A VARIANCE MODELING FRAMEWORK BASED ON VARIATIONAL AUTOENCODERS FOR SPEECH ENHANCEMENT

SUPPORTING DOCUMENT

Simon Leglaive[1] *Laurent Girin*[1,2] *Radu Horaud*[1]

[1]Inria Grenoble Rhône-Alpes, France [2]Univ. Grenoble Alpes, Grenoble INP, GIPSA-lab, France

This document provides the proof of Proposition 1 in [1]. We recall that we want to solve the following optimization problem:

$$\min_{\mathbf{H}_b \in \mathbb{R}_+^{K_b \times N}} \mathcal{C}(\mathbf{H}_b) = \hat{\mathcal{C}}(\mathbf{H}_b) + \breve{\mathcal{C}}(\mathbf{H}_b), \tag{1}$$

with

$$\hat{\mathcal{C}}(\mathbf{H}_b) = \frac{1}{R}\sum_{r=1}^{R}\sum_{(f,n)\in\mathbb{B}} \ln\left(g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + (\mathbf{W}_b\mathbf{H}_b)_{f,n}\right); \tag{2}$$

$$\breve{\mathcal{C}}(\mathbf{H}_b) = \frac{1}{R}\sum_{r=1}^{R}\sum_{(f,n)\in\mathbb{B}} \frac{|x_{fn}|^2}{g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + (\mathbf{W}_b\mathbf{H}_b)_{f,n}}, \tag{3}$$

and where $\mathbb{B} = \{0,...,F-1\} \times \{0,...,N-1\}$; $x_{fn} \in \mathbb{C}$; $g_n \in \mathbb{R}_+$; $\mathbf{z}_n^{(r)} \in \mathbb{R}^L$; $\sigma_f^2 : \mathbb{R}^L \mapsto \mathbb{R}_+$ and $\mathbf{W}_b \in \mathbb{R}_+^{F\times K_b}$. Similarly as in [2], we use the auxiliary function technique. Let us first recall its principle.

Definition 1 (Auxiliary function)**.** The $\mathbb{R}_+^{K_b\times N} \times \mathbb{R}_+^{K_b\times N} \mapsto \mathbb{R}_+$ mapping $\mathcal{G}(\mathbf{H}_b, \tilde{\mathbf{H}}_b)$ is an auxiliary function to $\mathcal{C}(\mathbf{H}_b)$ if and only if

$$\forall (\mathbf{H}_b, \tilde{\mathbf{H}}_b) \in \mathbb{R}_+^{K_b\times N} \times \mathbb{R}_+^{K_b\times N}, \qquad \mathcal{C}(\mathbf{H}_b) \leq \mathcal{G}(\mathbf{H}_b, \tilde{\mathbf{H}}_b); \tag{4}$$

$$\forall \mathbf{H}_b \in \mathbb{R}_+^{K_b\times N}, \qquad \mathcal{C}(\mathbf{H}_b) = \mathcal{G}(\mathbf{H}_b, \mathbf{H}_b). \tag{5}$$

In other words, $\mathcal{G}(\mathbf{H}_b, \tilde{\mathbf{H}}_b)$ is an upper bound of $\mathcal{C}(\mathbf{H}_b)$ which is tight for $\tilde{\mathbf{H}}_b = \mathbf{H}_b$. The original minimization problem can be replaced by an alternate minimization of this upper bound; from an initial point $\mathbf{H}_b^\star$ we iterate:

$$\mathbf{H}_b^\star \leftarrow \underset{\mathbf{H}_b \in \mathbb{R}_+^{K_b\times N}}{\arg\min}\ \mathcal{G}(\mathbf{H}_b, \mathbf{H}_b^\star). \tag{6}$$

This procedure corresponds to the majorize-minimize (MM) algorithm [3], which by construction leads to a monotonic decrease of $\mathcal{C}(\mathbf{H}_b)$. Moreover, its convergence properties are the same as the ones of the expectation-maximization algorithm [4].

Now let us rewrite Proposition 1 of [1] in the following equivalent form:

Proposition 1 (Auxiliary function to $\mathcal{C}(\mathbf{H}_b)$)**.**
The function $\mathcal{G}(\mathbf{H}_b, \tilde{\mathbf{H}}_b)$ defined below is an auxiliary function to $\mathcal{C}(\mathbf{H}_b)$.

$$\mathcal{G}(\mathbf{H}_b, \tilde{\mathbf{H}}_b) = \hat{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b) + \breve{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b), \tag{7}$$

with

$$\hat{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b) = \frac{1}{R}\sum_{r=1}^{R}\sum_{(f,n)\in\mathbb{B}} \ln\left(g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}\right) + \frac{(\mathbf{W}_b\mathbf{H}_b)_{f,n} - \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}}{g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}}; \tag{8}$$

$$\breve{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b) = \frac{1}{R}\sum_{r=1}^{R}\sum_{(f,n)\in\mathbb{B}} |x_{fn}|^2 \left(\frac{g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right)}{\left(g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}\right)^2} + \sum_{k=1}^{K_b} \frac{w_{b,fk}\tilde{h}_{b,kn}^2}{h_{b,kn}\left(g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}\right)^2} \right), \tag{9}$$

and where for all $(f,n) \in \mathbb{B}$, $k \in \{1,...,K_b\}$, $w_{b,fk} = (\mathbf{W}_b)_{f,k}$, $h_{b,kn} = (\mathbf{H}_b)_{k,n}$ and $\tilde{h}_{b,kn} = (\tilde{\mathbf{H}}_b)_{k,n}$.

Proof. The proof will be done in two parts.

Concave part. We first prove that $\hat{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b)$ is an auxiliary function to $\hat{\mathcal{C}}(\mathbf{H}_b)$. The condition $\hat{\mathcal{G}}(\mathbf{H}_b, \mathbf{H}_b) = \hat{\mathcal{C}}(\mathbf{H}_b)$ is trivially met. Let $\mathbf{h}_{b,n} \in \mathbb{R}_+^{K_b}$ (respectively $\tilde{\mathbf{h}}_{b,n} \in \mathbb{R}_+^{K_b}$) denote the n-th column of $\mathbf{H}_b$ (respectively $\tilde{\mathbf{H}}_b$). The criterion $\hat{\mathcal{C}}(\mathbf{H}_b)$ in equation (2) can be decomposed as

$$\hat{\mathcal{C}}(\mathbf{H}_b) = \frac{1}{R}\sum_{r=1}^{R}\sum_{(f,n)\in\mathbb{B}} \hat{\mathcal{C}}_{fn}^{(r)}(\mathbf{h}_{b,n}); \tag{10}$$

where

$$\hat{\mathcal{C}}_{fn}^{(r)}(\mathbf{h}_{b,n}) = \ln\left(g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + (\mathbf{W}_b\mathbf{H}_b)_{f,n}\right). \tag{11}$$

We prove that $\hat{\mathcal{C}}(\mathbf{H}_b) \leq \hat{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b)$ by majorizing each term $\hat{\mathcal{C}}_{fn}^{(r)}(\mathbf{h}_{b,n})$. As the composition of a concave function and a linear function, $\hat{\mathcal{C}}_{fn}^{(r)}(\mathbf{h}_{b,n})$ is a concave function, so it can be majorized by its tangent (first order Taylor expansion) at an arbitrary point $\tilde{\mathbf{h}}_{b,n} \in \mathbb{R}_+^{K_b}$:

$$\hat{\mathcal{C}}_{fn}^{(r)}(\mathbf{h}_{b,n}) \leq \hat{\mathcal{G}}_{fn}^{(r)}(\mathbf{h}_{b,n}, \tilde{\mathbf{h}}_{b,n}) := \hat{\mathcal{C}}_{fn}^{(r)}(\tilde{\mathbf{h}}_{b,n}) + \nabla^\top \hat{\mathcal{C}}_{fn}^{(r)}(\tilde{\mathbf{h}}_{b,n})(\mathbf{h}_{b,n} - \tilde{\mathbf{h}}_{b,n}), \tag{12}$$

where ∇ denotes the gradient operator. From (11), this upper bound can be further developed as follows:

$$\begin{aligned}\hat{\mathcal{G}}_{fn}^{(r)}(\mathbf{h}_{b,n}, \tilde{\mathbf{h}}_{b,n}) &= \ln\left(g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}\right) + \sum_{k=1}^{K_b} \frac{w_{b,fk}}{g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}}(h_{b,kn} - \tilde{h}_{b,kn}) \\ &= \ln\left(g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}\right) + \frac{(\mathbf{W}_b\mathbf{H}_b)_{f,n} - \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}}{g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}}.\end{aligned} \tag{13}$$

From (10), (12), (13) and (8), we have:

$$\hat{\mathcal{C}}(\mathbf{H}_b) \leq \frac{1}{R}\sum_{r=1}^{R}\sum_{(f,n)\in\mathbb{B}} \hat{\mathcal{G}}_{fn}^{(r)}(\mathbf{h}_{b,n}, \tilde{\mathbf{h}}_{b,n}) = \hat{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b), \tag{14}$$

which completes the first part of the proof.

Convex part. We now prove that $\check{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b)$ is an auxiliary function to $\check{\mathcal{C}}(\mathbf{H}_b)$. The condition $\check{\mathcal{G}}(\mathbf{H}_b, \mathbf{H}_b) = \check{\mathcal{C}}(\mathbf{H}_b)$ is trivially met. The criterion $\check{\mathcal{C}}(\mathbf{H}_b)$ in equation (3) can be decomposed as

$$\check{\mathcal{C}}(\mathbf{H}_b) = \frac{1}{R}\sum_{r=1}^{R}\sum_{(f,n)\in\mathbb{B}} |x_{fn}|^2 \check{\mathcal{C}}_{fn}^{(r)}(\mathbf{h}_{b,n}); \tag{15}$$

where

$$\check{\mathcal{C}}_{fn}^{(r)}(\mathbf{h}_{b,n}) = \frac{1}{g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + (\mathbf{W}_b\mathbf{H}_b)_{f,n}}. \tag{16}$$

We prove that $\check{\mathcal{C}}(\mathbf{H}_b) \leq \check{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b)$ by majorizing each term $\check{\mathcal{C}}_{fn}^{(r)}(\mathbf{h}_{b,n})$. Let us introduce the set of variables $\left\{\phi_{k,fn}^{(r)} \in [0,1]\right\}_{k=0}^{K_b}$ defined by:

$$\phi_{0,fn}^{(r)} = \frac{g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right)}{g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}}; \tag{17}$$

$$\phi_{k,fn}^{(r)} = \frac{w_{b,fk}\tilde{h}_{b,kn}}{g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b\tilde{\mathbf{H}}_b\right)_{f,n}}, \forall k \in \{1, ..., K_b\}. \tag{18}$$

It is straightforward to verify that $\sum_{k=0}^{K_b} \phi_{k,fn}^{(r)} = 1$. As the composition of a convex function and a linear function, $\check{\mathcal{C}}_{fn}^{(r)}(\mathbf{h}_{b,n})$ is a convex function. Therefore, using Jensen's inequality we have[1]:

$$\check{\mathcal{C}}_{fn}^{(r)}(\mathbf{h}_{b,n}) \leq \check{\mathcal{G}}_{fn}^{(r)}(\mathbf{h}_{b,n}, \tilde{\mathbf{h}}_{b,n}) := \phi_{0,fn}^{(r)} \frac{\phi_{0,fn}^{(r)}}{g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right)} + \sum_{k=1}^{K_b} \phi_{k,fn}^{(r)} \frac{\phi_{k,fn}^{(r)}}{w_{b,fk}h_{b,kn}}. \tag{19}$$

[1] Note that $\left\{\phi_{k,fn}^{(r)}\right\}_{k=0}^{K_b}$ are actually functions of $\tilde{\mathbf{h}}_{b,n}$.

Injecting (17) and (18) in (19), $\breve{\mathcal{G}}_{fn}^{(r)}(\mathbf{h}_{b,n}, \tilde{\mathbf{h}}_{b,n})$ can be further developed as follows:

$$\breve{\mathcal{G}}_{fn}^{(r)}(\mathbf{h}_{b,n}, \tilde{\mathbf{h}}_{b,n}) = \frac{g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right)}{\left(g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b \tilde{\mathbf{H}}_b\right)_{f,n}\right)^2} + \sum_{k=1}^{K_b} \frac{w_{b,fk} \tilde{h}_{b,kn}^2}{h_{b,kn}\left(g_n \sigma_f^2\left(\mathbf{z}_n^{(r)}\right) + \left(\mathbf{W}_b \tilde{\mathbf{H}}_b\right)_{f,n}\right)^2}. \tag{20}$$

From (15), (19), (20) and (9), we have:

$$\breve{\mathcal{C}}(\mathbf{H}_b) \leq \frac{1}{R} \sum_{r=1}^{R} \sum_{(f,n) \in \mathbb{B}} |x_{fn}|^2 \breve{\mathcal{G}}_{fn}^{(r)}(\mathbf{h}_{b,n}, \tilde{\mathbf{h}}_{b,n}) = \breve{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b), \tag{21}$$

which completes the second part of proof.

Finally, from (14) and (21):

$$\mathcal{C}(\mathbf{H}_b) = \hat{\mathcal{C}}(\mathbf{H}_b) + \breve{\mathcal{C}}(\mathbf{H}_b) \leq \hat{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b) + \breve{\mathcal{G}}(\mathbf{H}_b, \tilde{\mathbf{H}}_b) = \mathcal{G}(\mathbf{H}_b, \tilde{\mathbf{H}}_b), \tag{22}$$

which completes the proof.

1. REFERENCES

[1] Simon Leglaive, Laurent Girin, and Radu Horaud, "A variance modeling framework based on variational autoencoders for speech enhancement," Proc. of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2018.

[2] Cédric Févotte and Jérôme Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

[3] David R. Hunter and Kenneth Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[4] C.F. Jeff Wu, "On the convergence properties of the EM algorithm," *The Annals of statistics*, pp. 95–103, 1983.