsymbol{\Sigma}_{ac}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\Sigma}_{ac})^{-1}$ $\mathbf{B}_{s,t} = \boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{ac}(\boldsymbol{\Sigma}_t - \boldsymbol{\Sigma}_{ac}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{ac})^{-1}$ $\mathbf{C}_{s,t} = \boldsymbol{\Sigma}_t^{-1} - (\boldsymbol{\Sigma}_t - \boldsymbol{\Sigma}_{ac}\boldsymbol{\Sigma}_s^{-1}\boldsymbol{\Sigma}_{ac})^{-1}$ $D_{s,t} = -\frac{1}{2}\log\begin{vmatrix}\boldsymbol{\Sigma}_s & \boldsymbol{\Sigma}_{ac} \\ \boldsymbol{\Sigma}_{ac} & \boldsymbol{\Sigma}_t\end{vmatrix} + \frac{1}{2}\log\begin{vmatrix}\boldsymbol{\Sigma}_s & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_t\end{vmatrix}$ $\boldsymbol{\Sigma}_t = \mathbf{V}\mathbf{V}^{\mathsf{T}} + \mathbf{U}_t\mathbf{U}_t^{\mathsf{T}} + \boldsymbol{\Sigma}$

Experiments

- **Evaluation dataset:** Common evaluation conditions 2 of NIST SRE 2012 core set (truncated to range from 1-42 seconds).
- **Parameterization:** 19 MFCCs together with energy plus their 1st and 2nd derivatives → 60-Dim
- **UBM:** gender-dependent, 1024 mixtures
- **Total Variability Matrix:** gender-dependent, 500 total factors
- **I-Vector Preprocessing:**
  - Whitening by WCCN then length normalization
  - Followed by LDA (500-dim → 200-dim) and WCCN
- **PLDA and PLDA-UP with 150 speaker factors**
- **Fast Scoring Systems:**
  - System 1: Using Utterance duration
  - System 2: Using the mean of diagonal element of UU^T
  - System 3: Using the largest eigenvalue of UU^T

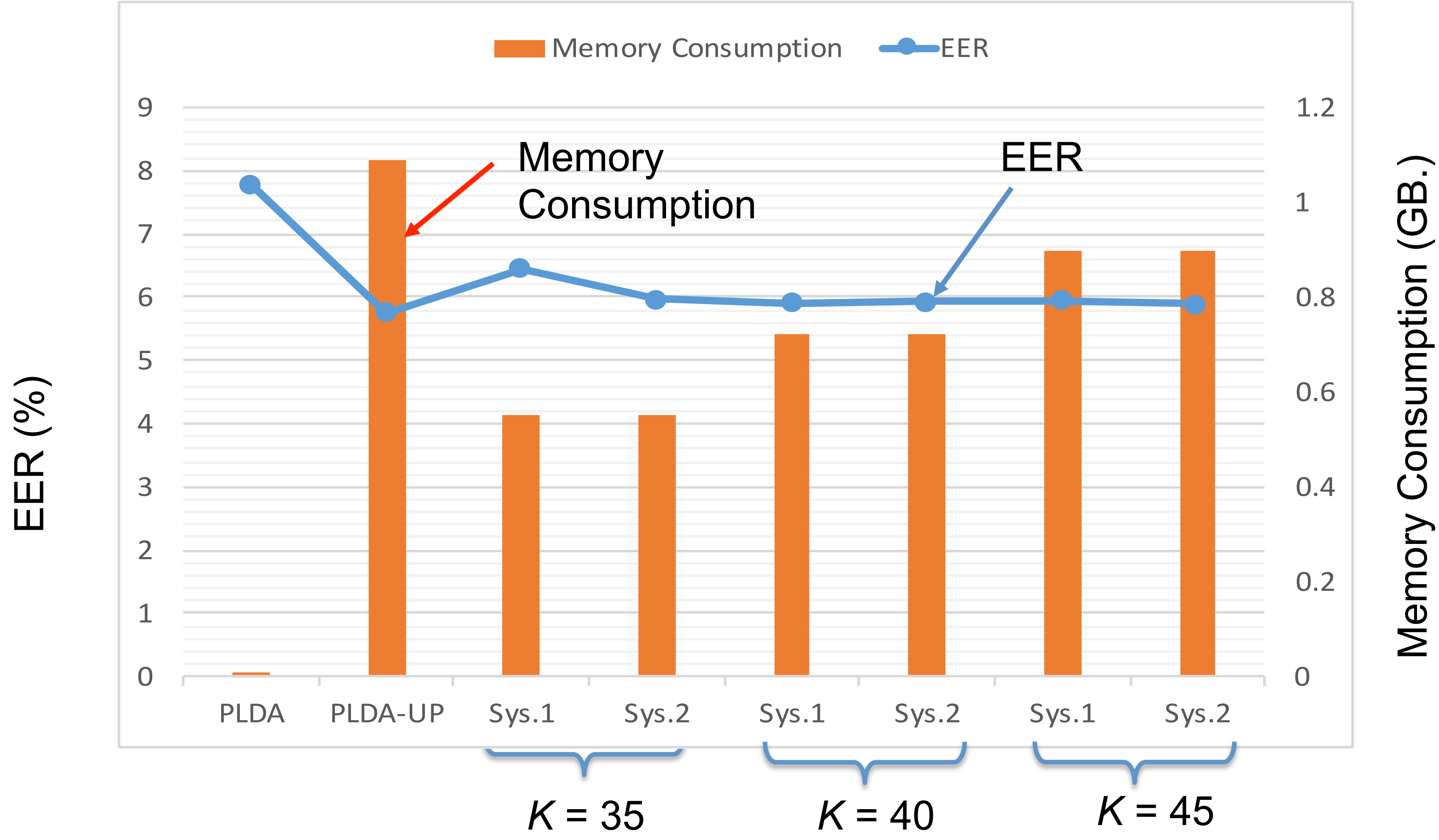
Comparing Scoring Time and EER



Sys 1: Use utterance duration

Sys 2: Use the mean of diagonal element of UU^T

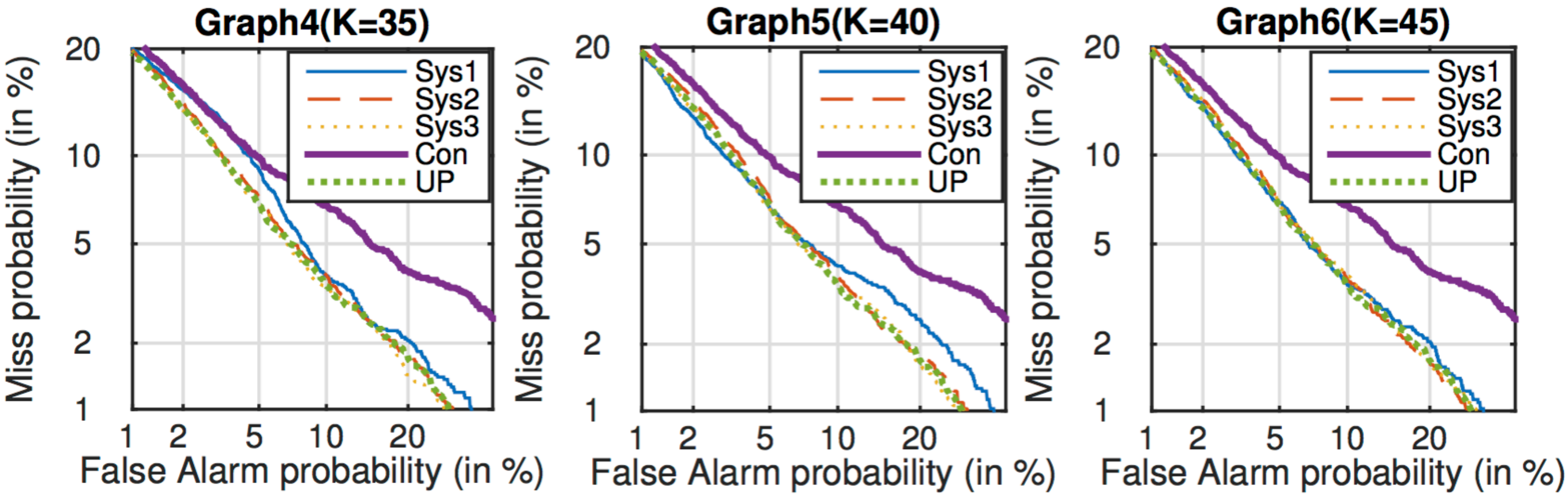
Comparing Memory Consumption



Sys 1: Use Utterance duration

Sys 2: Use the mean of diagonal elements of UU^T

DET Curves



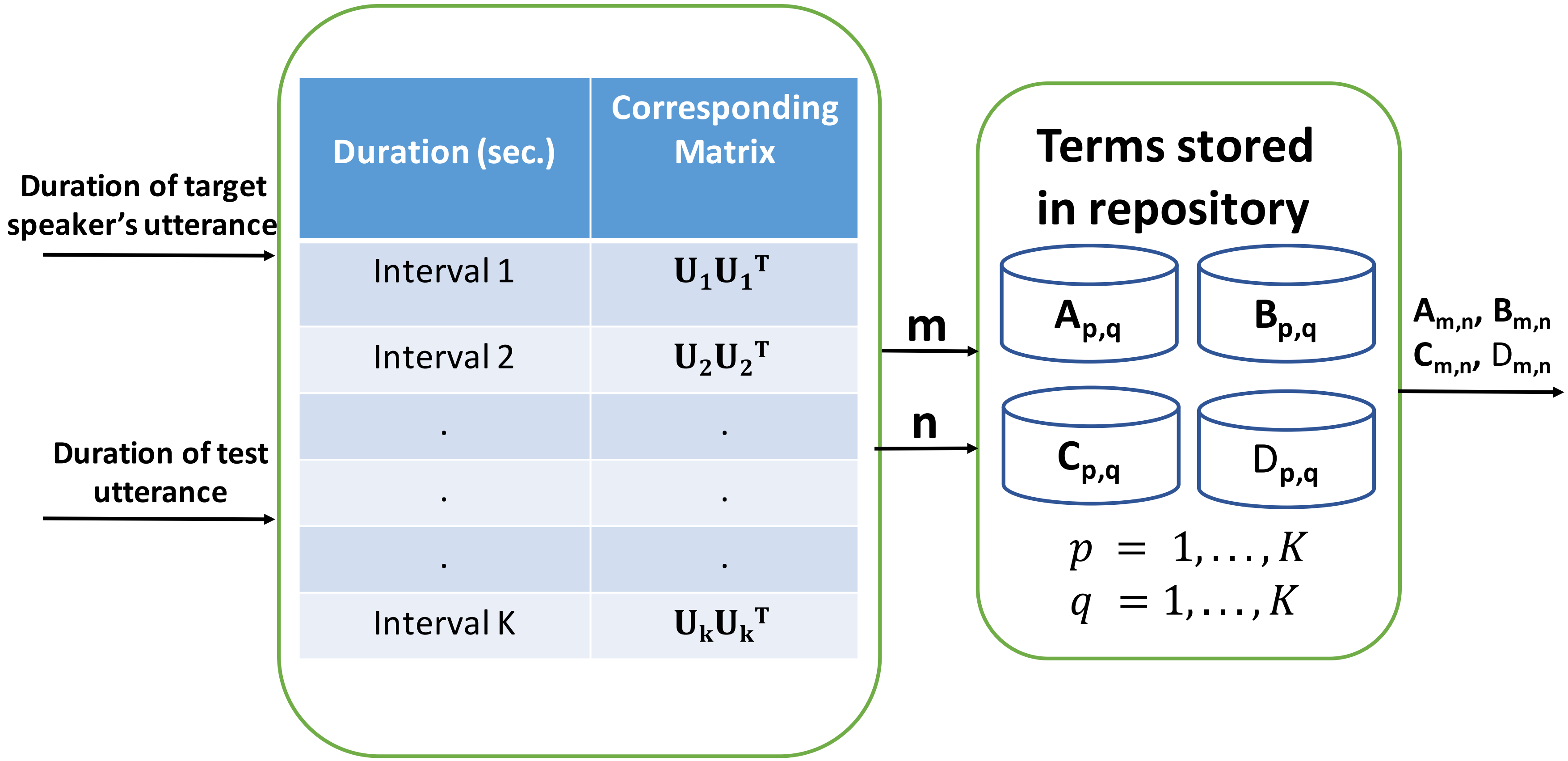
Sys 1: Fast scoring based on utterance duration
Sys 2: Fast scoring based on the mean of diagonal element of UU^T
Sys 3: Fast scoring based on the largest eigenvalue of UU^T
Con: Conventional PLDA
UP: PLDA with UP (without fast scoring)

Other than the problematic Sys 1 (using duration), DET curves show that fast scoring Systems can perform as good as PLDA-UP.

Conclusions

- We proposed a fast scoring method for PLDA with uncertainty propagation.
- Session-dependent loading matrices in UP were substituted by length-dependent matrices. Thus, pre-computations are possible.
- **Experiments confirm that the proposed method can perform as well as standard UP with only 2.3% of scoring time (Sys .1 K=45).**

Fast Scoring for PLDA-UP



Duration (sec.)	Corresponding Matrix
Interval 1	$\mathbf{U_1 U_1^T}$
Interval 2	$\mathbf{U_2 U_2^T}$
.	.
.	.
.	.
Interval K	$\mathbf{U_k U_k^T}$

Results and Discussion

- Performance of conventional PLDA, PLDA-UP and fast scoring systems.

Method	K	Male(CC2)					
		EER(%)			minDCF		
		Sys1	Sys2	Sys3	Sys1	Sys2	Sys3
Fast Scoring Systems	20	6.21	7.02	6.17	0.640	0.685	0.654
	25	6.07	6.35	6.00	0.635	0.658	0.646
	30	5.96	6.07	5.93	0.632	0.632	0.648
	35	6.45	5.97	5.91	0.633	0.631	0.643
	40	5.91	5.93	5.85	0.641	0.641	0.649
	45	5.95	5.89	5.96	0.633	0.642	0.636
PLDA	-	7.77			0.654		
PLDA-UP	-	5.75			0.644		

Time and Memory Consumption

Method	K	Male(CC2)			
		EER(%)	minDCF	Time(sec)	Mem.(GB)
PLDA	-	7.77	0.654	412	0.01
PLDA-UP	-	5.75	0.644	20729	1.09
Sys. 1	35	6.45	0.686	510	0.55
	40	5.91	0.658	492	0.72
	45	5.95	0.632	497	0.90
Sys. 2	35	5.97	0.631	6500	0.55
	40	5.93	0.641	6511	0.72
	45	5.89	0.642	6502	0.90