Exercise Problems: Information Theory and Coding

Exercise 9

1. An error-correcting Hamming code uses a 7 bit block size in order to guarantee the detection, and hence the correction, of any single bit error in a 7 bit block. How many bits are used for error correction, and how many bits for useful data? If the probability of a single bit error within a block of 7 bits is $p = 0.001$, what is the probability of an error correction failure, and what event would cause this?

2. What class of continuous signals has the greatest possible entropy for a given variance (or power level)? What probability density function describes the excursions taken by such signals from their mean value?

3. What does the Fourier power spectrum of this class of signals look like? How would you describe the entropy of this distribution of spectral energy?

4. Suppose that a continuous communication channel of bandwidth W Hertz and a high signal-to-noise ratio, which is perturbed by additive white Gaussian noise of constant power spectral density, has a channel capacity of C bits per second. Approximately how much would C be degraded if suddenly the added noise power became 8 times greater?

Solution:

1. An error-correcting Hamming code with a 7 bit block size uses <u>3 bits</u> for error correction and <u>4 bits</u> for data transmission. It would fail to correct errors that affected more than one bit in a block of 7; but in the example given, with $p = 0.001$ for a single bit error in a block of 7, the probability of two bits being corrupted in a block would be about 1 in a million.

2. The family of continuous signals having maximum entropy per variance (or power level) are Gaussian signals. Their probability density function for excursions x around a mean value μ, when the power level (or variance) is σ^2, is:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

3. The Fourier power spectrum of this class of signals is flat, or white. Hence these signals correspond to "white noise." The distribution of spectral energy has uniform probability over all possible frequencies, and therefore this continuous distribution has maximum entropy.

4. The channel capacity C in bits per second would be reduced by about <u>$3W$</u>, where W is the channel's bandwidth in Hertz, if the noise power level increased eight-fold. This is because the channel capacity, in bits per second, is

$$C = W \log_2 \left(1 + \frac{P}{N_0 W}\right)$$

If the signal-to-noise ratio (the term inside the logarithm) were degraded by a factor of 8, then its logarithm is reduced by -3, and so the overall capacity C is reduced by $3W$. The new channel capacity C' could be expressed either as:

$$C' = C - 3W$$

or as a ratio that compares it with the original undegraded capacity C:

$$\frac{C'}{C} = 1 - \frac{3W}{C}$$

Exercise 10

1. Consider a noiseless analog communication channel whose bandwidth is 10,000 Hertz. A signal of duration 1 second is received over such a channel. We wish to represent this continuous signal exactly, at all points in its one-second duration, using just a finite list of real numbers obtained by sampling the values of the signal at discrete, periodic points in time. What is the length of the shortest list of such discrete samples required in order to guarantee that we capture all of the information in the signal and can recover it exactly from this list of samples?

2. Name, define algebraically, and sketch a plot of the function you would need to use in order to recover completely the continuous signal transmitted, using just such a finite list of discrete periodic samples of it.

3. Explain why smoothing a signal, by low-pass filtering it *before* sampling it, can prevent aliasing. Explain aliasing by a picture in the Fourier domain, and also show in the picture how smoothing solves the problem. What would be the most effective low-pass filter to use for this purpose? Draw its spectral sensitivity.

4. Consider a noisy analog communication channel of bandwidth Ω, which is perturbed by additive white Gaussian noise whose power spectral density is N_0. Continuous signals are transmitted across such a channel, with average transmitted power P (defined by their expected variance). What is the channel capacity, in bits per second, of such a channel?

5. If a continuous signal $f(t)$ is *modulated* by multiplying it with a complex exponential wave $\exp(i\omega t)$ whose frequency is ω, what happens to the Fourier spectrum of the signal?

   Name a very important practical application of this principle, and explain why modulation is a useful operation.

   How can the original Fourier spectrum later be recovered?

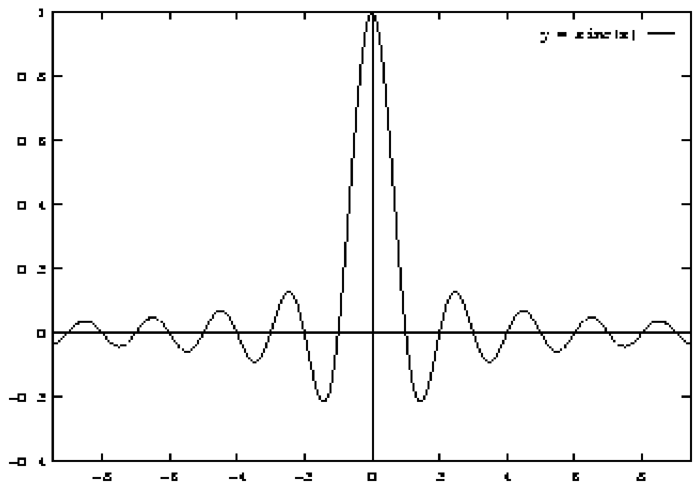
6. Which part of the 2D Fourier Transform of an image, the amplitude spectrum or the phase spectrum, is indispensable in order for the image to be intelligible?

   Describe a demonstration that proves this.

Solution:

1. $2\omega \mathrm{T} = \underline{20{,}000}$ discrete samples are required.

2. The <u>sinc</u> function is required to recover the signal from its discrete samples, defined as: $\mathrm{sinc}(x) = \dfrac{\sin(\pi x)}{\pi x}$. Each sample point is replaced by scaled copies of this function.



3. The Nyquist Sampling Theorem tells us that aliasing results when the signal contains Fourier components higher than one-half the sampling frequency. Thus aliasing can be avoided by removing such frequency components from the signal, by low-pass filtering it, before sampling the signal. The ideal low-pass filter for this task would have a strict cut-off at frequencies starting at (and higher than) one-half the planned sampling rate.

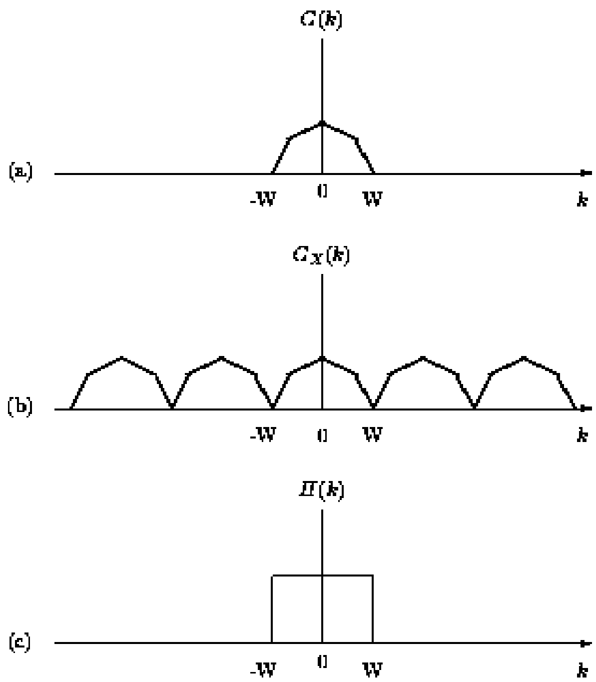
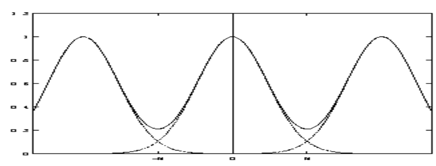


Figure 5: Band-limited signals: (a) Spectrum of $g(x)$. (b) Spectrum of $g_X(x)$. (c) Ideal filter response for reconstruction



4. The channel capacity is $\Omega \log_2 \left(1 + \frac{P}{N_0 \Omega}\right)$ bits per second.

5. Modulation of the continuous signal by a complex exponential wave $\exp(i\omega t)$ will shift its entire frequency spectrum upwards by an amount ω.

   All of AM broadcasting is based on this principle. It allows many different communications channels to be multi-plexed into a single medium, like the electromagnetic spectrum, by shifting different signals up into separate frequency bands.

   The original Fourier spectrum of each of these signals can then be recovered by demodulating them down (this removes each AM carrier). This is equivalent to multiplying the transmitted signal by the conjugate complex exponential, $\exp(-i\omega t)$.

6. The phase spectrum is the indispensable part. This is demonstrated by crossing the amplitude spectrum of one image with the phase spectrum of another one, and *vice versa*. The new image that you see looks like the one whose phase spectrum you are using, and not at all like the one whose amplitude spectrum you've got.

Exercise 11

The signal-to-noise ratio SNR of a continuous communication channel might be different in different parts of its frequency range. For example, the noise might be predominantly high frequency hiss, or low frequency rumble. Explain how the information capacity C of a noisy continuous communication channel, whose available bandwidth spans from frequency ω_1 to ω_2, may be defined in terms of its signal-to-noise ratio as a function of frequency, $\mathrm{SNR}(\omega)$. Define the bit rate for such a channel's information capacity, C, in bits/second, in terms of the $\mathrm{SNR}(\omega)$ function of frequency.

(Note: This question asks you to generalise beyond the material lectured.)

Solution:

The information capacity C of any tiny portion $\Delta\omega$ of this noisy channel's total frequency band, near frequency ω where the signal-to-noise ratio happens to be $\mathrm{SNR}(\omega)$, is:

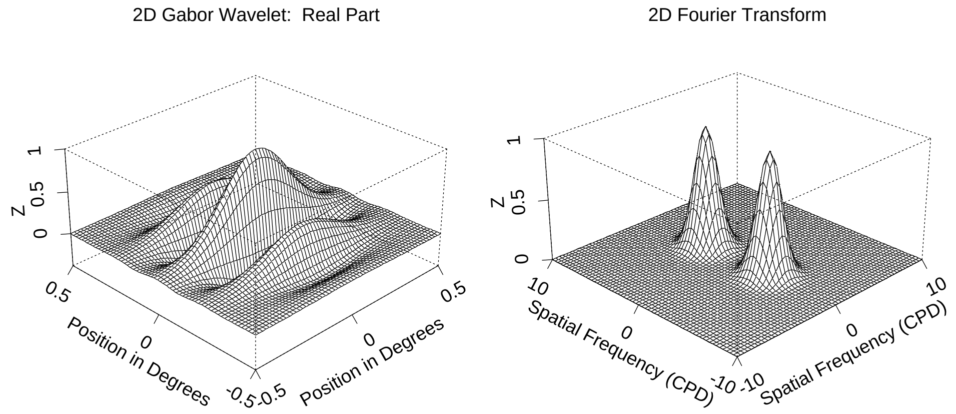
$$C = \Delta\omega \log_2(1 + \mathrm{SNR}(\omega))$$

in bits/second. Integrating over all of these small $\Delta\omega$ bands in the available range from ω_1 to ω_2, the total capacity in bits/second of this variable-SNR channel is therefore:

$$C = \int_{\omega 1}^{\omega 2} \log_2(1 + \mathrm{SNR}(\omega)) d\omega$$

Exercise 12

(a) Explain why the real-part of a 2D Gabor wavelet has a 2D Fourier transform with two peaks, not just one, as shown in the right panel of the Figure below.



(b) Show that the set of all Gabor wavelets is closed under convolution, *i.e.* that the convolution of any two Gabor wavelets is just another Gabor wavelet. [HINT: This property relates to the fact that these wavelets are also closed under multiplication, and that they are also self-Fourier. You may address this question for just 1D wavelets if you wish.]

(c) Show that the family of sinc functions used in the Nyquist Sampling Theorem,

$$\text{sinc}(x) = \frac{\sin(\lambda x)}{\lambda x}$$

is closed under convolution. Show further that when two different sinc functions are convolved, the result is simply whichever one of them had the lower frequency, *i.e.* the smaller λ.

(d) For each of the four classes of signals in the left table below, identify its characteristic spectrum from the right table. ("Continuous" here means supported on the reals, *i.e.* at least piecewise continuous but not necessarily everywhere differentiable. "Periodic" means that under multiples of some finite shift the function remains unchanged.) Give your answer just in the form 1-A, 2-B, etc. Note that you have 24 different possibilities.

Class	*Signal Type*
1.	continuous, aperiodic
2.	continuous, periodic
3.	discrete, aperiodic
4.	discrete, periodic

Class	*Spectral Characteristic*
A.	continuous, aperiodic
B.	continuous, periodic
C.	discrete, aperiodic
D.	discrete, periodic

(e) Define the Kolmogorov algorithmic complexity K of a string of data. What relationship is to be expected between the Kolmogorov complexity K and the Shannon entropy H for a given set of data? Give a reasonable estimate of the Kolmogorov complexity K of a fractal, and explain why it is reasonable.

Solution:

(a) The real-part of a 2D Gabor wavelet, as shown in the figure, has the functional form

$$f(x,y) = e^{-(x^2/\alpha^2 + y^2/\beta^2)} \cos(u_0 x + v_0 y)$$

But using the identity $\cos(\theta) = \frac{1}{2}(e^{i\theta} + e^{-i\theta})$ we see that the expression above is actually twice the sum of two complex Gabor wavelets, namely

$$g(x,y) = e^{-(x^2/\alpha^2 + y^2/\beta^2)} e^{i(u_0 x + v_0 y)}$$

and

$$h(x,y) = e^{-(x^2/\alpha^2 + y^2/\beta^2)} e^{-i(u_0 x + v_0 y)}$$

These each have a 2D Gaussian Fourier transform, one centred at (u_0, v_0) and the other centred at $(-u_0, -v_0)$, added together as seen in the figure.

(b) The Fourier Transform of a 1-D Gabor wavelet has exactly the same functional form, but with the parameters simply interchanged or inverted:

$$F(k) = e^{-(k-k_0)^2 a^2} e^{ix_0(k-k_0)}$$

(In other words, Gabor wavelets are self-Fourier.) It is obvious that the product of any two Gabor wavelets $f(x)$ will still have the functional form of a Gabor wavelet. Therefore the product's Fourier transform will also preserve this general form. Hence (using the convolution theorem of Fourier analysis), it follows that the family of Gabor wavelets are also closed under convolution.

(c) When two functions are convolved together, their Fourier transforms are just multiplied together to give the Fourier transform of the result of the convolution. Conveniently, the Fourier transform of any sinc function as given with frequency parameter λ is a (zero-centred) rectangular pulse function whose width is 2λ and 0 outside. Multiplying two such pulse functions together just creates another pulse function, and therefore the result of the convolution is again just a sinc function. But more specifically, because the product of the two pulse functions in frequency is simply whichever one of them was narrower, the resulting sinc function is just whichever one had the smaller λ (lower frequency).

(d)

1-A. Example: a Gaussian function, whose Fourier transform is also Gaussian.
2-C. Example: a sinusoid, whose Fourier transform is two discrete delta functions.
3-B. Example: a delta function, whose Fourier transform is a complex exponential.
4-D. Example: a comb sampling function, whose Fourier Transform is also a comb function.

(e) The Kolmogorov algorithmic complexity K of a string of data is defined as the length of the shortest binary program that can generate the string. Thus the data's Kolmogorov complexity is its "Minimal Description Length."

The expected relationship between the Kolmogorov complexity K of a set of data, and its Shannon entropy H, is that approximately $K \approx H$.
Because fractals can be generated by extremely short programs, namely iterations of a mapping, such patterns have Kolmogorov complexity of nearly $K \approx 0$.