Errata – An Introduction to Single-User Information Theory

F. Alajaji and P.-N. Chen

August 24 2018

• p. 6, Theorem 2.1: replace “p ∈ [0, 1]” with “p ∈ (0, 1]” and replace “0 ≤ p ≤ 1” with “0 < p ≤ 1”
• p. 7, lines 8-10: replace
  \[
  k_r \leq I(1/2) I(1/n) \leq k + 1
  \]
  with
  \[
  k_r \leq I(1/2) I(1/n) < k + 1
  \]
  and replace
  \[
  \log_b n^k \leq \log_b 2^r \leq \log_b n^{k+1} \quad \Leftrightarrow \quad \frac{k}{r} \leq \frac{\log_b(2)}{\log_b(n)} \leq \frac{k + 1}{r}
  \]
  with
  \[
  \log_b n^k \leq \log_b 2^r < \log_b n^{k+1} \quad \Leftrightarrow \quad \frac{k}{r} \leq \frac{\log_b(2)}{\log_b(n)} < \frac{k + 1}{r}
  \]
• p. 10, line above Lemma 2.4: a space should be inserted before “(its proof is left as an exercise)”
• p. 48, Problem 2.15: assume that \(X\) and \(\hat{X}\) have a common alphabet \(\mathcal{X}\) and that \(Z\) and \(\hat{Z}\) have a common alphabet \(\mathcal{Z}\).
• p. 65, three lines above Observation 3.7: the strict inequality should be equality
• p. 82, lines 10-11: replace “\(\ell_{\text{max}}\) should be less than” with “\(\ell_{\text{max}}\) should be no larger than”
• p. 90, line 6 of Observation 3.35: replace “one can get” with “one may get”
• p. 97, line 9 of item 2: “\(L = 3\)” should be “\(L = 7\)”
• p. 101, Problem 3.12: source symbols \(x\) should be replaced with sourcewords \(x^n\) as in Theorem 3.27 (but the upper bound on the average code rate is unchanged).
• p. 114, line 2 of Definition 4.5: the second “code” is redundant
• p. 127, line 13 of Section 4.4: after “capacity of the BEC” add “(see Example 4.22 in Section 4.5)”
• p. 149, line 4 of the third item: the linebreak should be removed.
• p. 172, Definition 5.11: a logarithm is missing in the expectation; i.e., we have \(D(X\|Y) = E \left[ \log_2 \frac{f_X(X)}{f_Y(X)} \right] \).
• p. 180, line 3 of Theorem 5.20: replace “\(S_{X^n} = \mathbb{R}^n\)” with “\(S_{X^n} \subseteq \mathbb{R}^n\)”
• p. 181: line 1 of the scalar case, replace “\(S_X = \mathbb{R}\)” with “\(S_X \subseteq \mathbb{R}\)”. Also in the proof, the three integrals should be over “\(S_X\)” instead of “\(\mathbb{R}\)”
• p. 203, line 5 of Observation 5.39: “radom fading” should be “random fading”
• p. 207, first line after (5.7.3): replace “(or equivalently (5.7.3))” with “(or equivalently (5.7.3))”
• p. 225: the logarithm in Definition 6.11 should be in base 2
• p. 234, bottom line: the logarithm should be in base 2
• p. 246, the upper bound result of Theorem 6.29 should be “\(\log_2 \left( \frac{1}{D} + 1 \right) \)” instead of “\(\log_2 \frac{1}{D} \)” Thus for the sake of completeness, the introductory paragraph of Section 6.4.3, Theorem 6.29 and its proof are revised as follows.

We herein focus on the rate-distortion function of continuous memoryless sources under the absolute error distortion measure. In particular, we provide the expression of the rate-distortion function for Laplacian sources with parameter \(\lambda\) (i.e., with variance \(2\lambda^2\)) and derive an upper bound on the rate-distortion function of arbitrary zero-mean real-valued
sources with absolute mean \( \lambda \) (i.e., \( E||Z|| = \lambda \)). When \( \lambda/D \gg 1 \), the upper bound approaches the rate-distortion function of Laplacian sources; hence in this low-distortion regime, Laplacian sources maximize the rate-distortion function (while Theorem 6.26 shows that Gaussian sources maximize the rate-distortion function under the squared error distortion measure for all distortion values). It is worth pointing out that in image coding applications, the Laplacian distribution is a good model to approximate the statistics of transform coefficients such as discrete cosine and wavelet transform coefficients [315, 375]. Finally, analogously to Theorem 6.27, we obtain a Shannon lower bound on the rate-distortion function under the absolute error distortion.

**Theorem 6.29** Under the additive absolute error distortion measure, namely
\[
\rho_n(z^n, \hat{z}^n) = \sum_{i=1}^{n} |z_i - \hat{z}_i|,
\]
the rate-distortion function of a memoryless Laplacian source \( \{Z_i\} \) with mean zero and parameter \( \lambda > 0 \) (variance \( 2\lambda^2 \) and pdf \( f_Z(z) = \frac{1}{2\lambda} e^{-\frac{|z|}{\lambda}}, z \in \mathbb{R} \)) is given by
\[
R(D) = \begin{cases} 
\log_2 \left( \frac{\lambda}{D} \right), & \text{for } 0 < D \leq \lambda \\
0, & \text{for } D > \lambda.
\end{cases}
\]
Furthermore, the rate-distortion function of any continuous memoryless source \( \{Z_i\} \) with a pdf of support \( \mathbb{R} \), zero mean and \( E||Z|| = \lambda \) (where \( \lambda > 0 \) is a fixed parameter) satisfies
\[
R(D) \leq \log_2 \left( \frac{\lambda}{D} + 1 \right) \text{ for } 0 < D \leq \lambda.
\]

**Proof:** For \( 0 < D \leq \lambda \), a zero-mean Laplacian \( Z \) with parameter \( \lambda \) and any \( f_{\hat{Z}|Z} \) satisfying \( E||Z - \hat{Z}|| \leq D \),
\[
I(Z; \hat{Z}) = h(Z) - h(Z|\hat{Z}) \\
= \log_2 (2e\lambda) - h(Z - \hat{Z}|\hat{Z}) \\
\geq \log_2 (2e\lambda) - h(Z - \hat{Z}) \quad \text{(by Lemma 5.14)} \\
\geq \log_2 (2e\lambda) - \log_2 (2eD) \\
= \log_2 \left( \frac{\lambda}{D} \right),
\]
where the last inequality follows since \( h(Z - \hat{Z}) \leq \log_2 (2eE||Z - \hat{Z}||) \leq \log_2 (2eD) \) by Observation 5.21 and the fact that \( E||Z - \hat{Z}|| \leq D \). Subject to \( D \leq \lambda \), we can choose independent \( V \) and \( \hat{Z} \) such that \( Z = V + \hat{Z} \) and \( V \) is a zero-mean Laplacian random variable with parameter \( D \) (i.e., \( E||V|| = D \)).\(^1\) Thus, \( h(Z - \hat{Z}|\hat{Z}) = h(V|\hat{Z}) = h(V) = \log_2 (2eD) \) and the lower bound \( \log_2 (\lambda/D) \) is achieved.

For \( D > \lambda \), let \( \hat{Z} \) satisfy \( \Pr(\hat{Z} = 0) = 1 \) and be independent of \( Z \). Then \( E||Z - \hat{Z}|| \leq E||Z|| + E||\hat{Z}|| = \lambda < D \).

For this choice of \( \hat{Z} \), \( R(D) \leq I(Z; \hat{Z}) = 0 \) and hence \( R(D) = 0 \). This completes the derivation of \( R(D) \) for a Laplacian \( Z \).

We next prove the upper bound on \( R(D) \) for an arbitrary \( Z \) with mean zero and \( E||Z|| = \lambda \). Since
\[
R(D) = \min_{f_{\hat{Z}|Z}: E||Z - \hat{Z}|| \leq D} I(Z; \hat{Z}),
\]
we have that for any \( f_{\hat{Z}|Z} \) satisfying the distortion constraint,
\[
R(D) \leq I(Z; \hat{Z}) = I(f_{Z}, f_{\hat{Z}|Z}).
\]
\(^1\)Indeed, with a zero-mean Laplacian \( Z \) and \( Z = V + \hat{Z} \), where \( V \) and \( \hat{Z} \) are independent of each other, we can write
\[
\frac{1}{1 + \lambda^2 t^2} = \frac{1}{1 + D^2 t^2} \quad \Rightarrow \quad \phi_{\hat{Z}}(t) = \phi_{\hat{Z}}(t) = \frac{1 + D^2 t^2}{1 + \lambda^2 t^2} = \frac{D^2}{\lambda^2} + \left( 1 - \frac{D^2}{\lambda^2} \right) \frac{1}{1 + \lambda^2 t^2},
\]
where \( \phi_{\hat{Z}}(t) = E[e^{jt\hat{Z}}] \) is the characteristic function of \( \hat{Z} \) (where \( j = \sqrt{-1} \)). Thus \( \hat{Z} \) equals zero with probability \( D^2/\lambda^2 \) and is zero-mean Laplacian distributed with parameter \( \lambda \) with probability \( 1 - D^2/\lambda^2 \).
For \(0 < D \leq \lambda\), choose \(\hat{Z} = Z + W\), where \(W\) is a zero-mean Laplacian random variable that is independent of \(Z\) and that satisfies \(E[|W|] = D\). Thus

\[E[|\hat{Z}|] = E[|Z + W|] \leq E[|Z|] + E[|W|] = \lambda + D\]

and


Hence this choice of \(\hat{Z}\) satisfies the distortion constraint. We can therefore write for this \(\hat{Z}\) that

\[R(D) \leq I(Z; \hat{Z}) = h(\hat{Z}) - h(\hat{Z}|Z)
= h(\hat{Z}) - h(W + Z | Z)
= h(\hat{Z}) - h(W | Z)
= h(\hat{Z}) - h(W) \quad \text{(by independence of } Z \text{ and } W)
= h(\hat{Z}) - \log_2(2e D)
\leq \log_2[2e(\lambda + D)] - \log_2(2e D) \quad \text{(by Observation 5.21)}
= \log_2 \left(\frac{\lambda}{D} + 1\right).\]

\(\square\)

**Note:** (Tighter upper bounds) A tighter upper bound than \(\log_2 \left(\frac{\lambda}{D} + 1\right)\) can be shown using a similar proof with the exception that \(E[|\hat{Z}|]\) above is determined exactly; it is as follows:

\[R(D) \leq \log_2 \left( g \left( \frac{\lambda}{D} \right) \right) \quad \text{for } 0 < D \leq \lambda,\]

where, setting \(U = Z/\lambda\) (with pdf \(f_U(u) = \lambda f_Z(\lambda u), u \in \mathbb{R}\)), the function

\[g(a) := \int_{-\infty}^{\infty} f_U(u) \int_{-\infty}^{\infty} |s + au| \frac{1}{2} e^{-|s|} ds du, \quad a \geq 1\]

can be numerically computed. Furthermore, it can be shown that if the source has a symmetric pdf (i.e., \(f_Z(z) = f_Z(-z)\) for all \(z \in \mathbb{R}\)), then \(g(a) \leq g(1) + a - 1\) for \(a \geq 1\); resulting in the following simpler upper bound (that is still tighter than \(\log_2 \left(\frac{\lambda}{D} + 1\right)\)):

\[R(D) \leq \log_2 \left( \frac{\lambda}{D} + b \right) \quad \text{for } 0 \leq D \leq \lambda,\]

where \(b := 2 \int_{0}^{\infty} e^{-u} f_U(u) du \leq 1\). For example, if \(Z\) is uniform over \([-2\lambda, 2\lambda]\), then \(b = \frac{1}{2}(1 - e^{-2}) = 0.432\).

- p. 247, line 4 of Observation 6.31: replace “\(d(\cdot, \cdot)\)” with “\(d(\cdot)\)”
- p. 252, 257 & 258, Tables 6.1-6-3: The top of the last column should be \(D_{SL} = 10^{-2}\).
- p. 265, Property 4 of Property A.10: replace “\((\forall A \in \mathbb{R})\)” with “\((\forall L \in \mathbb{R})\)”
- p. 282: the last line should end with a period instead of a comma
- p. 288, statement in parentheses in Theorem B.13: replace “\(^-\)” with “\(\mu\)”
- p. 292, Theorem B.16 and footnote 11: replace “\(\mathcal{O} \in \mathbb{R}^m\)” with “\(\mathcal{O} \subset \mathbb{R}^m\)”