Quantisation

Quantisation involves a map of the real line into a discrete set of quantisation levels. Given a set of $M$ quantisation levels $\{S_0, S_1, \ldots, S_{M-1}\}$, any real value $x$ is mapped into the quantised level $Q(x)$ closest to it: $Q(x) = S_i$ if $|x - S_i| < |x - S_j|$ for every $j \neq i$. Ties may be broken in any convenient fashion, for instance, by picking $i$ to be the smaller of the two arguments among two competing quantisation levels both equidistant from and at minimal distance to $x$.

Quantisation results in an irrevocable loss of information. A measure of this loss is the quantisation error $N_Q = x - Q(x)$. An evocative picture may be built up by thinking of sending a signal $x$ into a noisy channel (the quantiser) and obtaining a noisy response $Q(x) = x - N_Q$ consisting of the signal $x$ corrupted by an additive, signal-dependent “noise” term $N_Q = N_Q(x)$ at the output of the channel. The quantisation error is hence sometimes picturesquely, if not quite accurately, called quantisation noise.

If we increase the number of quantisation levels we say the the quantisation is finer; alternatively, if we decrease the number of quantisation levels we say that the quantisation is coarser. Clearly, the finer the quantisation the lower the potential error. Can we quantify how much increasing the number of quantisation levels buys us in performance?

Uniform Quantisation

Consider the simplest setting where the signal amplitude $x$ takes values in a bounded range $-A \leq x \leq A$. Suppose the quantiser has $M$ equispaced levels $S_0, S_1, \ldots, S_{M-1}$ available to it. Let $\Delta$ denote the spacing between successive quantisation levels so that, for each level $S_i$, there is a region of $\pm \Delta$ around $S_i$ which is mapped into it: $x \mapsto Q(x) = S_i$ if, and only if, $-\Delta < x - S_i \leq \Delta$. Thus, $M = 2A/2\Delta$ (where we ignore round-off issues) or, to put it another way, $\Delta = A/M$. The squared quantisation error $N_Q^2 = (x - Q(x))^2$ may be taken as indicative of the quantisation noise power. What is the average quantisation noise power in this setting?

If the number of levels $M$ is large, i.e., the quantisation is fine, each quantisation interval $[S_i - \Delta, S_i + \Delta]$ would be small. In such cases it may be permissible to imagine that the range of signal values in any quantisation interval are uniformly distributed in that interval. The mean squared quantisation noise

---

1Noise usually refers to random, uncontrolled, external perturbations which cause distortion or error. In this case, however, the “noise” is of our own making—we incur the quantisation error with malice aforesaid.

2We've tacitly adopted the tie breaking rule suggested above: in case of a tie, we assign $x$ to the lower quantisation level.
quantisation error would then be given by a simple weighted average of the squared error across the quantisation interval. We may identify this quantity as the average quantisation noise power. Let’s compute this quantity explicitly. Given that the sample value \( x \) falls in the \( i \)th quantisation interval, we have \( Q(x) = S_i \) whence \( N_Q = x - S_i \) so that the average quantisation noise power is given by

\[
P_{N_Q} = \frac{N^2_Q}{2\Delta} \int_{S_i-\Delta}^{S_i+\Delta} (x - S_i)^2 \, dx = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} N^2 \, dN \quad (N \leftarrow x - S_i)
\]

\[
= \frac{1}{2\Delta} \left. \frac{N^3}{3} \right|_{-\Delta}^{\Delta} = \frac{2\Delta^3}{6\Delta} = \frac{A^2}{3M^2}.
\]

In this setting, the signal power \( x^2 \) has a peak value of \( A^2 \) and an average value \( P_s \) which is typically of the form \( cA^2 \) for some constant \( c \) which is determined by the statistical distribution of the signal amplitudes across the range \(-A \leq x \leq A\). Consider, for instance, the sampling of a sinusoid \( x(t) = A \cos(2\pi f_0 t + \theta) \), of unknown phase \( \theta \) and frequency \( f_0 = 1/T \), at a rate much higher than the frequency of the sinusoid. The elementary trigonometric identity \( \cos^2 \theta = \frac{1}{2} (1 + \cos 2\theta) \) shows that the average power evidenced in a sample is then given by

\[
P_s = \frac{1}{T} \int_{-T/2}^{T/2} x^2(t) \, dt = \frac{1}{T} \int_{-T/2}^{T/2} A^2 \cos^2(2\pi f_0 t + \theta) \, dt \\
= \frac{A^2}{2T} \int_{-T/2}^{T/2} dt + \frac{A^2}{2T} \int_{-T/2}^{T/2} \cos(4\pi f_0 t + 2\theta) \, dt \\
= \frac{A^2}{2T} t \bigg|_{-T/2}^{T/2} + \frac{A^2}{2T} \sin(4\pi f_0 t + 2\theta) \bigg|_{-T/2}^{T/2} \\
= \frac{A^2}{2}.
\]

This is just the familiar result that the root-mean-square value of a sinusoid of amplitude \( A \) is \( A^2/2 \).

A measure of the relative impact of the distortion inherited under quantisation is given by the ratio of the average signal power \( P_s \) to the average (quantisation) noise power \( P_{N_Q} \), i.e., the signal-to-noise ratio \( \text{SNR} = P_s/P_{N_Q} \). For the case of a sinusoidal signal and a fine quantisation (large number of quantisation levels) we hence obtain

\[
\text{SNR} = \frac{P_s}{P_{N_Q}} = \frac{A^2/2}{A^2/3M^2} = \frac{3M^2}{2}.
\]
Equivalently, in terms of decibels, the signal-to-noise ratio becomes

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \frac{P_s}{P_N} = 10 \log_{10} \frac{3M^2}{2} = 20 \log_{10} M + 10 \log_{10} \frac{3}{2} \approx 20 \log_{10} M + 1.76 \text{ dB}.$$ 

The constant offset factor of 1.76 dB is an artifact specific to the case of a sinusoidal signal and reflects how large the mean signal power is with respect to the peak signal power. Things don't change materially for a general signal: for a wide range of signals, the mean signal power is a constant fraction of the peak signal power, $P_s = cA^2$ for some constant $c$. If, for instance, the signal is uniformly distributed between $-A \leq x \leq A$ then the mean signal power is $P_s = A^2/3$ (why? ruminate over (+) for a moment) so that $c = 1/3$ in this case. For a sinusoidal signal on the other hand, $P_s = A^2/2$, as we just saw, so that $c = 1/2$.

In general, then, when the quantisation is sufficiently fine the signal-to-noise ratio (in dB) is of the form

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \frac{cA^2}{A^2/3M^2} = 20 \log_{10} M + 10 \log_{10} 3c \text{ dB}$$

where $c$ is some fixed positive constant. The impact of increasing the number of levels on the efficacy of quantisation is now evident:

**Slogan.** Doubling the number of quantisation levels occasions a 6 dB improvement in the signal-to-quantisation noise ratio

Equivalently, each of the $M$ quantisation levels can be represented unambiguously by a distinct sequence of $n = \lceil \log_2 M \rceil$ bits where $\lceil x \rceil$ denotes the smallest integer bigger than or equal to $x$. Again ignoring integer round-off considerations, $M = 2^n$ so that the signal-to-quantisation noise ratio (in dB) may be written in the form

$$\text{SNR}_{\text{dB}} = 20 \log_{10} 2^n + 10 \log_{10} 3c = 20n \log_{10} 2 + 10 \log_{10} 3c \approx 6n + 10 \log_{10} 3c \text{ dB}.$$ 

Which leads us to:

**Slogan, alternative version.** Each extra bit of quantisation improves the signal-to-quantisation noise ratio by 6 dB.

This, however, is not the end of the story. In addition to the number of levels, the designer also has available as a design parameter, at least in principle, the distribution of these levels. Is there any advantage to departing from the intuitive and easily implementable regular, uniformly spaced quantisation strategy?
Nonuniform Quantisation

Uniform quantisation cuts up the range of possible sample values into a set of $M$ equispaced levels. Two considerations suggest, however, that it may be advantageous to consider nonuniform quantisation strategies.

Uniform quantisation discriminates relatively against low signal amplitudes. What does this mean? For any quantisation level $S_i$, the mean distortion $\overline{N_Q(x)} = x - S_i$ of signals falling in the interval $S_i - \Delta < x \leq S_i + \Delta$ corresponding to that level is surely the same for each of the uniformly spaced quantisation levels. However, the relative mean distortion $(x - S_i)/x$ of signals falling in the $i$th quantisation interval depends on $i$ and indeed, this relative distortion is much higher at low amplitude levels. If it is important to keep the relative distortion small in an application then more quantisation levels will have to be expended at small amplitudes vis-à-vis large amplitudes.

Signal values may not be uniformly spread across their range. In practice, signal values may spend a disproportionately large part of time in a relatively small part of their range with rare excursions outside. Expending more quantisation levels in more "populated" subranges would then tend to substantially improve the quantisation error most of the time. Admittedly, the occasional rare signal excursion into a less populated subrange will be treated shabbily by the procedure as relatively few quantisation levels will be expended there. However, with proper allocation of levels, the rare times when a large signal excursion into a sparsely populated area causes a large quantisation error will be more than compensated for by the much increased precision in signal representation most of the time when the signal stays in a highly populated area.

These considerations prompt a weaker slogan.

**Slogan.** Quantisation should be finer in regions where signal values are more likely to occur and coarser in regions where signal excursions are rare.

If signal statistics are known, one can take a somewhat more rigourous, if computationally intensive, tack to the simple idea espoused in the slogan: one formally seeks to optimise the selection of quantisation levels $S_0, \ldots, S_{M-1}$ so as to maximise the signal-to-noise ratio. The resulting optimal nonlinear quantiser is called the Lloyd-Max quantiser.

Of course, the implementation complexity of nonuniform quantisers is much more severe than that of uniform quantisers. In many applications, however, the resulting improvements in performance are so dramatic that the extra cost may be cheerfuly borne.
In some cases, if the nonuniformity in the quantisation is regular enough, the complexity of implementation can be reduced somewhat. The idea is to first pass the signal value through a nonlinear preprocessor \( y = F(x) \) whose function is to expand ranges where the signal spends more time and compress ranges where the signal spends less time—the preprocessor is frequently just called a compressor—followed by a uniform quantiser with equispaced quantisation levels. By expanding ranges where the signal is more likely to be these ranges encounter more quantisation levels in the uniform quantiser than they ordinarily would; likewise, by compressing less populated areas of the signal range into smaller regions, these regions likewise see fewer quantisation levels of a uniform quantiser than they would otherwise. An example of a compressor is shown in the adjoining figure. Observe how a relatively small range around the origin of the \( x \)-axis is expanded into a much larger range along the \( y \)-axis by the function. The range around \( x = 1 \), however, is compressed into a much smaller range along the \( y \)-axis.

To recover the original quantised levels the receiver simply reverses the process by computing the compressor inverse function \( F^{-1}(y) \), a process naturally called the expander. The combination of the compressor \( F(x) \) and the expander \( F^{-1}(y) \) is called a compander.

A good example of the efficacy of nonuniform quantisers is seen in the quantisation of speech samples. Low amplitudes are the norm in samples taken from speech waveforms with high amplitude excursions being rather rare. A nonuniform quantiser which partitions lower amplitudes much more finely than higher amplitudes is hence indicated. A compression law such as the one shown above followed by a uniform quantiser will do the trick. Indeed, this is the basis for the commercial nonuniform quantisation of speech samples in pulse code modulated (PCM) systems. In North America and Japan the compression law used is known as the \( \mu \)-law and implements a function of the form

\[
y = F(x) = \frac{\log(1 + \mu x)}{\log(1 + \mu)},
\]

while the expander implements the function

\[
x = F^{-1}(y) = \frac{1}{\mu} \left((1 + \mu)^y - 1\right).
\]

In these expressions \( x \) represents the sample value normalised to take values in the range \( 0 \leq x \leq 1 \) and \( y \) is the normalised output of the compressor. The nonnegative parameter \( \mu \) controls the degree of compression:
the compression law varies from a linear form $y = x$ when $\mu = 0$ to a highly nonlinear form which magnifies small amplitudes and compresses large amplitudes when $\mu$ is large. A family of $\mu$-law curves is shown alongside. The value $\mu = 255$ is used in commercial North American companding systems.

A similar companding system implemented in Europe uses a compression function of the form

$$y = F(x) = \begin{cases} \frac{Ax}{1+\log A} & \text{if } 0 \leq x \leq \frac{1}{A}, \\ 1 + \log(Ax) & \text{if } \frac{1}{A} \leq x \leq 1. \end{cases}$$

As before, the input signal values have been normalised to take values between 0 and 1. The compression function is commonly called the $A$-law. Observe that when $A = 1$ we just have uniform quantisation (no compression); successively higher values of $A$ increase the level of nonlinearity, hence the compression.

Commercial speech quantisers built on such principles can achieve signal-to-noise ratio improvements of 24 to 30 dB over uniform quantisers. Well worth the extra cost involved in implementing the companding operation.