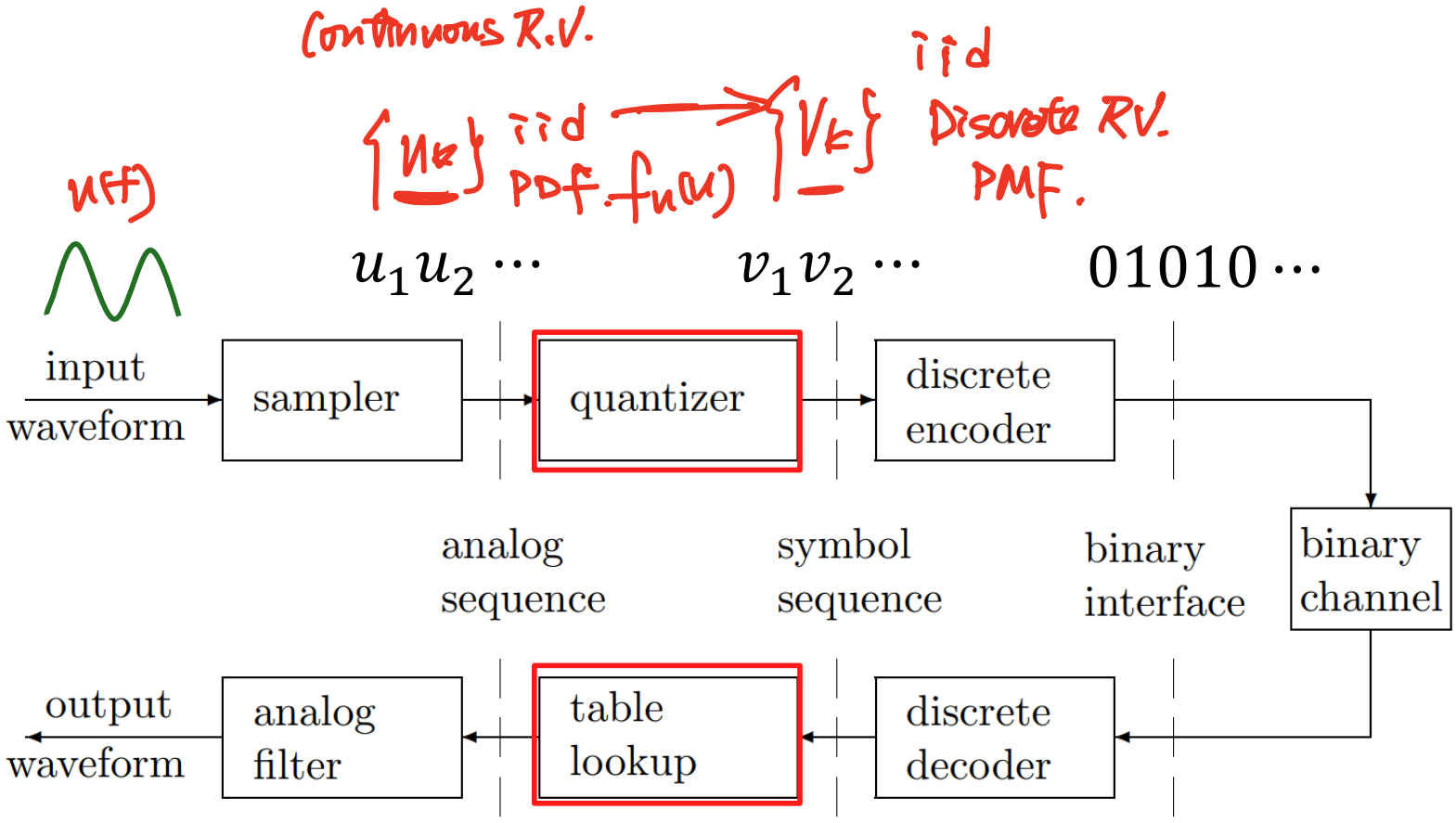
EE140 Introduction to Communication Systems

Lecture 12

Instructor: Prof. Lixiang Lian
ShanghaiTech University, Fall 2025

Quantization



Quantization

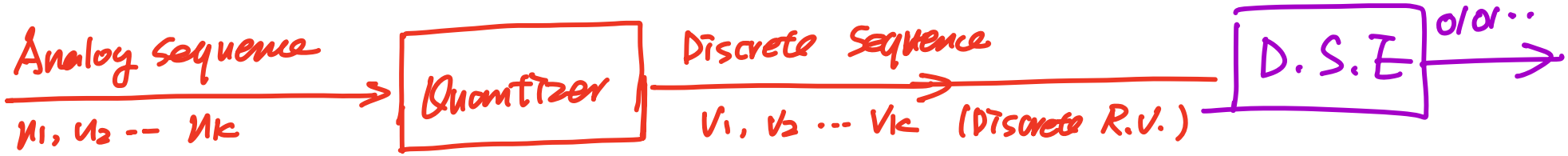
Objective: Map the incoming sequence $U_1, U_2, \cdots$ of analog rvs into a sequence of discrete rvs $V_1, V_2, \cdots$, where V_m should represent U_m with as little distortion as possible.

- Scalar Quantization: Each analog RV in the sequence is quantized independently of the other RVs.
- Vector Quantization: The analog sequence is first segmented into blocks of n RVs each; then each n-tuple is quantized as a unit.
- Mean-squared distortion: $E[(U-V)^2]$

Quantization

M-level Quantizer / Size of Q.A.

$R(\bar{L}) \geq H(V)$



$\{u_k\}$ iid continuous R.V.

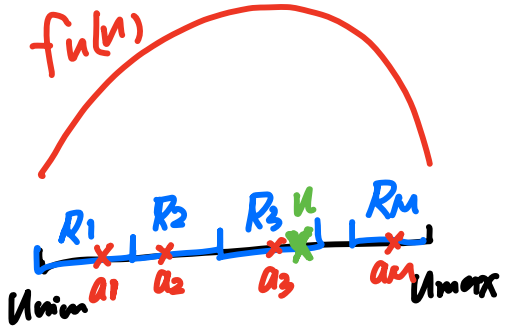
$U \sim f_U(u)$ (PDF)

given $u \sim f_U(u)$, $\boxed{M, \{R_j\}_{j=1}^{M}, \{a_j\}_{j=1}^{M}}$

Quantization: $U \longrightarrow V \in \{a_1, \cdots a_M\}$

$f_U(u)$ PMF Quantization Alphabet

if $u \in R_j$, $V(u) = a_j$, $\forall j$



PMF of V:

$P(V=a_1) = \int_{R_1} f_U(u)du = P(U \in R_1)$

$P(V=a_2) = \int_{R_2} f_U(u)du = P(U \in R_2)$

$\vdots$

$P(V=a_M) = \int_{R_M} f_U(u)du = P(U \in R_M)$

① entropy of V: $H(V)$

$(H(V) \leq R(\bar{L}))$

② Quantization Error

MSE (Mean-Squared Error/ Distortion)

$D = E_U[|U-V|^2]$

$\propto \int |u(t) - \tilde{v}(t)|^2 dt$

MSE: given $f_U(u)$, Quantizer

Method 1: $MSE = E_U\left[|U - V(U)|^2\right] = \int |u - V(u)|^2 f_U(u)\,du$

$$= \int_{R_1} |u-a_1|^2 f_U(u)\,du + \int_{R_2} |u-a_2|^2 f_U(u)\,du + \cdots \int_{R_M} |u-a_M|^2 f_U(u)\,du$$

$$= \sum_{j=1}^{M} \int_{R_j} |u-a_j|^2 f_U(u)\,du$$

Method 2: $MSE = E_V\left[E_{U|V}\left[|U - V(u)|^2 \,|\, V\right]\right]$

$$= \sum_{j=1}^{M} P(V=a_j)\, E_{U|V=a_j}\left[|U-V|^2 \,|\, V=a_j\right]$$

$$= \sum_{j=1}^{M} P(V=a_j) \int |u-a_j|^2 f_{U|V=a_j}(u)\,du$$

$P(V=a_j) \triangleq Q_j$

$\int |u-a_j|^2 f_{U|V=a_j}(u)\,du \triangleq MSE_j$

$f_{U|U\in R_j}(u) \triangleq f_j(u)$

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(A)}, & \text{if } A \text{ is true} \\ 0, & \text{otherwise} \end{cases}$$

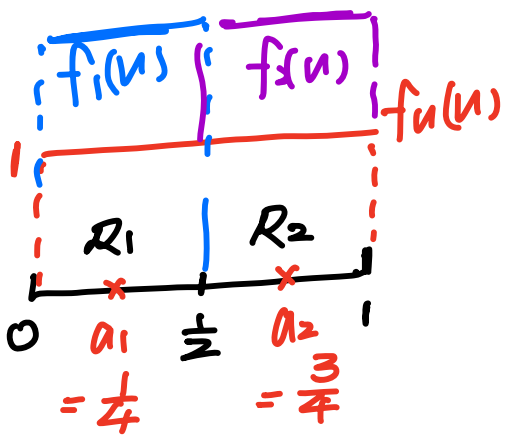
$$f_j(u) = \begin{cases} \frac{f_U(u)}{Q_j}, & \text{if } u \in R_j \\ 0, & \text{otherwise} \end{cases}$$

$$= \sum_{j=1}^{M} Q_j \int_{R_j} |u-a_j|^2 \frac{f_U(u)}{Q_j}\,du$$

$$= \sum_{j=1}^{M} \int_{R_j} |u-a_j|^2 f_U(u)\,du$$

Example

1. $M=2$ level Quantizer. $u \sim U[0,1]$



① $MSE = \int_0^{\frac{1}{2}} (u-\frac{1}{4})^2 du + \int_{\frac{1}{2}}^{1} (u-\frac{3}{4})^2 du$

$= \int_{-\frac{1}{4}}^{\frac{1}{4}} |u'|^2 du' + \int_{-\frac{1}{4}}^{\frac{1}{4}} |u'|^2 du' = \frac{1}{48}$

② $f_1(u) = \begin{cases} \frac{1}{\frac{1}{2}} = 2, & \text{if } u \in R_1 \\ 0, & \text{otherwise} \end{cases}$ $f_2(u) = \begin{cases} 2, & u \in R_2 \\ 0, & \text{otherwise} \end{cases}$

$MSE = \underbrace{\frac{1}{2}}_{Q_1} \underbrace{\int_0^{\frac{1}{2}} |u-\frac{1}{4}|^2 \times 2 \, du}_{MSE_1} + \underbrace{\frac{1}{2}}_{Q_2} \underbrace{\int_{\frac{1}{2}}^{1} |u-\frac{3}{4}|^2 \times 2 \, du}_{MSE_2}$

$= \frac{1}{48}$

2 uniform Quantizer

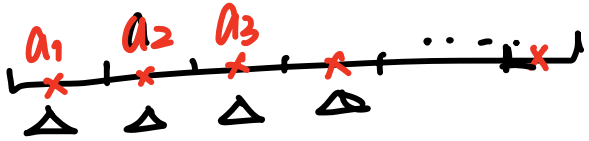
$f_u(u) \sim$ uniform, Δ, {middlepoint}

$f_j(u) = \begin{cases} \frac{f_u(u)}{f_u(u)\Delta} = \frac{1}{\Delta}, & \text{if } u \in R_j \\ 0, & \text{otherwise} \end{cases}$

$MSE = \sum_j Q_j \int_{a_j-\frac{\Delta}{2}}^{a_j+\frac{\Delta}{2}} |u-a_j|^2 \frac{1}{\Delta} \, du$

$= \sum_j Q_j \underbrace{\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} |u'|^2 \frac{1}{\Delta} \, du'}_{MSE_j}$

$= \frac{\Delta^2}{12}$



PMF: $P(V=a_j) = \frac{1}{M}, \forall j.$

$H(V) = \log_2 M.$

$M \uparrow \begin{cases} H(V) = \log_2 M \ (H(V) \le R_s) \uparrow & R \uparrow \\ MSE = \frac{\Delta^2}{12} & \Delta \downarrow \ MSE \downarrow \end{cases}$ Trade off

Quantization , $f_u(u)$, M

Scalar

Vector

① $\min \quad MSE = E[|u-v|^2]$

$\{R_j\}, \{a_j\}$ $\Rightarrow$ Lloyd-Max Algorithm

s.t. M

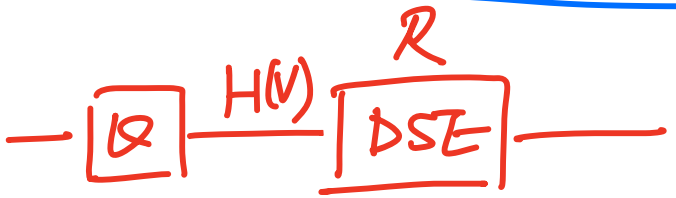
Scalar

Vector

② $\min \quad MSE = E[|u-v|^2] \quad \{a_j\}$

s.t. $H(V) = R \quad \{R_j\}$

$\Rightarrow$ Entropy Code Quantization

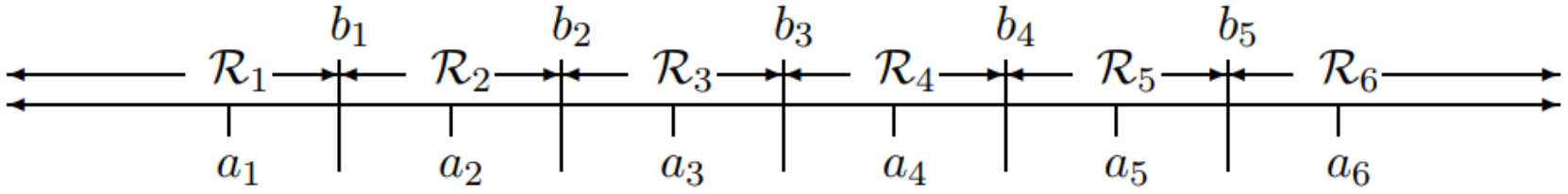


$H(V) \leq R$

$\eta = \frac{H(V)}{R}$

Scalar Quantization

- M level scalar quantization:
  - Partition the region R into M subsets $\mathcal{R}_1, \cdots, \mathcal{R}_M$, called quantization region.
  - Each region $\mathcal{R}_j$ is represented by a representation point $a_j \in R$.
- Scalar Quantization $\{v(u): R \to R\}$: $v(u) = a_j, \text{for } u \in \mathcal{R}_j$.



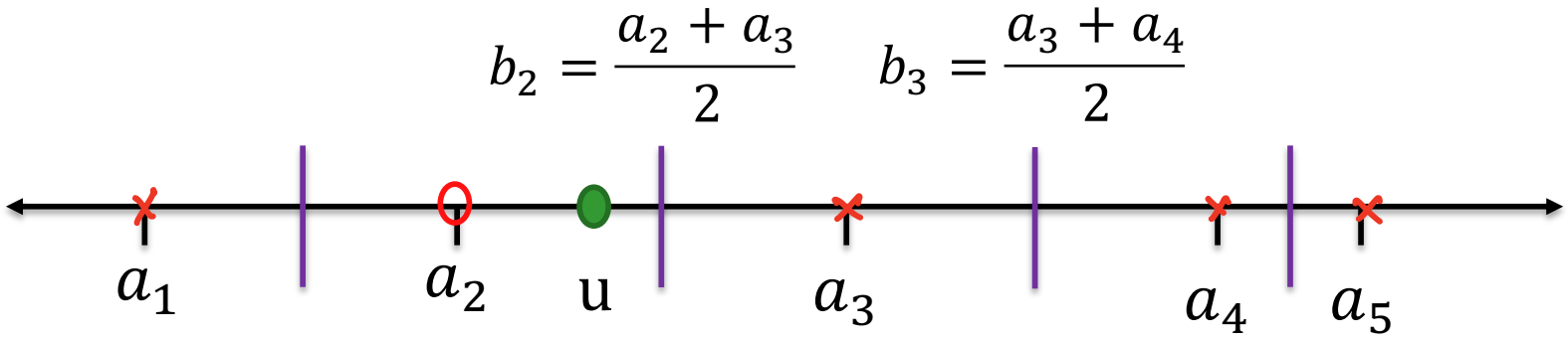
- Quantized output: $V_k = v(U_k)$ is a discrete rv with alphabet $\{a_1, \cdots, a_M\}$.

Scalar Quantization

- Q: how to choose the regions and representation points?
- Mean square distortion: $\boxed{\text{MSE} = E[(U-V)^2]}$
- Aim: Given pdf $f_U(u)$ and alphabet size M, choose $\{\mathcal{R}_j, 1 \le j \le M\}$ and $\{a_j, 1 \le j \le M\}$ to minimize MSE.
- Explore it into two ways:
  - Given a set of $\{a_j, 1 \le j \le M\}$, how should the intervals $\{\mathcal{R}_j, 1 \le j \le M\}$ be chosen?
  - Given a set of intervals $\{\mathcal{R}_j, 1 \le j \le M\}$, how to choose $\{a_j, 1 \le j \le M\}$?

Scalar Quantization

- Choice of $\{\mathcal{R}_j\}$ given $\{a_j\}$



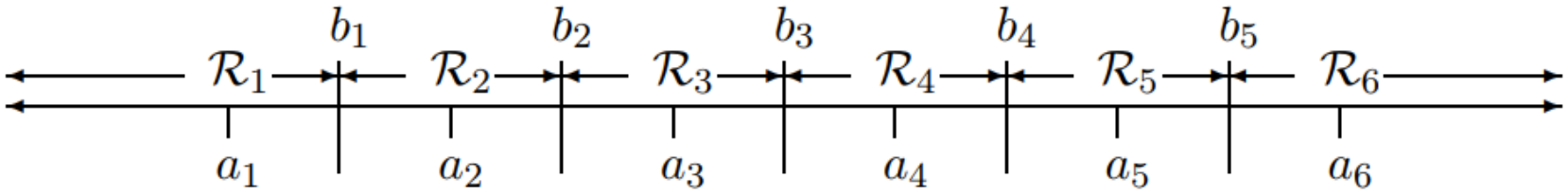
$$\min_j \left(u - a_j\right)^2$$

The boundary of b_j between $\mathcal{R}_j$ and $\mathcal{R}_{j+1}$ must lie halfway between a_j and a_{j+1}.

$$b_j = \frac{a_j + a_{j+1}}{2}$$

Scalar Quantization

- Choice of $\{a_j\}$ given $\{\mathcal{R}_j\}$



Method 1

$$\text{MSE} = E[(U-V)^2] = \int_{-\infty}^{\infty} f(u)(u-v)^2 du = \sum_{j=1}^{M} \int_{\mathcal{R}_j} f(u)(u-a_j)^2 du$$

$$= \sum_{j=1}^{M} Q_j \int_{\mathcal{R}_j} f_j(u)(u-a_j)^2 du = \sum_{j=1}^{M} Q_j \int_{-\infty}^{\infty} f_j(u)(u-a_j)^2 du$$

Method 2

$\int_{-\infty}^{\infty} f_j(u)(u-a_j)^2 du$: MSE$_j$

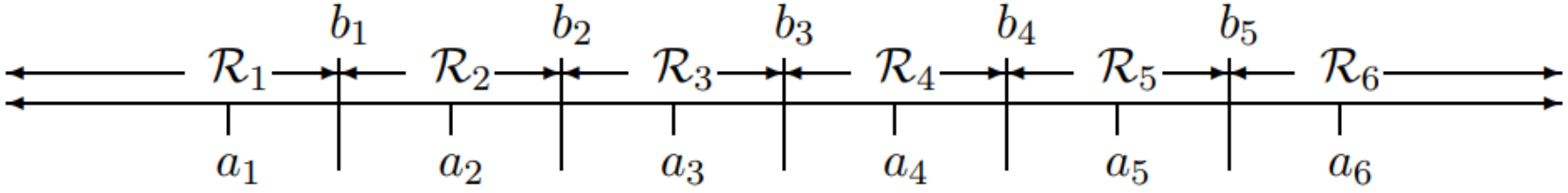
where $Q_j = \Pr(U \in \mathcal{R}_j) = \int_{b_{i-1}}^{b_j} f(u)du$ and

the conditional pdf of U given that $\{u \in \mathcal{R}_j\}$

$$f_j(u) = \begin{cases} f(u)/Q_j, & \text{if } u \in \mathcal{R}_j \\ 0, & \text{otherwise} \end{cases}$$

Scalar Quantization

- Choice of $\{a_j\}$ given $\{\mathcal{R}_j\}$



$$\text{MSE} = \sum_{j=1}^{M} Q_j \int_{-\infty}^{\infty} f_j(u)(u - a_j)^2 du$$

$U_{(j)} \sim f_j(u)$

Let $U_{(j)}$ be a new RV with pdf $f_j(u)$, then

$$\text{MSE} = \sum_{j=1}^{M} Q_j E[(U_{(j)} - a_j)^2]$$

MSE$_j$

$\min_{\{a_j\}} \sum_{j=1}^{M} Q_j E|U_{(j)} - a_j|^2$

$\Updownarrow$

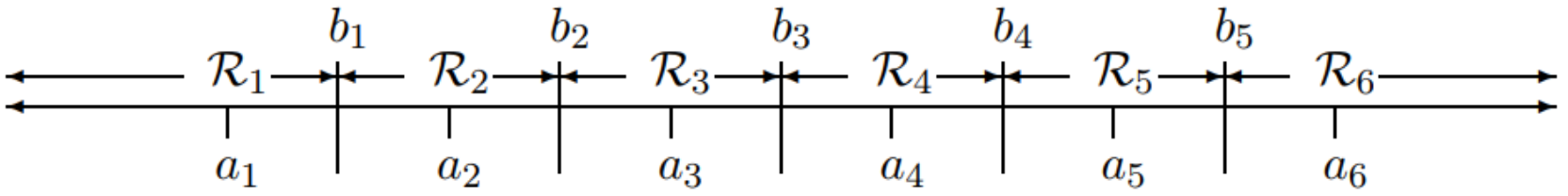
$\min_{a_j} E|U_{(j)} - a_j|^2$

$$\min_{a_j} E[(U_{(j)} - a_j)^2] = E[(U_{(j)}^2 - (E[U_{(j)}])^2 + (E[U_{(j)}])^2 - 2a_j U_{(j)} + a_j^2]$$
$$= \text{Var}(U_{(j)}) + (E[U_{(j)}] - a_j)^2$$

Thus,

$$a_j = E[U_{(j)}] = \int_{R_j} u f_j(u) du = \frac{\int_{R_j} u f(u) du}{\int_{R_j} f(u) du}$$

Lloyd-Max Algorithm



An optimal scalar quantizer must satisfy both $b_j = (a_j + a_{j+1})/2$ and $a_j = E[U_{(j)}]$.

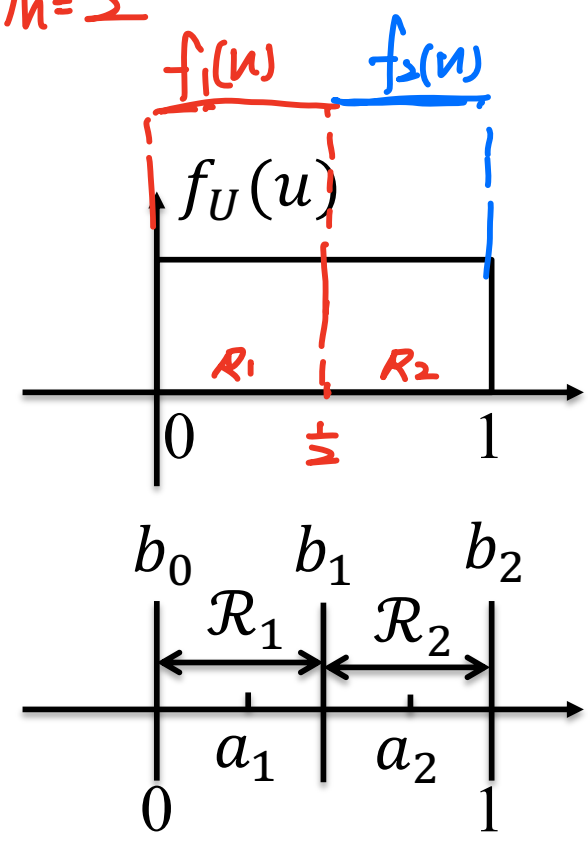
MSE

1. Choose $a_1 < a_2 < \cdots < a_M$
2. set $b_j = (a_j + a_{j+1})/2$ for $1 \le j \le M-1$
3. Set $a_j = E[U_{(j)}] = \frac{\int_{\mathcal{R}_j} u f(u) du}{\int_{\mathcal{R}_j} f(u) du}$ where $\mathcal{R}_j = (b_{j-1}, b_j]$ for $1 \le j \le M-1$
4. Iterate on 2 and 3 until improvement is negligible.

It find local min, not necessarily global min.

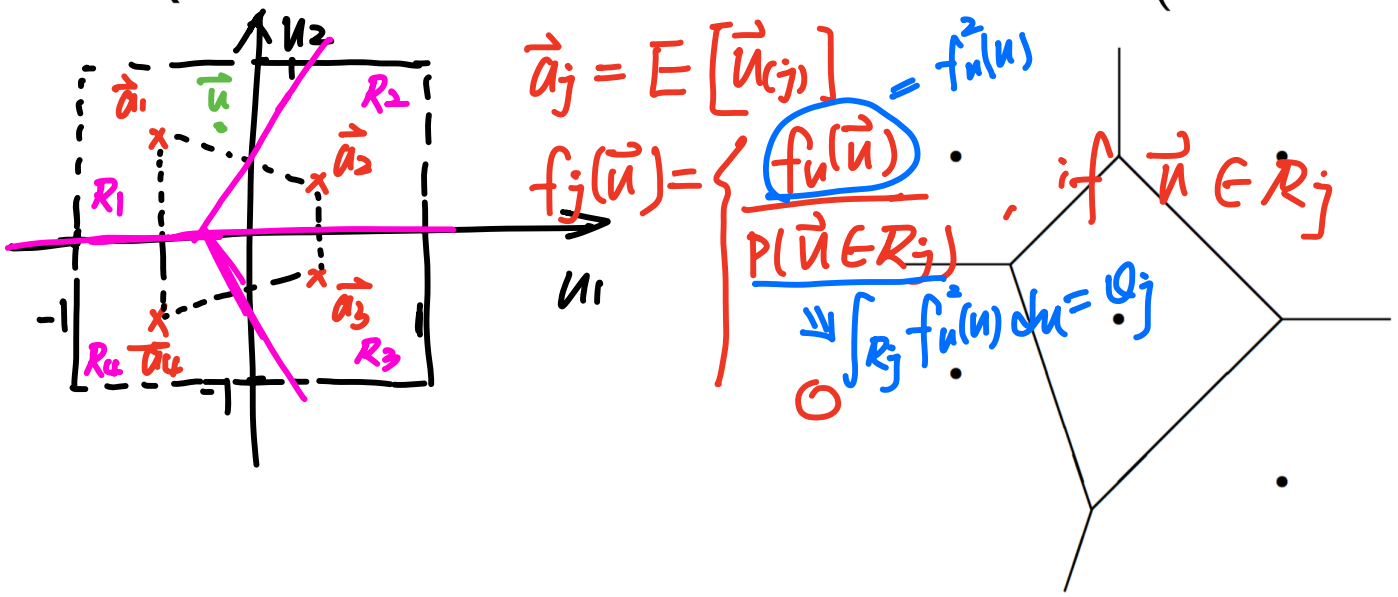
Scalar Quantization

- Example: Analog source U_k is uniformly distributed between 0 and 1. Find the 1 bit quantizer that minimizes the MSE. Find the MSE.
- Sol:
- M=2, b1=1/2.
- $f_j(u) = \dfrac{f_U(u)}{\int_{R_j} f_U(u)du_u} = \begin{cases} \frac{1}{\Delta} = 2, 0 \le u \le \frac{1}{2}, j = 1 \\ \frac{1}{\Delta} = 2, \frac{1}{2} \le u \le 1, j = 2 \end{cases}$
- a1=1/4,a2=3/4.
- MSE= $E(U-V)^2 = \int_0^{\frac{1}{2}} (u-a_1)^2 f_U(u)du + \int_{\frac{1}{2}}^{1} (u-a_2)^2 f_U(u)du = \dfrac{\Delta^2}{12} = \dfrac{1}{48}$.



Vector Quantization

Quantize n source variables at a time. (In scalar quantization, $n = 1$)



Given $\{(a_j, a'_j)\}$, how to choose $\{\mathcal{R}_j\}$

- The square error is $(u - a_j)^2 + (u' - a'_j)^2$, the point $\{a_j, a'_j\}$ which is the closest to (u, u') in Euclidean distance should be chosen.
- $\{\mathcal{R}_j\}$ contains points that are closer to (a_j, a'_j) than any other representation points, i.e., Voronoi regions.

Given a set of Voronoi region, how to find the $\{a_j, a'_j\}$?.

- Choose $\{a_j, a'_j\}$ to be the conditional means within those regions.

Vector Quantization

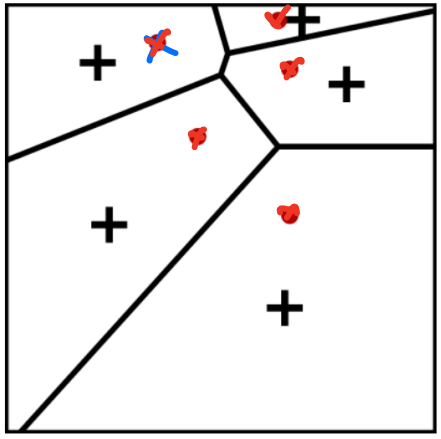


Figure: Iteration 1

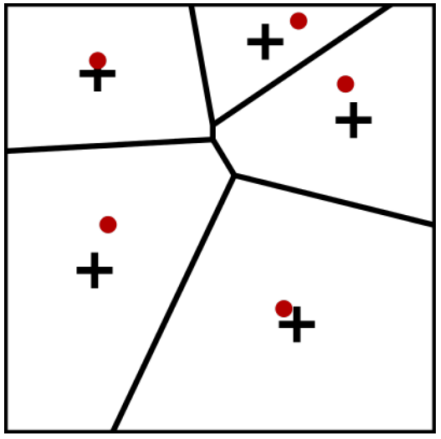


Figure: Iteration 2

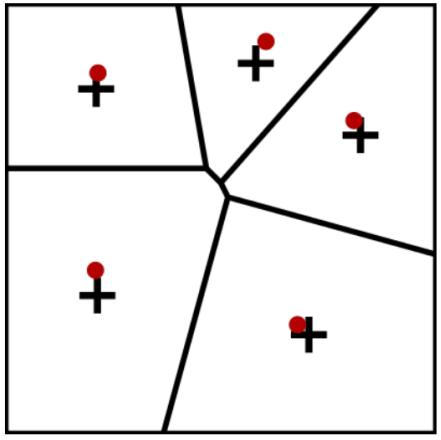


Figure: Iteration 3

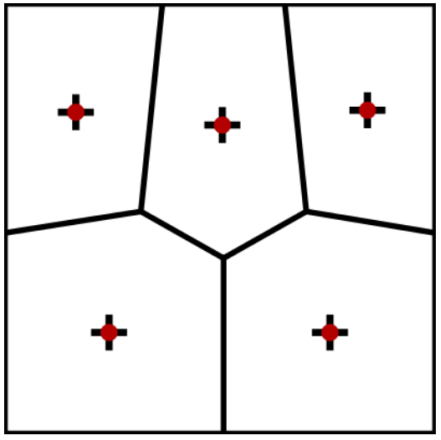
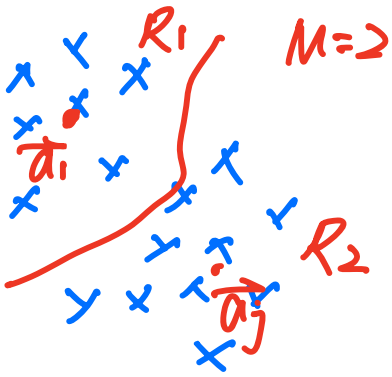


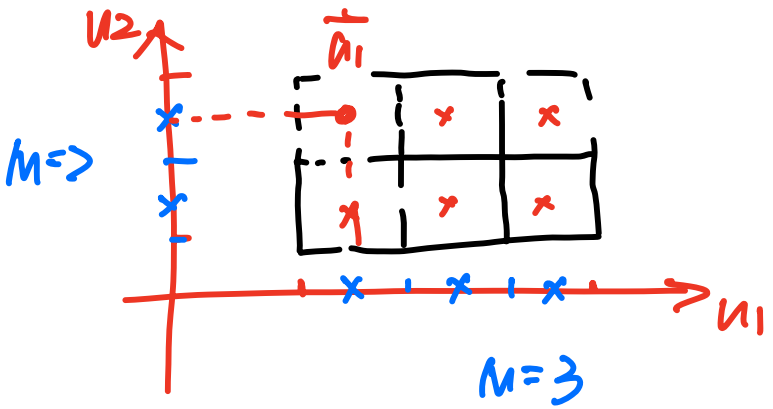
Figure: Iteration 4



Vector Quantization

- Popular research topic, related to deep learning algorithm [(1) clustering (2) Neural Discrete Representation Learning (VAE / VQ-VAE)
- Quantizing complexity goes up exponentially with n
- For vector quantization, the problem of local minimum is even worse.
- Reduction in MSE with increasing n is quite modest
- Samples are highly dependent
- Application: Video, Audio

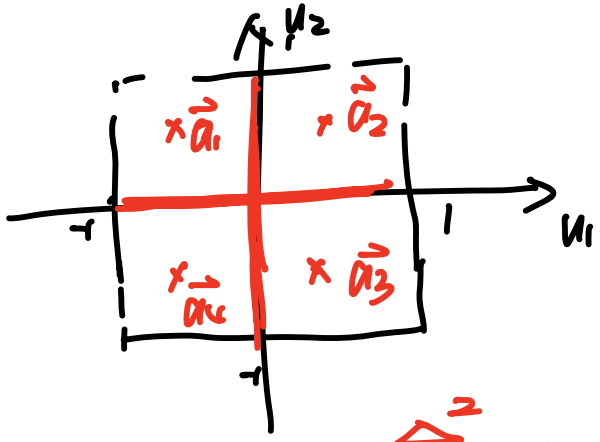
$u_1, u_2 \cdots u_K$ $\quad n=2 \quad M=6$



rectangular Quantizer

$u_1\ u_2 \cdots \quad u_K \quad n=2 \quad M=4$

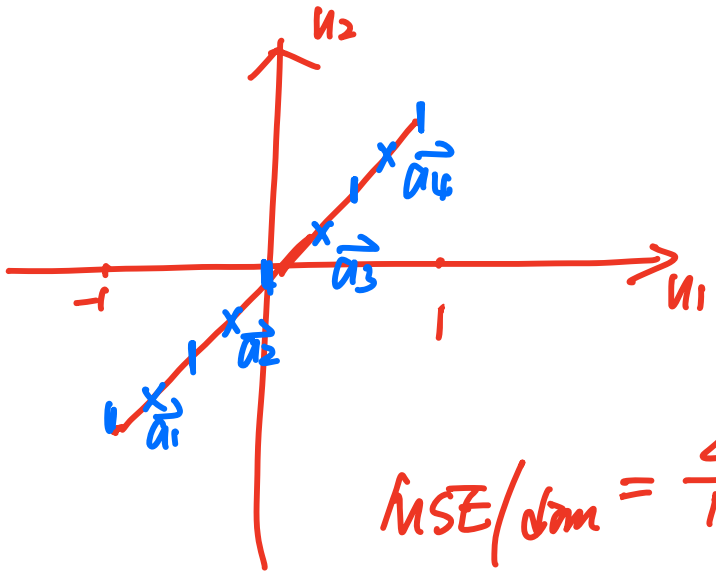
① [iid. $u_k \sim U[-1,1]$



$MSE = \frac{\Delta^2}{12} \quad (\Delta = 1)$

② $u_2 = u_1, \ u_3 = u_4 \cdots$

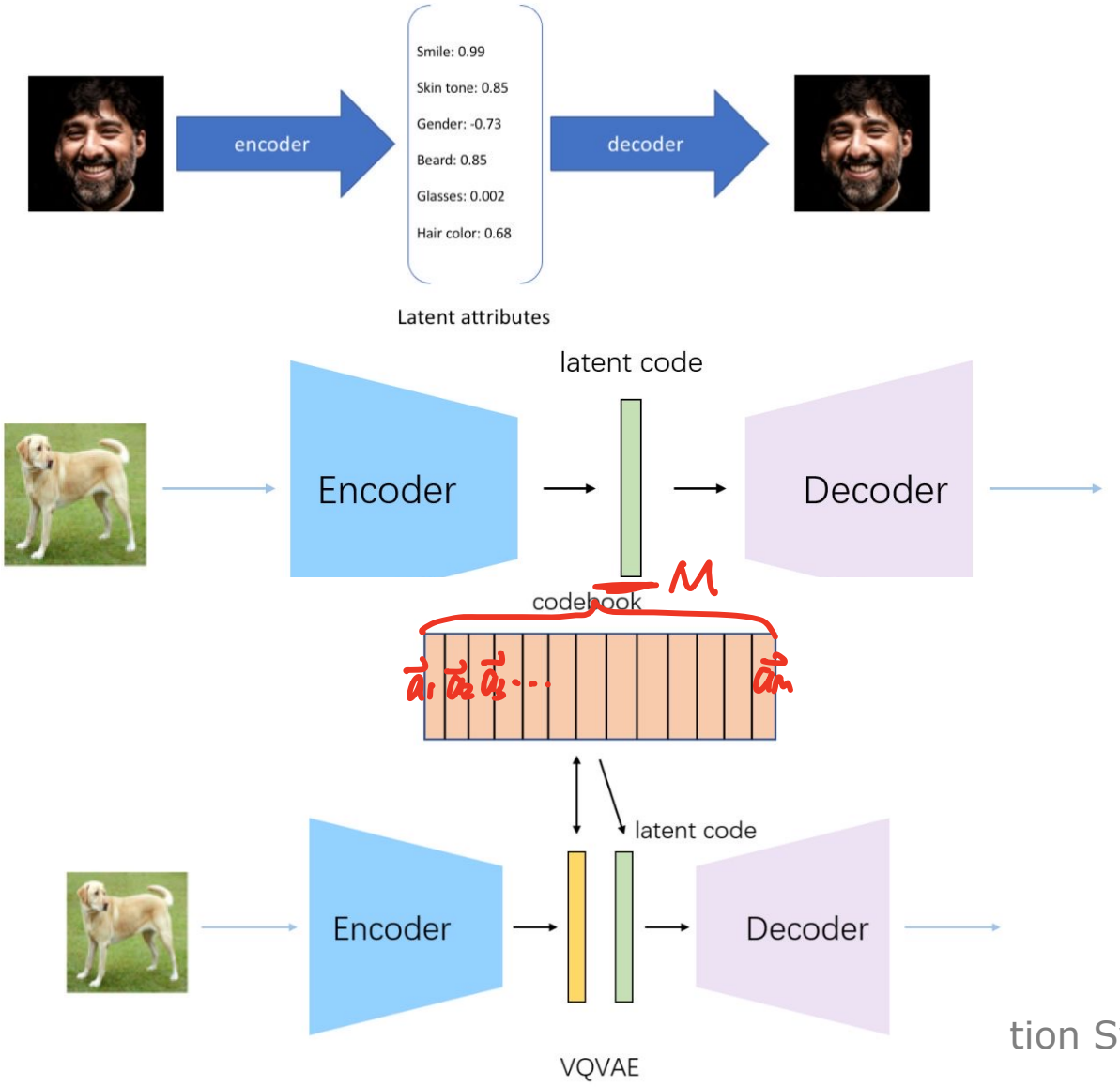
$u_1 \sim U[-1,1]$



$MSE/dim = \frac{\Delta^2}{12} \quad (\Delta = \frac{1}{2})$

Vector Quantization

- Image Reconstruction / Data Transmission

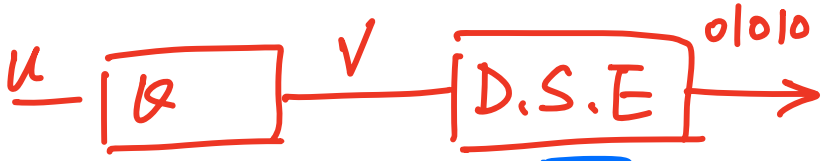


Entropy-Code Quantization

Quantization:

- Source sequence (samples): $(U_1, U_2, \ldots)$
- Quantizer's Output: $(V_1, V_2, \ldots)$, with $V_k \in \{a_1, a_2, \ldots, a_M\}$.
- If $U_k \in \mathcal{R}_j$, then $V_k = a_j$

(PMF)



$I_{min} = R \geq H(V)$ (entropy bound)

Discrete Source Encoding:

$P(V_k = a_j) = Q_j, \forall j$

$$a_1 \longrightarrow 0$$
$$a_2 \longrightarrow 10$$
$$a_3 \longrightarrow 11$$

$\bar{L}_{\min}$ minimum average code length: $\bar{L}_{\min} \approx H(V)$

We should minimize MSE for given entropy $H(V)$ rather than a given number of representation points.

Entropy-Code Quantization

- **Problem:** For a signal u with given pdf $f_U(u)$, find a quantizer such that

what is Entropy-Code Quantization

① given H(V), minimize MSE

② Q → V → D.S.E/Entropy coding

$$\begin{aligned} &\text{minimize} && E[(U-V)^2] \\ &\text{subject to} && R = H(V) \end{aligned}$$

{Rj} {aj} {Rj}

where $H(V) = -\sum_{i=1}^{M} p(a_i)\log(p(a_i))$ with $p(a_i) = \int_{\mathcal{R}_j} f(u)du$

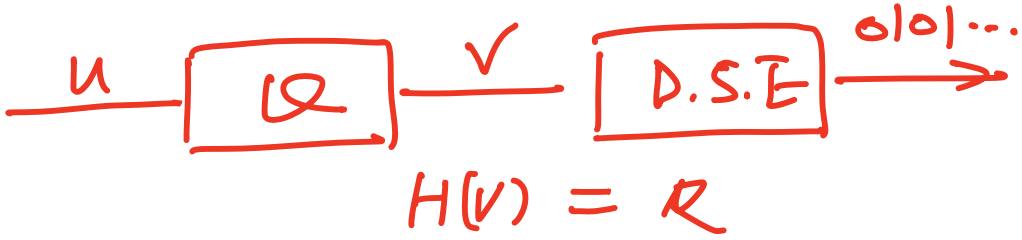
Trim: $\mathcal{R} = H(V)$

- **Solution:** Lagrangian cost function [Skip, See P. 85 in Gallager'Book]

$$L = \quad E[(U-V)^2] + \lambda H(V)$$

- **Compare with Lloyd-Max:** $H(V)$ depends on $p(a_i)$, not $\{a_j\}$
  - For given $\{\mathcal{R}_j\}$, $a_j = E[U_{(j)}]$ minimizes MSE without changing $H(V)$.
  - For given $\{a_j\}$, b_j minizes MSE, but may change $H(V)$

Assumption : High-rate Quantizer



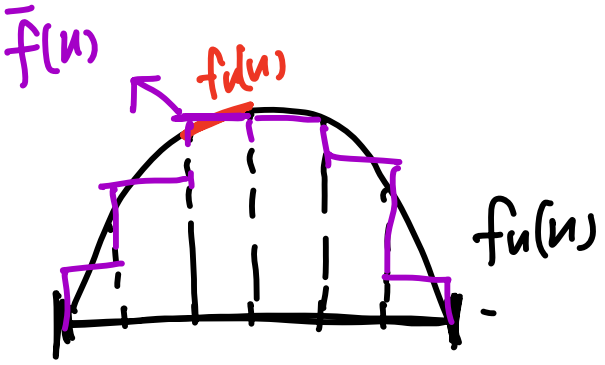
$H(V) = R$

① R is Large

② $H(V)$ is Large

③ M is Large

④ MSE is small



⑤* Quantization Region is Small →

$$f_U(u) \approx \bar{f}(u) \Leftarrow \boxed{\int_{R_j} f_U(u)\,du = \bar{f}(u)\,|R_j|}$$

constant

give $h(U)$, coding rate R.

opt. High-rate, Entropy-coded Quantizer

{ Uniform: Δ?

$H(V) = h(U) - \log_2 \Delta \le R$

$\Delta \ge 2^{h(U) - R}$

$M = \frac{}{}$ a_j = Midpoint

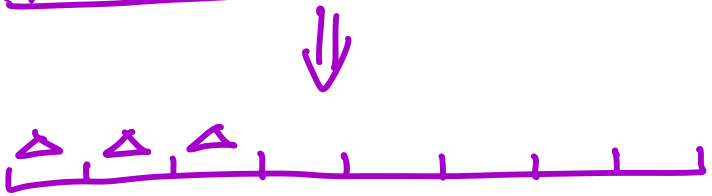
$MSE^* \triangleq \frac{1}{12}\Delta^2$

P&S.

High-rate quantizer (Large $H(V)$)

$\Rightarrow$ uniform scalar quantizer is a nearly optimal entropy-coded quantizer

$\Rightarrow$ **uniform scalar quantizer** approx. minimizes MSE, s.t. entropy constraint



$\Rightarrow$ (optimal high-rate entropy-coded quantizer)

min MSE

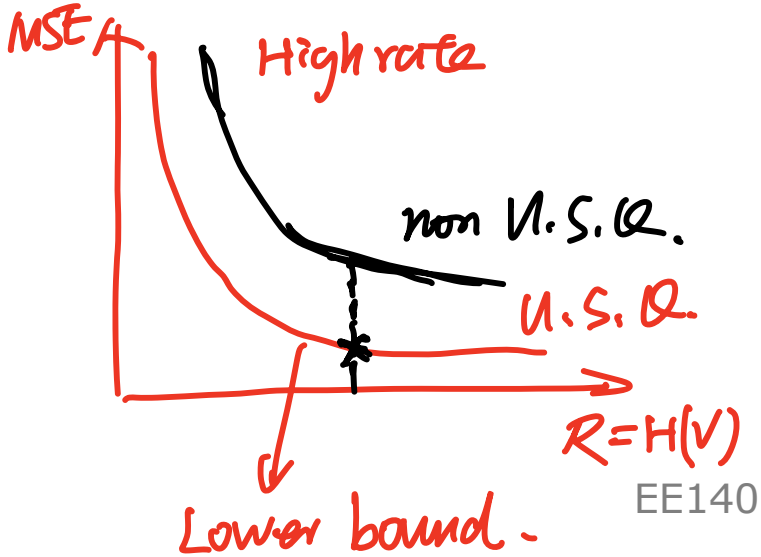
$H(V)$ is Large

$H(V)=R$

min MSE

s.t. $H(V)=R$

Highrate : $H(V)$ is Large



① given u, Δ, { MSE =
H(V) =

② { trade off perf. { MSE
H(V)
given u, R, $\Rightarrow$ opt. E.C.Q. { Δ = ? MSE = ?
{ a_{ij} = ?

Uniform Scalar Quantizer

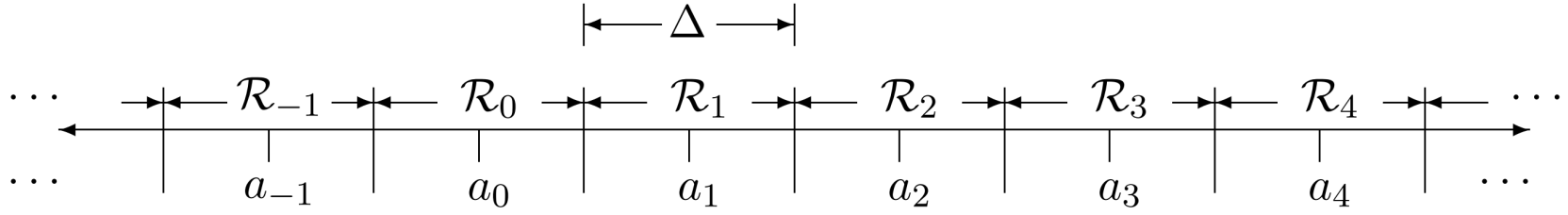
- Uniform scalar quantizer:
  - All quantization intervals have the same lengths.
- High-rate quantizer:
  - Quantization regions can be sufficiently small, pdf $f_U(u)$ is almost constant within each region.
  - Target entropy is large (high rate).
- Uniform scalar quantizer is nearly optimal in terms of MSE within the class of scalar quantizer when using
  - High-rate quantizers ($H(V)$ is large)
  - Entropy-coded quantization ($H(V) = R$)
  - Proof: See P. 85 in Gallager'Book

High-Rate Entropy-Coded Quantization

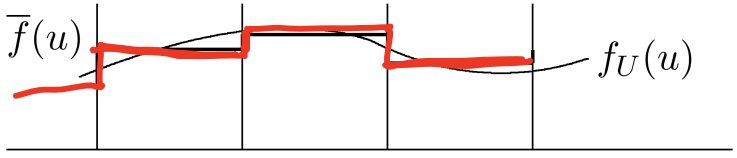
- Analyze the performance of Uniform scalar quantizer in the limit of high rate.

MSE=
H(V)

- Assume each quantization interval $\mathcal{R}_j$ has same length Δ.



- Assume Δ is small enough that $f_U(u)$ is appro. constant within each $\mathcal{R}_j$.



$$\overline{f}(u) = \frac{\int_{\mathcal{R}_j} f_U(u)du}{\Delta} \quad \text{for} \quad u \in \mathcal{R}_j$$

$$\Delta\overline{f}(u) = \Pr(\mathcal{R}_j) \text{ for all integer } j \text{ and all } u \in \mathcal{R}_j$$

- High rate assumption: $f_U(u) \approx \bar{f}(u) = \frac{\Pr(\mathcal{R}_j)}{\Delta}$ for all $u \in \mathcal{R}_j$.

High-Rate Entropy-Coded Quantization

- High rate assumption: $f_U(u) \approx \bar{f}(u)$ for all $u \in R$.
  - Conditional pdf:

$$f_{U|\mathcal{R}_j}(u) \approx \begin{cases} 1/\Delta, & u \in \mathcal{R}_j; \\ 0, & u \notin \mathcal{R}_j. \end{cases}$$

$$f_j(u) = \frac{f_U(u)}{\int_{R_j} f_U(u)\,du} \approx \frac{\bar{f}(u)}{\bar{f}(u)\cdot\Delta}$$

  - Conditional mean a_j: center of interval $\mathcal{R}_j$.

- MSE:

$$\mathrm{MSE}(\mathcal{R}_j) \approx \sum_{j=1}^{M} \Pr(\mathcal{R}_j)\frac{\Delta^2}{12} = \frac{\Delta^2}{12}$$

$$MSE_j = \int_{a_j-\frac{\Delta}{2}}^{a_j+\frac{\Delta}{2}} |u-a_j|^2 \frac{1}{\Delta}\,du = \frac{\Delta^2}{12}$$

  - If the decoding is successful, the MSE in reconstructing the original sequence is $\mathrm{MSE} \approx \frac{\Delta^2}{12}$.

- Entropy of V: $p(a_j) = \int_{R_j} f(u)du = \bar{f}(u)\Delta$

$$H(V) = \sum_j -p(a_j)\log p(a_j) = \sum_j \int_{\mathcal{R}_j} -f_U(u)du\log[\bar{f}(u)\Delta]$$

$$= \int_{-\infty}^{\infty} -f_U(u)\log[\bar{f}(u)\Delta]du$$

Using high-rate approximation $f_U(u) \approx \bar{f}(u)$

$$= \int_{-\infty}^{\infty} -f_U(u)\log[\bar{f}(u)]du - \log\Delta \rightarrow$$

High-Rate Entropy-Coded Quantization

- High rate assumption: $f_U(u) \approx \bar{f}(u)$ for all $u \in R$.
  - Conditional pdf:

$$f_{U|\mathcal{R}_j}(u) \approx \begin{cases} 1/\Delta, & u \in \mathcal{R}_j; \\ 0, & u \notin \mathcal{R}_j. \end{cases}$$

  - Conditional mean a_j: center of interval $\mathcal{R}_j$.
- MSE:

$$\text{MSE}(\mathcal{R}_j) \approx \sum_{j=1}^{M} \Pr(\mathcal{R}_j) \frac{\Delta^2}{12} = \frac{\Delta^2}{12}$$

  - If the decoding is successful, the MSE in reconstructing the original sequence is $\text{MSE} \approx \frac{\Delta^2}{12}$.
- Entropy of V:

$$\begin{aligned} H(V) &\approx \int_{-\infty}^{\infty} -f_U(u) \log[f_U(u)] du - \log \Delta \\ &= h(U) - \log \Delta \end{aligned}$$

$$P(V=a_j) = \int_{R_j} f_u(u)\,du = \bar{f}(u)\Delta$$

$$H(V) = \sum_j -P(V=a_j)\log P(V=a_j)$$

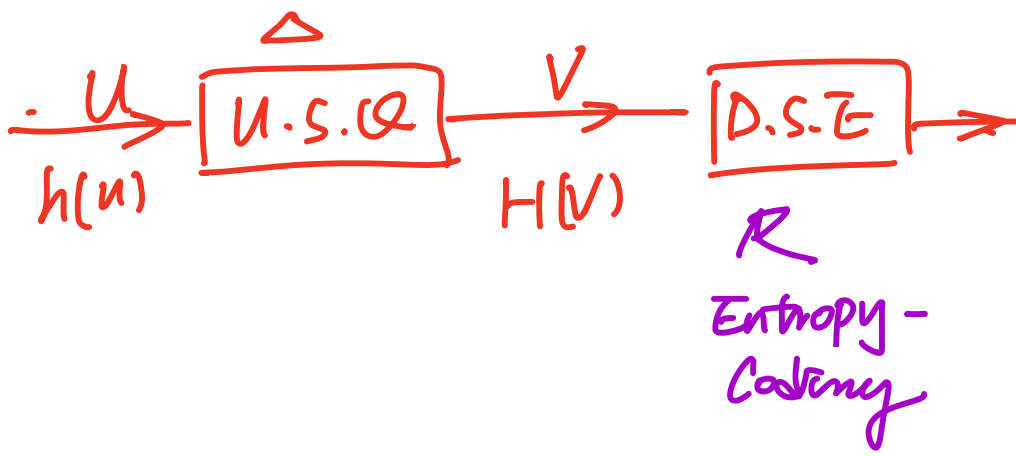
$$= \sum_j -\int_{R_j} f_u(u)\,du \log\left(\bar{f}(u)\Delta\right)$$

$$= -\int f_u(u)\log\left(\bar{f}(u)\Delta\right)\,du$$

High-rate

$$\approx -\int f_u(u)\log f_u(u)\,du - \log\Delta$$

$$= h(u) - \log\Delta$$



High rate

Average coding rate $\bar{R} = \bar{L}_{min} = H(V) \approx h(u) - \log\Delta$ bits/symbol

$\bar{L}\uparrow \rightarrow h(u)\uparrow \quad \Delta\downarrow$

$MSE\downarrow \approx \frac{\Delta^2}{12}\downarrow$

$\bar{L} = h(u) - \log_2 \Delta$ } Trade off.

$MSE = \frac{\Delta^2}{12}$

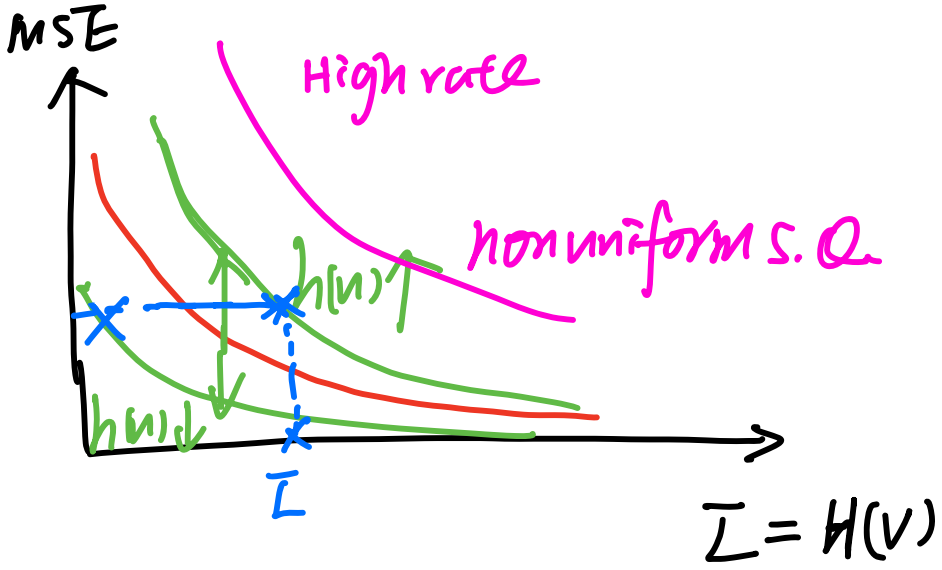
$\Rightarrow \log_2 \Delta = h(u) - \bar{L}$

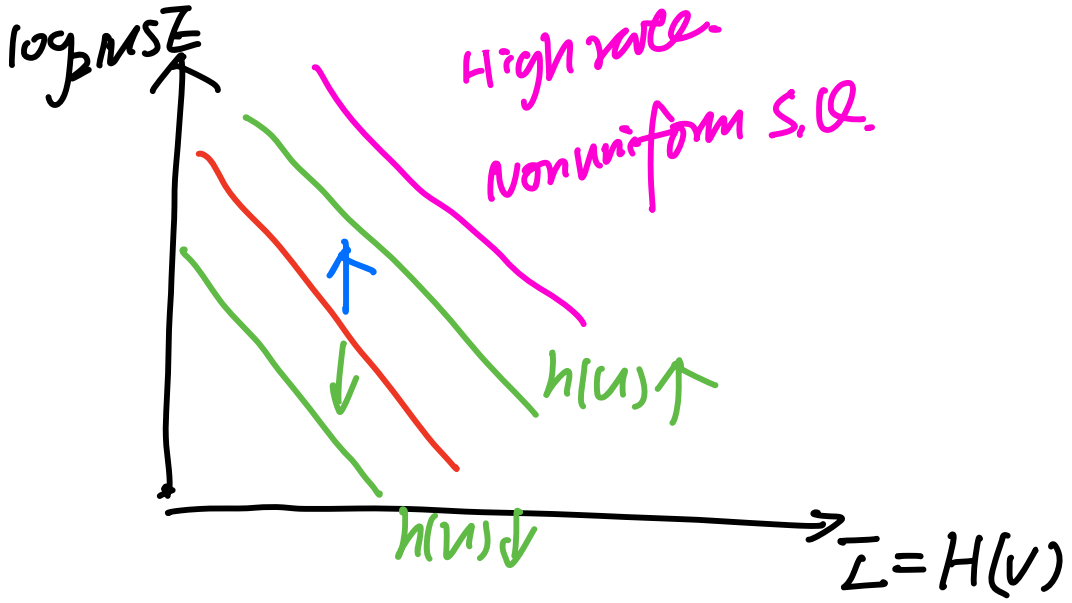
$\Delta = 2^{h(u)-\bar{L}}$

$\log_2 MSE = 2\log_2 \Delta - \log_2 12$

$= 2(h(u) - \bar{L}) - \log_2 12$

$MSE = \frac{2^{2(h(u)-\bar{L})}}{12}$





① $h(u)$: shift

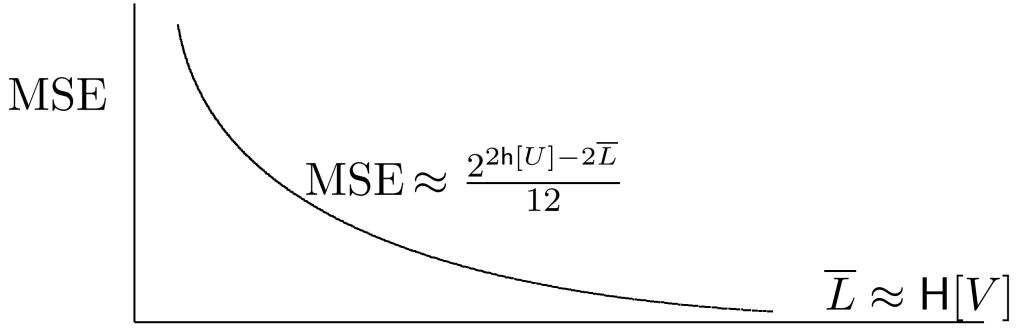
② Δ : $\Delta \downarrow \frac{1}{2}\Delta$ $\bar{L}+1$ bits/symbol $\frac{1}{4}MSE$ $\downarrow$ 6 dB

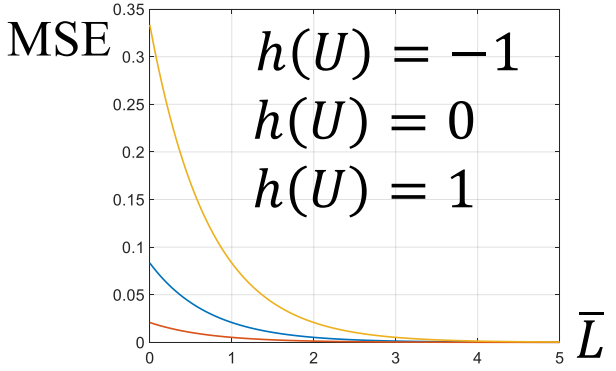
③ High rate same $\bar{L} = H(V)$: $MSE_{N.U.S.Q} > MSE_{U.S.Q}$.

High-Rate Entropy-Coded Quantization

- Average Code Length:
  - $\bar{L} \approx H(V) \approx h(U) - \log \Delta$ (bits/symbol)
  - $\bar{L}$ depends only on $h(U)$ and Δ, not on $f_U(u)$ or M.
- Trade off between coding rate and MSE

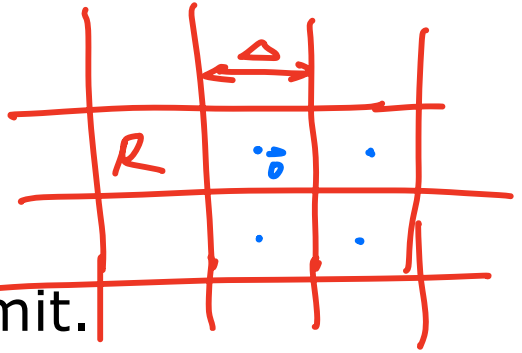
$$\overline{L} \approx \mathrm{h}(U) - \log \Delta; \qquad \mathrm{MSE} \approx \frac{\Delta^2}{12}$$





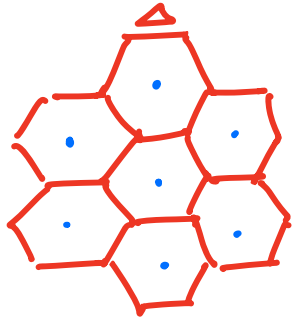
Reduction in Δ by a factor of 2, reduce MSE by a factor of 4, increase the $\bar{L}$ by 1bit/symbol. (Each additional bit per symbol decrease the MSE by 6 dB.

High-Rate 2D Quantizer



- Uniform 2D quantizer is optimal in the high-rate limit.
  - Proof: See P. 85 in Gallager'Book

$$f_j(\vec{u}) = \begin{cases} \frac{f(\vec{u})}{f(\vec{u})A(R)} & \text{if } u \in R_j \\ 0 \end{cases}$$



- Similar as 1D quantizer, we have

$$\sum_{j=1}^{M} P(V=\vec{a}_j)\,MSE_j$$

- MSE per dimension:

$$MSE_j = \int_R \|\vec{u}\|^2 \frac{1}{A(R)}\, d\vec{u}$$

$$\text{MSE} \approx \frac{1}{2}\int_{\mathcal{R}} \|\boldsymbol{u}\|^2 \frac{1}{A(\mathcal{R})}\, d\boldsymbol{u}$$

$A(\mathcal{R}) = \int_{\mathcal{R}} d\boldsymbol{u}$ is the area of basic cell (square, hexagon, etc.)

  - Square: $A(\mathcal{R}) = \Delta^2$, MSE=$\Delta^2/12$
  - Hexagon: $A(\mathcal{R}) = 3\sqrt{3}\Delta^2/2$, MSE=$5\Delta^2/24$

$\left(\frac{\text{MSE}}{A(\mathcal{R})}\right)_{\text{Sqaure}} / \left(\frac{\text{MSE}}{A(\mathcal{R})}\right)_{\text{Hex}} = 1.0392$. For a given $\bar{L}$, hexagonal quantizer is slightly better. Thus, uniform scalar quantizers (square quantizer) is good enough.

High-Rate 2D Quantizer

- Entropy of Uniform 2D quantizer

  With

$$p_j = \bar{f}(\boldsymbol{u})A(\mathcal{R}) = \int_{\mathcal{R}_j} f(\boldsymbol{u})d\boldsymbol{u}$$

  Entropy of $\boldsymbol{V}$:

$$\begin{aligned} H(\boldsymbol{V}) &= -\sum_j p_j \log p_j \\ &= -\sum_j \int_{\mathcal{R}_j} f(\boldsymbol{u}) \log[\bar{f}(\boldsymbol{u})A(\mathcal{R})]d\boldsymbol{u} \\ &= -\int f(\boldsymbol{u})[\log \bar{f}(\boldsymbol{u}) + \log A(\mathcal{R})]d\boldsymbol{u} \\ &\overset{(a)}{\approx} 2h(U) - \log A(\mathcal{R}) \end{aligned}$$

  where (a) follows from $\bar{f}(\boldsymbol{u}) \approx f(\boldsymbol{u})$ and $H(\boldsymbol{U}) = H(U_1U_2) = 2H(U)$.
  Thus, average rate per symbol:

$$\bar{L} \approx \frac{H(\boldsymbol{V})}{2} \approx h(U) - \frac{1}{2}\log A(\mathcal{R})$$

Trade off: $\bar{L}$ / MSE

$A(R)\downarrow$ $\bar{L}\uparrow$ MSE$\downarrow$

$A(R)\downarrow$ $\frac{1}{4}A(R)$

$\bar{L}\uparrow$ 1 bit/symbol

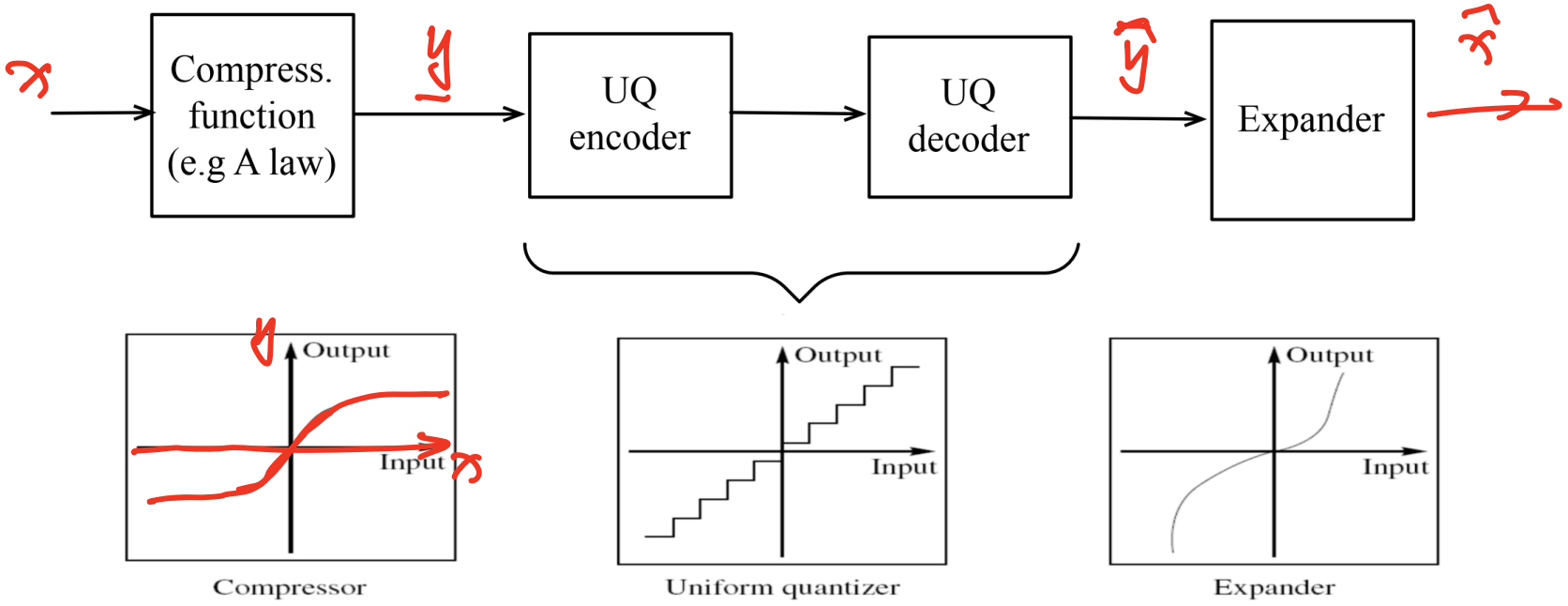
MSE$\downarrow$ 6 dB

give $\bar{L}$ (given $A(R)$)

MSE Hex < MSE square.

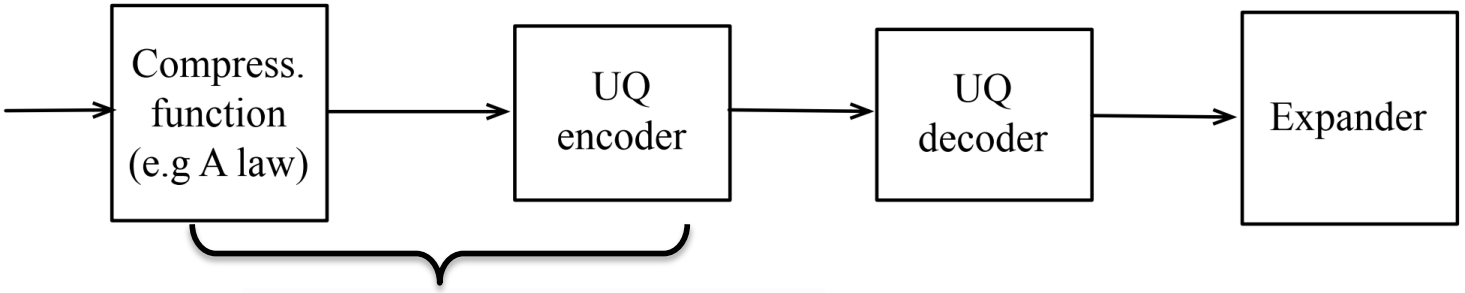
Quantization in Telephony Systems

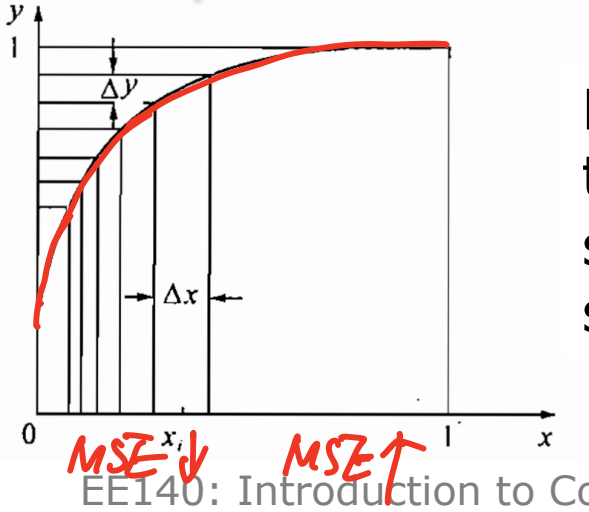
- Given input x, first use compressor by mapping $x \to y$, then use uniform quantizer
- To restore the signal, use a device with a characteristic complementary to the compressor, called expander.
- compander=compressor+expander



Quantization in Telephony Systems

- Given input x, first use compressor by mapping $x \to y$, then use uniform quantizer
- To restore the signal, use a device with a characteristic complementary to the compressor, called expander.
- compander=compressor+expander





Nonuniform quantizer: the quantization noise is small for smaller input signal

Quantization in Telephony Systems

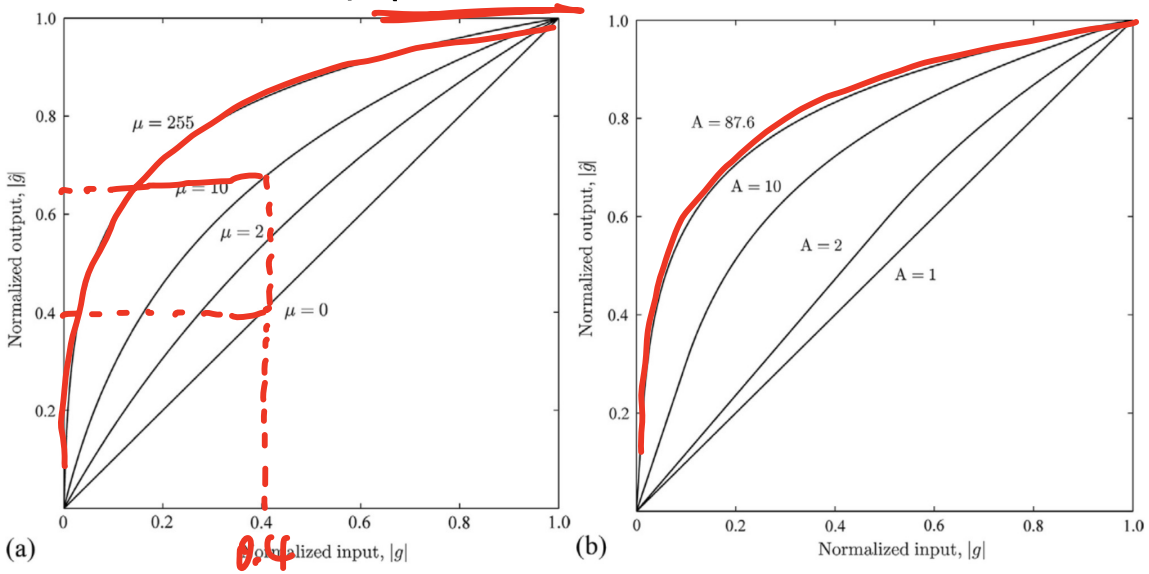
- μ law (Applied in North America and Japan): for $\mu > 0$

$$y = \frac{\log(1 + \mu|x|)}{\log(1 + \mu)}$$

- A law (Applied in elsewhere): for $A > 0$

$$y = \begin{cases} \frac{Ax}{1+\ln A}, & 0 < x \leq \frac{1}{A} \\ \frac{1+\ln Ax}{1+\ln A}, & \frac{1}{A} < x \leq 1 \end{cases}$$

A=87.6, μ=255.



Assume $M = 2^8 = 256$

If x=0.4

$M_{\mu=0} = 256 \times 0.4 \approx 102$

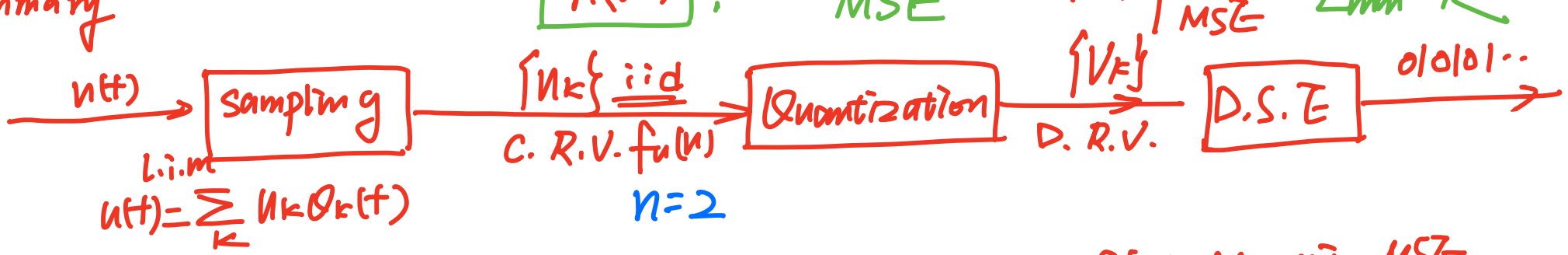
$M_{\mu=10} = 256 \times 0.65 \approx 166$

$M_{A=10} = 256 \times 0.71 \approx 181$

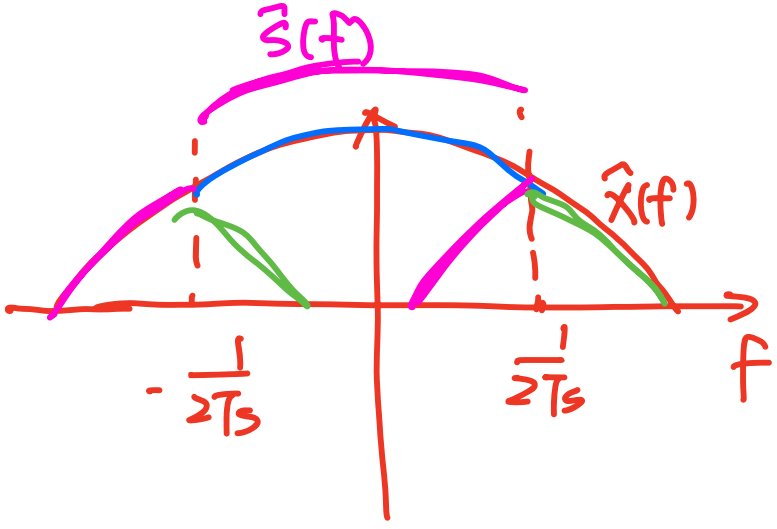
Summary

$h(t)$ Δ MSE

PMF $\begin{cases} H(V) \\ MSE \end{cases}$ $I_{min}=R$



- Time limited / unlimited (F.S.E)
- Band limited / unlimited (sampling Theory) ($T_s = \frac{1}{2W}$)
- Aliasing ($W > \frac{1}{2T_s}$, $T_s > \frac{1}{2W}$, $f_s < 2W$)



- $f_X(x)$, M, min MSE
  - ⇒ Lloyd-Max Alg ($\{b_j\}, \{a_j\} = E[X_j]$)
- $f_X(x)$, R, min MSE s.t $H(V) = R$
  - (High-rate)
  - ⇓
  - uniform-scalar quantizer
  - ⇓

Trade off $\begin{cases} MSE^* = \_\_\_\_\_ (h(x)) \\ H(V)^* = I_{min} = \_\_\_\_\_ \end{cases}$

Thanks for your kind attention!

Questions?