Version JL/2

EGT3
ENGINEERING TRIPOS PART IIB: SOLUTIONS

Tuesday 30 April 2024   2 to 3:40

Module 4F8

IMAGE PROCESSING AND IMAGE CODING

*Answer not more than **three** questions.*

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

Answers to questions in each section should be tied together and handed in separately.

*Write your candidate number **not** your name on the cover sheet.*

STATIONERY REQUIREMENTS
Single-sided script paper

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed
Engineering Data Book

10 minutes reading time is allowed for this paper at the start of the exam.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

You may not remove any stationery from the Examination Room.

1 (a) The 2D Fourier Transform is written as:

$$\boxed{F(\omega_1,\omega_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u_1,u_2)\, e^{-j(\omega_1 u_1+\omega_2 u_2)}\, du_1\, du_2}$$

So, shifting by (a, b) will give:

$$F^{'}(\omega_1,\omega_2) = \iint_{-\infty}^{\infty} f(u_1 - a, u_2 - b)\, e^{-j(\omega_1 u_1+\omega_2 u_2)}\, du_1\, du_2$$

Let $u_1 - a = u_1^{'}$ and $u_2 - b = u_2^{'}$

$$\begin{aligned}
F^{'}(\omega_1,\omega_2) &= \iint_{-\infty}^{\infty} f(u_1^{'}, u_2^{'})\, e^{-j[(\omega_1(u_1^{'}+a)+\omega_2(u_2^{'}+b)]}\, du_1^{'}\, du_2^{'} \\
&= e^{-j(a\omega_1+b\omega_2)} \iint f(u_1^{'}, u_2^{'})\, e^{-j(\omega_1 u_1^{'}+\omega_2 u_2^{'})}\, du_1^{'}\, du_2^{'} \\
&= e^{-j(a\omega_1+b\omega_2)}\, F(\omega_1,\omega_2)
\end{aligned}$$

$$\therefore \quad \boxed{f(u_1 - a, u_2 - b) \Leftrightarrow e^{-j(a\omega_1+b\omega_2)}\, F(\omega_1,\omega_2)}$$

[15%]

(b) If we rotate our coordinate system as given ($\mathbf{u'} = R\mathbf{u}$), the FT will be given by

$$F^{'}(\omega_1,\omega_2) = \iint_{-\infty}^{\infty} f(u_1\cos\phi + u_2\sin\phi, -u_1\sin\phi + u_2\cos\phi)\, e^{-j(\omega_1 u_1+\omega_2 u_2)}\, du_1\, du_2$$

Now put $u_1' = u_1\cos\phi + u_2\sin\phi$, $u_2' = -u_1\sin\phi + u_2\cos\phi$ noting that $du_1'\, du_2' = |J| du_1\, du_2$, where the jacobian J is 1 (as we simply have a rotation):

$$\begin{aligned}
F^{'}(\omega_1,\omega_2) &= \iint_{-\infty}^{\infty} f(u_1^{'}, u_2^{'})\, e^{-j[\omega_1(u_1'\cos\phi - u_2'\sin\phi)]}\, e^{-j[\omega_2(u_1'\sin\phi+u_2'\cos\phi)]} du_1^{'}\, du_2^{'} \\
&= \iint_{-\infty}^{\infty} f(u_1{'}, u_2{'})\, e^{-j(\omega_1' u_1' + \omega_2' u_2')} du_1{'}\, du_2{'}
\end{aligned}$$

since $[u_1, u_2]^T = R^T[u_1', u_2']^T = [u_1'\cos\phi - u_2'\sin\phi, u_1'\sin\phi + u_2'\cos\phi]^T$, and where $\omega_1' = \omega_1\cos\phi + \omega_2\sin\phi$, $\omega_2' = -\omega_1\sin\phi + \omega_2\cos\phi$.

From this we see that this is just the rotated FT. The sense of the rotation is in the same sense as the spatial plane rotation. [20%]

(c) Take the FT of the line:

$$F(\omega_1, \omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(u_2)\, e^{-j(\omega_1 u_1 + \omega_2 u_2)}\, du_1\, du_2 = \int e^{-j\omega_1 u_1} du_1 = 2\pi\delta(\omega_1)$$

Telling us that the FT of horizontal line which is the u_1-axis is the vertical line in the fourier domain which is the ω_2 axis – the line and its FT are therefore at 90 degrees. [20%]

(d) As above, do the direct FT to give

$$F(\omega_1, \omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(u_2 - [mu_1 + c])\, e^{-j(\omega_1 u_1 + \omega_2 u_2)}\, du_1\, du_2 = \int e^{-j\omega_1 u_1}\, e^{-j\omega_2 (mu_1 + c)} du_1$$

$$= e^{-j\omega_2 c} \int e^{-j(\omega_1 + m\omega_2) u_1} du_1 = 2\pi e^{-j\omega_2 c}\, \delta(\omega_1 + m\omega_2)$$

Therefore, in the Fourier domain, we have a line through the origin, $\omega_2 = -\frac{1}{m}\omega_1$, and a phase factor, $e^{-j\omega_2 c}$. The line in the fourier domain is at right angles to the line in the spatial domain, and the phase factor gives us information about how the line through the origin $u_2 = mu_1$ is translated.
In the previous parts we found that the FT of the u_1-axis was the ω_2 axis, ie $2\pi\delta(\omega_1)$. We also saw that if we rotate in the spatial domain we rotate in the Fourier domain, and that a shift in the spatial domain induces a phase factor in the Fourier domain. So, if we rotate the u_1-axis to the line $u_2 = mu_1$, the FT will be

$$2\pi\delta(\omega_1 + m\omega_2)$$

ie we rotate so that in the Fourier domain our line is at 90 degrees. If we then translate a distance c in the spatial domain along the u_2 axis, we pick up a factor of $e^{-j\omega_2 c}$. Hence we see how to obtain the result above via the answers to the previous parts. [20%]

(e) If we have a finite line (point pair) rather than an infinite line, this can be expressed as the multiplication of the infinite line with a top hat function (1D) along the line. Therefore, the FT of a finite line would be the convolution of the two FTS. ie, we would have the FT of the line (a perpendicular line in the frequency plane) with a sinc oriented along the line. [5%]

(f) The image in figure 1 is made up of 5 finite lines ..therefore the FT's will be made up of 5 lines through the origin (in the (ω_1, ω_2) plane) convolved with a sinc (there will

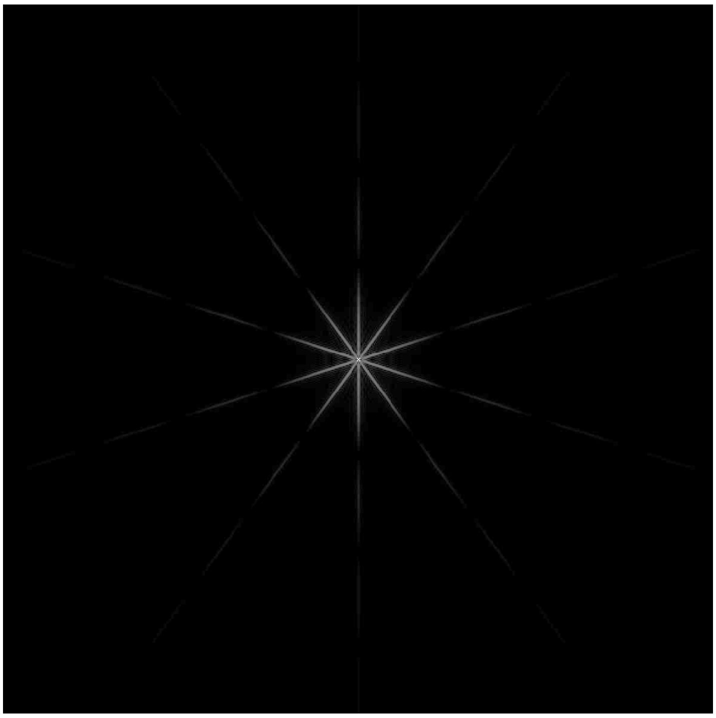


Fig. 1

be a phase factor which is difficult to display). The amplitude of the FT will therefore look something like the image in figure 1

[20%]

2 (a) Given the extent of the filter shown in Fig.2, we know that we are avoiding aliasing. Following the question, we first of all use the form for the impulse response (normalised freqs) for the *alternative ideal bandpass filter*:

$$h(n_1, n_2) =$$
$$\frac{1}{(\pi)}\left[\Omega_{U1}\text{sinc}(\Omega_{U1}n_1) - \Omega_{L1}\text{sinc}(\Omega_{L1}n_1)\right]\frac{1}{(\pi)}\left[\Omega_{U2}\text{sinc}(\Omega_{U2}n_2) - \Omega_{L2}\text{sinc}(\Omega_{L2}n_2)\right]$$

Now put $\Omega_{L1} = 0$ and $\Omega_{L2} = \Omega$ and $\Omega_{U1} = \Omega_{U2} = 2\Omega$ in the above expression (and move away from normalised freqs):

$$h(n_1, n_2) = \frac{\Delta_1\Delta_2}{(\pi^2)}[2\Omega\text{sinc}(2\Omega n_1\Delta_1)][2\Omega\text{sinc}(2\Omega n_2\Delta_2) - \Omega\text{sinc}(\Omega n_2\Delta_2)]$$

Now check that the above result can be obtained via the subtraction of two lowpass filters. First look at the impulse response of the ideal LP filter with $\Omega_{U2} = \Omega_{U1} = 2\Omega$ (and $\Omega_{L2} = \Omega_{L1} = 0$):

$$h_1(n_1, n_2) = \frac{\Delta_1\Delta_2}{(\pi^2)}[4\Omega^2\text{sinc}(2\Omega n_2\Delta_2)\text{sinc}(2\Omega n_1\Delta_1)]$$

Now look at the impulse response of the ideal LP filter with $\Omega_{U2} = \Omega, \Omega_{U1} = 2\Omega$ (and $\Omega_{L2} = \Omega_{L1} = 0$):

$$h_2(n_1, n_2) = \frac{\Delta_1\Delta_2}{(\pi^2)}[2\Omega^2\text{sinc}(\ \Omega n_2\Delta_2)\text{sinc}(2\Omega n_1\Delta_1)]$$

From the above expressions, it is clear that $h = h_1 - h_2$, as required. [25%]

(b) (i) Here, our distortion L can be modelled by convolving with a psf h. Suppose we firstly ignore the noise d.

$$y(n_1, n_2) \approx \sum_{m_1}\sum_{m_2} h(m_1, m_2)x(n_1 - m_1, n_2 - m_2)$$

x and$y$ are convolved $\Longrightarrow$ take the Fourier transform of each side of the above to give:

$$Y(\omega_1, \omega_2) = H(\omega_1, \omega_2)X(\omega_1, \omega_2)$$

where: $H(\omega_1, \omega_2) = \sum_{n_2=-\infty}^{\infty}\sum_{n_1=-\infty}^{\infty} h(n_1, n_2)e^{-j(\omega_1 n_1 + \omega_2 n_2)}$

$$\therefore X(\omega_1, \omega_2) = \frac{Y(\omega_1, \omega_2)}{H(\omega_1, \omega_2)}$$

the inverse filter is therefore $1/H$. If H has zeros then we encounter significant problems – can get around some of these using the *generalised inverse filters*. [15%]

(ii) In the MaxEnt method the prior is given by:

$$Pr(\mathbf{x}) \propto e^{\alpha S}$$

where the *cross entropy* S of the image is given by

$$S(\mathbf{x}, \mathbf{m}) = \sum_i \left[x_i - m_i - x_i \ln \left(\frac{x_i}{m_i} \right) \right]$$

where **m** is the *measure* on an image space (*the model*) to which the image **x** defaults in the absence of data. To see that the global maximum of S occurs at $\mathbf{x} = \mathbf{m}$, differentiate S wrt a pixel x_j:

$$\frac{\partial S}{\partial x_j} = 1 - \ln \frac{x_j}{m_j} - \frac{x_j m_j}{x_j m_j} = -\ln \frac{x_j}{m_j} = 0$$

which requires $x_j = m_j$ for all j. So in the absence of data, the result defaults to **m** – we often take **m** as a constant (non-zero) image. [20%]

(c) (i) The FT of the sampled image G_s (with sampling intervals of Δ_1, Δ_2) is given by

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

where G is the FT of the unsampled image.
Thus the spectrum of the sampled image is the spectrum of the original image repeated at gridpoints (spacing Ω_1,Ω_2). Therefore the spectrum will look like:

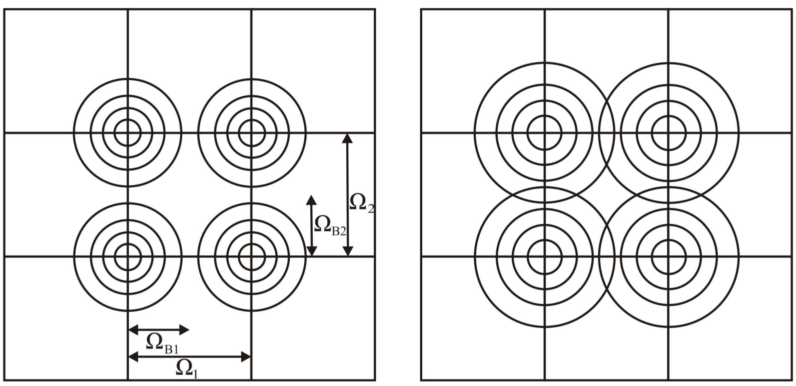


Fig. 2

 (cont.

These repetitions of the spectrum are referred to as *aliasing*. If our sampling, $\Omega_i = 2\pi/\Delta_i$, is greater than twice Ω_{Bi} (the highest freqs in image) we do not have overlap – if not, the spectra overlap and this lead to *aliasing artefacts*. [15%]

(ii) Sketch to be done – pretty trivial.
Unaliased implies that $\Omega_i > 2\Omega_{ci}$, where $\Omega_i = 2\pi/\Delta_i$ and Ω_{ci} are the highest freqs in the i direction ($i = 1, 2$).
Suppose image is of width and height (a_1, a_2), then $\Delta_i = a_i/n_i$, where $n_1 = 512$, $n_2 = 256$.
Highest ω_1 freq is $\Omega_{c1} = 200\Delta\omega_1$, and the highest ω_2 freq is $\Omega_{c2} = 120\Delta\omega_2$, where $\Delta\omega_i = 2\pi/(n_i\Delta_i) = 2\pi/a_i$.
Therefore if we are to downsample to (n'_1, n'_2) so as just to avoid aliasing, we have

$$\Delta'_i = \frac{a_i}{n'_i}, \quad \frac{2\pi}{\Delta'_i} > 2\Omega_{ci}$$

which means:

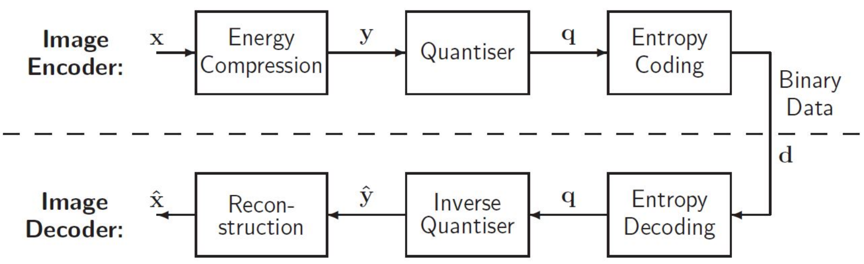
$$\frac{2\pi n'_1}{a_1} > 2\left(200\,\frac{2\pi}{a_1}\right)$$

and

$$\frac{2\pi n'_2}{a_2} > 2\left(120\,\frac{2\pi}{a_2}\right)$$

which tells us that $n'_1 > 400$ and $n'_2 > 240$. Therefore the max downsampling we can do in order to avoid aliasing is (400×240).
[A simpler argument involving ratios is fine] [25%]

3 (a) The figure below shows the main blocks in any image coding system. **x** is a monochrome (Y) image. Elements of **x** are $x(\mathbf{n}) \equiv x(n_1, n_2)$, where n_1 runs over rows and n_2 runs over columns.



[5%]

(b) Let us apply T to 2×2 blocks of X via $TX_{ij}T^T$, where X_{11} is the first block, X_{12} is the second block, etc.

$$TX_{11}T^T = TX_{12}T^T = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$TX_{21}T^T = TX_{12}T^T = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 2 & 2 \\ 3 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ -1 & 0 \end{bmatrix}$$

Therefore

$$Y = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 5 & 0 & 5 & 0 \\ -1 & 0 & -1 & 0 \end{bmatrix}$$

Now rearrange the image Y into a regrouped image Y'

$$Y' = \begin{bmatrix} 2 & 2 & 0 & 0 \\ 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 \end{bmatrix}$$

Thus we see that we have non-zero entries in the top LH 2×2 subimage and non-zero entries in the bottom LH 2×2 subimage – corresponding to having low frequency content (top two rows of X) and Lo-Hi (low pass horizontal, high pass vertical) frequency content (bottom two rows). This is exactly what we would expect. [35%]

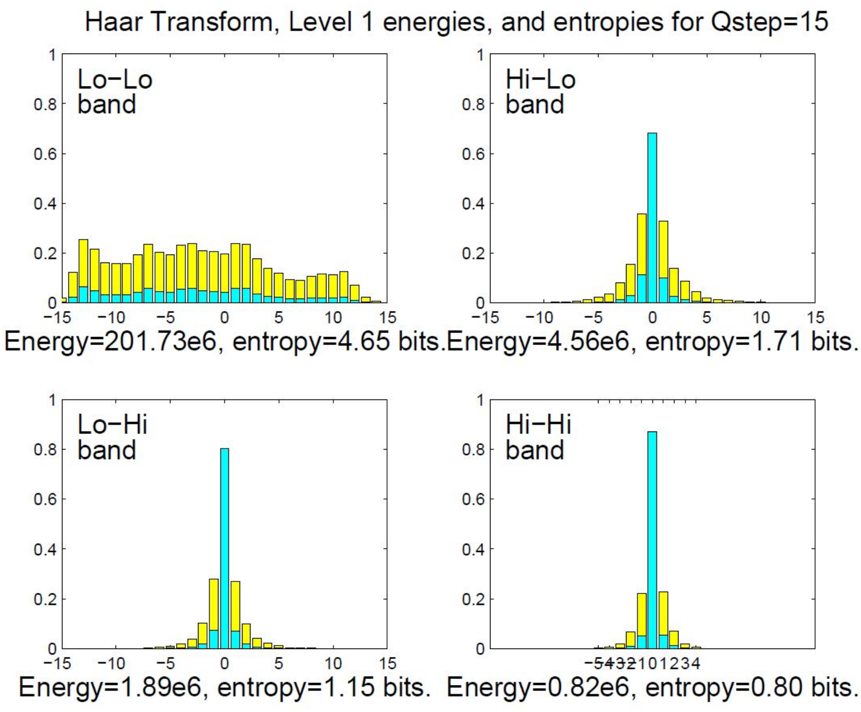
 (cont.

(c) If

$$Y' = \begin{bmatrix} Y'_{11} & Y'_{12} \\ Y'_{21} & Y'_{22} \end{bmatrix}$$

and we take the sum of the squared entries as a measure of energy, we see that $E_{11} = 58$, $E_{12} = 0$, $E_{21} = 2$, $E_{22} = 0$. So we basically have most of the image energy in the Lo-Lo sub-band and some energy in the Lo-Hi sub-band. The energy of the original image is $8 + 14 + 36 = 58$, of course, but now we have concentrated the energy into just two sub-bands. When we quantise these sub-bands, we can quantise the sub-bands with low energy more coarsely. We can also apply a further level of transform to the Lo-Lo band to further concentrate the energy. It is the selective quantisation of these sub-bands that enables us to effectively compress the image. [15%]

(d) Applying a 2×2 Haar transform to a typical natural image will give a distribution of energies/entropies over sub-bands as illustrated below (this is for Lenna with $Qstep = 15$:



Any similar looking sketches will be fine. [15%]

(e) The *approximate entropy* is arrived at by assuming that the distribution of the non-Lo-Lo sub-bands can be approximated by a laplacian pdf:

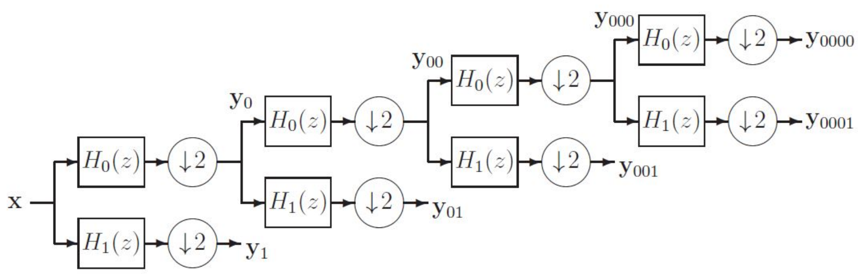
$$p(x) = \frac{1}{2x_0} e^{-|x|/x_0}$$

If we form the variance of this pdf ($\int x^2 p(x) dx$) we find that it is $2x_0^2$. Thus the SD (σ) is $\sqrt{2}x_0$, meaning that $x_0 = \frac{\sigma}{\sqrt{2}}$.
Q is the step size of the uniform quantiser used. [15%]

(f) The $n \times n$ DCT (for $n > 2$) gives us more basis functions and therefore better frequency resolution when we decompose into subbands. Exeperimentally the 8×8 DCT is optimal for standard compression.

If we take the 8-point DCT we can see that odd rows possess even symmetry about their centres and the even rows possess odd symmetry. This enables us to write the transform as two lots of 4×4 matrix multiplications. The same symmetry arguments can be applied again on one of the 4×4 matrices. In this way the number of operations required to perform the DCT can be more than halved. [15%]

4 (a) Such a filterbank is sketched below.



Assume that the max freq f_{max} is $\frac{1}{2}f_s$ (Nyquist), so that the passband of our original N samples is $0 - \frac{1}{2}f_s$.
At the first level the highpass signal has $N/2$ samples and goes from $\frac{1}{4}f_s - \frac{1}{2}f_s$.
At the second level the highpass signal has $N/4$ samples and goes from $\frac{1}{8}f_s - \frac{1}{4}f_s$.
At the third level the highpass signal has $N/8$ samples and goes from $\frac{1}{16}f_s - \frac{1}{8}f_s$.
At the fourth level the highpass signal has $N/16$ samples and goes from $\frac{1}{32}f_s - \frac{1}{16}f_s$. [20%]

(b) The inverse tree is just the reverse of the tree shown in Part (a) (upsample before $G_0(z)$ and $G_1(z)$).
the Perfect Reconstruction (PR) conditions are:

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

and

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

[15%]

(c) Substitute into first PR equation:

$$G_0(z)H_0(z) + G_1(z)H_1(z) = G_0(z)H_0(z) + H_0(-z)G_0(-z)$$

$$\implies P(z) + P(-z) = 2$$

This requires *all $P(z)$ terms in even powers of z to be zero*, except the z^0 term which should be 1. The $P(z)$ terms in odd powers of z may take any desired values since they cancel out. [15%]

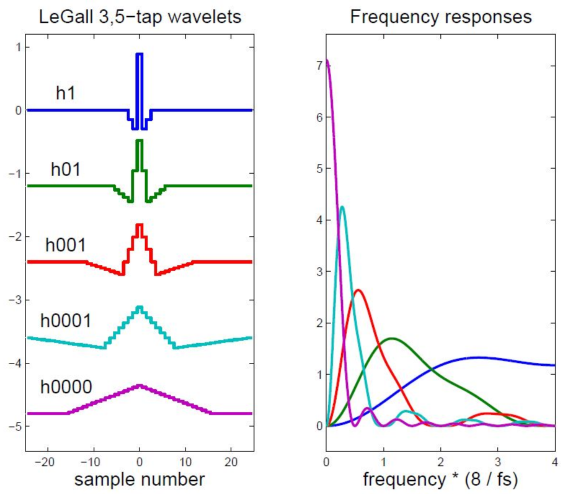
(d)

$$P_0(Z) = (1+Z)^2\,(1+aZ) = 1+(2+a)Z+(1+2a)Z^2+aZ^3$$
$$= 1+\frac{3}{2}Z-\frac{1}{2}Z^3 \text{ where } a=-\tfrac{1}{2} \text{ to suppress the term in } Z^2$$

..since we know we can only have odd powers of z.

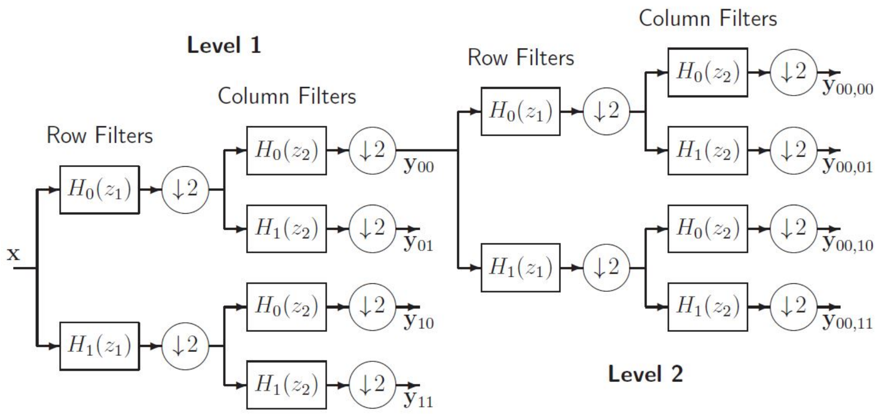
Suppose we allocate the factors of P_0 such that $(1+Z)$ gives H_0 and $(1+Z)(1+aZ)$ gives G_0. This split gives the LeGall (3,5) tap filters. [Other answers are acceptable]. [20%]

(e)



[20%]

(f)



[10%]

END OF PAPER