On Perceptual Lossy Compression: The Cost of Perceptual Reconstruction and An Optimal Training Framework Supplementary Material

This supplemental material first provides the proof of lemma and theorem 1, and then gives the derivation of equation (20). Finally, degeneration problem is discussed and we provide a pre-training trick to solve it.

A. Proof of Lemma 1

Suppose $\mathbf{B}^*$ is an optimal solution to (10), which satisfies $\langle \mathbf{W}, \mathbf{B}^* \rangle \leq D$ and $\sum_{i=1}^{m} b_{ij}^* = \sum_{i=1}^{m} b_{ji}^* = p(X = x_j), 1 \leq j \leq m$. Since Δ is symmetric, $w_{ij} = \Delta(x_i, x_j) = \Delta(x_j, x_i) = w_{ji}$ holds for any $1 \leq i, j \leq m$, which means $\mathbf{W} = \mathbf{W}^T$. Thus, we have

$$\left\langle \mathbf{W}, \mathbf{B}^{*T} \right\rangle = \left\langle \mathbf{W}^T, \mathbf{B}^{*T} \right\rangle = \langle \mathbf{W}, \mathbf{B}^* \rangle \leq D. \tag{A.1}$$

Denote the (i, j)-th element of $\mathbf{B}^{*T}$ by b'_{ij}, it follows that $b'_{ij} = b^*_{ji}$, so that

$$\begin{aligned} \sum_{i=1}^{m} b'_{ij} = \sum_{i=1}^{m} b^*_{ji} = p(X = x_j) \\ \sum_{i=1}^{m} b'_{ji} = \sum_{i=1}^{m} b^*_{ij} = p(X = x_j). \end{aligned} \tag{A.2}$$

Then, it is easy to see that $\mathbf{B}^{*T}$ is also a feasible solution to (10). Meanwhile, it can be justified that $\mathbf{B}^{*T}$ is also an optimal solution to (10) since the objective satisfies

$$\begin{aligned} G_{p_X}(\mathbf{B}^{*T}) &= 2H(X) + \sum_{i=1}^{m} \sum_{j=1}^{m} b'_{ij} \log b'_{ij} \\ &= 2H(X) + \sum_{i=1}^{m} \sum_{j=1}^{m} b^*_{ji} \log b^*_{ji} \\ &= G_{p_X}(\mathbf{B}^*). \end{aligned} \tag{A.3}$$

Next, denote $\mathbf{B}_0 := (\mathbf{B}^* + \mathbf{B}^{*T})/2$, we show that $G_{p_X}(\mathbf{B}_0) = G_{p_X}(\mathbf{B}^*)$. First, $\mathbf{B}_0$ is a feasible solution of (10) as it satisfies the constraints

$$\begin{aligned} \langle \mathbf{W}, \mathbf{B}_0 \rangle = \left\langle \mathbf{W}^T, \frac{\mathbf{B}^* + \mathbf{B}^{*T}}{2} \right\rangle = \frac{\langle \mathbf{W}, \mathbf{B}^* \rangle + \left\langle \mathbf{W}, \mathbf{B}^{*T} \right\rangle}{2} \leq D \\ \sum_{i=1}^{m} b_{0ij} = \sum_{i=1}^{m} \frac{b^*_{ij} + b'_{ij}}{2} = \frac{1}{2} \left(\sum_{i=1}^{m} b^*_{ij} + \sum_{i=1}^{m} b'_{ij} \right) = p(X = x_j) \\ \sum_{i=1}^{m} b_{0ji} = \sum_{i=1}^{m} \frac{b^*_{ji} + b'_{ji}}{2} = \frac{1}{2} \left(\sum_{i=1}^{m} b^*_{ji} + \sum_{i=1}^{m} b'_{ji} \right) = p(X = x_j). \end{aligned} \tag{A.4}$$

Meanwhile, the objective function $G_{p_X}(\mathbf{B}_0)$ can be expressed as

$$\begin{aligned} G_{p_X}(\mathbf{B}_0) &= 2H(X) + \sum_{i=1}^{m}\sum_{j=1}^{m} b_{0ij}\log b_{0ij} \\ &= 2H(X) + \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{b^*_{ij}+b'_{ij}}{2}\log\frac{b^*_{ij}+b'_{ij}}{2}. \end{aligned} \tag{A.5}$$

Notice that the function $f(x) = x\log x$ is strictly convex in $(0,1)$. Thus we have

$$\frac{b^*_{ij}+b'_{ij}}{2}\log\frac{b^*_{ij}+b'_{ij}}{2} \le \frac{1}{2}\left(b^*_{ij}\log b^*_{ij} + b'_{ij}\log b'_{ij}\right), \tag{A.6}$$

where the equality holds if and only if $b^*_{ij} = b'_{ij}$. Then, it follows that

$$\begin{aligned} G_{p_X}(\mathbf{B}_0) &= 2H(X) + \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{b^*_{ij}+b'_{ij}}{2}\log\frac{b^*_{ij}+b'_{ij}}{2} \\ &\le \frac{1}{2}\left[\left(2H(X) + \sum_{i=1}^{m}\sum_{j=1}^{m} b^*_{ij}\log b^*_{ij}\right) + \left(2H(X) + \sum_{i=1}^{m}\sum_{j=1}^{m} b'_{ij}\log b'_{ij}\right)\right] \\ &= \frac{1}{2}\left[G_{p_X}(\mathbf{B}^*) + G_{p_X}(\mathbf{B}^{*T})\right] = G_{p_X}(\mathbf{B}^*). \end{aligned} \tag{A.7}$$

Recall that $\mathbf{B}^*$ is an optimal solution, hence $G_{p_X}(\mathbf{B}^*) \le G_{p_X}(\mathbf{B}_0)$, which together with (A.7) leads to $G_{p_X}(\mathbf{B}_0) = G_{p_X}(\mathbf{B}^*)$. Thus, $\mathbf{B}_0$ is an optimal solution and for any $1 \le i,j \le m$ we have

$$\frac{b^*_{ij}+b'_{ij}}{2}\log\frac{b^*_{ij}+b'_{ij}}{2} = \frac{1}{2}\left(b^*_{ij}\log b^*_{ij} + b'_{ij}\log b'_{ij}\right), \tag{A.8}$$

Furthermore, since $f(x) = x\log x$ is strictly convex in $(0,1)$, we have $b'_{ij} = b^*_{ij}$ for any $1 \le i,j \le m$ and hence $\mathbf{B}^* = \mathbf{B}^{*T}$, which finally results in Lemma 1.

B. Proof of Theorem 1

Let X be a memoryless stationary source, $Y = (X_1, X_2, ..., X_t)$ be a source sequence of length t, L and Q be the encoder and decoder, respectively, with which the compressed representation is $Z = L(Y)$ and the output of the encoder is $\hat{Y} = Q(Z)$. Since $F_t(D,0)$ defined in (15) is non-increasing on D, in the case of squared-error distortion, we consider its inverse form for convenience as

$$\begin{aligned} &\min_{L,Q} \frac{1}{t}\mathbb{E}\left[\left\|Y - \hat{Y}\right\|^2\right] \\ &s.t. \quad Z = L(Y), \hat{Y} = Q(Z), \\ &\qquad\;\; H(Z) \le tR, d(p_Y, p_{\hat{Y}}) \le 0, \end{aligned} \tag{B.1}$$

which minimizes the MSE distortion under constraints on the average bit-rate and distribution divergence (perception quality).

For convenience in the sequel analysis, we define the joint distribution matrix of Y and Z as $\mathbf{L} \in \mathbb{R}^{m\times n}$ with the (i,j)-th element being $l_{i,j} = p_{Y,Z}(y_i, z_j)$. Similarly, we define the joint distribution matrix of $\hat{Y}$ and Z as $\mathbf{Q} \in \mathbb{R}^{m\times n}$ with the (i,j)-th element being $q_{i,j} = p_{\hat{Y},Z}(y_i, z_j)$. In fact, $\mathbf{L}$ and $\mathbf{Q}$ are the joint distribution matrices of the encoder and decoder, respectively.

Next we show that for any optimal encoder-decoder pair (L^*, Q^*) to (B.1) with joint distribution matrices $(\mathbf{L}^*, \mathbf{Q}^*)$, the encoder-decoder pairs with joint distribution matrices $(\mathbf{L}^*, \mathbf{L}^*)$ and $(\mathbf{Q}^*, \mathbf{Q}^*)$ are also optimal to(B.1).

First, for an optimal encoder-decoder pair (L^*, Q^*) to (B.1), we have $H(L^*(Y)) \leq tR$. Let the alphabet of Y be $\{y_1, y_2, ..., y_m\}$ and the alphabet of Z be $\{z_1, z_2, ..., z_n\}$, and suppose that $p(L^*(Y) = z_j) = h_j, 1 \leq j \leq n$. Then we consider the following formulation

$$\begin{aligned} &\min_{L,Q} \frac{1}{t}\mathbb{E}[\left\|Y - \hat{Y}\right\|^2] \\ &s.t. \quad Z = L(Y),\ \hat{Y} = Q(Z),\ d(p_Y, p_{\hat{Y}}) \leq 0, \\ &\qquad\ p_Z(z_j) = h_j,\ 1 \leq j \leq n. \end{aligned} \tag{B.2}$$

It is easy to see that the feasible region of problem (B.2) is a subset of the feasible region of problem (B.1) and (L^*, Q^*) is optimal to both (B.1) and (B.2). Thus any optimal solution of problem (B.2) must be an optimal solution of problem (B.1). Therefore, to justify that the encoder-decoder pairs with joint distribution matrices $(\mathbf{L}^*, \mathbf{L}^*)$ and $(\mathbf{Q}^*, \mathbf{Q}^*)$ are optimal to (B.1), it is enough to justify that $(\mathbf{L}^*, \mathbf{L}^*)$ and $(\mathbf{Q}^*, \mathbf{Q}^*)$ are optimal to (B.2).

Obviously, the constraint $p_Z(z_j) = h_j$ in (B.2) can be expressed as $\sum\limits_i l_{i,j} = \sum\limits_i q_{i,j} = h_j$. Besides, since Y and $\hat{Y}$ have the same distribution under perfect perception constraint, the constraint $d(p_Y, p_{\hat{Y}}) \leq 0$ can be expressed as $\sum\limits_j l_{i,j} = \sum\limits_j q_{i,j} = p_Y(y_i)$. Now, we rewrite the objective function of (B.2) as

$$\begin{aligned} \frac{1}{t}\mathbb{E}[\left\|Y - \hat{Y}\right\|^2 &= \frac{1}{t}\sum_{y,\hat{y}} p_{Y,\hat{Y}}(y,\hat{y})\|y - \hat{y}\|^2 \\ &= \frac{1}{t}[\sum_y p_Y(y)y^T y + \sum_{\hat{y}} p_{\hat{Y}}(\hat{y})\hat{y}^T\hat{y} - 2\sum_{y,\hat{y}} p_{Y,\hat{Y}}(y,\hat{y})y^T\hat{y}], \end{aligned} \tag{B.3}$$

where $\sum\limits_y p_Y(y)y^T y$ is constant for fixed source, and $\sum\limits_{\hat{y}} p_{\hat{Y}}(\hat{y})\hat{y}^T\hat{y} = \sum\limits_y p_Y(y)y^T y$ for the perfect perception constraint. Hence, minimizing the objective function of (B.2) is to equivalent to maximizing $\sum\limits_{y,\hat{y}} p_{Y,\hat{Y}}(y,\hat{y})y^T\hat{y}$, for which we have

$$\begin{aligned} \sum_{y,\hat{y}} p_{Y,\hat{Y}}(y,\hat{y})y^T\hat{y} &= \sum_{y,\hat{y},z} p_{Y,\hat{Y},Z}(y,\hat{y},z)y^T\hat{y} \\ &\overset{(a)}{=} \sum_{y,\hat{y},z} p_Z(z)p_{Y|Z}(y|z)p_{\hat{Y}|Z}(\hat{y}|z)y^T\hat{y} \\ &= \sum_j h_j[\sum_i p_{Y|Z}(y_i|z_j){y_i}^T \sum_k p_{\hat{Y}|Z}(\hat{y}_k|z_j)\hat{y}_k] \\ &= \sum_j h_j\mathbb{E}(Y|Z = z_j)^T\mathbb{E}(\hat{Y}|Z = z_j) \end{aligned} \tag{B.4}$$

where in (a) we used the property of Markov chain $Y \to Z \to \hat{Y}$ that Y and $\hat{Y}$ are independent under condition Z. Hence, using the joint distribution representation $(\mathbf{L}, \mathbf{Q})$ of the encoder-decoder pair (L, Q), the problem (B.2) can be equivalently reformulated as

$$\begin{aligned} &\max_{\mathbf{L},\mathbf{Q}} \sum_j h_j\mathbb{E}(Y|Z = z_j)^T\mathbb{E}(\hat{Y}|Z = z_j) \\ &s.t. \quad \sum_i l_{i,j} = \sum_i q_{i,j} = h_j, 1 \leq j \leq n \\ &\qquad\ \sum_j l_{i,j} = \sum_j q_{i,j} = p_Y(y_i), 1 \leq i \leq m \\ &\qquad\ 0 \leq l_{i,j}, q_{i,j} \leq 1, 1 \leq i \leq m, 1 \leq j \leq n. \end{aligned} \tag{B.5}$$

Accordingly, the joint distribution matrix pair $(\mathbf{L}^*, \mathbf{Q}^*)$ corresponding to the optimal encoder-decoder pair (L^*, Q^*) is an optimal solution to (B.5). Recall that $\mathbf{L}$ is the probability matrix of $p_{Y,Z}$ and $\mathbf{Q}$ is the probability matrix of $p_{\hat{Y},Z}$, hence

$\mathbb{E}(Y|Z=z_j)$ and $\mathbb{E}(\hat{Y}|Z=z_j)$ are functions of $\mathbf{L}$ and $\mathbf{Q}$. Define

$$f_j(\mathbf{L}) := \mathbb{E}(Y|Z=z_j) = \sum_i \frac{l_{i,j}}{h_j} y_i, \tag{B.6}$$

$$f_j(\mathbf{Q}) := \mathbb{E}(Y|Z=z_j) = \sum_i \frac{q_{i,j}}{h_j} y_i, \tag{B.7}$$

and

$$\begin{aligned} F(\mathbf{L},\mathbf{Q}) :&= \sum_j h_j \mathbb{E}\left(Y|Z=z_j\right)^T \mathbb{E}\left(\hat{Y}|Z=z_j\right) \\ &= \sum_j h_j f_j(\mathbf{L})^T f_j(\mathbf{Q}). \end{aligned} \tag{B.8}$$

Next, we show $(\mathbf{L}^*,\mathbf{L}^*)$ and $(\mathbf{Q}^*,\mathbf{Q}^*)$ are also optimal solutions to (B.5).

Because $(\mathbf{L}^*,\mathbf{Q}^*)$ is an optimal solution to (B.5), the optimal objective value of (B.5) is $F(\mathbf{L}^*,\mathbf{Q}^*)$. Since the constraints of $\mathbf{L}$ and $\mathbf{Q}$ are the same, it is easy to see that $(\mathbf{L}^*,\mathbf{L}^*)$ and $(\mathbf{Q}^*,\mathbf{Q}^*)$ are both feasible solutions to (B.5). Meanwhile, we have

$$F(\mathbf{L}^*,\mathbf{L}^*) = \sum_j h_j f_j(\mathbf{L}^*)^T f_j(\mathbf{L}^*) = \sum_j h_j \|f_j(\mathbf{L}^*)\|^2, \tag{B.9}$$

$$F(\mathbf{Q}^*,\mathbf{Q}^*) = \sum_j h_j f_j(\mathbf{Q}^*)^T f_j(\mathbf{Q}^*) = \sum_j h_j \|f_j(\mathbf{Q}^*)\|^2. \tag{B.10}$$

Summing up (B.9) and (B.10) yields

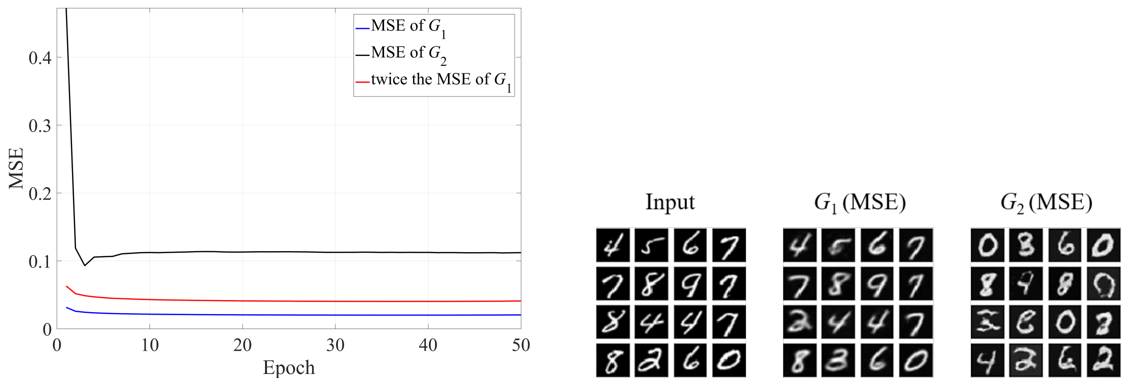
$$\begin{aligned} F(\mathbf{L}^*,\mathbf{L}^*) + F(\mathbf{Q}^*,\mathbf{Q}^*) &= \sum_j h_j \left(\|f_j(\mathbf{L}^*)\|^2 + \|f_j(\mathbf{Q}^*)\|^2\right) \\ &\geq 2\sum_j h_j \,\|f_j(\mathbf{L}^*)\|\,\|f_j(\mathbf{Q}^*)\| \\ &\overset{(b)}{\geq} 2\sum_j h_j |f_j(\mathbf{L}^*)^T f_j(\mathbf{Q}^*)| \\ &\overset{(c)}{\geq} 2\sum_j h_j f_j(\mathbf{L}^*)^T f_j(\mathbf{Q}^*) \\ &= 2F(\mathbf{L}^*,\mathbf{Q}^*), \end{aligned} \tag{B.11}$$

where in (b) we used the Cauchy inequality and (c) is due to the non-negativity of h_j. Since $(\mathbf{L}^*,\mathbf{Q}^*)$ is an optimal solution to (B.5), $F(\mathbf{L}^*,\mathbf{Q}^*) \geq F(\mathbf{L},\mathbf{Q})$ holds for any $(\mathbf{L},\mathbf{Q})$ under the constraint of (B.5), which together with (B.11) implies $F(\mathbf{L}^*,\mathbf{L}^*) = F(\mathbf{Q}^*,\mathbf{Q}^*) = F(\mathbf{L}^*,\mathbf{Q}^*)$. Therefore, $(\mathbf{L}^*,\mathbf{L}^*)$ and $(\mathbf{Q}^*,\mathbf{Q}^*)$ are also optimal solutions to (B.5)

Thus, for any source length t, there exist optimal solutions to (B.5) satisfying

$$p_{Y,Z} = p_{\hat{Y},Z},\ p_{Y|Z} = p_{\hat{Y}|Z}, \tag{B.12}$$

which finally results in Theorem 1.



(a) MSE loss versus training epoch

(b) Visual comparison of the output

Figure 1. **Illustration of a typical degeneration case.**

C. Derivation of equation (20)

Equation (20) can be straightforwardly derived as

$$\begin{aligned}
\frac{1}{t}\mathbb{E}[\left\|Y-\hat{Y}\right\|^2 &= \sum_{y,\hat{y}} p_{Y,\hat{Y}}(y,\hat{y})\|y-\hat{y}\|^2 \\
&\overset{(d)}{=} \frac{1}{t}\sum_{y,\hat{y},z} p_{Y,\hat{Y},Z}(y,\hat{y},z)\left\|y-\mathbb{E}[Y|z]+\mathbb{E}[\hat{Y}|z]-\hat{y}\right\|^2 \\
&\overset{(e)}{=} \frac{1}{t}\sum_{y,z} p_{Y,Z}(y,z)\|y-\mathbb{E}[Y|z]\|^2 + \frac{1}{t}\sum_{\hat{y},z} p_{\hat{Y},Z}(\hat{y},z)\left\|\mathbb{E}[\hat{Y}|z]-\hat{y}\right\|^2 \\
&\overset{(f)}{=} \frac{2}{t}\mathbb{E}\left[\|Y-\mathbb{E}[Y|Z]\|^2|Z\right],
\end{aligned} \tag{C.1}$$

where (d) is due to $\mathbb{E}[Y|Z]=\mathbb{E}[\hat{Y}|Z]$, (e) is due to

$$\begin{aligned}
\sum_{y} p_{Y|Z}(y|z)(y-\mathbb{E}[Y|Z]) &= \sum_{y} p_{Y|Z}(y|z)y-\mathbb{E}[Y|Z] \\
&= \mathbb{E}[Y|Z]-\mathbb{E}[Y|Z]=0
\end{aligned} \tag{C.2}$$

$$\begin{aligned}
\sum_{\hat{y}} p_{\hat{Y}|Z}(\hat{y}|z)(\hat{y}-\mathbb{E}[\hat{Y}|Z]) &= \sum_{\hat{y}} p_{\hat{Y}|Z}(\hat{y}|z)\hat{y}-\mathbb{E}[\hat{Y}|Z] \\
&= \mathbb{E}[\hat{Y}|Z]-\mathbb{E}[\hat{Y}|Z]=0
\end{aligned} \tag{C.3}$$

and (f) is due to the same distribution of Y and $\hat{Y}$.

D. Degenerate problem

Figure 1 shows a degeneration case in training G_2, where the MSE of G_2 converges to a value deviates largely from the 2-fold MSE of G_1. From Fig. 7(b), while the output numbers of G_1 are correct, those of G_2 are incorrect though more clear. It means that the bit stream from E contains enough information for correctly reconstructing the numbers, but the trained model G_2 tends to generate numbers randomly. This problem is typically encountered in adversarial training, due to that the the alternating training procedure converges to a poor point. To address this problem, we pre-train the discriminator J to discriminate between $(x_i, E(x_i))$ and $(x_j, E(x_i))$ with $i \neq j$, where x_i and x_j are samples of X. Intensive experiments show that this strategy can effectively reduce the occurrence of the degeneration problem.