SUMMARY OF SHANNON DISTORTION-RATE THEORY[1]

Consider a stationary source X with $f_k(\underline{x})$ as its kth-order pdf.

Recall the following OPTA function definitions:

$\delta(k,R)$ = least dist'n of k-dim'l fixed-rate VQ's w. rate $\leq R$

$\delta(k,n,R)$ = least dist'n of k-dim'l VQ's w. nth-order block lossless coding and rate $\leq R$

$$\delta(R) = \inf_k \delta(k,R) = = \inf_{k,n} \delta(k,n,R)$$

= least dist'n of VQ's with rate $\leq R$, any dimension and fixed- or variable-rate coding

These functions describe the best possible performance of VQ's.

High-Resolution Theory enabled us to find concrete formulas for them (the Zador-Gersho formulas) for the case that R is large.

Shannon's distortion-rate theory enables one to find $\delta(R)$ for ANY value of R. However, it does not allow us to find $\delta(k,R)$ or $\delta(k,n,R)$, not even for some R's. The key result is the following.

[1]This is also called "rate-distortion" theory.

Shannon's Distortion-Rate Theorem

For a stationary, ergodic source with finite variance.

$$\delta(R) = \mathcal{D}(R)$$

OPTA function = Shannon's DRF

where

$$\mathcal{D}(R) \triangleq \lim_{k\to\infty} \mathcal{D}(k,R) = \text{Shannon's "distortion-rate function"}$$

$$\mathcal{D}(k,R) \triangleq \inf_{q \in \mathcal{Q}_k(R)} E \frac{1}{k} ||\underline{X}-\underline{Y}||^2$$

$\underline{X} = (X_1 \ldots X_k)$ random variables from source

$\underline{Y} = (Y_1 \ldots Y_k)$ random variables from *test channel* q with $\underline{X}$ as input

$\mathcal{Q}_k(R)$ = set of conditional probability densities, called "test channels"

$$= \left\{ q(\underline{y}|\underline{x}) : \frac{1}{k} \int f(\underline{x}) q(\underline{y}|\underline{x}) \log_2 \frac{f(\underline{x})q(\underline{y}|\underline{x})}{f(\underline{y})} d\underline{x}\, d\underline{y} \leq R \right\}$$

$\left(\frac{1}{k} \times \right.$ "information" $I(\underline{X};\underline{Y})$ given by $\underline{Y}$ about $\left. \underline{X} \right)$

$E \frac{1}{k} ||\underline{X}-\underline{Y}||^2$ is computed wrt to joint density $f(\underline{x},\underline{y}) = f(\underline{x}) q(\underline{y}|\underline{x})$

- $\delta(R)$ is defined by a minimum over actual quantizers.
- $\mathcal{D}(R)$ is defined by a minimum over hypothetical conditional probability distributions.
- There is no straightforward connection between the $\delta(R)$ and $\mathcal{D}(R)$.
- This theorem is one of the deep and central results of information theory.
- Its proof can be found in information theory textbooks.

  As does most of information theory, the proof uses the asymptotic equipartition property, which in turn derives from the law of large numbers. Time permitting, we'll sketch some ideas of the proof later in the semester.
- The theorem says two things:

  Positive statement: For any R, there exist VQ's with rate R or less having MSE arbitrarily close to $\mathcal{D}(R)$.

  (The proof shows there exist fixed-rate codes.)

  Negative statement: For any R, every VQ with rate R or less (fixed- or variable-rate) has MSE greater than or equal to $\mathcal{D}(R)$.
- Unfortunately, this theorem does not indicate how large the dimension needs to be to able to attain distortion close to $\mathcal{D}(R)$.
- Fortunately, Zador's theorem does enable us to learn how large the dimension needs to be, at least for large R, which is why we have focused in this course on Zador's rather than Shannon's theorem.

- The test channels introduced in the definition of $\mathcal{D}(R)$ are not to be considered codes or any other part of an actual physical system.
- Although the definition of $\mathcal{D}(R)$ is quite complex, there are cases, such as Gaussian sources, where it can be reduced to a closed form or parametric expression.
- In other cases, the "Blahut algorithm" can at least be used to compute $\mathcal{D}(k,R)$, and if k is large, $\mathcal{D}(k,R) \cong \mathcal{D}(R)$. Unfortunately, the Blahut algorithm becomes very complex for large k. So in practice it is extremely difficult to compute $\mathcal{D}(R)$, except in special cases such as IID or Gaussian sources.
- Because $\mathcal{D}(R)$ can be so difficult ot compute, upper and lower bounds to it have been developed, which can serve as approximations.
- Shannon's theorem is often stated in the following equivalent form:

$$\gamma(D) = \mathcal{R}(D)$$

  where $\gamma(D)$ is the rate vs. distortion OPTA function, defined as the least rate of any lossy source code with distortion D or less (it is the inverse of $\delta(R)$), and $\mathcal{R}(D)$ is the "Shannon rate-distortion function", which is the inverse of $\mathcal{D}(R)$. In fact, Shannon originally stated the theorem in this form, and the subject is usually called "rate-distortion theory".

- The theorem generalizes to other measures of distortion between vectors of the form

$$d(\underline{x},\underline{y}) = \frac{1}{k}\sum_{i=1}^{k} d(x_i,y_i)$$

where $d(x,y)$ is some distortion measure between individual samples. $d(\underline{x},\underline{y})$ is called a per-letter distortion measure.

THE COMPLEMENTARY NATURE OF SHANNON DISTORTION-RATE THEORY AND ZADOR'S HIGH-RESOLUTION THEORY

Consider Fixed-Rate Coding

- Shannon Theory:

  For large k and any R, $\delta(k,R) \cong \delta(R) = \mathcal{D}(R)$

- High-Resolution Theory:

  For large R and any k: $\delta(k,R) \cong Z(k,R)$

- For large k and large R, they agree:

  $$\delta(R) = \mathcal{D}(R) \cong \delta(k,R) \cong Z(k,R)$$

- Important Note:

  $$\delta(k,R) \neq \mathcal{D}(k,R)$$

  All we can say is

  $\delta(k,R) > \mathcal{D}(k,R)$, for all k,R

  $\delta(k,R) \cong Z(k,R)$, when k,R large

RELATIONSHIPS BETWEEN THE DISTORTION-RATE FUNCTION AND THE ZADOR FUNCTION

The following can be shown directly from definitions (they also follow from what we know about the operational significance of $\mathcal{D}(k,R)$ and $Z(k,R)$):

- $\mathcal{D}(k,R) \geq \frac{1/2\pi e}{m_k^*} Z(k,R)$

  The ratio of the left and right sides goes to one as $R \to \infty$.

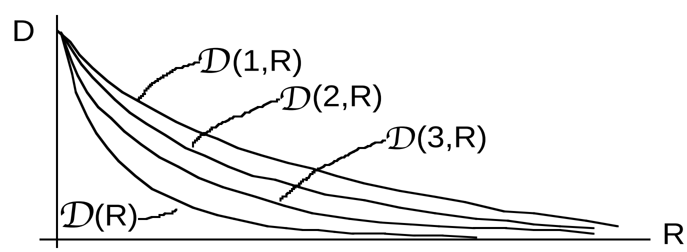
- $\mathcal{D}(R) \geq Z(R)$

  The ratio of the left and right sides goes to one as $R \to \infty$.
  Sometimes they are equal for sufficiently large values of R.

- The above inequalities are called Shannon-Lower Bounds.
  They are restated and proved later in Property 12.

PROPERTIES OF THE DISTORTION-RATE FUNCTION

These properties are derived by directly using and manipulating the definitions of $\mathcal{D}(k,R)$ and $\mathcal{D}(R)$.



1. $\mathcal{D}(0) = \mathcal{D}(k,0) = \sigma^2$

2. $\mathcal{D}(R) > 0$ and $\mathcal{D}(k,R) > 0$, $R \geq 0$

3. $\mathcal{D}(R)$ and $\mathcal{D}(k,R)$ decrease montonically to zero as R increases.

4. $\mathcal{D}(R)$, $\mathcal{D}(k,R)$ are convex (and consequently continuous) functions of R.

5. The $\mathcal{D}(k,R)$'s are subadditive. That is for any k, m, R

   $$\mathcal{D}(k+m,R) \leq \frac{k}{k+m}\mathcal{D}(k,R) + \frac{m}{k+m}\mathcal{D}(m,R)$$

   From which it follows that

   $$\mathcal{D}(R) \leq \mathcal{D}(nk,R) \leq \mathcal{D}(k,R) \leq \mathcal{D}(1,R) \text{ for all } k.$$

   $$\mathcal{D}(R) = \inf_k \mathcal{D}(k,R)$$

   Thus, $\mathcal{D}(k,R)$'s tend to decrease with k, but not necessarily monotonically.

6. $\mathcal{D}(R) = \mathcal{D}(1,R)$ when the source is IID.

7. For an IID Gaussian source

$$\mathcal{D}(R) = \mathcal{D}(1,R) = \sigma^2 2^{-2R}$$

Derivation: First recall that for an IID Gaussian source, the first-order differential entropy is $h_1 = \frac{1}{2} \log 2\pi e \sigma^2$. Therefore, the Shannon lower bound gives

$$\mathcal{D}(1,R) \geq \frac{1/2\pi e}{m_1^*} Z(1,R) = \frac{1/2\pi e}{m_1^*} m_1^* 2^{-2h_1-2R} = \sigma^2 2^{-2R}$$

The derivation is completed by showing $\mathcal{D}(1,R) \leq \sigma^2 2^{-2R}$. This is accomplished by verifying that the test channel

$$q(y|x) = \frac{1}{\sqrt{2\pi b}} \exp\left\{-\frac{(y-ax)^2}{2b}\right\}, \quad \text{with } a = 1-2^{-2R} \text{ and } b = 2^{-2R}(1-2^{-2R})\sigma^2,$$

has $I(X;Y) \leq R$ and $E(X-Y)^2 = \sigma^2 2^{-2R}$. It then follows from the definition of $\mathcal{D}(1,R)$ that

$$\mathcal{D}(1,R) \leq E(X-Y)^2 = \sigma^2 2^{-2R}.$$

8. For a first-order AR Gaussian source with correlation coef. ρ

$$\mathcal{D}(R) = Z(R) = \sigma^2(1-\rho^2)\, 2^{-2R} \quad \text{for } R \geq R_o \triangleq \frac{1}{2} \log_2 (1+|\rho|)^2$$

No closed form expression for other R's.
This property follows from the next.

9. For a stationary Gaussian source with power spectral density $S(\omega)$,

$$\mathcal{D}(R) = Z(R) = Q\, 2^{-2R}, \quad R \geq R_o \triangleq \frac{1}{2} \log_2 \frac{Q}{S_{min}}$$

where S_{min} is the minimum value of $S(\omega)$, and Q is the mean-squared error of the best linear prediction of X_i based on all past values of X:

$$Q = \exp\left\{\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln S(\omega)\, d\omega\right\}$$

For $R \geq R_o$, there is no closed form expression for $\mathcal{D}(R)$. However, the following parametric expression applies for all values of R.

For any θ, $0 \leq \theta \leq S_{max}$, where S_{min} is the max value of $S(\omega)$,

$$D_\theta = \mathcal{D}(R_\theta),$$

where

$$R_\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max\left\{0, \frac{1}{2} \log_2 \frac{S(\omega)}{\theta}\right\} d\omega$$

$$D_\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\{\theta, S(\omega)\}\, d\omega$$

Interpretation: For a given θ, all frequences ω for which $S(\omega) \geq \theta$, contribute $\frac{1}{2} \log \frac{S(\omega)}{\theta}$ to the rate, and θ to the distortion.

All frequencies ω for which $S(\omega) < \theta$ are discarded (e.g. filtered out). They contribute 0 to the rate, and $S(\omega)$ to the distortion.

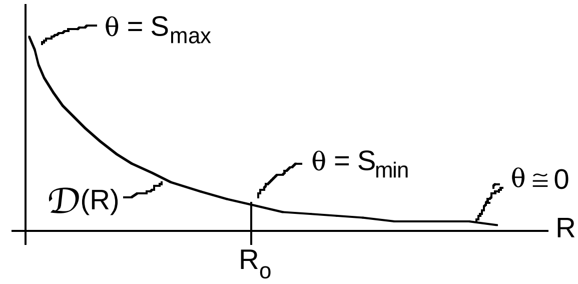
Special cases:

$\theta = 0 \;\Rightarrow\; R_\theta = \infty,\; D_\theta = 0$

$\theta = S_{max} \;\Rightarrow\; R = 0,\; D_\theta = \sigma^2$

$\theta \le S_{min} \;\Rightarrow\; R_\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{2}\log_2 \frac{S(\omega)}{\theta}\, d\omega,$

$D_\theta = \theta.$



$$\Rightarrow \mathcal{R}(D) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{2}\log_2 \frac{S(\omega)}{D}\, d\omega = \frac{1}{2}\log_2 \frac{Q}{D}, \quad D \le S_{min}$$

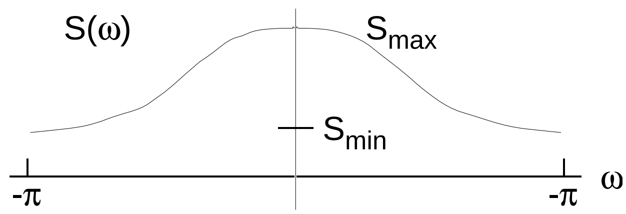
$$\Rightarrow \mathcal{D}(R) = Q\, 2^{-2R}, \quad R \ge \frac{1}{2}\log_2 \frac{Q}{S_{min}} = R_o.$$

For the AR source, Property 8 follows from Property 9:

$$S(\omega) = \sigma^2 \frac{1-\rho^2}{1-2\rho\cos(\omega)+\omega^2},$$

$$S_{min} = \sigma^2 \frac{1-|\rho|}{1+|\rho|}, \quad S_{max} = \sigma^2 \frac{1+|\rho|}{1-|\rho|},$$

$$Q = \sigma^2(1-\rho^2), \quad R_o = \frac{1}{2}\log_2 (1+|\rho|)^2$$



10. There are a few other sources for which $\mathcal{D}(R)$ can be computed analytically.

    For other sources, $\mathcal{D}(R)$ must be computed numerically. The most well known agorithm is that of Blahut for computing $\mathcal{D}(k,R)$.

    Because it is hard to compute, various upper and lower bounds have been found for $\mathcal{D}(k,R)$ and $\mathcal{D}(R)$. Two are given below.

11. An upper bound: For any source, $\mathcal{D}(R)$ and $\mathcal{D}(k,R)$ are bounded from above by the corresponding functions for a Gaussian source with the same autocorrelation function (equivalently, same power spectral density).

    It follows that Gaussian sources are the hardest to compress among those sources with a given autocorrelation function.

12. Shannon lower bounds

Let X be a stationary source. For any k and R

$$\mathcal{D}(k,R) \geq \frac{1}{2\pi e} 2^{2h_k - 2R} = \frac{1/2\pi e}{m_k^*} Z(k,R)$$

$$\mathcal{D}(R) \geq \frac{1}{2\pi e} 2^{2h_\infty - 2R} = Z(R)$$

where

$h_k = \frac{1}{k} h(X_1,...,X_k)$ is the kth order differental entropy of X

$h_\infty = \lim_{k\to\infty} h_k$ is the differential entropy rate of X

m_k^* = minimum value of any valid inerital profile

$Z(k,R)$ and $Z(R) = \lim_{k\to\infty} Z(k,R)$ are Zador functions

Note: The ratio of the left and right sides of each bound can be shown to go to one as $R \to \infty$. Sometimes equality hold for large R.

Derivation: To derive lower bound to $\mathcal{D}(k,R)$, consider any test channel q that is allowed to be used in the definition of $\mathcal{D}(k,R)$, i.e. any q s.t.

$$\frac{1}{k} I(\underline{X};\underline{Y}) \triangleq \frac{1}{k} \int f(\underline{x})\, q(\underline{y}|\underline{x}) \log_2 \frac{f(\underline{x})q(\underline{y}|\underline{x})}{f(\underline{y})}\, d\underline{x}\, d\underline{y} \leq R.$$

We will show:

$$E \frac{1}{k} ||\underline{X}-\underline{Y}||^2 \geq \frac{1}{2\pi e} 2^{2h_k - 2R}. \qquad (*)$$

Since this holds for any valid q, it holds for the q that minimizes $E \frac{1}{k} ||\underline{X}-\underline{Y}||^2$. Therefore, for the minimizing q

$$\mathcal{D}(k,R) = E \frac{1}{k} ||\underline{X}-\underline{Y}||^2 \geq \frac{1}{2\pi e} 2^{2h_k - 2R} = \frac{1/2\pi e}{m_k^*} Z(k,R)$$

where the last equality comes from the definition of $Z(k,R)$.
This derives the Shannon lower bound to $\mathcal{D}(k,R)$.

The shannon lower bound to $\mathcal{D}(R)$ follows by taking the limit as k grows to infinity.

It remains only to derive (*), which we do using the following lemma from information theory.

Fano's Lemma for MSE:

If $\underline{X}$ and $\underline{Y}$ are k-dimensional random vectors, then

$$E \frac{1}{k} ||\underline{X}-\underline{Y}||^2 \geq \frac{1}{2\pi e} 2^{2\frac{1}{k} h(\underline{X}|\underline{Y})}$$

By manipulating the defining formula for $I(\underline{X};\underline{Y})$, one may straightforwardly show

$$h(\underline{X}|\underline{Y}) = h(\underline{X}) - I(\underline{X};\underline{Y}) = k\, h_k - I(\underline{X};\underline{Y}) \quad \text{by the definition of } h_k$$

$$\geq k\, h_k - kR \quad \text{by the choice of } q$$

Substituting this into the lower bound given in Fano's Lemma yields (*), and finishes the derivation.