The Sampling Theorem and the Bandpass Theorem

by D.S.G. Pollock
University of Leicester
Email: stephen_pollock@sigmapi.u-net.com

The Shannon–Nyquist Sampling Theorem

According to the Shannon–Whittaker sampling theorem, any square integrable piecewise continuous function $x(t) \longleftrightarrow \xi(\omega)$ that is band-limited in the frequency domain, with $\xi(\omega) = 0$ for $|\omega| > \pi$, has the series expansion

(1) $$x(t) = \sum_{k=-\infty}^{\infty} x_k \frac{\sin\{\pi(t-k)\}}{\pi(t-k)} = \sum_{k=-\infty}^{\infty} x_k \psi_{(0)}(t-k),$$

where $x_k = x(k)$ is the value of the function $x(t)$ at the point $t = k$. It follow that the continuous function $x(t)$ can be reconstituted from its sampled values $\{x_t, t \in \mathcal{I}\}$.

Proof. Since $x(t)$ is a square-integrable function, it is amenable to a Fourier integral transform, which gives

(2) $$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \xi(\omega)e^{\mathrm{i}\omega t}d\omega, \quad \text{where} \quad \xi(\omega) = \int_{-\infty}^{\infty} x(t)e^{-\mathrm{i}\omega t}dt.$$

But $\xi(\omega)$ is a continuous function defined of the interval $(-\pi, \pi]$ that may also be regarded as a periodic function of a period of 2π. Therefore, $\xi(\omega)$ corresponds to a discrete aperiodic function in the time domain—which is to say that the relationship $x(t) \longleftrightarrow \xi(\omega)$ entails the discrete-time Fourier transform—and $\xi(\omega)$ may be expanded as

(3) $$\xi(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{-\mathrm{i}k\omega}, \quad \text{where} \quad c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \xi(\omega)e^{\mathrm{i}k\omega}d\omega.$$

By comparing (2) with (3), we see that the coefficients c_k are simply the ordinates of the function $x(t)$ sampled at the integer points; and we may write them as

(4) $$c_k = x_k = x(k).$$

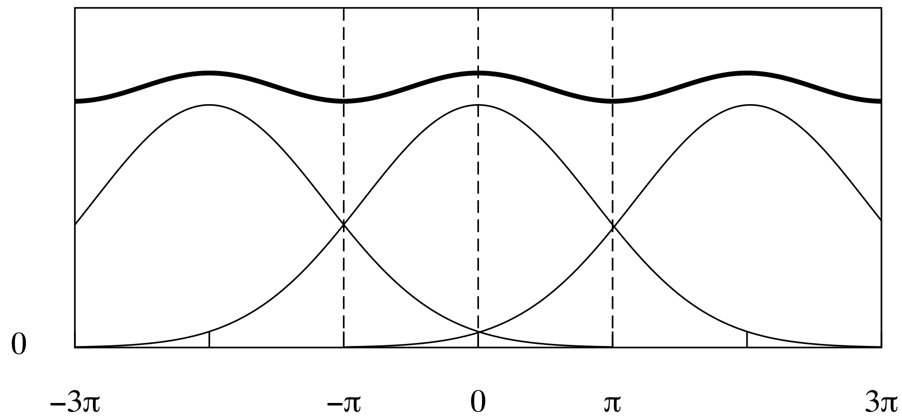

Figure 1. The figure illustrates the aliasing effect of regular sampling. The bell-shaped function supported on the interval $[-3\pi, 3\pi]$ is the spectrum of a continuous-time process. The spectrum of the sampled process, represented by the heavy line, is a periodic function of period 2π.

Next, we must show how the continuous function $x(t)$ may be reconstituted from its sampled values. Using (4) in (3) gives

$$\xi(\omega) = \sum_{k=-\infty}^{\infty} x_k e^{-\mathrm{i}k\omega}. \tag{5}$$

Putting this in (2), and taking the integral over $(-\pi, \pi]$ in consequence of the band-limited nature of the function $x(t)$, gives

$$x(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\sum_{k=-\infty}^{\infty} x_k e^{-\mathrm{i}k\omega}\right\} e^{\mathrm{i}\omega t} d\omega = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty} x_k \int_{-\pi}^{\pi} e^{\mathrm{i}\omega(t-k)} d\omega. \tag{6}$$

The integral on the RHS is evaluated as

$$\int_{-\pi}^{\pi} e^{\mathrm{i}\omega(t-k)} d\omega = 2\frac{\sin\{\pi(t-k)\}}{t-k}. \tag{7}$$

Putting this into the RHS of (5) gives the result of (1).

Imaging and Aliasing

Let $\xi_s(\omega)$ be the transform of the sampled sequence $\{x_t; t = 0, \pm 1, \pm 2, \ldots\}$. Then, at an integer point t, there is $x_t = x(t)$ and, therefore,

$$x_t = \frac{1}{2\pi}\int_{-\infty}^{\infty} \xi(\omega) e^{\mathrm{i}\omega t} d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \xi_s(\omega) e^{\mathrm{i}\omega t} d\omega. \tag{8}$$

The equation of the two integrals implies that

$$\xi_s(\omega) = \sum_{j=-\infty}^{\infty} \xi(\omega + 2j\pi). \tag{9}$$

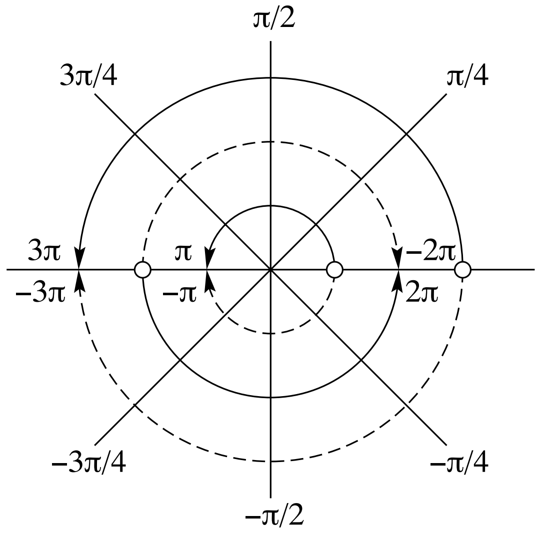

Figure 2. A diagram to illustrate the aliasing of frequencies when the Nyquist frequency is at π radians per sample interval. The arcs with the broken lines correspond to negative frequencies.

Thus, the periodic function $\xi_s(\omega)$ is obtained by wrapping $\xi(\omega)$ around a circle of circumference of 2π and adding the coincident ordinates. Alternatively, the periodic extension of $\xi_s(\omega)$ can be envisaged as the consequence of overlaying repeated copies of the function $\xi(\omega)$, with each copy shifted an integral multiple of 2π, which is the sampling frequency. This is illustrated in Figure 1. The creation of succesively displaced copies of $\xi(\omega)$ is commonly described as a process of imaging.

Unless $\xi(\omega)$ is band limited to the Nyquist frequency interval $[-\pi, \pi]$, the effect of wrapping and overlaying will be to create sample spectrum that differs from and which misrepresents the spectrum of the underlying continuous signal.

The elements of the signal that lie outside the Nyquist range will be misrepresented by elements that do lie within the range and which are described as their aliases. Thus, in the process of sampling, all frequencies will be mapped into the interval $[-\pi, \pi]$ according to a conversion described by

$$(10) \qquad \omega \longrightarrow \omega' = \begin{cases} \{(\omega + \pi) \bmod 2\pi\} - \pi, & \text{if } \omega > 0; \\ \{(\omega - \pi) \bmod 2\pi\} + \pi, & \text{if } \omega < 0. \end{cases}$$

This conversion can be illustrated by Figure 2, which show the effects of wrapping.

In the diagram, the points on the inner circle correspond to frequency values within the Nyquist interval $[-\pi, \pi]$. Those on the outer circles correspond to frequencies in the intervals $[-2\pi, -\pi] \cup [\pi, 2\pi]$ and $[-3\pi, -2\pi] \cup [2\pi, 3\pi]$ respectively. The values of the aliased frequencies ω' are to be found at the points where the radii that pass through the points ω on the outer circles intersect with the inner circle. Further concentric circles can be added to the diagram to accomodate higher frequencies.

Sampling at an Arbitrary Rate

The sampling theorem shows that a band-limited continuous signal can be perfectly reconstructed from a sequence of samples if the highest frequency of the signal does not exceed half the rate of sampling.

In the statement of the theorem, the sampling interval has been taken as fixed and it is defined to be the unit interval. It has been revealed that the highest detectable frequency in the sampled data is the Nyquist frequency of π radians per interval.

An alternative statement is appropriate if it is the maximum frequency in the continuous signal that is fixed and if it is required to determine the minimum rate of sampling necessary for capturing all of the information therein. This is a circumstance that usually prevails in communications engineering; and it leads to an alternative presentation of the sampling theorem.

Imagine that the maximum frequency of the signal $\omega_c = 2\pi W$, where W is a number of hertz or cycles per second. Then, to capture the information, the sampling must be at a rate of no less than $2W$, which implies a sampling interval of $1/(2W)$ seconds. In this case, the signal is represented by

$$x(t) = \frac{1}{2\pi}\int_{-\omega_c}^{\omega_c} \xi(\omega)e^{\mathrm{i}\omega t}d\omega \tag{11}$$

and the generic sampled value, taken at intervals of $1/(2W)$ seconds and indexed by $k \in \mathcal{I}$ is

$$x_k = x\left(\frac{k}{2W}\right) = \frac{1}{2\pi}\int_{-2\pi W}^{2\pi W} \xi(\omega)e^{\mathrm{i}\omega k/(2W)}d\omega. \tag{12}$$

Then, the continuous signal can be represented by

$$x(t) = \sum_{k=-\infty}^{\infty} x_k \frac{\sin\{\pi(2Wt-k)\}}{\pi(2Wt-k)} = \sum_{k=-\infty}^{\infty} x_k \frac{\sin(\omega_c t - k\pi)}{\omega_c t - k\pi}. \tag{13}$$

Impulses and Impulse Trains

An alternative proof of the sampling theorem is available which is based on the idea that a sampled sequence can be generated by modulating a continuous signal by an impulse train.

An impulse in continuous time, located at the point $t = 0$, is a generalised function for which

$$\delta(t) = 0 \quad \text{for all} \quad t \neq 0 \qquad \text{and} \qquad \int_{-\infty}^{\infty} \delta(t) = 1. \tag{14}$$

An essential property of this so-called Dirac delta function is the *sifting property* whereby

$$(15) \qquad f(\tau) = \int_{-\infty}^{\infty} f(t)\delta(t-\tau)dt.$$

The Fourier transform of the Dirac function is given by

$$(16) \qquad \delta(t-\tau) \longleftrightarrow e^{-\mathrm{i}\omega\tau} = \int_{-\infty}^{\infty} e^{-\mathrm{i}\omega t}\delta(t-\tau)dt.$$

When $\tau = 0$, this becomes a constant function that is dispersed over the entire line, which shows that every frequency is needed in order to synthesise the impulse.

It is also possible to define a Dirac function in the frequency domain as a single impulse located at $\omega = \omega_0$ with an area of 2π:

$$(17) \qquad 2\pi\delta(\omega-\omega_0) \longleftrightarrow e^{\mathrm{i}\omega_1\tau} = \frac{1}{2\pi}\int_{-\infty}^{\infty} 2\pi\delta(\omega-\omega_0)e^{\mathrm{i}\omega t}d\omega.$$

In describing the periodic sampling of a continuous-time signal, it is useful consider a train of impulses separated by a time period of T. (Here, we are using T to denote the length of time between the sample elements, as opposed to the number of elements within a finite sample; and we are free to normalise this length by setting $T = 1$.) The impulse train is represented by the function

$$(18) \qquad g(t) = \sum_{j=-\infty}^{\infty} \delta(t - jT),$$

which is both periodic and discrete. The periodic nature of this function indicates that it can be expanded as a Fourier series

$$(19) \qquad g(t) = \sum_{j-\infty}^{\infty} \gamma_j e^{i\omega_1 jt}.$$

The coefficients of this expansion may be determined by integrating over just one cycle. Thus

$$(20) \qquad \gamma_j = \frac{1}{T}\int_0^T \delta(t)e^{-i\omega_1 jt}dt = \frac{1}{T},$$

wherein $\omega_1 = 2\pi/T$ represents the fundamental frequency. On setting $\gamma_j = T^{-1}$ for all j in the Fourier-series expression for $g(t)$ and invoking the result under (17), it is found that the Fourier transform of the continuous-time impulse train $g(t)$ is the function

$$\begin{aligned}\gamma(\omega) &= \frac{2\pi}{T}\sum_{j=-\infty}^{\infty}\delta\Big(\omega - j\frac{2\pi}{T}\Big)\\ &= \omega_1\sum_{j=-\infty}^{\infty}\delta\big(\omega - j\omega_1\big).\end{aligned} \tag{21}$$

Thus it transpires that a periodic impulse train $g(t)$ in the time domain corresponds to a periodic impulse train $\gamma(\omega)$ in the frequency domain. Notice that there is an inverse relationship between the length T of the sampling interval in the time domain and the length $2\pi/T$ of the corresponding interval between the frequency-domain pulses.

An Alternative Proof of the Sampling Theorem

The mathematical representation of the sampling process depends upon the periodic impulse train or sampling function $g(t)$ defined under (18). The period T is the sampling interval, whilst the fundamental frequency of this function, which is $\omega_1 = 2\pi/T$, is the sampling frequency.

The activity of sampling may be depicted as a process of amplitude modulation wherein the impulse train $g(t)$ is the carrier signal and the sampled function $x(t)$ is the modulating signal. In the time domain, the modulated signal is described by the following *multiplication* of $g(t)$ and $x(t)$:

$$\begin{aligned}x_s(t) &= x(t)g(t)\\ &= \sum_{j=-\infty}^{\infty} x(t)\delta(t - jT).\end{aligned} \tag{22}$$

In most cases, one should be free to set $T = 1$, which is to say that the sample interval can be regarded as a unit in time. Then, it is worthwhile to observe that, unless we replace $x(t)$ by x_t, there is no distinction in notation between the case, in continuous time, of a function modulated by a train of Dirac impulses and the case, in discrete time, of a sequence of elements indexed by $t \in \mathcal{I} = \{0, \pm 1, \pm 2, \ldots\}$, each multplied, redundantly, by a unit impulse.

The Fourier transform $\xi_s(\omega)$ of $x_s(t)$ is the *convolution* of the transforms of $x(t)$ and $g(t)$, which are denoted by $\xi(\omega)$ and $\gamma(\omega)$ respectively. Thus,

$$\begin{aligned}\xi_s(\omega) &= \int_{-\infty}^{\infty} x_s(t)e^{-\mathrm{i}\omega t}dt\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\gamma(\lambda)\xi(\omega - \lambda)d\lambda.\end{aligned} \tag{23}$$

Substituting the expression for $\gamma(\lambda)$ from (20), gives

$$\begin{aligned}\xi_s(\omega) &= \frac{\omega_1}{2\pi}\int_{-\infty}^{\infty}\xi(\omega-\lambda)\left\{\sum_{j=-\infty}^{\infty}\delta\big(\lambda-j\omega_1\big)\right\}d\lambda \\ &= \frac{1}{T}\sum_{j=-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}\xi(\omega-\lambda)\delta\big(\lambda-j\omega_1\big)d\lambda\right\} \\ &= \frac{1}{T}\sum_{j=-\infty}^{\infty}\xi\big(\omega-j\omega_1\big).\end{aligned} \tag{24}$$

The final expression indicates that $\xi_s(\omega)$, which is the Fourier transform of the sampled signal $x_s(t)$, is a periodic function consisting repeated copies of the transform $\xi(\omega)$ of the original continuous-time signal $x(t)$. Each copy is shifted by an integral multiple of the sampling frequency $\omega_1 = 2\pi/T$ before being superimposed. Observe that equation (9), which was the previous expression of the result, is obtained is by setting $T = 1$.

A more explicit derivation of the result is obtained by setting $g(t) = T^{-1}\sum_j e^{\mathrm{i}\omega_1 jt}$ within $x_s(t) = x(t)g(t)$ to give

$$\begin{aligned}\xi_s(\omega) &= \int_{-\infty}^{\infty}x(t)g(t)e^{-\mathrm{i}\omega t}dt \\ &= \frac{1}{T}\sum_{j=-\infty}^{\infty}\int_{-\infty}^{\infty}x(t)e^{-\mathrm{i}(\omega-\omega_1 j)t}dt\end{aligned} \tag{25}$$

Imagine that $x(t)$ is a band-limited signal whose frequency components are confined to the interval $[0,\omega_c]$, which is to say that the function $\xi(\omega)$ is nonzero only over the interval $[-\omega_c,\omega_c]$. If

$$\frac{2\pi}{T} = \omega_1 > 2\omega_c, \tag{26}$$

then the successive copies of $\xi(\omega)$ will not overlap; and therefore the properties of $\xi(\omega)$, and hence those of $x(t)$, can be deduced from those displayed by $\xi_s(\omega)$ over the interval $[0,\omega_1]$. In principle, the original signal could be recovered by passing its sampled version through an ideal lowpass filter which transmits all components of frequency less that ω_1 and rejects all others.

If, on the contrary, the sampling frequency is such that $\omega_1 < 2\omega_c$, then the resulting overlapping of the copies of $\xi(\omega)$ will imply that the spectrum of the sampled signal is no longer simply related to that of the original; and no linear filtering operation can be expected to recover the original signal from its sampled version. The effect of the overlap is to confound the components of the original process which have frequencies greater that π/T with those of frequencies lower than π/T; and this is described as the aliasing error.

The foregoing results are expressed in the famous sampling theorem which is summarised as follows:

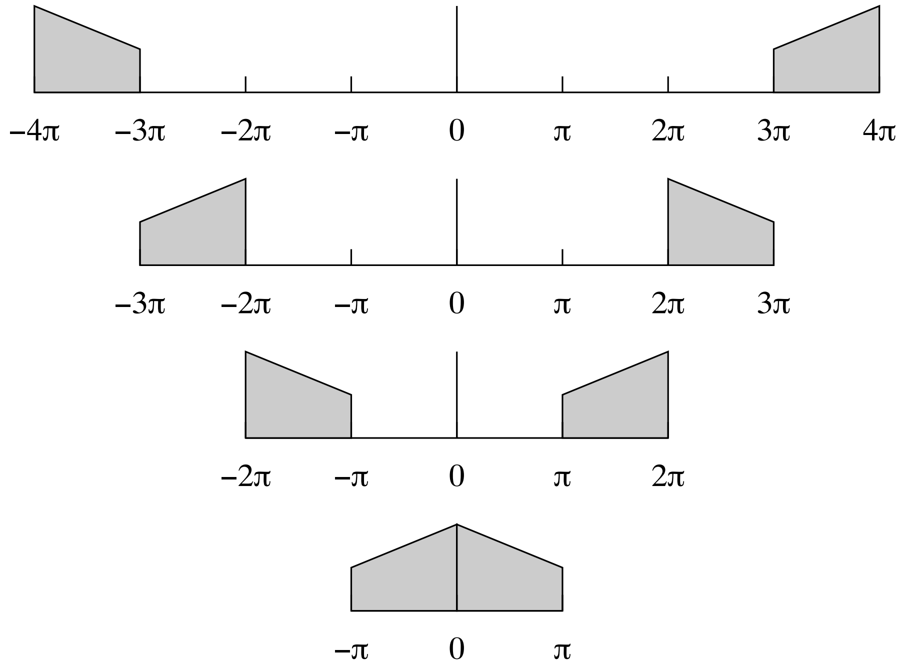

Figure 3. A diagram showing the spectra of four continuous processes which produce the same image when sampled at the rate of 2π.

(27) Let $x(t)$ be a continuous-time signal with a transform $\xi(\omega)$ which is zero-valued for all $\omega > \omega_c$. Then $x(t)$ can be recovered from its samples provided that the sampling rate $\omega_1 = 2\pi/T$ exceeds $2\omega_c$.

An alternative way of expressing this result is to declare that the rate of sampling sets an upper limit to the frequencies which can be detected in an underlying process. Thus, when the sampling provides one observation in T seconds, the highest frequency which can be detected has a value of π/T radians per second. This is the so-called Nyquist frequency.

The Bandpass Sampling Theory

In some cases, a signal is supported on the frequency intervals $(-f_L, -f_U)$ and (f_U, f_L), with $f_U \neq 0$. Then, it may be possible to capture all of the information in the signal by sampling it at a rate that is significantly lower than $2f_U$, which is the rate that is indicated by the classical Shannon–Nyquist sampling theorem.

To understand this possibility, one should consider Figure 3, which depicts four spectral structures, on the intervals $(-n\pi, [1-n]\pi) \cup ([n-1]\pi, n\pi); n = 1, 2, 3, 4$. When sampled at the Nyquist rate of 2π, the four processes generate identical sequences. This rate of sampling rate is appropriate to a signal that is supported on the interval $(-\pi, \pi)$, which is described as the base band. Provided that the frequency location of the true spectrum is known, full information on the underlying signal, which will permit its reconstruction, will be obtained by sampling at the rate of 2π.

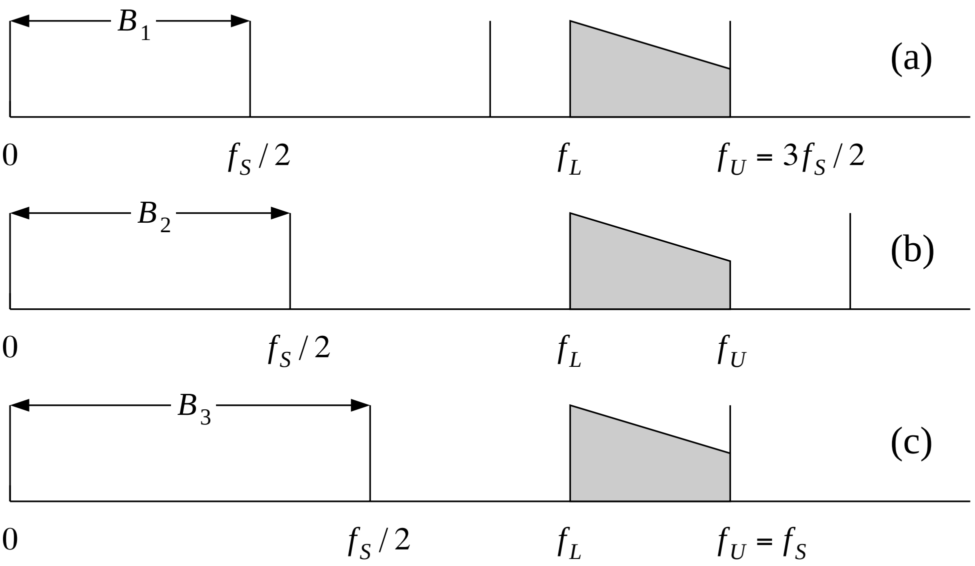

Figure 4. The sampling rate increases gradually from $f_S = 2B_1$, depicted in (a), to $f_S = 2B_2$, depicted in (b). Then, it jumps by $2\Delta = f_U - f_L$ to become $f_S = 2B_3$, depicted in (c).

Although, ostensibly, Figure 3 portrays the specta of four distinct signals, these might be construed as components of a single signal that have been extracted by a process of filtering. (The diagram should then be modified to indicate that the limiting frequency of the overall signal is π.) The bandwith of this overall signal is equally divided among is components. A conclusion to be drawn from the diagram is that, in these circumstances, the minimum rate of sampling is proportional to the bandwidth of the component signals. It is unrelated to the locations of their bands, which determine the frequencies of the constituent elements.

In general, in order to exploit the possibilities of bandpass sampling, the requirement is to determine a base band $(-B, B)$, measured in hertz, such that $(f_L, f_U) \in ([n-1]B, nB)$, where $n \in \{0, 1, 2 \ldots\}$ has an integer value. The condition that (f_L, f_U) lies in such an interval implies that

$$f_U \leq nB, \quad f_L \geq (n-1)B \quad \text{with} \quad 1 \geq n \geq [f/(f_U - f_L)],$$

where $[f_U/W]$ denotes the integer quotient of the division of f_U by W. The condition on the rate of sampling $f_S = 2B$ can be written concisely as

$$\frac{2f_U}{n} \leq f_S \leq \frac{2f_L}{n-1}. \tag{28}$$

The sampling rates that fulfil this condition vary in a discontinuous manner. In Figure 4, the values of f_U and f_L are to be regarded as fixed, while the width of the baseband or, equivalently, the sampling rate, increases from one tranche to the next.

The figure illustrates the discontinuity that occurs with an increasing sampling rate when the spectral structure on the interval (f_U, f_L) is crossed by the

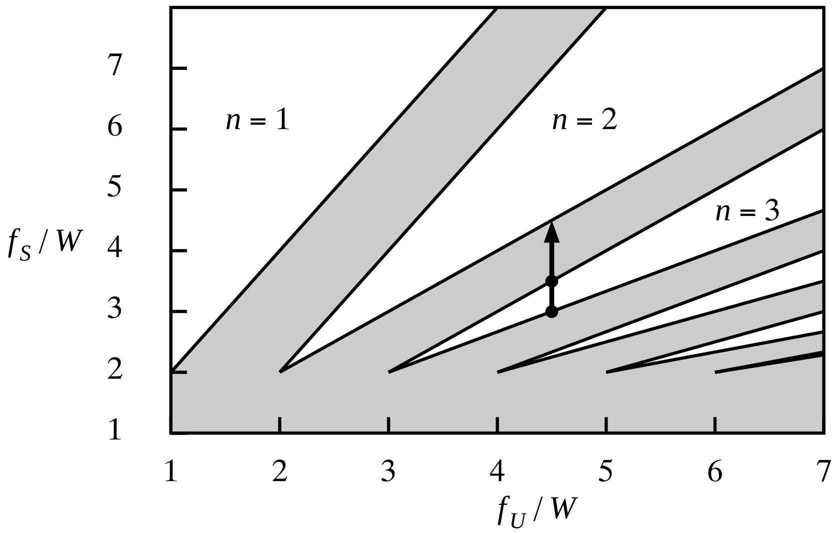

Figure 5. The band positions associated with various sampling rates. The vertical axis measures the sampling rate normalised by the bandwidth $W = f_U - f_L$. The horizontal axis measures the upper limit of the sampling band normalised by W. The base of the vertical arrow corresponds to position (a) of Figure 4, the point in the middle corresponds to position (b) and the tip corresponds to position (c).

lower bound of the band $([n-1]B_j, nB_j)$ which, up to this point, has been increasing continuously in width. The discontinuity occurs in the transition from (b) to (c).

At that point, the width of the band increases abruptly such that f_U becomes adjacent to the upper bound of the new band. If the width of the band before the jump was B_j, then the width after the jump will be $B_{j+1} = B_j + \Delta$, with $\Delta = (f_U - f_L)/(n-1)$. Figure 5 shows the allowed rates of sampling, which correspond to the white regions, and the disallowed rates of sampling, which correspond to the shaded regions. The discrete jumps in the sample rates correspond to the vertical distances within the shaded bands.