Derivations of Variation Bayes Equations in SNR- and Length-Invariant PLDA

Man-Wai MAK

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

Abstract

This document derives the optimal solutions of the variational posterior densities of latent factors in the SNR-invariant PLDA models.

Keywords: PLDA, Variational Bayes, Variational inference

---

1. SNR-Invariant PLDA

- Classical PLDA: $\mathbf{x}_{ij} = \mathbf{m} + \mathbf{V}\mathbf{h}_i + \boldsymbol{\epsilon}_{ij}$
- By adding an SNR factor to the conventional PLDA, we have SNR-invariant PLDA:

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

  where $\mathbf{U}$ denotes the SNR subspace, $\mathbf{w}_k$ is an SNR factor, and $\mathbf{h}_i$ is the speaker (identity) factor for speaker i.

- Note that it is not the same as PLDA with channel subspace:

$$\mathbf{x}_{ij}^k = \mathbf{m} + \mathbf{V}\mathbf{h}_i + \mathbf{G}\mathbf{r}_{ij} + \boldsymbol{\epsilon}_{ij},$$

  where $\mathbf{G}$ defines the channel subspace and $\mathbf{r}_{ij}$ represents the channel factors.

2. Problem Statement

Denote $\underline{\mathbf{w}} = [\mathbf{w}_1, \ldots, \mathbf{w}_K]$ and $\underline{\mathbf{h}} = [\mathbf{h}_1, \ldots, \mathbf{h}_S]$. In variational Bayes, we factorize the joint posterior as follows:

$$\ln p(\underline{\mathbf{h}}, \underline{\mathbf{w}}|\mathcal{X}) \approx \ln q(\underline{\mathbf{h}}) + \ln q(\underline{\mathbf{w}}) = \sum_{i=1}^{S} \ln q(\mathbf{h}_i) + \sum_{k=1}^{K} \ln q(\mathbf{w}_k)$$

where

$$\begin{aligned} \ln q(\underline{\mathbf{h}}) &= \mathbb{E}_{q(\underline{\mathbf{w}})}\{\ln p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})\} + \text{const} \\ \ln q(\underline{\mathbf{w}}) &= \mathbb{E}_{q(\underline{\mathbf{h}})}\{\ln p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})\} + \text{const} \end{aligned} \tag{1}$$

where $\mathbb{E}_{q(\underline{\mathbf{w}})}$ means taking expectation with respect to $\underline{\mathbf{w}}$ using the variational posterior $q(\underline{\mathbf{w}})$.

3. Derivation of VB Posteriors

We want to approximate the true joint posteriors $p(\underline{\mathbf{h}}, \underline{\mathbf{w}}|\mathcal{X})$ by a distribution $q(\underline{\mathbf{h}}, \underline{\mathbf{w}})$ such that

$$q(\underline{\mathbf{h}}, \underline{\mathbf{w}}) = q(\underline{\mathbf{h}})q(\underline{\mathbf{w}}).$$

The derivation of the VB posteriors begins with minimizing the KL-divergence

$$\begin{aligned} \mathcal{D}_{\text{KL}}(q(\underline{\mathbf{h}}, \underline{\mathbf{w}})||p(\underline{\mathbf{h}}, \underline{\mathbf{w}}|\mathcal{X})) &= \int\int q(\underline{\mathbf{h}}, \underline{\mathbf{w}}) \ln \left[\frac{q(\underline{\mathbf{h}}, \underline{\mathbf{w}})}{p(\underline{\mathbf{h}}, \underline{\mathbf{w}}|\mathcal{X})}\right] \mathrm{d}\underline{\mathbf{h}}\mathrm{d}\underline{\mathbf{w}} \\ &= -\int\int q(\underline{\mathbf{h}}, \underline{\mathbf{w}}) \ln \left[\frac{p(\underline{\mathbf{h}}, \underline{\mathbf{w}}|\mathcal{X})}{q(\underline{\mathbf{h}}, \underline{\mathbf{w}})}\right] \mathrm{d}\underline{\mathbf{h}}\mathrm{d}\underline{\mathbf{w}} \\ &= -\int\int q(\underline{\mathbf{h}}, \underline{\mathbf{w}}) \ln \left[\frac{p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})}{q(\underline{\mathbf{h}}, \underline{\mathbf{w}})p(\mathcal{X})}\right] \mathrm{d}\underline{\mathbf{h}}\mathrm{d}\underline{\mathbf{w}} \\ &= -\mathcal{L}(q) + \ln p(\mathcal{X}) \end{aligned} \tag{2}$$

where $\mathcal{L}(q)$ is the variational Bayes lower bound (VBLB). The lower bound is given by:

$$\begin{aligned} \mathcal{L}(q) &= \int\int q(\underline{\mathbf{h}}, \underline{\mathbf{w}}) \ln \left[\frac{p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})}{q(\underline{\mathbf{h}}, \underline{\mathbf{w}})}\right] \mathrm{d}\underline{\mathbf{h}}\mathrm{d}\underline{\mathbf{w}} \\ &= \int\int q(\underline{\mathbf{h}}, \underline{\mathbf{w}}) \ln p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})\mathrm{d}\underline{\mathbf{h}}\mathrm{d}\underline{\mathbf{w}} - \int\int q(\underline{\mathbf{h}}, \underline{\mathbf{w}}) \ln q(\underline{\mathbf{h}}, \underline{\mathbf{w}})\mathrm{d}\underline{\mathbf{h}}\mathrm{d}\underline{\mathbf{w}} \end{aligned}$$

$$
\begin{aligned}
&= \int\int q(\underline{\mathbf{h}})q(\underline{\mathbf{w}}) \ln p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})\mathrm{d}\underline{\mathbf{h}}\mathrm{d}\underline{\mathbf{w}} - \int\int q(\underline{\mathbf{h}})q(\underline{\mathbf{w}}) \ln q(\underline{\mathbf{h}})\mathrm{d}\underline{\mathbf{h}}\mathrm{d}\underline{\mathbf{w}} \\
&\quad - \int\int q(\underline{\mathbf{h}})q(\underline{\mathbf{w}}) \ln q(\underline{\mathbf{w}})\mathrm{d}\underline{\mathbf{h}}\mathrm{d}\underline{\mathbf{w}} \\
&= \int\int q(\underline{\mathbf{h}})q(\underline{\mathbf{w}}) \ln p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})\mathrm{d}\underline{\mathbf{h}}\mathrm{d}\underline{\mathbf{w}} - \int q(\underline{\mathbf{h}}) \ln q(\underline{\mathbf{h}})\mathrm{d}\underline{\mathbf{h}} \\
&\quad - \int q(\underline{\mathbf{w}}) \ln q(\underline{\mathbf{w}})\mathrm{d}\underline{\mathbf{w}}. \qquad (3)
\end{aligned}
$$

Note that the first term in Eq. 3 can be written as

$$
\begin{aligned}
\int\int q(\underline{\mathbf{h}})q(\underline{\mathbf{w}}) \ln p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})\mathrm{d}\underline{\mathbf{h}}\mathrm{d}\underline{\mathbf{w}} &= \int q(\underline{\mathbf{h}}) \left[\int q(\underline{\mathbf{w}}) \ln p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})\mathrm{d}\underline{\mathbf{w}}\right] \mathrm{d}\underline{\mathbf{h}} \\
&= \int q(\underline{\mathbf{h}})\mathbb{E}_{q(\underline{\mathbf{w}})}\{\ln p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})\}\mathrm{d}\underline{\mathbf{h}}.
\end{aligned} \qquad (4)
$$

Define a distribution of $\underline{\mathbf{h}}$ as

$$
q^*(\underline{\mathbf{h}}) \equiv \frac{1}{Z} \exp\left\{\mathbb{E}_{q(\underline{\mathbf{w}})}\{\ln p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})\}\right\}, \qquad (5)
$$

where Z is to normalize the distribution. Using Eq. 5 and Eq. 4, Eq. 3 can be written as:

$$
\begin{aligned}
\mathcal{L}(q) &= \int q(\underline{\mathbf{h}}) \ln q^*(\underline{\mathbf{h}})\mathrm{d}\underline{\mathbf{h}} - \int q(\underline{\mathbf{h}}) \ln q(\underline{\mathbf{h}})\mathrm{d}\underline{\mathbf{h}} - \int q(\underline{\mathbf{w}}) \ln q(\underline{\mathbf{w}})\mathrm{d}\underline{\mathbf{w}} + \ln Z \\
&= -\int q(\underline{\mathbf{h}}) \ln \left[\frac{q(\underline{\mathbf{h}})}{q^*(\underline{\mathbf{h}})}\right] \mathrm{d}\underline{\mathbf{h}} + \mathcal{H}(q(\underline{\mathbf{w}})) + \ln Z \\
&= -\mathcal{D}_{\mathrm{KL}}(q(\underline{\mathbf{h}})||q^*(\underline{\mathbf{h}})) + \mathcal{H}(q(\underline{\mathbf{w}})) + \ln Z.
\end{aligned}
$$

$\mathcal{L}(q)$ will attain its maximum when the KL-divergence vanishes, i.e.,

$$
\ln q(\underline{\mathbf{h}}) = \ln q^*(\underline{\mathbf{h}}) = \mathbb{E}_{q(\underline{\mathbf{w}})}\{\ln p(\underline{\mathbf{h}}, \underline{\mathbf{w}}, \mathcal{X})\} + \mathrm{const}. \qquad (6)
$$

Now, we write Eq. 2 as

$$
\ln p(\mathcal{X}) = \mathcal{L}(q) + \mathcal{D}_{\mathrm{KL}}(q(\underline{\mathbf{h}}, \underline{\mathbf{w}})||p(\underline{\mathbf{h}}, \underline{\mathbf{w}}|\mathcal{X})).
$$

Because KL-divergence is non-negative, we may maximize the data likelihood $p(\mathcal{X})$ by maximizing the VB lower bound $\mathcal{L}(q)$, which can be achieved by Eq. 6. A similar treatment can be applied to $q(\underline{\mathbf{w}})$.