ECE 8873
Data Compression & Modeling

Lecture 5: Lossy Compression, Distortion Measures

School of Electrical and Computer Engineering
Georgia Institute of Technology
Spring, 2004

Fundamentals

- A quantization/coding scheme consists of
  - An encoder $\alpha: \quad A^* \to B^* = \{0,1\}^*$

    If $x \in A$ and $\alpha(x)$ appear as a parsable units in their respective domain, we can define the average code length $E[l(\alpha(x))]$ per input "unit," which can be sample, pixel, word (variable number of letters), sequence, area, or time.
  - A decoder $\beta: \quad B^* \to \hat{A}^*, \ \beta(\alpha(x)) = \hat{x} \in \hat{A}$

    Usually $\hat{A} \subset A, \ \beta(\alpha(x)) = \hat{x} \in \hat{A}$ but $\hat{x}$ may not equal x.
  - A reproduction codebook $C = \{\beta(i), \ i \in I\}$

    Each $\beta(i), \ i \in I$ is called a codeword and the codeset I has elements represented in integer index without loss of generality. Why?

Objectives & Concerns

- Given $\|I\|$, find the encoder and the decoder such that the discrepancy between x and $\hat{x}$ is minimum.
- How to measure discrepancy? Use "distortion measure" $d(x,\hat{x})$.
- Instead of looking at $E[l(\alpha(x))]$, we are interested in

$$E[d(X,\hat{X})] = E[d(X,\beta(\alpha(X)))] = D(\alpha,\beta)$$

For the distortion measure to be useful, it should be:

- Easy to compute;
- Tractable (at least for consistency);
- Meaningful for perception or in specific applications.

General Distance or Dissimilarity Measure

- Point to point
- Point to set
- Set to set
- Dissimilarity measure between probability distribution or density functions

Distortion Measures

- A distortion measure is a real non-negative function that maps two variables, possibly vector valued, to a scalar that indicates the degree of dissimilarity between the two variables.
- The space the two variables are in may be a metric space. A space V is *metric* if the *distance* function $d(x,y)$ between two points A and B is characterized by the following properties:

  For $x \in V$, $y \in V$, and $z \in V$
  - $0 \le d(x,y) < \infty$, and $d(x,y) = 0$ if and only if $x = y$ (0 distance only when the two points coincide)
  - $d(x,y) = d(y,x)$ (the distance is symmetric)
  - $d(x,y) + d(y,z) \ge d(x,z)$ (triangle inequality.)
- But, a general measure of dissimilarity may not satisfy these properties.

Lp Norm

- Let $\mathbf{x}$ be a N-dimensional vector, $\mathbf{x} = (x_1, x_2, \cdots, x_N)$. The length of $\mathbf{x}$ can be defined in many ways. The L_p norm of $\mathbf{x}$ is defined as

$$\|\mathbf{x}\|_p = \left(\sum_{n=1}^{N} |x_n|^p \right)^{1/p}$$

- Or, it may be normalized: $$\|\mathbf{x}\|_p = \left(\frac{1}{N} \sum_{n=1}^{N} |x_n|^p \right)^{1/p}$$
  - When p=1, it is the mean absolute value.
  - When p=2, it is the RMS value.
  - When p=∞, it is referred to as "Chebyshev," "supremum," "minimax," or "uniform" norm.

Lp Norm as Distortion Measure

- Let $\mathbf{x}$ and $\mathbf{y}$ be two N-dimensional vectors,

$$\mathbf{x} = (x_1, x_2, \cdots, x_N) \qquad \mathbf{y} = (y_1, y_2, \cdots, y_N)$$

- The difference of the two vectors is

$$\mathbf{z} = \mathbf{x} - \mathbf{y} = (x_1 - y_1, x_2 - y_2, \cdots, x_N - y_N) = (z_1, z_2, \cdots, z_N)$$

- The normalized L_p of $\mathbf{z}$ can be used as a distortion measure between $\mathbf{x}$ and $\mathbf{y}$:

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_p = \left(\frac{1}{N} \sum_{n=1}^{N} |x_n - y_n|^p \right)^{1/p}$$

- When p=1, it is the city block distance.
- When p=2, it is the (normalized) Euclidean distance.
- When p=∞, it is referred to as "Chebyshev," or peak distance.

Weighted Euclidean Distance

- Euclidean distance $$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{n=1}^{N} (x_n - y_n)^2}$$
- Different significance may be manifested along different dimensions, thus weighting

$$d_W(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{n=1}^{N} w_n (x_n - y_n)^2} = (\mathbf{x} - \mathbf{y})^t \mathbf{W} (\mathbf{x} - \mathbf{y})$$

where W is a diagonal matrix. In general, it may not be.

- Normalized weighted Euclidean distance:

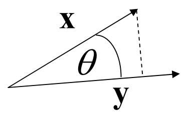
$$d_W(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{n=1}^{N} w_n (x_n - y_n)^2 \left(\sum_{n=1}^{N} w_n \right)^{-1}}$$

- How to introduce general weighting?
- The role of covariance matrix – if W is the inverse covariance matrix, the effect is to "equalize" the statistical spread along each data dimension – Mahalanobis distance.

Projection, Dot Product, Inner Product

- Projection of a vector $\mathbf{x}$ onto another vector $\mathbf{y}$ can be used to measure the dissimilarity between $\mathbf{x}$ and $\mathbf{y}$.
- The dot product or inner product (Hilbert space) between $\mathbf{x}$ and $\mathbf{y}$ is defined as

$$\mathbf{x} \bullet \mathbf{y} = |\mathbf{x}||\mathbf{y}|\cos\theta = \sum_{i=1}^{N} x_i y_i$$



- $\mathbf{x} \bullet \mathbf{y}$ attains maximum when $\theta = 0$

$$\max(\mathbf{x} \bullet \mathbf{y}) = |\mathbf{x}||\mathbf{y}|$$

- Many functions can be constructed as dissimilarity; e.g.

$$1 - \left|\frac{\mathbf{x} \bullet \mathbf{y}}{|\mathbf{x}||\mathbf{y}|}\right| = 1 - \cos\theta$$

(Discuss properties)

More on Projection

- Normalization by individual vector's norm

$$d(\mathbf{x}, \mathbf{y}) = 1 - \left|\frac{\mathbf{x} \bullet \mathbf{y}}{|\mathbf{x}||\mathbf{y}|}\right| = 1 - \cos\theta$$

- Normalization to resemble 0-mean 1-var random samples (Pearson Correlation)

$$d(\mathbf{x}, \mathbf{y}) = 1 - R(\mathbf{x}, \mathbf{y}) \quad \text{where } R(\mathbf{x}, \mathbf{y}) = \frac{1}{L} \frac{\mathbf{x} - \eta_x}{\sigma_x} \bullet \frac{\mathbf{y} - \eta_y}{\sigma_y}$$

- Still other normalization conventions

Other Forms of Dissimilarity Measure

- Fractional Norm

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_p = \left(\frac{1}{N} \sum_{n=1}^{N} |x_n - y_n|^p\right)^{1/p}, \quad p \in (0,1)$$

- Ratio oriented (scalar)

$$d(x, y) = \left\|xy^{-1} - x^{-1}y\right\|_p$$

Possible project assignment: collect and implement all distance or distortion measures

Other Distances for Probability Measures

- Variational distance (L1 distance between two distributions)

$$D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N} |x_i - y_i| \quad \text{where } \mathbf{x} \text{ and } \mathbf{y} \text{ are two distribution functions}$$

- Bhattacharyya distance (can be applied to pair of classes)

$$D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N} \sqrt{x_i y_i} \quad x_i, y_i \geq 0$$

- Harmonic mean

$$D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N} \frac{2 x_i y_i}{x_i + y_i}$$

- Kullback-Liebler number (and symmetrized versions)

$$D(p_1 \| p_2) = \int p_1(x) \log \frac{p_1(x)}{p_2(x)} dx$$

Ali-Silvey Class of Distances for PDFs

$d(p_1, p_2) = f\left(E_{p_1}\left[C(\gamma(X))\right]\right)$

where $\gamma(X) = \dfrac{p_2(X)}{p_1(X)}$ is the likelihood ratio, C is a convex function, E_{p_1} is the expectation with respect to p_1 and $f(.)$ is a non-decreasing function

If $C(x) = x \log x$ and $f(x) = x$,
it becomes the Kullback-Leibler distance.

If $C(p_1, p_2) = \max\limits_{0 \le t \le 1}\left(-\log \rho(t)\right)$, $\rho(t) = \int [p_1(x)]^{1-t}[p_2(x)]^t \, dx,$
it becomes the Chernoff distance.

Waveform Distortion Measures

$\mathbf{x} = (x_1, x_2, \cdots, x_i, \cdots) \qquad \mathbf{y} = (y_1, y_2, \cdots, y_i, \cdots)$

- Euclidean distance $d(\mathbf{x},\mathbf{y}) = \sum_i (x_i - y_i)^2$
- Signal-to-(coding)-noise ratio $SNR = 10\log_{10} \dfrac{\sum_i x_i^2}{\sum_i (x_i - y_i)^2}$
- SNR with gain optimization
  $SNRG = -10\log_{10}\left(1 - \left|\dfrac{\mathbf{x}\bullet\mathbf{y}}{|\mathbf{x}||\mathbf{y}|}\right|^2\right)$
- Segmental signal-to-noise ratio

$r_n = \rho + \dfrac{\sum_{i=-I}^{I} x_{n+i}^2}{\varepsilon + \sum_{i=-I}^{I}(x_{n+i} - y_{n+i})^2} \qquad SSNR = 10\log_{10}(r_n - \rho)$

- Average segmental signal-to-noise ratio

$\overline{SSNR} = 10\log_{10}\left(10^{N^{-1}\sum_{n=1}^{N}\log_{10} r_n} - \rho\right) \qquad \overline{SSNR} = N^{-1}\sum_{n=1}^{N} 10\log_{10} r_n \quad \text{if} \quad \rho = 0$

Spectral Distortion Measures

- Nonlinear distortion
  - THD (total harmonic distortion): the ratio of total rms power of the newly created harmonics due to non-linear distortion to that of the original signal power, expressed in dB.
- Linear distortion
  - Equivalent to result of filtering;
  - Measured in a number of ways, particularly when perceptual significance is taken into account.

Spectral Distance & Log Spectral Distance

Power spectral density: $\mathbf{x} \Leftrightarrow X(\omega) \quad \mathbf{y} \Leftrightarrow Y(\omega)$

- Lp norm of spectral difference
  $d^p(\mathbf{x},\mathbf{y}) = \int_{-\pi}^{\pi} \left|X(\omega) - Y(\omega)\right|^p \dfrac{d\omega}{2\pi}$
- Lp norm of log spectral difference
  $V(\omega) = \log X(\omega) - \log Y(\omega)$
  $d^p(\mathbf{x},\mathbf{y}) = \int_{-\pi}^{\pi} \left|V(\omega)\right|^p \dfrac{d\omega}{2\pi}$

p=1 ➔ mean absolute log spectral distance
p=2 ➔ mean root mean square log spectral distance
p=1 ➔ mean peak log spectral distance

Cepstral Distances

$$\log X(\omega) = \sum_{n=-\infty}^{\infty} c_{xn} e^{-j\omega n}$$

$c_{xn} = c_{-xn}$ are real and called the cepstral coefficients or cepstrum.

From Parseval's Theorem

$$d_2^2 = \int_{-\pi}^{\pi} \left| \log X(\omega) - \log Y(\omega) \right|^2 \frac{d\omega}{2\pi} = \sum_{n=-\infty}^{\infty} (c_{xn} - c_{yn})^2$$

Truncated cepstral distance:

$$d_2^2(L) = \sum_{n=-L}^{L} (c_{xn} - c_{yn})^2$$

Truncating the cepstrum amounts to low-quefrency (lowpass) liftering and will cause smoothing on the log spectrum.

Cepstrum of Model Spectrum

$$X(\omega) \Leftrightarrow \frac{\sigma^2}{|A(e^{j\omega})|^2} \qquad A(z) = \sum_{n=0}^{p} a_n z^{-n}$$

For a stable all-pole filter, $\log A(z^{-1})$ is analytic inside the unit circle and can be represented by the Laurent expansion

$$\log \frac{\sigma}{A(z)} = \log \sigma + \sum_{n=1}^{\infty} c_n z^{-n}$$ LPC cepstrum

Differentiating both sides wrt z^{-1} and equating the coefficients of like powers of z^{-1} :

$$c_n = -a_n - \frac{1}{n} \sum_{k=1}^{n-1} k c_k a_{n-k} \quad \text{for } n > 0 \qquad a_0 = 1 \text{ and } a_k = 0 \text{ for } k > p$$

$$\log \frac{\sigma^2}{|A(z)|^2} = \sum_{n=-\infty}^{\infty} c_n z^{-n} \quad \text{where } c_0 = \log \sigma^2, \text{and } c_{-n} = c_n$$

Cepstrum

$$X(z) = \frac{G z^r \left[\prod_{k=1}^{p_1} (1 - u_k z^{-1}) \right] \left[\prod_{k=1}^{p_2} (1 - v_k z) \right]}{\left[\prod_{k=1}^{p_3} (1 - z_k z^{-1}) \right] \left[\prod_{k=1}^{p_4} (1 - w_k z) \right]}$$

Cepstrum is a decaying sequence. The significance of cepstrum deminishes as the index gets higher. Truncation is thus justified.

$$c_n = \begin{cases} \log G & n = 0 \\ n^{-1} \left[\sum_{k=1}^{p_3} z_k^n - \sum_{k=1}^{p_1} u_k^n \right] & n > 0 \\ n^{-1} \left[\sum_{k=1}^{p_4} w_k^{-n} - \sum_{k=1}^{p_2} v_k^{-n} \right] & n < 0 \end{cases}$$

$$E\{c_n^2\} \approx \frac{1}{n^2}$$

$$|c_n| < \gamma \frac{\lambda^{|n|}}{|n|}$$

λ is maximum absolute value of $u_k, v_k, w_k,$ and z_k

Weighted Cepstrum Distance

$$E\{c_n^2\} \approx \frac{1}{n^2}$$

Weighting by inverse variance $d_{2W}^2 = \sum_{n=-\infty}^{\infty} n^2 (c_n - c_n')^2$

$$\frac{d}{d\omega} \log S(\omega) = \sum_{n=-\infty}^{\infty} - jnc_n e^{-jn\omega}$$

$$d_{2W}^2 = \int_{-\pi}^{\pi} \left| \frac{d \log S(\omega)}{d\omega} - \frac{d \log S'(\omega)}{d\omega} \right|^2 \frac{d\omega}{2\pi}$$

→ Spectral slope distance

$\{nc_n\}_n$ is called the root power sum for LPC cepstrum

Root Power Sum

$$A(z)=\prod_{i=1}^{p}(1-z_iz^{-1}) \qquad \log[1/A(z)]=\sum_{i=1}^{p}\sum_{n=1}^{\infty}\frac{z_i^n}{n}z^{-n}=\sum_{n=1}^{\infty}c_nz^{-n}$$

Root power sum or
Group delay spectrum

$$\sum_{i=1}^{p}\frac{z_i^n}{n}=c_n \quad \text{or} \quad \sum_{i=1}^{p}z_i^n=nc_n$$

$$\log[1/A(z)]=\log[1/|A(z)|]+j\arg[1/A(z)]$$

$$\log[1/|A(z)|]=\mathrm{Re}\left\{\sum_{i=1}^{p}\sum_{n=1}^{\infty}\frac{z_i^n}{n}z^{-n}\right\}=\sum_{n=1}^{\infty}\left(\sum_{i=1}^{p}\frac{z_i^n}{n}\right)\mathrm{Re}[z^{-n}]$$

$$\arg[1/A(z)]=\mathrm{Im}\left\{\sum_{i=1}^{p}\sum_{n=1}^{\infty}\frac{z_i^n}{n}z^{-n}\right\}=\sum_{n=1}^{\infty}\left(\sum_{i=1}^{p}\frac{z_i^n}{n}\right)\mathrm{Im}[z^{-n}]$$

$$-\frac{d}{d\omega}\arg[1/A(e^{j\omega})]=\sum_{n=1}^{\infty}\left(\sum_{i=1}^{p}z_i^n\right)\mathrm{Re}(e^{-jn\omega})$$

Group delay spectrum = transform of RPS sequence

Likelihood Distortion

- Itakura-Saito Distortion

$$V(\omega)=\log X(\omega)-\log Y(\omega)$$

$$d_{IS}(\mathbf{x},\mathbf{y})=\int_{-\pi}^{\pi}\left[e^{V(\omega)}-V(\omega)-1\right]\frac{d\omega}{2\pi}=\int_{-\pi}^{\pi}\frac{X(\omega)}{Y(\omega)}\frac{d\omega}{2\pi}-\log\frac{\sigma_{x,\infty}^2}{\sigma_{y,\infty}^2}-1$$

$\sigma_{x,\infty}^2,\sigma_{y,\infty}^2$ are "1-step" prediction errors of X and Y, respectively

- Linear prediction:

$X(\omega)$ is speech spectrum and $Y(\omega)=\sigma^2/|A(e^{j\omega})|^2$

$$d_{IS}\left(X,\frac{\sigma^2}{|A|^2}\right)=\frac{1}{\sigma^2}\int_{-\pi}^{\pi}X(\omega)|A(e^{j\omega})|^2\frac{d\omega}{2\pi}-\log\frac{\sigma_{x,\infty}^2}{\sigma^2}-1$$

$$\frac{1}{\sigma^2}\int_{-\pi}^{\pi}X(\omega)|A(e^{j\omega})|^2\frac{d\omega}{2\pi}=\frac{\mathbf{a}^t\mathbf{R}_p\mathbf{a}}{\sigma^2}$$

$$\mathbf{a}^t\mathbf{R}_p\mathbf{a}=r_x(0)r_a(0)+2\sum_{n=1}^{p}r_x(n)r_a(n)$$

$$r_a(n)=\sum_{i=0}^{p-n}a_ia_{i+n} \quad \text{for } n=0,1,2,\cdots,p$$

Comparison of Cepstral & Likelihood D

$$V(\omega)=\log X(\omega)-\log Y(\omega)$$

$$d_{IS}(\mathbf{x},\mathbf{y})=\int_{-\pi}^{\pi}\left[e^{V(\omega)}-V(\omega)-1\right]\frac{d\omega}{2\pi}$$

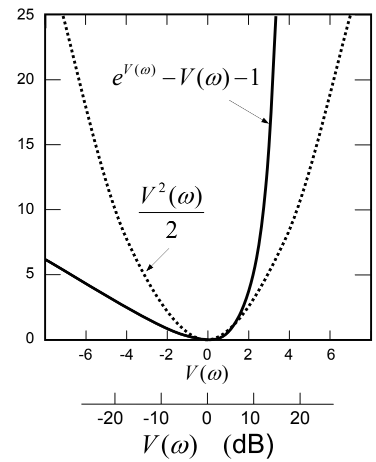
$$d_2^2(\mathbf{x},\mathbf{y})=\int_{-\pi}^{\pi}\left|V(\omega)\right|^2\frac{d\omega}{2\pi}$$

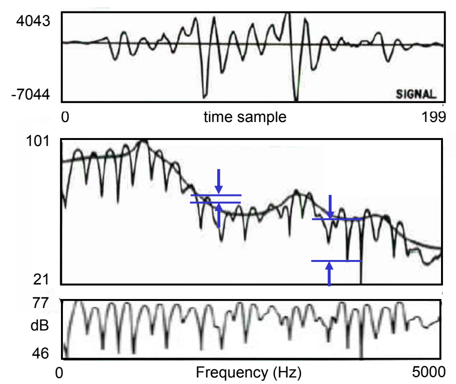
$e^V-V-1\approx e^V$ for $V>>1$ and $e^V-V-1\approx -V$ for $V<<1$

$e^V-V-1\approx\frac{V^2}{2!}+\frac{V^3}{3!}+\cdots\approx\frac{V^2}{2}$ for $|V|<<1$

$d_{IS}(\mathbf{x},\mathbf{y})\approx d_2^2(\mathbf{x},\mathbf{y})$ for small distortion

Implicit Weighting



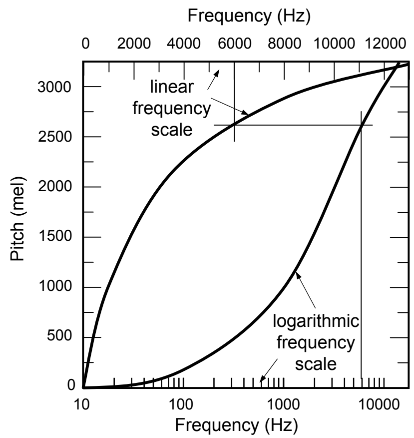


Spectral matching with weighting

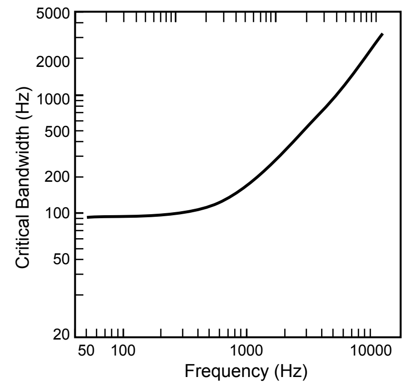
Perceptual Significance

- End quality of a compression system is to be judged by human listeners (for speech & audio) or viewers (for image & video)
- Human judgment involves “subjective” preferences.
- However, physiological studies provide clues, explanations, and justifications for incorporating perceptual effects in a distortion measure.
- Perceptual dimensions:
  - Representation – e.g., mel frequency scale vs. linear scale, all-pole, …
  - Weighting – emphasis on perceptually important dimensions such as formants …
  - Bit allocation

Nonlinear Perceptual “Frequency” Scale



Pitch perception



Loudness perception

Concept of Difference Limen (DL)

- DL is the perceptual sensitivity threshold – the minimum deviation or difference along a certain dimension that is perceptually noticeable; sometimes also called just noticeable difference (JND)
- Examples
  - Vowel formant frequencies: 3-5%
  - Formant amplitude:
    - ◊ 2^{nd} formant: ~3dB
    - ◊ Overall intensity: ~1.5dB (also 1^{st} formant)
    - ◊ Harmonics in spectral valleys: 13-(-∞)
  - Formant bandwidth: 20-40%
  - Fundamental frequency: 0.3-0.5%
  - Light intensity: ~2% (Weber ratio)

Perceptual Masking - Asymmetry