Dr J Lasenby and Prof N G Kingsbury

Solutions: 4F8 2009

ENGINEERING TRIPOS PART IIB

Monday 27April 2009 9 to 10.30

Module 4F8

IMAGE PROCESSING AND IMAGE CODING

1 (a) Taking the inverse FT of the ideal frequency response will give an impulse response which does not have finite support – to remedy this we multiply by a *window function* which forces the impulse response coefficients to zero for (n_1, n_2) outside R_h, the desired support region. The actual filter frequency response $H(\omega_1, \omega_2)$ is then given by the **convolution** of the desired frequency response $H_d(\omega_1, \omega_2)$ with the window function spectrum $W(\omega_1, \omega_2)$.

This is exactly as we should expect since we multiply in the spatial domain and must therefore convolve in the frequency domain.

Thus the effect of the window is to smooth H_d – clearly we would prefer to have the mainlobe width of $W(\omega_1, \omega_2)$ small so that H_d is changed as little as possible. We also want sidebands of small amplitude so that the ripples in the (ω_1, ω_2) plane outside the region of interest are kept small.

The two most popular methods of forming 2d windows from 1d windows are

(i) Taking the product of 1d windows:

$$w(u_1, u_2) = w_1(u_1)\, w_2(u_2)$$

(ii) Rotating a 1d window:

$$w(u_1, u_2) = w_1(u)|_{u=\sqrt{(u_1^2+u_2^2)}}$$

[10%]

(b) First find the FT of w_1

$$W_1(\omega_1) = \int_{-U_1}^{U_1} \mathrm{e}^{-j\omega_1 u_1} du_1$$

$$= \left[\frac{\mathrm{e}^{-j\omega_1 u_1}}{-j\omega_1}\right]_{-U_1}^{U_1}$$

$$= 2U_1 \ \mathrm{sinc}\omega_1 U_1$$

$W(\omega_2)$ will take precisely the same form so that the required spectrum will be the product of W_1 and W_2.

$$W(\omega_1, \omega_2) = 4U_1U_2 \ \mathrm{sinc}\omega_1 U_1 \ \mathrm{sinc}\omega_2 U_2$$

(cont.

Spectrum looks like:

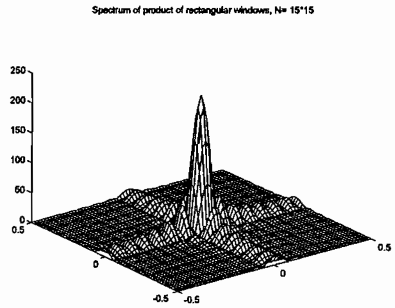


Fig. 1

where, U_1 is (for illustrative purposes) taken as 2.5π in the above sketch and units on ω_1 and ω_2 axes are in units of 2π. Only sketch of general shape and some indication of where zeros of sidebands are $\left(\frac{n\pi}{U_i}\right)$ required for answer.

[25%]

(c) Repeat above with the cosine window. First find the FT of w_1

$$W_1(\omega_1) = \int_{-U_1}^{U_1} \cos\left(\frac{\pi u_1}{U_1}\right) e^{-j\omega_1 u_1} du_1$$

$$= \frac{1}{2}\int_{-U_1}^{U_1} e^{ju_1(\pi/U_1-\omega_1)} + e^{-ju_1(\pi/U_1+\omega_1)} du_1$$

$$= \frac{1}{2}\left[\frac{e^{ju_1(\pi/U_1-\omega_1)}}{j(\pi/U_1-\omega_1)} - \frac{e^{-ju_1(\pi/U_1+\omega_1)}}{j(\pi/U_1+\omega_1)}\right]_{-U_1}^{U_1}$$

$$= U_1\{\ \mathrm{sinc}(\pi-\omega_1 U_1) + \ \mathrm{sinc}(\pi+\omega_1 U_1)\} \qquad (1)$$

As before, $W(\omega_2)$ will take precisely the same form so that the required spectrum will be the product of W_1 and W_2.

$$W(\omega_1,\omega_2) = U_1 U_2\{\ \mathrm{sinc}(\pi-\omega_1 U_1) + \ \mathrm{sinc}(\pi+\omega_1 U_1)\}\{\ \mathrm{sinc}(\pi-\omega_2 U_2) + \ \mathrm{sinc}(\pi+\omega_2 U_2)\}$$

(TURN OVER for continuation of SOLUTION 1

Spectrum along ω_1 axis looks like:

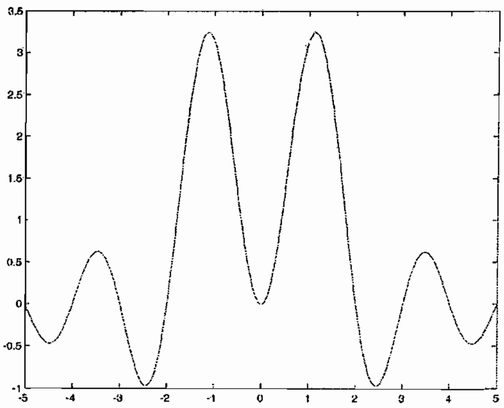


Fig. 2

(equation 1 drawn with $U_1 = \pi$). Require zero crossings in answer.

As we can see from the above plot, the spectrum of the cosine window has a wide main lobe with a significant depression at $\omega = 0$ – even though the sidelobes are fairly low, the mainlobe characteristics are not desirable. [25%]

(d) We can deduce the spectrum of this superposition of windows from the above results:

$$W_1(\omega_1) = U_1\,(2\alpha \;\mathrm{sinc}\omega_1 U_1 + \beta\{\;\mathrm{sinc}(\pi - \omega_1 U_1) + \;\mathrm{sinc}(\pi + \omega_1 U_1)\})$$

And similarly for $W_2(\omega_2)$. If $w_1(0) = 1$ we have

$$\alpha + \beta = 1$$

If $W_i(5\pi/(2U_i) = 0$ then we have that

$$U_1\,(2\alpha \;\mathrm{sinc}5\pi/2 + \beta\{\;\mathrm{sinc}(3\pi/2) + \;\mathrm{sinc}(7\pi/2)\}) = 0$$

so that $W_i(5\pi/(2U_i) = 0$ gives us

(cont.

$$\frac{2\alpha}{5\pi/2} - \beta(2/(3\pi) + 2/(7\pi)) = 0$$

and therefore

$$\beta[1 + 50/42] = 1$$

giving $\beta = 42/92 = 0.46$ and $\alpha = 50/92 = 0.54$

[30%]

(e) Note that the above values of α and β are in fact precisely those used in the Hamming window. It is clearly desirable to have the window function take the value 1 when centred on $(u_1, u_2) = 0$, as we do not want to introduce an overall scale. Also, given the form of the spectrum W the zero crossings are at integer multiples of $n\pi/U_i$, $n \geq 2$ and the sidelobe peaks are at $n\pi/(2U_i)$ where $n > 3$ and is odd. Thus, ensuring that the first main sidelobes have their peaks going to zero, ie the sidelobes with peaks at $\pm\frac{5\pi}{U_i}$, is clearly going to suppress the largest sidelobes and hence their effect (see this by superposing the plots). [10%]

2 (a) (i)

$$s(u_1, u_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \delta(u_1 - n_1\Delta_1, u_2 - n_2\Delta_2)$$

and

$$\boxed{G_s(\omega_1, \omega_2) = \frac{1}{\Delta_1 \Delta_2} \sum_{p_1=-\infty}^{\infty} \sum_{p_2=-\infty}^{\infty} G(\omega_1 - p_1\Omega_1, \omega_2 - p_2\Omega_2)}$$

where $\Omega_i = \frac{2\pi}{\Delta_i}$. It can therefore be seen that the Fourier transform or spectrum of the sampled 2d signal is the periodic repetition of the spectrum of the unsampled 2d signal – precisely analogous to the 1d case. It is therefore clear that for a bandlimited 2d signal, we must sample at more than twice the largest frequencies in the signal to keep these copies of the FT separate. [15%]

(ii) Given Part (i) we know that to avoid aliasing we need to sample at twice the largest frequencies in the signal. Therefore

(TURN OVER for continuation of SOLUTION 2

$$\Omega_1 = \frac{2\pi}{\Delta_1} > 2\Omega$$

and

$$\Omega_2 = \frac{2\pi}{\Delta_2} > 10\Omega$$

Therefore 2Ω and 10Ω are the minimum sampling frequencies required.

[15%]

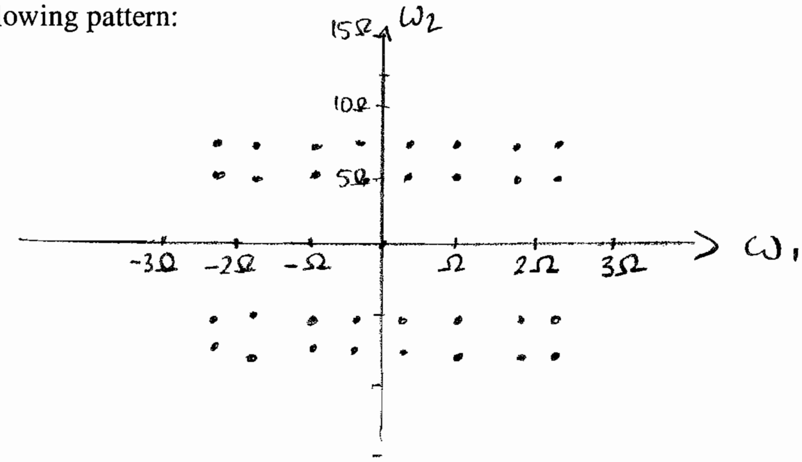
(iii) The above implies that we require

$$\Delta_1 < \frac{\pi}{\Omega}$$

and

$$\Delta_2 < \frac{\pi}{5\Omega}$$

Note that given value of $\Delta_1 (= 3\pi/(2\Omega))$ is greater than the above critical value, while given value of $\Delta_2 (= \pi/(6\Omega))$ is less than above critical value. Thus we will get aliasing. Since the spectrum of the sampled signal will be the spectrum of the original signal (which is 4 delta functions at $\pm\Omega$ and $\pm 5\Omega$) repeated at every interval of the sampling frequency, we get the following pattern: [15%]



A variety of forms of sketch are acceptable (ie 2D or 3D).

(b) (i) The likelihood is obtained by using the fact that the noise is Gaussian:

(cont.

$$P(\mathbf{y}|\mathbf{x}) \propto e^{-\frac{1}{2}\mathbf{d}^T N^{-1}\mathbf{d}} = e^{-\frac{1}{2}(\mathbf{y}-L\mathbf{x})^T N^{-1}(\mathbf{y}-L\mathbf{x})}$$

[15%]

(ii) Again, using the fact that we can regard **x** as a gaussian random variable:

$$P(\mathbf{x}) \propto e^{-\frac{1}{2}\mathbf{x}^T C^{-1}\mathbf{x}}$$

From Bayes' theorem,

$$P(\mathbf{x}|\mathbf{y}) \propto P(\mathbf{y}|\mathbf{x})P(\mathbf{x})$$

which is then given by

$$P(\mathbf{x}|\mathbf{y}) \propto e^{-\frac{1}{2}\left[\mathbf{y}-L\mathbf{x})^T N^{-1}(\mathbf{y}-L\mathbf{x})+\mathbf{x}^T C^{-1}\mathbf{x}\right]}$$

[15%]

(iii) The conventional expression for the Wiener filter contains the following quantities:

$H(\boldsymbol{\omega})$ is the Fourier transform of the point-spread function $h(\mathbf{n})$ – it is associated with the distortion matrix L.

$P_{xx}(\boldsymbol{\omega})$ is the *power spectrum* (Fourier transform of the autocorrelation function $R_{xx}(\mathbf{n})$) of the (assumed) spatially stationary process $x(\mathbf{n})$. It is associated with the covariance of the gaussian random variable **x**, C.

$P_{dd}(\boldsymbol{\omega})$ is the *power spectrum* (Fourier transform of the autocorrelation function $R_{dd}(\mathbf{n})$) of the (assumed) spatially stationary process $d(\mathbf{n})$. It is associated with the covariance of the noise, **d**, assumed gaussian, N. [25%]

(TURN OVER

3 (a) The Z transform of the original sampled signal y_n is

$$Y(z) = \sum_{n=-\infty}^{+\infty} y_n z^{-n}$$

therefore

$$Y(-z) = \sum_{n=-\infty}^{+\infty} y_n (-z)^{-n}$$

so that

$$\frac{1}{2}[Y(z) + Y(-z)] = \frac{1}{2}\sum_{n=-\infty}^{+\infty} y_n z^{-n}(1 + (-1)^n)$$

$(1 + (-1)^n)$ is 2 if n is even and 0 if n is odd, therefore the RHS of the above equation is $\widehat{Y}(z)$. [20%]

(b) Analysing the block diagram in the figure and substituting the above result, we get:

$$\begin{aligned}
\widehat{X}(z) &= \frac{1}{2}G_0(z)[Y_0(z) + Y_0(-z)] + \frac{1}{2}G_1(z)[Y_1(z) + Y_1(-z)] \\
&= \frac{1}{2}G_0(z)H_0(z)X(z) + \frac{1}{2}G_0(z)H_0(-z)X(-z) \\
&\quad + \frac{1}{2}G_1(z)H_1(z)X(z) + \frac{1}{2}G_1(z)H_1(-z)X(-z) \\
&= \frac{1}{2}X(z)[G_0(z)H_0(z) + G_1(z)H_1(z)] \\
&\quad + \frac{1}{2}X(-z)[G_0(z)H_0(-z) + G_1(z)H_1(-z)]
\end{aligned}$$

If we require $\widehat{X}(z) \equiv X(z)$ – the Perfect Reconstruction (PR) condition – then the antialiasing condition to eliminate the terms in $X(-z)$ is:

$$G_0(z)H_0(-z) + G_1(z)H_1(-z) \equiv 0$$

and the PR condition is:

$$G_0(z)H_0(z) + G_1(z)H_1(z) \equiv 2$$

(cont.

In an image coder, these conditions ensure that the forward and inverse wavelet transforms do not introduce any distortions in the output image. The only source of distortion is the coefficient quantiser. [25%]

(c) If $H_1(z) = zG_0(-z)$ and $G_1(z) = z^{-1}H_0(-z)$, then the LHS of the anti-aliasing condition becomes:

$$G_0(z)\,H_0(-z) + z^{-1}H_0(-z)\,(-z)\,G_0(z) = G_0(z)\,H_0(-z) - H_0(-z)\,G_0(z) = 0$$

Hence this condition is satisfied. The PR condition becomes:

$$G_0(z)H_0(z) + z^{-1}H_0(-z)zG_0(-z) = G_0(z)H_0(z) + H_0(-z)G_0(-z) = P(z) + P(-z)$$

Hence if $P(z) + P(-z) = 2$, the PR condition is also satisfied. [20%]

(d) For the given filters:

$$\begin{aligned} P(z) &= G_0(z)H_0(z) = \frac{1}{2}(z+2+z^{-1})\,(az^2+bz+c+bz^{-1}+az^{-2}) \\ &= \frac{1}{2}[az^3+(2a+b)z^2+(a+2b+c)z+(2b+2c)+(c+2b+a)z^{-1}+(b+2a)z^{-2}+az^{-3}] \end{aligned}$$

If $P(z) + P(-z) = 2$, then the coefs of z^2 and z^{-2} must be zero and the coef of z^0 must be 1 (odd terms will cancel). Hence

$$2a+b=0 \quad \text{and} \quad \frac{1}{2}(2b+2c) = b+c = 1$$

If, in addition, $H_0(z) = 0$ at $z = -1$, $2a-2b+c = 0$.

Hence $-b-2b+1-b = 0$, and so $b = 1/4$, $a = -b/2 = -1/8$ and $c = 1-b = 3/4$.

Since $H_1(z) = zG_0(-z)$, and $G_1(z) = z^{-1}H_0(-z)$, we have

$$H_1(z) = \frac{1}{2}\left(-z^2+2z-1\right)$$

$$G_1(z) = -\frac{1}{8}z - \frac{1}{4} + \frac{3}{4}z^{-1} - \frac{1}{4}z^{-2} - \frac{1}{8}z^{-3}$$

[15%]

(e)

$$G_0(z) = \frac{1}{2}(z+2+z^{-1}) = \frac{1}{2}z^{-1}(z+1)^2$$

(TURN OVER for continuation of SOLUTION 3

So G_0 has two zeros at $z = -1$, which means it is a lowpass filter.

$$\begin{aligned}
H_0(z) &= az^2 + bz + c + bz^{-1} + az^{-2} \\
&= \frac{1}{8}(-z^2 + 2z + 6 + 2z^{-1} - z^{-2}) \\
&= \frac{1}{8}z^{-2}(z+1)^2(-z^2 + 4z - 1) \\
&= \frac{-1}{8}z^{-2}(z+1)^2(z - 2 + \sqrt{3})(z - 2 - \sqrt{3})
\end{aligned}$$

This also has two zeros at $z = -1$ and two zeros that are well away from the unit circle, at $2 \pm \sqrt{3}$. The latter two zeros therefore only have a small effect on the frequency response, and so this is also a lowpass filter.

Now $H_1(z) = zG_0(-z)$ and $G_1(z) = z^{-1}H_0(-z)$, so these filters will have zeros at the negative of the zeros of the corresponding filters G_0 and H_0.

Hence H_1 will have two zeros at $z = +1$ and be a highpass filter. G_1 will have two zeros at $z = +1$ and two zeros that are well away from the unit circle, at $-2 \pm \sqrt{3}$; and hence G_1 will also be highpass.

Good performance is achieved in a wavelet transform image coder if the filters are of the above types, because the wavelet coefficients are generated by outputs from H_1 at each stage, and will have least energy (and hence greatest compression performance) if H_1 is a strongly highpass filter. This is because typical image signals have most of their energy at low frequencies. [20%]

4 (a) The sketch below shows the sensitivity of the human eye to luminance and chrominance (Y and (U,V)) – the horizontal scale is spatial frequency and the vertical scale is contrast sensitivity (ratio of maximum visible range of intensities to the minimum discernable peak-to-peak intensity variation at the specified frequency).

We see that the maximum sensitivity to Y occurs for spatial frequencies around 5 cycles/degree and we note that the eye has very little response to anything above about 100 cycles/degree. Also note that sensitivity to luminance drops off at low spatial frequencies (in the absence of time variations) and that the maximum chrominance sensitivity occurs at much lower spatial frequencies than for Y. The chrominance sensitivities indeed fall off above about 1 cycle/degree.

(cont.

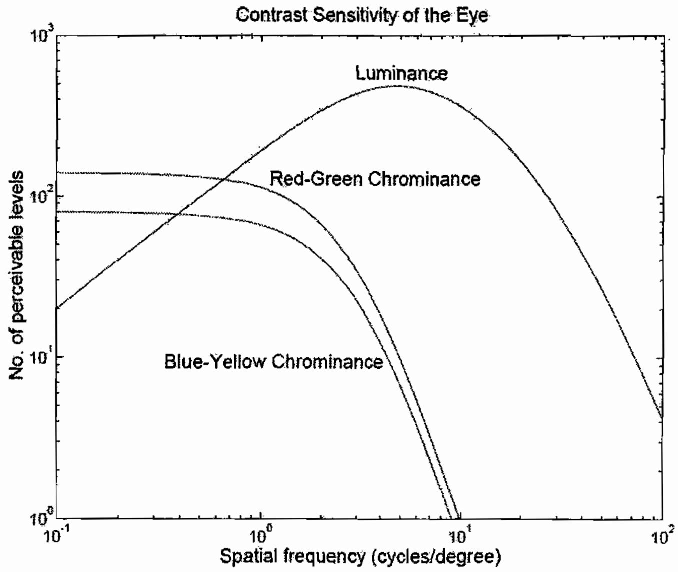


Fig. 3

From these facts we can see why it is better to convert to the YUV domain before attempting image compression:

- The U and V components may be sampled at a lower rate than Y due to the narrower bandwidth required
- The U and V components may be quantised more coarsely than Y due to the lower contrast sensitivity.

[15%]

(b) Under the system used in JPEG the luminance (Y) of a pixel is obtained from its RGB components from:

$$Y = 0.3R + 0.6G + 0.1B$$

The chrominance of a pixel in this system is defined by two components U and V which are given by:

$$U = 0.5(B - Y) \quad \text{and} \quad V = 0.625(R - Y)$$

(TURN OVER for continuation of SOLUTION 4

If C is the matrix that maps $\mathbf{x}_{rgb} = [R \;\; G \;\; B]$ onto $\mathbf{x}_{yuv} = [Y \;\; U \;\; V]$ and C^{-1} is its inverse, then the top row of the matrix equation $\mathbf{x}_{yuv} = C\mathbf{x}_{rgb}$ is given by

$$Y = C_{11}R + C_{12}G + C_{13}B$$

and comparing this with the above 'definition' of Y we see that $C_{11} = 0.3, \;\; C_{12} = 0.6, \;\; C_{11} = 0.1$, as required.

From the equation for C we also have $\mathbf{x}_{rgb} = C^{-1}\mathbf{x}_{yuv}$. So, in order to pick out the first column of C^{-1} we can take the vector $[Y \;\; 0 \;\; 0]$ so that

$$[R \;\; G \;\; B]^T = Y[C_{11}^{-1} \;\; C_{21}^{-1} \;\; C_{31}^{-1}]^T$$

But if $U = V = 0$ we must have $B = Y$ and $R = Y$, which then tells us that $0.6Y = 0.6G$ so that $R = G = B = Y$ giving

$$Y[1 \;\; 1 \;\; 1]^T = Y[C_{11}^{-1} \;\; C_{21}^{-1} \;\; C_{31}^{-1}]^T$$

as required, ie no chrominance components lead to an equal contribution from each of R, G, B. [25%]

(c) The image size is $768 \times 1024 = (3 \times 2^8)(1 \times 2^{10}) = 3 \times 2^{18}$.

Hence the luminance component (Y) of the YUV image is also of this size.

Since the U and V images are subsampled 2 : 1 in each direction, they each have 3×2^{16} pixels.

The number of bits required to code an image is $\geq$ *entropy* $\times$ *no. pixels*, thus

Y image requires $\geq 1.3 \times (3 \times 2^{18}) = 3.9 \times 2^{18}$ bits

U or V images each require $\geq 0.6 \times (3 \times 2^{16}) = 1.8 \times 2^{16}$ bits

Hence the total number of bits for Y, U, V images is $\geq (3.9 \times 2^2 + 1.8 + 1.8) \times 2^{16} = 19.2 \times 2^{16} = 1.2 \times 2^{20} = 1.2$ Mbit

In practice the number of bits could be a little higher than this (due to the use of non-ideal codes) or a little lower if higher-order correlations in the data can be exploited.

The proportion of chrominance (U, V) bits is $\frac{2 \times 1.8}{19.2} = 18.75\%$. [30%]

(d) *Run* determines the number of zero-valued coefficients which precede each non-zero coefficient.

(cont.

Size represents the $\log_2$ of the magnitude of the non-zero coefficient (rounded down to an integer value).

Additional Bits are an n-bit binary code that specifies the value of a coefficient of size n within the range 2^{n-1} to $(2^n - 1)$ and -2^{n-1} to $-(2^n - 1)$.

Typically no more than 15 states are needed for *Run* and no more than about 12 states are needed for *Size*, so it is feasible to create a 2-dimensional Huffman code for these two parameters combined.

This code can also take advantage of the strong correlation that exists between longer runs and small sizes (and vice-versa). Hence the entropy of the 2d code is less than that for the two parameters independently, and the bit rate can be lower than the basic entropy of the *Run* and *Size* data separately. [30%]

END OF SOLUTIONS

Engineering Tripos Part IIB 2009

Numerical Solutions: 4F8 Image Processing and Image Coding

Q1

(b) $W(\omega_1, \omega_2) = 4U_1U_2 \text{ sinc}\omega_1 U_1 \text{ sinc}\omega_2 U_2$

(c) $W(\omega_1, \omega_2) = U_1U_2\{ \text{ sinc}(\pi - \omega_1 U_1) + \text{ sinc}(\pi + \omega_1 U_1)\}\{ \text{ sinc}(\pi - \omega_2 U_2) + \text{ sinc}(\pi + \omega_2 U_2)\}$

(d) $\alpha = 0.54$, $\beta = 0.46$.

Q2

(a)(i) $s(u_1, u_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \delta(u_1 - n_1\Delta_1, u_2 - n_2\Delta_2)$

and

$G_s(\omega_1, \omega_2) = \frac{1}{\Delta_1 \Delta_2} \sum_{p_1=-\infty}^{\infty} \sum_{p_2=-\infty}^{\infty} G(\omega_1 - p_1\Omega_1, \omega_2 - p_2\Omega_2)$

where $\Omega_i = \frac{2\pi}{\Delta_i}$.

(ii) $\Omega_1 = \frac{2\pi}{\Delta_1} > 2\Omega$ and $\Omega_2 = \frac{2\pi}{\Delta_2} > 10\Omega$

(b)(i) $P(\mathbf{y}|\mathbf{x}) \propto e^{-\frac{1}{2}\mathbf{d}^T N^{-1}\mathbf{d}} = e^{-\frac{1}{2}(\mathbf{y}-L\mathbf{x})^T N^{-1}(\mathbf{y}-L\mathbf{x})}$

(ii) $P(\mathbf{x}|\mathbf{y}) \propto e^{-\frac{1}{2}\left[(\mathbf{y}-L\mathbf{x})^T N^{-1}(\mathbf{y}-L\mathbf{x}) + \mathbf{x}^T C^{-1}\mathbf{x}\right]}$

Q3

(b) The antialiasing condition is: $G_0(z)H_0(-z) + G_1(z)H_1(-z) \equiv 0$

and the PR condition is: $G_0(z)H_0(z) + G_1(z)H_1(z) \equiv 2$

(c) $b = 1/4$, $a = -1/8$ and $c = 3/4$.

$H_1(z) = \frac{1}{2}(-z^2 + 2z - 1)$ and $G_1(z) = -\frac{1}{8}z - \frac{1}{4} + \frac{3}{4}z^{-1} - \frac{1}{4}z^{-2} - \frac{1}{8}z^{-3}$

(e) G_0 has two zeros at $z = -1$

$H_0(z)$ has two zeros at $z = -1$ and two zeros that are well away from the unit circle, at $2 \pm \sqrt{3}$.

H_1 will have two zeros at $z = +1$
G_1 will have two zeros at $z = +1$ and two zeros that are well away from the unit circle, at $-2 \pm \sqrt{3}$.

Q4

(c) Y image requires 3.9×2^{18} bits

U or V images each require $= 1.8 \times 2^{16}$ bits

Total number of bits for Y, U, V images is $\times 2^{16} = 1.2 \times 2^{20} = 1.2$ Mbit

The proportion of chrominance (U, V) bits is 18.75%.

JL
May 2009