Algumas considerações sobre quantificação e distorção de quantificação

Francisco Vaz

Abstract **– Definitions of 1-D quantizer, quantization distortion and optimal quantizer are presented as well as examples of uniform and non uniform quantizers. Results of quantization distortion for several types of signals characterized by probability density function are also presented enabling to point some rules for a good choice of the quantizer**

Resumo **– No texto apresentam-se as definições de quantificador, distorção de quantificação e quantificador óptimo a uma dimensão. Estas noções são em seguida exemplicadas para quantificadores uniformes e não uniformes e são apresentados resultados da distorção de quantificação para vários tipos de sinal caracterizados pela sua função de distribuição que permitem evidenciar algumas regras para uma boa escolha de quantificadores**

Keywords **– quantizer, uniform quantizer, optimal quantizer, quantization distortion, k-means algorithm**

Palavras chave **– quantificador, quantificador uniforme, quantificador óptimo, distorção de quantificação, algoritmo das k-médias**

I. INTRODUÇÃO

Os sinais provenientes de sistemas físicos consideram-se em geral como variáveis contínuas. Para serem procesados digitalmente é necessário proceder à sua discretização temporal, feita habitualmente por uma amostragem uniforme. A grandeza das amostras continua a ser representável por um número real dentro de um intervalo equivalente à gama de variação do sinal. Para permitir o processamento digital esta representação contínua deve ser substituida por uma outra onde se use apenas um número finito e, em geral, reduzido de valores ou códigos. A esta operação chama-se quantificação e é realizável directamente por circuitos electrónicos que recebem a amostra do sinal e a transformam num valor numérico que a representa. No texto que se segue far-se-á uma revisão de quantificação, de distorção de quantificação e de quantificador óptimo.

II. QUANTIFICAÇÃO

Definição 1: Sendo $C = \{c_i, i = 1, 2, \ldots, N\} \subset \mathbb{R}$, um quantificador é uma aplicação de $\mathbb{R} \to C$ tal que

$$Q : x \in \mathbb{R} \to Q(x) \in C$$

Ao conjunto finito C chama-se livro de códigos e a $b = \log_2 N$ chama-se resolução em bits do quantificador. Associado ao livro de códigos está uma partição dos números reais em conjuntos $R_i, i = 1, 2, \ldots, N$, definidos por

$$R_i = \{x \in \mathbb{R} : Q(x) = c_i\} = \left\{Q^{-1}(c_i)\right\}$$

com $\mathbb{R} = \bigcup_{i=1}^{N} R_i$, e $\forall_{i,j} R_i \cap R_j = \varnothing$.

Definindo um conjunto de extremos $x_0 < x_1 < \ldots < x_N$ com $x_0 = -\infty$ e $x_N = \infty$ pode-se associar a cada elemento da partição um intervalo $(x_i, x_{i+1}]$ da recta real.

Aos subconjuntos R_i chamam-se células e podem ser de dois tipos - abertas ou fechadas - conforme os seus extremos sejam finitos ou não. Um quantificador pode ser visto como o resultado combinado de duas operações, uma de codificação e outra de descodificação. Designando por $I = \{1, 2, \ldots, N\}$ o conjunto dos índices e definindo codificação ENC como a aplicação $ENC : x \in \mathbb{R} \to i \in I$ e descodificação DEC como a aplicação $DEC : i \in I \to x \in C \subset \mathbb{R}$ a quantificação surge como a composição $DEC \circ ENC$.

$$Q(x) = DEC(ENC(x))$$

O processo de quantificação modifica o sinal original. Para caracterizar esta modificação introduz-se a distorção de quantificação que é uma medida de distância entre o valor original x e o valor quantificado $Q(x)$.

$$d(x, Q(x)) = \begin{cases} 0 & sse\ x = Q(x) \\ > 0 & x \neq Q(x) \end{cases}$$

e gozando das propriedades

1. Comutativa: $d(x, y) = d(y, x)$
2. Desigualdade triangular: $d(x, y) \leq d(x, u) + d(u, y)$
3. Invariância $\forall_z d(x, y) = d(x + z, y + z)$

Sendo $f_x(x)$ a função de densidade de probabilidade da variável aleatória que representa o sinal a quantificar, o seu valor médio será

$$D = E\left[d(x, Q(x))\right] = \int_{-\infty}^{+\infty} d(x, Q(x)) f_x(x) dx$$

e uma forma habitual de avaliar o desempenho do quantificador será determinada pela relação sinal ruído logarítmica, expressa em dB

$$SNR = 10 \log_{10} \frac{\sigma_x^2}{D}$$

Atendendo às definições anteriores e às regras gerais de probabilidades o valor médio da distorção pode ainda ser expresso por

$$D=\sum_{i=1}^{N}\int_{R_i} d(x,Q(x))f_x(x)dx = \qquad (1)$$

$$=\sum_{i=1}^{N} P\left[x\in R_i\right] E\left[d(x,Q(x))|x\in R_i\right]$$

É usual utilizar como medida a distância euclideana:

$$d(x,Q(x))=(x-Q(x))^2$$

e neste caso

$$D=\sum_{i=1}^{N}\int_{x_{i-1}}^{x_i}(x-c_i)^2 f_x(x)dx \qquad (2)$$

O erro de quantificação será

$$e=x-Q(x)$$

Para um conjunto discreto de amostras $x_1,\ldots,x_M$ igualmente prováveis e para o caso de distância euclideana a equação 1 conduz a

$$D=\sum_{i=1}^{N}\frac{M_i}{M}\sum_{x_k\in R_i}\frac{(x-c_i)^2}{M_i}$$

em que M_i é o número de elementos de R_i. Simplificando facilmente se conclui que

$$D=\frac{1}{M}\sum_{k=1}^{M}(x_k-Q(x_k))^2$$

a distorção iguala o valor médio quadrático das amostras do erro de quantificação.

III. QUANTIFICADOR UNIFORME

Um caso importante é o do quantificador uniforme que obedece às seguintes condições

$$c_{i+1}-c_i=\Delta, i=1,\ldots,N-1$$
$$x_{i+1}-x_i=\Delta, i=1,\ldots,N-1$$

isto é, intervalos todos iguais e códigos igualmente afastados. O livro de códigos será

$$c_i=x_i-\Delta/2, i=1,\ldots,N-1$$
$$c_N=x_{N-1}+\Delta/2$$

O erro de quantificação está limitado ao intervalo $[-\Delta/2,\Delta/2]$ para valores de x que vão de $x_{\min}=x_1-\Delta$ a $x_{\max}=x_{N-1}+\Delta$. Pode-se portanto definir estes valores limites tendo em consideração a gama de variação do sinal e determinar $\Delta=(x_{\max}-x_{\min})/N=2x_{\max}/N$, se a gama de variação for simétrica, isto é, $x_{\min}=-x_{\max}$. Na figura 1 apresenta-se, como exemplo, um quantificador de 3 bits, em que se escolheu $\Delta=1$ e um livro de códigos $C=\{-4,-3,\cdots,0,\cdots,3\}$

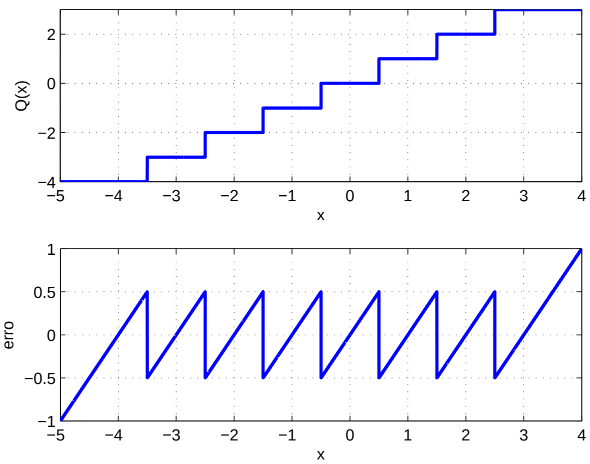

Fig. 1 - Aplicação de um quantificador uniforme de 3 bits a um sinal; sinal, sinal quantificado e erro de quantificação

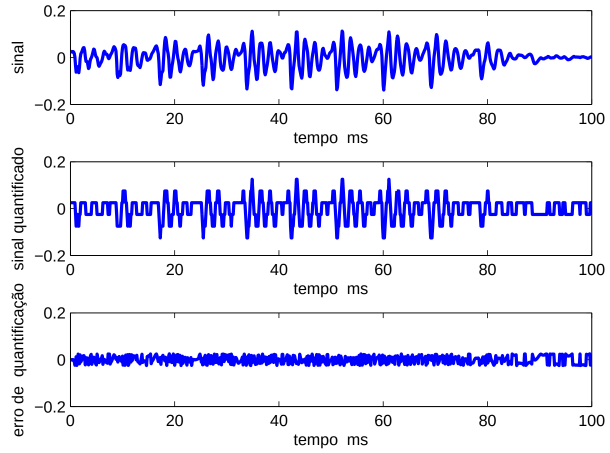

Fig. 2 - Aplicação de um quantificador uniforme de 3 bits a um sinal; sinal, sinal quantificado e erro de quantificação

O resultado da aplicação de um quantificador de 3 bits a um sinal está exemplificado a seguir, na figura 2

Admitindo que o sinal de entrada é uniformemente distribuido na gama $[-x_{\max},x_{\max}]$ isto significa que a sua densidade de probabilidade vale

$$f_x(x)=\begin{cases}\dfrac{1}{2x_{\max}} & |x|\le x_{\max}\\ 0 & |x|>x_{\max}\end{cases}$$

o valor da distorção será dado por

$$D = \sum_{i=1}^{N} \int_{x_{i-1}}^{x_i} (x-c_i)^2 f_x(x) dx =$$

$$D = \frac{1}{2x_{\max}} \sum_{i=1}^{N} \int_{c_i-\Delta/2}^{c_i+\Delta/2} (x-c_i)^2 dx =$$

$$= \frac{1}{2x_{\max}} \sum_{i=1}^{N} \frac{2}{3}\left(\frac{\Delta}{2}\right)^3 = \frac{1}{2x_{\max}} N \frac{2}{3}\left(\frac{\Delta}{2}\right)^3$$

e atendendo a que $\Delta/2 = x_{\max}/N$ a distorção será então

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

Este resultado pode ser interpretado como se o erro fosse uma variável aleatória uniformemente distribuida no intervalo $[-\Delta/2, \Delta/2]$ e a distorção fosse a sua variância. Definindo ainda factor de carga como

$$\gamma = \frac{x_{\max}}{\sigma_x}$$

a distorção poderá ser expressa como

$$D = \frac{1}{3}\left(\frac{\gamma \sigma_x}{N}\right)^2 = \frac{1}{3}\frac{\gamma^2 \sigma_x^2}{2^{2b}}$$

e a relação sinal-ruído será então

$$SNR = 20b \log_{10} 2 + 10 \log_{10} \frac{3}{\gamma^2} = 6.03b + 10 \log_{10} \frac{3}{\gamma^2}$$

A SNR para sinais uniformemente distribuidos na gama de variação cresce então 6dB por cada bit, e degrada-se se o factor de carga for elevado. Por outro lado haverá forte degradação se o sinal ultrapassar a gama dinâmica do quaantificador.

Para se caracterizar o desempenho do quantificador perante um sinal com uma determinada distribuição é necessário determinar a respectiva distorção integrando a equação 2. Recorrendo a técnicas de integração numérica mostra-se na figura 3 o desempenho de quanticadores de 3 a 10 bits, em função do factor de carga e para sinais com diferentes tipos de densidade de probabilidade: densidade uniforme, normal e de Laplace obedecendo, respectivamente, às leis:

$$f_x(x) = \begin{cases} 1/(s\sqrt{3}\sigma_x & |x| \leq \sqrt{3}\sigma_x \\ 0 & |x| > \sqrt{3}\sigma_x \end{cases}$$

$$f_x(x) = \frac{1}{\sqrt{2}\sigma_x} e^{-\sqrt{2}|x|/\sigma_x}$$

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-x^2/(2\sigma_x^2)}$$

Pode-se observar que as curvas na zona de factores de carga maiores são quase lineares e estão espaçadas por 6dB. Para factores de carga mais baixos, a distorção por efeito de limitação do quantificador faz degradar muito rapidamente o desempenho. As situações extremas são para as distribuições uniforme (melhor desempenho) e de Laplace (pior desempenho).

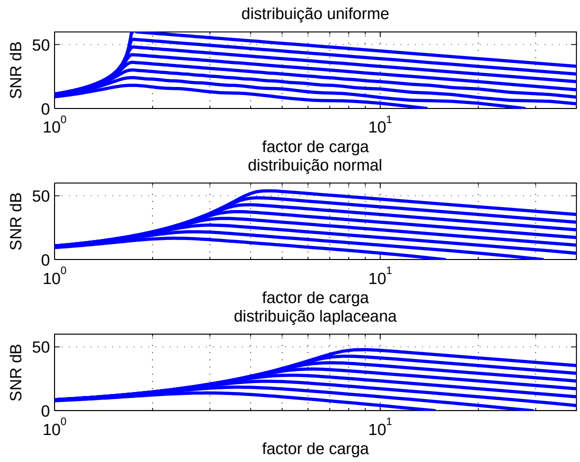

Fig. 3 - Curvas de distorção em função do factor de carga para quantificadores uniformes de 3 a 10 bits e para sinais com distribuição uniforme, normal e laplaceana

Dados estes resultados pode-se concluir que a melhor relação SNR de quantificadores uniformes se consegue para sinais uniformemente distribuidos na gama do quantificador. Ora os sinais reais não cumprem este requesito pelo que se deve pensar em métodos que permitam uma aproximação a estas condições.

IV. Quantificador óptimo

Definição 2: O quantificador óptimo é o quantificador que minimiza a distorção

Dada a definição de distorção, para uma dada densidade de probabilidade é preciso especificar os códigos c_i e as células R_i da partição para se caracterizar o quantificador óptimo. É possível demonstrar que para um livro de códigos C a partição óptima obedece à condição

$$R_i = \{x : d(x, c_i) \leq d(x, c_j), i \neq j\}, i = 1, \ldots N$$

o que equivale a afirmar que

$$d(x, Q(x)) = \min_i d(x, c_i)$$

Esta expressão mostra que dada uma amostra x, o código a atribuir é aquele que torna mínima a distorção/ distância da amostra ao código. Por esta razão se afirma que a partição óptima implementa a regra do vizinho mais próximo: de todo o livro de códigos escolhe-se o código mais próximo (no sentido de menor distorção) da amostra.

Definição 3: Centróide de um conjunto não vazio $A \subset \mathbb{R}$ é o ponto $a \in A$ que minimiza a distorção média entre x e a se $x \in A$, ou seja

$$cent(A) = a \text{ se}$$
$$\forall y \in A \quad E\left[d(x,a)|x \in A\right] \leq E\left[d(x,y)|x \in A\right]$$

Esta definição de centróide implica que os centróides de uma partição $R_i, i = 1, \ldots, N$ constituem os elementos do livro de códigos óptimo

$$cent(R_i) = c_i, i = 1, \ldots, N$$

Se a medida de distorção for o erro quadrático $(x - c_i)^2$ a mimização em ordem a c_i da equação 2 conduz facilmente ao resultado

$$c_i = \frac{\int_{R_i} x f_X(x)dx}{\int_{R_i} f_X(x)dx}$$

O problema que se coloca é o de determinar o livro de códigos óptimos dada a sua dimensão N e a caracterização estatística dos dados através de uma função de densidade de probabilidade. A solução pode ser iterativa, isto é, dado um livro de códigos pode-se determinar a partição óptima através da regra do vizinho mais próximo e em seguida determinar um novo livro de códigos a partir dos centróides da partição. Este processo iterativo dá origem a um algoritmo bem conhecido e designado de k-médias:

1. Escolha arbitrariarmente um livro de códigos C_1 e determine a distorção inicial D_i, faça $m = 1$
2. Determine a partição óptima $R_i, i = 1, \ldots, N$ pela regra dos vizinhos mais próximos
3. Determine o novo livro de códigos C_{m+1} e a distorção D_{m+1}
4. Se a a melhoria na distorção atingir um valor pré-definido pare. Caso contrário faça $m \leftarrow m + 1$ e regresse ao ponto 2

Um possível critério de paragem pode ser baseado na variação relativa da distorção $(D_m - D_{m+1})/D_m$. Contudo deve se referir que esta função da distorção não é necessariamente monótona. No caso de se dispôr de amostras $x_1, \ldots x_t$ e de se usar o erro quadrático, o cálculo dos centróides é simplificado pois

$$c_i = \frac{\int_{R_i} x f_X(x)dx}{\int_{R_i} f_X(x)dx} \simeq \frac{\sum_{x_k \in R_i} x_k \frac{1}{t_i}}{\sum_{x_k \in R_i} \frac{1}{t_i}} = \frac{1}{t_i} \sum_{x_k \in R_i} x_k$$

com t_i = número de elementos da classe R_i, ou seja os centróides são as médias aritméticas dos elementos de cada classe da partição. O algoritmo das k-médias será então:

1. Escolha um livro de códigos C_1 aleatoriamente entre as amostras e determine a distorção inicial D_i, faça $m = 1$
2. Distribua os elementos do conjunto de treino pelas N classes R_i da partição usando a regra do vizinho mais próximo.
3. Determine o novo livro de códigos C_{m+1} calculando as médias para cada R_i e a nova distorção D_{m+1}
4. Se $(D_m - D_{m+1})/D_m \leq L$ pare . Caso contrário faça $m \leftarrow m + 1$ e regresse ao ponto 2

A figura 4 mostra a aplicação do algoritmo das k-médias na obtenção de livros de código óptimos com $N = 4$. Foram usadas em cada caso 10000 amostras de distribuição uniforme no intervalo $[-1, 1]$ e distribuição normal de média nula e variância unitária, respectivamente, e o algoritmo pára quando a variação relativa da distorção for inferior a 10^{-6}. Note que para a distribuição uniforme o livro de códigos que se obtem é igual ao do codificador uniforme, o mesmo não sucedendo no outro caso.

A. Quantificador não uniforme

Um quantificador uniforme é, em geral, subóptimo, como se viu na secção anterior, e é uma consequência da distibuição estatística das amostras a quantificar. Uma maneira de se ultrapassar esta limitação ou pelo menos, de se se aproximar das condições óptimas, é usar uma quantificação não uniforme. Na quantificação não uniforme a variável de entrada começa por ser transfomada por uma função monótona e não linear f , sendo o resultado desta transformação quantificado por um quantificador uniforme Q_u. Os valores assim obtidos são em seguida tranformados pela função inversa f^{-1}.

Desta forma um quantificador não uniforme fica definido como uma aplicação de $x \in \mathbb{R} \rightarrow Q(x) \in C$ em que $Q(x) = f^{-1}(Q_u(f(x)))$.

A função f é escolhida por for forma a expandir as baixas amplitudes (onde o desempenho do quantificador uniforme é em geral mau) e comprimir as amplitudes elevadas (onde o desempenho do quantificador uniforme é em geral bom).

As funções mais usadas são do tipo logarítmico e apresenta-se a seguir uma muito usadas em aplicações de fala conhecida por lei-μ. Lei-μ

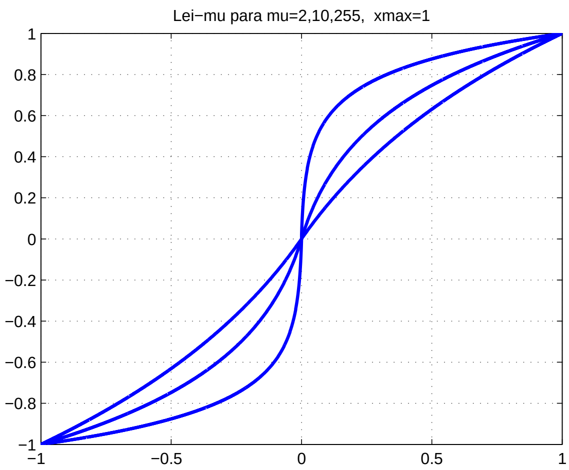

Fig. 5 - Lei μ de compressão

$$f(x) = x_{\max} \frac{\ln\left(1 + \mu\,|x|\,/x_{\max}\right)}{\ln\left(1 + \mu\right)} sign(x), \quad |x| \leq x_{\max}$$

a que corresponde uma função inversa:

$$f^{-1}(y) = \frac{x_{\max}}{\mu} \left(\exp(\frac{\ln(1+\mu).\,|y|}{x_{x\max}} \right) sign(y), \quad |y| \leq x_{\max}$$

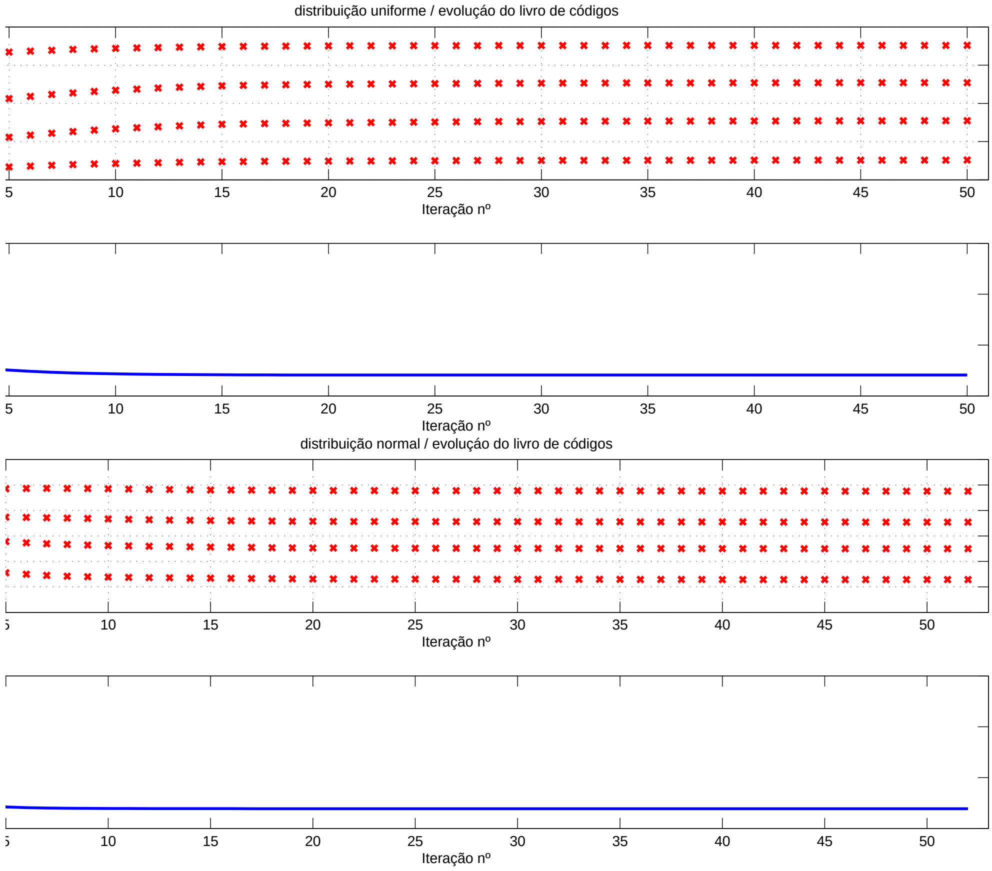

Fig. 4 - Algoritmo das k-médias: Evolução do livro de códigos e da distorção para distribuição uniforme e distribuição normal

Apresenta-se em seguida na figura 6 um quantificador não uniforme de 4 bits em que se usou a Lei-μ com $\mu = 10$ e $x_{\max} = 1$.

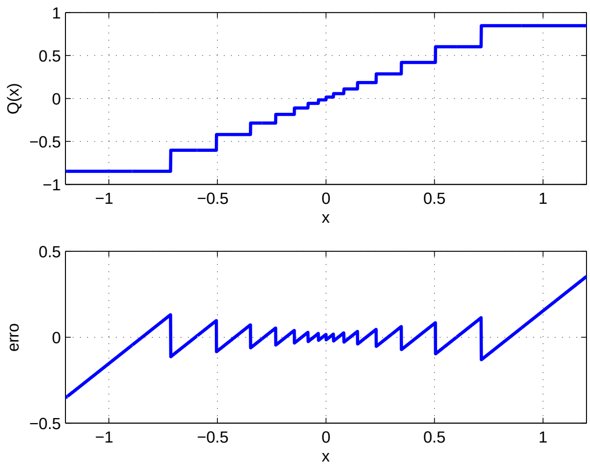

Fig. 6 - Quantificador não uniforme de 4 bits função de transferência e erro de quantificação

e na figura 7 um exemplo de quantificação não uniforme com 4 bits e $\mu = 255$

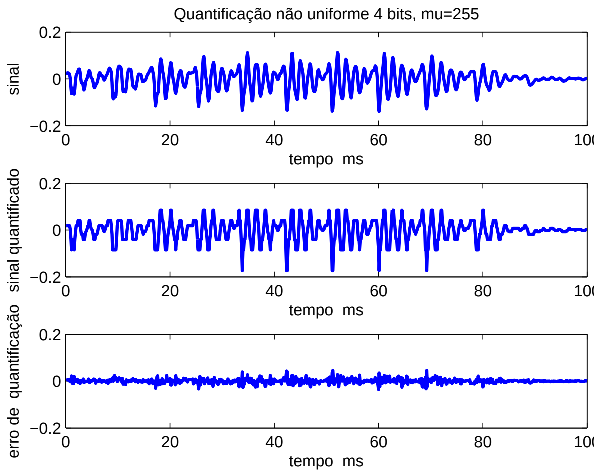

Fig. 7 - Qplicação de um quantificador não uniforme de 4 bits, $\mu = 255$ a um sinal: sinal quantificado e erro de quantificação

O efeito desta transformação é o de concentrar os códigos nas baixas amplitudes (figura 6) possibilitando uma melhor definição do sinal nessa zona. Já as as amplitudes altas são prejudicadas. Portanto este tipo de transformação só pode interessar se as distribuições

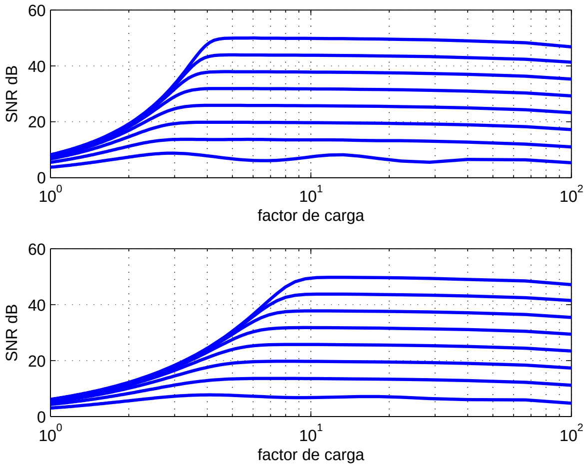

Fig. 8 - Curvas de distorção em função do factor de carga para quantificadores não uniformes de 3 a 10 bits, $\mu = 255$ e para sinais com distribuição normal e laplaceana

das amostras se concentram nas zonas de baixa amplitude, sendo as altas pouco frequentes, sendo exemplo os sinais de distribuição gausseana e laplaceana.

Esta análise qualitativa é confirmada quantitativamente efectuando a integração numérica da equação 2 da distorção, resultado que se apresenta na figura 8 onde se mostram os gráficos da SNR em função do factor de carga, para quantificadores não uniformes de 3 a 10 bits, usando a Lei- μ com $\mu = 255$ e para um sinal de distribuição gausseana e laplaceana. É claro da figura que a SNR se mantem alta e quási constante para uma elevada gama de valores do sinal de entrada, não piorando significativamente para valores altos do factor de carga.