《多媒体系统导论》期末速通教程

5. 音频的压缩

5.1 音频

[**人的声音频率**]

(1) 听觉频率: $20 \sim 2\mathrm{e}4\ \mathrm{Hz}$.

(2) 发声频率: $85 \sim 1100\ \mathrm{Hz}$.

[**音频信号的滤波**] 对音频信号滤波, 以消除不需要的频率, 一般保留 $50 \sim 10\mathrm{k}\ \mathrm{Hz}$ 的信号.

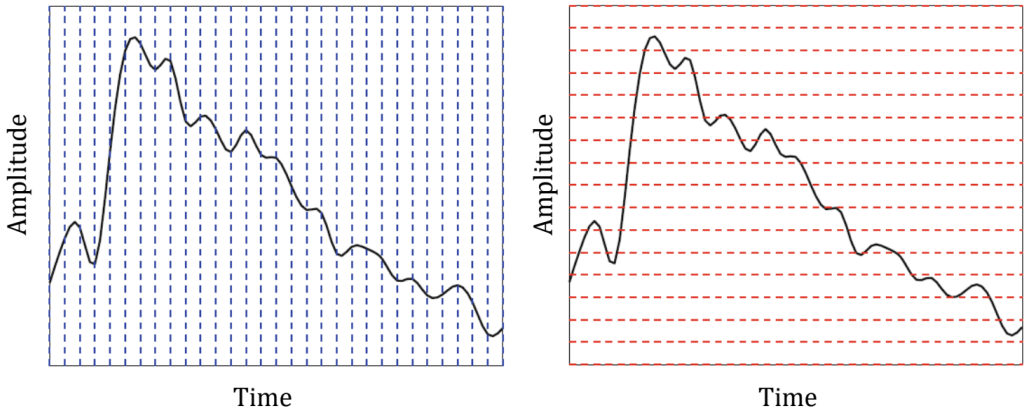
[**声音数字化**]

(1) 原因:

① 声波是压力波, 取连续值.

② 为将声波作为数字信号处理, 需将模拟信号**数字化**, 即将连续值转化为一系列的离散值.

(2) 过程:



① 采样: 在时间方向 (横向) 采样.

② 量化: 在幅度方向 (纵向) 量化.

[**Nyquist 理论**]

(1) **Nyquist 理论**:

① 采样率至少为信号最高频率的两倍, 该频率称为 **Nyquist 采样率**.

② 若信号大小 $\in [f_1, f_2]$, 则采样率至少为 $2 \cdot (f_2 - f_1)$.

(3) Nyquist 采样率的一半称为 **Nyquist 频率**.

(4) 采样频率 f_{sampling} 、真频 f_{true} 和假频 f_{alias} 的关系: $\begin{cases} f_{\text{alias}} = f_{\text{sampling}} - f_{\text{true}} \\ f_{\text{true}} < f_{\text{sampling}} < 2 \cdot f_{\text{true}} \end{cases}$.

[**例**]

[1] 何时会出现假频?

[2] 设真实频率为 $22.05\ \mathrm{Hz}$, 采样频率为 $33.075\ \mathrm{Hz}$. 求假频.

[**解**]

[1] $f_{\mathrm{true}} < f_{\mathrm{sampling}} < 2 \cdot f_{\mathrm{true}}$, 其中 f_{true} 为真实频率、f_{sampling} 为采样频率.

[2] 因 $f_{\mathrm{true}} = 22.05\ \mathrm{Hz} < f_{\mathrm{sampling}} = 33.075\ \mathrm{Hz} < 44.1\ \mathrm{Hz} = 2 \cdot f_{\mathrm{true}}$, 故有假频.

$$f_{\mathrm{alias}} = f_{\mathrm{sampling}} - f_{\mathrm{true}} = (33.075 - 22.05)\ \mathrm{Hz} = 11.025\ \mathrm{Hz} .$$

[**信噪比**, **Signal to Noise Ratio**, **SNR**]

(1) 常用单位: 分贝 (dB) , $1\ \mathrm{dB} = \dfrac{1}{10}\ \mathrm{bel}$.

(2) 以 dB 为单位时的定义: $\mathrm{SNR} = 10 \log_{10} \dfrac{V_{\mathrm{signal}}^2}{V_{\mathrm{noise}}^2} = 20 \log_{10} \dfrac{V_{\mathrm{signal}}}{V_{\mathrm{noise}}}$.

[**例**] 设声音信号电压为 $10\ \mathrm{V}$, 噪声电压为 $0.1\ \mathrm{V}$, 求 SNR .

[**解**] $\mathrm{SNR} = 20 \log_{10} \dfrac{V_{\mathrm{signal}}}{V_{noise}} = 20 \log_{10} \dfrac{10}{0.1} = 40\ \mathrm{dB}$.

[**信号量化噪声比**, **Signal to Quantization Noise Ratio**, **SQNR**]

(1) **量化误差**指真实值与最近的量化值的差, 不超过离散间距的一半.

(2) 定义: 量化精度为 N 位时, $\mathrm{SQNR} = 20 \log_{10} \dfrac{V_{\mathrm{signal}}}{V_{\mathrm{quan\_noise}}} \approx 6.02 \cdot N\ \mathrm{dB}$,

该定义实际上是**信号量化噪声比峰值** (**Peak Signal-to-Quantization-Noise Ratio**, **PSQNR**).

(3) $6.02 \cdot N$ 是最坏情况. 若输入信号是正弦信号, 则 $\mathrm{SQNR} = (6.02 \cdot N + 1.76)\ \mathrm{dB}$.

N 越大, SQNR 越大, 对模拟信号的逼近越精确, 能提供的音质越好.

[**例**] 某电脑有一块 16 位的声卡.

[1] 16 位指什么?

[2] 求 SQNR .

[**解**]

[1] 16 位指**模数转换器** (**Analog-to-Digital Converter**, **ADC**) 的位数, 表示声卡将模拟信号转化为数字信号时使用 16 位量化器, 它可将模拟信号分为 $2^{16} = 65536$ 个离散的级别.

[2] $\mathrm{SQNR} \approx (6.02 \cdot N + 1.76)\ \mathrm{dB} = (6.02 \times 16 + 1.76)\ \mathrm{dB} = 98.08\ \mathrm{dB}$.

[**Weber 定律**]

(1) **Weber 定律**: 要产生同样的感知, 所需要的增幅与原来的绝对值成比例, 即 $\Delta\text{Response} \propto \dfrac{\Delta\text{Stimulus}}{\text{Stimulus}}$.

(2) 推论: 声音的强度越大, 就需要更大的振幅使人感受到声音的变化.

(3) 应用: 利用该感知特性设计非均匀量化方案 μ 律和 A 律, s.t. 信噪比在输入信号范围内分布更均匀.

[**例**] 若人能感受到重量从 10 到 11 的变化, 则从 20 开始时, 需要 22 才能感受到重量的变化.

[**量化**]

(1) 线性量化: 采样存储为均匀分布的离散值.

(2) 非线性量化: 用更多位表示人耳最灵敏的声音区域.

[**例**] 选择采样频率为 $22.05\ \text{kHz}$ 、量化位数为 16 的录音参数, 在不使用压缩下, 计算 $2\ \text{min}$ 的立体声占的存储空间.

[**解**] $\text{每秒样本数} = \text{采样频率} \times 1\ \text{s} = 22.05\ \text{kHz} \times 1\ \text{s} = 22050$.

$\text{样本大小} = 16\ \text{bit} = 2\ \text{B}$.

$\text{单声道每秒大小} = \text{每秒样本数} \times \text{样本大小} = 22050 \times 2\ \text{B} = 44100\ \text{B}$.

立体声有 2 个声道, 则 $\text{每秒的大小} = 2 \times \text{单声道每秒大小} = 2 \times 44100\ \text{B} = 88200\ \text{B}$.

$\text{总大小} = \text{每秒的大小} \times \text{时间} = \dfrac{88200\ \text{B} \times 2 \times 60}{1024 \times 1024}\ \text{MB} \approx 10.09\ \text{MB}$.

[**例**] 简述频率遮蔽和时间遮蔽, 以及它们在音频编码中的应用.

[**答**]

(1) 定义:

① 频率遮蔽: 两声音频率相近时, 较强的声音会遮蔽较弱的声音, 使得人耳难察觉较弱的声音.

② 时间遮蔽: 两声音时间相近时, 较强的声音会遮蔽较弱的声音.

分类:

i) 前遮蔽: 较强的声音在出现前一段时间内会遮蔽较弱的声音.

ii) 后遮蔽: 较强的声音在消失后一段时间内会遮蔽较弱的声音.

(2) 应用:

① 有损音频压缩中去掉了某些被遮蔽的声音, 以减少信息量.

② MPEG 音频利用遮蔽建立了多维查找表, 该表记录被频率遮蔽或时间遮蔽的频率分量, 以压缩音频数据.

[**例**] 在 Yanny 和 Laurel 声音信号中, 为何有人只能听到 Yanny , 有人只能听到 Laurel ?

[**答**]

(1) 频率遮蔽现象. "Yanny" 主要在高频, "Laurel" 主要在低频.

(2) 不同人对高频和低频的敏感度不同, 与听觉系统和年龄有关.

(3) 随年龄增大, 听力下降, 首先失去对高频的感知, 故年龄小的人更可能听到 Yanny , 年龄大的人更可能听到 Laurel .

5.2 音频的量化与传输

5.2.1 脉冲编码调制 (PCM)

[**音频的编码**]

(1) 策略: 对音频信号, 除用扩展压缩音频信号的 μ 律外, 还可用当前时刻的信号和前一时刻的信号逐差来消除信号中的即时冗余.

(2) 优点:

① 降低传输的数据量.

② 差值集中在很小的范围内, 无损压缩更高效, 压缩率高.

(3) 分类:

① 生成音频的量化采样的方法统称**脉冲编码调制** (**Pulse Code Modulation**, **PCM**) .

② PCM 的差分版本称为**差分脉冲编码调制** (**Differential Pulse Code Modulation**, **DPCM**) .

③ DPCM 的粗略但有效的版本称为**增量调制** (**Delta Modulation**, **DM**) .

④ DPCM 的自适应版本称为**自适应差分脉冲编码调制** (**Adaptive Differential Pulse Code Modulation**, **ADPCM**) .

[**脉冲编码调制**, **Pulse Code Modulation**, **PCM**]

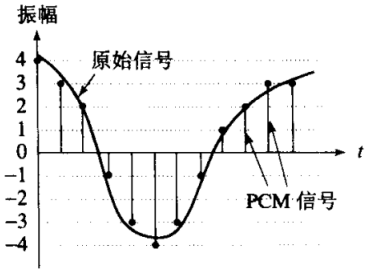
(1) 工作: 将模拟信号转化为数字信号, 即**模/数转换**.

(2) 步骤:

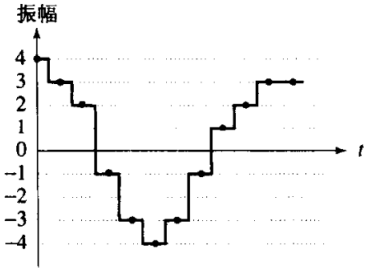
① 采样.

② 量化.

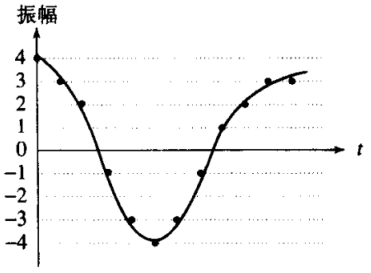
(3) 数/模转换示意图:



a）原始模拟信号及其相应的 PCM 信号

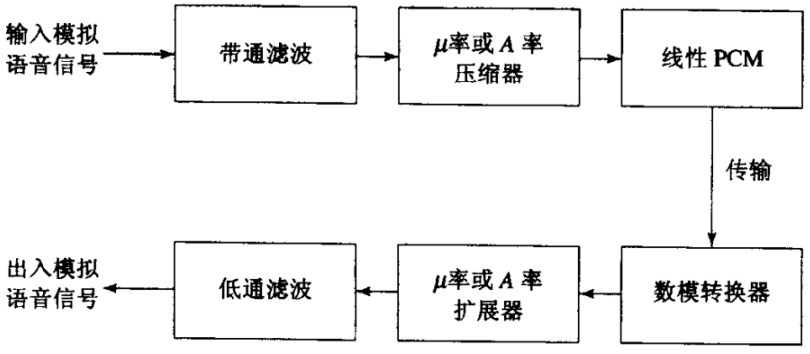


b）解码的阶梯信号



c）低通滤波后重构的信号

(4) PCM 的编解码过程的示意图:



(5) 压缩的三个阶段:

① 变换: 将输入数据转化为一个易于压缩或压缩效果更好的表示.

② 失真:

(i) 失真主要在量化中产生.

(ii) 因只使用有限数量的重构层, 其数量远小于原始信号的值的数量, 故量化过程必然损失信息.

③ 编码: 为每个输出层设置码字, 生成二进制流.

[**例**] 什么是模拟信号、数字信号? 它们的最主要区别是什么?

[**答**]

(1) 定义:

① 模拟信号是连续变化的信号, 其幅值可在一定范围内取任意值.

模拟信号常用于表示物理量的变化, 如声音、温度、光强等.

② 数字信号是离散的信号, 其幅值只能在有限、离散的值中变化.

数字信号常用于计算机和数字电子设备.

(2) 最主要区别:

① 模拟信号连续, 可在任意时间点取任意值.

② 数字信号离散, 只能在特定时间点取有限的值.

5.2.2 差分编码与无损预测编码

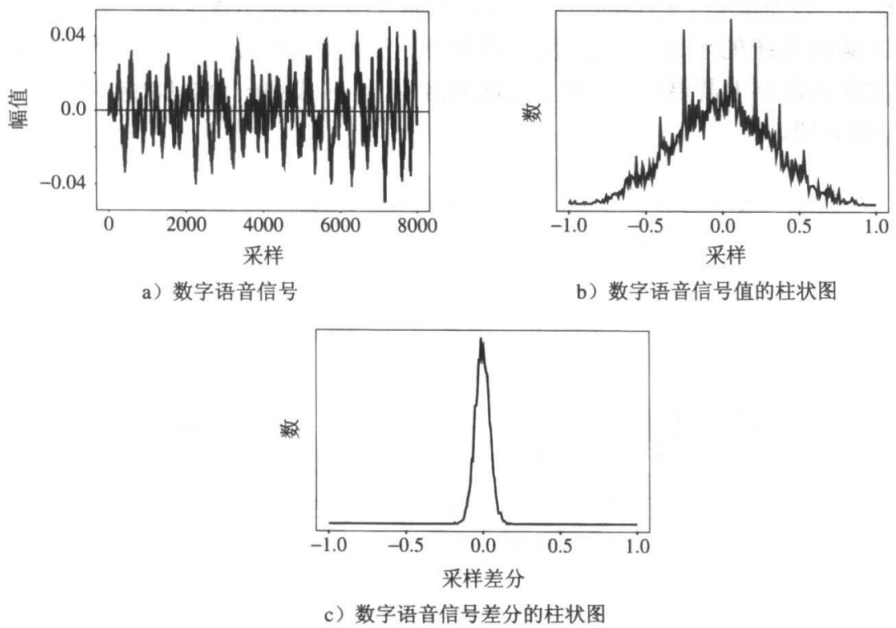
[**差分编码**]

(1) 原因:

① 消除信号的即时冗余.

② 得到在零点附近更集中的直方图, 进而可为出现频率高的符号分配短码字, 为出现频率低的符号分配长码字, 减小存储空间.

例:



(2) 使得差分的直方图更集中的方法: 为出现频率高的符号分配短码字, 为出现频率低的符号分配长码字.

[**无损预测编码**]

(1) 策略:

① 预测下一样本值与当前样本值相等.

② 用 PCM 发送采样值和预测值的误差.

(2) 形式化表示: 注意整数的差是整数.

① 输入信号: f_n .

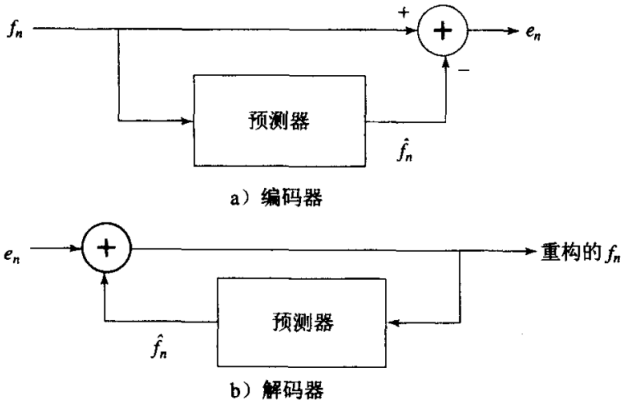
② 预测值: $\hat{f}_n = f_{n-1}$.

③ 误差: $e_n = f_n - \hat{f}_n$.

为 s.t. e_n 尽量小, 用以前的一些 f_i 值组成的函数能更好地预测 f_n , 如 $\hat{f}_n = \sum_{k=1}^{2\sim4} a_{n-k} \cdot f_{n-k}$.

(3) 特别大的差分值用两个新的码字 SU (Shift-Up) 和 SD (Shift-Down) 表示.

(4) 示意图:



[**例**] 音频的无损预测编码中, 设采样值范围为 $[0, 255]$, 则差分值范围为 $[-255, 255]$.将差分值 $[-15, 16]$ 定义为码字, 定义 SU 、SD 分别表示增加或减小 32 . 写出如下差分值的编码:

[1] 100 .

[2] -65 .

[**解**]

[1] 100 编码为 $\mathrm{SU}, \mathrm{SU}, \mathrm{SU}, 4$.

[2] -65 编码为 $\mathrm{SD}, \mathrm{SD}, -1$.

[**例**] 对输入序列 $[f_1,\cdots,f_5]=[21,22,27,25,22]$ 作预测编码, 取预测值 $\hat{f}_n=\left\lfloor\dfrac{f_{n-1}+f_{n-2}}{2}\right\rfloor\ (n\geq 3)$.

[**解**] 误差 $e_n=f_n-\hat{f}_n$.

(1) $\hat{f}_2=f_1=21$, $e_2=f_2-\hat{f}_2=22-21=1$.

(2) $\hat{f}_3=\left\lfloor\dfrac{f_2+f_1}{2}\right\rfloor=\left\lfloor\dfrac{22+21}{2}\right\rfloor=21$, $e_3=f_3-\hat{f}_3=27-21=6$.

(3) $\hat{f}_4=\left\lfloor\dfrac{f_3+f_2}{2}\right\rfloor=\left\lfloor\dfrac{27+22}{2}\right\rfloor=24$, $e_4=f_4-\hat{f}_4=25-24=1$.

(5) $\hat{f}_5=\left\lfloor\dfrac{f_4+f_3}{2}\right\rfloor=\left\lfloor\dfrac{25+27}{2}\right\rfloor=26$, $e_5=f_5-\hat{f}_5=22-26=-4$.

5.2.3 差分脉冲编码调制 (DPCM)

[**差分脉冲编码调制**, **Differential Pulse Code Modulation**, **DPCM**]

(1) 策略: 在无损预测编码的基础上增加量化.

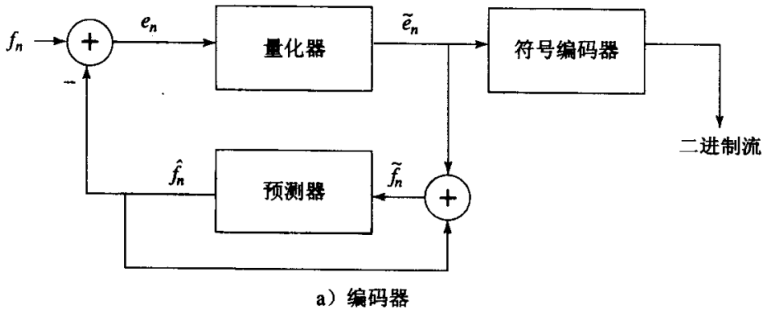
(2) 符号:

① f_n 量化并重构的信号: $\widetilde{f}_n$.

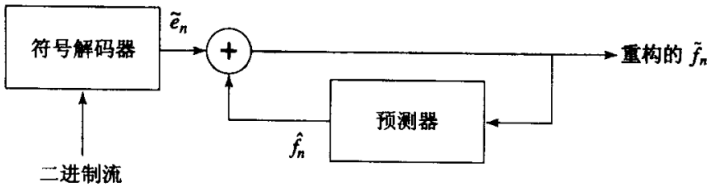
② 量化后的误差值: $\tilde{e}_n$.

(3) DPCM 方程组: $\begin{cases}\hat{f}_n=\text{function\_of}\left(\widetilde{f}_{n-1},\widetilde{f}_{n-2},\cdots\right)\\ e_n=f_n-\hat{f}_n\\ \tilde{e}_n=Q[e_n]\\ \text{传输 }\tilde{e}_n\text{ 的码字}\\ \text{重建: }\widetilde{f}_n=\hat{f}_n+\tilde{e}_n\end{cases}$, 使用熵编码生成 $\tilde{e}_n$ 的码字.

(4) 示意图:



a) 编码器



b) 解码器

(5) 生成 $\hat{f}_n$ 时需用到之前的一些预测值 $\widetilde{f}_i$, 故需缓存这些值.

(6) 量化噪声 $f_n-\widetilde{f}_n$ 等于误差 e_n 因量化产生的误差 $e_n-\tilde{e}_n$.

[**证**] $e_n-\tilde{e}_n=\left(f_n-\hat{f}_n\right)-\left(\widetilde{f}_n-\hat{f}_n\right)=f_n-\widetilde{f}_n$.

(7) 因预测器使用重构的量化信号值 $\widetilde{f}_n$, 故编码器中的预测器实现了解码器的功能.

[例] DPCM 编码中, 取预测器 $\hat{f}_n=\operatorname{trunc}\left(\dfrac{\widetilde{f}_{n-1}+\widetilde{f}_{n-2}}{2}\right)$, 则误差 $e_n=f_n-\hat{f}_n\in\mathbb{Z}$. 取量化器 $\tilde{e}_n=Q[e_n]=16\cdot\operatorname{trunc}\left(\dfrac{255+e_n}{16}\right)-256+8$, 则重构值 $\widetilde{f}_n=\hat{f}_n+\tilde{e}_n\in\mathbb{Z}$.

函数 $\operatorname{trunc}(x)$ 表示对 x 向零取整, 如 $\operatorname{trunc}(1.23)=1$, $\operatorname{trunc}(-1.23)=-1$, $\operatorname{trunc}(0)=0$.

注意到 $e_n\in[-255,255]$, 即有 511 个不同的误差值, 量化器 Q 将该范围分为 32 块, 除最后一块包含 15 个值外, 其它的块都包含 16 个值, 每块的代表值为其对应的 15 ~ 16 个取值的中间值.

量化表如下, 其中 4 位编码以阶梯函数的形式映射到 32 个重构值.

e_n 的范围	量 化 值
-255~-240	-248
-239~-224	-232
⋮	⋮
-31~-16	-24
-15~0	-8
1~16	8
17~32	24
⋮	⋮
225~240	232
241~255	248

在上述配置下, 对输入信号值 $[f_1,\cdots,f_5]=[130,150,140,200,230]$ 作 DPCM.

[解]

(1) 假设 $\hat{f}_1=f_1=130$, 则 $e_1=f_1-\hat{f}_1=0$.

假设初始量化误差 $\tilde{e}_1=0$, 则第一个重构值 $\widetilde{f}_1=\hat{f}_1+\tilde{e}_1=130+0=130$.

(2) 取 $\hat{f}_2=f_1=130$, 则 $e_2=f_2-\hat{f}_2=150-130=20$.

$\tilde{e}_2=Q[e_2]=24$, 则 $\widetilde{f}_2=\hat{f}_2+\tilde{e}_2=130+24=154$.

(3) $\hat{f}_3=\operatorname{trunc}\left(\dfrac{\widetilde{f}_2+\widetilde{f}_1}{2}\right)=\operatorname{trunc}\left(\dfrac{154+130}{2}\right)=142$,

则 $e_3=f_3-\hat{f}_3=140-142=-2$.

$\tilde{e}_3=Q[e_3]=-8$, 则 $\widetilde{f}_3=\hat{f}_3+\tilde{e}_3=142-8=134$.

(4) $\hat{f}_4=\operatorname{trunc}\left(\dfrac{\widetilde{f}_3+\widetilde{f}_2}{2}\right)=\operatorname{trunc}\left(\dfrac{134+154}{2}\right)=144$,

则 $e_4=f_4-\hat{f}_4=200-144=56$.

$$\tilde{e}_4=Q[e_4]=16\cdot\operatorname{trunc}\left(\frac{255+e_4}{16}\right)-256+8=16\cdot\operatorname{trunc}\left(\frac{255+56}{16}\right)-256+8$$

$=16\times19-256+8=56$, 则 $\widetilde{f}_4=\hat{f}_4+\tilde{e}_4=144+56=200$.

(5) $\hat{f}_5 = \text{trunc}\left(\dfrac{\widetilde{f}_4 + \widetilde{f}_3}{2}\right) = \text{trunc}\left(\dfrac{200+134}{2}\right) = 167$,

则 $e_5 = f_5 - \hat{f}_5 = 230 - 167 = 63$.

$$\tilde{e}_5 = Q[e_5] = 16 \cdot \text{trunc}\left(\frac{255+e_5}{16}\right) - 256 + 8 = 16 \cdot \text{trunc}\left(\frac{255+63}{16}\right) - 256 + 8$$

$= 16 \times 19 - 256 + 8 = 56$, 则 $\widetilde{f}_5 = \hat{f}_5 + \tilde{e}_5 = 167 + 56 = 223$.

综上, 编码结果:

i	1 **(假设值)**	2	3	4	5
$\hat{f}_i$	130	130	142	144	167
e_i	0	20	-2	56	63
$\tilde{e}_i$	0	24	-8	56	56
$\widetilde{f}_i$	130	154	134	200	223

5.2.4 增量调制 (DM)

[**增量调制**, **Delta Modulation**, **DM**]

(1) **DM** 是 DPCM 的简化版, 常用作快速的模/数转换器.

(2) 策略:

① 使用唯一的量化误差值, 该值可为正数或负数.

② 使用 1 位的编码器使得原始信号编码后的输出呈阶梯状.

(3) DM 方程组: $\begin{cases} \hat{f}_n = \widetilde{f}_{n-1} \\ e_n = f_n - \hat{f}_n = f_n - \widetilde{f}_{n-1} \\ \tilde{e}_n = \begin{cases} +k, & e_n > 0 \\ -k, & e_n \le 0 \end{cases} \ (k \in \text{Const.}) \\ \widetilde{f}_n = \hat{f}_n + \tilde{e}_n \end{cases}$. 注意预测只涉及一个时延.

(4) 适用场景: 信号基本维持常值而很少变动的场景.

(5) 缺点:

① 信号变化剧烈时, DM 效果不好.

改进: 提高采样频率至 Nyquist 采样率的若干倍.

② 信号波形很陡峭时, 阶梯状的逼近效果不好.

改进: 自适应地改变步长 k , 即**自适应 DM** .

[**例**] 对输入信号值 $[f_1, \cdots, f_4] = [10, 11, 13, 15]$ 作 DM , 取步长 $k = 4$.

[**解**]

(1) 假设 $\widetilde{f}_1 = f_1 = 10$.

(2) $\hat{f}_2 = \widetilde{f}_1 = 10$, 则 $e_2 = f_2 - \hat{f}_2 = 11 - 10 = 1 > 0$,

则 $\tilde{e}_2 = +k = 4$, 进而 $\widetilde{f}_2 = \hat{f}_2 + \tilde{e}_2 = 10 + 4 = 14$.

(3) $\hat{f}_3 = \widetilde{f}_2 = 14$, 则 $e_3 = f_3 - \hat{f}_3 = 13 - 14 = -1 \le 0$,

则 $\tilde{e}_3 = -k = -4$, 进而 $\widetilde{f}_3 = \hat{f}_3 + \tilde{e}_3 = 14 - 4 = 10$.

(4) $\hat{f}_4 = \widetilde{f}_3 = 10$, 则 $e_4 = f_4 - \hat{f}_4 = 15 - 10 = 5 > 0$,

则 $\tilde{e}_4 = +k = 4$, 进而 $\widetilde{f}_4 = \hat{f}_4 + \tilde{e}_4 = 10 + 4 = 14$.

综上, 编码结果为 $[10, 14, 10, 14]$.

5.2.5 自适应差分脉冲编码调制 (ADPDM)

[**自适应差分脉冲编码调制**, **Adaptive Differential Pulse Code Modulation**, **ADPDM**]

(1) 策略: 自动调整编码机制, 更好地适应输入值.

(2) 方式:

① **前向自适应量化**: 利用输入信号的特点.

② **后向自适应量化**: 利用量化输出信号的特点. 若量化误差过大, 需修改非均匀 Lloyd-Max 量化器.

(3) 自适应编码:

预测器一般是以前的重构量化信号值 $\widetilde{f}_i$ 的函数, 称预测器使用的以前的预测值的个数为预测器的**阶**.

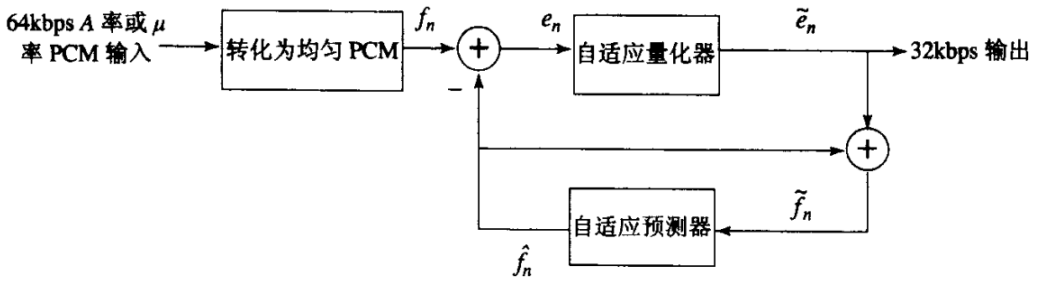
若预测器需要 M 个以前的预测值, 则预测器需要 M 个系数 $a_1, \cdots, a_M$, 即 $\hat{f}_n = \sum_{i=1}^{M} a_i \cdot \widetilde{f}_{n-i}$.

通过最小二乘问题 $\min \sum_{n=1}^{M} \left(f_n - \hat{f}_n \right)^2$ 求 $a_i \ \ (i = 1, \cdots, M)$ 的最优解.

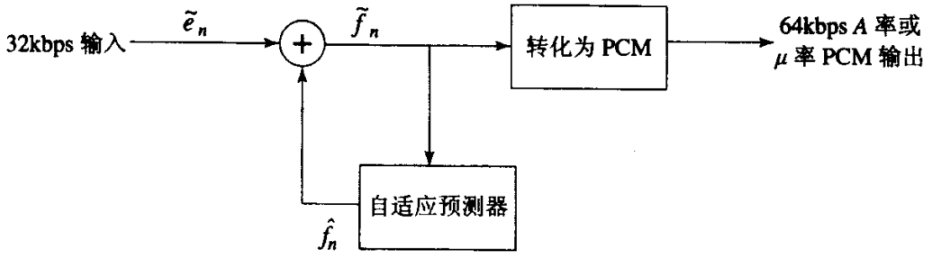
在上述问题中不使用预测重构值 $\widetilde{f}_n$, 而使用原始信号值 f_n,

问题转化为 $\min \sum_{n=1}^{M} \left(f_n - \sum_{i=1}^{M} a_i \cdot f_{n-i} \right)^2$, 令一些 a_i 取 0, 得到易解的 M 个方程.

(4) 示意图:



a）编码器



b）解码器