Lattice vector quantization

Let $\{\mathbf{u}_1,...,\mathbf{u}_n\}$ be a set of linearly independent vectors in R^n.

Then the lattice Λ generated by $\{\mathbf{u}_1,...,\mathbf{u}_n\}$ is a set of all points of the form

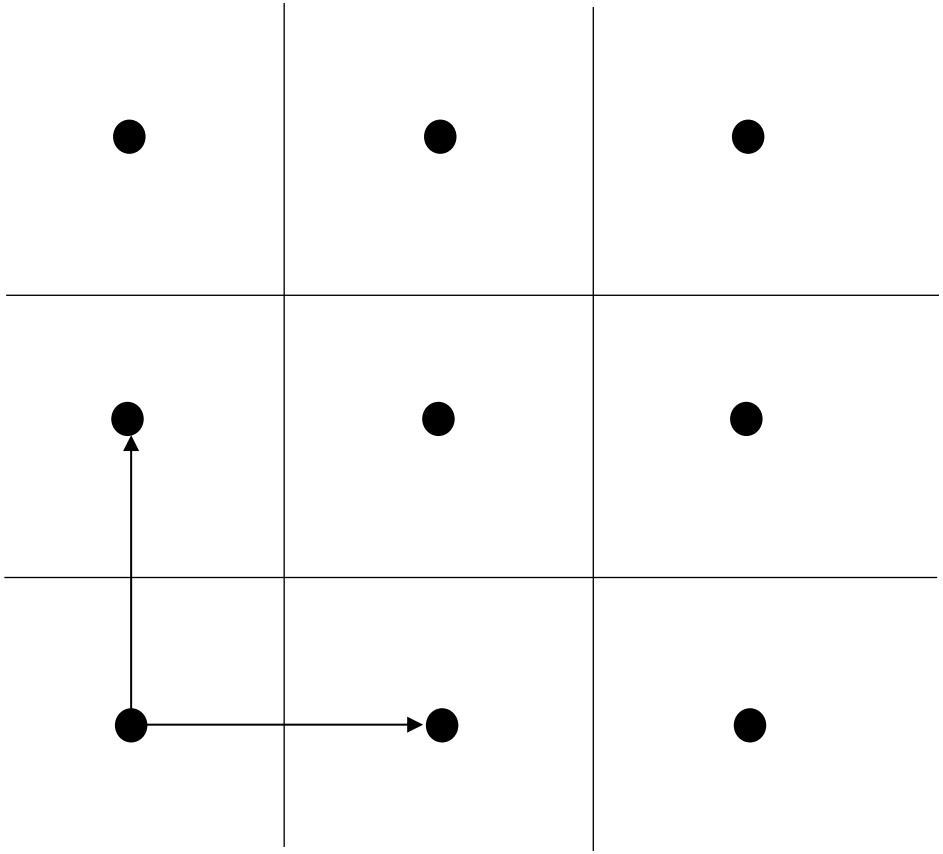
$$\mathbf{y} = \sum_{i=1}^{n} c_i \mathbf{u}_i, \quad \text{where } c_i \text{ are integers.}$$

The vectors $\{\mathbf{u}_i\}$ is a basis for n dimensional lattice.

The matrix $\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 \\ ... \\ \mathbf{u}_n \end{pmatrix}$ is a generator matrix of the lattice.

Any vector of the lattice $\mathbf{y} = \mathbf{cU}, \ \mathbf{c} = (c_1,...,c_n)$.

Voronoi cells and basis vectors of the square lattice



Lattice vector quantization

The hexagonal lattice A_2 has the generator matrix

$$\mathbf{U} = \begin{pmatrix} 1 & 0 \\ 1/2 & \sqrt{3}/2 \end{pmatrix}.$$

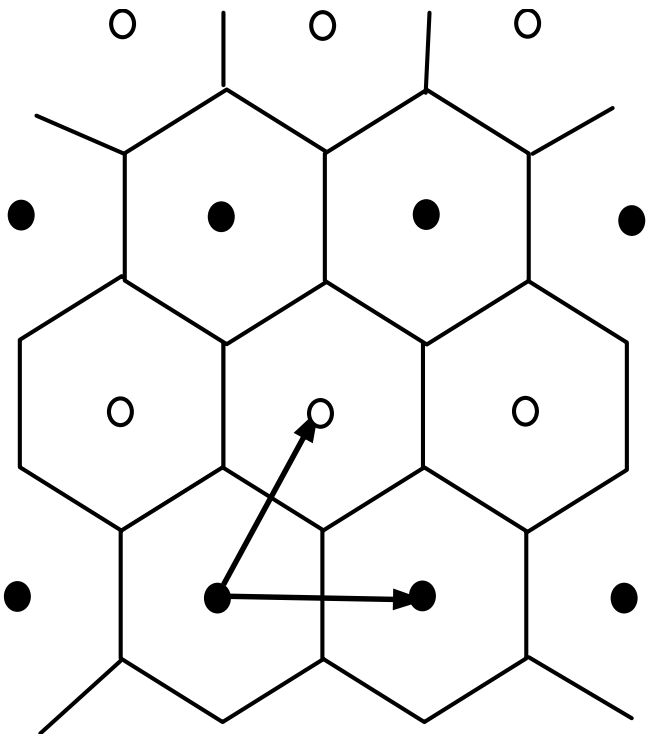
Any vector $\mathbf{y} = (y_1, y_2)$ of this lattice can be written

$$y_1 = c_1 + c_2/2$$
$$y_2 = c_2\sqrt{3}/2.$$

Lattice quantizer is a vector quantizer whose codebook is a lattice.

The approximating vectors are centroids of the Voronoi polyhedrons which are of the same size and shape, t.i. congruent.

Lattice vector quantization



Voronoi cells of the hexagonal lattice

Lattice vector quantization

A_2 is a union of the scaled rectangular lattice and its translate :

Any vector of S^2 has the form $\mathbf{y}_1 = k_1, k_2\sqrt{3}$ (3.1)

and any vector of its translate has the form

$$\mathbf{y}_2 = (k_1 + 1/2, k_2\sqrt{3} + \sqrt{3}/2) = \mathbf{y}_1 + (1/2, \sqrt{3}/2). \quad (3.2)$$

Let $\mathbf{x} = (-5.6, 0.82)$. In (3.1) it is quantized to $(-6.0, 0)$.

In (3.2) the approximation is $(-5.5, \sqrt{3}/2)$.

The best approximation is $(-5.5, \sqrt{3}/2)$.The corresponding error $\|\mathbf{x} - \mathbf{y}\|^2 / 2 = 0.0058..$ The output: $c_1 = -6,\ c_2 = 1.$

Lattice quantizer

The special class of lattices are lattices based on linear codes.

Let C be an (n,k) binary linear code. Then the lattice $\Lambda(C)$ is defined $\Lambda(C)=\{\mathbf{y}\in Z^n \,|\, \mathbf{y}\equiv \mathbf{c} (\mathrm{mod}\, 2) \text{ for some } \mathbf{c}\in C\}$.

Assume that the generator matrix is systematic and given as

$$G=(I|B).$$

Then $U=\begin{pmatrix} I & B \\ 0 & 2I \end{pmatrix}$ and $\Lambda(C)=\bigcup_{i=0}^{2^k-1}(\mathbf{c}_i+2Z^n)$.

Lattice quantizer. Quantization procedure

For each of 2^k cosets of $2Z^n$ do the following :

- Subtract the corresponding codeword from the input vector $\mathbf{d}_i = \mathbf{x} - \mathbf{c}_i$
- Scalar quantize each component of $\mathbf{d}_i$ with step 2 and obtain the quantized vector $\mathbf{q}_i$
- Compute the quantization error $\|\mathbf{x} - \mathbf{y}_i\|^2$, $\mathbf{y}_i = 2\mathbf{q}_i + \mathbf{c}_i$.
- Find the closest pair $(\mathbf{q}_i, \mathbf{c}_i)$ minimizing the error.
- Keep or transmit $(\mathbf{q}_i, i)$.

Lattice quantizer. Dequantization procedure.

- Reconstruct the approximating vector $\hat{\mathbf{d}}_i$

$$\hat{\mathbf{d}}_i = 2\mathbf{q}_i$$

- Add the corresponding codeword to the approximating vector $\mathbf{y}_i = \hat{\mathbf{d}}_i + \mathbf{c}_i$.

Lattice quantizer. Example.

Let C be a (3,2) linear block code with the generator matrix

$$G = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

The generator matrix of the lattice based on C is

$$U = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

$$\mathbf{x} = (0.4, 1.2, 3.7)$$

$\mathbf{c}_0 = (0,0,0)$	$\mathbf{q}_0 = (0,1,2)$	$\hat{\mathbf{d}}_0 = (0,2,4)$	$\mathbf{y}_0 = (0,2,4)$	$D_0 = 0.297$
$\mathbf{c}_1 = (0,1,1)$	$\mathbf{q}_1 = (0,0,1)$	$\hat{\mathbf{d}}_1 = (0,0,2)$	$\mathbf{y}_1 = (0,1,3)$	$D_1 = 0.230$
$\mathbf{c}_2 = (1,0,1)$	$\mathbf{q}_2 = (0,1,1)$	$\hat{\mathbf{d}}_2 = (0,2,2)$	$\mathbf{y}_2 = (1,2,3)$	$D_2 = 0.497$
$\mathbf{c}_3 = (1,1,0)$	$\mathbf{q}_3 = (0,0,2)$	$\hat{\mathbf{d}}_3 = (0,0,4)$	$\mathbf{y}_3 = (1,1,4)$	$D_3 = 0.163$

$i = 3$

Lattice quantizer

- The quantization procedure can be interpreted as a method for choosing the best sequence of n scalar values among 2^k allowed sequences, where each approximating value is generated by one of the two scalar quantizers.

- The first scalar quantizer has the approximating values:…,-4,-2,0,2,4,…. The second quantizer has the approximating values …,-3,-1,1,3,…

- Finding the quantized vector = exhaustive search among 2^k codewords. The computational complexity can be reduced by using code trellis and the Viterbi decoding algorithm.

Elements of rate-distortion theory. Rate-distortion function.

Each quantization procedure is characterized by the average distortion and by quantization rate.

The goal of compression system design is to optimize the rate-distortion tradeoff. In order to compare different quantizers the rate-distortion function $R(D)$ is introduced.

We say that for a given source a quantizer with rate-distortion function $R_1(D)$ is better than the other quantizer with $R_2(D)$ for $D = D_0$ if $R_1(D_0) \le R_2(D_0)$.

The theoretical limit of rate-distortion functions is information rate-distortion function $H(D)$.

Rate-distortion function for memoryless source

Theorem 3.1 The rate-distortion function $H(D)$ for $N(0,\sigma^2)$ source with squared error distortion is

$$H(D)=\begin{cases}\frac{1}{2}\log_2(\sigma^2/D), & \mathrm{D}\le\sigma^2\\ 0, & \mathrm{D}>\sigma^2\end{cases}, \qquad (3.3)$$

where $\mathrm{N}(0,\sigma^2)$ denotes the Gaussian variable with zero mean value and variance σ^2.

In general case $H(D)$

- Is non-increasing and convex downwards function of D.
- There exists some value D_0 such that $H(D)=0$, for all $\mathrm{D}\ge D_0$.

Rate-distortion function for memoryless source

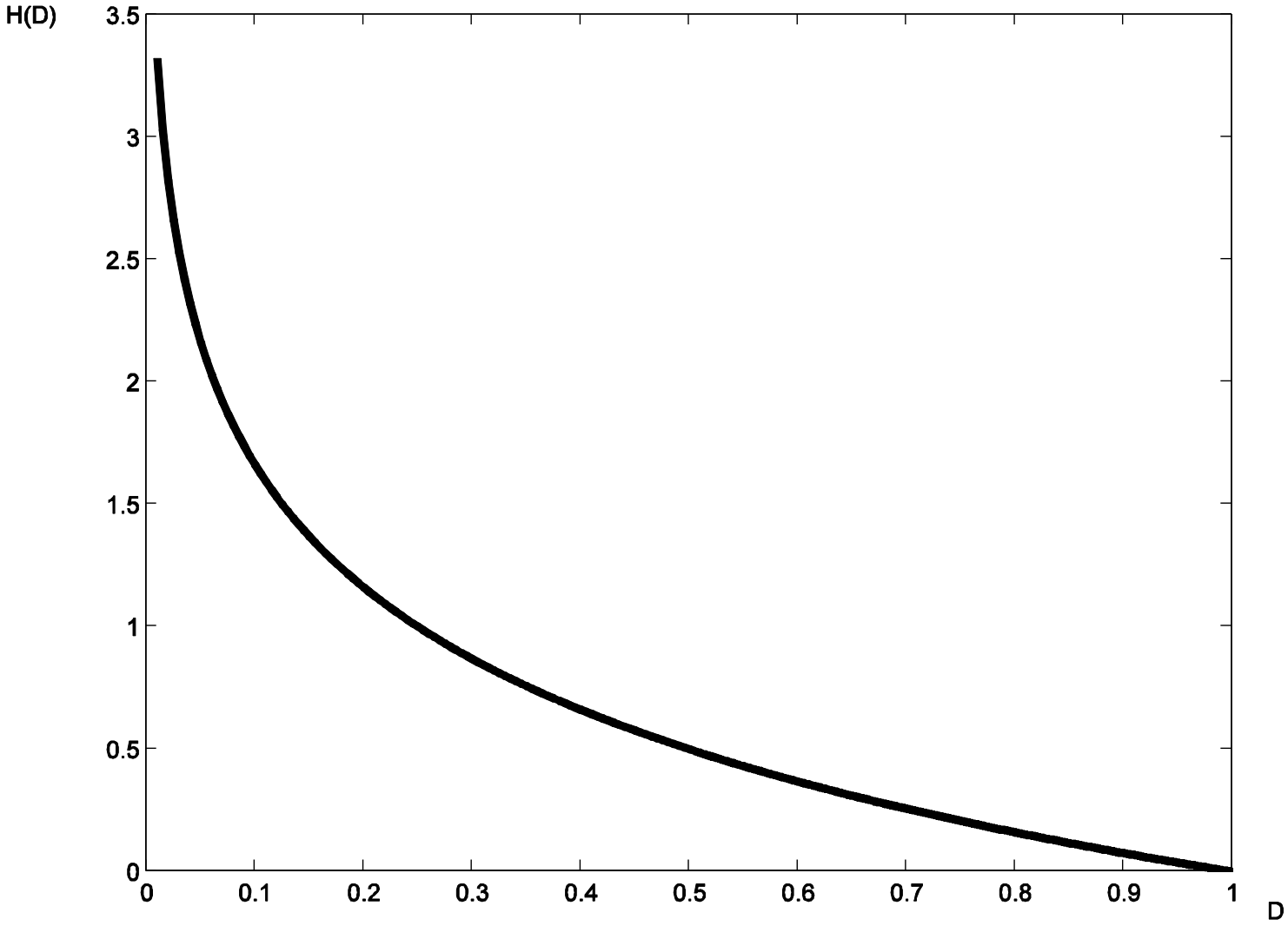


Fig.3.1

Rate-distortion function for source with memory

Consider a source which generates the stationary random Gaussian process of discrete time.

That is for any $n = 1,2,...$ a random vector $\mathbf{X} = (X_1,...,X_n)$ at the output of the source is Gaussian vector with covariance matrix Λ_n and vector of average values $\mathbf{m}$.

Its pdf has the form

$$f_n(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\left|\Lambda_n\right|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})\Lambda_n^{-1}(\mathbf{x}-\mathbf{m})^T\right\},$$

$$\Lambda_n = E\left\{(\mathbf{X}-\mathbf{m})^T(\mathbf{X}-\mathbf{m})\right\} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1n} \\ \dots & \dots & \dots & \dots \\ \lambda_{n1} & \lambda_{n2} & \dots & \lambda_{nn} \end{pmatrix}.$$

Rate-distortion function for source with memory

$\lambda_{ij} = E\{(X_i - m_i)(X_j - m_j)\}$ is the covariance moment of X_i and X_j. λ_{ii} is the variance of X_i.

The property of stationarity means that the n dimensional pdfs of vectors $\mathbf{X} = (X_1, ..., X_n)$ and $\mathbf{X}_j = (X_{j+1}, ..., X_{j+n})$ are identical. It follows from stationarity that

$\lambda_{ij} = \lambda_{ji} = \lambda_{|i-j|}$ and

$$\Lambda_n = \begin{pmatrix} \lambda_0 & \lambda_1 & \dots & \lambda_{n-1} \\ \dots & \dots & \dots & \dots \\ \lambda_{n-1} & \lambda_{n-2} & \dots & \lambda_0 \end{pmatrix}$$

is Toeplitz's matrix.

Rate-distortion function for source with memory

Let $|i-j| = \tau$. Assume that $\mathbf{m = 0}$ and that $\lim_{\tau \to \infty} \lambda_\tau = 0$.

The Fourier series expansion

$$\sum_{\tau=-\infty}^{\infty} \lambda_\tau e^{-j2\pi f\tau} = N(f), \quad -1/2 \le f \le 1/2$$

is called power spectral density of the random process generated by a stationary source. It characterizes how the power of the process is distributed over frequencies.

$$\lambda_\tau = \int_{-1/2}^{1/2} N(f) e^{j2\pi f\tau} df.$$

Rate-distortion function for source with memory

The main properties of the power spectral density are:

- $N(f)$ is a real function
- $N(f) = 2\sum_{\tau=1}^{\infty} \lambda_\tau \cos(2\pi f \tau) + \lambda_0$ since $\lambda_\tau = \lambda_{-\tau}$
- It follows from the previous property that $N(f)$ is even function and thereby $\lambda_\tau = \int_{-1/2}^{1/2} N(f)\cos(2\pi f \tau) df$
- $N(f) \geq 0$ for all $f \in [-1/2, 1/2]$.
- $\lambda_0 = \int_{-1/2}^{1/2} N(f) df$ is the variance of the random process with power spectral density $N(f)$.

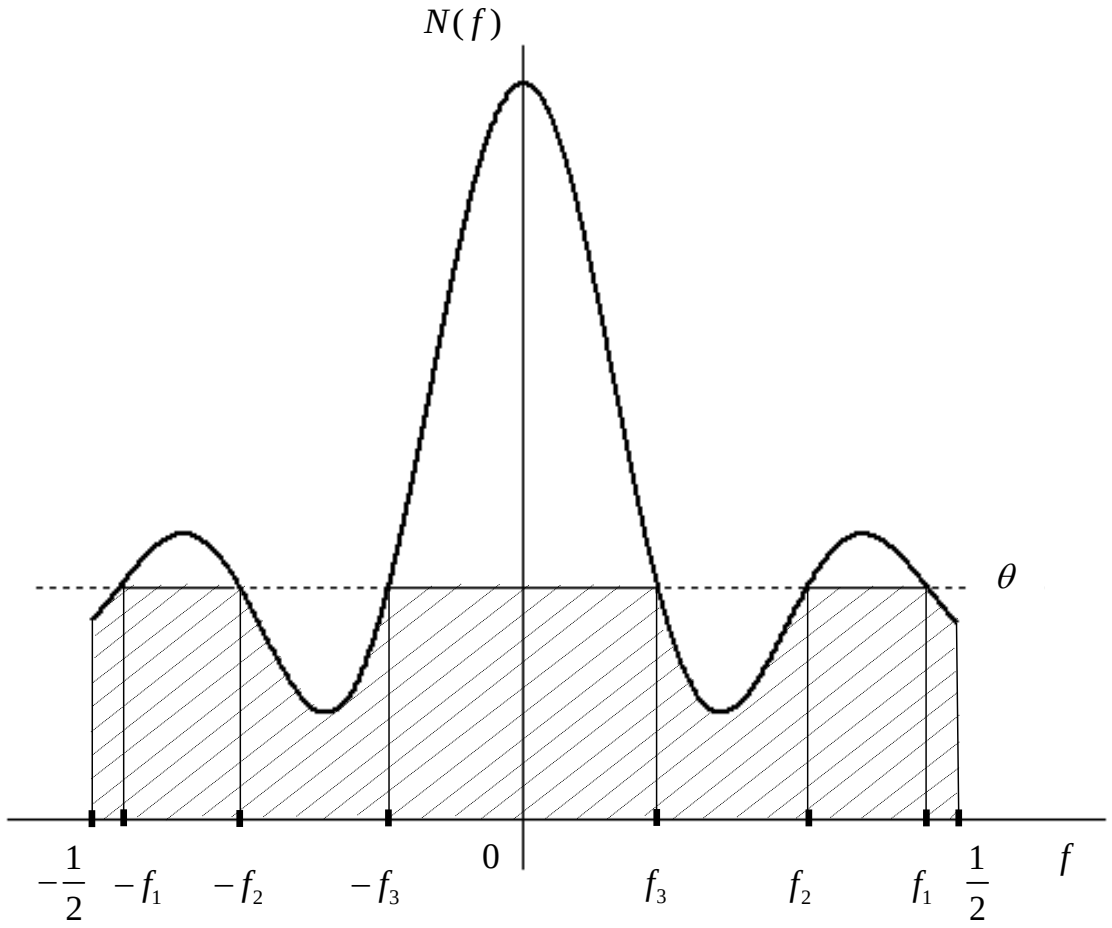
Rate-distortion function for source with memory

Theorem 3.2 The rate-distortion function $H(D)$ with squared error distortion for the discrete time stationary random Gaussian process with bounded and integrable spectral density $N(f)$ is computed as

$$H(D) = \frac{1}{2} \int_{-1/2}^{1/2} \log_2 \left\{ \max \left\{ 1, \frac{N(f)}{\theta} \right\} \right\} df, \tag{3.4}$$

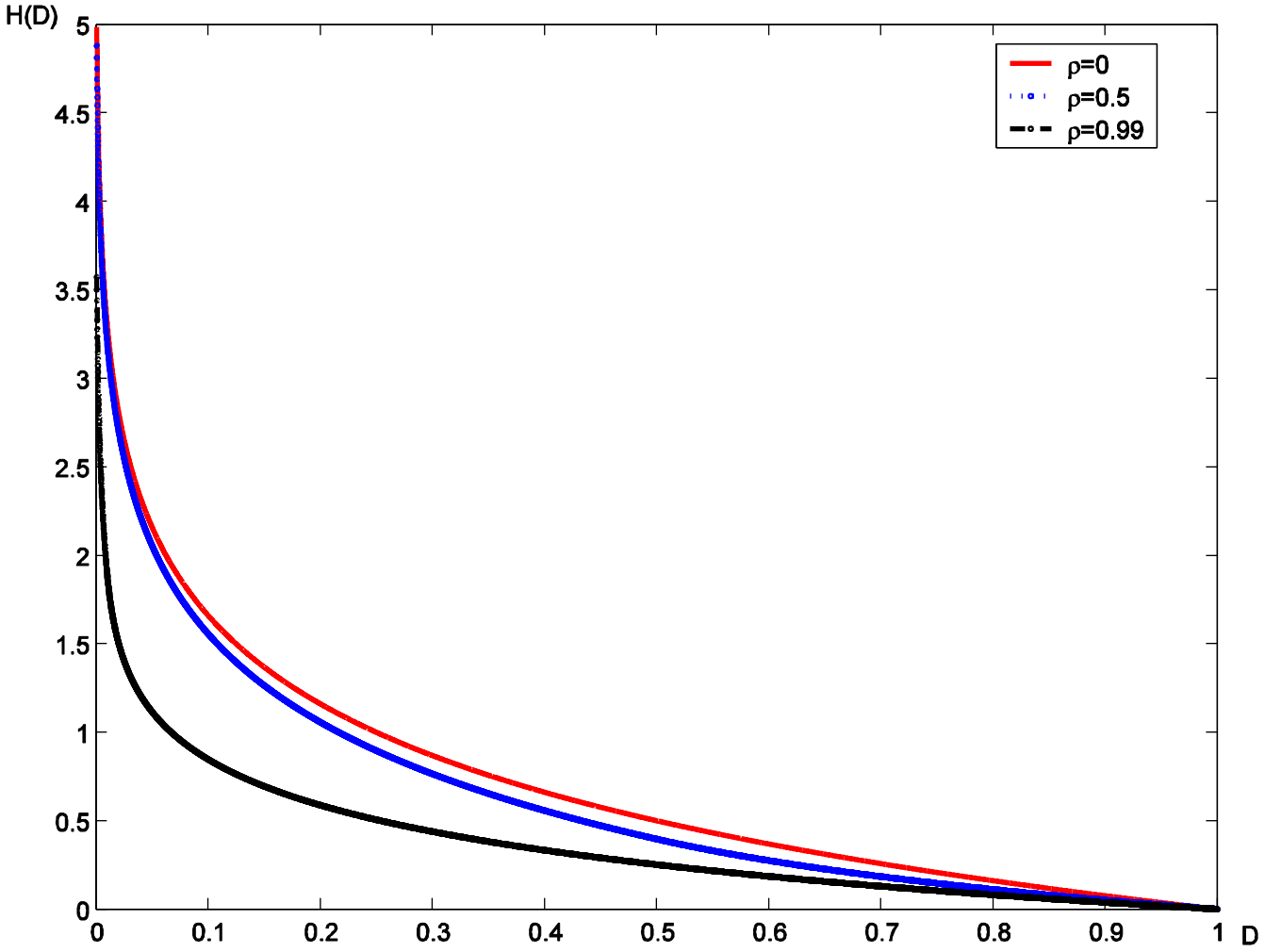
$$\int_{-1/2}^{1/2} \min \{ \theta, N(f) \} df = D.$$

Rate-distortion function for source with memory



Interpretation (3.4) by using “water-filling”.

Rate-distortion function for source with memory



Coding theorems

Theorem 3.3 For discrete time stationary continuous source with rate-distortion function $H(D)$ with respect to the squared error $d(x,y)=(x-y)^2$ there exists such n_0 that for all $n > n_0$ and for $\delta_1 > 0$ and $\delta_2 > 0$ there exists an (R, D_n) code of codelength n with coderate $R \leq H(D)+\delta_1$ for which the MSE D_n is less than or equal to $D+\delta_2$.

Theorem 3.4 (Converse of Th.3.3). For the source from Theorem 3.3 there does not exist a code for which simultaneously the MSE would be less than or equal to D and $R < H(D)$.

Comparison of quantization procedures

For the uniform scalar quantizer if the quantization step δ is small enough we can assume that pdf is constant inside each cell, that is $P(y_i) \approx \delta f(y_i)$. Then we obtain

$$D \approx \sum_j \int_{\Delta_j} (x - y_j)^2 f(y_j) dx \approx \sum_j \frac{P(y_j)}{\delta} \int_{y_j - \delta/2}^{y_j + \delta/2} (x - y_j)^2 dx = \frac{\delta^2}{12}.$$

The rate of such a quantizer: $$R(D) = h(X) - \frac{1}{2}\log_2(12D),$$

$h(X) = -\int_X f(x)\log_2 f(x)dx$ is the differential entropy of X.

For a memoryless source: $H(D) \geq R_L(D) = h(X) - \frac{1}{2}\log_2(2\pi eD)$

$$R(D) - R_L(D) \leq \frac{1}{2}\log_2 \frac{\pi e}{6} \approx 0.2546$$

Comparison of quantization procedures

Assume that the number of cells of the lattice quantizer is large and pdf is constant within each cell $S_i,\ i=1,\ldots,M$

$$D_n \approx \frac{1}{n}\frac{\int_S \|\mathbf{x}\|^2 d\mathbf{x}}{\mathrm{Vol}(S)},$$

$$R \approx h(X) - \frac{1}{n}\log_2 \mathrm{Vol}(S),$$

$$\mathrm{Vol}(S) = \int_S d\mathbf{x}$$ is the volume of the Voronoi cell S.

$$R(D) \approx R_L(D) + \frac{1}{2}\log_2 2\pi e G_n,$$

$$G_n = \frac{1}{n}\frac{\int_S \|\mathbf{x}\|^2 d\mathbf{x}}{\mathrm{Vol}(S)^{1+2/n}}$$

is the normalized second moment of the Voronoi region

Comparison of quantization procedures

The following bounds on G_n hold

$$\frac{1}{(n+2)\pi}\Gamma\left(\frac{n}{2}+1\right)^{2/n} \le G_n \le \frac{1}{n\pi}\Gamma\left(\frac{n}{2}+1\right)^{2/n}\Gamma\left(1+\frac{2}{n}\right),$$

$$G_n = \frac{1}{(n+2)\pi}\Gamma\left(\frac{n}{2}+1\right)^{2/n}$$ is the normalized second moment of the n-dimensional sphere .

If $n \to \infty$ $G_n \to \frac{1}{2\pi e} = 0.058550...$ and $R(D) \approx R_L(D)$

for a memoryless source or $R(D) \approx H(D)$ for a Gaussian memoryless source.

Pdf of the generalized Gaussian distribution

$$f(x)=\frac{\alpha\gamma(\alpha,\sigma)}{2\Gamma(1/\alpha)}\exp\{-(\gamma(\alpha,\sigma)|x-m|^{\alpha}\},$$

where m is the mathematical expectation, σ^2 is the variance,

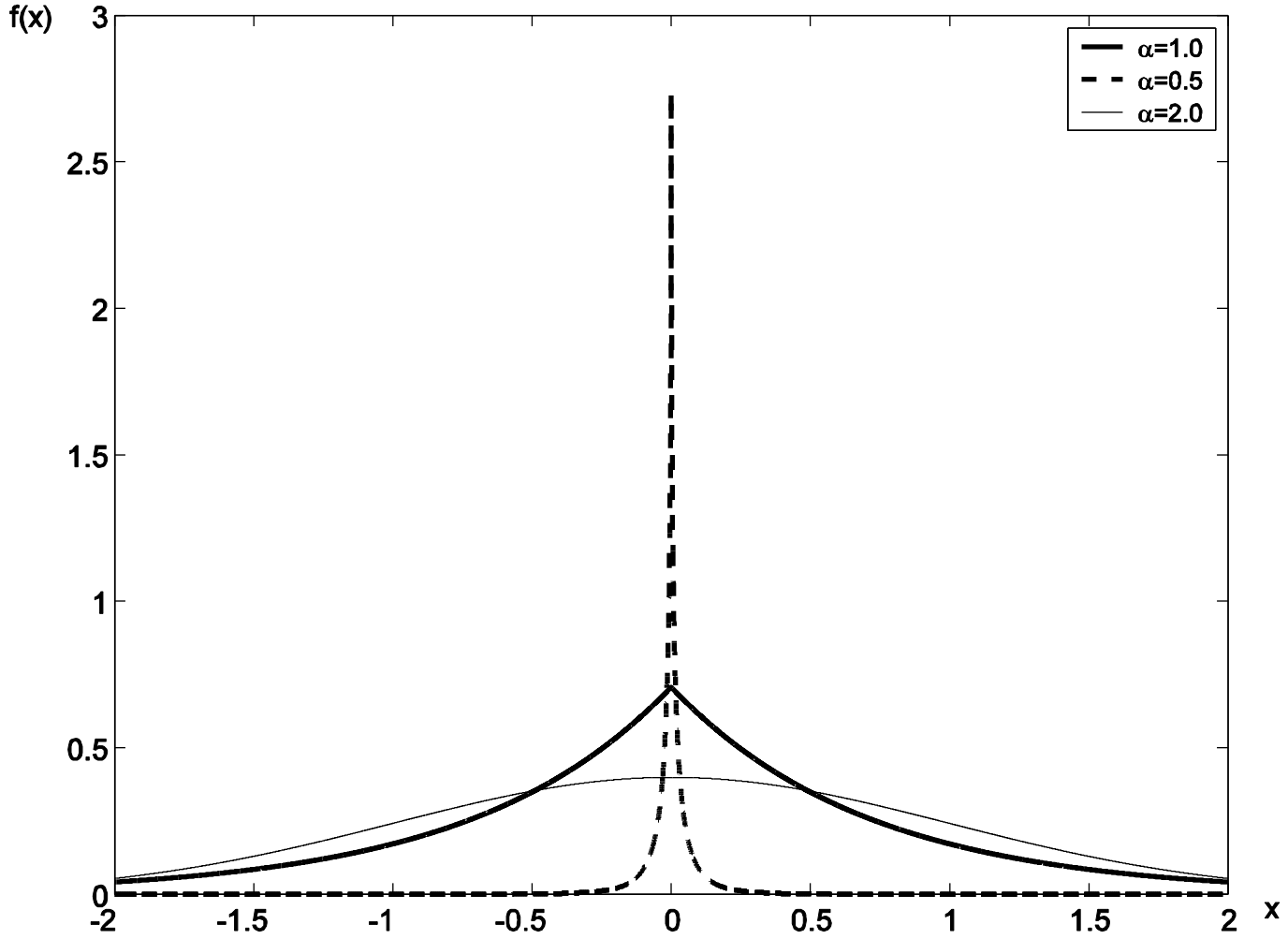
α is a parameter,

$$\gamma(\alpha,\sigma)=\sigma^{-1}\left[\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}\right].$$

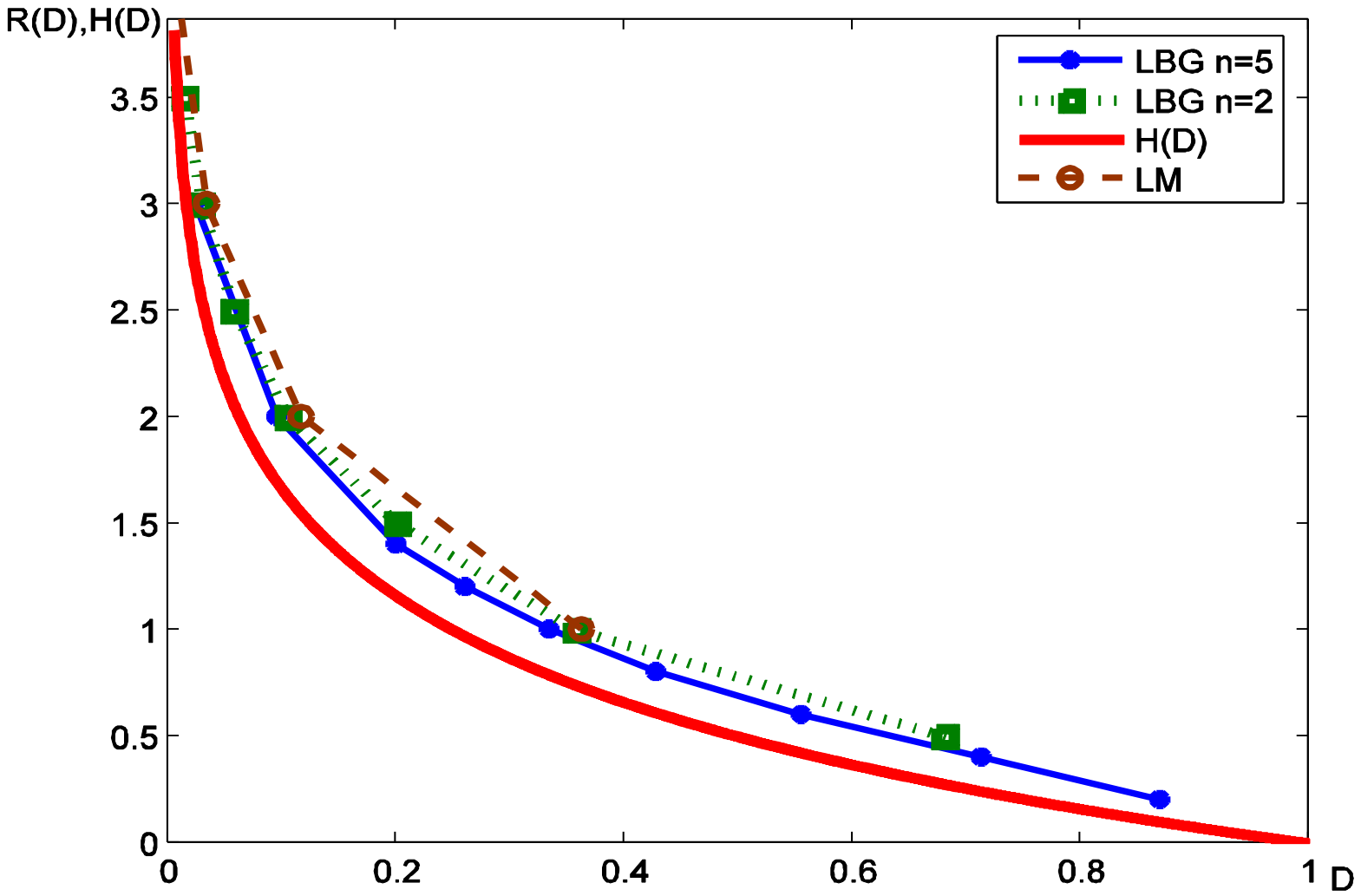
$$\Gamma(x)=\int_0^{\infty}e^{-t}t^{x-1}dt,$$

$$\Gamma(1)=1,\Gamma(x+1)=x\Gamma(x),$$
$$\Gamma(n+1)=n!,$$
$$\Gamma(1/2)=\sqrt{\pi}.$$

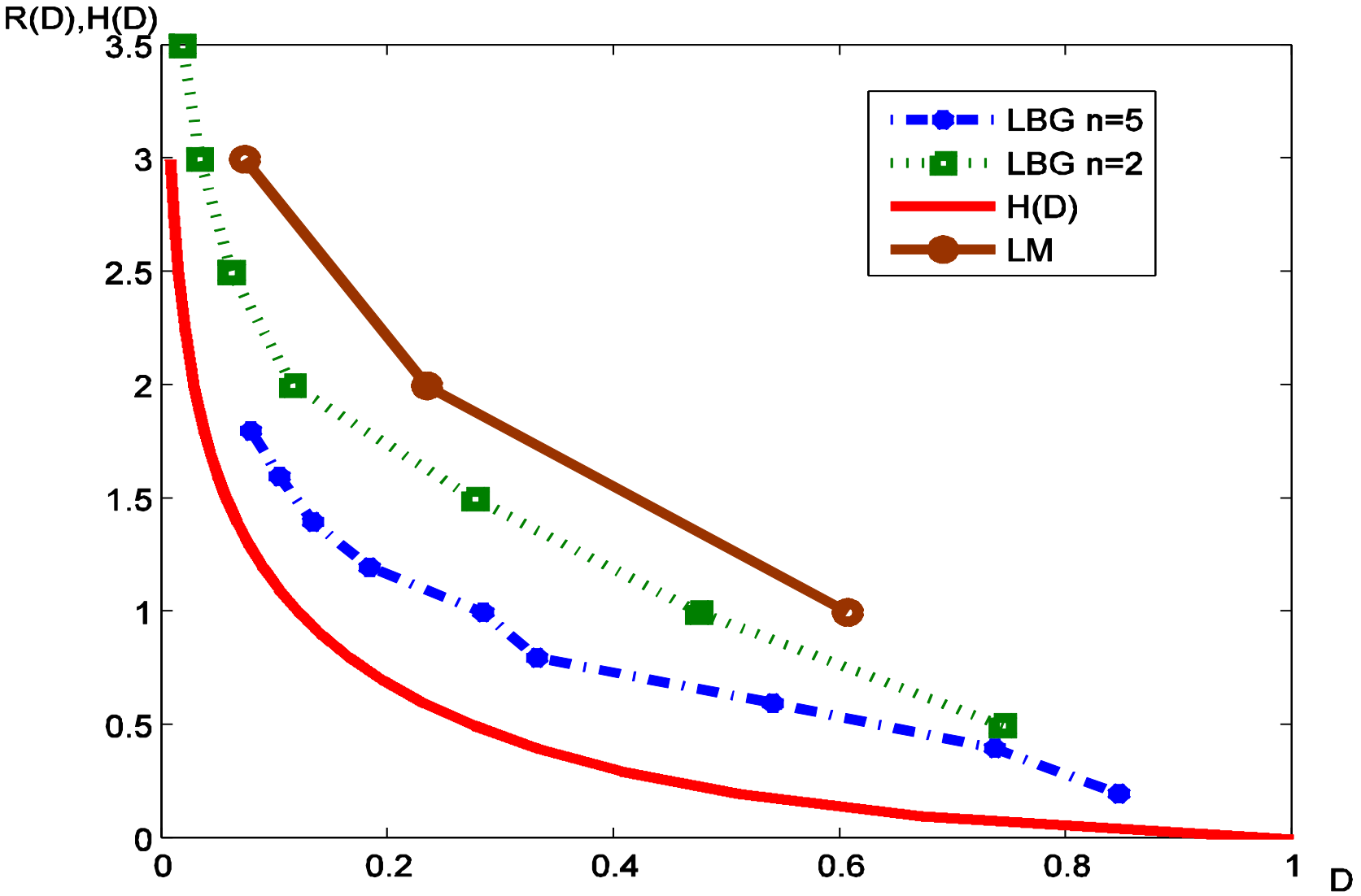
Pdf of the generalized Gaussian distribution



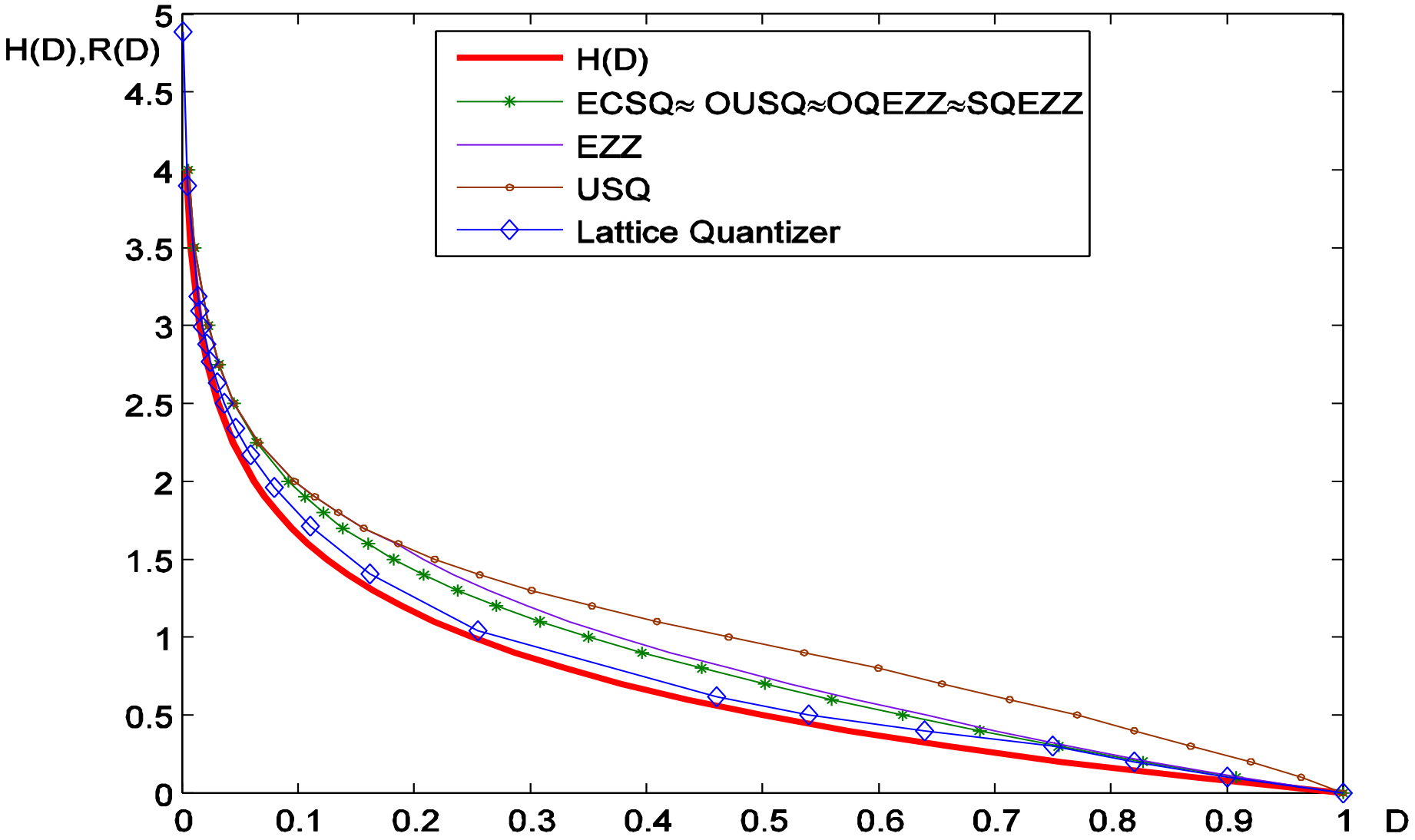
Comparison of fixed-rate quantizers ($\alpha = 2.0$)



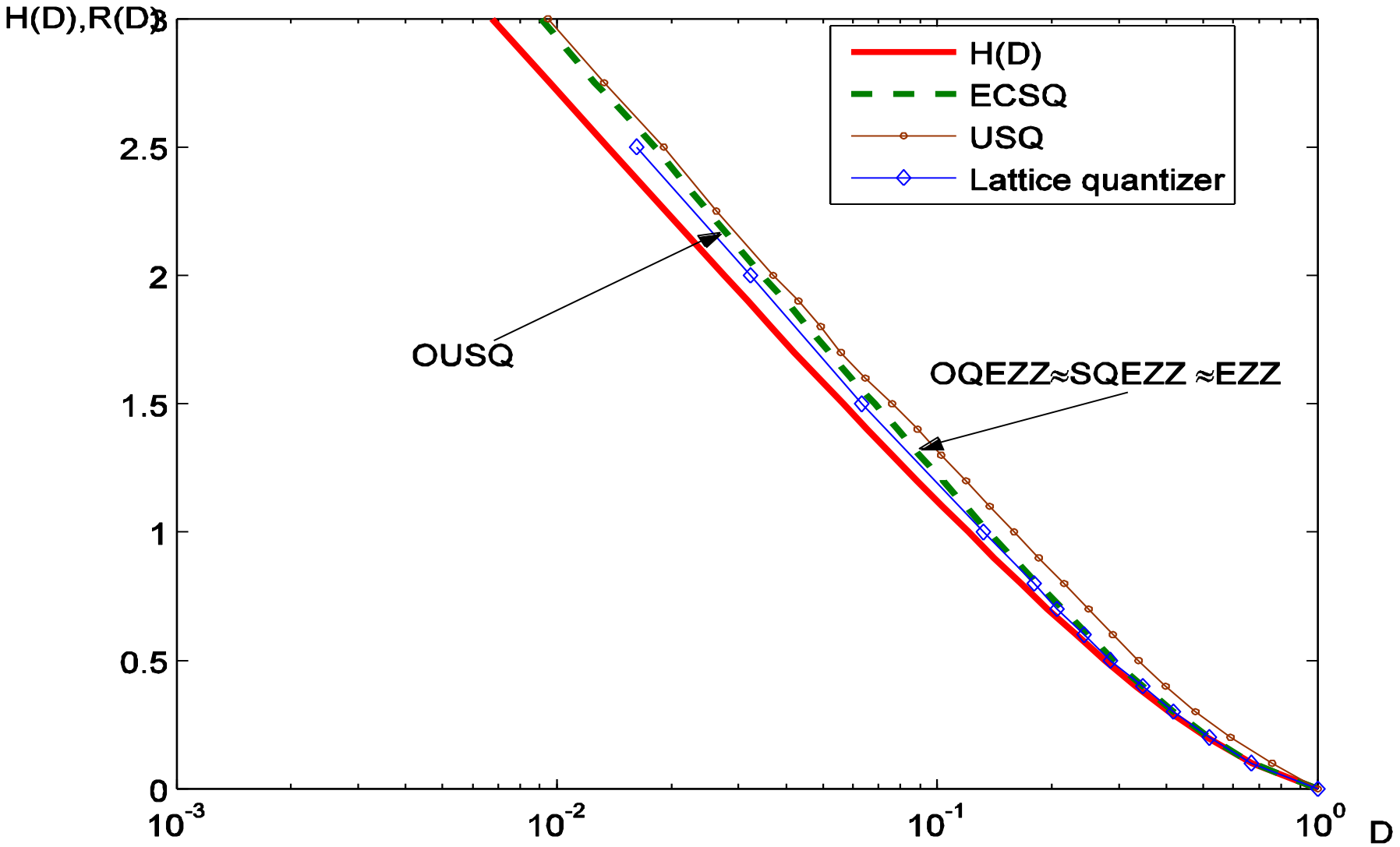
Comparison of fixed-rate quantizers ($\alpha = 0.5$)



Comparison variable-rate quantizers($\alpha = 2.0$)



Comparison variable-rate quantizers($\alpha = 0.5$)



Comparison of variable-rate quantization procedures

Fixed-rate quantizers:

•Nonuniform scalar quantizer is always better than uniform scalar quantizer

•Using scalar nonuniform quantizer it is impossible to obtain coderate less than 1bit/sample

•Vector quantization reduces redundancy $\max_D(R(D)-H(D))$ compared to scalar quantizer. Besides that VQ provides coderates less than 1 bit/sample

•VQ for the Gaussian stationary process with $n \to \infty$ might lead to the curve $R(D)$ lying below $H(D)$ for the Gaussian memoryless source.

Comparison of quantization procedures

For variable-rate quantizers:

- For rates greater than 2 bits/sample uniform quantizer (rounding off) and optimal scalar quantizer (ECSQ) provide almost the same $R(D)$ which is close to $R_L + 0.255$
- The uniform (rounding off) quantizer followed by a variable-length coder can provide coderates less than 1 bit/sample
- For rates less than 2 bit/sample uniform quantizer (rounding off) has worse performance than ECSQ.
- For rates less than 2 bits/sample suboptimal scalar quantizers with extended zero zone and optimal uniform quantizer have performances rather close to the performances of ECSQ.
- The rate-distortion function of the lattice quantizer coincides with $R(D)$ of ECSQ for low rates. For high rates the lattice quantizer wins 0.05-0.1 bits/sample compared to ECSQ.

Comparison of quantization procedures

•For memoryless sources VQ reduces redundancy compared to scalar quantization.

•For sources with memory all types of scalar quantizers provide rate-distortions which are rather far from the theoretical limit for the given source.

•VQ for sources with memory can reduce a quantization rate compared to scalar quantization but at the cost of an unacceptable increase of computational complexity.

Characteristics of digital speech, audio, image, and video signals

Speech and audio				
Speech/audio type	Frequency band	Sampling rate	Bits/sample	Uncompressed bit rate
Speech	200-3200Hz	8 kHz	16	128kb/s
CD audio	20-20000Hz	44.1kHz	16x2 channels	1.41 Mb/s
Still image				
Image type	Pixels per frame	Bits/Pixel		Uncompressed bit rate
FAX	3120×2040	1		6.36 Mb
VGA	640×480	8		2.46 Mb
XVGA	1024×768	24		18.87 Mb
Video				
Video type	Pixels per frame	Frame per Second	Bits/pixel	Uncompressed bit rate
NTSC	700×525	30	16	176.4 Mb/s
PAL	833×625	25	16	208.3 Mb/s
CIF	352×288	15	12	18.2 Mb/s
QCIF	176×144	10	12	3.0 Mb/s
HDTV	1280×720	60	12	622.9 Mb/s
HDTV	1920×1080	30	12	745.7 Mb/s