Supplementary Material for:
Weighted Variance Variational Autoencoder for Speech Enhancement

Ali Golmakani, Mostafa Sadeghi, Xavier Alameda-Pineda, and Romain Serizel

1 Generative model

Let us assume a latent code $\mathbf{z}_t \in \mathbb{R}^L$, an observed variable $\mathbf{s}_t \in \mathbb{C}^F$, and a latent positive weight $w_t > 0$. We will assume the following decomposition:

$$p_\theta(\mathbf{s}_t, \mathbf{z}_t, w_t) = p(\mathbf{s}_t|\mathbf{z}_t, w_t)p(w_t)p(\mathbf{z}_t) \tag{1}$$

with

$$p(\mathbf{z}_t) = \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{2}$$

$$p(w_t) = \mathcal{G}(w_t; \alpha, \beta), \tag{3}$$

$$p_\theta(\mathbf{s}_t|\mathbf{z}_t, w_t) = \mathcal{N}_c\Big(\mathbf{0}, \mathrm{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))/w_t\Big). \tag{4}$$

The Gamma distribution has the following pdf:

$$\mathcal{G}(w; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{\alpha-1} \exp(-\beta w), \tag{5}$$

where $\Gamma(.)$ is the gamma function.

2 Inference

To learn the set of parameters θ, we need to compute the posterior distribution as follows:

$$\begin{aligned}
p_\theta(\mathbf{z}_t, w_t|\mathbf{s}_t) &\propto p_\theta(\mathbf{s}_t|\mathbf{z}_t, w_t)p(w_t)p(\mathbf{z}_t) && (6)\\
&\propto \frac{w_t^F}{|\mathrm{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))|} \exp\left(-w_t \mathbf{s}_t^{\mathrm{H}} \mathrm{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))^{-1}\mathbf{s}_t\right) \times && (7)\\
&\quad \frac{\beta^\alpha}{\Gamma(\alpha)} w_t^{\alpha-1} \exp(-w_t\beta) p(\mathbf{z}_t) && (8)\\
&\propto w_t^{\alpha+F-1} \exp\left(-w_t(\beta + \mathbf{s}_t^{\mathrm{H}} \mathrm{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))^{-1}\mathbf{s}_t)\right) \times && (9)\\
&\quad \frac{\beta^\alpha}{\Gamma(\alpha)} p(\mathbf{z}_t) && (10)\\
&\propto \mathcal{G}(w; \alpha_t, \beta_t) \frac{\Gamma(\alpha_t)\beta^\alpha}{\Gamma(\alpha)\beta_t^{\alpha_t}} p(\mathbf{z}_t), && (11)
\end{aligned}$$

where we have defined $\alpha_t = \alpha + F$ and $\beta_t = \beta + \mathbf{s}_t^{\mathrm{H}} \mathrm{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))^{-1}\mathbf{s}_t$.

The equation above implies that we can write the exact distribution $p_\theta(w_t|\mathbf{z}_t, \mathbf{s}_t)$ (which is a Gamma distribution with parameters α_t and β_t) but not the one of the latent code $p_\theta(\mathbf{z}_t|\mathbf{s}_t)$. This last one is going to be approximated with the encoder $p_\theta(\mathbf{z}_t|\mathbf{s}_t) \approx q_\psi(\mathbf{z}_t|\mathbf{s}_t)$.

3 Learning

We start with the usual log probability for learning the parameters:

$$\log p_\theta(\mathbf{s}_t) = \mathbb{E}_{q(w_t,\mathbf{z}_t|\mathbf{s}_t)} \left\{ \log \frac{p_\theta(\mathbf{s}_t, w_t, \mathbf{z}_t)}{q(w_t, \mathbf{z}_t|\mathbf{s}_t)} + \log \frac{q(w_t, \mathbf{z}_t|\mathbf{s}_t)}{p_\theta(w_t, \mathbf{z}_t|\mathbf{s}_t)} \right\}. \tag{12}$$

This equality holds for any distribution q. It will hold in particular for a distribution of the form: $q(w_t, \mathbf{z}_t|\mathbf{s}_t) = q_\psi(\mathbf{z}_t|\mathbf{s}_t)p_\theta(w_t|\mathbf{z}_t, \mathbf{s}_t)$.

If we use this in the above expression, we obtain:

$$\log p_\theta(\mathbf{s}_t) = \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)} \left\{ \log \frac{p_\theta(\mathbf{s}_t, w_t, \mathbf{z}_t)}{q_\psi(\mathbf{z}_t|\mathbf{s}_t)p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)} + \log \frac{q_\psi(\mathbf{z}_t|\mathbf{s}_t)p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)}{p_\theta(w_t,\mathbf{z}_t|\mathbf{s}_t)} \right\}, \tag{13}$$

which simplifies to:

$$\log p_\theta(\mathbf{s}_t) = \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)} \left\{ \log \frac{p_\theta(\mathbf{s}_t, w_t, \mathbf{z}_t)}{q_\psi(\mathbf{z}_t|\mathbf{s}_t)p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)} + \log \frac{q_\psi(\mathbf{z}_t|\mathbf{s}_t)}{p_\theta(\mathbf{z}_t|\mathbf{s}_t)} \right\}, \tag{14}$$

In the second term, $p_\theta(\mathbf{z}_t|\mathbf{s}_t)$ does not have analytic expression. As in the case of standard VAE, this part (that boils down to a KL divergence) is ignored, giving the following variational lower bound.

$$\log p_\theta(\mathbf{s}_t) \geq \mathcal{L}(\widetilde{\Phi}; \mathbf{s}_t) = \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)} \left\{ \log \frac{p_\theta(\mathbf{s}_t, w_t, \mathbf{z}_t)}{q_\psi(\mathbf{z}_t|\mathbf{s}_t)p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)} \right\} \tag{15}$$

where $\widetilde{\Phi} = \{\theta, \psi, \alpha, \beta\}$. If we develop the VLB we obtain:

$$\mathcal{L}(\widetilde{\Phi}; \mathbf{s}_t) = \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)} \left\{ \log \frac{p_\theta(\mathbf{z}_t)}{q_\psi(\mathbf{z}_t|\mathbf{s}_t)} + \mathbb{E}_{p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)} \left\{ \log \frac{p_\theta(\mathbf{s}_t, w_t|\mathbf{z}_t)}{p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)} \right\} \right\} \tag{16}$$

While the first term is the standard VAE regularisation term, the other term is quite unusual. It can actually be decomposed into an entropy term and the more classical reconstruction term:

$$\mathcal{L}(\widetilde{\Phi}; \mathbf{s}_t) = -\mathcal{D}_{\text{KL}}(q_\psi(\mathbf{z}_t|\mathbf{s}_t) \| p_\theta(\mathbf{z}_t)) + \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)}\{\mathcal{H}(p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t))\} + \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)} \{\log p_\theta(\mathbf{s}_t, w_t|\mathbf{z}_t)\} \tag{17}$$

The entropy term can be obtained from standard formulae:

$$\mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)}\{\mathcal{H}(p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t))\} = \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)}\{\alpha_t - \log \beta_t + \log \Gamma(\alpha_t) + (1-\alpha_t)\Psi(\alpha_t)\} \tag{18}$$

where Ψ is the digamma function (it will be canceled out in the derivations). The reconstruction term can also be developed.

$$\begin{aligned}
&\mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)} \{\log p_\theta(\mathbf{s}_t|w_t,\mathbf{z}_t) + \log p_\theta(w_t|\mathbf{z}_t)\} && (19)\\
&= \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)} \left\{ -\log|\text{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))| + \mathbb{E}_{p_\theta(w_t|\mathbf{z}_t,\mathbf{s}_t)} \left\{ F \log w_t - w_t \mathbf{s}_t^{\text{H}} \text{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))^{-1}\mathbf{s}_t + \alpha \log \beta - \log \Gamma(\alpha) + \right.\right. && (20)\\
&\left.\left.(\alpha - 1)\log w_t - \beta w_t \right\}\right\} && (21)\\
&= \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)} \left\{ -\log|\text{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))| + F(\Psi(\alpha_t) - \log \beta_t) - \frac{\alpha_t}{\beta_t} \mathbf{s}_t^{\text{H}} \text{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))^{-1}\mathbf{s}_t + \alpha \log \beta - \log \Gamma(\alpha) + \right. && (22)\\
&\left.(\alpha - 1)(\Psi(\alpha_t) - \log \beta_t) - \beta \frac{\alpha_t}{\beta_t} \right\} && (23)\\
&= \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)} \left\{ -\log|\text{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))| + \alpha \log \beta - \log \Gamma(\alpha) + (\alpha_t - 1)(\psi(\alpha_t) - \log \beta_t) - \alpha_t \right\} && (24)
\end{aligned}$$

When adding up the entropy and the reconstruction losses, three terms simplify, namely: α_t, $\log \beta_t$ and $(1-\alpha_t)\Psi(\alpha_t)$ since they have different signs in the entropy and reconstruction losses.

$$\mathcal{L}(\widetilde{\Phi}; \mathbf{s}_t) = -\mathcal{D}_{\text{KL}}(q_\psi(\mathbf{z}_t|\mathbf{s}_t) \| p_\theta(\mathbf{z}_t)) + \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)}\{-\log|\text{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))| + \alpha \log \beta - \log \Gamma(\alpha) - \alpha_t \log \beta_t + \log \Gamma(\alpha_t)\} \tag{25}$$

We now use the following property of the Gamma function $\log \Gamma(t+1) = \log t + \log \Gamma(t)$. We also recall that $\alpha_t = \alpha + F$. Applying the property F times we obtain:

$$\log \Gamma(\alpha + F) = \log \Gamma(\alpha) + \sum_{f=0}^{F-1} \log(\alpha + f) \tag{26}$$

This simplifies even further the VLB:

$$\mathcal{L}(\widetilde{\Phi}; \mathbf{s}_t) = -\mathcal{D}_{\text{KL}}(q_\psi(\mathbf{z}_t|\mathbf{s}_t) \| p_\theta(\mathbf{z}_t)) + \mathbb{E}_{q_\psi(\mathbf{z}_t|\mathbf{s}_t)} \left\{ -\log|\text{diag}(\boldsymbol{\sigma}_\theta^2(\mathbf{z}_t))| + \alpha \log \beta - \alpha_t \log \beta_t + \sum_{f=0}^{F} \log(\alpha + f)) \right\}. \tag{27}$$