Summary of Shannon Rate-Distortion Theory

Consider a stationary source X with kth-order probability density function denoted $f_k(\underline{x})$.

Consider VQ with fixed-rate coding.

Recall the following OPTA function definitions.

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

$$\delta(R) = \inf_k \delta(k,R)$$

= least dist'n of VQ's with rate $\leq$ R and any dimension

These functions describe the best possible performance of fixed-rate 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 rate-distortion theory enables one to find $\delta(R)$ for ANY value of R. However, it does not allow us to find $\delta(k,R)$, not even for some R's. The key result is the following.

Shannon's Distortion-Rate Theorem

For a stationary, ergodic source with finite variance.

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

$$\text{OPTA function} = \text{Shannon's DRF}$$

where

$\mathcal{D}(R)$ = Shannon's "distortion-rate function"

$$= \lim_{k \to \infty} \mathcal{D}(k,R)$$

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

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

$\underline{Y} = (Y_1 ... Y_k)$ random variables from *test channel* q

$\mathcal{Q}_k(R)$ = set of test channels (cond'l prob. densities)

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

"mutual information" $I(\underline{X};\underline{Y})$

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

- This theorem is one of the deep and central results of information theory.

- Its proof can be found in information theory texts.

  As does most of information theory, it uses the asymptotic equipartition property, which follows from the law of large numbers.

  We'll sketch some ideas of the proof later.

- The theorem says two things:

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

  Negative statement: For any R, every VQ with rate R or less has MSE greater than or equal to D(R).

- The test channels introduced in the definition of D(R) are not to be considered codes or any other part of an actual physical system.

- Although the definition of 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 be used to compute it.

- The theorem can be generalized to show that no variable-rate code can do better than D(R). Thus D(R) is also the best performance of variable-rate codes. The theorem can even be generalized to show that other coding structures can do no better than D(R).

- Unfortunately, this theorem does not indicate how large the dimension needs to be to get good performance.

- The theorem generalizes to other distortion measures of the form

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

which are called per-letter distortion measures.

- The theorem is often stated as showing

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

i.e. the equality of the rate vs. distortion OPTA function, which is the inverse of $\delta(R)$, and the "Shanon rate-distortion function" $\mathcal{R}(D)$, which is the inverse of $\mathcal{D}(R)$. In fact, this is where the subject gets its name.

Shannon Rate-Distortion Theory and High-Resolution Theory Are Complementary

Consider Fixed-Rate Coding

- Shannon Theory:

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

- High-Resolution Theory:

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

- For large k and large R, they agree

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

- Important Note: $\mathcal{D}(k,R) \neq \delta(k,R)$

Relationships between distortion-rate functions and Zador-Gersho functions

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

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

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

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

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

  (In distortion-rate theory, the above inequalities are called Shannon-Lower Bounds to the distortion-rate function.)

PROPERTIES OF THE DISTORTION-RATE FUNCTION

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

2. $D(R) > 0$ and $D(k,R) > 0$, for all $R \geq 0$

3. $D(R)$ and $D(k,R)$ decrease montonically towards zero as R increases.

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

5. The $D(k,R)$'s are subadditive. That is for any k and m

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

   From which it follows that

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

   $$D(R) = \inf_k D(k,R)$$

   Thus, the $D(k,R)$'s tend to decrease with k but not necessarily monotonically.

6. $D(R) = D(1,R)$ when the source is IID.

7. For an IID Gaussian source

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

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

   $$D(R) = \sigma^2(1-\rho^2)\, 2^{-2R} \text{ for } R \geq \frac{1}{2}\log_2 (1+\rho)^2$$

   (No closed form expression for other R's.)

9. There are other sources for which $D(\mathrm{R})$ can be computed analytically (e.g. Gaussian but not IID, and IID Laplacian).

10. For other sources $D(\mathrm{R})$ must be computed numerically, and there is a numerical algorithm for computing $D(\mathrm{k,R})$ called the Blahut algorithm.

11. An upper bound: $D(\mathrm{R})$ and $D(\mathrm{k,R})$ are upper bounded by the corresponding functions for a Gaussian source with the same autocorrelation function (equivalently, power spectral density). In other words Gaussian sources are the hardest to compress of those sources with a given autocorrelation function.

12. There are other lower bounds besides the Shannon lower bound.

13. For a stationary Gaussian source with power spectral density $S(\omega)$, there is a parametric expression for the distortion-rate function.

Let S_{min} and S_{max} denote the min and max values of $S(\omega)$. The for any θ, $0 \le \theta \le S_{max}$, $D_\theta = \mathcal{D}(R_\theta)$, where

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

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

Special cases:

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

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

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

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

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

where Q is the minimum mean squared error of a lienar prediction of X_i based only past values:

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