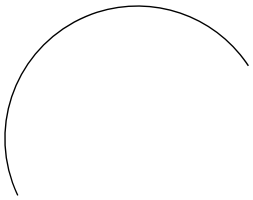
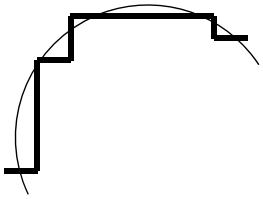
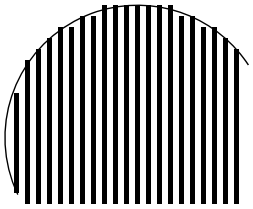
Types of Signals



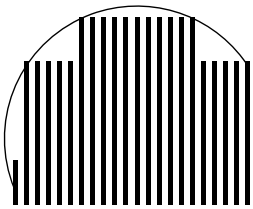
time - continuous & amplitude - continuous
(Analog)



time - continuous & amplitude - discrete



time - discrete & amplitude - continuous



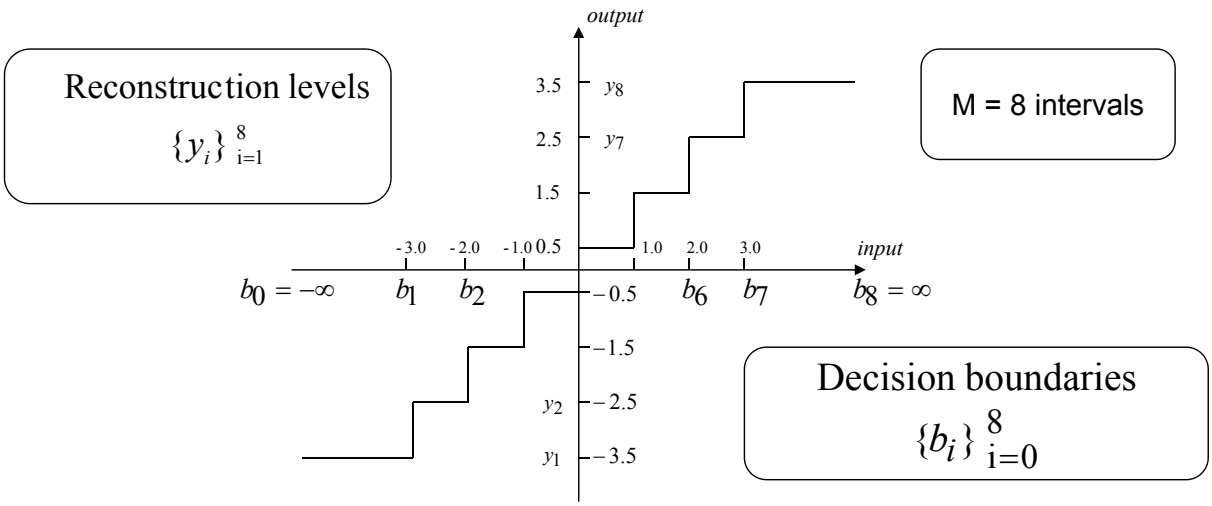
time - discrete & amplitude - discrete
(Digital)

- Time-discretization (sampling) : information preserving operation
- Amplitude-discretization (quantizing) : information lossy operation

Scalar Quantization (1)

- Quantization
  - the number of possible distinct source output values is generally much larger than the number of codewords available to represent them
  - the process of representing a large – possibly infinite – set of values with a much smaller set
- Scalar Quantization
  - the set of inputs and outputs of a quantizer are scalars
- Vector Quantization
  - the set of inputs and outputs of a quantizer are vectors
- Encoder Mapping [analog-to-digital (A/D) converter]
  - The encoder divides the range of values that the source generates into a number of intervals. Each interval is represented by a distinct codeword
- Decoder Mapping [digital-to-analog (D/A) converter]
  - For each codeword, the decoder generates a reconstruction value that best represents all the values in the corresponding interval.

Scalar Quantization (2)



- The input amplitude x is quantized as y_i if it falls into the interval $\{ b_{i-1} < x \le b_i \}$, i = 1,…, M, and the value y_i is used to represent all amplitudes of the interval.
  i.e. $Q(x) = y_i$ iff $b_{i-1} < x \le b_i$

Scalar Quantization (3)

- Quantization error (quantization distortion or quantization noise) : $x - Q(x)$
  - mean squared quantization error (msqe):

$$\sigma_q^2 = \int_{-\infty}^{\infty} (x - Q(x))^2 f_X(x)dx = \sum_{i=1}^{M} \int_{b_{i-1}}^{b_i} (x - y_i)^2 f_X(x)dx \qquad (1)$$

- The rate (R) of the quantizer is the average number of bits required to represent a single quantizer output
  - Fixed-length coding : $R = \lceil \log_2 M \rceil$.

Given an input $f_X(x)$ and M, find $\{b_i\}$ and $\{y_i\}$ so as to minimize σ_q^2.

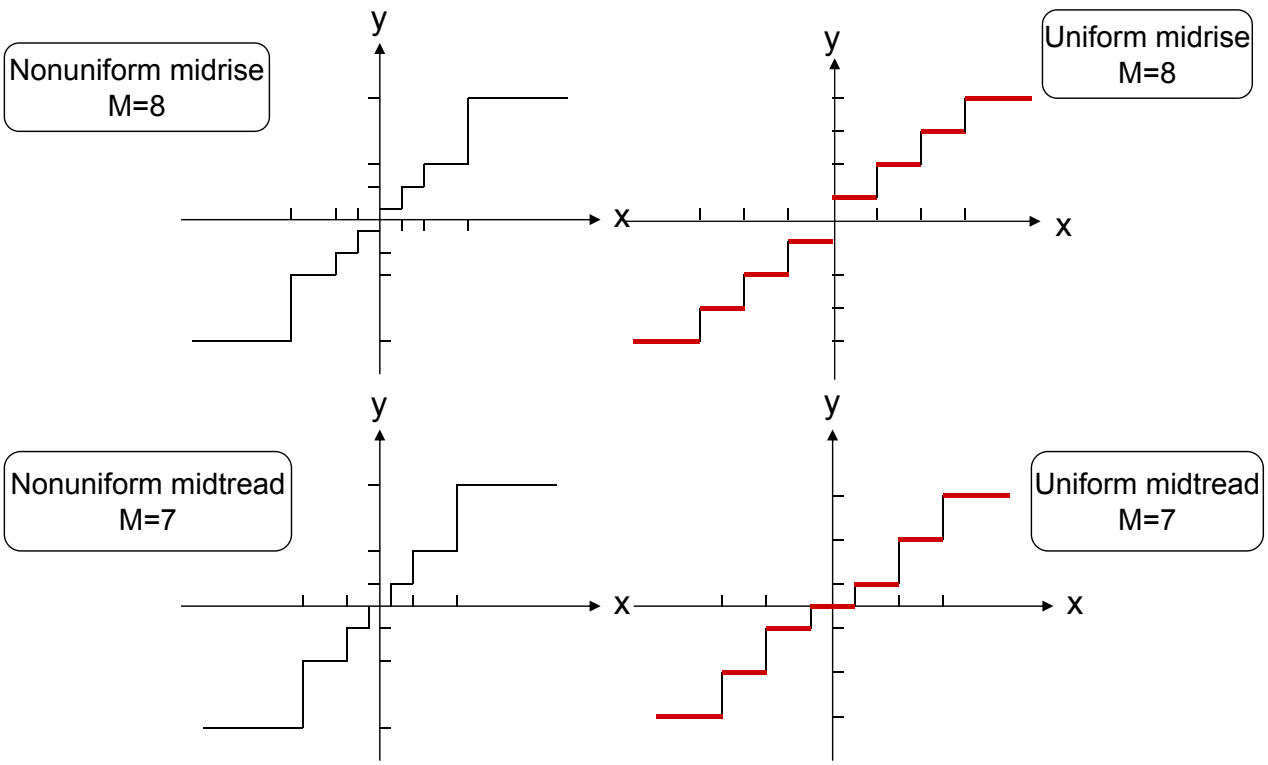
- Variable-length coding : $R = \sum_{i=1}^{M} l_i P(y_i)$, where l_i is the length of the codeword corresponding to y_i, and $P(y_i) = \int_{b_{i-1}}^{b_i} f_X(x)dx$.

Given an input $f_X(x)$, M, and

① $\sigma_q^2 \le D^*$, find $\{b_i\}$, $\{y_i\}$, and binary code so as to minimize R.

② $R \le R^*$, find $\{b_i\}$, $\{y_i\}$, and binary code so as to minimize σ_q^2.

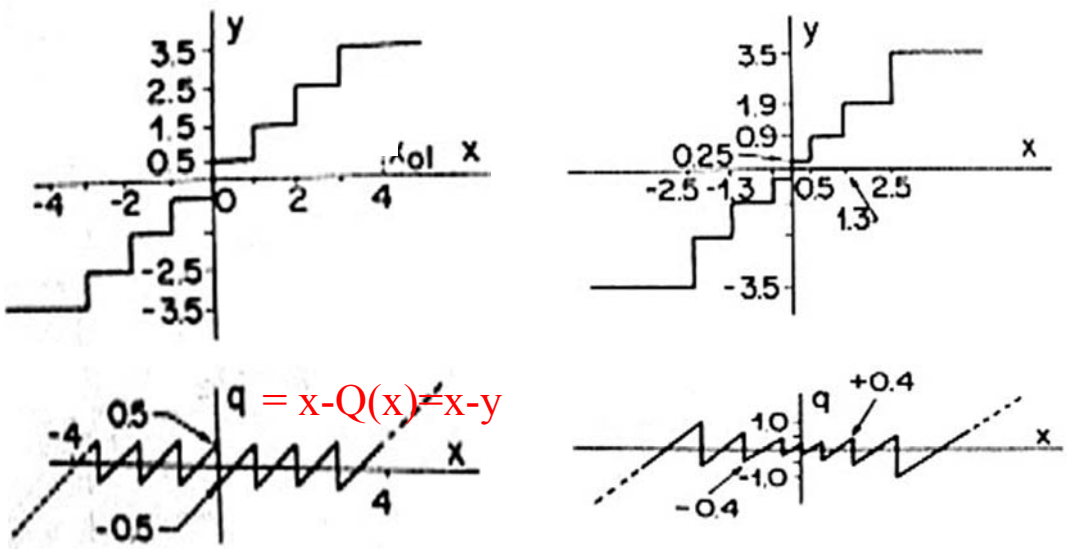
Quantizer Types



Quantization Error

Uniform quantizer

Nonuniform quantizer



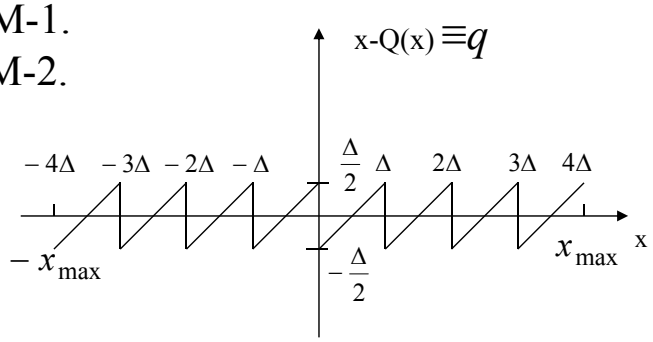
Uniform Quantizer (1)

- Uniform Quantizer

$y_{i+1} - y_i = \Delta$; for $i = 1, 2, \ldots, M-1$.

$b_{i+1} - b_i = \Delta$; for $i = 1, 2, \ldots, M-2$.

Q: why



1. Uniformly Distributed Source

$x \in [-X_{max}, X_{max}]$ with $-X_{max} = b_1 - \Delta$, $X_{max} = b_{M-1} + \Delta$

$$\Rightarrow \Delta = \frac{2X_{max}}{M}$$

$$\sigma_q^2 = \sum_{i=1}^{M} \int_{b_{i-1}}^{b_i} (x - y_i)^2 f_X(x)dx = 2\sum_{i=1}^{M/2} \int_{(i-1)\Delta}^{i\Delta} \left(x - \frac{2i-1}{2}\Delta\right)^2 \frac{1}{2X_{max}} dx$$

$= ?$

$$= \frac{1}{2X_{max}} \times 2 \times \frac{M}{2} \times \int_0^{\Delta} \left(x - \frac{\Delta}{2}\right)^2 dx = \frac{\Delta^2}{12} = \boxed{\frac{1}{\Delta}\int_{-\Delta/2}^{\Delta/2} q^2 dq}$$

Uniform Quantizer (2)

- Quantization error

$q = x - y_i = x - Q(x) \in \left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$

- Signal-to-Quantization noise ratio

$$SNR(dB) = 10\log_{10}\left(\frac{\sigma_s^2}{\sigma_q^2}\right) = 10\log_{10}\left(\frac{X_{max}^2}{3}\frac{12}{\Delta^2}\right) = 10\log_{10}\left(\frac{(2X_{max})^2}{\Delta^2}\right)$$

$$= 10\log_{10}(M^2) = 20\log_{10}(2^n) = 20n\log_{10}2$$

$$= 6.02n \quad dB$$

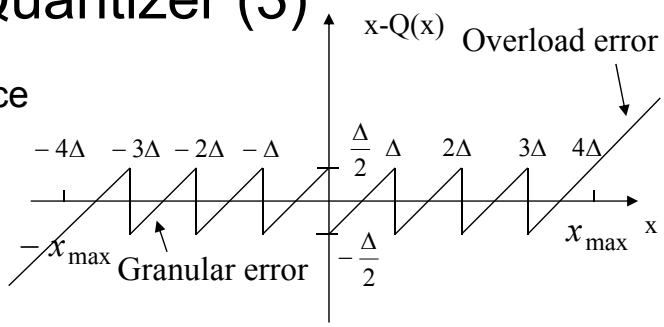
Every additional bit in the quantizer ⇒ the SNR will increase 6.02 dB

$$where \quad \sigma_s^2 = \int_{-x_{max}}^{x_{max}} x^2 \frac{1}{2X_{max}} dx = \frac{x_{max}^2}{3}$$

and $M = 2^n$ (i.e. an n-bits uniform quantizer)

Uniform Quantizer (3)

2. Nonuniformly Distributed Source



$$\sigma_q^2 = \sum_{i=1}^{M} \int_{b_{i-1}}^{b_i} (x - y_i)^2 f_X(x)dx$$

$$= \underbrace{2\sum_{i=1}^{\frac{M}{2}-1} \int_{(i-1)\Delta}^{i\Delta} \left(x - \frac{2i-1}{2}\Delta\right)^2 f_X(x)dx}_{\text{Granular error}} + \underbrace{2\int_{\left(\frac{M}{2}-1\right)\Delta}^{\infty} \left(x - \frac{M-1}{2}\Delta\right)^2 f_X(x)dx}_{\text{Overload error}}$$

?

- To find the optimal value of Δ

$$\frac{\partial \sigma_q^2}{\partial \Delta} = -\sum_{i=1}^{\frac{M}{2}-1} (2i-1)\int_{(i-1)\Delta}^{i\Delta} \left(x - \frac{2i-1}{2}\Delta\right) f_X(x)dx$$

$$-(M-1)\int_{(\frac{M}{2}-1)\Delta}^{\infty} \left(x - \frac{M-1}{2}\Delta\right) f_X(x)dx = 0$$

- $\Delta\uparrow \Rightarrow$ size of overload region $\downarrow$
  $\Rightarrow$ overload error $\downarrow$
- $\Delta\uparrow \Rightarrow$ granular error$\uparrow$

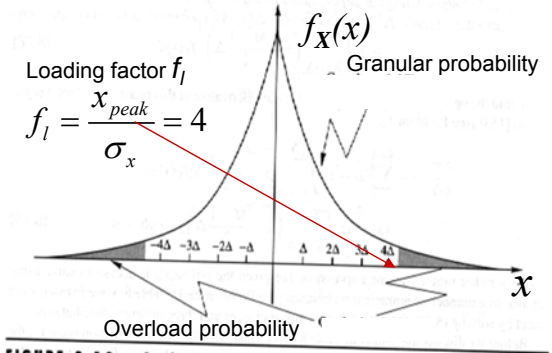


FIGURE 8.10 Overload and granular regions for a 3-bit uniform quantizer.

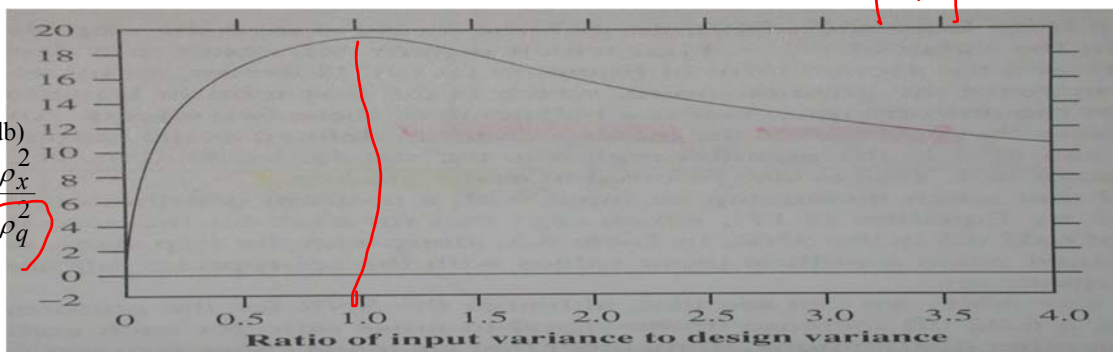
Uniform Quantizer (4)

TABLE 8.3 Optimum step size and SNR for uniform quantizers for different distributions and alphabet sizes [153, 4].

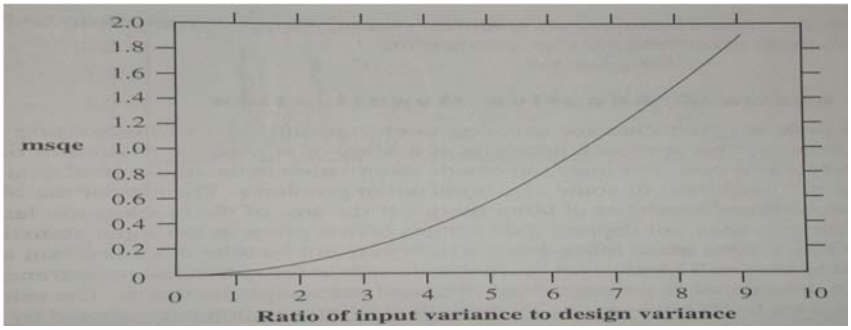
Alphabet Size (M)	Uniform Step Size (Δ)	Uniform SNR		Gaussian Step Size	Gaussian SNR		Laplacian Step Size	Laplacian SNR	
2	1.732	6.02		1.596	4.40		1.414	3.00	
4	0.866	12.04	6.02	0.9957	9.24	4.84	1.0873	7.05	4.05
6	0.577	15.58		0.7334	12.18		0.8707	9.56	
8	0.433	18.06		0.5860	14.27		0.7309	11.39	
10	0.346	20.02		0.4908	15.90		0.6334	12.81	
12	0.289	21.60		0.4238	17.25		0.5613	13.98	
14	0.247	22.94		0.3739	18.37		0.5055	14.98	
16	0.217	24.08		0.3352	19.36		0.4609	15.84	
32	0.108	30.10		0.1881	24.56		0.2799	20.46	

Uniform Quantizer (5)

$$\text{SNR(db)} = 10\log\frac{\rho_x^2}{\rho_q^2}$$



Effect of variance mismatch on the performance of 4-bit Gaussian uniform quantizer



Adaptive Quantization (AQ)

- Adaptation of quantizer
  - variances of input
  - pdf of input
  - mean of input

- Forward Adaptation (AQF) ← off-line
  - the quantizer adapts to changes in N future input variances e.g. at time n

$$\hat{\sigma}_q^2 = \frac{1}{N}\sum_{i=0}^{N-1} x_{n+i}^2$$

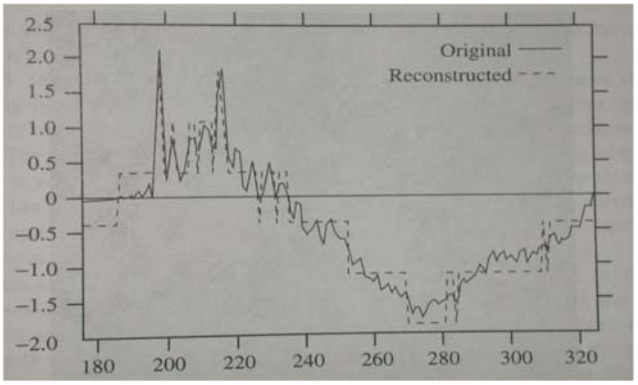
- Backward Adaptation (AQB) ← on-line
  - only the past quantized samples are used in adapting the quantizer
  - **Jayant quantizer** (quantization with one word memory)
  - adjust Δ after observing a single output

if the input falls in the outer levels ® Δ

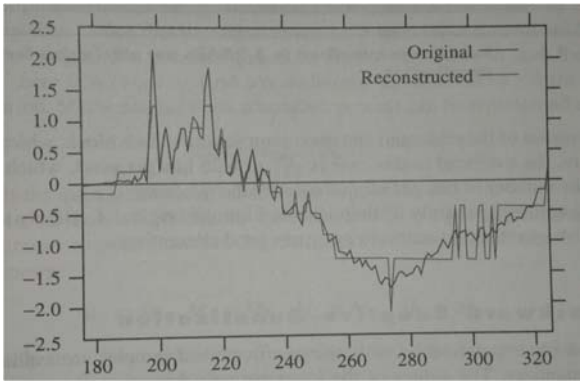
if the input falls in the inner quantizer levels ® Δ

Forward Adaptive Quantization

- Example
  - The step size is adjusted based on the statistics of the entire sequence
    - 16-bit per sample : original signal
    - 3-bit quantizer is used to quantize a segment of speech
    - 8-bit quantizer is used to quantize the side information (standard derivation for each block of 128 samples) (i.e 決定 step size,這時假設 pdf & mean 不變)



Original 16-bit speech and compressed 3-bit speech sequences



Original 16-bit speech and AQF quantized sequences

Backword Adaptive Quantization

- Jayant quantizer

– adaption : step size for the current input is modified based on the previous quantizer output $\Delta_n = M_{l(n-1)}\Delta_{n-1}$

Where $M_{l(n-1)}$ is the multiplier for quantization interval $l(n-1)$ at time $(n-1)$

- Example

– 3-bit unifrom quantizer

– input 0.1, -0.2, 0.2, 0.1, -0.3 ...

$\Delta_0 = 0.5$

$1 > M_0 \le M_1 \le M_2 \le M_3 > 1$

0.8 0.9 1 1.2

n	Δ_i	Input	Ouput level	Ouput
0	0.5	0.1	0	0.25
1	0.4	-0.2	4	-0.2
2	0.32	0.2	0	0.16
3	0.256	0.1	0	0.128
4	0.2048	-0.3	5	-0.3072

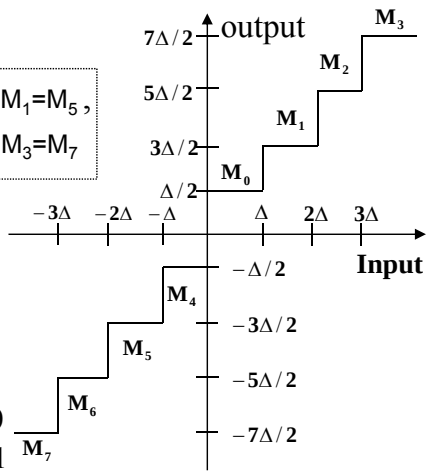
$\Delta_1 = M_0 \times \Delta_0$

$\Delta_2 = M_4 \times \Delta_1$

$\Delta_3 = M_0 \times \Delta_2$

$\Delta_4 = M_0 \times \Delta_3$

$M_0=M_4$, $M_1=M_5$, $M_2=M_6$, $M_3=M_7$



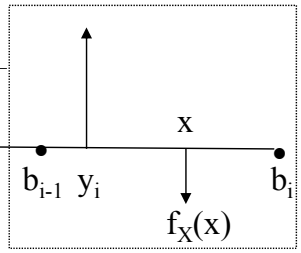
Nonuniform Quantization

Choose smaller qunatization intervals in regions of higher probability
& Choose larger quantization intervals otherwise

質心點

- PDF-Optimized Quantization

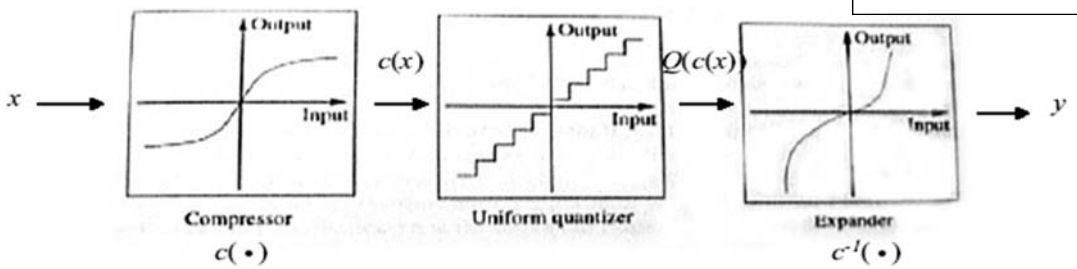
$$\sigma_q^2 = \sum_{i=1}^{M} \int_{b_{i-1}}^{b_i} (x - y_i)^2 f_x(x)dx \Rightarrow \begin{cases} \dfrac{\partial \sigma_q^2}{\partial y_i} = 0 \Rightarrow y_i = \dfrac{\int_{b_{i-1}}^{b_i} x f_x(x)dx}{\int_{b_{i-1}}^{b_i} f_x(x)dx} \\ \dfrac{\partial \sigma_q^2}{\partial b_i} = 0 \Rightarrow b_i = \dfrac{y_{i+1} + y_i}{2} \end{cases}$$



y_i = ? When $f_X(x)$ = constant

- Lloyd-Max algorithm
  - to calculate the optimal values b_i and y_i iteratively

- Companded Quantization



Theorem: f is continuous at [a,b]. If $F(x) = \int_c^x f(t)dt,\ c \in [a,b]$,

then $F'(x) = f(x)\ \forall x \in [a,b]$.

$$\frac{\partial\left[\int_{b_{i-1}}^{b_i} (x-y_i)^2 f_X(x)dx + \int_{b_i}^{b_{i+1}} (x-y_{i+1})^2 f_X(x)dx + \right]}{\partial b_i} = 0$$

$$\Rightarrow (b_i - y_i)^2 f_X(b_i) - (b_i - y_{i+1})^2 f_X(b_i) = 0$$

$$\Rightarrow \because f_X(b_i) \neq 0 \therefore b_i = (y_i + y_{i+1})/2.$$

$$F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} \simeq \frac{f(x) \cdot h}{h} = f(x)$$

$$\frac{\partial \sigma_q^2}{\partial y_i} = 0 \Rightarrow \int_{b_{i-1}}^{b_i} 2(x - y_i) f_X(x)dx = 0 \quad \forall i$$

$$\Rightarrow \int_{b_{i-1}}^{b_i} x f_X(x)dx - y_i \int_{b_{i-1}}^{b_i} f_X(x)dx = 0 \quad \forall i \Rightarrow y_i = \frac{\int_{b_{i-1}}^{b_i} x f_X(x)dx}{\int_{b_{i-1}}^{b_i} f_X(x)dx}$$

Lloyd-Max algorithm (1)

- For an M-level symmetric midrise quantizer
  - to obtain $\{y_1, y_2, \ldots, y_{M/2}\}$ and $\{b_1, b_2, \ldots, b_{(M/2)-1}\}$.
    where $b_0 = 0$ and $b_{M/2}$ = peak value of the input.

$$y_i = \frac{\int_{b_{i-1}}^{b_i} x f_X(x)\,dx}{\int_{b_{i-1}}^{b_i} f_X(x)\,dx}$$

$$b_i = \frac{y_{i+1} + y_i}{2}$$

$y_1 \rightarrow b_1$

$\downarrow$

y_2

$y_2 \rightarrow b_2$

$\downarrow$

y_3

$\vdots$

$y_{(M/2)-1} \rightarrow b_{(M/2)-1}$

$\downarrow$

$y_{M/2}$

(1) Pick y_1

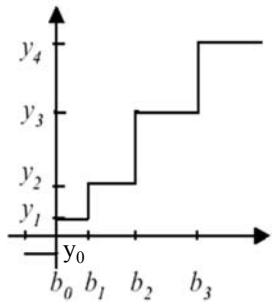
(2) Compute b_1 from $y_1 = \dfrac{\int_{b_0}^{b_1} x f_X(x)\,dx}{\int_{b_0}^{b_1} f_X(x)\,dx}$

(3) Compute $y_2 = 2b_1 - y_1$

(4) Repeat the process (1-3) until $y_{M/2}$ is obtained

(5) Check if $y_{M/2}$ is the centroid of the area of $f_X(x)$ between $b_{M/2-1}$ and $b_{M/2}$, namely,

$$y^*_{M/2} = \frac{\int_{b_{(M/2)-1}}^{b_{M/2}} x f_X(x)\,dx}{\int_{b_{(M/2)-1}}^{b_{M/2}} f_X(x)\,dx}$$



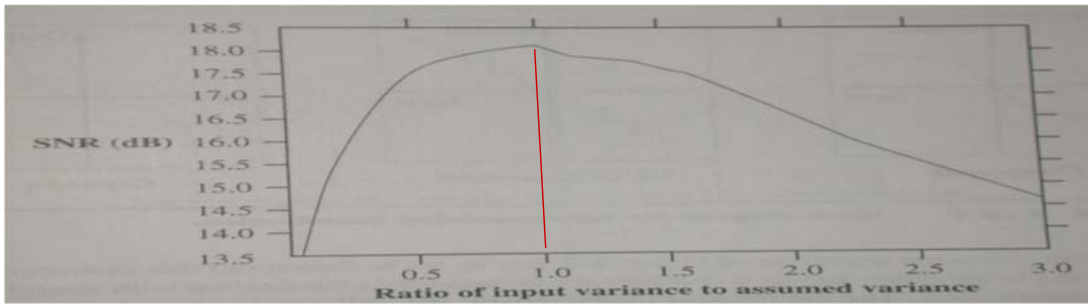
(6) If $|y_{M/2} - y^*_{M/2}|$ < threshold then stop
else $y_1 = y_1 - (y_{M/2} - y^*_{M/2})$ go to step 1

Q: $b_1 + (b_2 - b_1)/2 = y_2$? ×

Lloyd-Max algorithm (2)

TABLE 8.6 Quantizer boundary and reconstruction levels for nonuniform Gaussian and Laplacian quantizers.

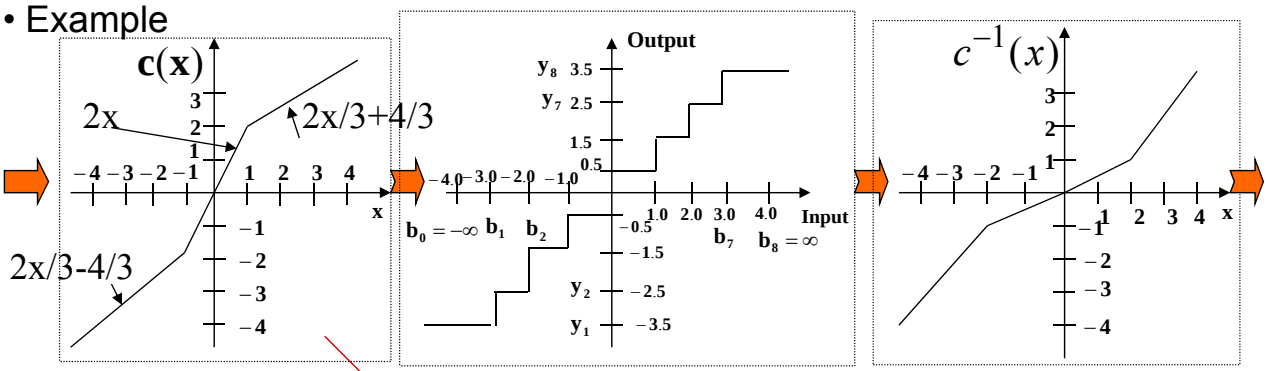
Levels (M)	Gaussian b_i	Gaussian y_i	Gaussian SNR	Laplacian b_i	Laplacian y_i	Laplacian SNR
4	0.0	0.4528		0.0	0.4196	
	0.9816	1.510	9.3 dB	1.1269	1.8340	7.54 dB
6	0.0	0.3177		0.0	0.2998	
	0.6589	1.0		0.7195	1.1393	
	1.447	1.894	12.41 dB	1.8464	2.5535	10.51 dB
8	0.0	0.2451		0.0	0.2334	
	0.7560	0.6812		0.5332	0.8330	
	1.050	1.3440		1.2527	1.6725	
	1.748	2.1520	14.62 dB	2.3796	3.0867	12.64 dB



Effect of mismatch on nonuniform quantization (4-bit Laplacian nonuniform quantizer)

Companded Quantization (1)

- Companding = *com*pressing and ex*pand*ing
- Example



input x=0.9

① uniform quantizer

$Q(x) = Q(0.9) = 0.5$

$q = 0.9 - 0.5 = 0.4$

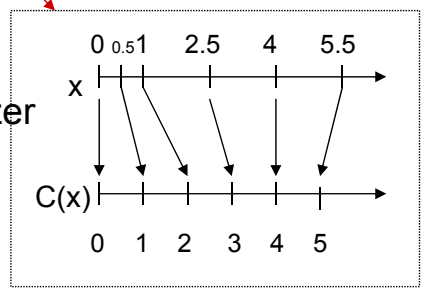
② companded quantizer

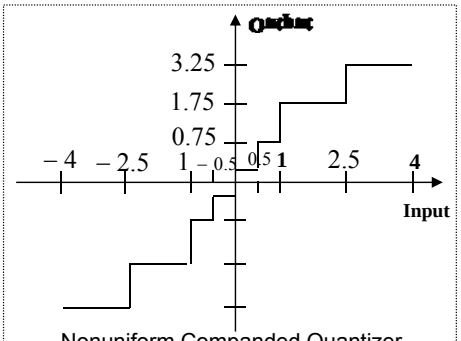
$c(x) = 1.8$

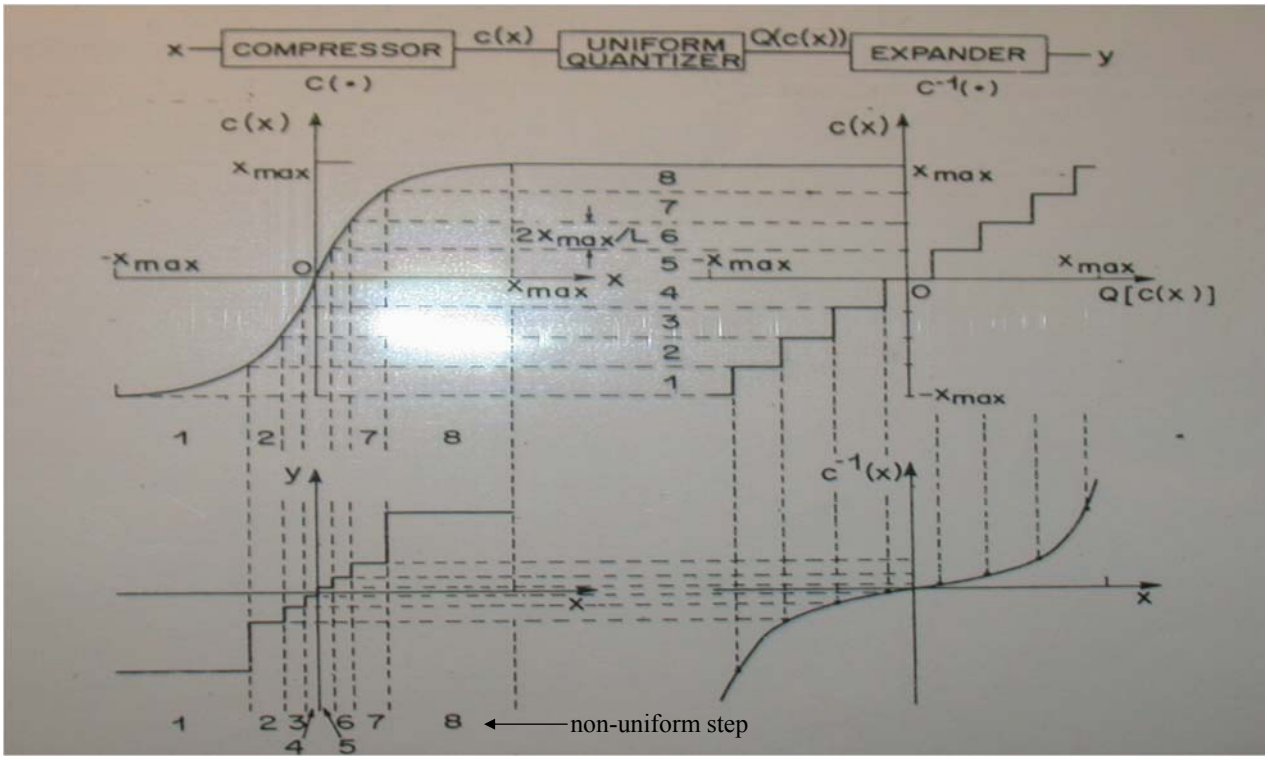
$Q(c(x)) = Q(1.8) = 1.5$

$c^{-1}(1.5) = 0.75$

$q = 0.9 - 0.75 = 0.15$





Nonuniform Companded Quantizer

Companded Quantization (2)