Chapter 6

Technical Specifications and Pre-processing of SAXS Images

Digital images are stored in a variety of numerical formats that determine the nature of the data to be analysed. Understanding these technical specifications is a central issue to the proper analysis of SAXS image data as it influences the nature of the subsequent statistical models that are developed. The technical details of the SAXS images to be analysed are reported upon in Section 6.1 of this thesis. The process of SAXS imaging produces artifacts and errors on the observed SAXS image, some of the more readily identified artifacts are presented in Section 6.2. The SAXS images are extensively processed before any further analysis in order to account for such experimental effects and artifacts. Section 6.3 discusses these 'pre-processing' operations and the impact that they may have on the results of any subsequent statistical analysis. A brief summary in Section 6.4 concludes the chapter.

6.1   Technical Specifications of the SAXS Images

The SAXS images used in this thesis were recorded using the SAXS imaging facilities at station 2.1 of the Daresbury Synchrotron Radiation Source. A multi-wire proportional counter detector was used to record an image of the pattern by summing the number of detecting x-ray photon counts on the detector over an exposure time of 300 seconds (Lewis *et al* 1997b; Lewis *et al* 2000). The detector has a high dynamic range in that it can record over 4 decades in intensity. The Intel IEEE Float format was used to record this large range of intensity values as real numbers with high precision (IEEE Standards Committee 754 1985; Daresbury Synchrotron Radiation

Source 2009 online). An image matrix (raster) of 512 rows and 512 columns is used to capture the pattern, with only one intensity value per entry. Each image was approximately one megabyte in size and stored in the BSL file format which is amenable to a range of x-ray diffraction specific functions (Daresbury Synchrotron Radiation Source 2009 online). The BSL file format used at station 2.1 consists of several components:

a) a header file written in ASCII containing information such as image title, the number of frames in the file, the names of the data files and,

b) a binary data file of the SAXS image intensities in unformatted, fixed length records as a stream of 4 byte floats with the Intel IEEE Float format for big-endian machines used to record the intensity values and,

c) a binary file describing the calibration operations performed

(Daresbury Synchrotron Radiation Source 2009 online). The image data was pre-processed to correct for a number of experimental parameters (see Section 6.3) using a in-house software package called PCDetpack (Lewis *et al* 2000). This software can import BSL file format data, perform specialist image processing operations (such as image intensity corrections for a variable incident x-ray beam flux) and record the data as a DET file which has a similar structure to the BSL format but allows a greater range of calibration operations.

6.2 Artifacts of the SAXS Imaging System

6.2.1 Image Formation & Noise

The detector builds up a digital image of a SAXS pattern by recording the counts of scattered photons at different positions on its surface. Stray photons scattered by objects other than the sample of interest may also be recorded by the detector. These stray photons add extra intensity counts on the image and can therefore be considered as additive noise. The counting nature of the detection process suggests that this noise can be modeled according to a Poisson statistical distribution. Detector noise can also be produced by internal electronic components. Lewis (1994) states that the noise (both background and internal) of the multi-wire proportional counter detectors used at Daresbury synchrotron radiation source is so low ($\approx 10^{-5}$ counts pixel^{-1} s^{-1} for a 1000 x 1000 pixel image) that the noise is swamped by diffuse scatter from the sample. Therefore, for the diagnosis of breast cancer using SAXS images the noise levels are assumed to be negligible.

6.2.2 Detector Point-Spread Function

The sample is imaged using an incident x-ray beam of finite width which has an influence on the observed SAXS image. The SAXS image observed is a convolution between the image of the x-ray beam on the detector plane, $I_{\mathrm{beam}}(\mathbf{h})$ and the 'ideal' image that would be obtained if the sample using a 'point-source' x-ray beam, $I_{\mathrm{ideal}}(\mathbf{h})$ (Le Flanchec *et al* 1996). That is, $I_{\mathrm{obs}}(\mathbf{h}) = I_{\mathrm{beam}}(\mathbf{h}) * I_{\mathrm{ideal}}(\mathbf{h})$, where $*$ denotes the convolution operation. The convolution of the 'ideal' SAXS image with the image of the incident x-ray beam filters and reduces the resolution of the observed SAXS image. The impact of the x-ray beam profile on the subsequent image analysis must be carefully considered. Comparisons of data collected between different experiments (such as that of Lewis *et al* (2000) and Round (2006) which use different beam profiles) must be done cautiously and the the effects of the convolution removed as much as possible.

6.2.3   Beam-Stop

The high incident x-ray fluxes used at synchrotron sources can damage multi-wire proportional counter detectors and result in incorrect functioning (Lewis 1994). As a result the primary X-ray beam must be attenuated using a lead beam-stop, which is visible in the centre of the SAXS image as a rectangular block of (approximately) uniform intensity. The beam-stop removes a proportion of the low- resolution (> 100 nm) information on the structure of the sample and excludes a proportion of potentially useful diagnostic information.

6.2.4   Capillary Flare

The tissue samples are placed in capillary tubes and stacked on a multiple sample holder in order to facilitate imaging. Careful horizontal alignment was performed using positioning motors that are attached to the sample holder to ensure that the sample was fully irradiated by the beam. Incorrect horizontal alignment results in an artifact known as 'capillary flare'. Figure 6.1 displays an example of this artifact, a sharp 'streak' across the equatorial direction of the image. This artifact destroys image information that could otherwise be used for tissue diagnosis. Fine control of the horizontal motors and a short 'test' exposure of the sample (in order to visually inspect the image) minimised the occurrence of this artifact.
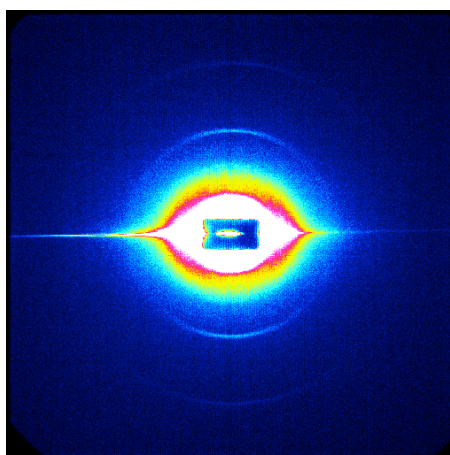

Figure 6.1: A breast tissue SAXS image with the capillary flare artifact present (the large intensity streak across the centre of the SAXS image).

6.2.5   Insufficient Exposure

An exposure time of 300 seconds was required in order to capture sufficient signal intensity of scattered photons from the structure of collagen. Images of poor quality are captured when the exposure time is insufficient. Both biological and experimental factors can be responsible for this problem. A tissue sample might have little collagen present which could lead to a low scattered photon flux. The structure of the collagen in the tissue sample can also be damaged or destroyed due to poor sample extraction and handling. Storing the samples for too long (on the order of months) in liquid nitrogen at -80 ^{0}C can cause some damage to the specimen structure, as might poor handling when placing the samples in capillary tubes in preparation for imaging. The samples can also suffer damage from the heat generated by the incident x-ray beam during the imaging process. To minimise the risk of such damage, a relatively large piece of sample is placed in the capillary tube (the incident x-ray beam has dimensions of approximately 1 x 1 mm^2 compared to the sample which has dimensions of approximately 1 mm wide x 10 mm long) and the tube oscillated through the x-ray beam throughout the exposure. This imaging technique minimises the risk of heat damage to the sample and produces an averaged SAXS image of the structure throughout the sample. Producing an averaged SAXS image has the additional benefit of minimising the influence of any one region of tissue in the sample. For instance, the tissue sample might be very heterogeneous consisting of regions of both adipose and collagenous tissue. Visual inspection of the specimen prior to imaging and other good experimental practices (such as watching the total scattered photon count on the detector as the sample is scanned) may not be able to accurately detect these different regions. Without the sample averaging technique, a poor quality SAXS image may simply result from sampling error in which the adipose region of the sample is incorrectly imaged by chance.

6.2.6   Cosmic Ray Artifact

Cosmic rays can also register on the detector and produce an intensity artifact in the SAXS image. Such an event is very rare but it has been know to occur during the SAXS imaging experiments used to collect this data. Foreign objects such as small pieces of shattered capillary glass can also influence the appearance of SAXS images. These artifacts are often easily identified (using regularly updated image monitors) and can often be corrected by running the exposure again.

6.2.7   Summary

The imaging process influences the quality of the SAXS image data that is obtained. It is important to identify and understand such processes in order to guard the quality of the data used in the development of a diagnostic model. Many of the problems encountered in the imaging process can be overcome using well designed experimental techniques and expert supervision. The SAXS image data collected in this thesis was performed by an experienced team (whose origins can be traced back to the work of Lewis *et al* (2000)) whose sample storage and handling techniques followed strict, well-understood protocols. Details of such protocols can be found in Round (2006). Trained scientists readily identified the presence of capillary flare and understood how to re-position the sample to mitigate its effect. The influence of the point-spread function (image of the incident x-ray beam) was controlled by the use of skilled scientists and technicians who could obtain consistent x-ray beam profiles across experiments. These counter-measures are not perfect and image artifacts (and other errors) can still make it through to the analysis stage. In this thesis, the data was carefully screened prior to analysis for any such errors using visual inspection, five number summaries and the careful study (varying bandwidth) of the density estimates of the intensity values in the image. The data (as supplied) was deemed to be of sufficient quality to develop a breast cancer diagnostic model.

6.3 Processing of SAXS Images to Correct for Experimental Effects

The raw SAXS images obtained from the synchrotron need to be corrected for a number of effects prior to further analysis. Those relevant to the analysis of breast tissue structure are the removal of detector wire artifacts, adjustment for detector flat field response, correction of the variation in the incident x-ray flux throughout the course of imaging, removal of background scattering and indexing of pixel location with object size. The details of these methods are poorly documented but they are well known and used by the x-ray diffraction community to pre-process SAXS images, this account is based primarily on the work of Round (2006). The data used in this thesis were supplied after these pre-processing operations had been done by Dr. K. Siu of Monash University, Melbourne, Australia.

6.3.1 Remove Wire Structure Algorithm

The wire-based structure of the multi-wire proportional detector introduces high-frequency line artifacts in the SAXS image. The remove wire structure algorithm was applied to remove these artifacts. A region of the image away from the beam-stop, that did not contain capillary flare artifacts or scattering rings was selected and the spectral components examined. High spatial frequency components of the image were identified and removed using a high-frequency filter. The form of this filter is 'fine-tuned' to the image at hand and assessed visually for optimum artifact removal. Figures 6.2 (a) and (b) display the result of the remove wire structure algorithm for a breast tissue SAXS image.

Figure 6.2: Removal Wire Structure Processing.
a) **Before**   b) **After**
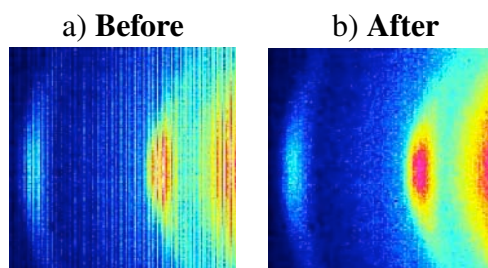
The subjectivity involved in this pre-processing technique is of great concern. Selection of the region used to extract the high spatial frequency components corresponding to the wire artifacts

is an issue of contention. The method relies on the user being able to visually identify image regions that do not contain scattering rings or artifacts. This method works under the inherent assumption that such features can be identified visually but given the high dynamic range of these images this may not be the case. User adjustments to the frequency filter are also of great concern, slightly different filters can be applied to different images and any apparent differences between images might be a result of such adjustments to the filter.

6.3.2 Detector Flat Field Response Correction

An image of the detector (*flat-field*) response to an iron 55 radioactive source was recorded during the imaging process and was used to correct for subtle differences in the detector response. Upon normalisation (achieved by dividing all image pixel magnitudes by the maximum magnitude of all of the image pixels) this flat-field response image can be used to suppress detector noise and other detector inhomogeneities. The correction is performed by dividing the pixel magnitudes in the SAXS image with the corresponding pixel magnitudes in the flat-field response image. This technique is often used in SAXS imaging and it might be very useful for the correction of location dependent differences in detector response. If detector noise is deemed to be a problem (even though Lewis (1994) indicated that it was not a problem for certain detectors used in SAXS imaging) then more sophisticated noise removal algorithms (such as those in Chapter 4) seem to be more appropriate than the detector flat-field response method.

6.3.3 Beam Intensity Normalisation

The incident x-ray beam at the synchrotron varies in intensity over time. This variation is due to both base usage and natural beam decay, as such the x-ray beam requires regular boosting. The beam intensity variation at the Daresbury synchrotron radiation source is on the order of a factor of two over a twenty-four hour period. This variation can influence the measured intensity of the image and can make comparisons between images collected under different beam intensities very difficult. Attenuation of the x-rays within the sample is another factor that influences the observed SAXS image intensity. Differences in tissue composition, quality and the amount of tissue packed in the capillary tube lead to such differences in sample attenuation. These variations are corrected for by the calculation of a normalisation factor. This factor is obtained by taking the reciprocal of the beam intensity as registered by an ion chamber post sample. In some cases a semi-transparent beam stop allowed the recording of the post-sample x-ray beam directly using

the SAXS image. The observed SAXS image is then multiplied by this correction factor.

Beam intensity normalisation is an important and effective image correction procedure. One complicating factor has been the use of different devices to record the transmitted x-ray beam intensity. The use of different semi-transparent beam stops in different x-ray imaging sessions is a problem in the analysis of the SAXS images (Round 2006). These semi-transparent beam stops have different levels of x-ray attenuation which influences the accuracy in the estimates of the normalisation factors. Comparison of the intensities in the SAXS images between different imaging sessions will be influenced by these errors.

6.3.4   Sample Holder Background Removal

Scattering from the capillary tube used to hold the samples contributes to the intensity of the SAXS images. To correct for this additional scattering component, a series of SAXS images from blank capillary tubes were recorded and an average image found. This average image was then subtracted from the SAXS image of the capillary loaded with the breast sample. The method assumes that the effects of the SAXS image of the capillary tube and the SAXS image of the tissue sample are additive. Subtraction of an average 'representative' SAXS capillary image also assumes little variation in the image between different capillaries. An alternative technique would have been to image the blank capillary prior to loading and imaging the sample. This approach would allow for a more precise measurement of the capillary tube scatter relevant to that sample.

6.3.5   Position Indexing

The position of each element (pixel) in the image array was indexed to the physical size of the scattering object. Pixels in the centre of the image correspond to larger object sizes and pixels towards the edge of the image to smaller object sizes. Indexing is achieved using a calibration SAXS image based on a sample of rat-tail collagen. This image, displayed in Figure 6.3 consists of a series of scattering rings along the meridian. These rings are associated with the axial D-repeat structure of fibrillar collagen, whose position can be indexed to an object size of $D = 67$ nm, when hydrated (Round 2006). The indexing operation is useful for those studies that extract physical parameters for classification, it is additional information not immediately available from the intensity values of the SAXS image.
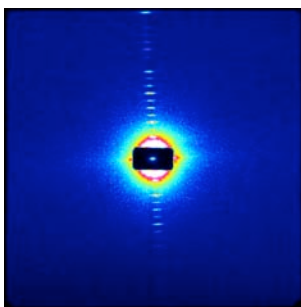
Figure 6.3: A false-colour SAXS image of hydrated rat tail collagen as obtained at Daresbury synchrotron radiation source station 2.1 with a camera length of 6.25m. Note the scattering rings on the meridian that correspond to the axial D-repeat structure.

6.4 Summary and Conclusions to Pre-processing of SAXS Images

The SAXS images are an array of real numbers that record the x-ray photon intensity as registered on the detector. The imaging process influences the record of the SAXS pattern (which is a physical electro-magnetic field). Factors such as detector noise and the point-spread function alter the appearance of the image. A number of pre-processing operations are performed in order to correct for well known imaging variables, it is important to understand these operations and any effects that they may have on the subsequent analysis. Improvements to the remove wire structure algorithm, beam intensity normalisation and the capillary scattering removal procedures are needed. Nonetheless, the current pre-processing operations are considered state of the art in the field and have been used to correct the data to greatest extent possible.

Chapter 7

Adaptive Transformation of SAXS Images

This chapter is the first of three chapters (Chapters 7, 8 and 9) that describe the original contributions of this thesis. The overall strategy of this project is depicted in Figure 7.1 and consists of a sequence of logical steps based upon the established knowledge to develop analytical methods with which to infer the presence of cancer in a tissue sample. The scientific evidence that was presented in Chapter 2 suggests a change in the structure of collagen when breast cancer is present. Imaging of this structure using X-ray diffraction techniques yields a SAXS pattern, which is an electro-magnetic field that varies both in space and time. The SAXS pattern is recorded by a detector and stored in a digital format that is referred to as a 'SAXS image'. It is these SAXS images that are analysed using statistical models to infer tissue state. Chapter 7 develops digital image processing methodology that filters the SAXS images in order to facilitate statistical modeling. Chapter 8 describes the statistical models that can be used to infer diseased (cancer) tissue state based upon the pre-filtered SAXS images. Application of these digital image processing and statistical modeling techniques to actual test data is then assessed in Chapter 9.
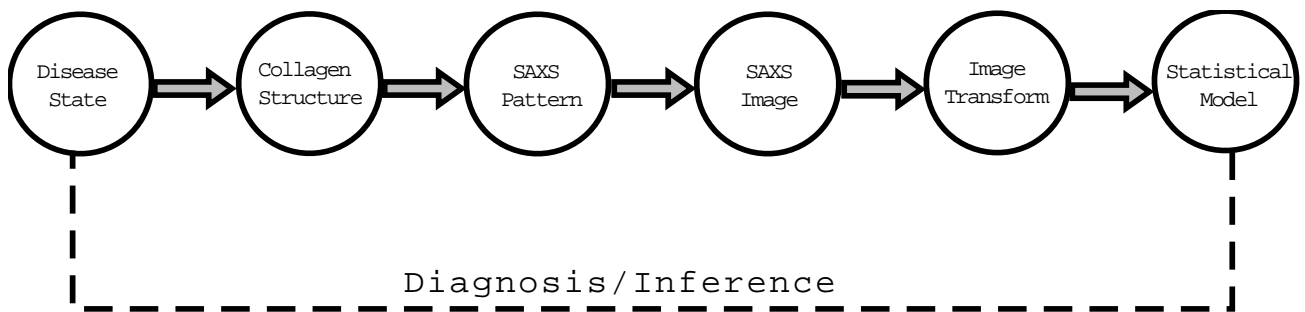

Figure 7.1: Strategy for the diagnosis of breast cancer using SAXS images of collagen structure.

The raw SAXS image data might contain many features and the use of a single filter function (for example a wavelet filter) might extract one feature at the expense of another useful feature. These undetected features may well contain vital information that enables accurate detection of disease. An alternative filter function could be used to extract such a diagnostic feature, but it may also come at the cost of the loss of other useful diagnostic features. As explained in Chapter 2, SAXS images of breast tissue contain a variety of features. It is unlikely that these features can be successfully extracted using a single image processing filter function. The use of a library of filter functions overcomes this difficulty by allowing the use of more than one function in the processing of the image. Features arise in different scale-locations. For each scale, the entire image data is analysed using each filter in turn and after each individual analysis, the significant coefficients are recorded. At the end of this process, there are sets of coefficients derived using each of the filter functions in the library. The next step is to reduce this collection to a single set of coefficients. The coefficients included in this set are those that correspond to the 'best' filter for each scale and location. The 'best' filter at each scale-location is determined using Bayes" Rule.

The content of this chapter concerns a transform which is composed of filter functions from a library of such filter functions which are designed specifically to capture different features from the SAXS image. The remaining of this chapter is organised as follows, Section 7.1 elaborates on the rationale for the image processing and transformation applied to the SAXS image. A case is made for the processing of the images using scale-space transforms such as wavelets. An adaptive scale-space image transformation that uses a library of filter functions with which to represent the image is presented in Section 7.2. The specific filter functions to be included in this library are examined in Section 7.3, with five new filter functions developed for the purpose of processing the SAXS images. The selection of the most appropriate filter function from the library with which to transform the image is discussed in Section 7.4. Since the primary objective of this project is successful classification of tissue state, filter functions from the library are compared based upon their ability to separate data belonging to different tissue groups in a training data set. Upper and lower bounds on the Bayes' probability of error for classification (calculated with the Jennsen-Shannon divergence measure) are used as a criterion to compare statistics based upon the coefficients of SAXS images processed with different functions from the library. At each scale and location of the transform, the filter function with the smallest probability of error for classification is selected as an analysis filter. The chapter is reviewed and concluded in Section 7.5.

The original contributions to the mathematical and statistical sciences found in this chapter are:

(i) An adaptive image transformation that allows the image to be represented across scale and space as a combination of different filter functions (Section 7.2) .

(ii) The development of five new filter functions that have properties that might be useful in image analysis (Section 7.3).

(iii) Combining adaptive image transformation, the selection of the most appropriate filter functions from a library and classification into the same framework using Bayes' probability of error for classification (Section 7.4).

7.1   Rationale for Image Transformation

In Section 4.3 and in Chapter 5, Fourier and wavelet methods were selected as one of several avenues of further investigation in this thesis. The Fourier methods focus on the extraction of *spatial* frequency information in the image and appear to be very useful in the analysis of regularly repeating structures in the SAXS image. The axial scattering ring features might be amenable to Fourier analysis. In contrast, the wavelet methods were very appealing because they allow modeling of the image over scale and space. This scale-space transformation provides an inherent hierarchical structure to the analysis, which matches well with the hierarchy of collagen structures within the breast tissue that produced the SAXS image.

Wavelet methods have further appeal in the analysis of SAXS images. Vidakovic (pages 9-10, 1999) demonstrated that wavelets can be used to separate a high-frequency signal that was super-imposed on a low frequency signal. This result suggests that wavelets can be used to decompose the signal into components from different structures of tissue across a variety of scales. The image information associated with the quasi-hexagonal packaging of collagen fibrils might be able to be separated from the information associated with the axial D-repeat structure of collagen fibrils. The separation of image information relevant to different physical features is a motivating reason for the use of the wavelet or other scale-space transforms in SAXS image analysis.

Vidakovic (1999) states on page 19, "*Being self-similar themselves, wavelets are especially apt to describe phenomena exhibiting self-similarity in different scales*". Wavelet (and related) methods are sensitive to self-similarity within an image and might be able to detect any departures from self-similarity. This could be very relevant in the diagnosis of breast cancer using SAXS images. Section 2.3 presented evidence for a breakdown in the structure of fibrillar collagen when breast tissue is malignant. Any similarity in the structure of collagen over several length scales could be lost as a result of such changes. The axial D-repeat structure of fibrillar collagen is likely to be self-similar over a range of scales. Figure 7.2 displays a diagram illustrating the periodic low and high density regions along a collagen fibril that are associated with the axial D-repeat structure. Section 2.2 explained that the axial D-repeat structure is of period 64 to 67 nanometres and is produced by the alignment of collagen molecules into a single fibril. The length of this fibril (when mature) is typically thousands of nanometres (Kadler *et al* 1996). Inspection of the diagram in Figure 7.2 over a variety of length scales suggests that the fibril has a self-similar structure. There is every reason to believe that this self-similarity will be contained

within the SAXS image of breast tissue collagen structure and that wavelet analysis might be able to detect any breakdown in this self-similarity associated with malignant conditions. The ability of wavelets (and other scale-space methods) to detect self-similar structures in the image data is another good reason for their inclusion in this thesis.
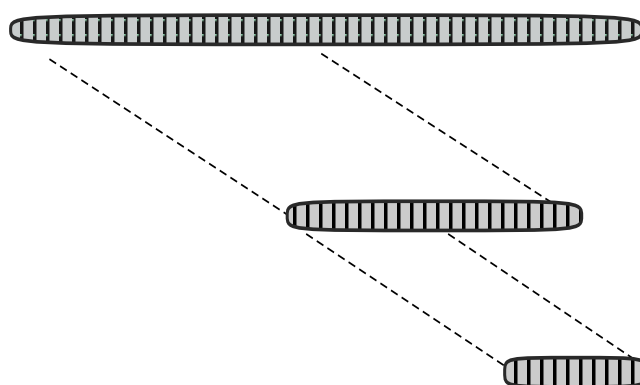

Figure 7.2: Diagram illustrating the self-similarity of the collagen fibril that is a result of the axial D-repeat structure. Examination of the fibril over a variety of scales reveals a structure that looks exactly the same. Note: high density regions are shaded in black.

7.2 Adaptive Transformation of SAXS images

This section develops a scale-space transformation of SAXS images. This transformation is simple in structure but seems well suited to the task at hand. Chapter 5 examined a range of image transformations such as the

 (i) The windowed Fourier transform (Section 5.1.2).

 (ii) The separable discrete wavelet transform (Section 5.2.2).

(iii) The steerable pyramid transform (Section 5.3.1).

Of particular interest were the adaptive image transformations (such as the matching pursuit algorithm of Section 5.4.2) that allow an image to be modeled with a variety of filter functions. The use of a library of filter functions in these transformations allows the selection of the image representation best suited to the task at hand. It also offers a way to combine Fourier and wavelet analysis functions in the same setting.

The adaptive image transformation approach will be adopted in this thesis to structure the data into an amenable multi-scale format. Candès & Donoho (2000) and Donoho (2001) argue against adaptive transforms on the grounds that non-adaptive transforms such as curvelets can achieve similar (optimal) m-term linear approximation rates of images. That is the number of medium to large magnitude coefficients is similar for both transforms when studying images with edges. While this may be true, the position of Candés & Donoho (2000) and Donoho (2001) fails to appreciate the flexibility offered by an adaptive transform. The library of filter functions can be selected at will and the cost functional selected for the task at hand. For instance, a sparse (optimal m-term linear approximation) representation may not be the objective of the study. Instead the researcher might be interested in the fractal properties of the image at a particular scale and the smooth trend at another. Rather than a sparse representation, one that separates the fractal and smooth signals seems far more important in this example. The focus of this thesis concerns the accurate classification of tissue state using SAXS images. Image transformations are used in order to develop a statistical model that yields accurate classification. The entire purpose of the transformation is to find a representation that produces superior classification results and structures the data in a form amenable to statistical analysis; this may or may not be the sparest image representation. For this reason, the adaptive image transformation approach is preferred over 'off-the-shelf' solutions such as the curvelet and contourlet transforms.

A drawback of current adaptive image transforms such as the matching and basis pursuit algorithms is their high computational complexity. This hinders practical implementation when large data sets are involved. The approach adopted in this thesis is a practical compromise. The transform coefficients are calculated rapidly for each function in the library. This creates a very redundant image representation and a potentially large storage requirement in a digital format. The image transform proposed in this thesis which we will refer to as the *adaptive image transform* is,

$$I(x,y) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{J}_{i,\boldsymbol{\gamma}} \, \xi_{i,\boldsymbol{\gamma}}(x,y) \, < \xi_{i,\boldsymbol{\gamma}}(x,y), I(x,y) > \tag{7.1}$$

where $\xi_{i,\boldsymbol{\gamma}}(x,y)$ is the ith out of the N filter functions in the library $\mathcal{L}$ and is indexed by a parameter vector, $\boldsymbol{\gamma}$. The particular form of the parameter vector, $\boldsymbol{\gamma}$ is dependent on the filter function used. For instance, a wavelet filter function, $\xi_{1,\boldsymbol{\gamma}_1}(x,y)$, could be indexed by scale paramater $a \, \epsilon \, \mathbb{R}$, location parameters $(b_x, b_y) \, \epsilon \, \mathbb{R}^2$ and orientation parameter $\theta \, \epsilon \, (0, 2\pi]$. The appropriate parameter vector in this case is $\boldsymbol{\gamma}_1 = (a, b_x, b_y, \theta)$. In contrast a Fourier filter function, $\xi_{2,\boldsymbol{\gamma}_2}(x,y)$

indexed by parameters $k \ \epsilon \ \mathbb{R}$ and $l \ \epsilon \ \mathbb{R}$ has a parameter vector $\boldsymbol{\gamma}_2 = (k, l)$. The transform co-efficients are found by the inner product of the filter function $\xi_{i,\boldsymbol{\gamma}_1}(x, y)$ with the image $I(x, y)$. That is via $< \xi_{i,\boldsymbol{\gamma}}(x, y), I(x, y) >$. Coefficients have largest magnitude when the features in the image correspond to the features of the filter function. By designing filter functions of different shapes (and mathematical properties), the image can be analysed for different types of features. Whilst some features may not be detected by a particular filter function, another filter function in the library is likely to find it.

The matrix $\mathbf{J}_{i,\boldsymbol{\gamma}}$ indicates whether the function $\xi_{i,\boldsymbol{\gamma}}(x, y)$ is included in the transform for parameters $\boldsymbol{\gamma}$, It is zero if the function is not included in the transform for these parameters and is one otherwise. Multiple functions from the library can also be assigned when they have the same parameters in the parameter vector $\boldsymbol{\gamma}$. Selection of the most appropriate function for each value of these parameters depends on a cost functional that is discussed in detail in Section 7.4.

In summary, the adaptive image transform represents an image as a linear combination of filter functions from a library of such functions, these functions can vary across parameters such as scale, position and direction. Within a transform scale, the functions used in the transform need not be of the same type across locations. For instance, a windowed Fourier and wavelet function are allowed to co-exist within the same scale.

As the focus of this thesis is on analysis, the inverse transformation is not required. Nonetheless it is noted that in certain cases recovery of the original image is possible from the coefficients. Successful recovery of the original image requires that the image information is preserved in the transformation process. One example of this is when $\mathbf{J}_{i,\boldsymbol{\gamma}} = \mathbf{1}$ for at least one of the i out of the N filter functions and when this filter function has a convenient algorithm to perform the inverse transformation. There is no need to recover the original image in this thesis and all that is required of the transform is that important image information useful for classification is extracted and put in a form useful for a statistical classification model. This is achieved by the careful selection of the matrix $\mathbf{J}_{i,\boldsymbol{\gamma}}$ via a cost functional based upon the probability of producing an error in data classification.

7.3 The library of transform functions

The library of functions in the transform library requires special attention, in particular families of functions that vary across scale, location and direction such as wavelets are of interest. The filter functions proposed in this thesis are similar to wavelets in that they can be designed to vary in scale, location, orientation and phase. The functions are of the form,

$$f_{\boldsymbol{\gamma}}(x,y) = f_a(x,y)\sin\left\{\omega_0[x\sin(\theta) + y\cos(\theta)] + \phi\right\} \tag{7.2}$$

for scale parameter $a \in \mathbb{Z}^+$, frequency parameter ω_0, location parameters (x,y), orientation parameter $\theta \in [0, 2\pi)$ and phase parameter $\phi \in [0, \frac{\pi}{2}]$. The form of the function $f_a(x,y)$ is of central interest and a collection of such functions can be stored in a library $\mathcal{L}$ of the transform. The filter function, $f_{\boldsymbol{\gamma}}(x,y)$, can then be created from the library function, $f_a(x,y)$, when the parameters, $\boldsymbol{\gamma} = (a, \omega_0, \theta, \phi)$ are specified using equation (7.2). The thesis will include both well-known as well as custom-designed filter functions in the transform library.

Functions sourced from the literature included in the library of filter functions are:

(i) Gabor functions (equation 5.3)

(ii) Mexican Hat wavelet (equation 7.3)

(iii) The Pet Hat wavelet (equation 7.4).

The Gabor function was defined in Section 5.1.2, it is the function used in the windowed Fourier transform. The Mexican hat is an isotropic wavelet that is rotation invariant and is defined as,

$$\Psi(\boldsymbol{\omega}) = \frac{|\boldsymbol{\omega}|^2}{a^2} \exp\left(-\frac{|\boldsymbol{\omega}|^2}{2a^2}\right) \tag{7.3}$$

for scale parameter $a \in \mathbb{R}^+$ and spatial frequency parameters $\boldsymbol{\omega} = (\omega_x, \omega_y)$ (equation 3.8, page 101, Antoine *et al* 2004). This wavelet has excellent resolution in the spatial coordinates but poor selectivity in both scale and direction (page 570, Arnéodo, Decoster & Roux 2000). Analysis using the Mexican hat wavelet will provide a set of coefficients that allows identification of the location in the image where a feature occurs but will provide limited information as to both its scale (for instance, whether it is sharp or diffuse) and orientation.

The Pet hat wavelet is an alternative to the Mexican hat that has better resolving power in scale. It is defined as,

$$\Psi(\boldsymbol{\omega}) = \begin{cases} \cos^2\left(\frac{\pi}{2}\log_2\frac{|\boldsymbol{\omega}|}{2\pi a}\right) & \text{for} & \pi < |\boldsymbol{\omega}|/a < 4\pi \\ 0 & \text{for} & |\boldsymbol{\omega}|/a < \pi \text{ and } |\boldsymbol{\omega}|/a > 4\pi \end{cases} \quad (7.4)$$

(adapted from equation 3.11, page 101, Arnéodo, Decoster & Roux 2000). It has superior ability in resolving the scale of an image feature because it resembles an annulus in the frequency plane, allowing only those frequencies between $(\pi < |\boldsymbol{\omega}|/a < 4\pi)$ to be included in the analysis. This is in contrast to the Mexican hat wavelet which is a product of a quadratic and an exponential term therefore allowing a range of frequencies to influence the analysis at a particular scale.

Five new custom-designed filter functions are introduced in this thesis and are included in the transform library. All of these custom-designed filters could be described as wavelet functions in two-dimensions as they can be shown to satisfy the modified admissibility condition (equation 5.13). These functions have been designed for specific tasks and allow the examination of the image from different viewpoints. The functions introduced in this thesis are:

 (i) Gamma filters - designed to emphasise smooth trends as well as to separate jumps in image intensity from changes in image phase.

 (ii) Mellin filters - designed to be a bandpass filter, allowing only a select window of frequencies into the analysis at a particular scale.

(iii) Chebyshev filters - designed to detect and extract smooth trends that resemble Bessel functions.

 (iv) Zeta filters - designed to detect power law trends in the SAXS data.

 (v) Witch's Hat filters - designed to apply a 'frequency comb', extracting both smooth and high frequency components.

Sections 7.3.1- 7.3.4 will discuss each of these functions and their advantages in depth. All filter functions will be developed at first in one-dimension and then extended to two-dimensions. Thus the library $\mathcal{L}$ consists of the three established functions in equations (5.3), (7.3) and (7.4) as well as the five new proposed filter functions above.
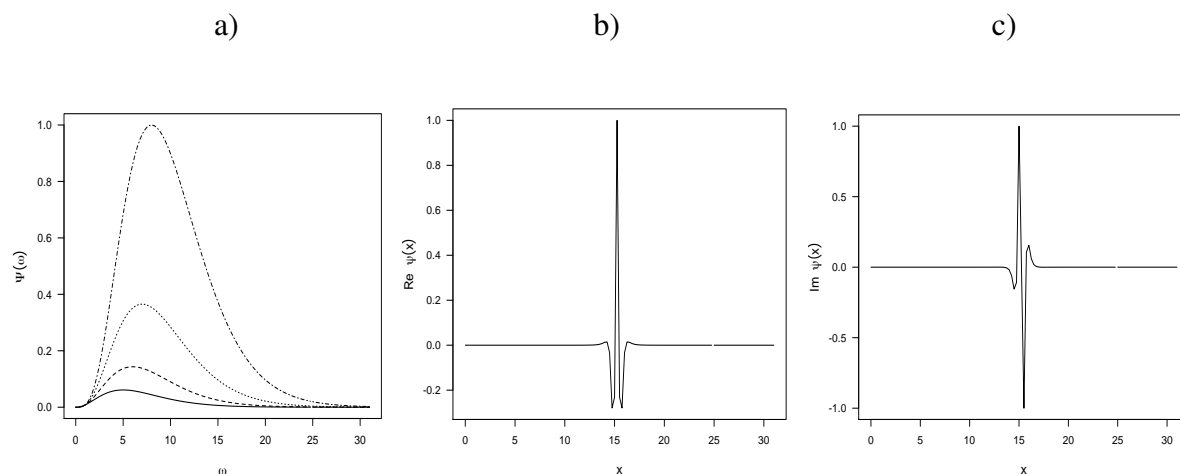
7.3.1   The Gamma Filter

Figure 7.3: The Gamma filter in one-dimension: (a) in the Fourier domain for scales $a = 5\text{-}8$, ($a = 5$ solid line, $a = 6$ long dash, $a = 7$ dots and $a = 8$ dot-dash), (b) the real component and (c) the imaginary component of the Gamma filter in the space domain for scale $a = 8$.

Figure 7.3 displays examples of the Gamma filters in both the space (x) and Fourier (ω) domains. These filters will be discussed in the following paragraphs. The Gamma filter is motivated by the need for a filter that has low energy at high (fine) scales and for which priority is given to smooth trends. It allows fine-tuning of the response with respect to scale such that the majority of high frequency spatial oscillations are supressed and only the largest of intensity edges contain significant energy. This type of filter might be useful in the analysis of SAXS images because it might allow better separation of the *smooth* background intensity trends in the image from the scattering ring features. It might also assist in isolating the background scatter from other smooth features in the image. The Gamma filter is also designed to detect both jumps and variations in image intensity. It can achieve this because of its response to spatial frequency with respect the scale and by the shape of the filter. Both of these properties influence the magnitude of the adaptive image transform coefficients (when the Gamma filter is used) at a particular scale.

Denote $\Psi(a, \omega)$ as a filter function that depends on frequency parameter $\omega \in \mathbb{R}$ and scale parameter $a \in \mathbb{R}^+$. The admissibility condition of equation (5.13) places requirements on those filter functions acceptable for wavelet analysis. The specific requirement of the admissibility condition is that the integral $\int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega$ is finite. Careful substitution of a function $f(a, \omega)$ (where parameter $a \in \mathbb{R}^+$ represents a scale term) in place of the $\frac{|\Psi(\omega)|^2}{\omega}$ term in the admissibility condition can be used to guarantee this condition is met for some general function dependent on both scale and frequency parameters. By equating $f(1, \omega) = \frac{|\Psi(\omega)|^2}{\omega}$ (for scale parameter $a = 1$) it is also possible to find the function $\Psi(1, \omega)$. This filter function $\Psi(\omega)$ is the *mother function* from which an entire family of similar filter functions can be formed using scale parameter $a \in \mathbb{R}^+$ and translation parameter $b \in \mathbb{R}$. This can be achieved in a manner similar to that in the integral wavelet transform of Section (5.2.1). Denote the space-domain representation of the mother Gamma filter as $\psi(a, x)$ (which is the inverse Fourier transform of $\Psi(a, \omega)$), then the *family* of one-dimensional Gamma filters is given by, $\psi_{a,b}(x) = a^{-1}\psi[a^{-1}(x - b)]$. In image analysis only the positive spatial frequencies, $\omega > 0$ are of interest so we can omit the negative frequencies $\omega < 0$ keeping in mind that the filter functions developed are designed to be suitable for a two-dimensional, not a one-dimensional wavelet analysis. Thus the admissibility condition can be slightly modified to be,

$$C_\psi = \int_0^{+\infty} \frac{|\Psi(a, \omega)|^2}{|\omega|} d\omega \quad \text{for } \omega \in \mathbb{R}^+, \ a \in \mathbb{R}^+ \tag{7.5}$$

(adapted from page 112, Prasad & Iyengar 1997).
Substituting $f(a, \omega) = \frac{|\Psi(a,\omega)|^2}{|\omega|} = \omega^{a-1}e^{-\omega}$ gives,

$$C_\psi = \int_0^\infty \omega^{a-1}e^{-\omega} d\omega \quad \text{for } a \in \mathbb{Z}^+ \tag{7.6}$$

which is a Gamma function of scale parameter, a so that

$$\begin{aligned} C_\psi &= \Gamma(a) \\ &= (a-1)! \ < \ \infty \end{aligned} \tag{7.7}$$

which is finite for $a \in \mathbb{R}^+$. Noting $\frac{|\Psi(a,\omega)|^2}{\omega} = \omega^{a-1}e^{-\omega}$ for $\omega > 0$, $a \in \mathbb{R}^+$, we can solve for $\Psi(a, \omega)$ to find the *Gamma filter*,

$$\Psi(a, \omega) = \sqrt{\omega^a e^{-\omega}} e^{i\varphi(\omega)}. \tag{7.8}$$

as defined in Fourier space. In this thesis, we set $\varphi(\omega) = 2\pi$, so that $e^{i\varphi(\omega)} = 1$. The inverse Fourier transform of $\Psi(a, \omega)$ is used to find the spatial version $\psi(a, x)$ version of the Gamma filter. The inverse Fourier transform of $\Psi(a, \omega) = \omega^{a/2}e^{-\omega/2}$ is,

$$
\begin{aligned}
\psi(a, x) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Psi(\omega)e^{i\omega x} d\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \omega^{a/2} e^{\omega(ix - 1/2)} d\omega.
\end{aligned}
\tag{7.9}
$$

Using *Mathematica* software (Wolfram Research 2008), the integral in equation (7.9) is

$$
\begin{aligned}
\psi(a, x) &= -\frac{1}{2\pi} \frac{\Gamma[\frac{a}{2} + 1, -(ix - \frac{1}{2})\omega]}{(\frac{1}{2} - ix)^{(\frac{a}{2}+1)}} \\
&= \kappa(x)^{-1}\Gamma[\frac{a}{2} + 1, -(ix - \frac{1}{2})\omega]
\end{aligned}
\tag{7.10}
$$

where $\kappa(x) = -2\pi(\frac{1}{2} - ix)^{\frac{a}{2}+1}$, $x \neq \frac{1}{2i}$. Thus the spatial version of the Gamma filter can be described as an incomplete Gamma function $\Gamma[\frac{a}{2} + 1, -(ix - \frac{1}{2})\omega]$ that is weighted by a polynomial, $\kappa(x)$. Note that the incomplete Gamma function is (in general) defined as,

$$
\Gamma(n, x) = \frac{1}{\Gamma(n)} \int_0^x e^{-t}t^{n-1} dt
\tag{7.11}
$$

which is a truncated form of the ordinary Gamma function, $\Gamma(x)$ whose integral limits range from $[0, x]$ instead of $[0, \infty)$ (equation 6.5.1, page 260, Abramowitz & Stegun 1972). The smaller the value of x, the smaller the 'weight' that the integrand ($e^{-x}x^{n-1}$) contributes to the incomplete Gamma function and hence the Gamma filter. The polynomial $\kappa(x)$ moderates the influence of the incomplete Gamma function $\Gamma[\frac{a}{2} + 1, -(ix - \frac{1}{2})\omega]$ according to the scale parameter, a, the greater the magnitude of the scale parameter the greater the degree of the polynomial in $\kappa(x)$. This polynomial can also be viewed as a filter, the greater the scale, a, the greater the degree of the polynomial and hence the more sensitive the Gamma filter $\psi(a, x)$ is to x the spatial position parameter.

Observe that the Gamma filter $\psi(a, x)$ has both a real and imaginary component. Both of these components can be viewed as extracting different features from the data. The coefficients from a Gamma filter analysis can be expressed in polar form. Denote $d_a(x) = < f(x), \psi(a, x) >$ as the coefficients formed by taking the inner product of the data (expressed as a function $f(x)$)

with the Gamma filter $\psi(a, x)$. The coefficients can be represented in *modulus* ($|d_a(x)|$) and *phase* ($\theta_a(x)$) format via,

$$d_a(x) = |d_a(x)|e^{i\theta_a(x)}. \tag{7.12}$$

The modulus, $|d_a(x)|$, provides information on the presence of periodic structures in the data and is conveyed by the real component of the coefficients. In contrast, the phase $\theta_a(x)$, provides information on the relative position of periodic structures in the data and is contained within the imaginary components of the coefficients. Therefore, the Gamma filter $\psi(a, x)$ is designed to separate out the modulus and phase information in the data.

In order to plot both $\Psi(a, \omega)$ and $\psi(a, \omega)$, the magnitude of the filter is calculated at a large number of points at uniform spacing and over a wide range. For convenience, calculations were performed in the Fourier domain using $\Psi(a, \omega) = \sqrt{\omega^a e^{-\omega}}$ over the range $0 \leq \omega \leq 35$ radians. Calculations using the location domain version of the Gamma filter, $\psi(a, x)$ would involve the incomplete Gamma function $\Gamma[\frac{a}{2} + 1, -(ix - \frac{1}{2})\omega]$ and hence the calculation of definite integrals which can be computationally more demanding. The estimate of the Fourier domain representation of the Gamma filter $\Psi(a, \omega)$ is displayed over a range of scales (a = 5-8) in Figure 7.3(a). This figure indicates that as the scale parameter, a, increases the height and spread in frequency (ω) of the wavelet increases. The inverse discrete Fourier transform can be applied to the discrete set of points that was used to estimate $\Psi(a, \omega)$. The result is a discrete set of points that provides an estimate of $\psi(a, \omega)$. By the sampling theorem (Section 4.1), this estimate can assumed to be a reasonable estimate of $\psi(a, \omega)$ under the assumption that the data was sampled sufficiently fine and over a wide enough range. Figures 7.3(b) and (c) display the estimates of the real and imaginary components of the Gamma filter $\psi(a, x)$ for scale $a = 8$. The real component of $\psi(a, x)$ is associated with the modulus $|d_a(x)|$ and the imaginary component of $\psi(a, x)$ with the phase $\theta_a(x)$. Both components of the filter are highly localised in space indicating that they would be good at detecting edges or discontinuities in a signal. Observe the differences in the filter function $\psi(a, x)$ shape, the real component in Figure 7.3(b) consists of one sharp *spike* (large jump in magnitude) surrounded by two small *wells* (dips in magnitude). In contrast, the imaginary component in Figure 7.3(c) has two small dips and two large jumps. These differences in filter function shape between the real and imaginary components of $\psi(a, x)$ are very important because they indicate that very different features are extracted by the modulus and phase of the coefficients.

Coefficient energy, E, of the integral wavelet transform is defined by,

$$E = \int_{-\infty}^{\infty} |f(x)|^2 dx = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{|\mathcal{W}_\psi[f(x)(a,b)]|^2}{a^2} da\, db \tag{7.13}$$

where $\mathcal{W}_\psi[f(x)(a,b)]$ is the one-dimensional equivalent of the integral wavelet transform for scale parameter $a \in \mathbb{Z}^+$ and location parameter $b \in \mathbb{R}$ (page 47, Vidakovic 1999). The coefficient C_ψ weights the energy of the transform to that of the signal. The admissibility condition and the coefficient C_ψ are related to Parseval's identity (theorem 2.2.2, page 27, Vidakovic 1999), which states that the energy calculated from the data and the coefficients is the same. The Gamma filter weights the energy of coefficients according to scale, $C_\psi = (a-1)!$, thus $E \propto \frac{1}{(a-1)!}$. Therefore, the energy of the signal $f(x)$ becomes much smaller than the energy of the transform as scale increases. At higher resolutions (larger scales) the coefficients become much larger as compared to those at lower resolutions. In two-dimensions, the adaptive image transform can be used instead of the integral wavelet transform but the general properties of the energy of the Gamma filter coefficients remains the same. The response can also be altered such that filter prefers low-frequency trend terms by substituting $\alpha^\star = (J-a)$ for a in equation (7.8), where J is the number of scales considered in the transform. Thus, the Gamma filter offers a flexible method to emphasise certain signal components at the expense of others via the scale parameter.

The Gamma filter of equation (7.8) can be extended to two-dimensions using the tensor product. The two-dimensional Gamma filter is defined as,

$$\Psi(a, \omega_1, \omega_2) = \Psi(a, \omega_1) \cdot \Psi(a, \omega_2) \tag{7.14}$$

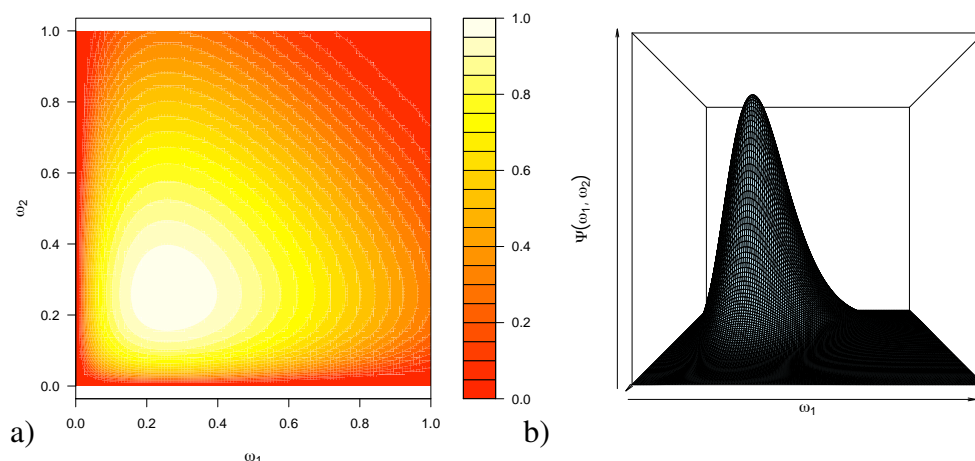$$\text{for } a \in \mathbb{R},\ \omega_1 \in \mathbb{R},\ \omega_2 \in \mathbb{R}.$$

Figure 7.4: The two-dimensional Gamma filter (for scale $a = 8$) in Fourier space: (a) image perspective (b) three-dimensional perspective.

Figures 7.4(a) and (b) display two perspectives of the two-dimensional Gamma filter $\Psi(a, \omega_1, \omega_2)$ in Fourier space. The filter function is broad, diffuse and is centered on a set spatial frequency. The real and imaginary components of the two-dimensional Gamma filter are displayed in Figures 7.5(a)-(d) in both the image and three-dimensional perspective. The real component is highly localised in space, the imaginary component is also highly localised but has two sharp positive-valued spikes along one diagonal and two sharp negative-valued spikes along the opposite diagonal. Both components are useful in the detection of intensity spikes in the image. The real component would match well (and hence produce a large inner product/ high energy coefficients) with the axial ring features of the SAXS image. The dips around the central spike may also assist with the analysis of SAXS images. At the appropriate scale, these lobes may capture and remove the effect of the even-ordered meridional scattering rings (which are not as prominent in the image) thereby enhancing the analysis of the odd-ordered meridional scattering rings (that are visible in Figure 2.2). The imaginary component of the Gamma filter is also very useful for the analysis of SAXS images. It would also match very well with the axial scattering ring features but would extract pairs of scattering rings. This is very useful as the meridional scattering rings features always occur in pairs in the SAXS image. Furthermore, the imaginary component of the Gamma filter conveys information on the position of the scattering rings in the

SAXS image. Coefficient energy is large only when the imaginary component of the wavelet and the image match each other well. At a particular scale, the imaginary component of the Gamma filter will extract the contribution to image energy of a scattering ring pair. The filters at adjacent scales will not match as well and there will be a distinct decrease in coefficient energy. Therefore, comparison of the imaginary component of the Gamma filter coefficients across scales for different SAXS images (particularly those from different tissue groups) might allow the detection of any differences in the position of their scattering rings.
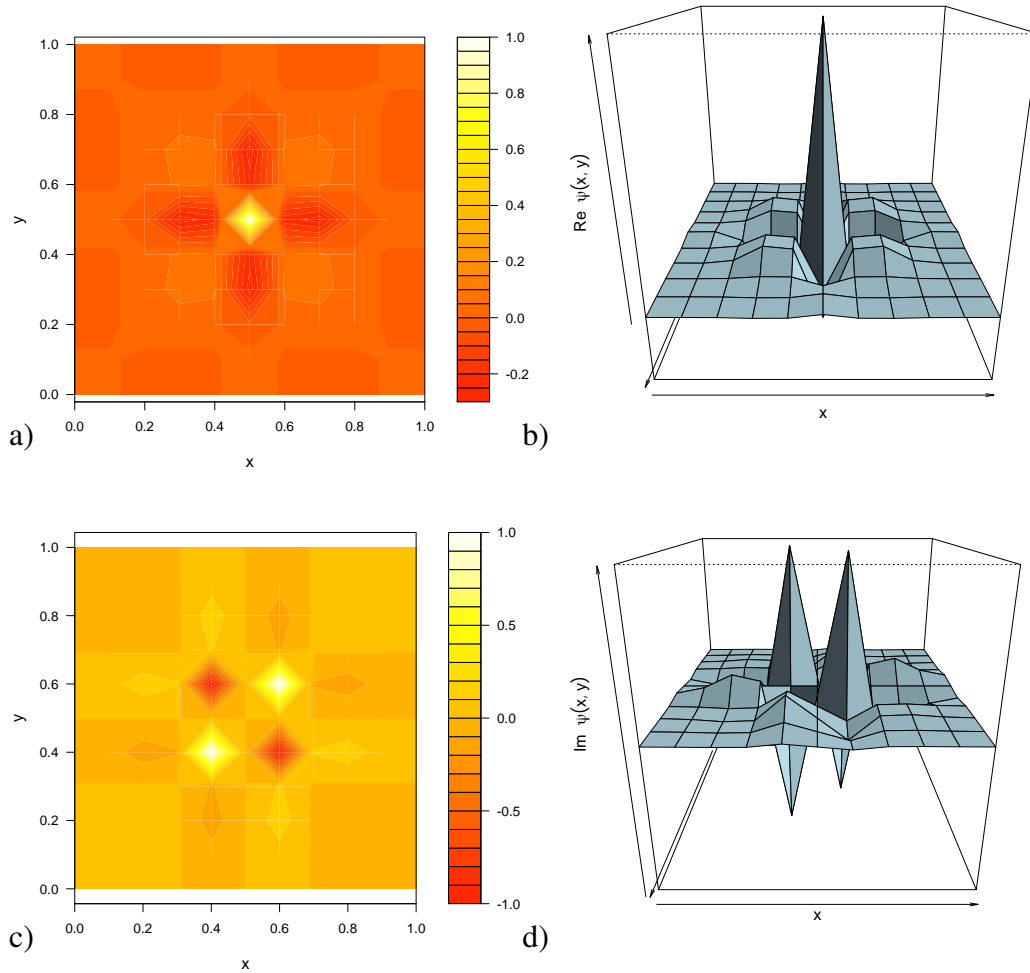

Figure 7.5: The two-dimensional Gamma filter (for scale $a = 8$) in spatial coordinates in both the image and three-dimensional perspectives for the real (a)-(b) and the imaginary components (c)-(d) respectively.
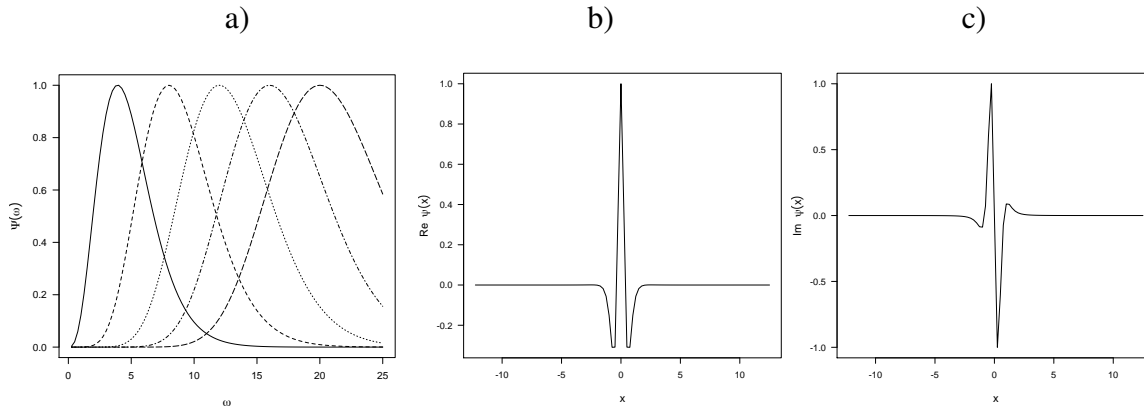
7.3.2 The Mellin filter

Figure 7.6: The one-dimensional Mellin filter for: (a) $\Psi(a, \omega)$ for scales a = 1-5 in frequency space and (b) the real and (c) imaginary components of $\psi(a, x)$ in space domain. The solid line in (a) corresponds to scale a =1,the short dash to scale a =2, dots to a =3, dot-dash to $a = 4$ and the long-dash to $a = 5$.

The Mellin filter is motivated by the desire to have a filter function that is similar to the Gamma filter in the location domain but is more specific bandpass properties in the frequency domain. This filter is designed to decay rapidly and smoothly from a particular reference frequency, $\omega_{(0,a)}$. The reference frequency is scale dependent which ensures that there is limited overlap between the filter functions at low resolutions and those at higher resolutions. The name 'Mellin filter' is motivated by the Mellin transform which is defined as,

$$g(z) = \int_0^{+\infty} f(\omega)\omega^{z-1}d\omega \qquad (7.15)$$

for parameter $z \in \mathbb{C}$ (page 305 Erdélyi *et al* 1954). Clever choice of the function $f(\omega)$ leads to a filter function with properties that the user desires such as the ability to extract high intensity peaks from the data. Equation (10) on page 313 of volume 1 of Erdélyi *et al* (1954) states that,

$$\int_0^{+\infty} (e^\omega - 1)^{-2}\omega^{z-1}d\omega = \Gamma(z)\big[\zeta(z-1) - \zeta(z)\big] \qquad (7.16)$$

for $\Re(z) > 2$ where $\Gamma(z)$ is the Gamma function and $\zeta(z) = \sum_{k=1}^{+\infty} k^{-z}$ ($\Re(z) > 1$) the Riemann zeta function (equation 23.2.1, page 807, Abramowitz & Stegun 1972). Thus, equation (7.16) provides an analytical form for the Mellin transform (equation 7.15) of $f(\omega) = (e^\omega - 1)^{-2}$.

A substitution is now made to relate the Mellin transform to the modified admissibility condition, observe that if we substitute $z = 4a + 1$ into equation (7.16) (where $a \; \epsilon \; \mathbb{Z}^+$ is the scale parameter of the filter function) then,

$$
\begin{aligned}
g(4a + 1) &= \int_0^\infty (e^\omega - 1)^{-2} \omega^{4a} d\omega \\
&= \Gamma(4a + 1) \big[\zeta(4a) - \zeta(4a + 1) \big] \\
&= (4a)! \frac{(2\pi)^{4a}}{2(4a)!} |\mathbf{B}_{4a}|
\end{aligned}
\tag{7.17}
$$

as

$$
\Gamma(4a + 1) = (4a)!
\tag{7.18}
$$

using $\Gamma(n + 1) = n!$ for $n \; \epsilon \; \mathbb{Z}$ (equation 6.1.6, page 255, Abramowitz & Stegun 1972) and

$$
\zeta(4a) = \frac{(2\pi)^{4a}}{2(4a)!} |\mathbf{B}_{4a}|
\tag{7.19}
$$

and

$$
\zeta(4a + 1) = \frac{(2\pi)^{4a+1}}{2(4a + 1)!} |\mathbf{B}_{4a+1}| = 0
\tag{7.20}
$$

where $\mathbf{B}_{4a}$ and $\mathbf{B}_{4a+1}$ are Bernoulli numbers (section 23.1 and equation 23.1.2, page 804, Abramowitz and Stegun 1972). The results in equations (7.19) and (7.20) can be demonstrated using the following rule,

$$
\zeta(2n) = \frac{(2\pi)^{2n}}{2(2n)!} |\mathbf{B}_{2n}|, \qquad n \; \epsilon \; \mathbb{Z}
\tag{7.21}
$$

(equation 23.2.16, page 807, Abramowitz & Stegun 1972). Substituting $2n = 4a$ into equation (7.21) gives the result in equation (7.19). Observe that, $\zeta(4a + 1) = 0$ because

$$
\mathbf{B}_{2n^\star + 1} = 0, \qquad n^\star \; \epsilon \; \mathbb{Z}
\tag{7.22}
$$

(equation 23.1.19, page 805, Abramowitz & Stegun 1972). Substitution of $n^\star = 2a$ demonstrates that $\mathbf{B}_{2n^\star+1} = \mathbf{B}_{2(2a)+1} = \mathbf{B}_{4a+1} = 0$. Thus from equation (7.20), $\zeta(4a + 1) = 0$.

By selecting,

$$\frac{|\Psi(a,\omega)|^2}{\omega} = (e^\omega - 1)^{-2}\omega^{4a+1} \tag{7.23}$$

the modified admissibility condition (equation 5.13) is satisfied because,

$$\begin{aligned}
C_\psi &= \int_0^{+\infty} \frac{|\Psi(a,\omega)|^2}{\omega}d\omega \\
&= \int_0^{+\infty} (e^\omega - 1)^{-2}\omega^{4a+1}d\omega \\
&= \Gamma(4a+1)\big[\zeta(4a) - \zeta(4a+1)\big] \\
&= (4a)!\frac{(2\pi)^{4a}}{2(4a)!}|\mathbf{B}_{4a}| < \infty.
\end{aligned} \tag{7.24}$$

Therefore,

$$|\Psi(a,\omega)|^2 = (e^\omega - 1)^{-2}\omega^{4a+1} \tag{7.25}$$

and the definition of the Mellin filter function is

$$\Psi(a,\omega) = (e^\omega - 1)^{-1}\omega^{2a+\frac{1}{2}}e^{i\varphi(\omega)} \tag{7.26}$$

for $\omega > 0$ and some function $\varphi(\omega) \epsilon L_1(\mathbb{R})$. In equation (7.26), for $\varphi(\omega) = 2\pi$ (and $e^{i\varphi\omega} = 1$), the filter function $\psi(a,x)$ is complex in the location (x) domain. An analytical form of $\psi(a,x)$ has not been found and can only be expressed as,

$$\begin{aligned}
\psi(a,x) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\Psi(a,\omega)e^{i\omega x}d\omega \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty}(e^\omega - 1)^{-1}\omega^{2a+\frac{1}{2}}e^{i\omega x}d\omega.
\end{aligned} \tag{7.27}$$

This integral appears intractable, nevertheless highly accurate graphs of the function $\psi(a,x)$ can be drawn and understanding of this wavelet developed using the graphs. Figure 7.6(a) displays the Mellin filter function $\Psi(a,\omega)$ for the scales $a = 1-5$. Note how the filter associated with each scale is centered around a particular frequency ($\omega_{(0,a)}$) and that this frequency increases as the scale parameter increases. The lowest scale ($a = 1$) and the highest scale displayed ($a = 5$) barely overlap, which indicates that these two functions filter out very different frequencies in the data. Figures 7.6(b) and (c) display the real and imaginary components of the Mellin filter $\psi(a,x)$ for scale $a = 1$ respectively. The real component of the filter is sharply defined around a

central location and is non-zero over a finite interval which means that is has *compact support*. The real component of the filter is also symmetric about the point $x = 0$, which indicates that signal features either side of the central peak of the filter are given equal weighting when calculating transform coefficients. In contrast, the imaginary component of the filter is not symmetric about $x = 0$. The region of the imaginary component of the Mellin filter corresponding to $x < 0$ is the exact opposite of the region corresponding to $x > 0$. The real component of the Mellin filter seems better suited to detecting jumps in signal/image intensity whilst the imaginary component seems better suited to detecting change in polarity [1] (from positive to negative or vice versa) in signal/image intensity. Therefore, the Mellin filter allows for multi-scale analysis that is focused on detecting two types of signal/image characteristics. Using the Mellin filter for image analysis provides sets of real and imaginary coefficients that describes the jumps in intensities and the changes in polarity of features respectively.

The two-dimensional version of the Mellin filter is found using the tensor product, $\Psi(a, \omega_1, \omega_2) = \Psi(a, \omega_1) \cdot \Psi(a, \omega_2)$. The frequency space version of the Mellin filter $\Psi(a, \omega_1, \omega_2)$ is displayed in Figure 7.7(a) as an image and in Figure 7.7(b) as a three-dimensional perspective. The Mellin filter in the frequency domain is a disc of finite size centered on reference frequencies $(\omega_{1,0,a}, \omega_{2,0,a})$. Figure 7.7(c) displays the real component of the Mellin filter (for scale $a = 1$) it is a sharp spike of finite size. The imaginary component of the Mellin filter is displayed in Figure 7.7(d), it consists of two sharp large magnitude peaks that alternate between positive and negative values and small side lobes that surround these peaks. Therefore the two-dimensional Mellin filter is localised over a region in the frequency domain, highly localised in the spatial domain and able to identify change in image intensity as well as change in image polarity. As with the Gamma filter, the imaginary component measures the combined positive and negative lobes that result in SAXS images via diffraction from a physical structure.
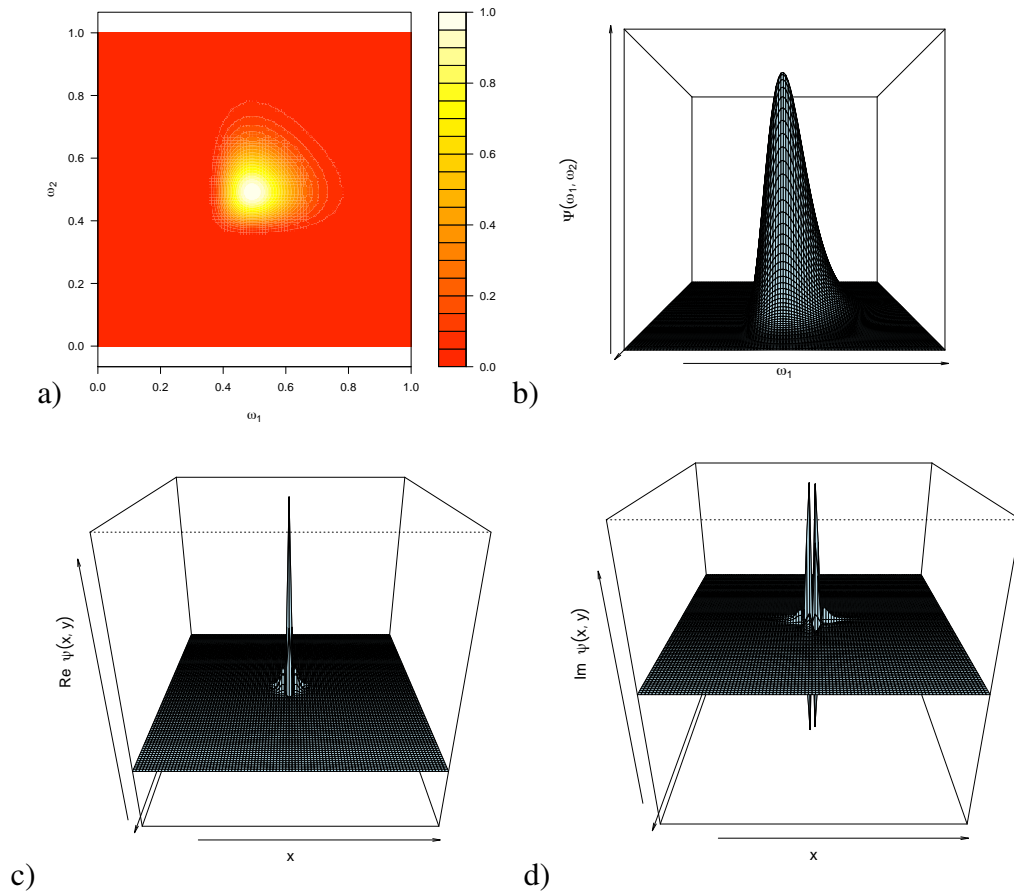
---

[1] a feature of diffraction images.

Figure 7.7: The two-dimensional Mellin filter $\Psi(a, \omega_1, \omega_2)$ for scale $a = 1$ in frequency space in the (a) image and (b) three-dimensional perspectives and (c) the real $\Re[\psi(a, x, y)]$ and (d) $\Im[\psi(a, x, y)]$ the imaginary components in the spatial domain.
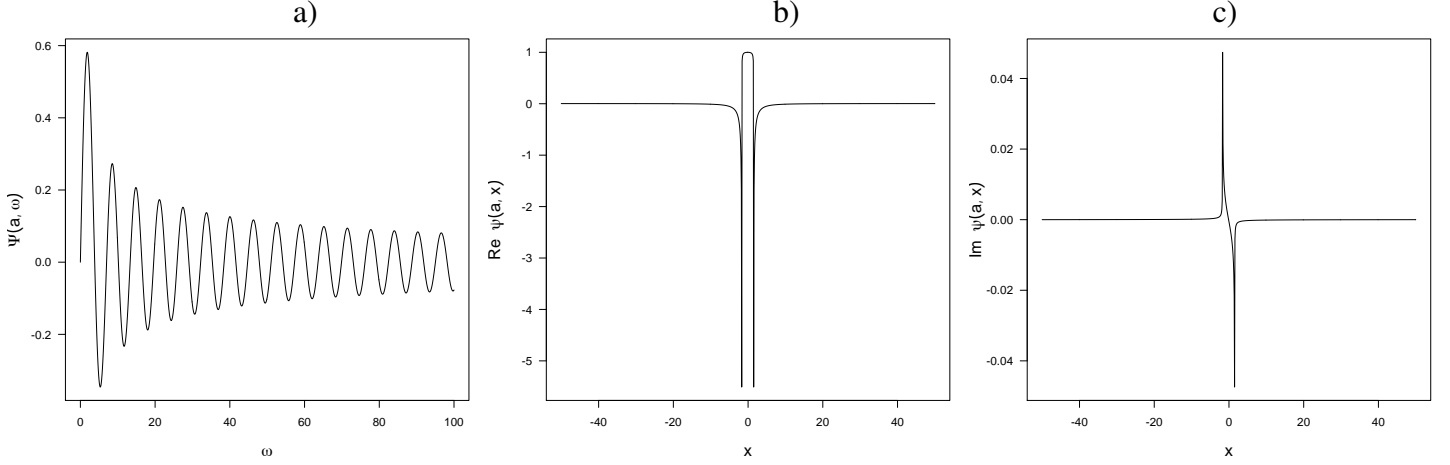
7.3.3   The Chebyshev Filter

Figure 7.8: The one-dimensional Chebyshev filter for (a) $\Psi(a,\omega)$ for scale $a = 1$ in frequency space and (b) the real and (c) imaginary components of $\psi(a,x)$ in location domain.

The Chebyshev filter is designed to detect and extract trends that resemble Bessel functions of the first kind, $J_\gamma(z)$ in the data. This filter function might be useful for examining the equatorial scattering rings in the SAXS images as Lewis *et al* (2000) extracted one-dimensional profiles of the the intensity of these features and discovered that Bessel functions of the first kind provide a good fit. The Bessel function of the first kind is defined as,

$$J_\gamma(z) = \frac{2(\frac{1}{2}z)^\gamma}{\pi^{\frac{1}{2}}\Gamma(\gamma + \frac{1}{2})} \int_0^1 (1 - t^2)^{\gamma - \frac{1}{2}} \cos(zt)dt \tag{7.28}$$

$(\gamma \, \epsilon \, \mathbb{C} \, , \Re(\gamma) \, > \, -\frac{1}{2})$ (equation 9.1.1, page 358, Abramowitz & Stegun 1972). It is a solution to the following differential equation,

$$z^2 \frac{d^2w}{dz^2} + z\frac{dw}{dz} + (z^2 - \gamma^2)w = 0 \tag{7.29}$$

for z and $w \, \epsilon \, \mathbb{C}$ (equation 9.1.1, page 358, Abramowitz & Stegun 1972). The fact that Bessel functions of the first kind, $J_\gamma(z)$, are solutions to a differential equation (equation 7.29) implies that they have constraints on their derivatives (and hence on the rate of oscillation of z about the w axis). Analysis using the Chebyshev filter is confined to detecting very specific trends in the data that conform to the smoothness specified by equation (7.29) and can therefore be described using Bessel functions of the first kind, $J_\gamma(z)$.

The Chebyshev polynomial, $T_n(x)$ is defined as

$$T_n(x) = \frac{1}{4\pi i} \int_\gamma \frac{(1-t^2)t^{-n-1}}{1-2xt+t^2} dt \tag{7.30}$$

for $n \in \mathbb{Z}^+$, $x \in \mathbb{R}$ and γ is a contour corresponding to the unit circle (equation 14, page 397, Weisstein 2003). The Bessel function of the first kind, $J_n(x)$ (where the subscript γ has been replaced by the subscript n to specifically indicate that the indices are integer values) is related to the Chebyshev polynomial $T_n(x)$ via:

$$T_n(i\frac{d}{dx}) = \frac{i^{-n} J_n(x)}{J_0(x)} \tag{7.31}$$

(modified from equation 21, page 397, Weisstein 2003). Therefore the nth Chebyshev polynomial, $T_n(x)$ can be understood to be related to the nth Bessel function of the first kind $J_n(x)$, scaled by $J_0(x)$, the Bessel function of the first kind of order zero and the constant i^{-n} (for fixed n).

To develop a filter function, $\psi(a, x)$ that is capable of extracting trends in the data that are related to Bessel functions of the first kind, $J_n(x)$ observe that

$$\int_0^{+\infty} \frac{J_n(x)}{x} dx = \frac{1}{2n} \tag{7.32}$$

where $\Re(n) > 0$ (equation 10, page 283, Jeffrey 2000). Recall the modified admissibility condition of equation (7.5) is

$$C_\psi = \int_0^\infty \frac{|\Psi(a,\omega)|^2}{\omega} d\omega \tag{7.33}$$

for $\omega \in \mathbb{R}^+$ and $a \in \mathbb{Z}^+$. Setting $|\Psi(a,\omega)| = J_a(\omega)$ (where $J_a(\omega)$ is a Bessel function of the first kind of order a indexed by frequency space parameter, ω).

Therefore,

$$\Psi(a, \omega) = J_a(\omega) \exp(i\varphi(\omega)). \tag{7.34}$$

For the image analysis applications in this thesis, the Chebyshev filter is designed to have only a real component in frequency space, so that $\exp[i\varphi(\omega)] = 1$. The modified admissibility condition of equation (7.5) is satisfied as,

$$\int_0^\infty \frac{|\Psi(\omega)|^2}{\omega} d\omega = \int_0^\infty \frac{J_a(\omega)^2}{\omega} d\omega = \frac{1}{2a} \; < \; \infty \tag{7.35}$$

using equation (7.32).

The Bessel function of the first kind, $J_a(\omega)$ and the Chebyshev polynomial, $T_n(x)$, are a Fourier transform pair. That is,

$$
\begin{aligned}
\psi(a, x) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Psi(a, \omega) e^{i\omega x} d\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} J_a(\omega) e^{i\omega x} d\omega \\
&= \frac{1}{2\pi(-1)^a i^a \pi}
\begin{cases}
0, & -\infty < x < 1 \\
(1-x^2)^{-\frac{1}{2}} T_a(x), & -1 < x < 1 \\
0, & 1 < x < \infty
\end{cases}
\end{aligned}
\tag{7.36}
$$

(using the table of Fourier transforms, equation 2, page 122, Erdélyi *et al* 1954). Note that the Chebyshev filter is non-zero over the domain $(-1, 1)$ which indicates that it has compact support. The Chebyshev filter derives its' name because it is proportional to the Chebyshev polynomial, $T_a(x)$ as evident in equation (7.37). Observe that the Chebyshev filter $\psi(a, x)$ also consists of a $(1 - x^2)^{-\frac{1}{2}}$ term which acts like a polynomial filter, modifying the degree of the Chebyshev polynomial $T_a(x)$. This polynomial filter resembles a 'well' (U-shaped graph of $f(x)$ against x) and has the least influence (weight) around $x = 0$ and the greatest influence around the extremes of the domain ($x = -1$ and $x = 1$).

In a manner similar to the Gamma and Mellin filter, the frequency space representation $\Psi(a, \omega)$ and the location domain representation $\psi(a, \omega)$ can be accurately plotted using software. The one-dimensional version of the Chebyshev filter is displayed in Figures 7.8(a)-(c) for scale $a = 1$. The frequency space representation, $\Psi(a, \omega)$ is displayed in Figure 7.8(a), it is equivalent to the Bessel function of the first kind $J_a(\omega)$. Further examples of Bessel functions of the first kind are available in Figure 17.1, on page 274 of Jeffrey (2000). Figure 7.8(a) reveals that the Chebyshev filter seeks out trends that resemble Bessel functions of the first kind in the frequency domain. The real $\Re[\psi(a, x)]$ and imaginary $\Im[\psi(a, x)]$ components of the Chebyshev filter in the location domain for scale $a = 1$ are displayed in Figures 7.8(b) and (c) respectively. The real component $\Re[\psi(a, x)]$ appears to be capable of extracting smooth 'bumps' in signal intensity similar in shape to the intensity profile of the equatorial scattering rings. Therefore, it may capture different signal information than either the Gamma or the Mellin filters. The imaginary component $\Im[\psi(a, x)]$ is very sharp and jagged in shape. Similar to the Mellin filter, it is useful in detecting changes in the polarity of signal intensity.
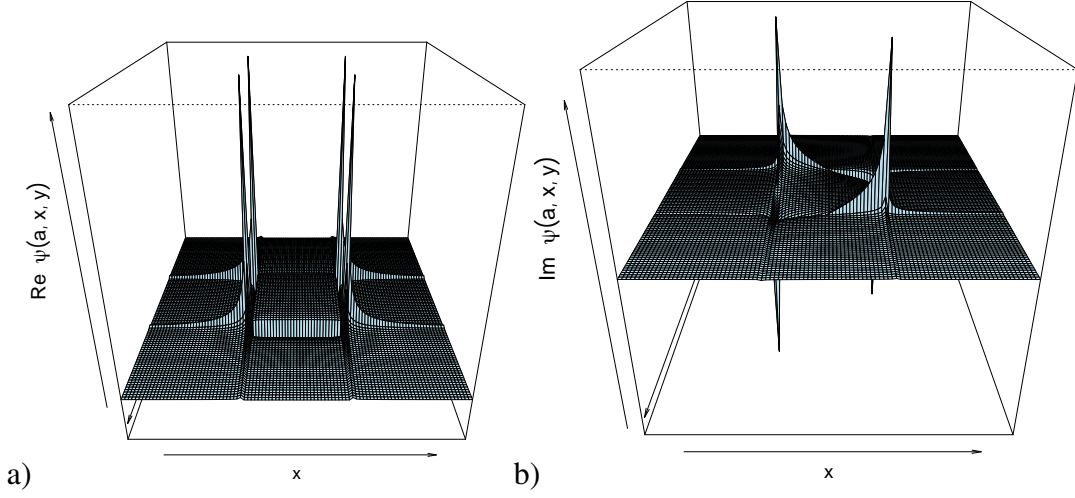
Figure 7.9: The two-dimensional Chebyshev filter for scale $a = 1$ in perspective view for the: (a) real and (b) imaginary components in the spatial domain.

The real and imaginary spatial components of the two-dimensional version of the Chebyshev filter $\psi(a, x, y)$ for scale $a = 1$ is displayed in Figures 7.9(a) and (b) respectively. The real component $\Re[\psi(a, x, y)]$ resembles a 'box' surrounded by very sharp edges. Similar to the one-dimensional version this component of the filter is best suited to extracting the larger and smoother features from an image. It is also suited to extracting those image features that resemble squares, rectangles or other regions of the image that have changes in image intensity magnitude and sign in two orthogonal directions (such as the beam-stop region of the SAXS image). The imaginary component of the Chebychev filter $\Im[\psi(a, x, y)]$ that is displayed in Figure 7.9(b) for scale $a = 1$, has both a positive $\{\Im[\psi(a, x, y)] > 0\}$ and negative $\{\Im[\psi(a, x, y)] < 0\}$ component. The peaks of the positive component are connected together along a smooth curve that resembles a ring. A similar geometry holds for the peaks of the negative component. Together, the positive and negative components allows the detection of a change in image intensity polarity along a scattering ring. It could allow precise determination of where a scattering ring 'jumps up' from the background intensity and where it 'falls down' back to the background intensity. Thus it would enable measurement of the width of the scattering ring. This could be very important if malignant breast tissue conditions results in a greater variation of the axial d-spacing or in the lateral organisation of collagen fibrils (refer to Section 2.2) as both may result in a greater width of the scattering rings in the SAXS images. Therefore, imaginary part $\Im[\psi(a, x, y)]$ of the Chebyshev filter might provide useful information in the analysis of SAXS images.
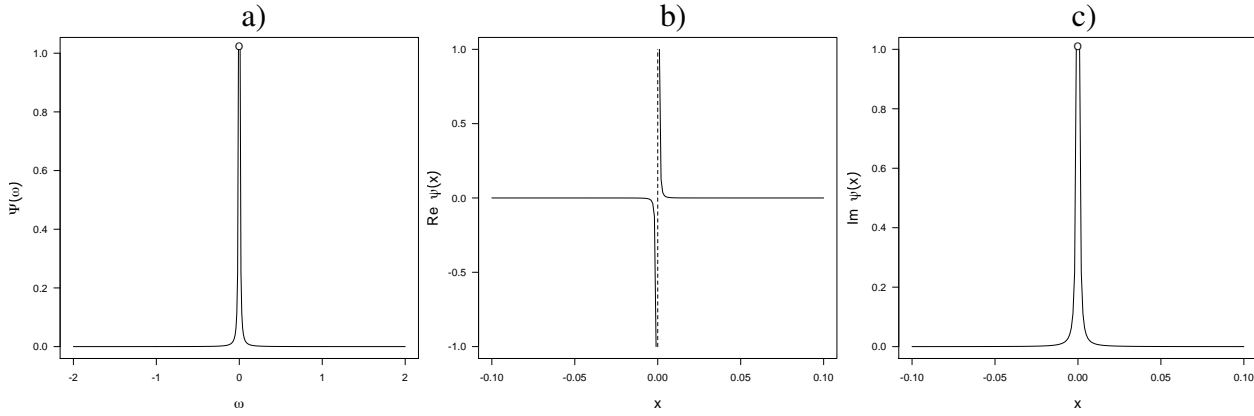
7.3.4 The Zeta filters

Figure 7.10: The one-dimensional Zeta filter for scale $a = 1$ in the: (a) $\Psi(a, \omega)$ frequency domain (the open dot indicates a discontinuity at $\omega = 0$) and (b) the real and (c) imaginary components of $\psi(a, x)$ in the location domain.

Zeta filters are designed to capture *power law* (a polynomial dependency of data magnitude with frequency) processes. These trends are of significance for the analysis of SAXS images because they describe relationships across scales and hence between spatial frequencies. The equatorial and meridional scattering rings in the SAXS images of breast tissue (refer to Figure 2.2) are due to diffraction from the physical structures of collagen over different length scales. A relationship might exist between these tissue structures for healthy breast tissue and this relationship may breakdown upon the onset of breast cancer. The SAXS image features are likely to convey this information and the Zeta filter is suitable for detection and description of such relationships.

The particular power-law processes that we are interested in analysing are those that can be described by a function of the form,

$$f(a, \omega) = \frac{\omega^{-2a}}{\zeta(2a)} \tag{7.37}$$

for scale parameter $a \in \mathbb{Z}^+$. In equation (7.37) the Zeta function $\zeta(2a)$ is a scaling function that is dependent on the scale parameter, a. The frequency dependency of the data is described by the polynomial ω^{-2a} and therefore the analysis function $f(a, \omega)$ examines power-law processes of the form $(1/\omega^{2a})$.

By taking the logarithm of $f(a, \omega)$ the scale-dependency of these processes is more explicit,

$$\log[f(a, \omega)] = -2a \log(\omega) + \log[1/\zeta(2a)]. \tag{7.38}$$

A linear relationship exists between $\log[f(a, \omega)]$ and $\log(\omega)$. Changing the scale parameter a alters the slope $(-2a)$ and the intercept $\log[1/\zeta(2a)]$, but the relationship still holds. Power law relationships are known to exist in SAXS data. The analysis of SAXS image intensity profiles of the form $I(h)$ against h ($|\mathbf{h}| = 4\pi \sin\theta/\lambda$ refer to Section 2.2) typically involves studying such power law processes. Porod's law states that $I(h) \propto h^{-4}$ (Section 2.4.4, page 45, Feigin & Svergun 1987). Studies of the adherence or deviation of SAXS intensity profiles to Porod's law can be used to infer the fractal nature of the surface of the object under study. When the frequency space representation $\Psi(a, \omega)$ of the Zeta filter is used to analyse SAXS data (or images) it is particularly adept at picking out Porod law trends (as they are proportional to minus two). By including a scale parameter a in the analysis it is possible to use the Zeta filter to search for a whole family of power law processes (not just the Porod law trends).

The Zeta filter, $\Psi(a, \omega)$ can be defined in frequency space as,

$$\Psi(a, \omega) = K_a \frac{\omega^{-2a}}{\zeta(2a)}. \tag{7.39}$$

where K_a is a constant selected such that $\int_{-\infty}^{\infty} \Psi(a, \omega)\, d\omega = 1$. The modified admissibility condition (of equation 7.5) nor the standard admissibility condition (of equation 5.13) are satisfied as,

$$
\begin{aligned}
\int_0^\infty \frac{|\Psi(a, \omega)|^2}{\omega} d\omega &= \int_0^\infty [K_a \frac{\omega^{-2a}}{\zeta(2a)}]^2 \omega^{-1} d\omega \\
&= [\frac{K_a}{\zeta(2a)}]^2 \int_0^\infty \omega^{-(4a^2+1)} d\omega
\end{aligned}
\tag{7.40}
$$

does not converge. Similarly,

$$\int_{-\infty}^{+\infty} \frac{|\Psi(a, \omega)|^2}{\omega} d\omega = [\frac{K_a}{\zeta(2a)}]^2 \int_{-\infty}^{+\infty} \omega^{-(4a^2+1)} d\omega \tag{7.41}$$

does not converge.

The consequence of this lack of convergence is that the Zeta filter can be used for a multi-scale analysis but the data cannot be reconstructed from the corresponding coefficients of an integral wavelet transform (or the adaptive image transform in two dimensions). In the case of the analysis of SAXS data (as in this thesis) such a situation is acceptable as there is no need to reconstruct the data from the coefficients.

The location space version of the Zeta filter, $\psi(a, x)$, can be found by applying the inverse Fourier transform.

$$
\begin{aligned}
\psi(a, x) &= \frac{K_a}{2\pi} \int_{-\infty}^{+\infty} \Psi(a, \omega) \exp(i\omega x) d\omega \\
&= \frac{K_a}{2\pi\zeta(2a)} \int_{-\infty}^{+\infty} \omega^{-2a} \exp(i\omega x) d\omega.
\end{aligned}
\tag{7.42}
$$

The integral $\int_{-\infty}^{+\infty} \omega^{-2a} \exp(i\omega x)\, d\omega$ can be solved by using,

$$
\int_{-\infty}^{+\infty} \frac{\exp(kt)}{t^m} dx = \frac{1}{m-1}\left[k\int_{-\infty}^{+\infty} \frac{\exp(kt)}{t^{m-1}} dt - \frac{\exp(kt)}{t^{m-1}}\right]
\tag{7.43}
$$

which is not defined for $k \neq 0.5$ (equation 404, Peirce 1899).

Thus,

$$
\begin{aligned}
\psi(a, x) &= \int_{-\infty}^{+\infty} \omega^{-2a} \exp(i\omega x) d\omega \\
&= \frac{1}{2a-1}\left[ix\int \frac{\exp(i\omega x)}{\omega^{2a-1}} d\omega - \frac{\exp(i\omega x)}{\omega^{2a-1}}\right] \\
&= \frac{-1}{2a-1}\left[\frac{\exp(i\omega x)}{\omega^{2a-1}} + i\omega^{3-2a} E_{2a-1}(-i\omega x)\right].
\end{aligned}
\tag{7.44}
$$

Where $E_n(z)$ is the exponential integral which is defined as,

$$
E_n(z) = \int_1^{\infty} \frac{\exp(-zt)}{t^n} dt
\tag{7.45}
$$

for $z \in \mathbb{C}$, $\Re(z) > 0$, $n \in \mathbb{Z}$ (equation 5.1.4, page 228, Abramowitz & Stegun 1972).

The Zeta filter $\psi(a, x)$ consists of an exponential term $\exp(i\omega x)$ weighted by the polynomial ω^{2a-1}. The larger the scale parameter, a, the greater the degree of this polynomial. An additional component is provided by the product of the polynomial ω^{3-2a} and the exponential integral $E_{2a-1}(-i\omega x)$. As scale increases, the degree of ω^{3-2a} decreases and the polynomial transfers from the role of function enhancer to function moderator. The exponential integral $E_{2a-1}(-i\omega x)$ depends upon the scale parameter, a, in such a way that as the scale increases the exponential integral function becomes flatter (smaller magnitude derivative with respect to x).

The Riemann zeta function, $\zeta(2a)$ can be found for $a \, \epsilon \, \mathbb{N}$ by noting that,

$$\zeta(2a) = \frac{(2\pi)^{2a}}{2(2a)!} |\mathbf{B}_{2a}| \tag{7.46}$$

where $\mathbf{B}_{2a}$ denotes the $(2a)$th Bernoulli number (equation 23.2.16, page 807, Abramowitz & Stegun 1972). Specific values of the Bernoulli number $\mathbf{B}_{2a}$ can be found using,

$$\mathbf{B}_{2a} = B_{2a}(0) = \frac{(-1)^{a-1}2(2a)!}{(2\pi)^{2a}} \sum_{k=1}^{\infty} \frac{1}{k^{2a}} \tag{7.47}$$

for $a \, \epsilon \, \mathbb{N}$ where $B_{2a}(0)$ is the Bernoulli polynomial $B_{2a}(x)$ evaluated at $x = 0$ (equation 23.1.1, page 804, Abramowitz & Stegun 1972). Once $\zeta(2a)$ is calculated specification of $\psi(a, x)$ is straightforward.

The one-dimensional Zeta filter is displayed in both the frequency and location domains in Figures 7.10(a)-(c) for scale parameter $a = 1$. The Zeta filter $\Psi(a, \omega)$ is highly localised in the frequency domain, and is symmetric around, but also discontinuous at $\omega = 0$. The real component of the Zeta filter, $\psi(a, x)$ is displayed in Figure 7.10(b), it appears useful in the detection of features that change polarity (from negative intensity to positive intensity). Note that the real component is not defined for $x = 0$. Figure 7.10(c) displays the imaginary component of the Zeta filter $\psi(a, x)$ for scale $a = 1$. It resembles the frequency space version $\Psi(a, \omega)$ of the Zeta filter and might be useful in the detection in jumps in signal intensity.
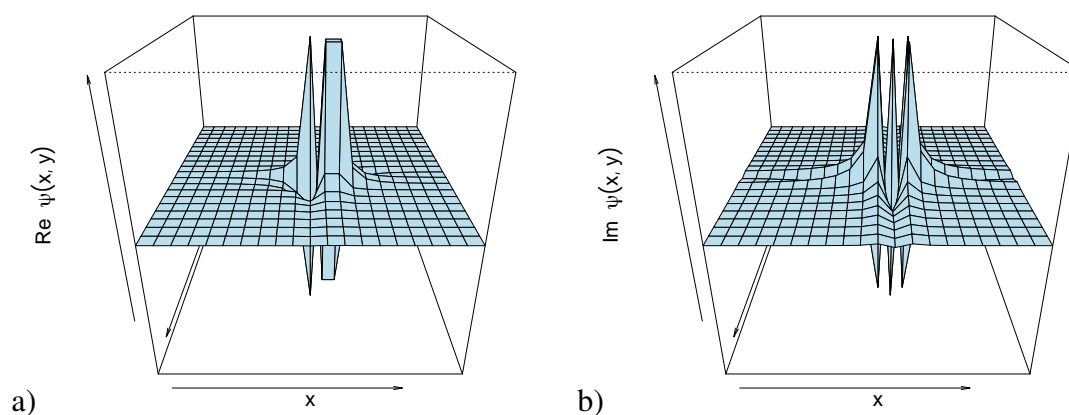
Figure 7.11: The two-dimensional Zeta filter $\psi(a, x, y)$ for scale $a = 1$ in the spatial domain: (a) real and (b) imaginary components.

The two-dimensional version of the Zeta filter is displayed in Figures 7.11(a) and (b). The real component of $\psi(a, x, y)$ in the spatial domain is displayed in Figure 7.11(a), it appears that it might be useful in detecting changes in the polarity of intensity for pairs of scattering ring features. It may be able to resolve the edges of closely spaced scattering ring features. Such a characteristic might be useful if the tissue sample under investigation contains several different populations of collagen that have different axial D-repeat periods. In this case, the axial scattering ring features might contain very fine scattering rings hidden within them. The real component $\psi(a, x, y)$ could be very effective at detecting such features. The imaginary component of the Zeta filter (for scale $a = 1$) is displayed in Figure 7.11(b), it appears useful in detecting a series of intensity peaks. The SAXS image of breast tissue collagen contain such series of intensity peaks along the meridional axis. The imaginary component of the two-dimensional Zeta filter might be effective at detecting such features.
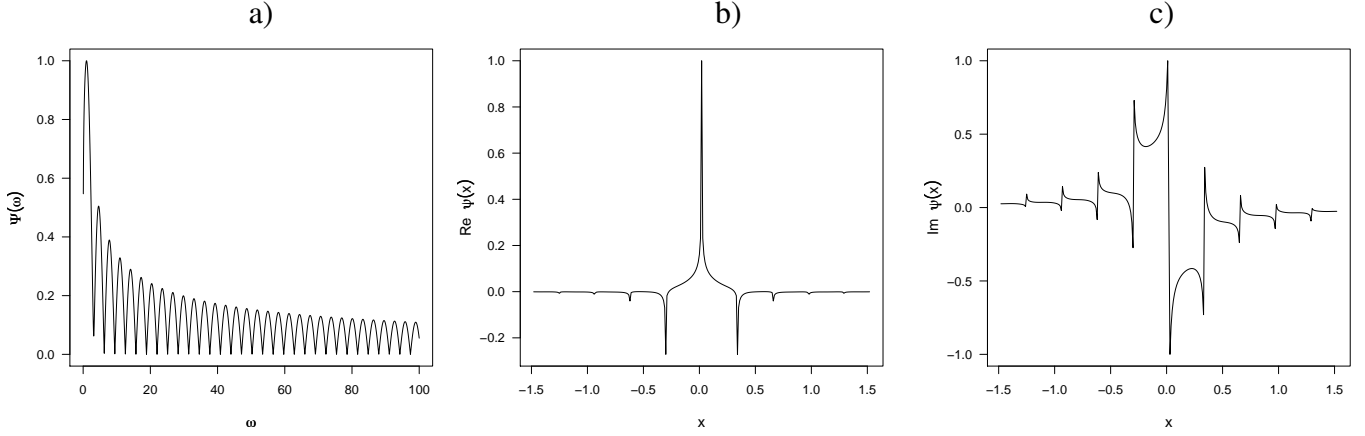
7.3.5   The Witch's Hat Filter

Figure 7.12: The one-dimensional Witch's Hat filter for scale $a = 1$ in the:
(a) Fourier domain and (b) the real and (c) imaginary components in the spatial domain.

The Witch's Hat filter function is designed to be a 'frequency-comb' filter, that is certain frequency components retained whilst others are attenuated. The series of frequency components are retained depends on the scale parameter, a. Both low and high frequency components are retained simultaneously, allowing the Witch's Hat filter to perform an analysis that examines the interaction between low and high frequency signal components. This is of significance in SAXS data analysis because the same physical structure (such as the axial D-repeat) can produce both low and high frequency features in the signal. The frequency components within the feature might then be coupled and dependent upon each other. Standard wavelet filters (such as those reported in Mallat (1998)) attempt to separate the low and high frequency information in a signal into different analysis bands. Such an approach could fail to detect frequency coupling within a signal and obscure potentially useful diagnostic information. Analysis using the Fourier transform (Section 5.1) may be able to detect such frequency coupling but will probably not be able to identify the features involved. The Witch's Hat filter overcomes both of these problems and is able to identify certain types of frequency coupling as well as those features/frequencies involved.

A frequency-comb filter can be designed using the following function,

$$\Psi(a, \omega) = \sqrt{\frac{\sin^2(a\omega)}{\omega} + \sqrt{\omega}\exp(-a\omega)} \qquad (7.48)$$

for $a \in \mathbb{R}^+$.

The $\frac{\sin^2(a\omega)}{\omega}$ term of $\Psi(a, \omega)$ provides the 'frequency comb' structure whilst the $\sqrt{\omega}\exp(-a\omega)$ adds a smooth decaying background trend. The dependency of $\sin^2(a, \omega)$ on the scale parameter a ensures that a whole family of different frequency relationships is considered. As the scale increases, the 'finer' this comb is, with frequencies of similar magnitude being alternatively retained or attenuated. By design the Witch's Hat filter $\Psi(a, \omega)$ hopes to capture scale dependent frequency couplings that can be described by the $\sin^2(a\omega)$ function.

The Witch's Hat filter, $\Psi(a, \omega)$ satisfies the modified admissibility condition (equation 7.5) as,

$$
\begin{aligned}
\int_0^\infty \frac{|\Psi(a, \omega)|^2}{\omega} d\omega &= \int_0^\infty \Big[\frac{\sin^2(a\omega)}{\omega^2} + \frac{\exp(-a\omega)}{\sqrt{\omega}}\Big] d\omega \\
&= \int_0^\infty \frac{\sin^2 a\omega}{\omega^2} d\omega + \int_0^\infty \frac{\exp(-a\omega)}{\sqrt{\omega}} d\omega \\
&= \frac{\pi a}{2} + \sqrt{\frac{\pi}{a}}
\end{aligned}
\tag{7.49}
$$

using $\int_0^\infty \frac{\sin^2 a\omega}{\omega^2} d\omega = \frac{\pi a}{2}$ for $a \; \epsilon \; \mathbb{R}^+$ and $\int_0^\infty \frac{\exp(-a\omega)}{\sqrt{\omega}} d\omega = \sqrt{\frac{\pi}{a}}$ for $a \; \epsilon \; \mathbb{R}^+$, (equations 486 and equation 497, page 63 Peirce 1899, Wolfram 2008).

Figures 7.12(a)-(c) display the one-dimensional version of the Witch's Hat filter for scale $a = 1$, Figure 7.12(a) is of the Witch's Hat filter, $\Psi(a, \omega)$ in the frequency space. Note its repetitive structure, a key reference frequency $\omega_{0,a}$ is retained, whilst all other frequencies are either substantially attenuated or removed. The amount of attenuation increases for higher frequencies and is dependent on scale due to the $\sin^2(a\omega)$ and $\exp(-ax)$ terms in equation (7.48). The plot of $\Psi(a, \omega)$ against ω in Figure 7.12(a) suggests that the Witch's Hat filter might be useful for extracting features that contain a number of coupled frequency components. The inverse discrete Fourier transform of $\Psi(a, \omega)$ was used to obtain a numerical estimate of the real and imaginary components of $\psi(a, x)$. The real component of the Witch's Hat filter is displayed in Figure 7.12(b), the shape of the filter function in this plot inspired the name '*Witch's Hat*'. This component of the filter appears to be useful in detecting sharp 'spikes' or rises in the intensity of a signal/image. Figure 7.12(c) displays the imaginary component of the Witch's Hat filter for scale $a = 1$, its interesting shape might be useful in the detection of a series of intensity 'steps' in an signal/image. The Witch's Hat filter provides another tool in the arsenal of functions that are available to analyse a signal or image.

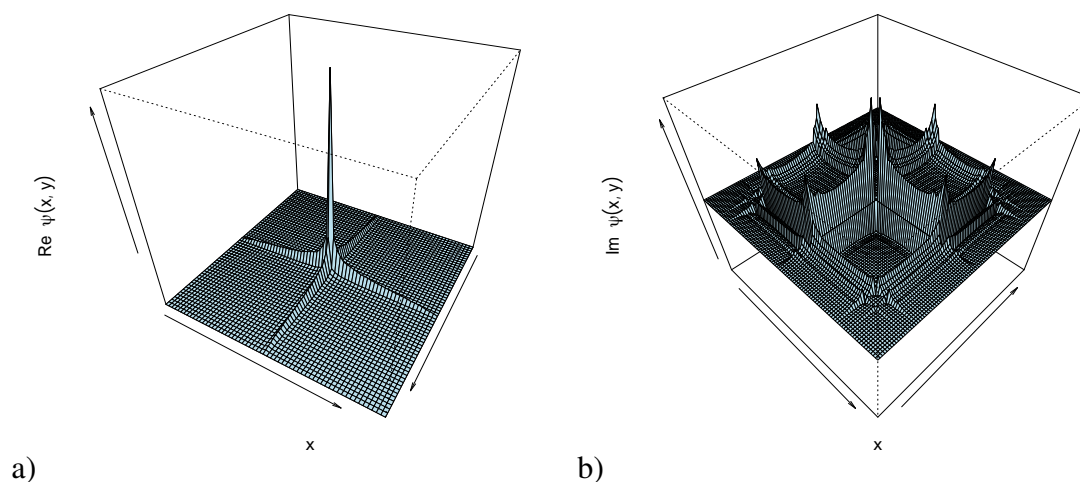a)   b)

Figure 7.13: The two-dimensional Witch's Hat filter $\psi(a, x, y)$ for scale $a = 1$ in the spatial domain: (a) real and (b) imaginary components.

The two-dimensional version of the Witch's Hat filter is displayed in Figures 7.13(a) and (b). The real component in Figure 7.13(a) consists of the combination of a cross-shaped and a sharp peaked structure. This component of the filter function might be useful in detection sharp rises in image intensity in certain directions. The imaginary component of the Witch's Hat filter (for scale $a = 1$) is displayed in Figure 7.13(b), it has two very jagged positive components along both diagonals and it might be useful in analysing the orientation of features in an image. The coefficients of the adaptive image transform are found by taking the inner product of the image with the wavelet analysis function. Large magnitude coefficients correspond to those image features that produce a large inner product with the analysis function, the greatest magnitude inner products occur when the image features resemble the relevant component of the filter analysis function. The real and imaginary components of the Witch's Hat filter look very different and appear to extract very different features from the image. In conclusion, the real and imaginary components of the adaptive image transformation using the Witch's Hat filter produce large magnitude coefficients for different image features but when studyied together provide a deeper understanding of image structure.

7.4   Function Selection for Pattern Recognition

A major challenge with the adaptive image transform is the need to select the most appropriate filter function from the library with which to represent the image over a range of scales and locations. The 'best' function with which to represent the image is often dependent on the objective of the analysis. For example, the best filter function for image compression may not correspond to the best filter function for image recognition. Furthermore selection of the 'best' function from a library of such functions can be a computationally demanding task (as discussed in Section 5.4.2). The challenge is therefore to find a method that selects those filter functions from the library that are best suited to the task of image recognition and to do so with minimal computation. The approach adopted in this thesis differs substantially from the best basis and matching pursuit methods that were designed for data compression (Coifman & Wickerhauser 1992; Mallat & Zhang 1993). For the task of image recognition, the upper and lower bounds on classification performance were calculated for each function in the library (across scales and locations) and were used as a means of function selection.

In order to select the best filter functions a training data set is required which consists of sufficient samples from each group to be separated. From this data set bounds on the probability of misclassification (the probability that a sample belonging to one group will be incorrectly classified as belonging to another group) can be calculated and used as a measure to compare different filter functions. The function with the smallest probability of misclassification is selected for inclusion in the adaptive image transform for that particular scale and location. The training data set consists of N images (of dimensions N_x by N_y) that can be separated into subsets according to a vector $\mathbf{G} = \{g_1, g_2, \ldots, g_n\}$ of groups. For instance, the data sets under investigation in this thesis can be separated according to whether they belong to the 'normal' g_N, the 'benign' g_B or the 'malignant' g_M groups. The data are assigned to these groups by the modeler and must be known in advance. For each of the $l = 1, \ldots, L$ filter functions (denoted $\psi_l(a, x, y)$) in the library $\mathcal{L} = \{\psi_1(a, x, y), \psi_2(a, x, y), \ldots, \psi_L(a, x, y)\}$ and for each of the $j = 1, \ldots, N$ images (denoted $I_j(x, y)$), the inner product $\mathbf{D}_{l,j,a} = < \psi_l(a, x, y), I_l(x, y) >$ is calculated at the scale $a \in \mathbb{Z}$. The result is an (1 by N by N_x by N_y) array of coefficient matrices denoted $\mathcal{D}_a$ that describes the analysis of the the lth filter function, on the jth image at the ath scale.

The structure of this array can be specified as follows,

$$
\mathcal{D}_a =
\begin{bmatrix}
\mathbf{D}_{1,1,a} & \cdots & \mathbf{D}_{L,1,a} \\
\mathbf{D}_{1,2,a} & \cdots & \mathbf{D}_{L,2,a} \\
\vdots & \ddots & \vdots \\
\mathbf{D}_{1,N,a} & \cdots & \mathbf{D}_{L,N,a}
\end{bmatrix}
\tag{7.50}
$$

where each entry describes a coefficient matrix $\mathbf{D}_{l,j,a}$, with the columns referring to coefficient matrices that are associated with each filter function and the rows to the image studied. This process is repeated across a range of scale parameters (as specified the scale parameter vector $\mathbf{a}^\star = (a_1, a_2, \ldots, a_m)$, $m \in \mathbb{Z}$) to produce the super-array of coefficients $\mathcal{D} = [\mathcal{D}_{a_1}, \mathcal{D}_{a_2}, \ldots, \mathcal{D}_{a_M}]$ whose entries are the coefficient arrays specified in equation (7.50) that are relevant to scale a_m where $m \in 1, \ldots, M$. Each entry within the array $\mathcal{D}$ is a coefficient that can be considered a feature which is denoted $z \in \mathbb{C}$. For the lth filter, the ath scale and the location specified by the pair (x, y), the value of the feature for each of the N images can be collected into a feature vector $\mathbf{z}_{l,a,x,y} = \{z_{l,1,a,x,y}, \ldots, z_{l,N,a,x,y}\}$. Now for each feature vector $\mathbf{z}_{l,a,x,y}$ we associate the known group label ($g_{n^\star} \in \mathbf{G}$, $n^\star \in 1, \ldots, n$) with the jth image to produce the pair $(z_{l,a,x,y}^j, g_{n^\star})$. In addition, the prior probabilities $\pi_{g_{n^\star}}$ for each of the n groups by,

$$
\pi_{g_{n^\star}} = \frac{N_g}{N}
\tag{7.51}
$$

where N_g is the total number of images from the group $g_{n^\star}$ and N is the total number of images in the training data set.

For the task of image recognition, the unconditional probability of misclassification for each group (denoted $\mathsf{pmc}_{g_{n^\star}}$) is of interest and is defined as,

$$
\mathsf{pmc}_{g_{n^\star}} = \Pr[\hat{c}(\mathbf{z}_{l,a,x,y}) \neq g_{n^\star},\ \hat{c}(\mathbf{z}_{l,a,x,y}) \in (g_1, g_2, \ldots, g_n) | G = g_{n^\star}]
\tag{7.52}
$$

where the symbol Pr denotes probability, the symbol G denotes the true state or group of the sample and $\hat{c}(\mathbf{z}_{l,a,x,y}) : \mathbb{C} \rightarrow (g_1, g_2, \ldots, g_n)$ denotes an arbitrary classification function that takes coefficients from the complex plane $\mathbb{C}$ and assigns a group label from the set G (adapted from equation 2.1, page 18, Ripley 1996) . Application of the classifier $\hat{c}(\mathbf{z}_{l,a,x,y})$ to the feature vector $\mathbf{z}_{l,a,x,y}$ produces a classification vector $\mathbf{C}_{l,a,x,y} = (\hat{g}_1, \hat{g}_2, \ldots, \hat{g}_K)$ of the estimate of the

appropriate group for each image using the coefficient information from the ith filter, at the ath scale and at location (x, y). If the classifier does not separate the groups perfectly, then $\hat{c}(\mathbf{z}_{l,a,x,y}) = \mathbf{C} \neq \mathbf{G}$. The total unconditional probability of misclassification pmc for all groups is then,

$$\text{pmc} = \Pr[\hat{c}(\mathbf{z}_{l,a,x,y}) \neq \mathbf{G}]. \tag{7.53}$$

In order to determine the probability of misclassification the posterior probability for each group $g_{n^\star}$ (under the assumption that the elements from the coefficient vector $\mathbf{z}_{l,a,x,y}$ is of a certain fixed magnitude denoted $z \in \mathbb{C}$) needs to be calculated. The elements in the vector $\mathbf{z}_{l,a,x,y}$ are indexed by parameter, $r = 1, \ldots, n$, so that the element $\mathbf{z}_{l,a,x,y}(r)$ is assigned value z if it belongs to group $g_{n^\star}$ (the group under consideration) and to the value zero if it is not in this group. Denote the set R to contain all of those indices r to which the value of z was assigned. Then the posterior probability for each group is defined as,

$$\begin{aligned} \Pr[G = g_{n^\star} | \mathbf{z}_{l,a,x,y}(r) = z, \ r \in R] &= p(g_{n^\star} | \mathbf{z}_{i,a,x,y}(r) = z) \\ &= \frac{\pi_{g_{n^\star}} p_{g_{n^\star}}(z)}{\sum_{l=1}^{n} \pi_{g_l} p_{g_l}(z)} \end{aligned} \tag{7.54}$$

where $p_{g_{n^\star}}(z)$ is the probability density function of the feature (coefficients of magnitude z) when only the data from the group $g_{n^\star}$ is considered and $\mathbf{z}_{l,a,x,y}(r)$ is the rth element of vector $\mathbf{z}_{l,a,x,y}$ (equation 2.4, page 19, Ripley 1996). Provided all of the classification errors are considered to be equally serious, the group with the largest value of $p(g_{n^\star} | \mathbf{z}_{l,a,x,y}(r) = z)$ (and hence the largest magnitude of $\pi_{g_n} p_{g_{n^\star}}(z)$) is deemed to be the most probable group to which the data belongs for that particular value of magnitude z. This method of assignment is known as *Bayes' rule*, it is the optimum method of classification and provides a limit on how well data from different groups can be separated.

The probability of misclassification pmc can be calculated from $p(g_{n^\star}|\mathbf{z}_{l,a,x,y}(r) = z)$ by noting that,

$$
\begin{aligned}
\mathsf{pmc} &= \mathrm{E}[1 - \max_{n^\star=1}^{n} p(g_{n^\star}|\mathbf{z}_{l,a,x,y}(r) = z), r \, \epsilon \, R] \\
&= \int_{-\infty}^{+\infty} p(z)[1 - \max_{n^\star=1}^{n} p(g_{n^\star}|\mathbf{z}_{l,a,x,y}(r) = z, r \, \epsilon \, R)]dz \qquad (7.55)
\end{aligned}
$$

where $p(z)$ is the probability density function of the data irrespective of group and when there is exact assignment of observations to groups that is there is no 'doubt' report option available to the classifier (adapted from page 20, Ripley 1996). Function selection based upon the probability of misclassification is closely related to Bayes' rule. Calculation of pmc involves using $[1 - \max_{n^\star=1}^{n} p(g_{n^\star}|\mathbf{z}_{l,a,x,y}(r) = z)]$ whilst Bayes' rule involves allocating the observation to the group with $\max_{n^\star=1}^{n} p(g_{n^\star}|\mathbf{z}_{l,a,x,y}(r) = z)$. The term $p(z)$ in equation (7.55) is the same irrespective of the group $g_{n^\star}$, thus it can be inferred that the larger the posterior probability $p(g_{n^\star}|\mathbf{z}_{l,a,x,y}(r) = z)]$ the smaller the probability of misclassification pmc. Different filter functions in the library are compared using the probability of misclassification, the function with the smallest probability of misclassification is selected for use in the adaptive image transform. By adopting this approach we ensure that the filter function with the highest posterior probability among all posterior probabilities calculated (the maximum of all $\max_{n^\star=1}^{n} p(g_{n^\star}|\mathbf{z}_{l,a,x,y}(r) = z)$) and the filter function that achieves the separation of the data (for at least one group) is selected.

Integration over $p(z)[1 - \max_{n^\star=1}^{n} p(g_{n^\star}|\mathbf{z}_{l,a,x,y}(r) = z, r \, \epsilon \, R)]$ in equation (7.55) can (at times) make the calculation of the probability of misclassification (pmc) either computationally demanding or infeasible. This obstacle is unacceptable in the adaptive image transformation where the probability of misclassification must be calculated for many filters over many scales and locations. A practical solution is to calculate bounds on the magnitude of the probability of misclassification and to use these as a guide for filter function selection. The bounds proposed are readily calculated and thus provide a realistic solution to the problem. When the conditional probability density function $p(g_{n^\star}|\mathbf{z}_{l,a,x,y}(r) = z)$ is denoted as $p(g_{n^\star}|z)$, the upper bound on the probability of misclassification is given by,

$$
\mathsf{pmc} \leq [H(\boldsymbol{\pi}) - JG(p(g_1|z), p(g_2|z), \ldots, p(g_n|z)] \qquad (7.56)
$$

where $H(\boldsymbol{\pi}) = H(\pi_1, \pi_2, \ldots, \pi_n) = -\sum_{i=1}^{n} \pi_i \log_2 \pi_i$ is a multi-dimensional version of Shannon's entropy (section 20, Shannon 1948) which considers the entropy induced by the relative

proportions of the groups and $JG[p(g_1|z), p(g_2|z), \ldots, p(g_n|z)]$ is the Jenssen-Shannon divergence which is defined as,

$$JG\big[p(g_1|z), p(g_2|z), \ldots, p(g_n|z)\big] = H\left[\sum_{i=1}^{n} \pi_i p(g_i|z)\right] - \sum_{i=1}^{n} \pi_i H\left[p(g_i|z)\right] \qquad (7.57)$$

where $H[p(g_i|z)] = -\sum_{i=1}^{n} p(g_i|z) \log_2 p(g_i|z)$ is the Shannon entropy function for the conditional probability density function $p(g_i|z)$. The Jenssen-Shannon divergence compares the entropy of the sum to the sum of the individual entropies. In heuristic terms the entropy of the entire data set (via $H[\sum_{i=1}^{n} \pi_i p(g_i|z)]$), irrespective of group, is compared to the combined effect of the entropies from each group (via $\sum_{i=1}^{n} \pi_i H(p(g_i|z))$). If there is little separation in the data (for a particular group), then both of these terms will be approximately equal and the Jenssen-Shannon divergence will be close to zero. If a particular group is clustered in a particular region of the feature space, then it will have a low entropy and hence the terms $H(\sum_{i=1}^{n} \pi_i p(g_i|z))$ and $\sum_{i=1}^{n} \pi_i H(p(g_i|z))$ will not be equal allowing the separation to be identified. The lower bound is given by,

$$\mathsf{pmc} \geq \frac{1}{4(n-1)}\big\{H(\boldsymbol{\pi}) - JG[p(g_1|z), p(g_2|z), \ldots, p(g_n|z)]\big\} \qquad (7.58)$$

(equation 5.4, page 149, Lin 1991).

The lower and upper bounds of the Bayes' probability of error for classification are readily computed using equations (7.57) and (7.58) provided that all the conditional probability density functions $p(g_{n^\star}|z)$ can be accurately estimated from the data, z. The density estimation from data problem is well studied in statistics and is well-documented, selected references include Wand & Jones (1995), Silverman (1986) and Scott (1992). These methods can be used in practice to estimate the probability of misclassification from data and thus select the filter functions from the library. In this thesis an alternative technique called exclusive-or (XOR) estimation was developed and implemented to overcome the computational challenges produced by the large number of density functions that need to be estimated. This technique is described in Section 8.2.2.

The **filter function selection algorithm** that is based upon the probability of misclassification can be incorporated in the adaptive image transform as follows,

i) For each filter function $\psi(a, x, y)$ in the library $\mathcal{L}$ of the adaptive image transform, calculate the coefficients, z and hence the bounds on the probability of misclassification provided by equations (7.57) and (7.58).

ii) Retain those coefficients that have the smallest lower bound on probability of misclassification. Set the coefficients that correspond to the same scale and location on the other filter functions in the library to zero.

iii) For those coefficients that were retained in the previous step, set to zero all of those whose lower bound exceeds 0.50.

(iv) Create the array $\mathcal{D}^\star$. This array is produced by recording the magnitude, filter index, scale and location of all the coefficients that remain non-zero from the previous three steps. A diagrammatic representation of the reduced array of coefficients is displayed in Figure 7.14.

(v) Remove the super-array $\mathcal{D}$ from computer memory but retain the array $\mathcal{D}^\star$.
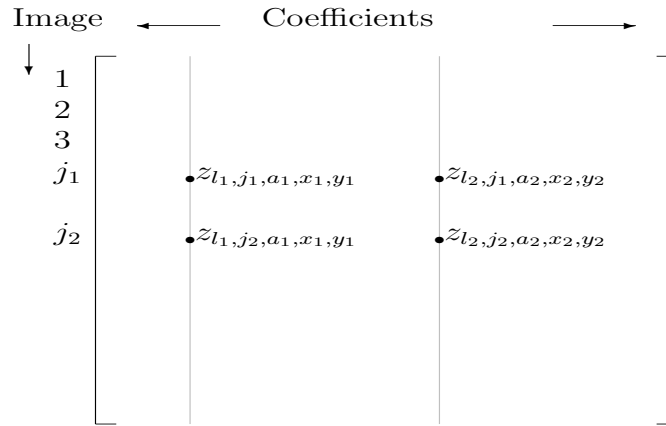

Figure 7.14: Schematic diagram to represent the array $\mathcal{D}^\star$. The coefficients $z_{l,a,x,y}^{j}$ that are selected by the filter function selection algorithm are recorded in this array where the rows correspond to different images and the columns correspond to different combinations of the parameters (l, a, x, y).

The filter function selection algorithm is applied to the super-array $\mathcal{D}$ to give a vastly reduced array (but still not small in general) of coefficients $\mathcal{D}^\star$. For the majority of cases the filter function selection algorithm will reduce the data to one complex coefficient for each scale-location, from the eight coefficients that come from the full library of filter functions in the transform. This is because one coefficient will often be optimal in the Baye's probability of misclassification. Sometimes coefficients from multiple filter functions will be retained as the algorithm may not be able to decide on a clear winner. In the array $\mathcal{D}^\star$, each column is the coefficient from a scale-location which best classifies the groups when classification is done independently. Also in the notation, the index of the filter function which produced that coefficient is retained so that cases with more than one coefficient for a scale-location can be identified. Thus the complex coefficients are given the notation $z_{l,a,x,y}^{j}$ where,

j	indicates the image number,
l	indicates the filter function,
a	indicates the scale and
x,y	indicates the location.

The filter function selection algorithm provides a method with which to select the best filter functions in the library for pattern recognition. It allows a diverse range of filter functions to be considered and their performance compared. A variety of statistical techniques can also be incorporated to estimate $p(g_{n^\star}|z)$ allowing further development and improvement of the algorithm.

7.5   Summary and Conclusion to the Adaptive Image Transform

This chapter has introduced five new functions- the Gamma, Mellin, Chebyshev, Zeta and Witch's Hat filters each of which was designed to be efficient at identifying a particular type of feature in a SAXS image. A transformation was proposed that uses these functions (plus another three from the literature) to analyse a digital image over scale and location. This transformation is adaptive in the sense that the most suitable of these functions can be selected to represent the image. Different functions can be used within the same transform allowing a flexible representation that is not offered by many other image transforms such as curvelets.

Use of Bayes" rule allows the problem of selection of the most appropriate filter function from the library to be tailored to classification problems. More precisely, it provides bounds on the Bayes' probability of error for classification for each of the *eight* filter functions in the transform. Across location and scale all of the filters are trialled. The image data for a particular scale-location is conveyed by the coefficients from the 'best' filter (out of the eight in the library) for the data. This approach to image modelling is different from other adaptive image transformation algorithms such as best orthogonal basis model which focus on sparse representations for image compression. Such a strategy is selected because the vast amount of salient information in an image is unlikely to be utilised if only a single filter was used. The library of filter functions provides a better chance of picking up image features that might be significant for detecting cancer. These features could otherwise be missed. This approach seems far more sensible than the *ad hoc* feature discovery methods that have been previously applied in the field of SAXS image analysis (refer to Chapter 3) and is driven by rigorous mathematical methodology that can be readily applied to others types of digital SAXS images, not just breast tissue. Such an approach circumvents the problem of having to search for promising diagnostic features every time a new application of SAXS technology is developed. Furthermore, we can be confident that the coefficients extracted using Bayes" rule in conjunction with the adaptive image transformation are the best available for classification within the scope of the filter functions in the library. The adaptive image transform detailed in this chapter provides features (coefficients) via a deterministic analysis, a statistical model must also be built to consider the stochastic part of the analysis and use these features to provide accurate diagnosis. The development of this statistical model will be the subject of the next chapter.