On the Use of Continuous Wavelet Transforms to Analyse Accelerometer Data Collected by the NAT Device to Characterise Parkinson's Disease

Garry Phillip Hollier

EngD

University of York

Computer Science

January, 2018

Abstract

Parkinson's disease (PD) is treated using drugs with powerful side-effects, and has similar symptoms to a range of other diseases. It is therefore important to diagnose PD accurately, and then to monitor its progression, to avoid mis-prescribing and under- and over-prescription of these drugs. Although diagnosis by experts in PD is fairly accurate, there is certainly room for improvement, and most initial diagnoses are carried out by non-expert medical staff with a much lower degree of accuracy. It is therefore desirable that automatic methods of diagnosis and monitoring be developed, and this Thesis examines the use of wavelets like those used in the Continuous Wavelet Transform (CWT), using wavelets extracted from the data itself, to try and detect features in the data pertaining to PD patients and controls.

We believe that we have successfully automatically detected distinct features, with the proviso that this has been done with very little data, opening up the possibility of tracking the development of the disease as different characteristics alter in importance, or of distinguishing subtypes of PD (analyses of the genome have determined that such subtypes exist). However, the CWTs generated by wavelets corresponding to individual features, do not suffice as a basis for diagnosis, as the features may only be intermittently present. In combination, either with other wavelets of the same type, or other methods entirely, diagnosis remains a possibility.

We used Neural Acquisition Tracker (NAT) devices to obtain triaxial accelerometer data as input to these methods, and will analyse the effect of finite bandwidth on their performance.

We believe the following list sums up the main novelty of our techniques: **Libraries of motion shapes:** at least within the field of the analysis of the motion of PD subjects; **Representing motion shapes:** by equivalence classes of piecewise polynomial wavelet triplets, which are invariant under rotations and reflections; **Distance functions:** working out the form of a distance based on the L_2 distance between individual functions, but which works on the equivalence classes mentioned above; **hierarchical k-medoids:** an approximation to k-medoids, using the clustering of cluster centres, which runs faster than k-medoids.

Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank the following for their material and moral support: my supervisors, Professor Jim Austin and Dr Chris Crispin-Bailey, The EPSRC (LSCITS), Cybula Ltd, John Charles Hollier, Steve Carter, Nick Freeman, John Daborn and Joyce Naylor

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Table of abbreviations

Abbreviation (A-M)	**Meaning**	**First occurrence**
ABN	Artificial Biological Network	p35
ADL	Activities of Daily Life	p40
AGN	Artificial Genetic Network	p35
AMN	Artificial Metabolic Network	p35
AUROC	Area Under the Receiver Operating Characteristic	p35
CGP	Cartesian Genetic Programming	p32
CWT	Continuous Wavelet Transform	p28
DWT	Discrete Wavelet Transform	p31
EA	Evolutionary Algorithm	p32
ec	Endpoint Continuity	p117
ECG	ElectroCardioGraphy	p22
ECWT	Equivalence Class of Wavelet Triplets	p63
ed	Endpoint Differentiability	p117
EEG	ElectroEncephaloGraphy	p22
EMG	ElectroMyoGraphy	p22
GoF	Goodness Of Fit	p73
ICRCGP	Implicit Context Representation Cartesian Genetic Programming	p33
iqr	InterQuartile Range	p199
MODWT/RDWT	Maximal Overlap Discrete Wavelet Transform/Redundant Discrete Wavelet Transform	p42
MRA	Multi-Resolution Analysis	p31

Abbreviation (N-*symbols*)	**Meaning**	**First occurrence**
NAT	Neural Activity Tracker	p19
PC	Principal Component	p40
PCA	Principal Component Analysis	p40
PD	Parkinson's Disease	p19
PRG	Portable Gait Rythmogram	p38
RMS	Root-Mean Square	p40
ROC	Receiver Operating Characteristic	p35
ROI	Region Of Interest	p34
SampEn	SAMPle ENtropy	p40
SVM	Support Vector Machine	p43
UPDRS	Unified Parkinson's Disease Rating Scale	p40
wec	Wrapped Endpoint Continuity	p117
wed	Wrapped Endpoint Differentiability	p117
%REC	PERCENTAGE RECurrence rate	p40

Chapter 1

Introduction

This research attempts to develop tools for the diagnosis and monitoring of PD (Parkinson's Disease).

More precisely, the aim of the research detailed here is to find methods of automatically analysing accelerometer data obtained by the patient wearing a NAT (Neural Activity Tracker) device (described below), so as to assist the reliable diagnosis of PD. The NAT is designed to be unobtrusive, and the initial idea is that the patient goes about their normal life while they wear the device.

If this work is successful, the same techniques can be used for monitoring patients to see the effect of treatment and to track the progression of the disease.

The NAT may also be used to gather other information, and an automatic analysis of the NAT accelerometer data will enable this information to be put into context.

1.1 Parkinson's Disease

PD is characterised (for this, and the rest of the section, see Wikipedia[1], [2]) by movement disorders which are collectively called parkinsonism, and include shaking, rigidity, slowness of movement and difficulties with gait.

1.1.1 Movement symptoms

The four main movement symptoms of PD are, according to Swinn *et al*[3], paucity of movement, rigidity, tremor and postural instability.

Paucity of movement is divided into three components: bradykinesia (slowness of movement), hypokinesia (reduced movement) and akinesia (inability to initiate movement).

Rigidity is resistance to movement about a joint, and is divided into "lead-pipe" and "cogwheeling" rigidity. In the former, the resistance is smooth, and in the latter it oscillates so that it feels as if movement is achieved through a mechanism involving cogs. It has been argued that rigidity is mainly due to bradykinesia, with a component of tremor in the case of cogwheeling rigidity.

Tremor is seen in about 75% of patients, and is characterised by shaking at 4-6Hz when the affected body parts are at rest. The tremor diminishes or vanishes when the body parts are in voluntary motion. Frequently, when the body part involved is the hand, the tremor appears to involve an invisible pill being rolled between forefinger and thumb.

Postural instability usually appears as the disease progresses, and has been used by some to signal the transition from mild to moderate PD. It, and problems connected with it, is described as the most disabling of the main movement symptoms of PD.

Although obvious abnormal patterns of movement at a small scale and *disregarding their speed of execution* seem to be characteristic of tremor, even being perceptible in "pill-rolling" tremor, and also of cogwheeeling rigidity, it is also possible that they are present in akinesia, postural instability and even bradykinesis, even though the latter is also likely to be characterised by patterns of motion which would be normal *were it not for their speed of execution.*

1.1.2 Other topics

In PD, a further characterisation is the presence of Lewy bodies in the neurons of the brain — these bodies consist of accumulations of the alpha-synuclein protein — and the death of dopamine-producing neurons in the midbrain region know as the substantia nigra. The term *Parkinson's disease* was initially reserved for cases where the cause of cell-death in the substantia nigra was unknown, but more links between PD and the genome are being discovered nowadays.

Dopamine is a chemical with a variety of roles in the body, but most importantly in this context it functions as a neuro-transmitter. In the brain, it is synthesised near to where it is used. Dopamine is unable to cross the blood-brain barrier, but its precursor, L-DOPA, can cross this barrier and is used to treat PD.

Nevertheless, only 5-10% of L-DOPA crosses the blood-brain barrier, which leads to overproduction of dopamine in other areas of the body and the production of side-effects such as an irregular heart-beat, nausea and joint stiffness. Prolonged use of L-DOPA also leads to involuntary movements differing from those of PD itself, and can produce punding (stereotypical movements which are compulsive rather than involuntary), and a compulsion, on the part of the patient, to overuse L-DOPA by exceeding the prescribed dosage.

In the long run, the use of L-DOPA therapy can suppress the body's own production of L-DOPA.

L-DOPA is transformed by the body to dopamine using a chemical called dopa decarboxylase, and agents such as carbidopa and benserazide, which inhibit dopa decarboxylase but are unable to cross the blood-brain barrier, are often given alongside L-DOPA. When this is done, the metabolism of L-DOPA outside the brain is reduced. Without reduction in dosage, this both

1. increases the amount of L-DOPA available for the transport across the blood-brain barrier; and

2. decreases the symptoms due to an excess of dopamine being present in processes utilising it elsewhere in the body.

In addition, 1 enables the reduction of the dosage of L-DOPA, and so the beneficial effect 2 is heightened.

The use of dopa decarboxylase inhibitors is intended to reduce the side-effects of L-DOPA, but they do persist at a reduced level, and, in some cases the combination of L-DOPA and carbidopa, or L-DOPA and benserazide, merely transforms the set of side effects associated with L-DOPA alone, reducing some, but even producing new side effects such as hallucinations or suicidal feelings.

Another characteristic of prolonged use of L-DOPA is the development of a rapid alternation of "on" and "off" phases in the patient's response to the therapy. In the "on" phase, the patient responds well to the drug and has much reduced motor symptoms, and in the "off" phase the response is low and the motor symptoms strong.

As the side effects of L-DOPA treatment, even alleviated with dopa decarboxylase inhibitors, are so onerous, medical practice is to delay its use by employing alternative drugs in the first few years after diagnosis.

The main alternative drugs are dopamine agonists and and MAO-B inhibitors.

Agonists are capable of performing at least one of the roles played the substance they are the agonist of — in this case, dopamine agonists like bromocriptine and pergolide act as neuro-transmitters — but possibly with reduced levels of effectiveness. In fact, dopamine agonists tend to be less effective at controlling PD motor symptoms than L-DOPA, but their side effects are milder (but may include hallucinations, drowsiness, insomnia, nausea and constipation) in the long term. However, in the short term, the side effects may actually be more serious than those of L-DOPA. There is also a tendency for the agonists to cause fewer side effects in younger patients, so they are more often felt to be useful for such patients than for older ones.

MAO-B is an enzyme which breaks down dopamine in the brain, and so its inhibitors elevate the concentration of dopamine there. Again, the replacement drugs are not as effective as L-DOPA in controlling PD motor symptoms, and have their own side effects.

As the disease progresses, both the dopamine agonists and MAO-B inhitiors' effects are not sufficient to alleviate the symptoms, and patients are moved to dopatherapy. However, both these drug classes have a role (which varies, depending on the patient response to them and L-DOPA) in managing the alternation between "on" and "off" phases.

There are several other diseases or conditions such that typical cases of these diseases and of PD have distinct symptoms, but atypical cases have a great overlap in symptoms. These diseases and conditions include essential tremor, several drug-induced conditions, defective copper metabolism and withdrawal, including alcohol withdrawal.

Given all this, it is important that diagnosis is accurate and rapid. However, most initial diagnosis is done by by front-line professionals, and is fairly inaccurate (75%), although expert diagnosis is much more accurate (92-4% — see Michael Lones *et al*[4] for diagnosis accuracy).

1.2 The NAT data

1.2.1 The NAT

The NAT miniaturised bio-signal data recorder is fully described in Austin *et al.*[5]. It is intended to be attached to a human or an animal, recording their movement through an accelerometer, as well as other data stemming from the analogue devices attached to its other channels.

The NAT measures 18 × 22 × 10mm (an idea of how small this is can be gleaned from Figure 1.1), and weighs less than 2.3g, including the single zinc-air battery cell. This compares to, e.g., a mouse length of 7.5-10cm (excluding the tail) and weight of 10-25g.

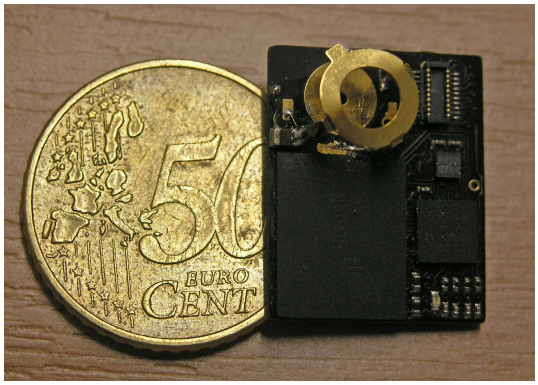


Figure 1.1: *NAT with 50 eurocent coin (picture taken from Austin* et al.*[5])*

Further specifications of the NAT are summarised in Table 1.1. As can be seen there, the accelerometer can be set to have a range of 2 or 8G (the 2G range is the one appropriate for mice studies).

The NAT can sample at rates between 100Hz and 2kHz, and both its recording capacity and the maximum length of a single recording session (dictated by battery life) increase as the sampling rate is reduced — from 24 hours to many days[1], and from 12 hours to 24 hours respectively.

The devices which may be attached to its four analogue channels include EEG (electroencephalography), ECG (electrocardiography) and EMG (electromyography) electrodes. In addition, one of the acceleration channels may be sacrificed to allow the attachment of an infra-red detector which can be used to pick up codes denoting events or for the determination of the zone occupied by an animal within its enclosure.

In Table 1.2 we present some comparision data for the devices whose use is discussed below. The UWE-PD2G, W-190L-PD2GT and M190-D2GT devices are manufactured by the Little Leonardo Co. Ltd of Tokyo. In addition, we include the Shimmer Platform, manufactured by Realtime Technologies Ltd.

Although this is partially masked by the reduced pressure for miniaturisation (which is sometimes expensive) when the devices are used on large animals, there is a discernible trend

[1] With the full specification version of the NAT

NAT Specifications		
Parameter	**Limits**	**Units**
Analogue inputs	4	channels
Bits per channel	11	bits
Accelerometer	3	axis
Bits per Acc. Axis	8	bits
Sample rate (max)	2000	Sa/Sec x 4 ch
Power at 2KHz	2.4	mA
Data Capacity	4	Gbits
Recording Time 2KSa/Sec	12	Hours
Analogue Range	±1000	uV
Accelerometer range	2 or 8	G (G-force)
Accelerometer sensitivity	18	mG at 2G range

Table 1.1: *The specification of the NAT — taken from Austin* et al.*[5]*

towards smaller devices as time passes. The NAT, at 13.9% of the mass of the lightest of the other devices (and with a "typical" length, as defined in the Table, of 63.5% of that of its nearest rival), exemplifies this trend.

1.2.2 Accelerometer data

In Figure 1.2 we show the directions of the heave, sway and surge accelerations. These are defined with respect to the tetrapod body plan, and the same definitions are used in the papers reviewed below.

When the NAT accelerometer is attached to the animal, it is clear that its axes correspond to the animal's body axes, and will register a combination of its linear and angular acceleration, in addition to the static acceleration of gravity. As it takes five independent variables (three for its path in space, two for its rotation as it follows that path) to fully describe the motion of a body, and the rotations affect the angles between the body and the space axes, it seems unlikely that the accelerations recorded by the NAT could be integrated to track the position of the animal, even if the constraints of anatomy compelled these variables to remain within a three-dimensional subspace of the full five-dimensional one.

Nevertheless, such an integration would be of little interest, and the body axes appear to be the appropriate ones to use as an animal's orientation in the horizontal plane will not affect either its behaviour nor the interpretation of it (in the absence of sensitivity to magnetic fields, at any rate).

An important characteristic of an accelerometer is the bandwidth of its response. Being made of some material, it cannot respond to extremely rapid vibrations, and its response in reaction to some excitation falls off as the characteristic frequency of the excitation increases. In general, the point at which the response is half the full response is called the bandwidth of the accelerometer.

The bandwidth of the NAT is 80Hz in all directions, and that of the Shimmer device

	Comparison of NAT with other devices				
Device	**Where described**	**Date of use**	**Geometric mean of known dimensions (cm)**	**Mass (g)**	**Capabilities [Relevant species]**
UWE-PD2G	Yoda *et al.*[6]	1998	3.37	60	Biaxial accelerometry, speed, depth [penguins]
W-190L-PD2GT	Tanaka, Takagi and Naito[7]	1999	3.72	60	Biaxial accelerometry, barometry, thermometry [salmon]
M190-D2GT	Watanuki *et al.*[8]	2003	2.49	16	Biaxial accelerometry, depth [shags]
Fleck2	Guo *et al.*[9]	2006	8.49**	?**	Triaxial accelerometry and magnetometry, GPS [cows]
HOBO® Pendant G	Moreau *et al.*[10]	2007	3.53	18	Triaxial accelerometry [goats]
Daily diary	Wilson *et al*[11], Shepard *et al.*[12]	2008†	2.91-4.81 [3.30]*	48-90 [21]*	(Not all used for all applications) Triaxial accelerometry and magnetometry, pressure, speed, internal and external temperature, 2 light intensity measurements, humidity, GPS [various, mostly megafauna]
Unnamed	Pastell *et al.*[13]	2009†	26.38	19	Triaxial accelerometry, radio transmission of data [cows]
Portable Gait Rhythmogram (PGR)	Terashi *et al.*[14]	2012†	4.58	80	Triaxial accelerometry [humans]
Shimmer Platform	‡	present	2.94	22	Triaxial accelerometry, various expansion modules available [humans]
NAT	here	present	1.58	2.3	Triaxial accelerometry and 4 analogue channels OR biaxial accelerometry, 1 IR detector and 4 analogue channels [mice]

Table 1.2: *Comparison of the NAT with other devices.* **Dimensions of the Daily Diary's (DD's) triaxial accelerometer and the rest of the device are given separately: range of box volume/weight for remainder of DD is outside square brackets, for triaxial accelerometer inside.* ** *No depth or weight information given.* †*Date of paper, not date of experiment.* ‡*http://www.shimmersensing.com/images/uploads/docs /Shimmer_Wireless_Sensor_Platform_Spec_Sheet.pdf (accessed 6th December 2013)*

is 150Hz in one direction, and 350Hz in each of the two remaining directions [15].

The NAT accelerometer data itself consists of a stream of values with constant separation in time (determined by the NAT sampling rate). The values themselves consist of multiples of $\frac{1}{18}$G, between $-2\frac{5}{18}$ and $2\frac{5}{18}$ (cf the nominal range of -2 to 2G), represented by signed 11-bit numbers.

The data we use in this Theses is collected from NATs sampling at 500Hz, attached to either the right or left wrist of our subjects. This represents an oversampling by a factor of 3.125, given that we know that the bandwidth of the accelerometer is 80Hz. We regard the accelerometer as a physical low-pass filter, so spectral content of its output at frequencies

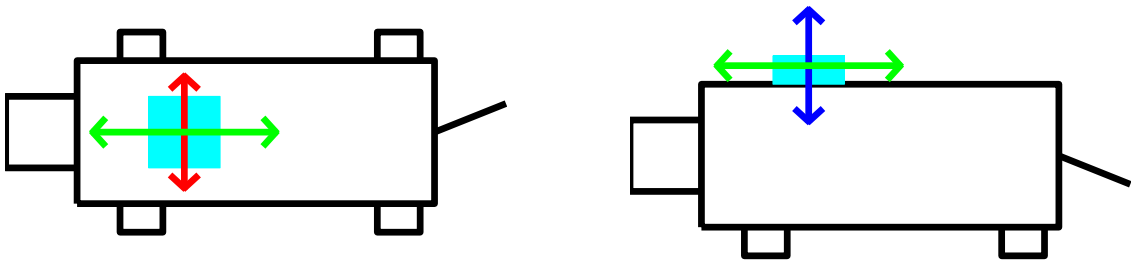


Figure 1.2: *Left: schematic of NAT on rodent, aerial view; right: schematic of NAT on rodent, side view. The blue arrow is in the direction of heave, the red, of sway and the green, of surge — see, e.g., Wilson, Shepard and Liebsch[11] for the definitions*

above 250Hz would be virtually zero. But a signal with no frequency content above 250Hz is perfectly reconstructible using a sampling rate of 500Hz by the well-known Shannon Sampling Theorem, with no aliassing.

1.2.3 Sample Fourier transforms of data

In Figure 1.3 we show the mean Fourier transform of data stemming from NATs attached to a PD subject's wrists.

For both attachment sites, the traces show the peaks of activity characteristic of PD at roughly 6Hz, although this is slightly reduced in frequency for the subject's left-hand. This is probably a real effect, as the data was collected over a reasonable length of time, although this is not certain (the data was not collected simultaneously).

In Figure 1.4, which displays the Fourier transform of similar data from two control subjects, the peak near 6Hz is absent, as would be expected.

In both Figures 1.3 and 1.4, we see that at least two of the accelerometer channels for each subject and each wrist detect a peak in the region of 2Hz, which we believe to be associated with walking.

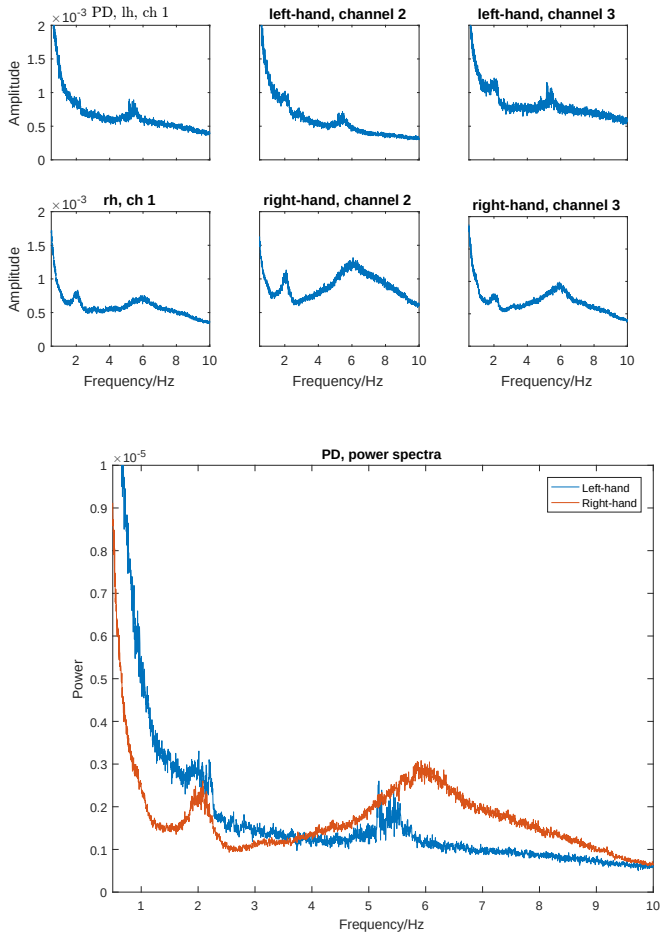


Figure 1.3: *The left-hand diagram shows the amplitude spectra for the accelerometer data stemming from a NAT attached to a PD subject's left (respectively right) wrist (top row) (respectively bottom row) for the three channels, averaged over 385 five-minute epochs (respectively 527 epochs). The right-hand diagram shows the power spectra summed over the three channels (left-hand attachment blue, right-hand red)*

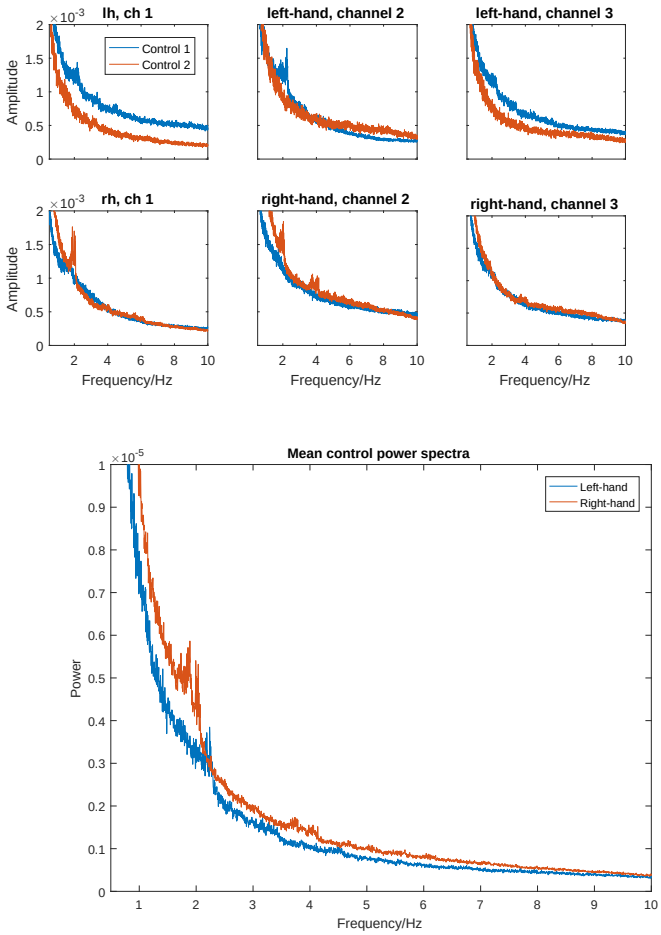


Figure 1.4: *The left-hand diagram shows the amplitude spectra for the accelerometer data stemming from a NAT attached to two control subjects' left (respectively right) wrists (top row) (respectively bottom row) for the three channels, averaged over 350 (control subject 1) and 105 (cs 2) five-minute epochs (respectively 441 and 427 epochs). The right-hand diagram shows the power spectra summed over the three channels and averaged over the two subjects*

1.2.4 Other data

The other data that it possible for the NAT to collect consists of (up to four) streams of real numbers (again represented by 11-bit numbers, and again evenly spaced according to the NAT sampling rate), stemming from the attached device, which might, for example, be EMG electrodes.

If an infra-red detector is added as a daughterboard, perhaps to detect signals specifying the varying location of the device at the times when data is collected, then its output is a stream of evenly spaced integer codes.

None of this other data is used in this Thesis, although it could be used in conjunction with the data that *is* used in future applications.

1.3 The Continuous Wavelet Transform

The initial attempt will be to use CWTs (Continuous Wavelet Transforms) to characterise accelerometer data from PD patients and controls, where the wavelet is derived from the PD patients' data itself.

With this in mind, we will construct a library of wavelet triplets (one for each accelerometer axis), matching all the windows in the data of a range of lengths in the data.[2] As this will result in an enormous library, it will be pruned to yield a much smaller one. Then, given new data, the scaleograms (the collections of inner products of the new data with the wavelets at each point in scale-time space) can be used to recognise behaviours similar to the ones present in the data used to generate the library.

The first step here, the construction of the library may be executed offline (so long as this is done in reasonable time), but it would be advantageous for the generalisation to monitoring if the second step could be done in real time.

In the context of a CWT, a *wavelet* ψ is simply an element of the set of square-integrable complex functions over $\mathbb{R}$, $L^2(\mathbb{R})$, s.t.

$$\|\psi\|^2 = \int_{-\infty}^{\infty} |\psi(x)|^2 \mathrm{d}x = 1 \tag{1.1}$$

and

$$\int_{-\infty}^{\infty} \psi(x) \mathrm{d}x = 0 \tag{1.2}$$

(the first condition implies that ψ is not identically zero, the second condition that both the real and imaginary parts of ψ are either identically zero a.e., or are sometimes positive and sometimes negative (so at most one part is identically zero a.e.) — it is a *wave*, and then the first condition implies that ψ, unlike the cosine or sine, doesn't "go on forever" — it is a wave*let*, and either has compact support, or dies off faster than $|x|^{-1}$ as $x \to \pm\infty$). In

[2] An alternative to this unsupervised learning is to find wavelet triplets matching examples of labelled data (supervised learning)

particular, in this context, there is no need for ψ and its translates and dilations to satisfy any orthonormality conditions.

Given such a ψ, and $f \in L^2(\mathbb{R})$, the CWT $W[f](a,b)$ of f is

$$W[f](a,b) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{\infty} f(x)\overline{\psi_{a,b}(x)}\mathrm{d}x = \frac{1}{|a|^{\frac{1}{2}}} \int_{-\infty}^{\infty} f(x)\overline{\psi\left(\frac{x-b}{a}\right)}\mathrm{d}x, \tag{1.3}$$

for $(a,b) \in (\mathbb{R} - \{0\}) \times \mathbb{R}$, where $\psi_{a,b}(x) = \frac{1}{|a|^{\frac{1}{2}}}\psi\left(\frac{x-b}{a}\right)$. Clearly, $\psi_{a,b}$ also satisfies the conditions given in the previous paragraph, and thus is equally a wavelet, so ψ is called the *mother* wavelet where confusion might arise.

Such a ψ is called *admissible* if

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\xi)|^2}{|\xi|}\mathrm{d}\xi < \infty, \tag{1.4}$$

where $\hat{\psi}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-ix\xi}\psi(x)\mathrm{d}x$ is the Fourier transform of ψ.

For an admissible wavelet, f can be reconstructed from its CWT via $f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W[f](a,b)\psi_{a,b}(x)\mathrm{d}b\frac{\mathrm{d}a}{a^2}$.

However, we do not require our wavelets to be admissible, as we merely wish to use wavelets to characterise our data, not provide an alternative representation of it.

For details about the definition of the CWT, see Rao and Bopardikar[16].

Chapter 2

Literature review

2.1 The Discrete Wavelet Transform

Hernández and Weiss[17] are the source for the material here.

The DWT (Discrete Wavelet Transform) is defined similarly to CWT in equation (1.3), with the restriction of $(a, b) \in \{(2^{-j}, 2^{-j}k) : (j, k) \in \mathbb{Z}^2\}$ and different conditions on ψ.

Relabelling $\psi_{2^{-j},2^{-j}k}$ by $\psi_{j,k}$, ψ must be such that

I $\langle \psi_{j,k}, \psi_{\ell,m} \rangle = \delta_{j\ell}\delta_{km}$;

II $\overline{[\{\psi_{j,k} : (j,k) \in \mathbb{Z}^2\}]} = L^2(\mathbb{R})$, i.e., the closure of the span of $\{\psi_{j,k} : (j,k) \in \mathbb{Z}^2\}$ is $L^2(\mathbb{R})$.

Clearly, condition I implies equation (1.1) and, if ψ is not badly behaved, it turns out that conditions I and II together imply equation (1.2), so ψ is suitable for a CWT if it is suitable for a DWT.

For the rest of this section, a wavelet will mean a ψ such that conditions I and II hold.

It also turns out that well-behaved wavelets can be derived from an MRA (Multi-Resolution Analysis).

An MRA is a sequence of subspaces $\{V_j : j \in \mathbb{Z}\}$ of $L^2(\mathbb{R})$ s.t.

I $V_j \subset V_{j+1}$;

II $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}\ \forall j \in \mathbb{Z}$;

III $\cap_{j\in\mathbb{Z}} V_j = \{0\}$;

IV $\overline{\cup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$;

V $\exists \phi \in V_0$ s.t. $\{\phi(x-k) : k \in \mathbb{Z}\}$ is an orthonormal basis for V_0.

The function ϕ, whose dyadic scalings and shifts are denoted $\phi_{j,k}(x) = 2^{\frac{j}{2}}\phi(2^j x - k)$, is called the *scaling function* for the MRA.

That such MRAs exist is demonstrated by the *Haar* function, $\phi_H(x) = \begin{cases} 1, & x \in \left[-\frac{1}{2}, \frac{1}{2}\right]; \\ 0, & \text{otherwise}: \end{cases}$

the subspaces are then $V_j = \overline{[\{\phi_{jk} : k \in \mathbb{Z}\}]}$.

Given an MRA, a wavelet ψ can be derived s.t. $\{\psi_{j,k} : j,k \in \mathbb{Z}\}$ is an orthonormal set and $V_{j+1} = V_j \oplus W_j$ and $\bigoplus_{j\in\mathbb{Z}} W_j = L^2(\mathbb{R})$ for $W_j = \overline{[\{\psi_{j,k} : k \in \mathbb{Z}\}]}$ (called the *detail space* at level j).

This wavelet is given by

$$\hat{\psi}(\xi) = \mathrm{e}^{\frac{1}{2}i\xi}\nu(\xi)\overline{m_0\left(\frac{1}{2}\xi + \pi\right)}\hat{\phi}\left(\frac{1}{2}\xi\right), \tag{2.1}$$

where $\nu(\xi)$ is an arbitrary 2π-periodic measurable function s.t. $|\nu(\xi)|$ and $m_0(\xi) = \frac{\hat{\phi}(2\xi)}{\hat{\phi}(\xi)}$ is the *low-pass filter* associated with ϕ.

In the case of the MRA with scaling function ϕ_H, the usual Haar wavelet $\psi_H(x) = \begin{cases} 1, & x \in \left[-\frac{1}{2}, 0\right); \\ -1, & x \in \left[0, \frac{1}{2}\right); \\ 0, & \text{otherwise :} \end{cases}$ is obtained by taking $\nu(\xi) = 1$; alternative versions can be obtained by making other choices.

The advantage of using a DWT over a CWT is that there exist algorithms analogous to the fast Fourier transform for calculating $\{\langle f, \psi_{j,k}\rangle : k \in \mathbb{Z}\}$ and $\{\langle f, \phi_{j,k}\rangle : k \in \mathbb{Z}\}$, given $\{\langle f, \psi_{j+1,k}\rangle : k \in \mathbb{Z}\}$ and $\{\langle f, \phi_{j+1,k}\rangle : k \in \mathbb{Z}\}$, providing that ϕ (and consequently ψ) has compact support. If ϕ has sufficiently good decay, the fast algorithms may still be used, at the cost of introducing some error.

The disadvantage of using a DWT in our context is that the conditions I and II are much more difficult to integrate into a pattern matching context than are the conditions given by equations (1.1) and (1.2).

The situation is similar when $\{\psi_{j,k} : j,k \in \mathbb{Z}\}$ is *not* an orthonormal system, but we know a *dual basis* $\{\tilde{\psi}_{j,k} : j,k \in \mathbb{Z}\}$ $\langle \tilde{\psi_{j,k}}, \psi_{\ell,m}\rangle = \delta_{j\ell}\delta_{km}$: compact support leads to fast algorithms (but with a more complicated derivation), rapid decay leads to approximate fast algorithms.

2.2 Parkinson's Disease

The purpose of reviewing some papers concerning the diagnosis of Parkinson's disease is that the NAT, or an assembly of NATs, could be used to monitor patients in their daily life in order to diagnose, and assess changes in, PD. The methods discussed in these papers could be compared with methods derived by using the approach described here, or combined with them.

2.2.1 Cartesian Genetic Programming and its Application to Medical Diagnosis, *Smith[18]*

The author describes various types of CGP (Cartesian Genetic Programming).

The simplest version deals with the application of an EA (Evolutionary Algorithm) to a population of sets of input nodes, intermediate processing nodes arranged in a grid and

a final output node, where the processing is controlled by a chromosome. The nodes are numbered from 0, starting with the input nodes, then the intermediate processing nodes and finally the output node. Each processing node has two inputs, equal to the outputs of two preceding nodes (if the grid is one-dimensional, each processing node has one immediate predecessor, and, if it is higher dimensional it has a whole layer of immediate predecessor nodes, but, in both cases, all predecessor nodes are candidates for providing the input, not just the immediate predecessor(s)) and an associated function from a finite, indexed, function table. Each function has (at least formally) two arguments. Each node is controlled by a gene of three integers, the first two indexing into the set of nodes, and the third into the function table. The node output is the result of applying the function specified by the third number to the inputs specified (in order) by the first two numbers in the gene. The chromosome consists of the set of genes for each processing node, plus the index of the processing node whose output is presented at the output node.

Evolution is achieved by cloning a grid with a randomly generated chromosome and mutating the numbers defining the chromosomes of the clones at a given rate. At each generation of the evolution process, the outputs of each clone given a set of inputs are compared to a set of desired outputs using some objective function, and the grid with the best value of this objective function is kept, while the remaining grids are discarded. The winning grid is then cloned and the next evolutionary step is carried out.

When some terminating condition (a target value of the objective function is achieved, or a number of generations have passed, for example), the process is terminated and the winning grid is used to classify new examples.

A variation of CGP is ICRCGP (Implicit Context Representation CGP). "Ordinary" CGP suffers from the defect that the evolution of programs is affected by the order in which genetic information is stored, thus potentially hiding the effect of the genes. ICRCGP seeks to avoid this by interpreting each node as an enzyme which binds to similar enzymes in the network.

Each enzyme (node) has two binding sites consisting of vectors of integers (for example, in the range 0-255) which are randomly assigned initially, and randomly mutated, as is the node's (enzyme's) function. If there are n node functions available, then the last n of the binding site vectors' entries correspond to these. The node's shape is a convex linear combination of its binding site vectors. For example, if the binding site vectors have m components, and the node's function is the kth, then s, the vector specifying the shape is given by

$$s = a_1 b_1 + a_2 b_2 + (1 - a_1 - a_2) \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 255 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{2.2}$$

where the b's are binding site vectors, fixed a_1 and a_2 satisfy $0 < a_1, a_2, 1 - a_1 - a_2 < 1$ and the 255 in the last vector is in the $(m - k + 1)$th place. Of course, rounding is employed to keep the components of s integral.

The node bindings are found by calculating the Euclidean distance between each enzyme shape and every binding site belonging to a node ahead of it in the grid. The enzyme pair with the smallest distance between the shape and binding site are bound, the binding site is deleted from a list of available sites, and the enzyme pair with the smallest distance between the shape and an available binding site is then selected for binding. This process proceeds until all binding sites are used up.

This node binding process replaces that of ordinary CGP, and everything proceeds as before[1].

The paper applies CGP to two areas of medicine, the diagnosis of PD, and the detection of breast cancer.

According to the paper, there is an initial misdiagnosis rate of 25% for Parkinson's, which underlines the need for new diagnostic tools.

Two tests were used to generate data from Parkinson's patients and controls: in the first, the patients had to trace out, on a digitalising tablet, a piecewise linear spiral with 5 turns per twist, and, in the second they had to tap the index finger and thumb together on alternate hands for 30s, while connected to electromagnetic sensors.

In both cases, an acceleration profile was derived from the measuring devices and a sequence of windows of ten successive values presented to the ten inputs of an ICRCGP with 30 processing nodes.

The results clearly show that the ICRCGP detected many more artefacts in the Parkinson's patients performance on these tests than in that of the controls, but, it is unclear whether the presented results relate to the re-presentation of the data on which the ICRCGP was trained, or to some test data.

The work on breast cancer concentrated on a subproblem, the classification of microcalcifications (small calcium-containing deposits). Some geometries of microcalcifications are indicative of malignancy, although most are benign.

ROIs (Regions Of Interest) containing microcalcifications were segmented from mammograms, and subjected to three procedures involving CGPs:

i) a 10-dimensional vector of statistical features of the ROI was presented to an ICRCGP with 10×3 intermediate processing nodes;

ii) the raw pixel values of the ROI were presented to an ICRCGP;

iii) the raw pixel values of 8×8 non-overlapping parts of a ROI were presented to a corresponding number of ICRCGPs, and their chromosomes were swapped, or one replaced by another, at a mutational rate, in addition to the usual evolution.

[1] From the diagrams it appears that the output node has two inputs and also processes these for the final result, although this is a small detail

Method i) produced an overall classification accuracy of 70% with an AUROC of 0.69, and method ii) an AUROC of 0.78. (The Area Under the ROC — AUROC — is equal to the probability that a subject with a malignant condition would score more highly than one with a benign condition. The ROC — Receiver Operating Characteristic — is a plot of the true positive rate vs the false negative rate, parameterised by the threshold defining the classification boundary).

2.2.2 Characterising neurological time series data using biologically motivated networks of coupled discrete maps, *Michael Lones* et al[4]

This paper describes evolved ABNs (Artificial Biological Networks), in particular AMNs (Artificial Metabolic Networks) and AGNs (Artificial Genetic Networks), and their application to the task of detecting the signs of PD. The data used in the study stemmed from 49 PD and 41 control subjects.

An AMN consists of the five element set $\{C, E, L_C, I_C, O_C\}$ where

C is the set of chemical concentrations, $\{c_0, \ldots, c_n\} \subset \mathbb{R}^{n+1}$;

E is the set of enzymes, $\{e_0, \ldots, e_q\}$, $e_i = \{S_i, P_i, m_i\}$, where:

- $S_i \subseteq C$ is the set of enzyme substrates;
- $P_i \subseteq C$ is the set of enzyme products;
- $m_i : \mathbb{R}^{|S_i|} \to \mathbb{R}^{|P_i|}$ is the substrate-product mapping;

L_C is an indexed set of initial concentrations, with $|L_C| = |C|$;

$I_C \subset C$ is the set of chemicals used as external inputs;

$O_C \subset C$ is the set of chemicals used as external outputs.

The AMN is executed by synchronously updating the elements of C, by taking the outputs of the elements of E when each is applied to its range (if a product belongs to the intersection of a set of P_i, its mean value over that set is taken). If necessary, the values of the c_i are uniformly scaled so that $\sum_{c_i \in C} c_i = \frac{1}{2}|C|$, to mimic mass conservation. After a given number of steps, the values of the c_i in O_C are taken to be the network's output.

The enzymes are selected from the following set of functions:

logistic map: $[0,1] \to [0,1], x \mapsto rx(1-x)$, $r \in [0,4]$;

Chirikov's standard map: $[0,1]^2 \to [0,1]^2, \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x + y - \frac{k}{2\pi}\sin(2\pi x) \\ y - \frac{k}{2\pi}\sin(2\pi x) \end{bmatrix} \mod 1$, $k \in [0,10]$;

baker's map: $[0,1]^2 \to [0,1]^2, \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{cases} \begin{bmatrix} 2x \\ \frac{y}{2} \end{bmatrix}, & x \in \left[0, \frac{1}{2}\right]; \\ \begin{bmatrix} 2-2x \\ 1-\frac{y}{2} \end{bmatrix}, & x \in \left[\frac{1}{2}, 1\right]; \end{cases}$

Arnold's cat map: $[0,1]^2 \to [0,1]^2, \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 2x+y \\ x+y \end{bmatrix} \mod 1,$

all of which (when iterated) have both regions with ordered and regions with chaotic behaviour, or the

logistic function $[0,1] \to [0,1], x \mapsto \frac{1}{1+e^{ax}}$.

An AGN consists of the four element set $\{G, L_G, I_G, O_G\}$ where

G is the set of genes, $\{g_0, \ldots, g_n\}$, $g_i = \{\lambda_i, R_i, f_i\}$, where:

$\lambda_i \in \mathbb{R}$ is the gene's expression level;

$R_i \subseteq G$ is the set of the gene's regulatory inputs;

$f_i : \mathbb{R}^{|R_i|} \to \mathbb{R}$ is the gene's regulatory function;

L_G is an indexed set of initial expression levels, with $|L_G| = |G|$;

$I_G \subset G$ is the set of genes used as external inputs;

$O_G \subset G$ is the set of genes used as external outputs.

The AGN is executed by synchronously updating the expression levels of the elements of G, by applying each gene's regulatory function to the expression levels of its inputs. After a given number of steps, the values of the expression levels of the genes in O_G are taken to be the network's output.

The regulatory functions are selected from the logistic map, Chirikov's map, the baker's map, Arnold's cat map and the logistic function.

Both types of ABN are applied to a time series of acceleration data in the same way: the concentration level/expression level of the single member of I_C/I_G is set to the value of a datum, the ABN is iterated for t_b steps, and another datum is presented. After all data has been presented, the ABN is iterated for another t_a, and then the value of the single member of O_C/O_G is read as the output.

In the EA for the ABNs, the connections between the enzymes/genes, the parameters of the functions, t_a and t_b are all encoded in a chromosome, and these chromosomes are evolved with a point mutation rate of 6% and crossover rate of 15%. Fitness was assessed by finding the AUROC over a set of training data.

The results for the best-performing ABN, when evaluated by the AUROC performance statistic, fall between the diagnostic performance of non-expert secondary carers (75%) and that of experts (92-4%), being only a few percent below the latter.

Thus, there seems to be great scope for the algorithms to be used at the stage of initial diagnosis.

The best-performing network is an AMN, and the rest of the paper consists of an analysis of the evolved AMNs.

The worst AMNs, in terms of the AUROC over the test data have short chromosomes, but the best ones have a mixture of short and long chromosomes. (This contrasts with the usual situation, where long chromosomes correspond to overtraining and poor generalisation.)

For ease of analysis, the best performing AMN with a short chromosome is chosen. It has an AUROC of 0.91 on the test data and four enzymes, three of which are chaotic. Knocking out any single enzyme reduces the AUROC substantially. The output chemical concentration is not directly manipulated by the enzymes, but is set by the changing concentrations of the other chemicals through the operation of the conservation law. When applied to the real acceleration data in the test set, the output concentration *eventually* falls when the data comes from PD subjects, but may remain high or even have a temporary increasing trend during the presentation of the data. Similarly, when the data comes from controls, the final concentration is high, but may fall below the optimal decision threshold for its final value during the processing of the data. This indicates that this trained AMN integrates the succession of abnormal features present across the trace.

Replacing real subject data with possibly perturbed sine waves shows that the system has sensitivity to frequencies falling below the normal[2], amplitudes falling below the normal[3] and to noise[4]. In particular, one of the internal (i.e., not the input nor the output) concentrations shows spikes in response to noise, which are presumably propagated through the AMN to lower the output concentration.

The impulse response of the network is very sensitive to the magnitude of the input. When there is no input, the network settles into a state with constant values of the concentrations, when the impulse has magnitude 0.4 the network settles into a state with different constant values of the concentrations, and when the magnitude is 0.5 or 0.6, the network settles to periodic behaviours of the concentrations (qualitatively different in each case).

Finally, if the conservation law is removed, the response to real data is chaotic.

The authors conclude that the chaotic behaviour is necessary to obtain the sensitivity of the AMN to the differing inputs stemming from PD patients and controls, but this behaviour needs to be damped by the conservation law.

2.2.3 Evolving computational dynamic systems to recognise abnormal human motor function, *Lones* et al*[19]*

This is primarily a results paper, summarising Lones *et al*[4]).

2.2.4 Evolving classifiers to recognise the movement characteristics of Parkinson's disease patients, *Lones* et al.*[20]*

This paper also describes ABNs as applied to the data in Lones *et al.*[4], but in more detail. This data was also processed using the CGPs (for brevity — they were actually the ICRCGPs)

[2]Obviously relating to bradykenesis — slow execution of movement
[3]Relating to movement fatigue
[4]Relating to tremor

described in Smith[18]. Data measuring rest movement from the same subjects was also gathered, to assess the presence of tremor.

The AUROCs corresponding to the ABNs and the CGPs were much higher than those based on extracting gross features, such as the mean amplitude or mean speed from the movement data time series, and the presence of tremor was not very consistent, so this would be of limited diagnostic value.

The study also found that the AUROCS associated with data from the subjects' dominant hands were higher, which tends to confirm that PD signs are stronger on the dominant side.

Finally, the paper finds that, although the AUROCs for ABNs tend to be higher than for CGPs, there is mileage in combining the two types of classifier.

2.2.5 Deficits in scaling of gait force and cycle in Parkinsonian gait identified by long-term monitoring of acceleration with the portable gait rhythmogram, *Terashi* et al*[14]*

The authors here discuss the use of their recently developed PRG (Portable Gait Rythmogram) in the detection of gait abnormality due to PD.

The PGR is a three-axis accelerometer and recording device which weighs 80g (*cf* NAT weight of <2.3g) and measures 80×60×20mm (*cf* 18×22×10mm for the NAT). It has a sampling rate of 100Hz (*cf* the NAT's range of 100-2000Hz), and can record 40 hours of data (*cf* the NAT's recording duration of 24 hours at this rate). In use, the PGR is attached to the patient's waist.

The authors starting point was that the variation in gait in response to changing circumstances is attenuated in PD patients, and they wished to investigate the manifestation of this attenuation in the acceleration traces.

The subjects were 40 PD patients able to walk unaided and with no dyskenesia while taking their medicine at peak dose (the PD patients were on a regime with "on" and "off" — no medication — phases), and 17 age- and height-matched controls.

Data was then collected as the subjects went about their normal daily life.

Step cycles in the data were detected by using a template from the triaxial accelerometer data (a_x, a_y, a_z), consisting of a series of p consecutive readings denoted $(a_x(1), a_y(1), a_z(1)), (a_x(2), a_y(2), a_z(2)), \ldots, (a_x(p), a_y(p), a_z(p))$ (with a relabelling of the data if necessary) corresponding to an instance of a cycle, and then finding the cross-correlation

$$CC(t) = \frac{\sum_{i=1}^{p}(a_x(i)a_x(i+t) + a_y(i)a_y(i+t) + a_z(i)a_z(i+t))}{\sqrt{\{\sum_{i=1}^{p}(a_x(i+t)^2 + a_y(i+t)^2 + a_z(i+t)^2)\}\{\sum_{i=1}^{p}(a_x(i)^2 + a_y(i)^2 + a_z(i)^2)\}}}$$

(where t may be negative if the relabelling is not trivial). The step cycles were then identified by finding the large peaks in the CC.

Then the mean scalar acceleration and cycle length were extracted from each gait cycle and then the mean of these quantities was taken over each period of ten minutes (or excluded

if there were fewer than 20 cycles in an hour). Each data pair averaged over ten minutes represented a data point in the further analysis.

The logarithms of the mean scalar acceleration and of the cycle length were found to have an approximately normal scatter for each subject, so a linear regression was carried out between these quantities (resulting in a power law between the raw quantities).

For both the mean acceleration and the cycle length, the standard deviations across the data points relating to each control subject were calculated, and then the mean of these standard deviations was calculated across the controls. These means (of the standard deviation for both the mean acceleration and cycle length) were defined to be the alteration ranges of the control subjects.

If the hypothesis that gait variability is attenuated in PD is correct, the standard deviation of the mean acceleration and of the cycle length for PD subjects would be expected to be lower, so 75% of the alteration range of these quantities calculated for control subjects was adopted as a threshold.

The authors found that the mean acceleration over a step cycle was significantly lower in PD patients, although there was no significant difference in the length of step cycle. For the mean acceleration, 16 out of 40 PD patients had a standard deviation below 75% of the alteration range, for the cycle length, 12 out of 40 PD patients fell below 75% of the alteration range, and 4 patients were below 75% of the alteration range for both quantities.

An analysis of the individual patient scattergrams of mean scalar acceleration and cycle length showed that these reflected the reduced variability of the patients with regard to one or both measures in a fairly simple way.

The authors are aware that the results may reflect the reduced physical activity of PD patients, and, as the patients were capable of attending their hospital on foot, they were asked to do so wearing the PGR, so as to provide data covering periods of at least moderately strenuous activity.

However, it is not clear whether the patients who fell into the 4 categories defined by the thresholds on the percentage alteration range for the two measures were suffering from different underlying impairments, or merely adopted different strategies for dealing with impairment.

2.2.6 EMG and acceleration signal analysis for quantifying the effects of medication in Parkinson's Disease, *Rissanen* et al.*[21]*

In this paper, Rissanen *et al.*[21] sought to evaluate the time profile of the effectiveness of various treatments for PD.

Nine patients, taking 2 or 3 drugs in various proportions from the list Madopar, Eldepryl, Stalevo, Sifrol, Sinemet, Ipsatol and Efexor were asked to undergo a test while in various conditions: 24 hours after the last medication (medicine "off") and 1, 2 and 3 hours after taking medication.

The test consisted of holding the elbow at 90° with the palm upwards for 30s. While this

was done, 1kHz tri-axal accelerometer and EMG (from the biceps) recordings were taken.

The data was divided into overlapping epochs of 2048ms, overlap 1536ms, and, after some filtering to remove artefacts, four features were extracted (giving 8 features when divided into left and right-hand versions). These features were:

1) EMG kurtosis[5], $k = \frac{E[(x-\mu)^4]}{\sigma^2}$;

2) %REC (PERCENTAGE RECurrence rate) of the EMG signal, which is a measure of the number of recurring structures in the data;

3) RMS (Root-Mean Square) of scalar acceleration;

4) SampEn (SAMPle ENtropy) of scalar acceleration,

$$-\ln \frac{\overline{\overline{\{(k,\ell) : \max\{|a_k - a_\ell|, |a_{k+1} - a_{\ell+1}|, |a_{k+2} - a_{\ell+2}|\} < r, 1 \le k < \ell \le 2046\}}}}{\overline{\overline{\{(k,\ell) : \max\{|a_k - a_\ell|, |a_{k+1} - a_{\ell+1}|\} < r, 1 \le k < \ell \le 2047\}}}},$$

where a is the scalar acceleration, r is some given constant and $\overline{\overline{S}}$ is the cardinality of S.

The authors quote in papers in which features 1) and 2) were higher for PD patients than for controls, and feature 4) was higher for controls than for PD patients.

The extracted 8-component feature vectors were subjected to a PCA (Principal Component Analysis) and PC1 (the first Principal Component) was found to explain 71% of the variation. Adding in further components did not increase the discrimination power of the method, so the authors made further analysis using PC1 alone.

If the components of the feature vector are $(k_r, k_\ell, \%\mathrm{REC}_r, \%\mathrm{REC}_\ell, \mathrm{RMS}_r, \mathrm{RMS}_\ell, \mathrm{SampEn}_r, \mathrm{SampEn}_\ell)$ (r and ℓ subscripts stand for right and left hand here), then the signature of the inner products of PC1 with its components is $+++++++--$, which is what would be expected given the correlations between the features and the presence or absence of PD mentioned above.

The patients where also assessed by using the section of the UPDRS (Unified Parkinson's Disease Rating Scale)[6] for motor function, and it was found that their score fell from its

[5] "For this measure, higher kurtosis means more of the variance is the result of infrequent extreme deviations, as opposed to frequent modestly sized deviations", or, to put it another way, higher kurtosis means the distribution is more heavy-tailed, Wikipedia[22]

[6] "The UPDRS is made up of the following sections:
Part I: evaluation of Mentation, behavior, and mood;
Part II: self-evaluation of the ADLs (Activities of Daily Life) including speech, swallowing, handwriting, dressing, hygiene, falling, salivating, turning in bed, walking, cutting food;
Part III: clinician-scored motor evaluation;
Part IV: Hoehn and Yahr staging of severity of Parkinson disease.
Part V: Schwab and England ADL scale,"
Wikipedia[23].
Part III includes assessments of speech, facial expression, tremor at rest, action tremor, rigidity, finger taps, hand movements, rotation of the hands and forearms so that the palms face downward, rotation of the hands and forearms so that the palms face upward, leg agility, rising from chair, posture, gait, postural stability and bradykinesia, European Parkinson's Disease Association[24]

high level in the "off" condition to a minimum two or three hours after medication, and then began to rise.

For all patients bar two, the PC1 values also fell then rose, although the timings were slightly different, so the PC1 value measures something different to UPDRS. In one of the two exceptions, the initial fall in the UPDRS score was reflected in the PC1 value; however, the latter continued to fall, which leaves open the possibility that the PC1 value could rise later. Thus it seems the method is a step towards an objective method for measuring the severity of the signs of PD.

One obvious question is what the handedness of the patients was, and, if they did not all have the same handedness, would the results have been any different had the features been split along dominant/non-dominant rather than left/right lines.

2.2.7 Work involving the author

Two papers, Austin *et al.*[5] and Bailey *et al.*[25] have been produced featuring Fourier analyses of data stemming from NATs attached to a PD and several non-PD subjects. Similar data is analysed below, but a feature of the Fourier analyses is that the PD subject has a peak in their trace around 6Hz, in line with expectations of heightened activity in the range 4-6Hz (see, e.g. National Parkinson Foundation[26]), and this will be relevant below.

2.3 Animal behaviour

Animal behaviour can be investigated through motion sensors, including accelerometers, and so papers recounting such investigations form an important thread in our literature survey. The techniques involved may, at least in principle, be transferred to the analysis of human motion, and we concentrate on those techniques here.

2.3.1 Prying into the intimate details of animal lives: use of a daily diary on animals, *Wilson, Shepard and Liebsch[11]*; Identification of animal movement patterns using tri-axial accelerometry, *Shepard* et al.*[12]*

Here, Wilson, Shepard and Liebsch[11] and Shepard *et al.*[12] discuss the use of a *daily diary*, a device which is attached to 37 species, mostly megafauna, and includes a triaxial accelerometer, as well as GPS location, magnometers and other measuring hardware. Their treatment of the outputs is consists of picking out by eye of patterns of acceleration consistent with various behaviours.

The authors point out that it is important in their methodology to separate the static acceleration due to gravity, and the dynamic acceleration due to the animals' movements. The static acceleration is useful for making deductions about animal posture from its distribution across the three axes. However, there is little detail on how this separation is achieved,

beyond saying that it is derived from a running mean of the sway and surge accelerometer readings.

A classifier may not require an explicit separation of static and dynamic acceleration to distinguish between behaviours, but it would be interesting to see what effect preprocessing the data in this way has on classification performance.

2.3.2 A wireless accelerometer system with wavelet analysis for assessing lameness in cattle, *Pastell* et al.*[13]*

To try to develop automatic methods to detect lameness in cattle, Pastell *et al.*[13] use the MODWT (Maximal Overlap DWT) to analyse accelerometer data. The MODWT is synonymous with the RDWT (Redundant DWT, Wikipedia[27]). The RDWT is "essentially an undecimated version of the DWT" (Fowler[28]), and consequently, the "details" and the "smooths" at each level have the same length as the original signal.

The authors attached a transmitting tri-axial accelerometer to each leg of a walking cow to obtain their data.

In order to remove the static acceleration due to gravity (to first order, at least), the authors transformed the trace from each channel by subtracting its mean.

Given the variances σ_L^2, σ_R^2 for the data stemming from paired legs, the authors define the symmetry of variance thus: $S_{\text{var}} = \frac{\min\{\sigma_L^2,\sigma_R^2\}}{\max\{\sigma_L^2,\sigma_R^2\}}$.

The use of MODWT enables the variance of the acceleration data to be split into parts which relate to different scales on which the signal can be considered. If $W_{j,t}$ is the tth component of the jth level "detail", then $\sigma_j^2 = \frac{1}{N}\sum_{t=0}^{N-1} W_{j,t}^2$ is the variance associated with that detail, and $S_{D_j} = \frac{\min\{\sigma_{Lj}^2,\sigma_{Rj}^2\}}{\max\{\sigma_{Lj}^2,\sigma_{Rj}^2\}}$ is the corresponding symmetry of variance.

The analysis was carried to the fourth level, and, given the experimental parameters, this meant that the time scales of the four levels were 40, 80, 160 and 320ms. The analysis was applied to each component of the acceleration, and to the scalar acceleration.

The authors found that there was a significant difference ($p < 0.001$ on the Wilcoxon[7] signed rank test) between S_{var} for sound and lame cows when this was calculated for forward acceleration of the hind legs.

The differences in S_{var} for the other channels and the scalar acceleration for the hind legs, and for all channels and scalar acceleration for the fore legs were not significant.

A finer analysis of the data was enabled by using the S_{D_j}; it was found that there were no significant differences for any combination of channels (or scalar acceleration), detail levels and paired legs, except for forward acceleration at detail level 1 for the hind legs, where again $p < 0.001$. Detail level 1 corresponds to a time-scale of 40ms, and this is around the duration of the gait stage in which the foot hits the ground.

[7] See [29] for the Wilcoxon rank sum test. We will use this test quite a lot later in the Thesis

2.3.3 Animal behaviour understanding using wireless sensor networks, *Guo* et al.*[9]*

Here, the authors examine the use of the Fleck2 hardware device, which can store and transmit data derived from its sensors — GPS, tri-axial accelerometer, tri-axial magnetometer and thermometer — for the understanding of cattle behaviour. The device is attached to cows using a collar which also carries GPS and RF antennæ, and batteries.

The authors use a moving mean of the acceleration as an approximation to the static acceleration caused by gravity, and then use this information to derive the dynamic acceleration.

Through a visual inspection of the traces of

1 the dynamic acceleration components;

2 the pitch and roll angles, derived from the static acceleration components, and the heading angle, derived from the pitch and roll and the magnetometer data;

3 the rate of change of the heading angle,

the authors were able to distinguish various behaviours. In 1, greater variance was visible during grooming and grazing; in 2, small rapid oscillations are seen in the pitch and roll angles during grooming and grazing; and, in 3, the magnitude of the trace is much higher in grooming and grazing.

The authors show that the classification of behaviours by eye from traces derived from their three data sets (GPS information; accelerometer and magnetometer information; navigation solution information) is feasible, but they do not automate this classification.

2.3.4 Cow behaviour pattern recognition using a three-dimensional accelerometer and Support Vector Machines, *Martiskainen* et al.*[30]*

Here, an attempt is made to classify cow behaviour using data from a collar-borne tri-axial accelerometer with an 8-bit AD converter which transmits data.

The accelerometers and associated apparatus were attached to the collars of these cows while they were housed in a barn with free access to a pasture. Ground truth was obtained via video recordings.

The data was windowed, resulting in samples of length 10s, consisting of 100 three-dimensional data points.

A 28-component feature vector was generated from these samples, including the mean, standard deviation, skew, maximum, minimum and energy for each accelerometer channel, plus the correlations between the channels. Each component of the feature vector was rescaled to lie in the range $[0, 1]$, and the result of these procedures was a set of 10^4 of these normalised 28-component feature vectors.

70% of these feature vectors were used to train a set of SVMs (Support Vector Machines), one for each of the behaviours to be identified. For each behaviour, the SVM is

trained to classify the feature vector into those appertaining to the behaviour or not appertaining to the behaviour. These binary classifiers were combined into a multiclass classifier using the "one-against-all" method (however, it is not clear which measure of confidence they used to determine the joint classification when a feature vector F was such that $\overline{\{i : \text{classifier for behaviour } B_i \text{ classes } F \text{ as } B_i\}} \neq 1$ — such a measure is necessary for the 'one-against-all" method. One such measure would be the distance from the decision boundary $w^{\mathrm{T}}x = b$.).

The remaining 30% of the feature vectors were used to assess the performance of the trained SVMs.

A "basic" SVM is a binary classifier which is trained on a linearly separable set of data, and consists of a hyperplane in the p-dimensional space in which the data lie. The data in one class lie on one side of the hyperplane and those in the other class lie on the other side. Previously unseen data points are then classified according to which side of the hyperplane they fall. The hyperplane itself is placed to maximise its minimum perpendicular distance to any training set data point.

The hyperplane is found by finding $w \in \mathbb{R}^p$, $b \in \mathbb{R}$ s.t. w minimises $\frac{1}{2}\|w\|^2$ subject to $y_i(w_i^{\mathrm{T}} - b) \geq 1$ for all N data points x_i, where $y_i \in \{-1, 1\}$ has a value dependent on the class to which x_i belongs. Then $w^{\mathrm{T}}x = b$ gives the hyperplane.

This constrained problem is equivalent to its dual: finding

$$\arg\min_{(w,b)} \max_{\alpha_1 \geq 0, \ldots, \alpha_N \geq 0} \left\{ \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i [y_i(w^{\mathrm{T}}x_i - b) - 1] \right\},$$

where the α_i are Lagrange multipliers.

It can be shown that $w = \sum_{i=1}^{N} y_i \alpha_i x_i$, and then the problem becomes that of finding

$$\arg \max_{\{k_1, \ldots, k_{N'}\} \subset \{1, \ldots, N\}, \alpha_{k_1} \geq 0, \ldots, \alpha_{k_{N'}} \geq 0} \left\{ \sum_{i=1}^{N'} \alpha_{k_i} - \frac{1}{2} \sum_{i=1}^{N'} \sum_{j=1}^{N'} \alpha_{k_i} \alpha_{k_j} y_{k_i} y_{k_j} x_{k_i}^{\mathrm{T}} x_{k_j} \right\},$$

where any solution $(\{k_1, \ldots, k_{\ell-1}, k_\ell, k_{\ell+1}, \ldots, k_{N'}\}, \alpha_{k_1}, \ldots, \alpha_{k_{\ell-1}}, \alpha_{k_\ell}, \alpha_{k_{\ell+1}}, \ldots, \alpha_{k_{N'}})$ such that $(\alpha_{k_1}, \ldots, \alpha_{k_{\ell-1}}, \alpha_{k_\ell}, \alpha_{k_{\ell+1}}, \ldots, \alpha_{k_{N'}}) = (\alpha_{k_1}, \ldots, \alpha_{k_{\ell-1}}, 0, \alpha_{k_{\ell+1}}, \ldots, \alpha_{k_{N'}})$ is disregarded (as it is equivalent to $(\{k_1, \ldots, k_{\ell-1}, k_{\ell+1}, \ldots, k_{N'}\}, \alpha_{k_1}, \ldots, \alpha_{k_{\ell-1}}, \alpha_{k_{\ell+1}}, \ldots, \alpha_{k_{N'}})$). Thus the problem reduces to a finite discrete maximisation.

Given a solution $(\{k_1, \ldots, \ldots, k_{N'}\}, \alpha_{k_1}, \ldots, \alpha_{k_{N'}})$, the points $x_{k_1}, \ldots, x_{k_{N'}}$ are called the *support points*, and $b = w^{\mathrm{T}}x_{k_i} - y_{k_i}$ for $i = 1, \ldots, N'$, or $b = \frac{1}{N'} \sum_{i=1}^{N'} [w^{\mathrm{T}}x_{k_i} - y_{k_i}]$. The two hyperplanes parallel to $w^{\mathrm{T}}x = b$ which pass through the support points form the boundary of the *margin*.

The "basic" SVM can be modified in two ways to cope with training sets which are not linearly separable.

The first is to use Cortes and Vapnik's soft margin, allowing some points to be misclassified and introducing a penalisation for this misclassification.

If a linear penalty is used, this is done by relaxing the constraints to $y_i(w^{\mathrm{T}}x_i - b) \geq 1 - \xi_i$,

$\xi_i \geq 0$ and changing the minimisation to that of $\frac{1}{2}\|w\|^2 + C\sum_i \xi_i$.

With a linear penalty, the problem reduces to

$$\arg \max_{\{k_1,\ldots,k_{N'}\}\subset\{1,\ldots,N\}, C\geq\alpha_{k_1}\geq 0,\ldots,C\geq\alpha_{k_{N'}}\geq 0} \left\{ \sum_{i=1}^{N'} \alpha_{k_i} - \frac{1}{2}\sum_{i=1}^{N'}\sum_{j=1}^{N'} \alpha_{k_i}\alpha_{k_j} y_{k_i} y_{k_j} x_{k_i}^{\mathrm{T}} x_{k_j} \right\},$$

so we again obtain a finite discrete maximisation.

The second modification to enable SVMs to deal with non-linearly separable training sets is to use the "kernel trick". This uses a map ϕ from the space in which the data lies to some higher dimensional space where the two classes *are* linearly separated (or, at least, closer to being so). The procedure for the "basic" SVM, or Cortes and Vapnik's modification, is then carried out in this higher dimensional subspace.[8]

The "kernel trick" consists of using $K(x_i, x_j) = \phi(x_i)^{\mathrm{T}}\phi(x_j)$ to carry out the SVM minimisations in the original space: because of Mercer's Theorem, which states that a symmetric function $K(x, y)$ can be expressed as an inner product if and only if K is positive semi-definite, i.e., $\int K(x, y)g(x)g(y)\mathrm{d}x\mathrm{d}y \geq 0\ \forall g$, the mapping ϕ to the higher space need not be known for symmetric positive semi-definite K.

(Information about SVM taken from Jordan[31], Wikipedia[32]).

Martiskainen *et al.* use SVMs with both soft margins and the kernel trick, with Gaussian kernel $K(x, y) = \exp(-\gamma\|x - y\|^2)$. They selected C and γ to maximise Cohen's $\kappa = \frac{P_a - P_c}{1 - P_c}$ over the training set, where P_a is the proportion of cases where two classifiers agree and P_c is the proportion where they would agree by chance, which measures the degree of agreement between between classifiers, in this case between ground truth and the joint decision of their classifiers.

The authors report good (over 80%) accuracy (ratio of the number of correct predictions to the total) for all the behaviours. However, in the case of rare behaviours (the acts of standing up or of sitting down), accuracy is dominated by predictions of the behaviour *not* occurring.

When we look at sensitivity (ratio of true positives to sum of true and false positives), and precision (ratio of true positives to sum of true positives and false negatives), the performance drops for the act of standing up (to 0.67 and 0.33 respectively — there appears to be a mistake in Table 4 of Martiskainen *et al.*) and for the act of sitting down (to 0 for both measures).

The sensitivity for the other behaviours ranged from 0.65-0.80, and the precision from 0.66-0.86, so there appears to be scope for improvement on these figures.

2.3.5 Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behaviour, *Moreau* et al.*[10]*

The authors of this paper attempt to classify goat behaviour using triaxial accelerometry readings. To this end, they used a HOBO® Pendant G triaxial accelerometer (58×33×23mm

[8] The kernel trick also makes the SVM a nonlinear classifier

weighing 18g including battery), attached to goats by i) a belt around the animal's chest; ii) a dog harness; iii) a collar around its neck. The weight of the attachments was in the range 242-264g.

In addition, a GPS tracker was attached to the goats' necks by a collar, and this assembly weighed 610g. The tracker played no role in the classification, but was intended to show where the goats' activities occurred.

The behaviours to be classified were: resting; eating; and walking. Eating was to be separated into grazing (with the head down) and browsing (with the head up).

The experiments took place in Germany (two goats), in mostly flat pasture, and in Oman (one goat), on mountainous terrain. Browsing activity was absent in the German experiments.

Ground truth was supplied by video recordings, and the accelerometer data was sampled at 1Hz (Germany) or 0.5Hz (Oman).

It is somewhat difficult to determine whether the raw data discussed in this paper is dynamic acceleration or total acceleration (it is labelled dynamic, but the derivation of the head position makes no sense if it is not total acceleration).

The method for distinguishing the top level activities was to transform the readings a_{vj} from each accelerometer channel ($v \in \{x, y, z\}$, j the index into the time-series) into $d_{vj} = |a_{vj} - a_{v,j-1}|$ and define the moving average $\mu_{vj} = \frac{d_{v,j-m_v}+d_{v,j-m_v+1}+\ldots+d_{vj}+\ldots+d_{w,j+m_v}}{2m_v+1}$. The activity at the time indexed by j is classed (according to channel v) as resting if $\mu_{vj} < \theta_{rev}$, eating if $\theta_{rev} \leq \mu_{vj} < \theta_{ewv}$ and walking if $\theta_{rwv} \leq \mu_{vj}$, for thresholds θ_{rev} and θ_{ewv} with $0 < \theta_{rev} < \theta_{ewv}$. The quality of the method for arbitrary m_v, θ_{rev} and θ_{ewv} is assessed by $S_v = \sum_j S_{vj}$ where

$$S_{vj} = \begin{cases} 1 - \frac{N_r}{N}, & \text{the } j\text{th time is correctly classified as resting;} \\ 1 - \frac{N_e}{N}, & \text{the } j\text{th time is correctly classified as eating;} \\ 1 - \frac{N_w}{N}, & \text{the } j\text{th time is correctly classified as walking;} \\ 0, & \text{the } j\text{th time is incorrectly classified,} \end{cases}$$

where N is the total number of observations, and $N_{r,e,w}$ are the number of observations of resting, eating and walking, respectively. This weighting gives each behaviour equal importance, compensating for their different degrees of representation in the data. Then, for each v, m_v, θ_{rev} and θ_{ewv} are chosen to maximise S_v.

The overall classification is arrived at by taking the majority classification (over the classifications according to x, y and z). If there is no majority, the classification with the largest value of S_v is taken as the overall classification.

Clearly, this method relies on there being more, or larger, or both more and larger, changes of acceleration (on the time scale of the sampling) as the animal's activity moves from resting to eating to walking.

The discrimination between grazing and browsing, for those times where the higher-level classification is "eating", depends on the position of the animal's head, and the accelerometer is attached to the animal in such a way that the x-component of the acceleration is most

affected by the head position. The method for automatically distinguishing these behaviours consequently uses just this channel.

A moving average of the raw values, $\nu_j = \frac{a_{x,j-m}+a_{x,j-m+1}+\ldots+a_{xj}+\ldots+a_{x,j+m}}{2m+1}$ is thresholded: if $\nu_j < \theta_{gb}$ the classification is "grazing"; if $\theta_{gb} \leq \nu_j$, "browsing". For arbitrary m, θ_{gb}, classification success is measured by $S = \sum_{\{j:\text{the } j\text{th moment is classified as eating}\}} S_j$, where

$$S_j = \begin{cases} 1 - \frac{N_g}{N_e}, & \text{the } j\text{th time is correctly classified as grazing;} \\ 1 - \frac{N_b}{N_e}, & \text{the } j\text{th time is correctly classified as browsing;} \\ 0, & \text{the } j\text{th time is incorrectly classified,} \end{cases}$$

where $N_{g,b}$ are the number of instances of browsing and grazing, respectively, in the ground truth.

m and θ_{gb} are then chosen to maximise S over the training set (generated solely from the Oman data, due to the absence of browsing in the German data).

The authors examined the effect of various factors:

Fixation: the true recognition rate $\left(\frac{\text{number of correctly classified examples}}{\text{number of examples of a behaviour}}\right)$ for walking was best ($\sim 55-58\%$) for the belt fixation; for eating it was best ($\sim 98-9\%$) for the collar; and for resting there was an overlap between the performances with the collar and with the harness, with the belt being not far behind ($\sim 85-98\%$ over all three fixations);

Generalisation across individuals: when data from one German goat was used to train and data from the same goat was used to test, the overall true recognition rate was 83%; when the training set and test set came from different German goats, the overall true recognition rate increased to 85%, which seems to be an indication of good generalisation. However, the increase was entirely due to an increased true recognition rate for resting, which masked a substantial fall in the same rate for walking, which suggests goats at rest are similar, but they have different walking styles;

Effect of sampling interval: the authors examined this by deleting samples from their datasets and found that the overall true recognition rate fell from 85% to 75% as the sampling frequency decreased from 1Hz to 0.05Hz. Also, the underestimation of the number of observations of resting increased, the overestimation of the number of walking observations became an underestimation and the overestimation of the number of eating observations increased; that is, the behaviour which might be expected to be intermediate in terms of the summed magnitude of changes in acceleration was predicted more frequently as the sampling became coarser and less information was available;

Generalisation across fixations: unsurprisingly, training on data from one method of fixation and testing on data from a different one resulted in a significant degradation in performance when compared to training and testing on the same method;

Generalisation across husbandry systems: when the training was done using the combined data from both German goats and the testing done with data from the Oman goat, the overall true recognition rate was 83%, the incidences of resting and walking behaviours were greatly underestimated and eating was greatly overestimated. When the training and test sets were exchanged, the overall true recognition rate was 87% and the only large over- or underestimation was that of the incidence of walking behaviour, which was greatly underestimated.

For grazing and browsing, the authors investigated three different splits into training and test sets of the data stemming from the Oman goat when it was fitted with the dog harness. The true recognition rate for grazing and browsing

$$\left(\text{i.e. } \frac{\text{number of correctly classified examples of grazing or browsing}}{\text{number of examples of grazing or browsing}}\right)$$

was 70-75%, whereas the estimate of the number of moments of grazing was 79-88%. The corresponding figure was 95-111% for browsing.

The authors discuss several sources of error. Firstly, goats change their behaviour quite abruptly, so the 10s intervals over which the animals' behaviour is manually classified to provide ground truth may miss some episodes of behaviour. Secondly, the use of moving averages blurs the boundary between behaviours. In particular, if a goat changes from resting to walking, the method may erroneously place a period of eating between the two behaviours actually present. Nevertheless, from practical considerations, the authors rule out a more-finely granulated ground truth and they also conclude that a reliable enough classification can be based on sampling at 0.2Hz, subject to further investigation.

A third source is the optimisation of the quantities S_w and S, which is an overall measure of the classification success, rather than of classification for each behaviour separately.

The authors finally discuss the use of these methods in combination with GPS to determine the exploitation of the terrain by the animals, and conclude that this is feasible.

2.3.6 A new technique for monitoring the behaviour of free-ranging Adélie penguins, *Yoda* et al.*[6]*

The authors here used biaxial accelerometry in conjunction with speed and depth measurements to compare the foraging behaviour of breeding Adélie penguins in regions with fast sea-ice and those without sea-ice. As an intermediate step in doing this, it was desired to identify the behaviours walking, tobogganing (the penguin lying flat on its belly and pushing itself forward by alternating thrusts by one foot and the other), standing on land, lying on land, resting at the water surface, porpoising (jumping out of the water) and diving (swimming more than one 1m below the surface).

The authors attached a device containing a logger, an biaxial accelerometer. a speedometer and a means for measuring depth to nesting penguins and sampled speed and depth measurements at 1Hz,and the acceleration channels at either 3.3Hz or 16Hz. The device

had a diameter of 20mm, a length of 122mm and a mass of 60g (cf Adélie penguin mass of ~ 4.5kg).. The device was aligned with the animals' body axes so as to either measure surge and heave, or surge and sway.

The static acceleration in the surge direction was estimated by taking a moving mean over 111 points according to the paper (although this seems to ignore the effect of the different sampling rates), and this was used to determine the posture of the bird (clearly, if the surge accelerometer channel registers 1g when the surge axis is stationary and vertical, it will register $\sin\theta \times$g when the surge axis is at a stationary angle of θ to the horizontal — i.e. $\pi - \theta$ to the vertical).

Calibration experiments on captive birds and observation of birds travelling between nesting sites and the sea enabled the acceleration/depth profiles to be established for various behaviours.

The static surge acceleration enabled the authors to distinguish between the two sets of behaviour where this is roughly constant: when the value is around 1g, the bird is standing or walking, and when it is around 0g, the bird is lying on land on or resting at the sea surface, or is tobogganning. These sets were also distinguishable from the set containing diving and porpoising, in which the static surge acceleration is not constant. This division into sets was automated.

Standing on land and walking were distinguished by the strong frequency component corresponding to stride frequency in the surge and sway acceleration trace for the latter, lying on land and resting at the sea surface were distinguished by the effect of sea waves on the traces for the latter, and tobogganing was distinguished by the presence of a frequency component corresponding to the alternate pushing by the feet. Diving was detected by the depth information, and porpoising has a distinctive acceleration profile of its own. These classifications were done by eye.

The authors found that Adélie penguins have a significantly (on the Mann-Whitney U-test, $P < 0.05$) different activity profile during foraging trips when sea ice is present to that when it is absent: although the time spent diving is not significantly different, the penguins in the sea-ice free area preferred to rest on water, those in the area with sea-ice rest (i.e., stand or lie) on land, and porpoising is an activity restricted to the ice-free area.

2.3.7 Regulation of stroke and glide in an foot-propelled avian glider, *Watanuki* et al.*[8]*

To understand how European shags (*Phalacocorax aristoteles*) regulate the thrusting and gliding phases of their dives, Watanuki *et al.* attached M190-D2GT data loggers (containing biaxial accelerometers and a barometer for depth measurement, with length 60mm, diameter 16mm and weight 16g) to the backs of the birds.

The M190-D2GT loggers sampled depth information at 1Hz, and accelerometry data at 64Hz. To ensure that the alignment of the logger to measure surge and heave accelerations was the appropriate one, four birds had the devices attached with the alternative orientation

measuring the surge and sway components, and it was found that the sway component had negligible variation during the ascent and descent portions of the dives.

Hence, the remaining ten shags had the logger attached in the surge/heave position.

The authors determined the static acceleration by using a low-pass two-band filter with a pass frequency (below which frequency components were passed) and a stop frequency (above which frequency components were stopped), with a transition between pass and stop between. They considered the stop-start pairs (0.1Hz, 0.5Hz), (1Hz, 1.5Hz), (2Hz, 3Hz), finding that the (0.1Hz, 0.5Hz) and the (2Hz,3Hz) pair risked confusing thrust and the changes in body orientation at the start of a dive (when body orientation changes rapidly), and in the middle of the dive (when body orientation changes slowly) respectively. Thus, they used a low-pass filter with pass frequency 1Hz and stop frequency 1.5Hz to obtain the static acceleration, and got the dynamic acceleration by subtracting this from the total acceleration.

The authors also made corrections to compensate for logger fixings which were at angles to the animals' body axes.

During the course of 100 dives spread over 9 of the birds, the rate of change of depth fell into three ranges: $\leq -0.6\text{ms}^{-1}$, $[-0.3, 0.3]\text{ms}^{-1}$ and $\geq 1\text{ms}^{-1}$, interpreted as relating to birds in the ascendent, bottom and descendent parts of the dive respectively.

The authors found that there was a significant (on the Mann-Whitney U-test, P¡0.01) difference between the weights of the male (mean: ~ 1.9kg) and female birds (mean ~ 1.6kg), but this was not reflected in significant sex differences in the median dive depth (10.2-43.3m, medians for different individuals) or dive duration (43-97s), nor in the maximum values of these quantities.

They also found an empirical means of detecting surge thrusts — these were said to occur if there was more than 0.16ms^{-2} change in values measured at $\frac{1}{64}$s intervals. During the ascendent and descendent phases of the dive, there is a thrust cycle, consisting of a series of thrusts, each followed by a glide.

Using their ability to identify body position, thrusts and thrust cycles, the authors are able to conclude that shags dive almost vertically (to minimise the time taken to get to the ocean floor and the food in its vicinity), and that in the descendent phase they decrease the average thrust with depth (as their buoyancy decreases with depth) by increasing the duration of the glide portion of the thrust cycle and keeping the thrust portion roughly constant in duration and magnitude (so that they continue to thrust with maximum efficiency).

2.3.8 Swimming speeds and buoyancy compensation of migrating adult chum salmon *Oncorhyncus keta* revealed by speed/depth/acceleration data logger, *Tanaka, Takagi and Naito[7]*

The authors here investigate the dives of salmon migrating towards their natal rivers for breeding. As these salmon do not feed in this phase of their life, the dives are for purposes other than feeding. These are believed to be thermoregulation and/or the acquisition of navigational clues.

The salmon were fitted with loggers for this investigation. In 1997, they were fitted with UWE-200PDT loggers, which record speed, depth and water temperature, and, in 1999, with W-190L-PD2GT loggers which record speed, depth, biaxial acceleration and temperature. It is the second experiment which is of most interest here.

The W-190L-PD2GT has a diameter of 21mm and a length of 117mm, and weighs 60g (the salmon used in the experiment weigh ∼2.5-∼4kg and are of length 58-70.5cm). It was attached to the fish in an orientation which enabled it to measure surge and sway acceleration. The sampling rates were 16Hz for the acceleration and 1Hz for the other variables.

To determine the static acceleration, the authors applied a low-pass filter with a cut-off of 0.8Hz — this cut-off was sufficiently low to enable the tailbeats to be assigned to the dynamic acceleration. Then the static acceleration could be used to determine the body-angle of the animal, as in other papers quoted here.

Tailbeats were recognised in the sway acceleration trace as having peaks and troughs of magnitude greater than 1ms^{-2} and an inter-trough separation less than 1s.

Using these tools, the authors were able to determine that the salmon do not appear to use their swim bladders to any great extent to maintain neutral or negative buoyancy during their dives, as they need to actively swim in the descent and ascent. The main function of the swim bladder appears to be to maintain neutral buoyancy in surface waters.

2.4 PD and CWT

It appears to be quite difficult to find references to PD and CWT (if we exclude cases where CWT means "Curved Walking Therapy"), presumably because DWT is considerably more efficient (especially if the CWT mother wavelet does not have a special form, like piecewise polynomial).

However, we have found three, plus one which could easily be extended to PD.

Muthuraman *et al.*[33] use the CWT with a scaled and shifted versions of a single mother wavelet to analyse PD accelerometry and EMG, alongside multitaper (i.e., with different filtering functions on the windows into the data) spectral analysis to determine whether the appearance of "first harmonics" (i.e., at twice the characteristic PD tremor frequency) with elevated amplitudes is correlated with the amplitude at the fundamental tremor frequency. Their answer is that it is not, and the first fundamental therefore appears to separately generated.

Sello *et al.*[34] examine the CWT of phonograms stemming from the respiratory sounds of various subjects, transforming the phonogram into a scaleogram. They believe that they have found clinically useful interpretations of this scaleogram, although they, too suffer from a lack of sufficient subjects! Although none of their subjects has PD, this research could easily be extended to encompass an investigation into the effects of PD on breathing.

Strambi *et al.*[35] use the correlation between two scaleograms, derived from the CWT of EMG data to examine the coherence of motion, while performing specific exercises, of PD subjects at different points in their medication cycle, and of control subjects. They note that

the coherence of the PD subjects' motion is reduced in comparison to controls, but to a lesser extent when the PD subjects are on medication. All of this is interpreted in terms of changes in the physiology of the PD subjects' muscles.

Revanian and Lockhart[36] investigate the use of CWTs to detect episodes of freezing of gait. By creating a scaleogram and summing across two ranges of scales, they obtain a pair of functions of time, one of which is believed to be high relative to the other during episodes of freezing of gait. By thresholding the ratio of these functions, each time in the trace can be classified as an episode, or not an episode, of freezing of gait. Ground truth was supplied by physiotherapists observing the subjects during the authors' experiments, and fairly good results were obtained.

These examples all use either a fixed mother wavelet, or one with just one parameter, which is not adjusted to fit any training data.

2.5 Summary

We have learnt from our review that there are a wealth of techniques that can be applied to the analysis human motion, including some that have already been used in this way, and some that have been developed in animal studies.

We have also learnt that there is a gap in the exploration of the range of possible techniques — using CWT-type wavelets to develop libraries of motion shapes which are independent of rotation, reflection and time dilation. We decided to investigate this gap as it relates to PD. Although there is the danger that bradykinesis will not be dealt with by these techniques, it can be dealt with by appending scales to the shape library members. We can consider this as a factorisation of the problems of analysing PD patterns of motion: one factor is the timescale, the other, the shape information. We investigate the latter here.

2.6 Hypothesis

We end this section with our hypothesis.

We hypothesise that:

1 it is possible, using the techniques to be developed, to distinguish between data stemming from PD patients and controls on the basis of patterns of motion;

2 these patterns of motion can be captured sufficiently well by the NAT and CWT analysis of the resultant data;

3 these patterns of motion are diverse enough to be easily distinguished from each other and frequently occur in the data.

Item 1 means that the methods here can be used to help diagnose PD, and item 3 opens up the possibility of using a changing balance of different patterns to monitor the development

of the disease and the effects of various treatments. We do not explore these possibilities here, as this would require a mass of data not currently available to us.

Although the limitations of the available data also make it impossible to *prove* this hypothesis, we will at least attempt to show that the hypothesis is feasible, by distinguishing between data stemming from *a* PD patient and controls.

Chapter 3

The Wavelets and Libraries

The purpose of this Chapter is to describe our method and provide the mathematical machinery behind it. More or less in parallel, we develop examples showing how this machinery works in practice.

In a simplified version of this method, we fit a wavelet[1] to data of one type (e.g., PD or control), recording the goodness of fit. We then identify an equivalence class[2] of wavelets to which this wavelet belongs. This equivalence class is then inserted into a library of such equivalence classes. After a large number of operations like this, we obtain a large library of equivalence classes, which we reduce

1 by putting a lower threshold on the goodness of fit to the data and removing those equivalence classes with a poor fit;

2 by removing members which are "too" close to other members[3].

After this trimming, we order the members of the library by their distance from the equivalence classes of wavelets fitted to a second class of data, with the most distant being preferred.

The process so far can be considered training.

In a classification context, the distance between the top-ranked members of the library and unseen data can be used as the basis for classifying it as the first or second type above.[4]

In a characterisation context, once we have a set of patterns of motion, represented by members of our library of equivalence classes, which are typical of the movement of the source of the first set of data, but atypical of the source of the second set, we can track the evolution of the strength with which the patterns are represented in future data collected from the first source.

Note that "source" here can mean one subject, but, in most contexts, it would be better if it meant "class of subjects".

[1] Actually, a wavelet triplet

[2] As defined below

[3] In terms of a distance also defined below

[4] The idea of a library of data and using a distance from its members and new data for classification, is, of course not new — see, e.g. the Profi trademark recognition project, [37]

3.1 Implementation

Our original implementation was in largely unvectorised Matlab®, but we vectorised[5] our code (making great use of the Matlab®built-in function `filter`), and, as a result the code for building a library of wavelet triplets now runs in hours rather than in days, and that for finding the activation of a three-dimensional acceleration time series with 2000 samples with respect to a given wavelet triplet with a fixed window length is about 6s, or roughly real time plus 50% with a sampling rate of 500Hz.

3.2 The CWT and accelerometer data

We wish to fit to our data a wavelet defined by

$$\psi(x) = \sum_{k=0}^{n} a_k f_k(x), \tag{3.1}$$

where

$$f_k(x) = \begin{cases} x^k, & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

For such a wavelet, conditions (1.1) and (1.2) become

$$a^{\mathrm{T}} H a = 1, \tag{3.3}$$

$$b^{\mathrm{T}} a = 0, \tag{3.4}$$

where $a \in \mathbb{R}^{n+1}$ is a vector whose kth component is a_{k-1}, and

$$b = \int_{-\frac{1}{2}}^{\frac{1}{2}} \begin{bmatrix} f_0(x) \\ f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix} \mathrm{d}x = \begin{bmatrix} 1 \\ 0 \\ \frac{1}{12} \\ \vdots \\ \frac{1+(-1)^n}{2^{n+1}(n+1)} \end{bmatrix} \in \mathrm{R}^{n+1} \text{ and} \tag{3.5}$$

$$\begin{aligned} H &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \begin{bmatrix} f_0(x)^2 & f_0(x)f_1(x) & \cdots & f_0(x)f_n(x) \\ f_0(x)f_1(x) & f_1(x)^2 & \cdots & f_1(x)f_n(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(x)f_n(x) & f_1(x)f_n(x) & \cdots & f_n(x)^2 \end{bmatrix} \mathrm{d}x \\ &= \begin{bmatrix} 1 & 0 & \frac{1}{12} & \cdots & \frac{1+(-1)^n}{2^{n+1}(n+1)} \\ 0 & \frac{1}{12} & 0 & \cdots & \frac{1-(-1)^n}{2^{n+2}(n+2)} \\ \frac{1}{12} & 0 & \frac{1}{80} & \cdots & \frac{1+(-1)^n}{2^{n+3}(n+3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1+(-1)^n}{2^{n+1}(n+1)} & \frac{1-(-1)^n}{2^{n+2}(n+2)} & \frac{1+(-1)^n}{2^{n+3}(n+3)} & \cdots & \frac{1}{2^{2n}(2n+1)} \end{bmatrix} \in \mathrm{R}^{(n+1)\times(n+1)}. \end{aligned} \tag{3.6}$$

As an aside (at this point we are not very interested in the reconstruction property of admissible wavelets), because $\int x^k \mathrm{e}^x \mathrm{d}x = (-1)^k k! \mathrm{e}^x \sum_{r=0}^{k} \frac{(-1)^r x^r}{r!} =$

[5]However, Mathworks have now built their JIT compiler into Matlab®, which they claim makes vectorisation rather than using loop constructions merely a matter of style, and not efficiency

$(-1)^k k!\left[1-\mathrm{e}^x\sum_{r=k+1}^{\infty}\frac{(-1)^r x^r}{r!}\right]$ (which can be proven by integration by parts and induction), we have

$$\begin{aligned}
\hat{f}_k(\xi) &= \frac{1}{\sqrt{2\pi}}\int_{-\frac{1}{2}}^{\frac{1}{2}} x^k \mathrm{e}^{-ix\xi}\mathrm{d}x \\
&= \frac{k!(-i)^{k+1}}{\sqrt{2\pi}\xi^{k+1}}\left[-\mathrm{e}^{\frac{1}{2}i\xi}\sum_{r=k+1}^{\infty}\frac{(-i)^r\xi^r}{2^r r!}+\mathrm{e}^{-\frac{1}{2}i\xi}\sum_{r=k+1}^{\infty}\frac{i^r\xi^r}{2^r r!}\right] \\
&= \frac{k!(-i)^{k+1}}{\sqrt{2\pi}\xi^{k+1}}\left[(\cos\tfrac{\xi}{2}+i\sin\tfrac{\xi}{2})\left\{\sum_{r=\lceil\frac{1}{2}(k+1)\rceil}^{\infty}\frac{(-1)^r\xi^{2r}}{2^{2r}(2r)!}-i\sum_{r=\lceil\frac{1}{2}k\rceil}^{\infty}\frac{(-1)^r\xi^{2r+1}}{2^{2r+1}(2r+1)!}\right\}-\right. \\
&\qquad\left.(\cos\tfrac{\xi}{2}-i\sin\tfrac{\xi}{2})\left\{\sum_{r=\lceil\frac{1}{2}(k+1)\rceil}^{\infty}\frac{(-1)^r\xi^{2r}}{2^{2r}(2r)!}+i\sum_{r=\lceil\frac{1}{2}k\rceil}^{\infty}\frac{(-1)^r\xi^{2r+1}}{2^{2r+1}(2r+1)!}\right\}\right] \\
&= \frac{2k!(-i)^{k+1}}{\sqrt{2\pi}\xi^{k+1}}\left[\cos\tfrac{\xi}{2}\sum_{r=\lceil\frac{1}{2}(k+1)\rceil}^{\infty}\frac{(-1)^r\xi^{2r}}{2^{2r}(2r)!}+\sin\tfrac{\xi}{2}\sum_{r=\lceil\frac{1}{2}k\rceil}^{\infty}\frac{(-1)^r\xi^{2r+1}}{2^{2r+1}(2r+1)!}\right] \\
&= \begin{cases}\frac{2k!(-i)^{k+1}}{2^{k+1}\sqrt{2\pi}}\left[\cos\frac{\xi}{2}\sum_{r=0}^{\infty}\frac{(-1)^{r+\frac{1}{2}k+1}\xi^{2r+1}}{2^{2r+1}(2r+k+2)!}+\sin\frac{\xi}{2}\sum_{r=0}^{\infty}\frac{(-1)^{r+\frac{1}{2}k}\xi^{2r}}{2^{2r}(2r+k+1)!}\right], & k \text{ even;}\\ \frac{2k!(-i)^{k+1}}{2^{k+1}\sqrt{2\pi}}\left[\cos\frac{\xi}{2}\sum_{r=0}^{\infty}\frac{(-1)^{r+\frac{1}{2}(k+1)}\xi^{2r}}{2^{2r}(2r+k+1)!}+\sin\frac{\xi}{2}\sum_{r=0}^{\infty}\frac{(-1)^{r+\frac{1}{2}(k+1)}\xi^{2r+1}}{2^{2r+1}(2r+k+2)!}\right], & k \text{ odd}\end{cases} \qquad (3.7)\\
&= \begin{cases}-i\left[\frac{1}{2^{k+1}(k+2)\sqrt{2\pi}}\xi+\mathrm{O}(\xi^3)\right], & k \text{ even;}\\ \frac{1}{2^k(k+1)\sqrt{2\pi}}+\mathrm{O}(\xi^2), & k \text{ odd.}\end{cases} \qquad (3.8)
\end{aligned}$$

Then, near $\xi=0$, $\hat{\psi}(\xi)=\sum_{k=0}^{n}a_k\hat{f}_k(\xi)=\sum_{k=0}^{\lfloor\frac{1}{2}n\rfloor}a_{2k}\hat{f}_{2k}(\xi)+\sum_{k=0}^{\lfloor\frac{1}{2}(n-1)\rfloor}a_{2k+1}\hat{f}_{2k+1}(\xi)=\frac{1}{\sqrt{2\pi}}\left\{-i\left[\sum_{k=0}^{\lfloor\frac{1}{2}n\rfloor}a_{2k}\frac{1}{2^{k+1}(k+2)}\xi+\mathrm{O}(\xi^3)\right]+\sum_{k=0}^{\lfloor\frac{1}{2}(n-1)\rfloor}a_{2k+1}\frac{1}{2^k(k+1)}+\mathrm{O}(\xi^2)\right\}$, so

$$\begin{aligned}
|\hat{\psi}(\xi)|^2 &= \frac{1}{2\pi}\left\{\left[\sum_{k=0}^{\lfloor\frac{1}{2}n\rfloor}a_{2k}\frac{1}{2^{k+1}(k+2)}\xi+\mathrm{O}(\xi^3)\right]^2+\left[\sum_{k=0}^{\lfloor\frac{1}{2}(n-1)\rfloor}a_{2k+1}\frac{1}{2^k(k+1)}+\mathrm{O}(\xi^2)\right]^2\right\} \\
&= \frac{1}{2\pi}\left[\sum_{k=0}^{\lfloor\frac{1}{2}(n-1)\rfloor}a_{2k+1}\frac{1}{2^k(k+1)}\right]^2+\mathrm{O}(\xi^2)=\frac{(b^{\mathrm{T}}a)^2}{2\pi}+\mathrm{O}(\xi^2)=\mathrm{O}(\xi^2). \qquad (3.9)
\end{aligned}$$

by the zi condition as given by equation (3.4).

If ψ is to be admissible, we need $\int_{-\infty}^{\infty}\frac{|\hat{\psi}(\xi)|^2}{|\xi|}\mathrm{d}\xi<\infty$. But $\int_{-\infty}^{\infty}\frac{|\hat{\psi}(\xi)|^2}{|\xi|}\mathrm{d}\xi=\left(\int_{-\infty}^{-1}+\int_{-1}^{1}+\int_{1}^{\infty}\right)\frac{|\hat{\psi}(\xi)|^2}{|\xi|}\mathrm{d}\xi\le\left(\int_{-\infty}^{-1}+\int_{1}^{\infty}\right)|\hat{\psi}(\xi)|^2\mathrm{d}\xi+\int_{-1}^{1}\frac{|\hat{\psi}(\xi)|^2}{|\xi|}\mathrm{d}\xi\le\int_{-\infty}^{\infty}|\hat{\psi}(\xi)|^2\mathrm{d}\xi+\int_{-1}^{1}\frac{|\hat{\psi}(\xi)|^2}{|\xi|}\mathrm{d}\xi=1+\int_{-1}^{1}\frac{|\hat{\psi}(\xi)|^2}{|\xi|}\mathrm{d}\xi<\infty$ by equation (3.9) and the ue condition given by equation (1.1), so our ψs are admissible.

3.2.1 Additional conditions

We can impose additional conditions on $\psi(x)$, such as

wrapped continuity (wc): ψ is continuous if the endpoints of its support are identified: $\psi\left(-\frac{1}{2}\right) = \lim_{x\to\frac{1}{2}} \psi(x)$;

endpoint continuity (ec): ψ is continuous: $\psi\left(-\frac{1}{2}\right) = \lim_{x\to\frac{1}{2}} \psi(x) = 0$;

wrapped differentiability (wd): ψ is differentiable if the endpoints of its support are identified: $\psi\left(-\frac{1}{2}\right) = \lim_{x\to\frac{1}{2}} \psi(x)$ and $\frac{\mathrm{d}}{\mathrm{d}x}\,\psi(x)|_{x=-\frac{1}{2}} = \lim_{x\to\frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}x}\psi(x)$;

endpoint continuity and wrapped differentiability (ec+wd): ψ is continuous, and differentiable if the endpoints of its support are identified: $\psi\left(-\frac{1}{2}\right) = \lim_{x\to\frac{1}{2}} \psi(x) = 0$ and $\frac{\mathrm{d}}{\mathrm{d}x}\,\psi(x)|_{x=-\frac{1}{2}} = \lim_{x\to\frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}x}\psi(x)$;

endpoint differentiability (ed): ψ is differentiable: $\psi\left(-\frac{1}{2}\right) = \lim_{x\to\frac{1}{2}} \psi(x) = 0$ and $\frac{\mathrm{d}}{\mathrm{d}x}\,\psi(x)|_{x=-\frac{1}{2}} = \lim_{x\to\frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}x}\psi(x) = 0$.

All of these conditions can be written as one or more equations of the form $c^{\mathrm{T}}a = 0$:

wc: $c_{c-}^{\mathrm{T}}a = 0$;

ec: $c_{c-}^{\mathrm{T}}a = c_{c+}^{\mathrm{T}}a = 0$;

wd: $c_{c-}^{\mathrm{T}}a = c_{d-}^{\mathrm{T}}a = 0$;

ec+wd: $c_{c-}^{\mathrm{T}}a = c_{c+}^{\mathrm{T}}a = c_{d-}^{\mathrm{T}}a = 0$;

ed: $c_{c-}^{\mathrm{T}}a = c_{c+}^{\mathrm{T}}a = c_{d-}^{\mathrm{T}}a = c_{d+}^{\mathrm{T}}a = 0$,

where

$$c_{c\pm} = \begin{bmatrix} f_0\left(-\frac{1}{2}\right) \\ f_1\left(-\frac{1}{2}\right) \\ \vdots \\ f_n\left(-\frac{1}{2}\right) \end{bmatrix} \pm \lim_{x\to\frac{1}{2}-} \begin{bmatrix} f_0(x) \\ f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix} = \begin{bmatrix} 1\pm 1 \\ \frac{-1\pm 1}{2} \\ \vdots \\ \frac{(-1)^n \pm 1}{2^n} \end{bmatrix}, \tag{3.10}$$

$$c_{d\pm} = \lim_{x\to-\frac{1}{2}+} \frac{\mathrm{d}}{\mathrm{d}x} \begin{bmatrix} f_0(x) \\ f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix}\Bigg|_{x=-\frac{1}{2}} \pm \lim_{x\to\frac{1}{2}-} \frac{\mathrm{d}}{\mathrm{d}x} \begin{bmatrix} f_0(x) \\ f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix} = \begin{bmatrix} 0 \\ 1\pm 1 \\ \vdots \\ \frac{(-1)^{n-1}\pm 1}{2^{n-1}} n \end{bmatrix}. \tag{3.11}$$

We will collect cs of the conditions of the form $c^{\mathrm{T}}a = 0$ we are applying in a matrix B, so they can be expressed in the single condition $B^{\mathrm{T}}a = 0$. For example, if we are applying wd to ψ, $B = [\, b\ c_{c-}\ c_{d-}\,]$.

We note that, for the condition sets we consider, if $B \in \mathbb{R}^{(n+1)\times p'}$ and $n \geq p'$, then the columns of B are linearly independent (shown by demonstrating that B has a non-singular $p' \times p'$ submatrix in Maple®).

3.2.2 Model function

We wish to represent the data $(x_k, y_k), k = 1, 2, \ldots, N$, shifted so that $x_1 = -\frac{1}{2}$ and $x_N = \frac{1}{2}$. Assuming that the x_k are evenly spaced, $x_k = (k-1)\Delta - \frac{1}{2}$ with $\Delta = \frac{1}{N-1}$, we define $x'_0 = -\frac{1}{2}$, $x'_k = \left(k - \frac{1}{2}\right)\Delta - \frac{1}{2}, k = 1, 2, \ldots, N-1$, $t'_N = \frac{1}{2}$ and

$$f(x) = \begin{cases} y_k, & x \in [x'_{k-1}, x'_k], \quad k = 1, 2, \ldots, N \\ 0 & \text{otherwise.} \end{cases} \tag{3.12}$$

Obviously, this piecewise constant function is L_2, and we use it to model our data.

3.2.3 Fitting a wavelet to data

We attempt to maximise the inner product $\langle f, \psi \rangle$ (which is real as f and ψ are real) with respect to a, subject to conditions (3.3) and (3.4), and possibly $p' - 1$ other conditions of the form $c^{\mathrm{T}}a = 0$, collected together with (3.4) in $B^{\mathrm{T}}a = 0$.

As

$$\begin{aligned} J(a) &= \langle f, \psi \rangle \\ &= \int_{-\infty}^{\infty} f(t)\overline{\psi(t)}\mathrm{d}t = \sum_{k=0}^{n} a_k \int_{-\frac{1}{2}}^{\frac{1}{2}} f_k(t) f(t)\mathrm{d}t = \sum_{k=0}^{n}\sum_{\ell=1}^{N} a_k y_\ell \int_{x'_{\ell-1}}^{x'_\ell} f_k(t)\mathrm{d}t \\ &= z^{\mathrm{T}}a, \end{aligned} \tag{3.13}$$

where

$$z = Ry \in \mathbb{R}^{n+1} \tag{3.14}$$

for $R \in \mathbb{R}^{(n+1)\times N}$ with elements

$$R_{k\ell} = \int_{x'_{\ell-1}}^{x'_\ell} f_k(t)\mathrm{d}t = \frac{(x'_\ell)^{k+1} - (x'_{\ell-1})^{k+1}}{k+1} \tag{3.15}$$

for $k = 0, 1, \ldots, n$, $\ell = 1, 2, \ldots, N$, $y = [y_1, y_2, \ldots, y_N]^{\mathrm{T}} \in \mathbb{R}^N$, this conditional maximisation is equivalent to finding the turning points of

$$\begin{aligned} Q(a, \lambda, \mu) &= J(a) + \lambda(a^{\mathrm{T}}Ha - 1) + \mu^{\mathrm{T}}B^{\mathrm{T}}a \\ &= \lambda a^{\mathrm{T}}Ha + (B\mu + z)^{\mathrm{T}}a - \lambda \end{aligned} \tag{3.16}$$

with respect to a and the Lagrange multipliers[6] $\lambda \in \mathbb{R}$ and $\mu \in \mathbb{R}^{p'}$, s.t. $J(a)$ has the largest value. As $J(a)$ is identically zero if $z = Bd$ for some nonzero vector d, we assume that this is not the case.

Solving $\frac{\partial}{\partial a^{\mathrm{T}}}Q(a, \lambda, \mu) = 2\lambda Ha + z + B\mu = 0$ yields

$$a = a^*(\lambda, \mu) = -\frac{1}{2\lambda}H^{-1}(z + B\mu),$$

[6]For the method of Langrange multipliers see Goldstein[38]

so (on the assumption that $\lambda \neq 0$; but, if $\lambda = 0$, $z = -B\mu$ for some μ, contrary to our assumption above)

$$Q(a^*(\lambda,\mu),\lambda,\mu) = -\frac{1}{4\lambda}(z+B\mu)^{\mathrm{T}}H^{-1}(z+B\mu) - \lambda,$$

and then $\frac{\partial}{\partial\mu}Q(a^*(\lambda,\mu^{\mathrm{T}}),\lambda,\mu) = -\frac{1}{2\lambda}B^{\mathrm{T}}H^{-1}(z+B\mu)$ so $\frac{\partial}{\partial\mu}Q(a^*(\lambda,\mu^{\mathrm{T}}),\lambda,\mu) = 0$, implies[7]

$$\mu = \mu^* = -(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}z,$$

and

$$\begin{aligned} Q(a^*(\lambda,\mu^*),\lambda,\mu^*) &= -\lambda - \frac{1}{4\lambda}z^{\mathrm{T}}\left[\mathrm{I} - B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}\right]^{\mathrm{T}}H^{-1}\times \\ &\qquad \left[\mathrm{I} - B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}\right]z \\ &= -\lambda - \frac{1}{4\lambda}\left[z^{\mathrm{T}}H^{-1}z - z^{\mathrm{T}}H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}z\right], \end{aligned}$$

and finally,

$$\frac{\partial}{\partial\lambda}Q(a^*(\lambda,\mu^*),\lambda,\mu^*) = 0 \Rightarrow \lambda = \lambda^* = \pm\frac{1}{2}\sqrt{z^{\mathrm{T}}H^{-1}z - z^{\mathrm{T}}H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}z}.$$

Consequently, $a = a^*(\lambda^*,\mu^*) = \mp\frac{H^{-1}\left[z-B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}z\right]}{\sqrt{z^{\mathrm{T}}H^{-1}z-z^{\mathrm{T}}H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}z}}$ and then $J(a) = J(a^*(\lambda^*,\mu^*)) = \mp\sqrt{z^{\mathrm{T}}H^{-1}z - z^{\mathrm{T}}H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}z}$, so we need to take $\pm = -$, and then we have[8]

$$a = a^* = a^*(\lambda^*,\mu^*)) = \frac{H^{-1}\left[z - B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}z\right]}{\sqrt{z^{\mathrm{T}}H^{-1}z - z^{\mathrm{T}}H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}z}}. \tag{3.17}$$

and

$$J(a) = J(a^*(\lambda^*,\mu^*)) = \sqrt{z^{\mathrm{T}}H^{-1}z - z^{\mathrm{T}}H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}z}. \tag{3.18}$$

Equations (3.17) and (3.18) simplify considerably if we impose no further conditions like, for example, wc, on ψ, so $B = b$. Note that b is the first column of H, so $H^{-1}b = e_0 \in \mathbb{R}^{n+1}$, the vector with 1 in the first place and zeros elsewhere.

---

[7] B is of full rank, so $Bz = 0 \Leftrightarrow z = 0$, which means that $B^{\mathrm{T}}H^{-1}B$ is positive definite and hence invertible for $n \geq p'$, where p' is the number of columns of B. Maple®calculations show that, for $n \geq p'$, $n \leq 12$, the inverse 1-norm condition number of $B^{\mathrm{T}}H^{-1}B$ is always greater than $10^9\epsilon_M$, where $\epsilon_M = 2.2204\times10^{-16}$ is the machine epsilon for the machine used in our calculations, so the worst of them is not near-singular, even if it is far from being excellently conditioned. (for a discussion of the condition number, see Burden and Faires[39])

[8] Maple®calculations show that, for $n \geq p'$, $n \leq 12$ and our Bs, $H^{-1} - H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}$ is positive semi-definite and of rank $n - p' + 1$. But the subspace $\mathrm{im}\,B = \{z : z = Bd, d \in \mathbb{R}^{p'}\}$ of $\mathbb{R}^{n+1}$ is of dimension p' for all B of full rank (and all the Bs we consider are of full rank), $[H^{-1} - H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}]Bd = 0$ and $H^{-1} - H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1} \in \mathbb{R}^{(n+1)\times(n+1)}$, so $\mathrm{im}\,B = \ker(H^{-1} - H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1})$. This all means that $z \notin \mathrm{im}\,B$ (in line with our assumption) implies $z^{\mathrm{T}}H^{-1}z - z^{\mathrm{T}}H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}z > 0$, at least for the conditions under consideration

Hence, in this special case,

$$a = \frac{1}{\sqrt{z^{\mathrm{T}}H^{-1}z - z_0^2}}[H^{-1}z - z_0 e_0], \tag{3.19}$$

and

$$J(a) = \sqrt{z^{\mathrm{T}}H^{-1}z - z_0^2}, \tag{3.20}$$

where we have assumed that $z^{\mathrm{T}}[H^{-1} - e_0 e_0^{\mathrm{T}}]z \neq 0$ throughout. If $H^{-1} - e_0 e_0^{\mathrm{T}}$ is positive semi-definite (which it is for $n \leq 10$), $z^{\mathrm{T}}[H^{-1} - e_0 e_0^{\mathrm{T}}]z = 0 \Leftrightarrow [H^{-1} - e_0 e_0^{\mathrm{T}}]z = H^{-1}[\mathrm{I} - He_0 e_0^{\mathrm{T}}]z = 0 \Leftrightarrow [\mathrm{I} - be_0^{\mathrm{T}}]z = z - z_0 b = 0$. But then either $z = 0$, or $a \perp b \Rightarrow a \perp z$, and, in both these cases, $J(a) = 0$ for all a satisfying the constraints.

The advantage of maximising $\langle f, \psi \rangle$ over minimising $\sum_{k=1}^{N}(y_k - \psi(x_k))^2$ is that the solution to the latter optimisation involves the inversion of a matrix dependent on the positions of the x_k for each sample, whereas the former requires the inversion of a matrix which is constant for each value of n. Even if the x_k are evenly spaced as they are here, in the minimisation a matrix inversion would have to be carried out for each pair (n, N).

In addition, fitting to an interval $\mathcal{I}$ over which f is constant rather than to a point $x \in \mathcal{I}$ will have a minor regularising effect, as, if the fitted curve changes too rapidly in an interval around x, the contributions to the error at points of $\mathcal{I} - \{x\}$ will be large.

Although H^{-1} can be calculated exactly, and its entries are all integers, the ratio between the smallest and largest magnitudes of its nonzero entries decreases fairly rapidly with n, resulting in a poorly conditioned matrix and restricting its use in double precision floating-point calculations when n exceeds about 8. We can use Matlab®'s variable precision arithmetic to push the practically usable n to higher values, about 14, at the cost of increased processing time.

3.2.4 Fitting wavelet triplets to the data

As our data at each time point consists of three components, $\mathbf{y}_k = [\ y_{k1} \quad y_{k2} \quad y_{k3}\ ]$, giving the acceleration for each accelerometer axis, we are naturally interested in fitting a three-dimensional entity to it, or rather the function $\mathbf{f} = [\ f_1 \quad f_2 \quad f_3\ ]$ representing it. We use wavelet triplets $\boldsymbol{\psi} = [\ \psi_1 \quad \psi_2 \quad \psi_3\ ]^{\mathrm{T}}$, where each component satisfies the condition of equation (1.2), $\int_{-\infty}^{\infty}\psi_j(x)\mathrm{d}x = b^{\mathrm{T}}a_j = 0$, but we replace the condition of equation (1.1) by

$$\|\psi_1\|^2 + \|\psi_2\|^2 + \|\psi_3\|^2 = a_1^{\mathrm{T}}Ha_1 + a_2^{\mathrm{T}}Ha_2 + a_3^{\mathrm{T}}Ha_3 = 1, \tag{3.21}$$

where, of course, $\psi_j(x) = \sum_{k=0}^{n} a_{jk}x^k$.

We adapt the method of the previous section to fit $\boldsymbol{\psi}$ to $\mathbf{f}$.

The analogue of equation (3.13) is $J(a_1, a_2, a_3) = z_1^{\mathrm{T}}a_1 + z_2^{\mathrm{T}}a_2 + z_3^{\mathrm{T}}a_3$, where each $z_j = Ry_j$ for the same R as previously, and the analogue of equation (3.16) is $Q(a_1, a_2, a_3, \lambda, \mu_1, \mu_2, \mu_3) = \lambda\sum_{j=1}^{3}a_j^{\mathrm{T}}Ha_j + \sum_{j=1}^{3}(B\mu_j + z_j)^{\mathrm{T}}a_j - \lambda$.

We quickly obtain

$$\mu_j = \mu_j^* = -(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}z_j,$$

and

$$a_j = a_j^{**}(\lambda) = a_j^*(\lambda, \mu_j^*) = -\frac{1}{2\lambda}H^{-1}(z_j + B\mu_j^*) = -\frac{1}{2\lambda}H^{-1}[\mathrm{I} - B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}]z_j,$$

so

$$\begin{aligned} Q^*(\lambda) &= Q(a_1^{**}(\lambda), a_2^{**}(\lambda), a_3^{**}(\lambda), \lambda, \mu_1^*, \mu_2^*, \mu_3^*) \\ &= -\frac{1}{4\lambda}\sum_{j=1}^{3} z_j^{\mathrm{T}}[H^{-1} - H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}]z_j - \lambda. \end{aligned}$$

The equation $\frac{\mathrm{d}}{\mathrm{d}\lambda}Q^*(\lambda) = 0$ and the fact that we are maximising $J(a_1, a_2, a_3)$ then determine

$$\lambda = -\frac{1}{2}\sqrt{\sum_{k=1}^{3} z_k^{\mathrm{T}}[H^{-1} - H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}]z_k},$$

so

$$a_j = \frac{H^{-1}[\mathrm{I} - B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}]}{\sqrt{\sum_{k=1}^{3} z_k^{\mathrm{T}}[H^{-1} - H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}]z_k}}z_j, \tag{3.22}$$

$$z_j^{\mathrm{T}}a_j = \frac{z_j^{\mathrm{T}}[H^{-1} - H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}]z_j}{\sqrt{\sum_{k=1}^{3} z_k^{\mathrm{T}}[H^{-1} - H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}]z_k}}, \tag{3.23}$$

and

$$J(a_1, a_2, a_3) = \sqrt{\sum_{k=1}^{3} z_k^{\mathrm{T}}[H^{-1} - H^{-1}B(B^{\mathrm{T}}H^{-1}B)^{-1}B^{\mathrm{T}}H^{-1}]z_k}. \tag{3.24}$$

As in the case of fitting a single wavelet to data, when there are no additional constraints and $B = b$, these simplify and

$$a_j = \frac{1}{\sqrt{\sum_{k=1}^{3}[z_k^{\mathrm{T}}H^{-1}z_k - z_{k0}^2]}}[H^{-1}z_j - z_{j0}e_0], \tag{3.25}$$

$$z_j^{\mathrm{T}}a_j = \frac{z_j^{\mathrm{T}}H^{-1}z_j - z_{j0}^2}{\sqrt{\sum_{k=1}^{3}[z_k^{\mathrm{T}}H^{-1}z_k - z_{k0}^2]}}, \tag{3.26}$$

and

$$J(a_1, a_2, a_3) = \sqrt{\sum_{k=1}^{3}[z_k^{\mathrm{T}}H^{-1}z_k - z_{k0}^2]}. \tag{3.27}$$

3.3 Using the fitted wavelet — ECWT libraries

3.3.1 Distance between equivalence classes

As we are primarily interested in the *shape* of motions characteristic of PD, rather than their orientation in space, our libraries will actually consist of *Equivalence Classes* of Wavelet Triplets, or ECWTs. Given a wavelet triplet, $\boldsymbol{\psi} = [\ \psi_1 \quad \psi_2 \quad \psi_3\ ]^{\mathrm{T}}$, its equivalence class in the present sense is the set of all $\boldsymbol{\psi}'$ related to $\boldsymbol{\psi}$ by a rotation (not necessarily a proper rotation, as reflections will also be allowed):

$$[\boldsymbol{\psi}] = \left\{ \sum_{j=1}^{3} \left[\ O_{1j}\psi_j \quad O_{2j}\psi_j \quad O_{3j}\psi_j\ \right]^{\mathrm{T}}, O \in \mathrm{O}(3) \right\}, \tag{3.28}$$

where O(3) is the set of orthogonal matrices operating on $\mathbb{R}^3$.

An appropriate distance between these equivalence classes is

$$\begin{aligned}
D([\boldsymbol{\psi}],[\boldsymbol{\psi}']) &= \min_{\boldsymbol{\psi}''\in[\boldsymbol{\psi}],\boldsymbol{\psi}'''\in[\boldsymbol{\psi}']} \|\boldsymbol{\psi}''-\boldsymbol{\psi}'''\| = \min_{O'',O'''\in\mathrm{O}(3)} \|O''\boldsymbol{\psi}-O'''\boldsymbol{\psi}'\| \\
&= \min_{O'',O'''\in\mathrm{O}(3)} \sqrt{\sum_{j=1}^{3} \|(O''\boldsymbol{\psi})_j-(O'''\boldsymbol{\psi}')_j\|^2} \\
&= \min_{O'',O'''\in\mathrm{O}(3)} \sqrt{\sum_{j,k,\ell=1}^{3} (O''_{jk}O''_{j\ell}\langle\psi_k,\psi_\ell\rangle - 2O''_{jk}O'''_{j\ell}\langle\psi_k,\psi'_\ell\rangle + O'''_{jk}O'''_{j\ell}\langle\psi'_k,\psi'_\ell\rangle)} \\
&= \min_{O'',O'''\in\mathrm{O}(3)} \sqrt{\sum_{j=1}^{3} \left[\|\psi_j\|^2 - 2\sum_{k,\ell=1}^{3} \langle O''_{jk}O'''_{j\ell}\psi_k,\psi'_\ell\rangle + \|\psi'_j\|^2 \right]} \\
&= \min_{O'',O'''\in\mathrm{O}(3)} \sqrt{2\left[1-\sum_{j,k,\ell=1}^{3} \langle O''_{jk}O'''_{j\ell}\psi_k,\psi'_\ell\rangle\right]} \\
&= \min_{O'',O'''\in\mathrm{O}(3)} \|O'''^{\mathrm{T}}O''\boldsymbol{\psi}-\boldsymbol{\psi}'\| = \min_{O\in\mathrm{O}(3)} \|O\boldsymbol{\psi}-\boldsymbol{\psi}'\|.
\end{aligned} \tag{3.29}$$

As the minimisation of $\|O\boldsymbol{\psi}-\boldsymbol{\psi}'\|$ is equivalent to the minimisation of $\|O\boldsymbol{\psi}-\boldsymbol{\psi}'\|^2 = 2[1-\langle O\boldsymbol{\psi},\boldsymbol{\psi}'\rangle] = 2[1-\langle O\boldsymbol{\psi},\boldsymbol{\psi}'\rangle]$, so it is equivalent to the maximisation of $\langle O\boldsymbol{\psi},\boldsymbol{\psi}'\rangle$, all extremisations here being with respect to O and subject to O being orthogonal.

But $\langle O\boldsymbol{\psi},\boldsymbol{\psi}'\rangle = \sum_{j,k=1}^{3} O_{jk}\langle\psi_k,\psi'_j\rangle = \sum_{j,k=1}^{3} O_{jk}a_k^{\mathrm{T}}Ha'_j$ (where $\psi_j(x) = \sum_{k=0}^{n} a_{jk}f_k(x)$, $\psi'_j(x) = \sum_{k=0}^{n} a'_{jk}f_k(x)$), so, writing $K_{jk} = a_j^{\mathrm{T}}Ha'_k$, we wish to maximise $\mathrm{Tr}(O^{\mathrm{T}}K)$, subject to O being orthogonal.

We first consider KK^{T}. If $c \in \mathbb{R}^3$, then $c^{\mathrm{T}}KK^{\mathrm{T}}c = \sum_{i,\ell,p=1}^{3}\sum_{j,k,m,q=0}^{n} c_i a_{ij}H_{jk}a'_{\ell k}a'_{\ell m}H_{mq}a_{pq}c_p = \left[\sum_{p=1}^{3} c_p a_p\right]^{\mathrm{T}} H \left[\sum_{p=1}^{3} a'_p a'^{\mathrm{T}}_p\right] H \left[\sum_{p=1}^{3} c_p a_p\right]$, so $c^{\mathrm{T}}KK^{\mathrm{T}}c = 0$ implies

1 $c = 0$;

2 or $\exists c \neq 0$ such that $H\left[\sum_{p=1}^{3} c_p a_p\right] = 0$. In this case, as H is nonsingular, we must have the linear dependence of the vectors a_p;

3 or $\exists c$ such that $H\sum_{p=1}^{3} c_p a_p \neq 0$ but $\left[H\sum_{p=1}^{3} c_p a_p\right]^{\mathrm{T}}\left[\sum_{p=1}^{3} a'_p a'^{\mathrm{T}}_p\right]\left[H\sum_{p=1}^{3} c_p a_p\right] = 0$. In this case, the vectors a'_p must be linearly dependent.

If both sets of vectors a_p and a'_p are linearly independent (and they almost always will be in practice), then neither 2 nor 3 holds, and KK^{T} will be nonsingular. Even if this is not the case, as the vectors a_p and a'_p are all nonzero, KK^{T} will be of rank at least 1. Moreover, as K has the same rank as KK^{T}, it will be of rank at least 1, and will be nonsingular if both the a_p and the a'_p are linearly independent.

Let $K = U_K D_K V_K^{\mathrm{T}}$ be a singular-value decomposition of K. Then U_K and V_K are orthogonal matrices and D_K is a diagonal matrix which can be chosen to have descending non-negative values along its diagonal (such a decomposition exists — see []). If K is of rank r, D_K will have r non-zero entries.

Then, as $\mathrm{Tr}(O^{\mathrm{T}}K) = \mathrm{Tr}(O^{\mathrm{T}}U_K D_K V_K^{\mathrm{T}}) = \mathrm{Tr}(V_K^{\mathrm{T}}O^{\mathrm{T}}U_K D_K) = \mathrm{Tr}(O_K^{\mathrm{T}}D_K)$, where $O_K = U_K^{\mathrm{T}}OV_K$ is orthogonal if and only if O is, maximising $\mathrm{Tr}(O^{\mathrm{T}}K)$ subject to the orthogonality of O is equivalent to finding the turning points of

$$Q(O_K, L) = \mathrm{Tr}(O_K^{\mathrm{T}}D_K) + \mathrm{Tr}(L(O_K^{\mathrm{T}}O_K - \mathrm{I})) = \sum_{j=1}^{3} O_{Kjj}D_{Kjj} + \sum_{j,k,\ell=1}^{3} L_{jk}O_{K\ell k}O_{K\ell j} - \sum_{j=1}^{3} L_{jj} \tag{3.30}$$

with respect to O_K and L, where L is a matrix of Lagrange multipliers.

Now $\frac{\partial}{\partial O_{Krs}}Q(O_K, L) = D_{Krs} - \sum_{j=1}^{3} L_{js}O_{Krj} + \sum_{k=1}^{3} L_{sk}O_{Krk}$, so $\frac{\partial}{\partial O_{Krs}}Q(O_K, L) = 0$ implies

$$D_K + O_K(L + L^{\mathrm{T}}) = 0, \tag{3.31}$$

At this point we assume that D_K is of full rank, so $L + L^{\mathrm{T}} = -O_K^{\mathrm{T}}D_K$ is invertible, and

$$O_K = -D_K(L + L^{\mathrm{T}})^{-1}. \tag{3.32}$$

Instead of continuing down the Lagrangian path, we impose orthogonality on the O_K of equation (3.32), in the form $O_K^{\mathrm{T}}O_K = \mathrm{I}$: $(L + L^{\mathrm{T}})^{-1}D_K^2(L + L^{\mathrm{T}})^{-1} = \mathrm{I}$ or $L + L^{\mathrm{T}} = (D_K^2)^{\frac{1}{2}}$, where $(D_K^2)^{\frac{1}{2}}$ is one of the square roots of the diagonal positive definite matrix D_K^2 (we will use $(D_K^2)^{-\frac{1}{2}}$ for the inverse of $(D_K^2)^{\frac{1}{2}}$). This means that $O_K = -D_K(D_K^2)^{-\frac{1}{2}}$ and $\mathrm{Tr}(O_K^{\mathrm{T}}D_K) = -\mathrm{Tr}((D_K^2)^{-\frac{1}{2}}D_K^2) = -\mathrm{Tr}((D_K^2)^{\frac{1}{2}})$. Obviously, $\mathrm{Tr}(O_K^{\mathrm{T}}D_K)$ is now maximised by choosing $(D_K^2)^{\frac{1}{2}} = -D_K$, so $O_K = \mathrm{I}$, $O = U_K V_K^{\mathrm{T}}$ and $\mathrm{Tr}(O_K^{\mathrm{T}}D_K) = \mathrm{Tr}(D_K)$. But the diagonal elements of D_K are the square roots of the eigenvalues of KK^{T}, so

$$D([\psi], [\psi']) = \sqrt{2\left[1 - \sum_{k=1}^{3} \mu_k^{\frac{1}{2}}\right]}, \tag{3.33}$$

where the μ_k are the 3 eigenvalues of KK^{T}.

If we do not assume that D_K is of full rank, then it is of rank 1 or 2, and, by equation (3.31), $L + L^{\mathrm{T}}$ will have the same rank.

If D_K and $L + L^{\mathrm{T}}$ are of rank 2, we can write $D_K = \left[\begin{smallmatrix} D_{K1} & 0 \\ 0 & 0 \end{smallmatrix}\right]$, $O_K = \left[\begin{smallmatrix} O_1 & p \\ q^{\mathrm{T}} & o_0 \end{smallmatrix}\right]$, where D_{K1} is non-singular and $D_{K1}, O_1 \in \mathbb{R}^{2\times 2}$, $p, q \in \mathbb{R}^2$ and $o_0 \in \mathbb{R}$. We can also replace $L + L^{\mathrm{T}}$ by a spectral decomposition $U_L D_L U_L^{\mathrm{T}}$, where U_L is orthogonal and D_L is diagonal, with the eigenvalues of $L + L^{\mathrm{T}}$ along its diagonal. These decompositions exist as $L + L^{\mathrm{T}}$ is symmetric (see any good book on linear algebra, or, for a proof including the infinite-dimensional analogue, Bachman and Narici[40]), and $U_L D_L U_L^{\mathrm{T}}$ may chosen such that $D_L = \left[\begin{smallmatrix} D_{L1} & 0 \\ 0 & 0 \end{smallmatrix}\right]$, and then we can put $U_L = \left[\begin{smallmatrix} U_1 & u \\ v^{\mathrm{T}} & u_0 \end{smallmatrix}\right]$, for non-singular D_{L1}, where $D_{L1}, U_1 \in \mathbb{R}^{2\times 2}$, $u, v \in \mathbb{R}^2$ and $u_0 \in \mathbb{R}$.

Equation (3.31) becomes $D_K + O_K U_L D_L U_L^{\mathrm{T}} = 0$ and leads to

$$\begin{bmatrix} O_1 & p \\ q^{\mathrm{T}} & o_0 \end{bmatrix} \begin{bmatrix} U_1 & u \\ v^{\mathrm{T}} & u_0 \end{bmatrix} \begin{bmatrix} D_{L1} & 0 \\ 0 & 0 \end{bmatrix} = - \begin{bmatrix} D_{K1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1 & u \\ v^{\mathrm{T}} & u_0 \end{bmatrix}$$

or

$$\begin{bmatrix} O_1 U_1 D_{L1} + p v^{\mathrm{T}} D_{L1} & 0 \\ q^{\mathrm{T}} U_1 D_{L1} + o_0 v^{\mathrm{T}} D_{L1} & 0 \end{bmatrix} = - \begin{bmatrix} D_{K1} U_1 & D_{K1} u \\ 0 & 0 \end{bmatrix}, \tag{3.34}$$

which means that

$$\begin{aligned} u &= 0 \text{ (as } D_{K1} \text{ is non-singular)}, \\ q^{\mathrm{T}} U_1 + o_0 v^{\mathrm{T}} &= 0 \text{ (as } D_{L1} \text{ is non-singular)} \\ \text{and } O_1 U_1 D_{L1} + p v^{\mathrm{T}} D_{L1} &= -D_{K1} U_1. \end{aligned} \tag{3.35}$$

Then the orthogonality of U_L means that

$$U_L^{\mathrm{T}} U_L = \begin{bmatrix} U_1^{\mathrm{T}} & v \\ 0 & u_0 \end{bmatrix} \begin{bmatrix} U_1 & 0 \\ v^{\mathrm{T}} & u_0 \end{bmatrix} = \begin{bmatrix} U_1^{\mathrm{T}} U_1 + v v^{\mathrm{T}} & u_0 v \\ u_0 v^{\mathrm{T}} & u_0^2 \end{bmatrix} = \begin{bmatrix} \mathrm{I} & 0 \\ 0 & 1 \end{bmatrix} \tag{3.36}$$

and we can conclude that $v = 0$, $u_0 = \pm 1$ and U_1 is orthogonal (and hence non-singular).

Using these facts in equations (3.35) leads to $q = 0$ and $O_1 U_1 D_{L1} = -D_{K1} U_1$. But $q = 0$ and the orthogonality of O_K leads to the conclusion that $p = 0$, $o_0 = \pm_1 1$ and O_1 is orthogonal (by the same argument used for the sub-components of U_L — we add suffixes to $\pm$, $\mp$ to indicate independent instances, but, of course $\mp_k 1 = -(\pm_k 1)$).

As we have $O_1 = -D_{K1} U_1 D_{L1}^{-1} U_1^{\mathrm{T}}$, the orthogonality of O_1 requires $D_{K1} U_1 D_{L1}^{-2} U_1^{\mathrm{T}} D_{K1} = \mathrm{I}$, which implies $U_1 D_{L1}^{-2} U_1^{\mathrm{T}} = D_{K1}^{-2}$. We may write $U_1 = \left[\begin{smallmatrix} \cos\theta & \sin\theta \\ \mp_2 \sin\theta & \pm_2 \cos\theta \end{smallmatrix}\right]$, $\theta \in [0, 2\pi)$ (such an expression holds for any member of O(2)), $D_{L1}^{-2} = \left[\begin{smallmatrix} \nu_1 & 0 \\ 0 & \nu_2 \end{smallmatrix}\right]$ and $D_{K1}^{-2} = \left[\begin{smallmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{smallmatrix}\right]$, and then $U_1 D_{L1}^{-2} U_1^{\mathrm{T}} = \left[\begin{smallmatrix} \cos\theta & \sin\theta \\ \mp_2 \sin\theta & \pm_2 \cos\theta \end{smallmatrix}\right] \left[\begin{smallmatrix} \nu_1 & 0 \\ 0 & \nu_2 \end{smallmatrix}\right] \left[\begin{smallmatrix} \cos\theta & \mp_2 \sin\theta \\ \sin\theta & \pm_2 \cos\theta \end{smallmatrix}\right] = \left[\begin{smallmatrix} \nu_1 \cos^2\theta + \nu_2 \sin^2\theta & (\mp_2 \nu_1 \pm_2 \nu_2) \sin\theta \cos\theta \\ (\mp_2 \nu_1 \pm_2 \nu_2) \sin\theta \cos\theta & \nu_1 \sin^2\theta + \nu_2 \cos^2\theta \end{smallmatrix}\right] = \left[\begin{smallmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{smallmatrix}\right] = D_{K1}^{-2}$, and this can only hold if

1 $\nu_2 = \nu_1$, and then $\kappa_1 = \kappa_2 = \nu_1$, $D_{L1} = \pm_3 \nu_1^{-\frac{1}{2}} \mathrm{I}$, $D_{K1} = \pm_4 \nu_1^{-\frac{1}{2}} \mathrm{I}$, so $O_1 = \mp_3 \pm_4 \mathrm{I}$ and $\mathrm{Tr}(O_K^{\mathrm{T}} D_K) = \mp_3 \pm_4 \mathrm{Tr}(D_K)$;

2 or $\sin\theta = 0$, so $\cos\theta = \pm_3 1$, $U_1 = \pm_3 \left[\begin{smallmatrix} 1 & 0 \\ 0 & \pm_2 1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} \nu_1 & 0 \\ 0 & \nu_2 \end{smallmatrix}\right] = \left[\begin{smallmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{smallmatrix}\right]$, which means $D_{K1} = \begin{bmatrix} \pm_4 \kappa_1^{-\frac{1}{2}} & 0 \\ 0 & \pm_5 \kappa_2^{-\frac{1}{2}} \end{bmatrix}$, $D_{L1} = \begin{bmatrix} \pm_6 \kappa_1^{-\frac{1}{2}} & 0 \\ 0 & \pm_7 \kappa_2^{-\frac{1}{2}} \end{bmatrix}$, and $O_1 = \left[\begin{smallmatrix} \mp_4 \pm_6 1 & 0 \\ 0 & \mp_5 \pm_7 1 \end{smallmatrix}\right]$. Consequently, $\mathrm{Tr}(O_K^{\mathrm{T}} D_K) = \mp_6 \kappa_1^{-\frac{1}{2}} \mp_7 \kappa_2^{-\frac{1}{2}}$;

3 or $\cos\theta = 0$, so $\sin\theta = \pm_3 1$, $U_1 = \pm_3 \left[\begin{smallmatrix} 0 & 1 \\ \mp_2 1 & 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} \nu_2 & 0 \\ 0 & \nu_1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} \kappa_1 & 0 \\ 0 & \kappa_2 \end{smallmatrix}\right]$. We then have $D_{K1} = \begin{bmatrix} \pm_4 \kappa_1^{-\frac{1}{2}} & 0 \\ 0 & \pm_5 \kappa_2^{-\frac{1}{2}} \end{bmatrix}$, $D_{L1} = \begin{bmatrix} \pm_6 \kappa_2^{-\frac{1}{2}} & 0 \\ 0 & \pm_7 \kappa_1^{-\frac{1}{2}} \end{bmatrix}$, $O_1 = \left[\begin{smallmatrix} \mp_4 \pm_7 1 & 0 \\ 0 & \mp_5 \pm_6 1 \end{smallmatrix}\right]$ and $\mathrm{Tr}(O_K^{\mathrm{T}} D_K) = \mp_7 \kappa_1^{-\frac{1}{2}} \mp_8 \kappa_2^{-\frac{1}{2}}$.

Clearly, $\mathrm{Tr}(O_K^{\mathrm{T}} D_K) = \mathrm{Tr}(D_K)$ is among the values of $\mathrm{Tr}(O_K^{\mathrm{T}} D_K)$ that can be obtained here by choice of the $\pm$s, and is maximal among these values, so equation (3.33) also holds when the rank of K is 2.

By an analogous argument, we will not be presented here, equation (3.33) also holds when that rank is 1.

Before continuing, we first check that D *is* a distance; that, is, for all equivalence classes $[\boldsymbol{\psi}_1], [\boldsymbol{\psi}_2], [\boldsymbol{\psi}_3]$ we have

1 $D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_2]) \geq 0$, and $D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_2]) = 0 \Leftrightarrow [\boldsymbol{\psi}_1] = [\boldsymbol{\psi}_2]$;

2 $D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_2]) = D([\boldsymbol{\psi}_2], [\boldsymbol{\psi}_1])$;

3 $D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_2]) + D([\boldsymbol{\psi}_2], [\boldsymbol{\psi}_3]) \geq D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_3])$.

1 and 2 are trivial consequences of the definitions of D and our equivalence classes, and we have

$$\begin{aligned}
\|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_2\| + \|\boldsymbol{\psi}_2 - \boldsymbol{\psi}_3\| &\geq \|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_3\| \geq D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_3]) \; \forall \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\psi}_3 \\
\Rightarrow \|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_2\| + \|\boldsymbol{\psi}_2 - O_3\boldsymbol{\psi}_3\| &\geq D([\boldsymbol{\psi}_1], [O_3\boldsymbol{\psi}_3]) = D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_3]) \; \forall \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\psi}_3; O_3 \in \mathrm{O}(3) \\
\Rightarrow \|O_1\boldsymbol{\psi}_1 - \boldsymbol{\psi}_2\| + \|\boldsymbol{\psi}_2 - O_3\boldsymbol{\psi}_3\| &\geq D([O_1\boldsymbol{\psi}_1], [\boldsymbol{\psi}_3]) = D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_3]) \\
&\qquad \forall \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\psi}_3; O_1, O_3 \in \mathrm{O}(3) \\
\Rightarrow D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_2]) + D([\boldsymbol{\psi}_2], [\boldsymbol{\psi}_3]) &\geq D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_3]) \; \forall \boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\psi}_3.
\end{aligned}$$

3.3.2 L_2 distance between offset triplets, and a further distance between equivalence classes

It is trivial to extend our distance D and equivalence classes from triplets of mother wavelets to triplets of child wavelets at scale 1 and non-zero shifts: using $\psi_{j;x_0}(x) = \psi_j(x - x_0)$, etc., equation (3.28) becomes

$$[\boldsymbol{\psi}_{;x_0}] = \left\{ \sum_{j=1}^{3} \begin{bmatrix} O_{1j}\psi_{j;x_0} & O_{2j}\psi_{j;x_0} & O_{3j}\psi_{j;x_0} \end{bmatrix}^{\mathrm{T}}, O \in \mathrm{O}(3) \right\}$$

(we note that $[\boldsymbol{\psi}'_{;x'_0}] = [\boldsymbol{\psi}_{;x_0}] \Rightarrow x'_0 = x_0$ — consider the supports of the functions involved), and equation (3.29) becomes

$$D([\boldsymbol{\psi}_{;x_0}], [\boldsymbol{\psi}'_{;x'_0}]) = \min_{\boldsymbol{\psi}''_{;x_0} \in [\boldsymbol{\psi}_{;x_0}], \boldsymbol{\psi}'''_{;x'_0} \in [\boldsymbol{\psi}'_{;x'_0}]} \|\boldsymbol{\psi}''_{;x_0} - \boldsymbol{\psi}'''_{;x'_0}\|.$$

Of course, the extension of these things to triplets of child wavelets at different scales is just as trivial, but we are not interested in this extension at this point.

Because of the way that shifts combine we have $D([\boldsymbol{\psi}_{1;x_0}], [\boldsymbol{\psi}_{2;x'_0}]) = D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_{2;x'_0 - x_0}])$, and, if $x_0 \in [0, 1]$, we have

$$\begin{aligned} \langle \psi_{1j}, \psi_{2j;x_0} \rangle &= \sum_{k,\ell=0}^{n} a_{1jk} a_{2j\ell} \int_{-\frac{1}{2}+x_0}^{\frac{1}{2}} x^k (x - x_0)^\ell \mathrm{d}x \\ &= a_{1j}^{\mathrm{T}} H_1(x_0) a_{2j}. \end{aligned} \tag{3.37}$$

A similar expression holds for x_0 outside of $[0, 1]$, and we have

$$H_{1k\ell}(x_0) = \begin{cases} \int_{-\frac{1}{2}+x_0}^{\frac{1}{2}} x^k (x - x_0)^\ell \mathrm{d}x, & x_0 \in [-1, 0]; \\ \int_{-\frac{1}{2}}^{\frac{1}{2}+x_0} x^k (x - x_0)^\ell \mathrm{d}x, & x_0 \in [0, 1]; \\ 0, & \text{otherwise.} \end{cases} \tag{3.38}$$

Thus,

$$\|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_{2;x_0}\|^2 = \|\boldsymbol{\psi}_1\|^2 + \|\boldsymbol{\psi}_{2;x_0}\|^2 - 2 \sum_{j=0}^{3} a_{1j}^{\mathrm{T}} H_1(x_0) a_{2j} = 2 \left[1 - \sum_{j=0}^{3} a_{1j}^{\mathrm{T}} H_1(x_0) a_{2j} \right] \tag{3.39}$$

By argument completely analogous that leading to equation (3.33), except that $H_1(x_0)$ plays the role of H, we obtain

$$D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_{2;x_0}]) = \sqrt{2 \left[1 - \sum_{k=1}^{3} \nu_k(x_0)^{\frac{1}{2}} \right]}, \tag{3.40}$$

where the $\nu_k(x_0)$ are the 3 eigenvalues of $K_1(x_0) K_1(x_0)^{\mathrm{T}}$, $K_{1k\ell} = a_{1k}^{\mathrm{T}} H_1(x_0) a_{2\ell}$.

We now show that it is not necessarily the case that $\min_{x_0} D([\boldsymbol{\psi}_{1;x_0}], [\boldsymbol{\psi}_2]) = D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_2])$.

In Figure 3.1 a), left-hand column of three diagrams, we display the triplets $\boldsymbol{\psi}_1$ (blue)

$$\begin{aligned}
\psi_{11}(x) &= \begin{cases} 2x, & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise,} \end{cases} \\
\psi_{12}(x) &= \begin{cases} \frac{1}{2}\sqrt{\frac{5}{3}}(12x^2-1), & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise,} \end{cases} \\
\psi_{13}(x) &= \begin{cases} \sqrt{\frac{7}{3}}x(20x^2-3), & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise,} \end{cases}
\end{aligned}$$

and $\boldsymbol{\psi}_2$ (red)

$$\begin{aligned}
\psi_{21}(x) &= \begin{cases} \frac{1}{8\sqrt{3}}(-1680x^4+360x^2-9), & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise,} \end{cases} \\
\psi_{22}(x) &= \begin{cases} \frac{1}{4}\sqrt{\frac{11}{3}}x(-1008x^4+280x^2-15), & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise,} \end{cases} \\
\psi_{23}(x) &= \begin{cases} \frac{1}{16}\sqrt{\frac{13}{3}}(-14784x^6+5040x^4-420x^2-5), & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

In the right-hand column, we show the squared pointwise differences between the components of these triplets. These differences are to be integrated and summed to yield $\|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_2\|^2$. As the set of component wavelets of these triplets is mutually orthogonal, $\|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_2\|^2 = 2$, and we also have $\|O\boldsymbol{\psi}_1 - \boldsymbol{\psi}_2\|^2 = 2$ for any $O \in \mathrm{O}(3)$, so $D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_2])^2 = \|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_2\|^2$.

In Figure 3.1 b), first column, we show $\boldsymbol{\psi}_{2;x_0}(x)$ (red), and $\boldsymbol{\psi}_1(x)' = O\boldsymbol{\psi}_1(x)$, where $O \in \mathrm{O}(3)$ is $\arg\min_{O' \in \mathrm{O}_3} \|O'\boldsymbol{\psi}_1 - \boldsymbol{\psi}_{2;x_0}\|$ (blue), for $x_0 = 0.1$. We also show $\boldsymbol{\psi}_1(x)$ (blue, dashed).

In the second column of Figure 3.1 b), we show $[\psi'_{1j}(x) - \psi_{2j;x_0}(x)]^2$ (magenta) and $[\psi_{1j}(x) - \psi_{2j;x_0}(x)]^2$ (cyan).

Figures 3.2 c) and 3.2 d) are repeats of Figure 3.1 b), but with $x_0 = 0.2$ and 0.3.

It may be possible to discern by inspection of Figure 3.1 b) and Figures 3.2 c) and d) that the area under the cyan curves is greater than that under the magenta curves in the right-hand columns of diagrams, but this is certainly the case.

In Table 3.1 we present the sums of the areas under the cyan curves ($\|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_{2;x_0}\|^2$) and of those under the magenta curves ($D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_{2;x_0}])^2$), for $x_0 = 0, 0.1, 0.2$ and 0.3, and we can see, as would be expected, that $D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_{2;x_0}])^2 \leq \|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_{2;x_0}\|^2$. Moreover, as promised, we can see that the minimum value of $D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_{2;x_0}])$ is not obtained by putting $x_0 = 0$.

In fact, as we can see from Figure 3.3, which plots $D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_{2;x_0}])^2$ and $\|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_{2;x_0}\|^2$ against x_0, the global minimum of $D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_{2;x_0}])$ is attained near $x_0 = \pm 2$. More accurate values of the positions of this minimum are given in Table 3.1.

Leaving behind these particular examples of $[\boldsymbol{\psi}_1]$ and $[\boldsymbol{\psi}_2]$, for arbitrary ECWTs we can

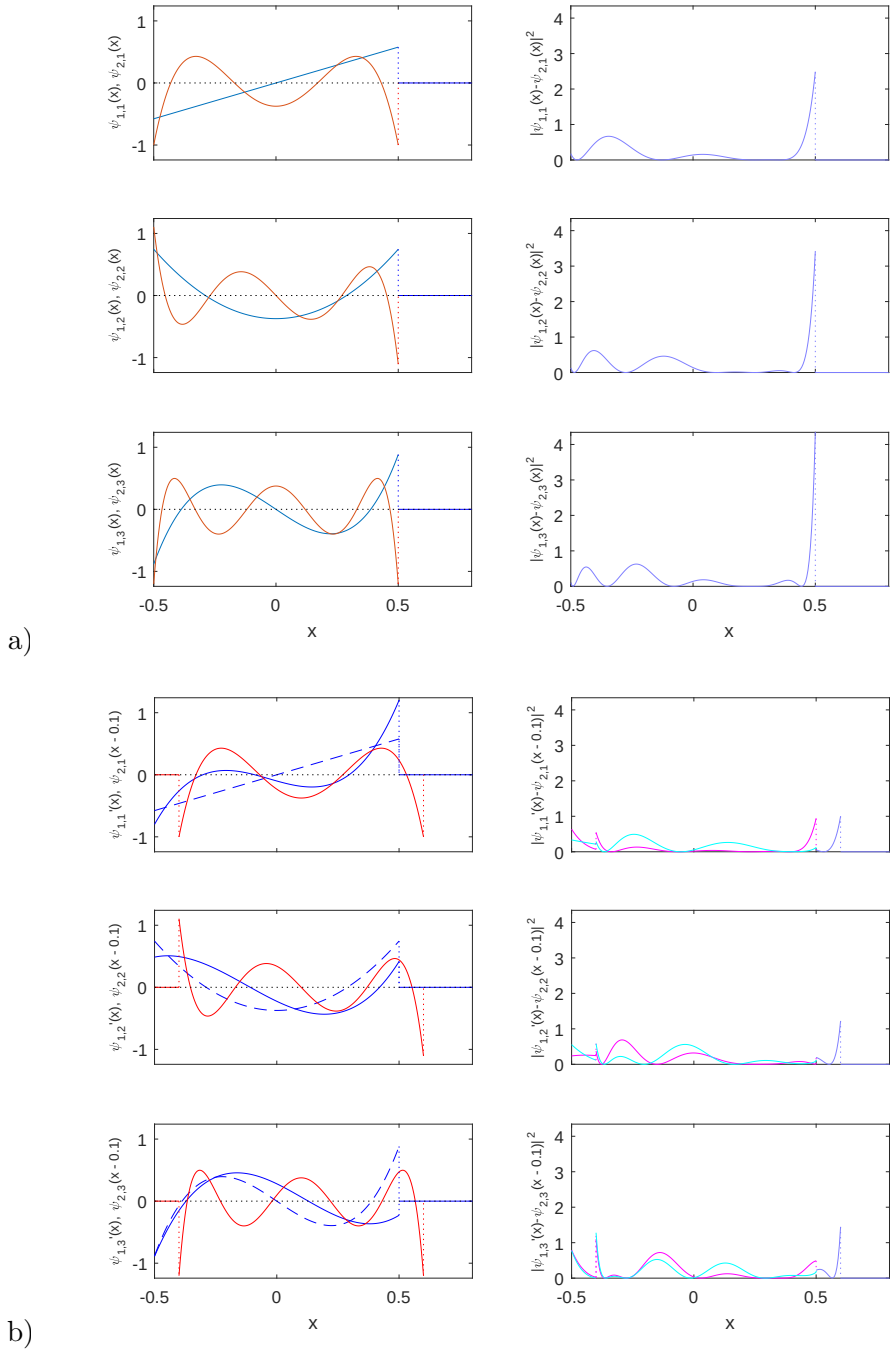
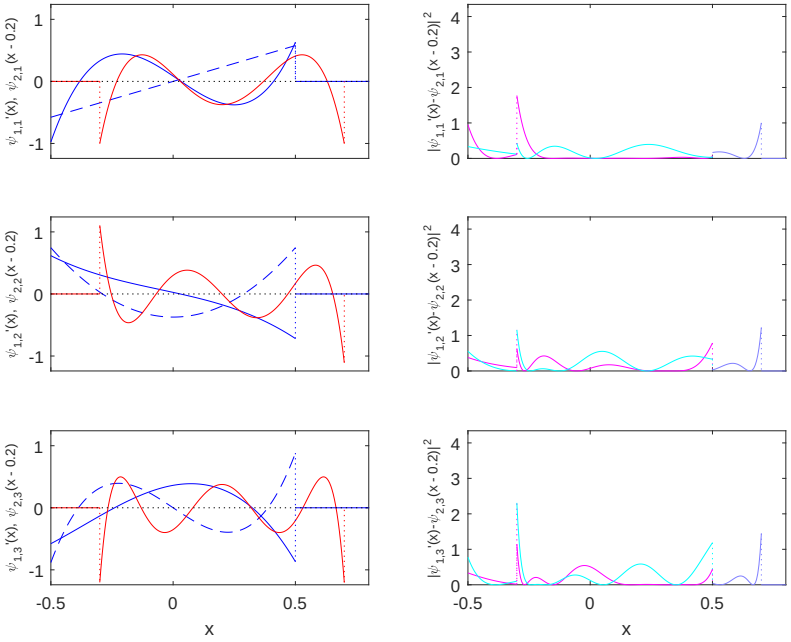


Figure 3.1: *a)* $x_0 = 0$ — *first column: component wavelets of the first triplet are in blue, those of the second are in red; second column: squared pointwise difference of wavelets in the first column; b)* $x_0 = 0.1$ — *first column: original wavelets of first triplet are dashed, the red wavelets are moved 0.1 to the right, and the solid blue curves represent the member of the equivalence class of the first triplet which is closest to the shifted second triplet; second column: squared pointwise differences between the red and blue dashed curves (cyan) and between the red and solid blue curves of the first column (magenta) (or blue where curves are coincident)*

c)



d)

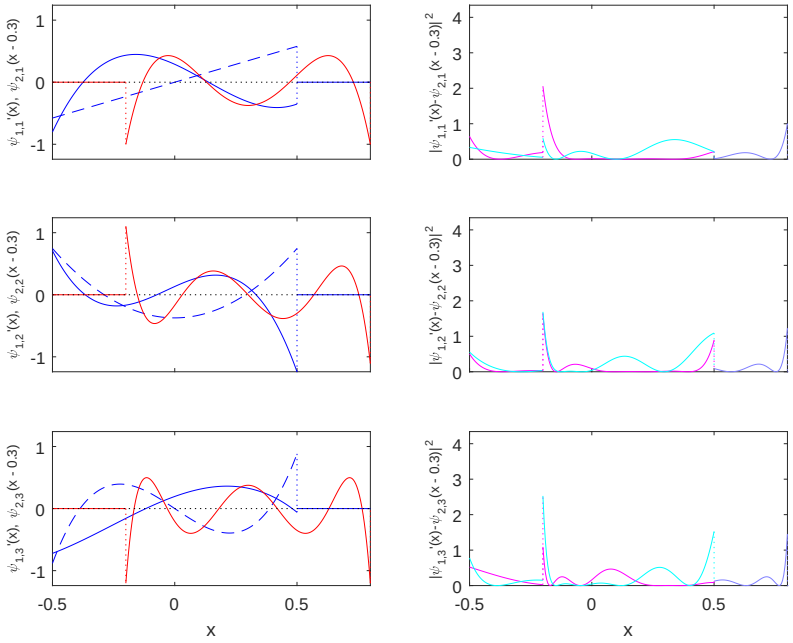


Figure 3.2: *c) and d) are analogous to Figure 3.1 b) for* $x_0 = 0.2, 0.3$

	Offset				
Square distance	0	0.1	0.2	0.3	± 0.2067
$\|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_{2;x_0}\|^2$	2	1.7685	2.2921	2.3588	2.3152
$D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}])^2$	2	1.5233	1.4659	1.5038	1.4655

Table 3.1: $\|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_{2;x_0}\|^2$ *and* $D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}])^2$ *for the values of* x_0 *of Figures 3.1 and 3.2, and for* $x_0 = \arg\min_{x_0} D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}])^2$

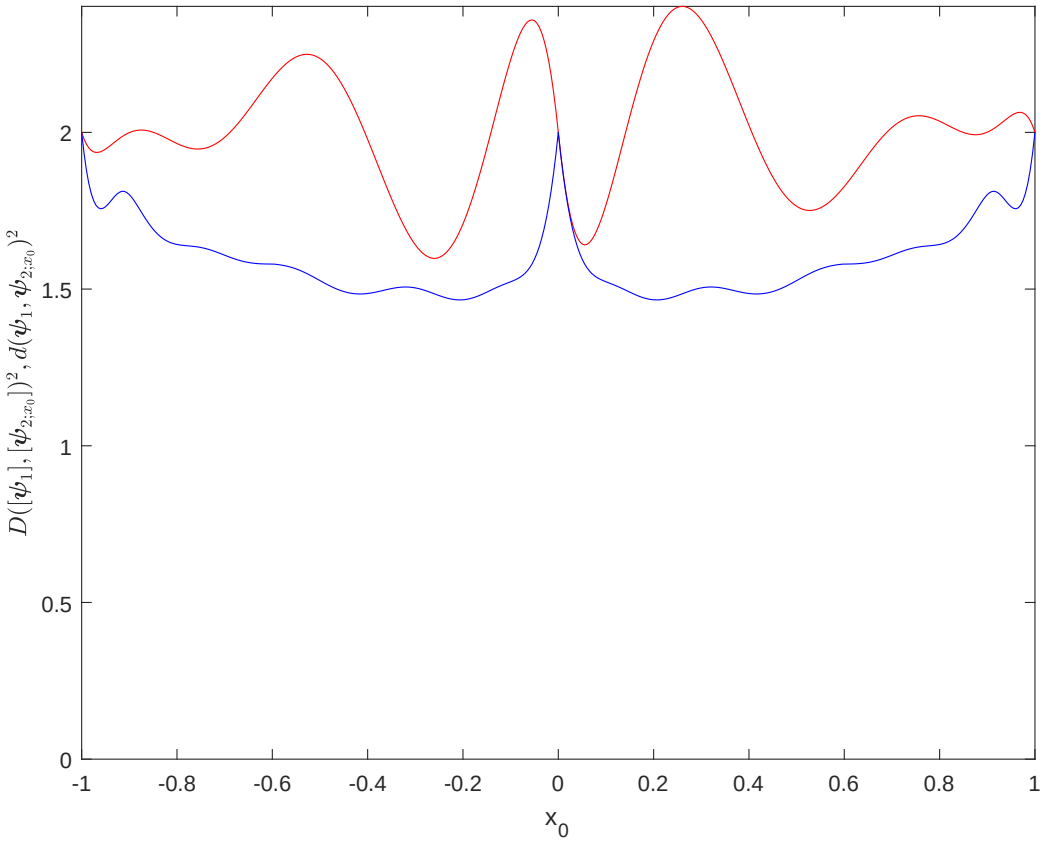


Figure 3.3: *Plots of* $D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}])^2$ *(blue) and* $\|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_{2;x_0}\|^2$ *(red) against* x_0

define

$$D^*([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_2]) = \min_{x_0} D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}]) \tag{3.41}$$

(we can use min instead of inf here, as we know x_0 is in the compact set $[-1,1]$).

We will discuss the motivation for this definition at a later point, but here we will satisfy ourselves that D^* is, in fact, a distance function.

Obviously, $D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}]) \geq 0 \ \forall x_0 \Rightarrow D^*([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_2]) \geq 0$. Moreover, if for a particular x_0, $D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}]) = 0$, then $[\boldsymbol{\psi}_1] = [\boldsymbol{\psi}_{2;x_0}]$. But this is impossible for $x_0 = 0$, as it would

imply that a non-zero polynomial was zero on a non-trivial interval. Hence, $D^*([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_2]) = 0 \Rightarrow\ \exists x_0 : D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}]) = 0 \Rightarrow D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_2]) = 0 \Rightarrow [\boldsymbol{\psi}_1] = [\boldsymbol{\psi}_2]$. That $D^*([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_2]) = D^*([\boldsymbol{\psi}_2],[\boldsymbol{\psi}_1])$ follows almost by definition.

We now need only prove $D^*([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_2]) + D^*([\boldsymbol{\psi}_2],[\boldsymbol{\psi}_3]) \geq D^*([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_3])$. But

$$\begin{aligned}
D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}]) + D([\boldsymbol{\psi}_2],[\boldsymbol{\psi}_{3;x'_0}]) &= D([\boldsymbol{\psi}_{1;-x_0}],[\boldsymbol{\psi}_2]) + D([\boldsymbol{\psi}_2],[\boldsymbol{\psi}_{3;x'_0}])\ \forall x_0, x'_0 \Rightarrow \\
D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}]) + D([\boldsymbol{\psi}_2],[\boldsymbol{\psi}_{3;x'_0}]) &\geq D([\boldsymbol{\psi}_{1;-x_0}],[\boldsymbol{\psi}_{3;x'_0}])\ \forall x_0, x'_0 \Rightarrow \\
D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}]) + D([\boldsymbol{\psi}_2],[\boldsymbol{\psi}_{3;x'_0}]) &\geq D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{3;x_0+x'_0}])\ \forall x_0, x'_0 \Rightarrow \\
D([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_{2;x_0}]) + D([\boldsymbol{\psi}_2],[\boldsymbol{\psi}_{3;x'_0}]) &\geq D^*([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_3])\ \forall x_0, x'_0 \Rightarrow \\
D^*([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_2]) + D^*([\boldsymbol{\psi}_2],[\boldsymbol{\psi}_3]) &\geq D^*([\boldsymbol{\psi}_1],[\boldsymbol{\psi}_3]).
\end{aligned}$$

3.3.3 A canonical representative of the equivalence class

As different representatives of the same equivalence class can look very different, comparisons of the ECWT will be difficult, unless we pick out a special representative of each class which depends continuously (in some sense) on the classes.

In order to make our definition of this canonical representative, we first need a reference wavelet, ψ_{ref}. We choose ψ_{ref} to be the wavelet (of our form) with underlying polynomial of least degree, and such that $\psi_{\text{ref}}(x)$ is nonnegative on some interval $\left[-\frac{1}{2}, -\frac{1}{2}+\epsilon\right)$. The zi and ue conditions suffice to determine $\psi_{\text{ref}}(x) = \begin{cases} -2\sqrt{3}x, & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise.} \end{cases}$

Given a wavelet triplet $\boldsymbol{\psi}$, we will find $\boldsymbol{\psi}' \in [\boldsymbol{\psi}]$ which maximises $\|\psi'_1\|^2$, as well as maximising $\|\psi'_2\|^2$ subject to $\|\psi'_1\|^2 = \max_{\boldsymbol{\psi}'' \in [\boldsymbol{\psi}]} \|\psi''_1\|^2$. Then $\|\psi'_3\|^2$ will be fixed at $1 - \|\psi'_1\|^2 - \|\psi'_2\|^2$. As the $\boldsymbol{\psi}'$ will only be determined up to the signs of its components, we will use ψ_{ref} to determine these signs.

First, let $\boldsymbol{\psi}' \in [\boldsymbol{\psi}]$. Then there exists $O \in \mathrm{O}(3)$ such that $\psi'_k = \sum_{j=1}^3 O_{kj}\psi_j$, so $\|\psi'_1\|^2 = \sum_{k,j=1}^3 O_{1k}O_{1j}a_k^{\mathrm{T}}Ha_j$. Writing $R_{jk} = a_k^{\mathrm{T}}Ha_j$ and letting O_1 be the first row of O, we wish to find O_1 maximising $O_1RO_1^{\mathrm{T}}$, subject to $O_1O_1^{\mathrm{T}} = 1$. Obviously, $O_1 = \pm_1 r_3^{\mathrm{T}}$ will maximise $\|\psi'_1\|^2$, where r_3 is a unit eigenvector of symmetric and positive semi-definite R corresponding to a maximal eigenvalue of R.

But there exists a unit eigenvector r_2 of R orthogonal to r_3, whose eigenvalue is either also maximal or is the largest among the submaximal eigenvalues. Setting the second row of O, $O_2 = \pm_2 r_2^{\mathrm{T}}$, where $\pm_2$ is independent of $\pm_1$, will maximise $\|\psi'_2\|^2$, under the condition that $\|\psi'_1\|^2$ is maximal.

Specifying the first two rows of $O \in \mathrm{O}(3)$ determines the third row up to sign, but it will be the case that $O_3 = \pm_3 r_1$, with the obvious notation.

Consequently, $\psi'_j = \pm_j \sum_{k=1}^3 r_{4-j,k}\psi_k$, and we now determine the signs by maximising $\langle\psi_j, \psi_{\text{ref}}\rangle = \pm_j \sum_{k=1}^3 r_{4-j,k}\langle\psi_k, \psi_{\text{ref}}\rangle = \pm_j \sum_{k=1}^3 r_{4-j,k} a_k^{\mathrm{T}}Ha_{\text{ref}} = \mp_j 2\sqrt{3}\sum_{k=1}^3 r_{4-j,k}a_k^{\mathrm{T}}H[\ 0\ \ 1\ \ 0\ \ 0\ \ \cdots\ ]^{\mathrm{T}} = \mp_j 2\sqrt{3}\sum_{k=1}^3 r_{4-j,k}a_k^{\mathrm{T}}[\ 0\ \ \frac{1}{12}\ \ 0\ \ \frac{1}{80}\ \ 0\ \ \cdots\ ]^{\mathrm{T}} = \mp_j 2\sqrt{3}\sum_{k=1}^3 r_{4-j,k}d_{\text{ref}}^{\mathrm{T}}a_k$ so we require $\mp_j d^{\mathrm{T}} r_{4-j} \geq 0$, where $d_{\text{ref}} = \left[\ 0\ \ \frac{1}{12}\ \ 0\ \ \frac{1}{80}\ \ 0\ \ \cdots\ \ \frac{1-(-1)^n}{2^{n+2}(n+2)}\ \right]^{\mathrm{T}}$,

$d = [\ d_{\text{ref}}^{\mathrm{T}} a_1 \quad d_{\text{ref}}^{\mathrm{T}} a_2 \quad d_{\text{ref}}^{\mathrm{T}} a_3\ ]^{\mathrm{T}}$. Hence, we set $\pm_j = -\operatorname{sgn} d^{\mathrm{T}} r_{4-j}$.

3.3.4 Generating the library

At this point, as we are finding the equivalence class containing the triplet closest to a function representing the data, we can allow that triplet to stand in for its class in our calculations.

We will suppose we are given a set of training data $(t_1, \mathbf{v}_1), (t_2, \mathbf{v}_2), \ldots, (t_M, \mathbf{v}_M)$ from which we are to derive a library. As our data consists of three traces, one for each channel of our tri-axial accelerometer, we will fit a wavelet, in the fashion described above, to each channel independently.

In the case of unsupervised learning, every window W_{wk} of length w in the training data, $(t_{k-w+1}, \mathbf{v}_{k-w+1}), (t_{k-w+2}, \mathbf{v}_{k-w+2}), \ldots, (t_k, \mathbf{v}_k)$, $k = w, w+1, \ldots, M$, for w in some set $\mathcal{S}_w$, has a wavelet triplet $\boldsymbol{\psi}^{(wk)}$ fitted to it. Of course, this would result in the production of a vast library if some pruning was not undertaken during the library construction process.

We first choose the desired size of our library, N_{lib}, and some $N_{\text{super}} \gg N_{\text{lib}}$. A third parameter of the library generation process is the positive integer, *peak half-width*, h_{peak}.

We then consider the quantity $G_{\text{fit}}(w, \ell) = \dfrac{\left[\sum_{j=1}^{3} \left|\left\langle f_j^{(w\ell)}, \psi_j^{(w\ell)} \right\rangle\right|\right]^2}{\sum_{j=1}^{3} \left\| f_j^{(w\ell)} - \bar{f}_j^{(w\ell)} \right\|^2}$ for the wavelet triplet $\boldsymbol{\psi}^{(w\ell)} = \left[\psi_1^{(w\ell)}, \psi_2^{(w\ell)}, \psi_3^{(w\ell)}\right]^{\mathrm{T}}$ fitted to $W_{w,\ell}$, the ℓth data window of width w, where $f_j^{(w\ell)}$ is the function modelling that data and $\bar{f}_j^{(w\ell)}$ is the mean of $f_j^{(w\ell)}$.

$\boldsymbol{\psi}_0 = [\psi_{01}, \psi_{02}, \psi_{03}]^{\mathrm{T}}$, where $\psi_{0j}(x) = \dfrac{f_j(x) - \bar{f}_j}{\sqrt{\sum_{k=1}^{3} \|f_k(x) - \bar{f}_k\|^2}}$ is a wavelet triplet, although not of our form, and $\max_{\boldsymbol{\psi}' \text{is a wavelet triplet}} \sum_{j=1}^{3} \langle f_j, \psi_j' \rangle = \max_{\boldsymbol{\psi}' \text{is a wavelet triplet}} \sum_{j=1}^{3} \langle f_j - \bar{f}_j, \psi_j' \rangle = \sum_{j=1}^{3} \langle f_j - \bar{f}_j, \psi_{0j} \rangle = \sqrt{\sum_{j=1}^{3} \|f_j - \bar{f}_j\|^2}$. This is an upper bound for $\sum_{j=1}^{3} \langle f_j, \psi_j \rangle$ for $\boldsymbol{\psi}$ of our form satisfying the conditions for a CWT, equations (1.1) and (1.2), so G_{fit} is a measure of the GoF (goodness of fit) of $\boldsymbol{\psi}$ to $\boldsymbol{f}$.

Using equation (3.24), we can write G_{fit} in terms of the transformed data z:

$$G_{\text{fit}}(w, \ell) = \frac{\sum_{j=1}^{3} z_j^{\mathrm{T}} [H^{-1} - H^{-1} B (B^{\mathrm{T}} H^{-1} B)^{-1} B^{\mathrm{T}} H^{-1}] z_j}{\sum_{j=1}^{3} \left\| f_j^{(w\ell)} - \bar{f}_j^{(w\ell)} \right\|^2}. \tag{3.42}$$

If $B = b$, we can use equation (3.27) and obtain the simpler expression

$$G_{\text{fit}}(w, \ell) = \frac{\sum_{j=1}^{3} [z_j^{\mathrm{T}} H^{-1} z_j - z_{j0}^2]}{\sum_{j=1}^{3} \left\| f_j^{(w\ell)} - \bar{f}_j^{(w\ell)} \right\|^2}. \tag{3.43}$$

instead.

We first wish to find the wavelet triplets which are a reasonable fit to the data. A secondary aim is to avoid selecting near-identical triplets.

To this end, for each w, we initialise the set of candidate indices $\mathcal{C}_w$ by

$$\begin{aligned}\mathcal{C}_w &= \{\ell : (\ell = w \text{ and } G_{\text{fit}}(w,\ell) \geq G_{\text{fit}}(w,\ell+1))\} \cup \\ &\quad \{\ell : (\ell = M \text{ and } G_{\text{fit}}(w,\ell) \geq G_{\text{fit}}(w,\ell-1))\} \cup \\ &\quad \{\ell : (w < \ell < M \text{ and } G_{\text{fit}}(w,\ell) \geq G_{\text{fit}}(w,\ell-1) \text{ and } G_{\text{fit}}(w,\ell) \geq G_{\text{fit}}(w,\ell+1))\},\end{aligned}$$

and $\mathcal{A}$ and $\mathcal{D}$ to $\emptyset$ where $\mathcal{A}$ is the set of accepted indices, and $\mathcal{D}$ the set of indices to be deleted.

Let $r_w(\ell)$ be the rank of $G_{\text{fit}}(w,\ell)$ when the $G_{\text{fit}}(w,\ell')$ are sorted in decreasing order[9].

Then, until $\mathcal{C}_w - \mathcal{D} = \emptyset$, find $\ell_0 = \arg\min_{\ell \in \mathcal{C}_w - \mathcal{D}}\{r_w(\ell)\}$ and set $\mathcal{A} := \mathcal{A} \cup \{\ell_0\}$ and $\mathcal{D} := \mathcal{D} \cup \{\ell : \ell_0 - h_{\text{peak}} \leq \ell \leq \ell_0 + h_{\text{peak}}\}$.

When this process has been completed, we set $\mathcal{C}_w := \mathcal{A}$. In this way, for each w we have selected the highest peaks in $G_{\text{fit}}(w,\ell)$ that are separated by at least h_{peak} steps in ℓ, and thus avoided choosing near identical wavelet triplets derived from data in neighbouring windows.

On the completion of similar processes for all w, set $\mathcal{C} = \bigcup_{w \in \mathcal{S}_w}\{(w,\ell) : \ell \in \mathcal{C}_w\}$, and $\mathcal{L}_{\text{super}} = \{\boldsymbol{\psi}^{(w\ell)} : (w,\ell) \in \mathcal{C}\}$. We then order and relabel $\mathcal{L}_{\text{super}}$ according to decreasing $G_{\text{fit}}(\boldsymbol{\psi}^{(w\ell)})$, so $G_{\text{fit}}(\boldsymbol{\psi}_1) = G_{\text{fit}}(\boldsymbol{\psi}^{(w_1\ell_1)}) \geq G_{\text{fit}}(\boldsymbol{\psi}_2) = G_{\text{fit}}(\boldsymbol{\psi}^{(w_2\ell_2)}) \geq \ldots$, where each $(w_k,\ell_k) \in \mathcal{C}$. The final step in constructing $\mathcal{L}_{\text{super}}$ is to discard any $\boldsymbol{\psi}_k$ such that $G_{\text{fit}}(\boldsymbol{\psi}_k) < G_{\text{fit}}(\boldsymbol{\psi}_{N_{\text{super}}})$, leaving $\mathcal{L}_{\text{super}} = \{\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_{N'_{\text{super}}}\}$. It is not necessarily the case that $N'_{\text{super}} = N_{\text{super}}$ (if, before this discard stage, $\overline{\overline{\mathcal{L}_{\text{super}}}} < N_{\text{super}}$, then $N'_{\text{super}} < N_{\text{super}}$ and, if $G_{\text{fit}}(\boldsymbol{\psi}_{N_{\text{super}}+1}) = G_{\text{fit}}(\boldsymbol{\psi}_{N_{\text{super}}})$, then $N'_{\text{super}} > N_{\text{super}}$), but, if the two quantities differ greatly, there is either insufficient data to build the full $\mathcal{L}_{\text{super}}$ or there are a pathological number of ties between values of $G_{\text{fit}}(\boldsymbol{\psi}_k)$.

In practice, because of memory considerations, we divide the input into epochs of a length which is much greater than our window lengths, but much shorter than the entire data series. For each window length, w, we process each epoch, adding the resultant triplets to $\mathcal{L}_{\text{super}}$, until it is of size N_{super}. Thereafter, each new $\boldsymbol{\psi}$ replaces $\boldsymbol{\psi}' = \arg\min_{\boldsymbol{\psi}'' \in \mathcal{L}_{\text{super}}} G_{\text{fit}}(\boldsymbol{\psi}'')$ in $\mathcal{L}_{\text{super}}$ if $G_{\text{fit}}(\boldsymbol{\psi}) > \min_{\boldsymbol{\psi}'' \in \mathcal{L}_{\text{super}}} G_{\text{fit}}(\boldsymbol{\psi}'')$.

In its simplest form, this division of the data into epochs leads to some loss of data, as if a data series contains L points, there are $L - w + 1$ windows of length w into it, so, if L_{epoch} is the length of an epoch and N_{epoch} is the number epochs in the data (assumed to be an integer), then the number of windows into all the epochs is $N_{\text{epoch}}(L_{\text{epoch}} - w + 1) = L - N_{\text{epoch}}(w-1) < L - w + 1$ if $N_{\text{epoch}} > 1$. Although this loss can be circumvented by carrying information from one epoch to its successor, we estimate that the data loss will only rarely be significant (at least if $L_{\text{epoch}} \gg w$), and do not do this.

There is also a tiny bias in this procedure towards $\boldsymbol{\psi}$ derived from the epochs and window lengths that are first processed, as they may find themselves in $\mathcal{L}_{\text{super}}$ before others with the same value of G_{fit}, which will not be able to displace them, but this bias is not likely to realised in practice.

[9]We accept the order returned by whichever sorting algorithm we choose when there are ties

Note that in the determination of the members of $\mathcal{L}_{\text{super}}$ we have not needed to calculate the component wavelets of the member triplets.

3.3.5 Reducing $\mathcal{L}_{\text{super}}$ to $\mathcal{L}$

We now wish to further prune $\mathcal{L}_{\text{super}}$ to produce our final library $\mathcal{L}$ while ensuring that the ECWTs in $\mathcal{L}$ are reasonably distinct from one another[10], and also representative of $\mathcal{L}_{\text{super}}$. One way of doing this is to use a clustering algorithm on $\mathcal{L}_{\text{super}}$, and select the cluster centres as $\mathcal{L}$, as the points close to the centres will be in their cluster and will be automatically excluded from $\mathcal{L}$. As this selection depends on the distances between members of $\mathcal{L}_{\text{super}}$, and not their Goodness of Fit, we call this stage of the process *selection on distance*.

We do not expect that the results of our method will depend critically on the goodness of the clustering here, so we regard it as permissible to set the number of centres in advance to the desired size of $\mathcal{L}$.

Of course, to be consistent with our implicit definition of similarity, we will need to cluster based on the distance given by equation (3.33).

But, to use equation (3.33), we need to explicitly calculate the wavelet triplets in $\mathcal{L}_{\text{super}}$ which are fitted to the data, as in the previous section, 3.3.4. In order to exploit the advantages listed in section 3.3.3, we then transform this wavelet triplet to the canonical representative of its equivalence class.

We would now like to use the k-medoids algorithm (Jin and Han[41]) on these equivalence classes, as this algorithm produces cluster centres which lie in the set being clustered, in line with our aim of producing a *subset* of $\mathcal{L}_{\text{super}}$ to use as $\mathcal{L}$.

Given a set of vectors $\mathcal{V}$, a distance measure d between them and the integer k, this attempts to find the subset $\mathcal{K}$ of $\mathcal{V}$ of cardinality $\overline{\overline{\mathcal{K}}} = k$ such that $\sum_{w\in\mathcal{K}}\sum_{v\in\mathcal{V}_w} d(v,w)$ is minimised, where $\mathcal{V}_w = \mathcal{V}_{w<} \cup \mathcal{V}_{w=}$, $\mathcal{V}_{w<} = \{v \in \mathcal{V} : d(v,w) < d(v,w')\forall w' \in \mathcal{K} - \{w\}\}$, $\mathcal{V}_{w=} = \{v \in \mathcal{V} : d(v,w) \leq d(v,w'), d_t(v,w) < d_t(v,w')\forall w' \in \mathcal{K} - \{w\}\} - \mathcal{V}_{w<}$ for some algorithm-dependent tie-breaking function d_t.

The input to the k-medoids algorithm would be the canonical representatives of the members of $\mathcal{L}_{\text{super}}$, the distance measure would be D^* (of equation (3.41)), and the k parameter would be be N_{lib}. The output would be the indices of $\mathcal{L}$ in $\mathcal{L}_{\text{super}}$, i.e., we would use the medoids as $\mathcal{L}$.

If $\overline{\overline{\mathcal{V}}}$ were reasonably small, an exhaustive search for $\mathcal{K}$ could be carried out, but, with sets of larger cardinality, algorithms, such as those supplied in Matlab®'s Statistics and Machine Learning Toolbox, which return a possibly sub-optimal $\mathcal{K}$, will need to be employed.

However, when using these approximate algorithms, N_{super} and N_{lib} are too large in combination for k-medoids to finish in a reasonable time, so we have had to employ a "divide and rule" strategy — we initially divide $\mathcal{L}_{\text{super}}$ into much smaller sets of neighbouring ECWTs, apply k-medoids to each of these sets in turn, and then unite the resulting medoids into a set

[10]The hope is that the movement patterns are also distinct, but this needs to be one of the things which are evaluated at a later stage

to which we apply a variant of k-medoids, so obtaining medoids of medoids. In other words, we achieve an hierarchical clustering.

The reason we use a *variant* of k-medoids for the higher-level clustering is that we need to take the cluster sizes after the lower-level clustering into account, in order to ensure that the higher-level centres reflect the lower-level clustering. For example, we do not wish to choose low-level medoids whose clusters are both small and isolated as high-level ones.

For this purpose, we use a tiling technique, similar to that employed in the SIVIA (*S*et *I*nversion *V*ia *I*nterval *A*nalysis) algorithm (see, e.g. Jaulin and Walter[42]).

Given a set of points $\mathcal{P} \subset \mathbb{R}^{n_p}$ such that $\overline{\overline{P}} = N_p$, and a target length $N_{\text{tar}} = \texttt{target}$ the algorithm is, in pseudo-code,

1 define a one-element list, `Tiles = [Tiles(1)]` containing a `tile` structure with fields

   `Min` consisting of a vector in $\mathbb{R}^{n_p}$ such that `Min`(i) $= \min_{x \in \mathcal{P}} x_i$;

   `Max` consisting of a vector in $\mathbb{R}^{n_p}$ such that `Max`(i) $= \max_{x \in \mathcal{P}} x_i$;

   `members` consisting of a vector in $\mathbb{N}^{N_p}$ with elements $1, 2, 3, \ldots, N_p$, representing the members of $\mathcal{P}$;

   `dims` consisting of a vector in $\mathbb{N}^{n_p}$ with elements $1, 2, 3, \ldots, n_p$, representing the dimensions of $\mathbb{R}^{n_p}$.

2 while `max(length(Tiles(1).members), length(Tiles(2).members),...)) > target`

   (a) find the first index, `i`, such that `length(Tiles(i).members) = max(length(Tiles(1).members), length(Tiles(2).members),...))`;

   (b) set `Tiles = [Tiles(1), Tiles(2),..., Tiles(i-1), Tiles(i+1),..., Tiles(length(Tiles)), split(Tiles(i))]`.

Here `split` is the function `newTiles = split(Tile)` which does the following:

1 for each `i` find `div(i) = (Tile.Min(i) + Tile.Max(i))/2`;

2 for each `i` find $\overline{\overline{\{x \in \mathcal{P} : \texttt{Tile.Min(i)} \leq x_{\texttt{i}} < \texttt{div(i)}\}}}$ and set `lMems(i)` to this;

3 for each `i` find $\overline{\overline{\{x \in \mathcal{P} : \texttt{div(i)} \leq x_{\texttt{i}} < \texttt{Tile.Max(i)}\}}}$ and set `uMems(i)` to this;

4 find the first `i0` such that `abs(uMems(Tile.dims(i0))-lMems(Tile.dims(i0))) = min(abs(uMems-lMems))`;

5 (a) if $\{x \in \mathcal{P} : \texttt{Tile.Min(Tile.dims(i0))} \leq x_{\texttt{i}} < \texttt{div(Tile.dims(i0)}\} = \emptyset$ then

      i. set

```
                    Tile1.Min  :=  Tile.Min,
                    Tile1.Max  :=  Tile.Max,
        Tile1.Min(Tile.dims(i0))  :=  div(Tile.dims(i0)),
                 Tile1.members  :=  Tile.members,
                    Tile1.dims  :=  [Tile.dims(1), Tile.dims(2),...,
                                      Tile.dims(i0-1),...
                                      Tile.dims(i0+1),...,
                                      Tile.dims(i0)];
```

ii. set `newTiles = [Tile1]`;

(b) else if $\{x \in \mathcal{P} : \texttt{div(Tile.dims(i0))} \leq x_{\texttt{i}} < \texttt{Tile.Max(Tile.dims(i0)}\} = \emptyset$

i. set

```
                    Tile1.Min  :=  Tile.Min,
                    Tile1.Max  :=  Tile.Max,
        Tile1.Max(Tile.dims(i0))  :=  div(Tile.dims(i0)),
                 Tile1.members  :=  Tile.members,
                    Tile1.dims  :=  [Tile.dims(1), Tile.dims(2),...,
                                      Tile.dims(i0-1),...
                                      Tile.dims(i0+1),...,
                                      Tile.dims(i0)];
```

ii. set `newTiles = [Tile1]`;

(c) else

i. set

```
                    Tile1.Min  :=  Tile.Min,
                    Tile1.Max  :=  Tile.Max,
        Tile1.Max(Tile.dims(i0))  :=  div(Tile.dims(i0)),
                 Tile1.members  :=  Tile.members,
                    Tile1.dims  :=  [Tile.dims(1), Tile.dims(2),...,
                                      Tile.dims(i0-1),...
                                      Tile.dims(i0+1),...,
                                      Tile.dims(i0)];
```

Also, set `Tile1.members` to the indices in `Tile.members` of $\{x \in \mathcal{P} : \texttt{Tile.Min(Tile.dims(i0))} \leq x_{\texttt{i}} < \texttt{div(Tile.dims(i0)}\}$;

ii. set

```
                    Tile2.Min  :=  Tile.Min,
                    Tile2.Max  :=  Tile.Max,
        Tile2.Min(Tile.dims(i0))  :=  div(Tile.dims(i0)),
                 Tile2.members  :=  Tile.members,
                    Tile2.dims  :=  [Tile.dims(1), Tile.dims(2),...,
                                      Tile.dims(i0-1),...
                                      Tile.dims(i0+1),...,
                                      Tile.dims(i0)];
```

Also set `Tile2.members` to the set difference of `Tile.members` and `Tile1.members`;

iii. set `newTiles = [Tile1, Tile2]`

(we have used Matlab®-like syntax in which, if `u` is a vector, `u(i)` is the ith component of the vector, and functions like `abs` and `min` operate over the components of any vector they are applied to).

Basically, this algorithm iteratively splits boxes ("tiles") containing points of a set into halves, choosing the (axis-parallel) hyperplane of division so that the two halves each contain as close to half the points as possible.

At each step, the box containing the most points is chosen for division. The purpose of the `dims` field is to (at least minimally) reduce the number of times that the divisions occur parallel to any particular hyperplane, as we suspect that thin slices of space will be undesirable.

The algorithm halts when the maximal number of points in a tile is sufficiently small, less than, say, $T_{\max}$. For our purposes, a number of the order of 150 appears to suffice[11].

An illustration of the tiling algorithm is given in Figure 3.4.

We now have obtained much smaller sets on which to use k-medoids with the distance D^*, the intersections of the set of ECWTs in $\mathcal{L}_{\text{super}}$ with the "boxes" resulting from the tiling algorithm, and we now apply that algorithm to these sets.

We first choose k_0 such that $k_0 n_t > N_{\text{lib}}$, where n_t is our final number of tiles (= number of subsets of $\mathcal{L}_{\text{super}}$ we are applying k-medoids to) by a factor > 1, and apply k-medoids with $k = k_0$ to each of our subsets. The resulting centres (bottom-level centres in our eventual two-level hierarchical clustering), will be placed into a set $\mathcal{L}'_{\text{super}}$, and a corresponding weighting will be assigned to each centre, equal to the number of equivalence classes in it[12].

When our algorithm is run with reasonable values of $T_{\max}$ and k_0, and calls `kmedoids` with Matlab®'s default parameters, it results in an execution time of a couple of hours over all of our subsets.

Our weighted version of k-medoids is simply Partition Around Medoids (see Jin and Han[41]) with the target for reduction, $\sum_{w\in\mathcal{K}}\sum_{v\in\mathcal{V}_w} d(v,w)$ replaced by $\sum_{w\in\mathcal{K}}\sum_{v\in\mathcal{V}_w} \mu_v d(v,w)$, where μ_v is the weight assigned to v. We implemented this with a similarly modified version of k-means++ (for improved initialisation; — for the original algorithm, see Arthur and Vassilvitskii[43]), and the only options we provided are the number of replicates (runs with independently selected random starting centres — obviously not *uniformly* random — that's the point of k-means++!) and the maximum number of iterations for each replicate.

Both modified algorithms are equivalent to the originals run with a number of duplicates for each point equal to its weight here, except that the computational complexity depends

[11] Although this tiling technique seems more compatible with Euclidean distances than the distance D^* of equation (3.41), we believe the subsequent use of variants of k-medoids will overcome any problems caused by this

[12] If any subset has cardinality less than k_0, we simply place all its points into $\mathcal{L}_{\text{super}}$, each with weighting 1

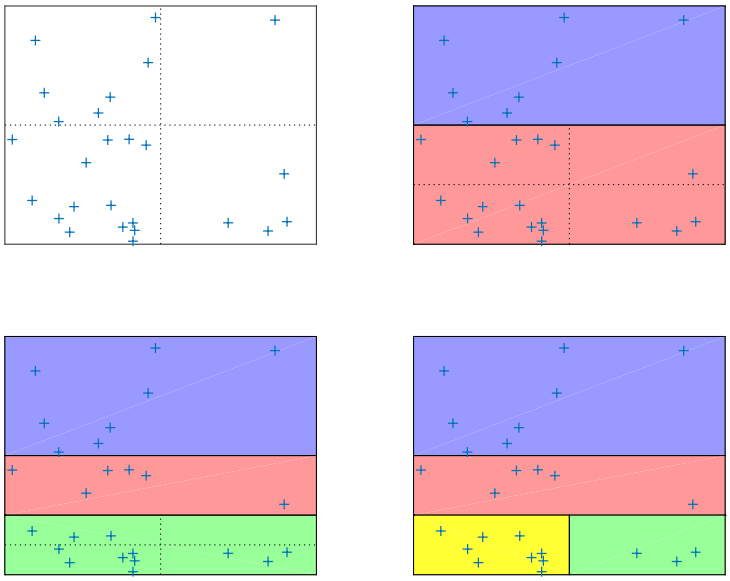


Figure 3.4: *Tiling algorithm: top left: original space to be split up, with possible split planes; top right: initial tiling, chosen for the most numerically even division. The posssible split planes are shown for the tile with the most members; bottom left: new tiling, chosen for the most even split. The possible split planes are shown for the tile with the most members; bottom right: final tiling. Although the possible split planes would have resulted in equally even divisions of the split tile, one was chosen on the basis of being orthogonal to the previous split plane*

on the number of unique points rather than the sum of their weights, which is much greater.

It seems plausible that the final use of the modified algorithm will frequently reunite clusters rent asunder by the initial tiling.

3.3.6 Example[13]: 1) fitting the triplets

As an example of this procedure, we carry it out on some data in the nominal ±2g range (so actually in $\left(-2\frac{5}{18}, 2\frac{5}{18}\right)$g), sampled at 500Hz, and stemming from a NAT attached to the right wrist of a PD subject, PD1, going about their normal daily activity for 6 hours (so the data consists of $6 \times 60 \times 60 \times 500 = 1.08 \times 10^7$ triplets of accelerations). The particular 6 hours was chosen to avoid saturation (readings equal to $\pm 2\frac{5}{18}$g). We used window lengths $76, 80, 84, \ldots, 140, 144$ (corresponding to $6.58, 6.25, 5.95, \ldots, 3.57, 3.47$Hz, and therefore including the ∼4-6Hz region which Fourier analysis has been shown to be important for this PD subject). We use wavelets without any further conditions beyond ue and zi, and set n to 5, N_{lib} to 100, and N_{super} to 10^4. Finally, we choose an h_{peak} of 10, roughly a tenth of the width of the windows, and an epoch length of ten minutes or 3×10^5 samples.

As the windows into the data move along the input, the components of z may be obtained by using the components of R (of equation 3.15) as a filter. In Figure 3.5, we show the filters for each component of z when the window length is 108. The filters for z_2 and z_4 are fairly similar in shape, and this similarity between those for z_3 and z_5 is even greater.

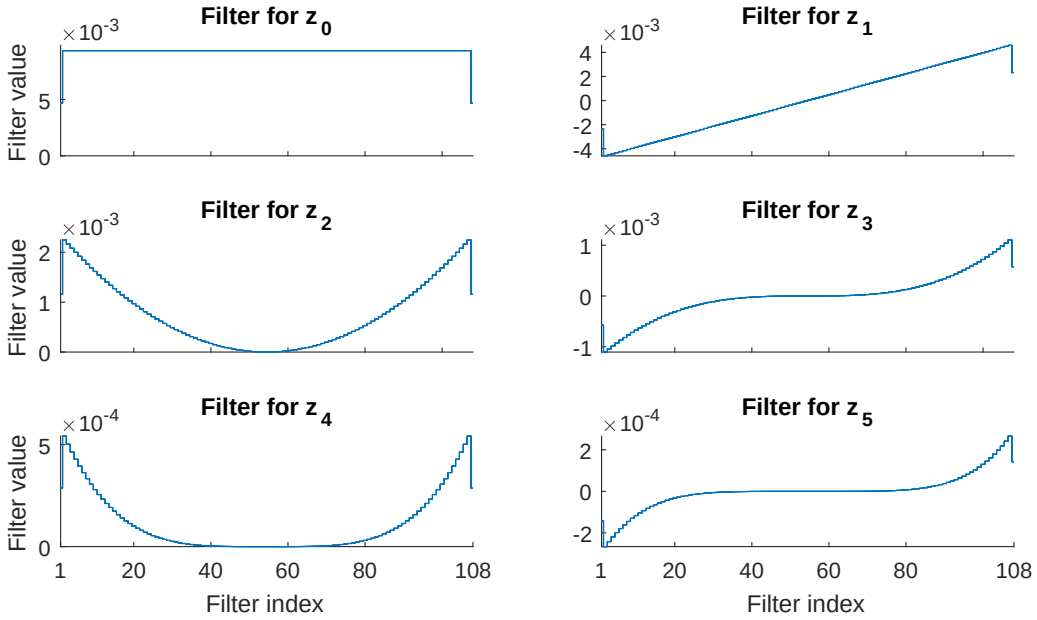


Figure 3.5: *Filter components, for $n = 5$, window length 108*

Figure 3.6 depicts sample input traces for the three acceleration channels (this sample is actually near where the best G_{fit} was obtained for windows of length 108, but we are getting ahead of ourselves here).

[13]The example has been developed in parallel with the software used to generate it. There may be minor discrepancies in the illustrative data as a result of this, but the purpose of the example is to illustrate, not to prove anything about the efficacy of the procedures, so these discrepancies are of no major importance. Nevertheless, it is only pressure of time that has prevented the reworking of the example from the start

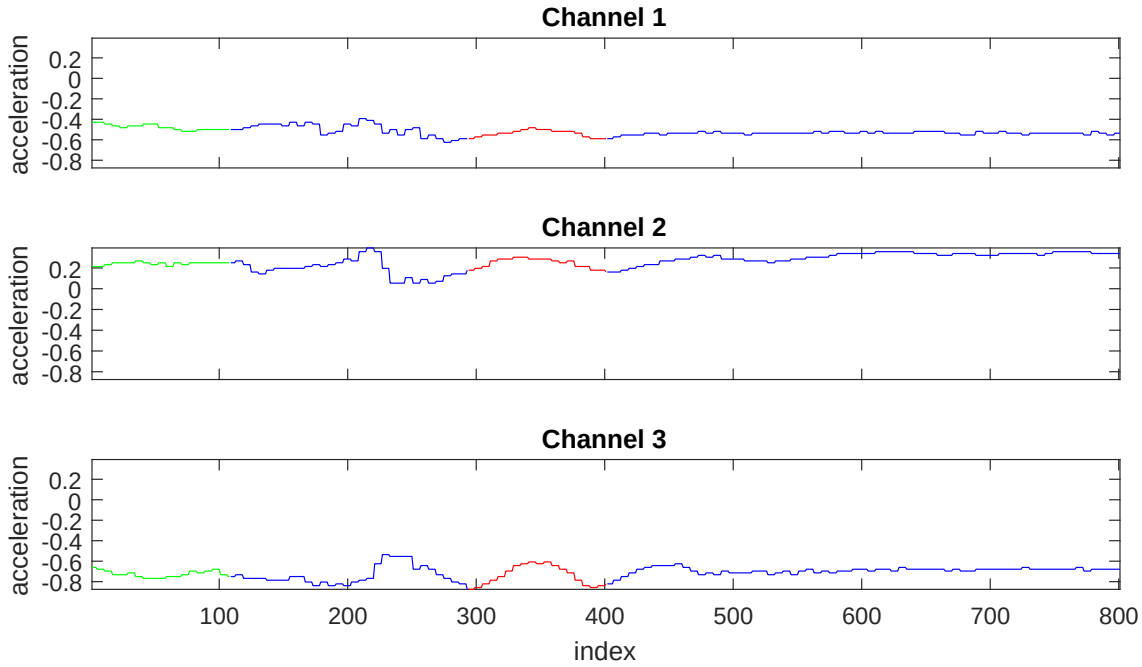


Figure 3.6: *Traces of the input near where the best G_{fit} for window length 108 was obtained. The green part of the traces contributes to the first values of z_j in Figure 3.7, the red part contributes to the values of z_j involved in the best G_{fit}*

Now we can derive our zs, by applying the filters of Figure 3.5 to the data of Figure 3.6. When we do so, we obtain Figure 3.7. Note that we drop the first 107 points of these traces, as they are derived from incomplete windows whose start point precedes that of the data; in other words, they are derived from a mixture of padding and actual input.

We can immediately see that the curves for z_2 and z_4 have a similar shape, reflecting the similarity in shape of the filters for these quantities, and that the same is true of the z_3 and z_5 curves, with the same explanation.

With the zs at our disposal, we can now find $G_{\text{fit}}(108, \ell)$, using equation (3.43). This is shown in Figure 3.8. The largest peak is actually the greatest value achieved by G_{fit} for window length 108 within our 6 hours of data. (For illustrative purposes, we have used $h_{\text{peak}} = 27$, a quarter of the window length, in the Figure, rather than $h_{\text{peak}} = 10$, which we have used in the example itself. The trace of G_{fit} is what was derived in the example, however.)

All of this can be done without actually calculating the wavelet triplet $\boldsymbol{\psi}$, but here, in Figure 3.9, we show the as for the wavelet triplet of best fit, and for the wavelet triplets derived from neighbouring windows, obtained *via* equation (3.25).

We note that the magnitude of the values of a_{3j} (gold) are large for even j at the triplet of best fit, compared to the other as at this point, and the as at most other points.

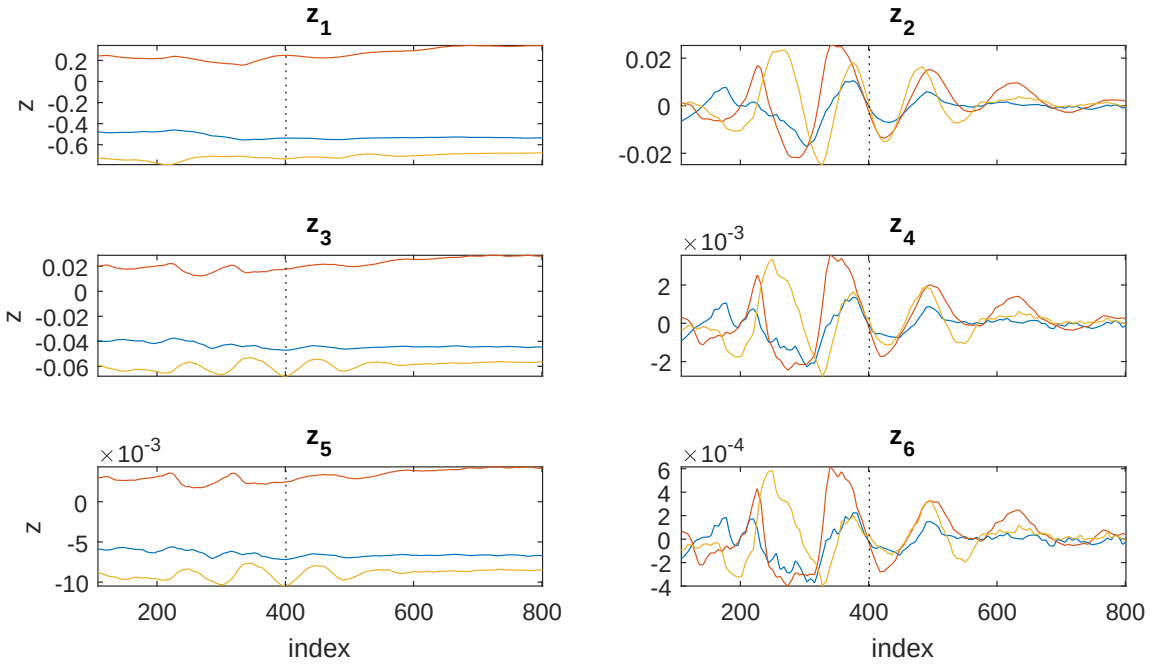


Figure 3.7: z_j, $j = 0$-5. *The zs for Channel 1 are drawn in blue, those for Channel 2 in red, and the gold traces relate to Channel 3. The dotted vertical line shows the location of the best value of G_{fit} for window length 108*

We can now plot the wavelet triplets themselves, and we have done this for the best-fitting wavelet triplet and its eight neighbours in time in Figure 3.10. In this Figure, we have vertically scaled the triplets to bring them closer to the data, which is legitimate, as our interest is in the wavelet content in the data, and such scaling is consistent with the approach. The vertical shifting for the same purpose is also permissible, as the mean of the data plays no role in the analysis of its shape.

It can be seen that the third component of the best fitting triplet (shown in magenta) is larger than the other components and we believe that it is visibly dominated by the component which is even about its midpoint, which is consistent with what has just been said regarding the values of the as at the best-fitted wavelet triplet.

Figure 3.10 confirms that triplets generated by data in neighbouring windows can be very similar, and supports our decision to drop triplets at maxima in the G_{fit} trace which are in the neighbourhood of larger maxima.

The similarity between the equivalence classes represented by the triplets of Figure 3.10 may be even greater, as the freedom to rotate the triplets will almost certainly remove some of their difference, although this is not particularly apparent in Figure 3.11, which displays the canonical representative triplets of the equivalence classes. Despite the fact that these representatives are now horizontally aligned, are vertically centred and have the same vertical scale, the curves are not appreciably closer together.

One thing is revealed by the use of the canonical representation is that most of the

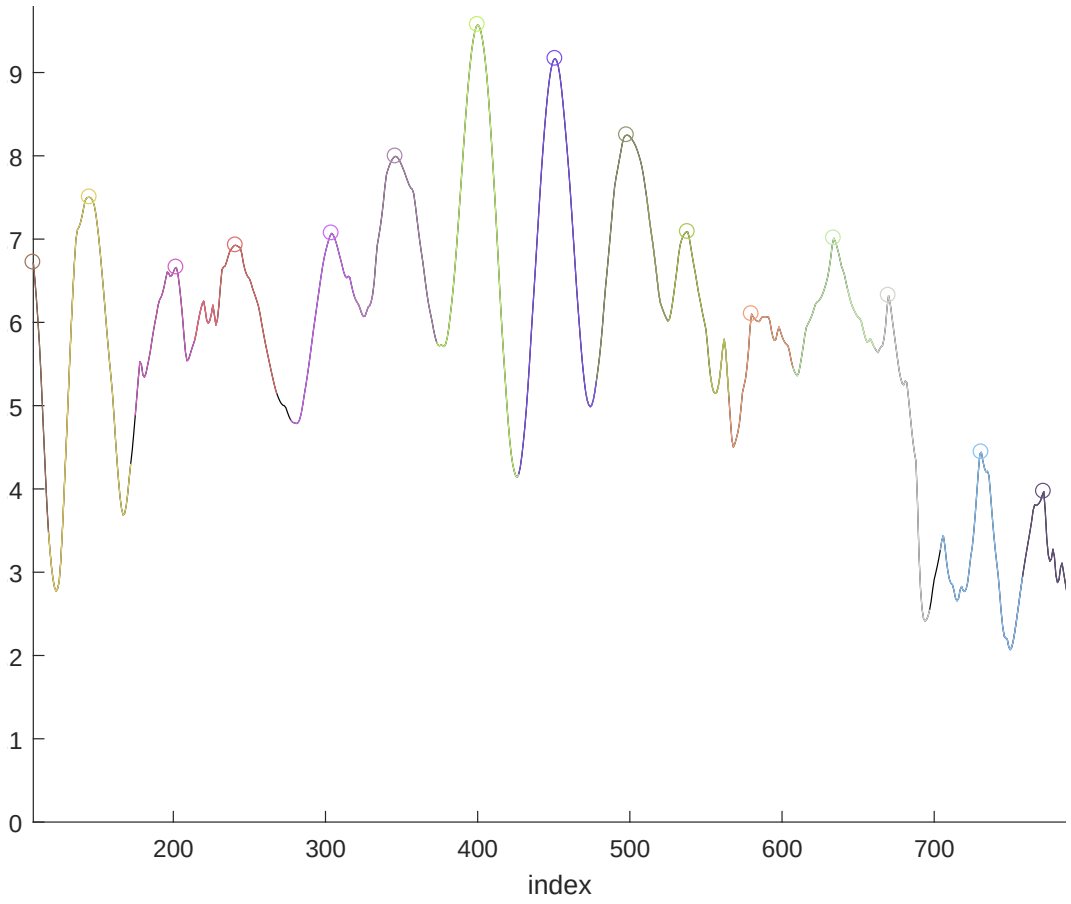


Figure 3.8: $G_{fit}(108, \ell)$, *near the the maximum value for window length 108,* $G_{fit}(108, 401) = 0.9510$. *Peaks which are not too near higher peaks (within a quarter of the window length 108 of them in this case) are ringed, and the curve in the region they "dominate" is in the same colour as the ring (the small "undominated" region is in black). The ringed peaks correspond to wavelets which are candidates for the initial library,* $\mathcal{L}_{super}$

acceleration, apparently especially at the best-fitted wavelet triplet (magenta), is in a plane, as $\|\psi_3'\|^2 = 0.0022$, compared to $\|\psi_1'\|^2 = 0.6985$ and $\|\psi_2'\|^2 = 0.2993$, for the best-fitting triplet.

For the neighbouring triplets, these numbers obey $0.0017 \leq \|\psi_3'\|^2 \leq 0.0081$, compared to $0.6583 \leq \|\psi_1'\|^2 \leq 0.7159$ and $0.2818 \leq \|\psi_2'\|^2 \leq 0.3336$ (so there are cyan curves in the lower diagram that are even flatter than the magenta one, so the motion is even more dominated by its components in a particular plane).

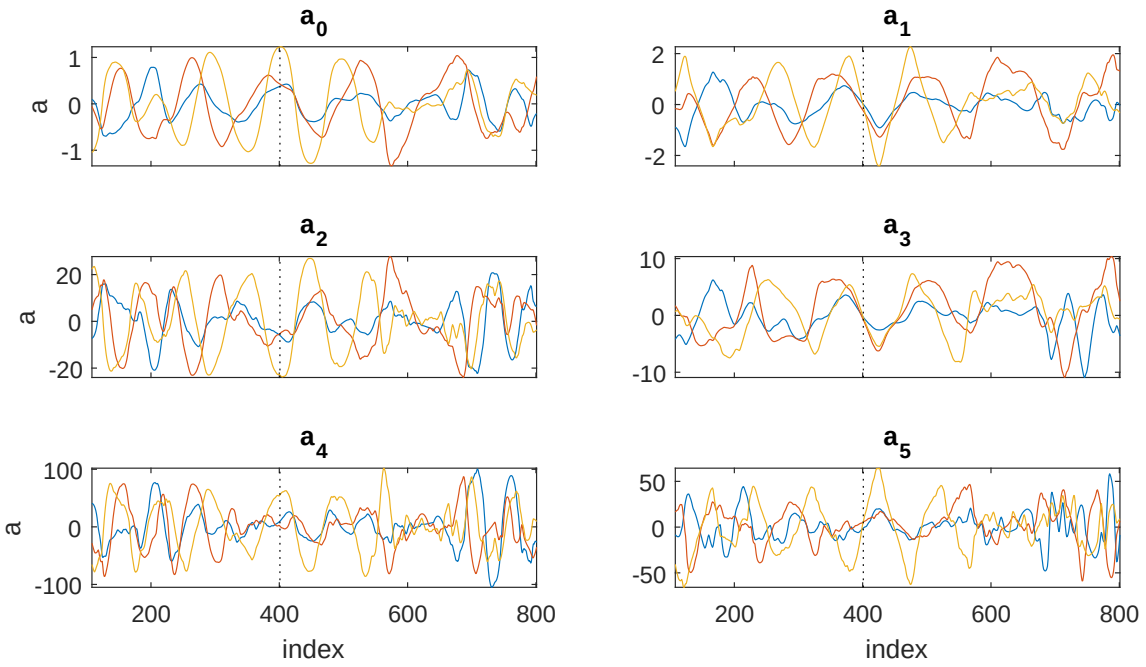


Figure 3.9: $a_{.j}$, $j = 0$-5. *The as for Channel 1 are drawn in blue, those for Channel 2 in red, and the gold traces relate to Channel 3. The dotted vertical line shows the location of the best value of* G_{fit} *for window length 108*

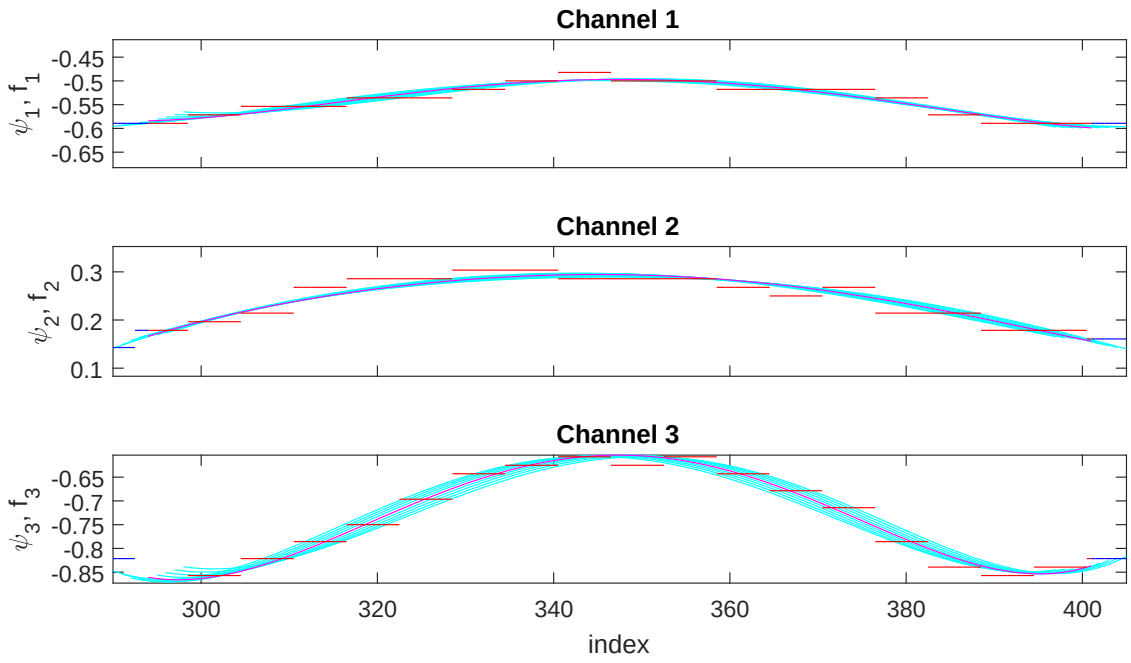


Figure 3.10: *The best-fitted (child) wavelet triplet for window length 108. The data supporting this triplet is represented by horizontal red lines, and the triplet itself is shown in magenta. Also shown are neighbouring triplets (cyan), and the data (blue) which supports these, but not the best-fitted triplet. Note that the triplets (not the individual wavelets) have been scaled and the individual wavelets have been shifted vertically to obtain the best visual match to the data, so zi and ue are apparently not obeyed. They are also aligned horizontally to the data which supports them. The vertical axes are to the same scale for each channel*

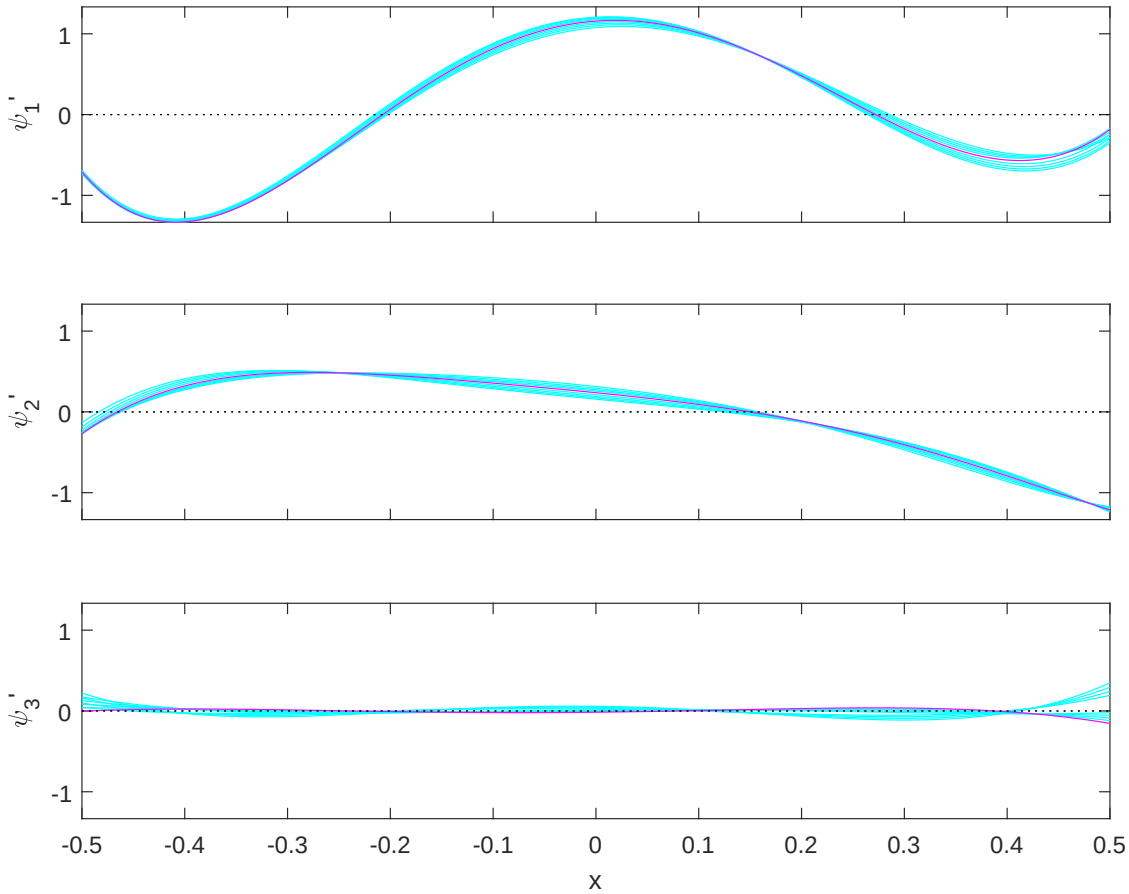


Figure 3.11: *The canonical representative wavelet triplets for the equivalence classes containing the wavelet triplets of Figure 3.10*

3.3.7 Example: 2) Building $\mathcal{L}_{\text{super}}$

Now that we have $G_{\text{fit}}(\boldsymbol{\psi})$ for our candidate members of $\mathcal{L}_{\text{super}}$, we can slim down the candidate list by choosing the local maxima of that quantity, and then removing such maxima dominated by neighbouring ones, as described above and depicted in Figure 3.8. Then we select from the remaining candidates by thresholding the actual value of G_{fit} to form the membership list of $\mathcal{L}_{\text{super}}$.

The next step is to actually calculate the members of $\mathcal{L}_{\text{super}}$, as the $\boldsymbol{\psi}$ will be needed to determine $\mathcal{L} \subset \mathcal{L}_{\text{super}}$.

The results of this process, $\mathcal{L}$, will then have some of its characteristics compared to those of $\mathcal{L}_{\text{super}}$ later.

3.3.8 Example: 3) Selecting $\mathcal{L}$

We have three stages in our reduction of $\mathcal{L}_{\text{super}}$ to $\mathcal{L}$:

1 the tiling of $\mathcal{L}_{\text{super}}$ to yield tiles each containing at most N_{tar} members of $\mathcal{L}_{\text{super}}$. We call the intersection of the sth tile and $\mathcal{L}_{\text{super}}$, $\mathcal{T}_s$. Of course, $\mathcal{L}_{\text{super}} = \bigcup_s \mathcal{T}_s$ and $\mathcal{T}_s \cap \mathcal{T}_r = \emptyset, r \neq s$;

2 the use of k-medoids on each of the $\mathcal{T}_s$ of sufficiently large cardinality, to produce (at most) k_0 low-level centres for each subset. We call these centres $\mu_{\text{low},q}$, and together they form $\mathcal{L}'_{\text{super}} \subset \mathcal{L}_{\text{super}}$;

3 and the use of a variant, weighted, k-medoid algorithm to produce the N_{lib} higher-level centres, called $\mu_{\text{high},i}$, which constitute $\mathcal{L} \subset \mathcal{L}'_{\text{super}} \subset \mathcal{L}_{\text{super}}$.

We select $N_{\text{tar}} = 150$, $k_0 = 3$, the number of replicates for the variant version of k-medoids to be 5, and the maximum number of iterations per replicate to be 10.

The tiling phase produces 96 tiles, none of which contain fewer than four members, so we apply Matlab®'s default k-medoids algorithm to all 96 of these subsets. This produces 288 lower-level centres for the weighted k-medoids algorithm to work on to produce $N_{\text{lib}} = 100$ clusters of lower-level centres.

Let $\mathcal{C}_{\text{low},q}$ be the cluster of points of $\mathcal{L}_{\text{super}}$ associated with $\mu_{\text{low},q}$ by the application of k-medoids in stage 2 above, and let $\mathcal{C}'_{\text{high},i}$ be the cluster of the $\mu_{\text{low},q}$s associated with $\mu_{\text{high},i}$ by the weighted k-medoids algorithm in stage 3. Then we can define the higher-level clusters of points by $\mathcal{C}_{\text{high},i} = \bigcup_{\{q:\mu_{\text{low},q} \in \mathcal{C}'_{\text{high},i}\}} \mathcal{C}_{\text{low},q}$.

In Figure 3.12 we present histograms of the cardinality of the $\mathcal{T}_s$s, the $\mathcal{C}_{\text{low},q}$s and the $\mathcal{C}_{\text{high},i}$s. The number of members in the $\mathcal{T}_s$s is in the range 54–150, with a peak at 130, with 5 of the 96 $\mathcal{T}_s$s having this many members. From a visual inspection, $\overline{\overline{\mathcal{T}_s}}$ has no obvious underlying approximate distribution. Its median value is 102.5, with first and third quartiles at 82.5 and 129.

Although k-means clustering as we apply it to a $\mathcal{T}_s$ only produces three centres, the clusters, $\mathcal{C}_{\text{low},q}$s, belonging to these centres do not have to be of equal cardinality, and the

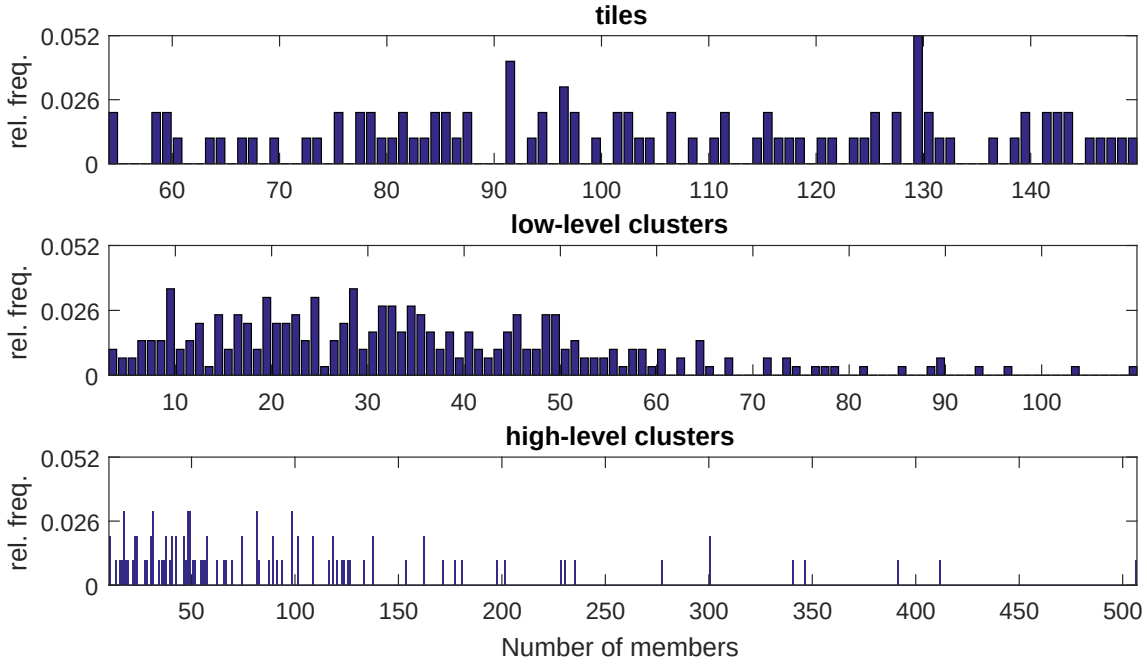


Figure 3.12: *Histograms of the cardinality of the various subsets of $\mathcal{L}_{super}$ (each bin in the histograms covers one value)*

central histogram of Figure 3.12 displays the distribution of these cardinalities. We can just about characterise this distribution as very roughly normal, with mean 34.72 and standard deviation 20.09, but with significant skewing (its skewness is 0.91). It is, of course, necessarily also truncated, as the minimum cardinality of a $\mathcal{C}_{\text{low},q}$ is 1, and its theoretical maximum (achieved when its containing $\mathcal{T}_s$ has its maximum theoretical cardinality of 150 and the clusters found in it include two one-member clusters) cardinality is 148. In this case, the maximum cardinality of $\mathcal{T}_s$ *is* 150, but the cardinality of $\mathcal{C}_{\text{low},q}$ is in the range 3-110.

The weighted version of k-means as applied to $\mathcal{L}'_{\text{super}}$ produces 100 high-level clusters as a by-product of the construction of $\mathcal{L}$. The empirical distribution of the cardinality of these high-level clusters, the $\mathcal{C}_{\text{high},i}$s, is shown in the bottom diagram in Figure 3.12. Again, this is very roughly normal, with mean 100 ($= \frac{N_{\text{super}}}{N_{\text{lib}}}$) and standard deviation 95.60, but with significant distortions (e.g., truncated to the range 10-507, skewness 1.098).

The final phase, using our weighted k-medoids algorithm, only takes at most four iterations in each of the five replicates we allowed, indicating that, in this case anyway, the low-level centres naturally fall into well-defined clusters (at least, in so far as "well-defined" is equivalent to "easily computable") at the higher level.

3.3.9 Example: 4) Characteristics of $\mathcal{L}$ and $\mathcal{L}_{\text{super}}$

In what follows, remember that $\mathcal{L}_{\text{super}}$ is selected only on a criterion derived from the GoF traces for each fixed window length, and $\mathcal{L}$ is selected from $\mathcal{L}_{\text{super}}$ only on the basis of

clustering.

In Figure 3.13, we show the regions in the trace from which the wavelets in $\mathcal{L}_{\text{super}}$ and $\mathcal{L}$ stem: the upper histogram refers to $\mathcal{L}_{\text{super}}$, the lower to $\mathcal{L}$. Interestingly, the upper diagram shows that the best-fitting wavelet triplets originate from four fairly well-defined regions in the trace (possibly where signal to noise is high?), around epochs 2, 17, 24 and 32, and the lower diagram shows that selection on distance only has a limited effect on these concentrations. The empirical distribution of the originating epoch is more concentrated around epochs 2, 26 and 32 for $\mathcal{L}$ (with most of the extra concentration coming from the epochs away from 2, 17, 24 and 32, plus a little from the epochs around 17). A possible explanation is that the ECWTs stemming from the epochs away from 2, 24 and 32 tend to be away from the centre of real clusters in the data, whereas those near 2, 24 and 32 are more centrally located in these real clusters, and tend to be selected as centres and thus make it into $\mathcal{L}_{\text{super}}$, and the other changes in the histograms have similar causes.[14].

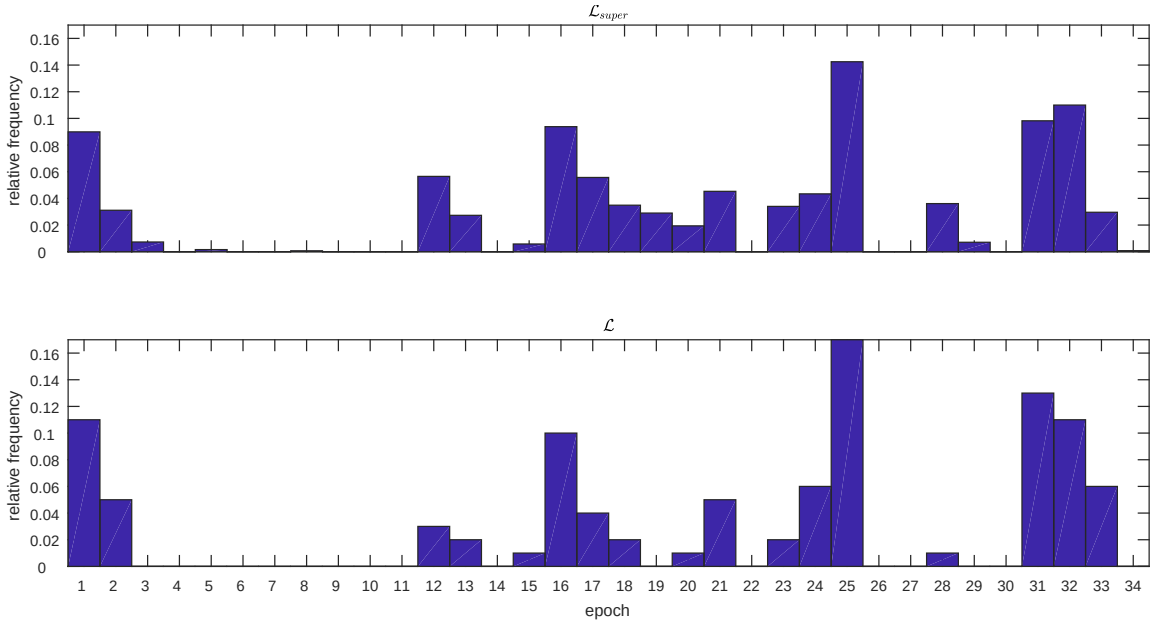


Figure 3.13: *Histogram of the window positions in the trace giving rise to the triplets in the library. Top: before selection on distance; bottom: after selection on distance*

In Figure 3.14, we show the window lengths giving rise to our triplets at the two stages of the process. The empirical selection probability for $\mathcal{L}_{\text{super}}$ (upper diagram) is roughly constant with respect to increasing window length over the first few lengths, and around a half of the triplets are associated with window lengths 76–98. The selection probability starts a steady decline from length 86. Except for minor details, the shorter the window length, the more likely it is to give rise to a triplet (and, in more detail, the selection probability[15] declines

[14]There is no bias towards any part of the interior — i.e., (one window length - 1) samples away from the start — of the trace in the algorithm

[15]The fact that there are marginally more ECWTs to be selected from for shorter window lengths is very likely of minor importance

roughly linearly with window length, after the initial roughly flat part of the histogram). This is probably due to the fitting process being easier if there are fewer values in the data to be fitted.

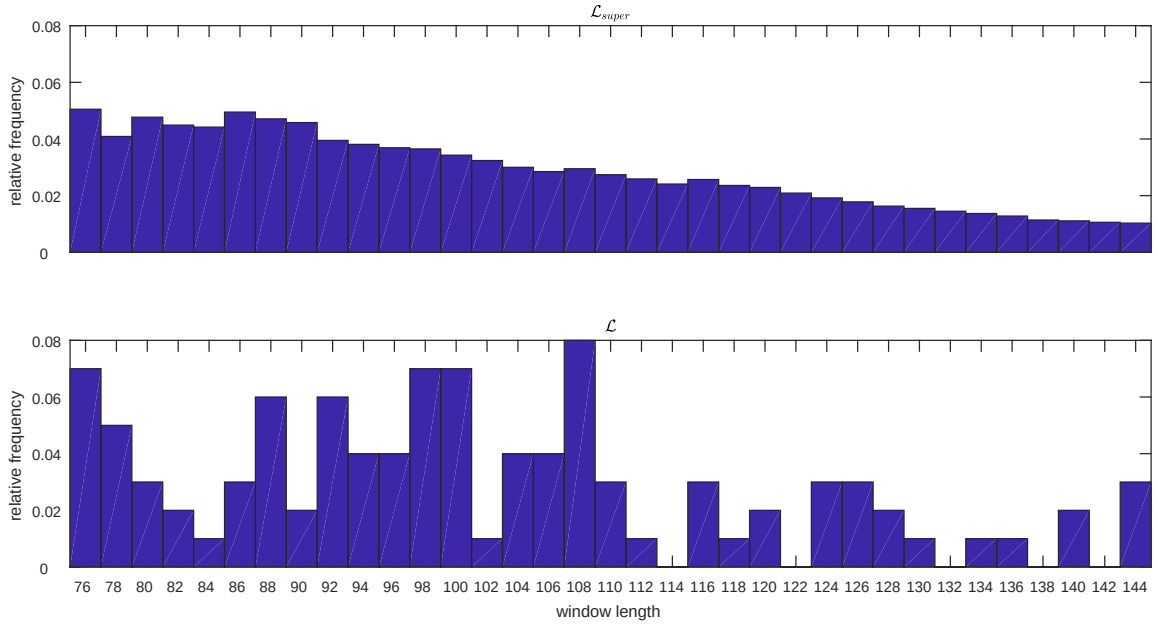


Figure 3.14: *Histograms of window lengths*

In $\mathcal{L}$, this pattern of a short rise, followed by a steady decline in selection probability with window length is apparently broken. In this particular data set, at least, there is peak for the smallest windows, then a sharp drop, followed by the ECWTs with windows in the second quarter of the range mostly having boosted empirical selection probabilities, so ECWTs generated on windows of length around 92-108 are more likely to be in the centre of high-level clusters. However, we cannot rule out the possibility that the boosting of these selection probabilities at the expense of the others is merely an artefact of a particular uniformly random selection of $\mathcal{L}$ from $\mathcal{L}_{\text{super}}$.

Figure 3.15 displays the GoF for the various triplets in the library. For $\mathcal{L}_{\text{super}}$, the frequency of the values of GoF decreases almost linearly with the value, becoming quite small for the rightmost bin, 0.9786-0.9806, although this does contain 0.6% of the ECWTs of $\mathcal{L}_{\text{super}}$. But, for $\mathcal{L}$, the relative frequencies of the bins in the second and fourth quarters of the range seem to be augmented at the expense of the others, in comparison to the histogram for $\mathcal{L}_{\text{super}}$. The median GoF for $\mathcal{L}$ is 0.9546, almost identical to that for $\mathcal{L}_{\text{super}}$, 0.9536, and the interquartile ranges are 0.0152 and 0.0153, respectively, showing that the GoF for $\mathcal{L}$ is very marginally more highly concentrated around its median that it is for $\mathcal{L}_{\text{super}}$.

These results show that, in this case at least, discarding ECWTs on the edge of our clusters neither substantially improves GoF, nor damages it.

However, without further investigation, it would be mere speculation to assert that this happens in general.

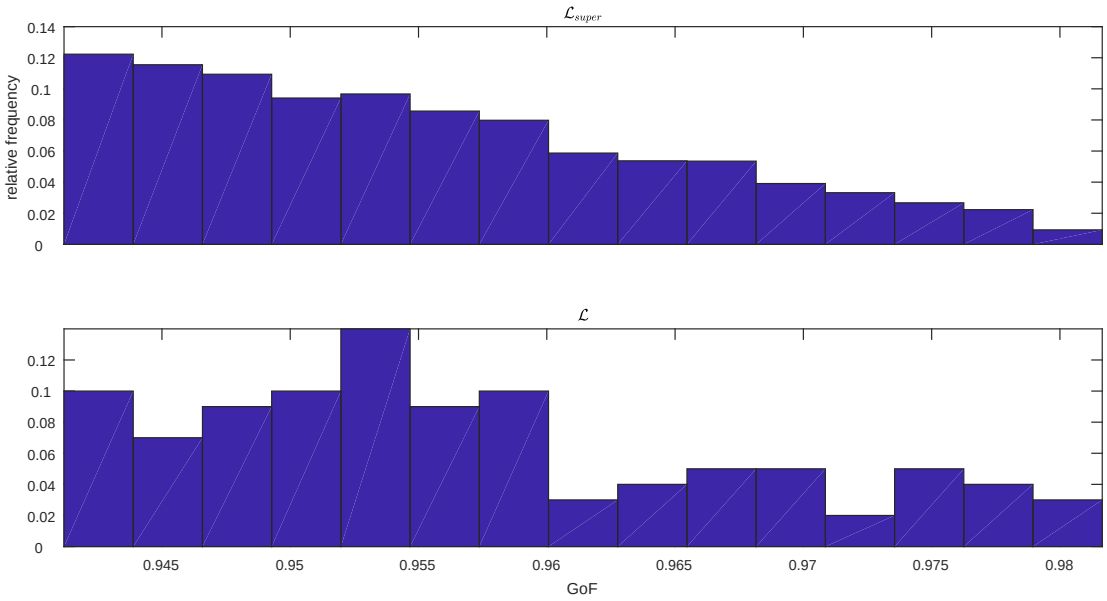


Figure 3.15: *Histograms of Goodness of Fit*

Figure 3.16 consists of scatter diagrams of the components of the canonical representatives of a sample of 1000 of the 10^5 member ECWTs of $\mathcal{L}_{\text{super}}$.[16] Because $\|\psi_1\|^2 \geq \|\psi_2\|^2 \geq \|\psi_3\|^2$ by definition, the points in Figure 3.16 have to lie on or below (left-hand two diagrams), or on or above (right-hand diagram) the black dotted lines, and because $\|\psi_1\|^2 + \|\psi_2\|^2 \leq 1$, they lie below the red dashed line in the leftmost diagram (and these restrictions on the possible placement of points will also be true of Figure 3.17).

As we maximise $\|\psi_1\|^2$ among the members of the equivalence class, and $\|\psi_2\|^2$ among the members for $\|\psi_1\|^2$ fixed at its maximum, and this suffices to fix $\|\psi_3\|^2$, if $\|\psi_3\|^2$ is close to 0, then the accelerations corresponding to $[\boldsymbol{\psi}]$ lie mostly in some plane, with respect to the axes of the NAT. Similarly, if both $\|\psi_2\|^2$ and $\|\psi_3\|^2$ are close to 0, the acceleration mostly lies along a line, again with respect to the NAT axes.

If the NAT rotates only a small amount over the duration of the relevant window into the data, then the acceleration will also mostly be in a plane (respectively, a line) with respect to co-ordinates in space.

We can see from Figure 3.16 that there are a lot of these "planar" equivalence classes in $\mathcal{L}_{\text{super}}$, and more than a negligible number of "linear" ones.

The analogue of Figure 3.16 for $\mathcal{L}$ is Figure 3.17, and we can see that the members of $\mathcal{L}$ appear to be representative of those of $\mathcal{L}_{\text{super}}$, in the sense of the ratios of the $\|\psi_j\|^2$.

We can generalise from the preceding and define three measures on $[\boldsymbol{\psi}]$: if $\boldsymbol{\psi}'$ is the canonical representative of $[\boldsymbol{\psi}]$, then

linearity: $\mathcal{D}_1([\boldsymbol{\psi}]) := \|\psi_1'\|^2$.

[16]Note that all pictures in Figure 3.16, and those in Figure 3.17, are to the same scale, and this is common to the horizontal and vertical axes

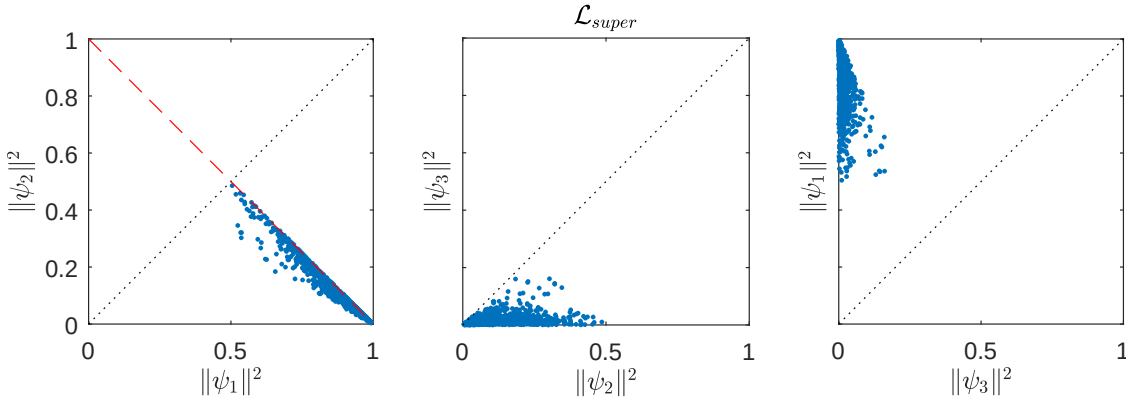


Figure 3.16: *Scatter diagrams for* $\|\psi_1\|^2$ *and* $\|\psi_2\|^2$*;* $\|\psi_2\|^2$ *and* $\|\psi_3\|^2$*; and* $\|\psi_3\|^2$ *and* $\|\psi_1\|^2$*, respectively, for the canonical representatives of a sample of members of* $\mathcal{L}_{super}$*. The black dotted lines are* $\|\psi_j\|^2 = \|\psi_k\|^2$ *for appropriate* j *and* k*, and the red dotted line is* $\|\psi_1\|^2 + \|\psi_2\|^2 = 1$

planarity: $\mathcal{D}_2([\boldsymbol{\psi}]) := \|\psi_2'\|^2$;

fullness (of dimensionality): $\mathcal{D}_3([\boldsymbol{\psi}]) := \|\psi_3'\|^2$;

As $\mathcal{D}_1([\boldsymbol{\psi}]) \geq \mathcal{D}_2([\boldsymbol{\psi}]) \geq \mathcal{D}_3([\boldsymbol{\psi}]) \geq 0$ and $\mathcal{D}_1([\boldsymbol{\psi}]) + \mathcal{D}_2([\boldsymbol{\psi}]) + \mathcal{D}_3([\boldsymbol{\psi}]) = 1$, values of $\mathcal{D}_1([\boldsymbol{\psi}])$ close to its maximum, 1, force $\mathcal{D}_2([\boldsymbol{\psi}])$ and $\mathcal{D}_3([\boldsymbol{\psi}])$ to be close to zero, values of $\mathcal{D}_2([\boldsymbol{\psi}])$ close to its maximum, $\frac{1}{2}$, reduce $\mathcal{D}_1([\boldsymbol{\psi}])$ to around $\frac{1}{2}$ and force $\mathcal{D}_3([\boldsymbol{\psi}])$ to be close to zero, and values of $\mathcal{D}_3([\boldsymbol{\psi}])$ close to its maximum, $\frac{1}{3}$, force $\mathcal{D}_1([\boldsymbol{\psi}]) + \mathcal{D}_2([\boldsymbol{\psi}])$ to be close to $\frac{2}{3}$.

Given all this, we can classify $[\boldsymbol{\psi}]$ as

near linear if $\mathcal{D}_1([\boldsymbol{\psi}]) \geq 2\mathcal{D}_2([\boldsymbol{\psi}]), 3\mathcal{D}_3([\boldsymbol{\psi}])$;

near planar if $2\mathcal{D}_2([\boldsymbol{\psi}]) > \mathcal{D}_1([\boldsymbol{\psi}])$, $2\mathcal{D}_2([\boldsymbol{\psi}]) \geq 3\mathcal{D}_3([\boldsymbol{\psi}])$;

(dimensionally) full if $3\mathcal{D}_3([\boldsymbol{\psi}]) > \mathcal{D}_1([\boldsymbol{\psi}]), 2\mathcal{D}_2([\boldsymbol{\psi}])$.

Figure 3.18 displays four pseudo-colour plots of the squared distances between ECWTs. The top left-hand diagram uses the distances in $\mathcal{L}$, and the other plots use distances within three different, uniformly randomly selected subsets of 100 members of $\mathcal{L}_{\text{super}}$.

In order to try and reveal any structure in the diagram, the members of of $\mathcal{L}$ are ordered: given a member of $\mathcal{L}$, we find the sum of its squared distances to the other members, and order the members in increasing values of this quantity. We order the subsets of $\mathcal{L}_{\text{super}}$ in a similar way.

In fact, there does seem to be some structure in the $\mathcal{L}$ diagram — for example, there are clearly two square regions along the main diagonal which are bluer than the remaining rectangular regions. The square regions are given by both indices in the range 43–62, and both indices in the range 67–75. Although this is consistent with the presence of at least two clusters in the high-level cluster centres (note that a different ordering of $\mathcal{L}$ might reveal different apparent clusters), a glance at the similar diagrams for the randomly selected members

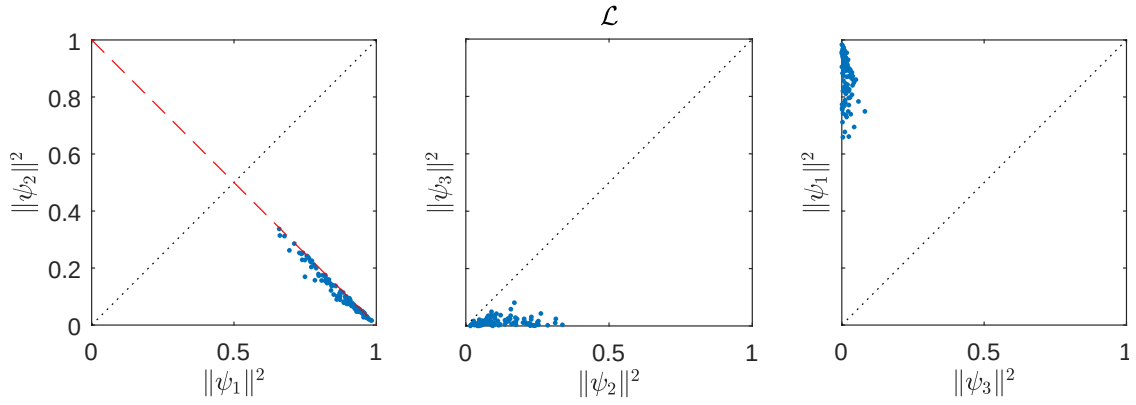


Figure 3.17:]
Analogue of Figure 3.16 for all of $\mathcal{L}$

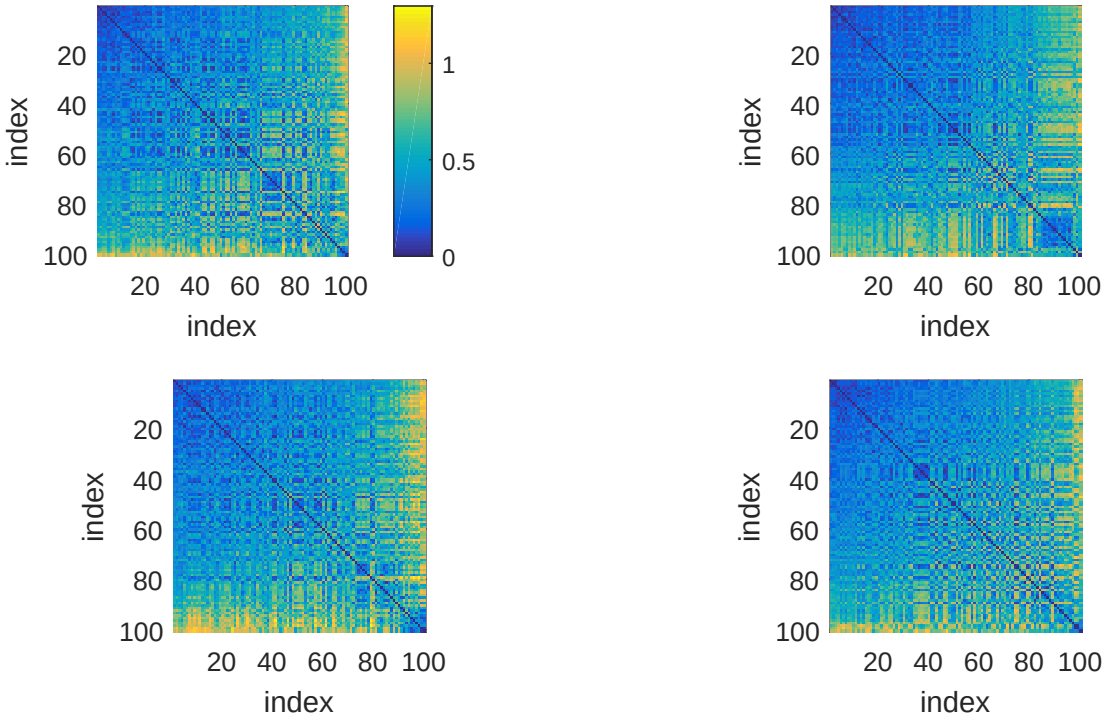


Figure 3.18: *Pseudocolour plots of the matrices of the squared distances between triplets. Upper left: for $\mathcal{L}$; remaining plots: for uniformly randomly selected subsets of 100 members of $\mathcal{L}_{super}$. The colour scale (given by the colour bar in the upper left picture) is the same for each plot*

of $\mathcal{L}_{\text{super}}$ shows that the impression of structure given by any of the diagrams is probably accidental, as the latter diagrams show "structures" which appear just as well-defined but are in fact different.

However, despite some smaller square regions along the main diagonal of the $\mathcal{L}_{\text{super}}$ diagrams, any consistent structure across these is restricted to the drift, and even this is less prominent than in the $\mathcal{L}$ diagram.

The squared D^* distances between distinct ECWTs in $\mathcal{L}$ are in the range 0.0282-1.2096 (*cf.* the theoretical maximum of 2), and these ranges for each of the subsets of $\mathcal{L}_{\text{super}}$ are 0.0025-1.0622, 0.0002-1.1037 and 0.0188-1.1433. The median square distance in $\mathcal{L}$ is 0.4209, and this quantity is 0.4455, 0.4073 and 0.4441 in the $\mathcal{L}_{\text{super}}$ samples. The interquartile distance of the squared distances is 0.3504 for $\mathcal{L}$, and 0.3433, 0.3253 and 0.3613 for the samples.

Without actually doing a statistical analysis, there appears to be no significant difference between the distribution of the distances within $\mathcal{L}$ and those within the samples of $\mathcal{L}_{\text{super}}$.

We might have expected that outliers would be eliminated from $\mathcal{L}_{\text{super}}$ when $\mathcal{L}$ was selected, and that cluster centres would not be too close together. These effects would have resulted in the distribution of the distances in $\mathcal{L}$ being narrower than the same distribution for $\mathcal{L}_{\text{super}}$. However, if such a narrowing effect exists, it is too small to be visible on a superficial examination.

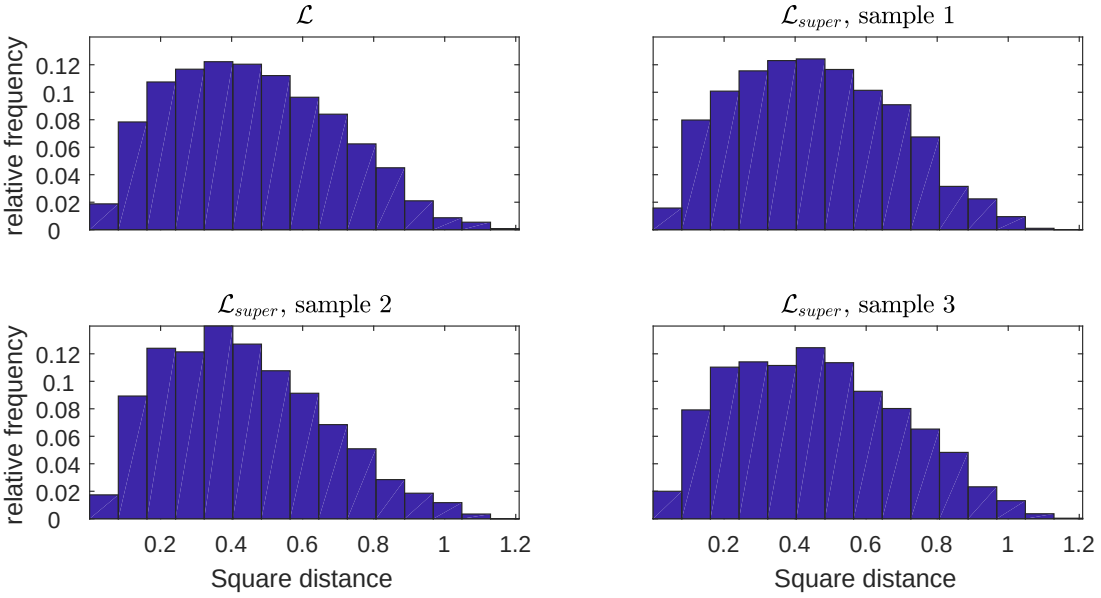


Figure 3.19: *Histograms of the squared distances between triplets. Upper left: for $\mathcal{L}$; remaining plots: for samples of $\mathcal{L}_{super}$*

To sum up, $\mathcal{L}$ appears to be representative of $\mathcal{L}_{\text{super}}$ in terms of the joint distribution of the squared magnitude of the three components of the canonical representatives and the relative importance of the parts of the trace giving rise to the members of $\mathcal{L}_{\text{super}}$ and $\mathcal{L}$. It appears that the pattern of window lengths relevant to $\mathcal{L}$ is different to that relevant to $\mathcal{L}_{\text{super}}$, with central window lengths being more prominent in the latter, and it is also true

that the central values of the GoF distribution seem boosted in $\mathcal{L}$ in comparison to those in the samples of $\mathcal{L}_{\text{super}}$. Also, contrary to expectation, the distribution of inter-member distances of $\mathcal{L}$ does not seem to be more peaked *vis-à-vis* that of $\mathcal{L}_{\text{super}}$.

We now turn to examining some members of $\mathcal{L}$ more closely, for illustrative purposes.

In Figure 3.20, we show the two members of the library which are the farthest apart, in terms of the distance $\mathcal{D}^*$, which turn out to be the 58th and 99th members of $\mathcal{L}$, $[\boldsymbol{\psi}_{58}]$ and $[\boldsymbol{\psi}_{99}]$, using the representatives of the equivalence classes which were derived from the data.

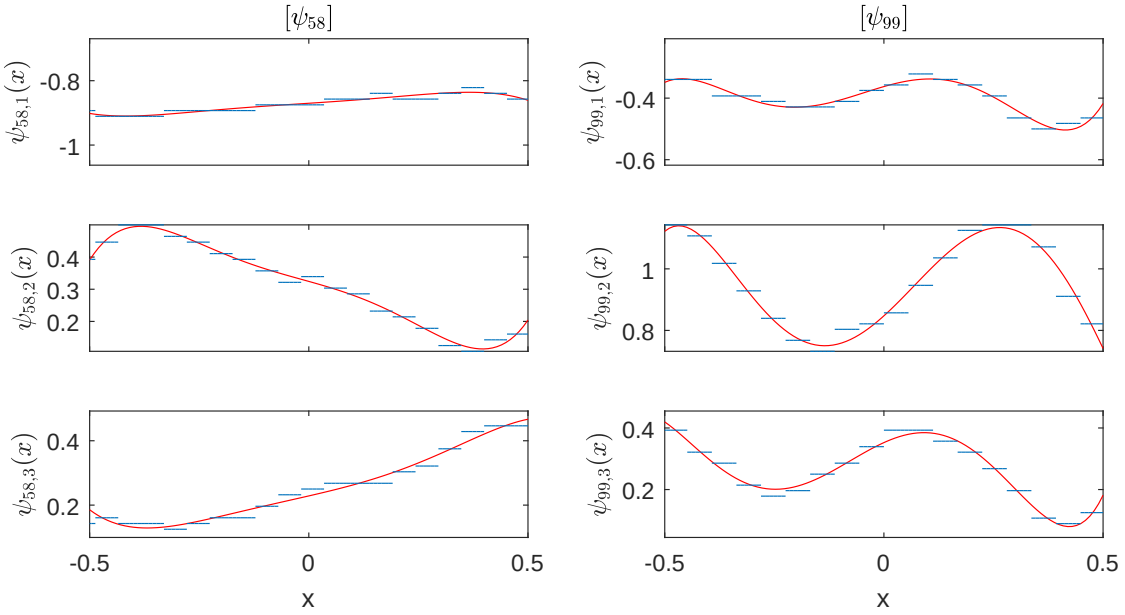


Figure 3.20: *The data-fitted representatives of the two ECWTs in the final library with the greatest separation, with data supporting their derivation (the wavelets are time-scaled to fit the window, and vertically shifted and scaled to be as close as possible to the data, and the difference between the minimum and maximal values on the vertical axis is constant for diagrams in the same column)*

In Figure 3.21, we use the canonical representatives, $\boldsymbol{\psi}'_{58}$ and $\boldsymbol{\psi}'_{99}$ (so $[\boldsymbol{\psi}'_{58}] = [\boldsymbol{\psi}_{58}]$ and $[\boldsymbol{\psi}'_{99}] = [\boldsymbol{\psi}_{99}]$) to show the same two members of the library.

In Figure 3.22, we observe that the areas between the horizontal axis and the curve $\psi'_{58,1}(x)\psi'_{99,1}(x)$ make positive and negative contributions of roughly equal magnitude to the integral over this products' support, and so the integral is small. The products $\psi'_{58,2}(x)\psi'_{99,2}(x)$ and $\psi'_{58,2}(x)\psi'_{99,2}(x)$ are relatively small in magnitude, so the integral in $\langle \psi'_{58,j}, \psi'_{99,j} \rangle = \int_{-\frac{1}{2}}^{\frac{1}{2}} \psi'_{58,j}(x)\psi'_{99,j}(x)\mathrm{d}x$ is also close to 0 for $j = 2, 3$. Therefore, ψ'_{58} and $\psi'_{99,}$ are near-orthogonal, and the squared L_2 distance between the canonical representatives of 1.8987 (note this is not the same as the D^* distance between the ECWTs, which we know is 1.2096 from the range of distances in $\mathcal{L}$ given above) comes as no surprise. The L_2 distance between the data-fitted representatives is 2.6664, which is again no surprise, as the canonical representatives of the ECWTs are, in general, closer together than the effectively randomly-selected representatives which are "chosen" by the original accelerometry data.

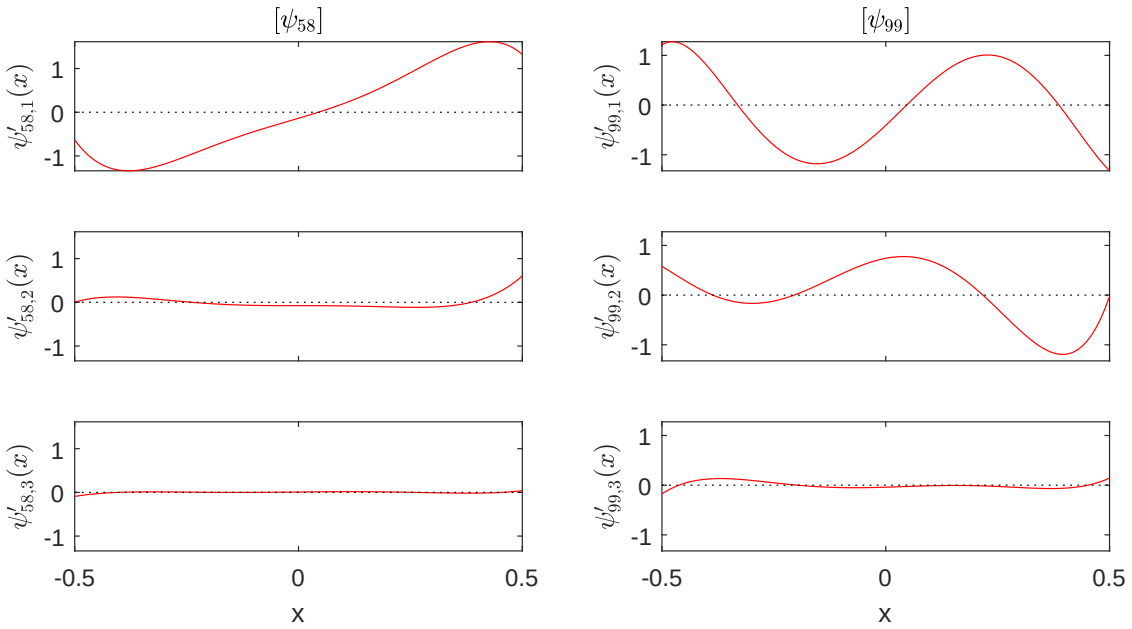


Figure 3.21: *The canonical representatives of the two ECWTs of Figure 3.20*

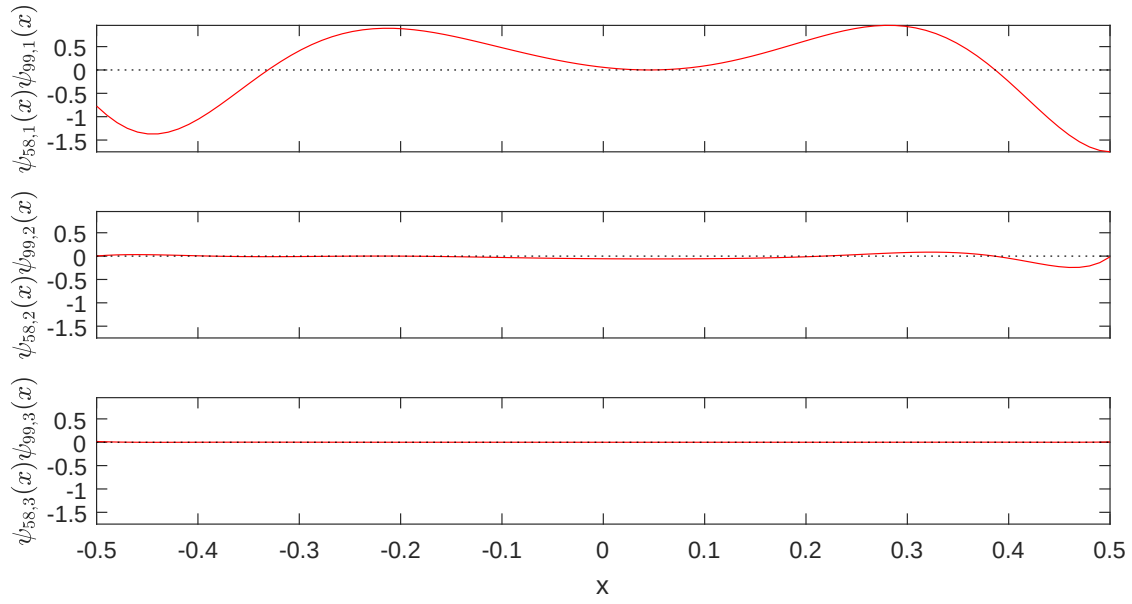


Figure 3.22: *Products of the components of the canonical representatives the two ECWTs of Figure 3.20*

We collect some properties of $[\boldsymbol{\psi}_{58}]$ and $[\boldsymbol{\psi}_{99}]$ in Table 3.2.

We see that $[\boldsymbol{\psi}_{58}]$ comes from an epoch in the the trace from whence 4% of ECWTs in $\mathcal{L}$ originate (and 16% originate from this epoch or its immediate neighbours — see Figure 3.13)

	$[\boldsymbol{\psi}_j]$		
Property	$[\boldsymbol{\psi}_{58}]$	$[\boldsymbol{\psi}_{99}]$	$[\boldsymbol{\psi}_{54}]$
epoch	17	25	31
window length	116	108	128
GoF	0.9814	0.9551	0.9488
$\|\psi'_{j,1}\|^2$	**0.9837**	0.6589	0.7494
$\|\psi'_{j,2}\|^2$	0.0160	**0.3372**	0.1694
$\|\psi'_{j,3}\|^2$	0.0003	0.0038	**0.0811**

Table 3.2: *Some properties of of* $[\boldsymbol{\psi}_{58}]$, $[\boldsymbol{\psi}_{99}]$ *and* $[\boldsymbol{\psi}_{54}]$. *The values in bold are maximised by their* $[\boldsymbol{\psi}_j]$

but $[\boldsymbol{\psi}_{99}]$ is extracted from a data epoch which yields over four times as many ECWTs (17%, 23% with its neighbouring epochs). The window length giving rise to $[\boldsymbol{\psi}_{58}]$ is of low-medium "popularity" in $\mathcal{L}$, at (3%, and 7% with the two neighbouring window lengths), while that giving rise to $[\boldsymbol{\psi}_{58}]$ is of maximal popularity (8%, 15% with neighbours). The two values of GoF are of low-mid (for ECWT 58) and mid-high (for ECWT 99) "popularity", while the values themselves are high and mid-high.

Finally, $[\boldsymbol{\psi}_{58}]$ has a high linearity score ($\mathcal{D}_1 = \|\psi'_{\cdot 1}\|^2$), and $[\boldsymbol{\psi}_{99}]$ a high planarity score. In fact, they have the highest such scores in $\mathcal{L}$, as shall see very shortly.

The major differences between $[\boldsymbol{\psi}_{58}]$ and $[\boldsymbol{\psi}_{99}]$ are in their dimensionality measures and their GoF, as neither emerges from a rare epoch, nor is extracted using an exceptional window length.

We now find the ECWTs with the maximum values in $\mathcal{L}$ of $\mathcal{D}_1([\boldsymbol{\psi}])$, $\mathcal{D}_2([\boldsymbol{\psi}])$ and $\mathcal{D}_3([\boldsymbol{\psi}])$. These are in fact $[\boldsymbol{\psi}_{58}]$, $[\boldsymbol{\psi}_{99}]$ and $[\boldsymbol{\psi}_{54}]$ respectively, so we needed only to extend Table 3.2 to include details of $[\boldsymbol{\psi}_{54}]$.

We display the data-fitted and canonical representatives of $[\boldsymbol{\psi}_{54}]$ in Figure 3.23. This ECWT does not stem from an unproductive epoch, nor is the window giving rise to it of a length which is isolated among the lengths giving rise to other ECWTs. However, its GoF is near the low end of the range for $\mathcal{L}$.

We have $D^*([\boldsymbol{\psi}_{58}], [\boldsymbol{\psi}_{99}])^2 = 1.2096$, $D^*([\boldsymbol{\psi}_{99}], [\boldsymbol{\psi}_{54}])^2 = 0.4821$ and $D^*([\boldsymbol{\psi}_{54}], [\boldsymbol{\psi}_{58}])^2 = 0.9287$, so, with this distance, the most "linear" of the three is closest to the "fullest", which is contrary to what might be expected if "planarity" is mid-way between "linearity" and "fullness". In other words, the ECWTs are capturing more than the dimensionality characteristics of the motion.

Some of the advantages of using the canonical representative are illustrated by comparing the two Figures 3.20 and 3.21, and the left and right columns of Figure 3.23. It would take very careful consideration of the data-fitted representatives of $[\boldsymbol{\psi}_{58}]$, $[\boldsymbol{\psi}_{99}]$ and $[\boldsymbol{\psi}_{54}]$ to decide that they stood for motion in a line, in a plane and neither, respectively (at least with respect to the NAT axes).

It is virtually impossible to deduce by merely looking at the data-fitted representatives

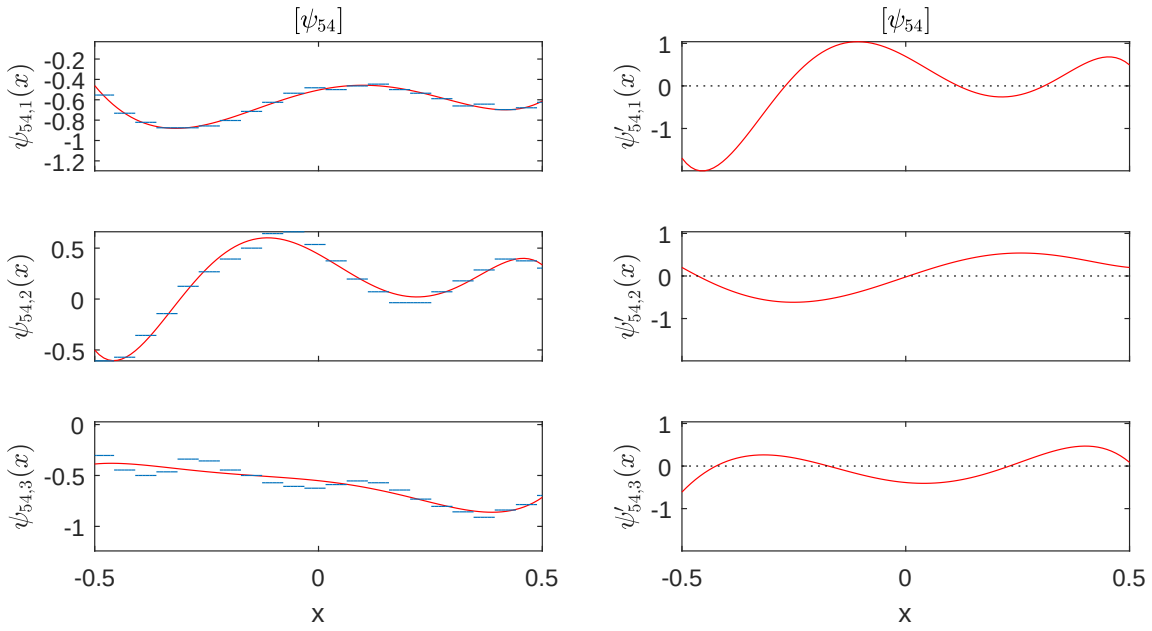


Figure 3.23: *The representatives of the ECWT in the final library with the greatest value of $\mathcal{D}_3$. On the left-hand side, we present the data-fitted representatives, with data supporting their derivation (the wavelets are time-scaled to fit the window, and vertically shifted and scaled to be as close as possible to the data; the vertical scaling of the* diagrams *is the same in all three diagrams), and on the right we have the canonical representatives (same axes in all cases on the right)*

that $[\boldsymbol{\psi}_{46}]$ can be be reduced to acceleration mostly in a single plane, or that $[\boldsymbol{\psi}_{82}]$ cannot be so reduced, but this can be done easily by looking at the canonical representatives.

In order to put the "dimensional" properties of Table 3.2 into context, we present the median, first and third quartile values in Table 3.3.

Property	**Median**	**1st quartile**	**3rd quartile**	**Interquartile distance**
$\|\psi'_{:,1}\|^2$	0.8913	0.8095	0.9390	0.1294
$\|\psi'_{:,2}\|^2$	0.0902	0.0516	0.1684	0.1168
$\|\psi'_{:,3}\|^2$	0.0101	0.0050	0.0223	0.0173

Table 3.3: *Some statistics for the $\mathcal{D}_j(=\|\psi'_{:,j}\|^2)$ over $\mathcal{L}$*

Thus, we see that $\mathcal{D}_1$ exceeds its 3rd-quartile value for $[\boldsymbol{\psi}_{58}]$ by about 0.42 times the interquartile distance, while it is below its 1st-quartile value for $[\boldsymbol{\psi}_{99}]$ (by 1.80 iqr) and $[\boldsymbol{\psi}_{54}]$ (0.46 iqr). $\mathcal{D}_2$ is 0.30 iqr below its 1st-quartile value for $[\boldsymbol{\psi}_{58}]$, while it is above its 3rd-quartile value for $[\boldsymbol{\psi}_{99}]$ (by 1.45 iqr). It is also very slightly above its 3rd quartile for $[\boldsymbol{\psi}_{54}]$. Finally, $\mathcal{D}_3$ is below its 1st-quartile value for $[\boldsymbol{\psi}_{58}]$ (0.27 iqr) and $[\boldsymbol{\psi}_{99}]$ (0.07 iqr), while it is above its 3rd-quartile level for $[\boldsymbol{\psi}_{54}]$ (3.40 iqr).

Therefore, $[\boldsymbol{\psi}_{58}]$, $[\boldsymbol{\psi}_{99}]$ and $[\boldsymbol{\psi}_{54}]$ have comparatively large or very large values of the dimensionality quantities they were selected for, and low values of the quantities they were not selected for, with the exception of $[\boldsymbol{\psi}_{54}]$ for $\mathcal{D}_2$ and $[\boldsymbol{\psi}_{99}]$ for $\mathcal{D}_3$.

According to our earlier definitions, 99% of the members of $\mathcal{L}$ are near linear, 0% are dimensionally full and the sole near planar member is $[\boldsymbol{\psi}_{99}]$. Note that the maximally full $[\boldsymbol{\psi}_{54}]$, which is closer (in terms of D^*) to near-linear $[\boldsymbol{\psi}_{58}]$ than it is to near-planar $[\boldsymbol{\psi}_{99}]$, is itself actually near linear.

3.3.10 Using the library

The activation

We define the activation at x of the ECWT $[\boldsymbol{\psi}]$ with respect to the window length w by $A^{(0)}_{T,[\boldsymbol{\psi}]}(x) := \frac{\max_{\boldsymbol{\psi}'\in[\boldsymbol{\psi}]}\left\{\sum_{j=1}^{3}\left|\langle f_j,\psi'_{T,x,j}\rangle\right|^2\right\}}{\sum_{j=1}^{3}\|f_j-\bar{f}_j\|^2}$, where $\mathbf{f}$ models the data in the window of length (in time) $T = w/S$, centred on x (S is the sample rate). As $[\,\pm\psi'_1\ \pm'\psi'_2\ \pm''\psi'_3\,] \in [\boldsymbol{\psi}]$ if $[\,\psi'_1\ \psi'_2\ \psi'_3\,] \in [\boldsymbol{\psi}]$, we can drop the absolute value here. Also, because of the way that the mother wavelets $\psi'_j = \psi'_{1,0,j}$ scale to give $\psi'_{T,x,j}$, if the independent variable in $\mathbf{f}$ is linearly transformed to yield $\mathbf{f}_{x,T}$, (necessarily) zero outside of $\left[-\frac{1}{2},\frac{1}{2}\right)$, we can write $A^{(0)}_{T,[\boldsymbol{\psi}]}(x) = \frac{\max_{\boldsymbol{\psi}'\in[\boldsymbol{\psi}]}\left\{\left[\sum_{j=1}^{3}\langle f_{x,T,j},\psi'_j\rangle\right]^2\right\}}{\sum_{j=1}^{3}\|f_{x,T,j}-\bar{f}_{x,T,j}\|^2} := A_{w,[\boldsymbol{\psi}]}(x)$.

Thus, if $\boldsymbol{\psi}$ was fitted to the data over a window of length w centred on x, then $G_{\text{fit}}(\boldsymbol{\psi}) = A_{w,[\boldsymbol{\psi}]}(x)$. In other words, G_{fit} is the activation of $[\boldsymbol{\psi}]$ on its own data.

As multiplying the magnitude of the data in our interval by c multiplies both $[\sum_{j=1}^{3}\langle f_{x,T,j},\psi'_j\rangle|^2$ and $\sum_{j=1}^{3}\|f_{x,T,j}-\bar{f}_{x,T,j}\|^2$ by c^2, $A_{w,[\boldsymbol{\psi}]}(x)$ can be interpreted as a monotonically increasing function of the relative content of $[\boldsymbol{\psi}]$ in $\mathbf{f}$ at scale T.

Now $\max_{\boldsymbol{\psi}'\in[\boldsymbol{\psi}]}\left\{\left[\sum_{j=1}^{3}\langle f_{x,T,j},\psi'_j\rangle\right]^2\right\} = \max_{O\in\mathrm{O}(3)}\left\{\left[\sum_{j=1}^{3}\sum_{k=1}^{3}O_{jk}\langle f_{x,T,j},\psi_k\rangle\right]^2\right\}$, where, if $f_{x,T,j}(u) = \begin{cases} y_{j\ell}, & u\in[x_{\ell-1},x_\ell],\quad \ell=1,2,\ldots,N;\\ 0, & \text{otherwise}\end{cases}$ (for $x_0=-\frac{1}{2}$, $x_N=\frac{1}{2}$), $\langle f_{x,T,j},\psi_k\rangle = \int_{-\frac{1}{2}}^{\frac{1}{2}} f_{x,T,j}(u)\psi_k(u)\mathrm{d}u = \sum_{q=0}^{n}\sum_{\ell=1}^{N}a_{jk}y_{j\ell}\frac{x_\ell^{q+1}-x_{\ell-1}^{q+1}}{q+1} = z_j^{\mathrm{T}}a_k$, for $z_j = Ry_j$, where R is given by equation (3.15).

This means maximising $\left[\sum_{j=1}^{3}\sum_{k=1}^{3}O_{jk}\langle f_{x,T,j},\psi_k\rangle\right]^2$ with respect to O, subject to $O\in\mathrm{O}(3)$ is a problem equivalent to finding the extrema of $\sum_{j=1}^{3}\sum_{k=1}^{3}O_{jk}z_j^{\mathrm{T}}a_k$ under the same conditions. But this is the problem of section 3.3.1, with $K_{jk} = a_j^{\mathrm{T}}Ha'_k$ replaced by $K_{jk} = z_j^{\mathrm{T}}a_k$.

Consequently, we have

$$A_{w,[\boldsymbol{\psi}]}(x) = \frac{1}{\sum_{j=1}^{3}\|f_{x,T,j}-\bar{f}_{x,T,j}\|^2}\sum_{k=1}^{3}\mu_k^{\frac{1}{2}}, \tag{3.44}$$

where the μ_k are the eigenvalues of KK^{T} for $K_{jk} = z_j^{\mathrm{T}}a_k$.

Selecting distinguishing library members

Reviewing our overall aim, we wish to find patterns of acceleration which are relatively frequent in the movement of PD subjects compared to non-PD ones. Constructing a library in the manner described above will, it is hoped, capture the movement patterns of PD subjects, including those which also occur in non-PD subjects. If we have captured patterns characteristic of PD, these and very similar patterns will occur much less frequently in data stemming from non-PD subjects.

In terms relating to our library, an ECWT which has high activations on new data (i.e.,

data different to the data it was extracted from) from PD patients, compared to the activations on data from non-PD subjects, is likely to be picking up an acceleration pattern characteristic of PD.

We quantify this criterion for picking distinguishing ECWTs as follows:

1 pick a window tolerance W;

2 for each ECWT $[\boldsymbol{\psi}]$ in the library, set $w_{[\boldsymbol{\psi}]}$ to the window length that gave rise to the data-fitted representative of $[\boldsymbol{\psi}]$ and set $\mathcal{S}_{[\boldsymbol{\psi}]} = \{w_{[\boldsymbol{\psi}]} - W, w_{[\boldsymbol{\psi}]} - W + 1, \ldots, w_{[\boldsymbol{\psi}]}, \ldots, w_{[\boldsymbol{\psi}]} + W - 1, w_{[\boldsymbol{\psi}]} + W\}$;

3 for each ECWT $[\boldsymbol{\psi}]$ and each $w \in \mathcal{S}_{[\boldsymbol{\psi}]}$, calculate $A_{w,[\boldsymbol{\psi}]}(x)$ for each x that is a centre of a (complete) window of length w into the new PD data;

4 repeat 3 for the windows into the non-PD data;

5 find the maximum $M_A = A_{w,[\boldsymbol{\psi}]}(x)$ over $w \in \mathcal{S}_{[\boldsymbol{\psi}]}$ and x in both data sets;

6 for each $\theta \in T$, where $T \subset (0, M_A)$ is a discrete set of reasonably large cardinality, find the number of (w, x) pairs with x in the new PD data for which $A_{w,[\boldsymbol{\psi}]}(x) > \theta$ and find the ratio, $\rho(\theta; \text{PD})$ of this number to the total number of (w, x) pairs considered;

7 repeat 6 for the non-PD data, to find $\rho(\theta; \text{non-PD})$ analogous to $\rho(\theta; \text{PD})$;

8 find $M_{\Delta\rho} = \max_{\theta \in T}[\rho(\theta; \text{PD}) - \rho(\theta; \text{non-PD})]$ and $\theta_0 = \arg\max_{\theta \in T}[\rho(\theta; \text{PD}) - \rho(\theta; \text{non-PD})]$;

9 find the ECWTs with the largest values of $M_{\Delta\rho}$ as candidates to distinguish PD data from non-PD data.

Alternatively, the last two steps can be replaced by:

8 find $M'_{\Delta\rho} = \max_{\{\theta \in T : \max\{\rho(\theta;\text{PD}), \rho(\theta;\text{non-PD})\} \geq \epsilon_\rho\}} \frac{\rho(\theta;\text{PD}) - \rho(\theta;\text{non-PD})}{\rho(\theta;\text{PD}) + \rho(\theta;\text{non-PD})}$ and $\theta'_0 = \arg\max_{\{\theta \in T : \max\{\rho(\theta;\text{PD}), \rho(\theta;\text{non-PD})\} \geq \epsilon_\rho\}} \frac{\rho(\theta;\text{PD}) - \rho(\theta;\text{non-PD})}{\rho(\theta;\text{PD}) + \rho(\theta;\text{non-PD})}$;

9 find the ECWTs with the largest values of $M'_{\Delta\rho}$ as candidates to distinguish PD data from non-PD data,

that is, we basically use the difference of $\rho(\theta; \text{PD})$ and $\rho(\theta; \text{non-PD})$ relative to their sum, instead of the raw difference.

The parameter W ensures that scales neighbouring the one which gave rise to an ECWT in the first place are considered, so that slower or faster versions of the relevant acceleration "shape" will be detected — necessary as a single subject may differ in the speed at which they execute such a shape, depending on how they feel, and also necessary for generalisations across subjects.

The parameter ϵ_ρ ensures that relative differences are not based on too few windows, as small random changes in a small denominator of a ratio can dramatically affect that ratio.

The outputs of these algorithms are library members and the corresponding thresholds for calculating the ρs.

We may also use $\Theta = \frac{1}{2}(\rho(\theta_0; \text{PD}) + \rho(\theta_0; \text{non-PD}))$ and $\Theta' = \frac{1}{2}(\rho(\theta'_0; \text{PD}) + \rho(\theta'_0; \text{non-PD}))$ as the starting point for finding classification boundaries.

Classifying on the activation

Once we have found the distinguishing members of our library, we can attempt to use them to classify sets of data.

Given further sets of data, we can calculate ρ on them in a manner completely analogous to the calculation of $\rho(\theta_0; \text{PD})$ and $\rho(\theta_0; \text{non-PD})$ in the previous section. Then this quantity can be thresholded to give a classification into PD and non-PD data. The quality of this classification will depend on the threshold and the empirical distribution of the ρ.

The later point — motivation for D^*

At the end of section 3.3.2, we said we would provide a motivation for the definition of D^*.

Consider $[\boldsymbol{\psi}_1]$, $[\boldsymbol{\psi}_2]$ such that $D^*([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_2]) = D([\boldsymbol{\psi}_1], [\boldsymbol{\psi}_{2;x_0}]) \approx 0$. Then, if $[\boldsymbol{\psi}_1]$ has a high activation on a particular window into the data, centred on x_1 and at a scale s, say, then $[\boldsymbol{\psi}_2]$ is likely to have a high activation at the same scale on a neighbouring window, centred on $x_1 + sx_0$.

Consequently, the number of high activations of $[\boldsymbol{\psi}_1]$ over a stretch of data is likely to be similar to the number of high activations of $[\boldsymbol{\psi}_2]$ over the same stretch and scale, reflecting the similarity of the shapes of $[\boldsymbol{\psi}_1]$ and the shifted version of $[\boldsymbol{\psi}_2]$. This redundancy is not what we want, so we will interpret the similarity of $[\boldsymbol{\psi}_1]$ and of $[\boldsymbol{\psi}_{2;x_0}]$ as a similarity of $[\boldsymbol{\psi}_1]$ and of $[\boldsymbol{\psi}_2]$. The mathematical expression of this interpretation is the use of D^* as a distance between ECWTs.

3.3.11 Example: 5) Activation, member selection and classification

In order to proceed with our example, we need 4 more data sets in addition to the six hours of data from the PD1 subject we have used so far. We will refer to that data as the *library* data. We have a further four hours of saturation-free data with the same characteristics (right-wrist attachment, nominal ±2g, sample rate 500Hz) from PD1, and eight hours of such data from a control subject, $C2$, which will will call PD- and non-PD-*selection* data. Finally, we have two hours of other saturation-free data from each of these subjects to serve as *test* data.

We set the parameter $W = 4$ and use several values of θ, tailored to the particular ECWT and selected in the following way: let M_A be the maximum value of the activation with respect to the particular ECWT being considered, over both selection-data sets and all the window lengths considered. M_A will not attain its upper bound of 1, as we are fitting polynomials to (in general, non-constant) piecewise constant functions — how close it gets depends on the

ECWT. Then we use $\theta_{0.85} = 0.85M_A$, $\theta_{0.9} = 0.9M_A$, $\theta_{0.95} = 0.95M_A$ and $\theta_{0.99} = 0.8M_A$ as the thresholds.

Activation

It turns out that when we evaluate the activations on our PD- and non-PD-selection data with respect to our library members, the greatest activation we find is 0.9395, with respect to $[\psi_{60}]$ at a scale corresponding to 72 samples (and so to 6.94Hz), despite this ECWT originally being derived from a window of 80 samples (6.25Hz), on a stretch of PD-selection data.

Figure 3.24 shows this activation at the scales considered with respect to ψ_{60} over a section of the PD-selection data surrounding the stretch of greatest activation.

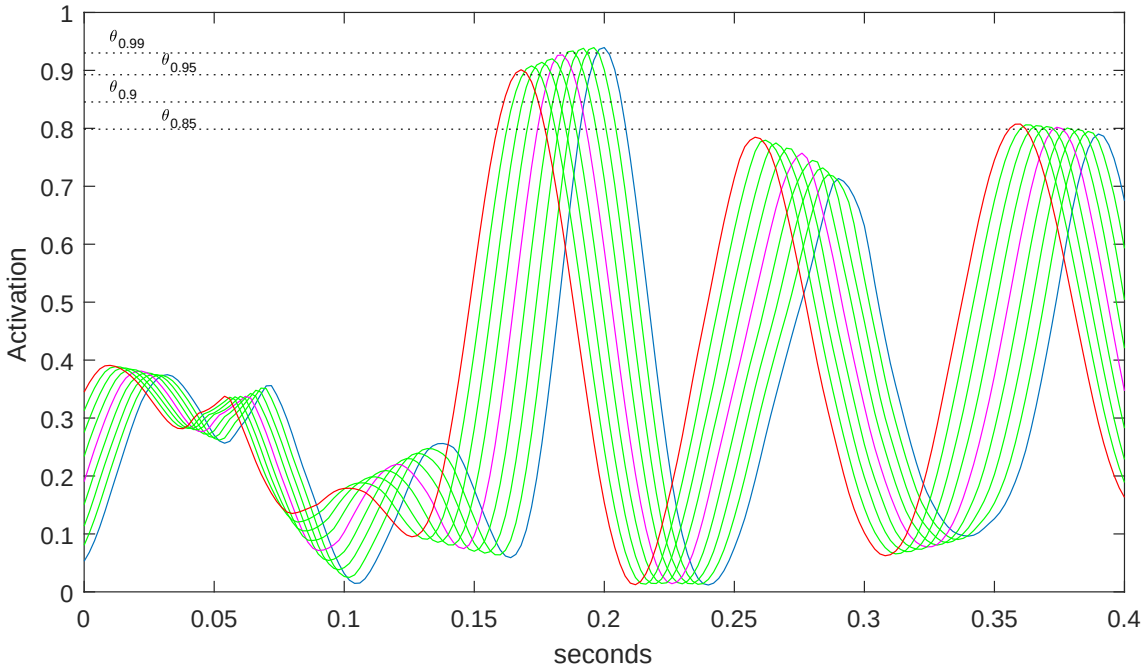


Figure 3.24: *The activations with respect to the 60th ECWT of our library, over the 0.4s surrounding the greatest activation. The blue line is for the scale 72 samples, the smallest considered for this ECWT, the magenta line is for the 80 sample scale, the scale on which this ECWT was originally derived, and the red line for the 88 sample scale, the largest scale we consider. The green lines correspond to the intermediate values*

Selection

In Figure 3.24 we also draw the horizontal lines corresponding to thresholds $\theta = \theta_{0.85} = 0.85M_A, \theta_{0.9} = 0.9M_A, \theta_{0.95} = 0.95M_A$ and $\theta_{0.99} = 0.9M_A$. For a particular θ, $\rho(\theta; \text{PD})$ is the count of all the windows corresponding to the points on all the curves which are above the θ-line divided by the total number of such windows (where, of course the diagram is extended to the whole of the trace). $\rho(\theta; \text{non-PD})$ is calculated in a similar fashion, using the

non-PD-selection data.

We then select the library members which yield the greatest $\max_\theta[\rho(\theta;\text{PD})-\rho(\theta;\text{non-PD})]$ or its relative version, and these are presented in Table 3.4.

	Best library members, $M_{\Delta\rho}$ criterion				**Best library members, $M'_{\Delta\rho}$ criterion**			
Rank	**Member number**	$M_{\Delta\rho}$	θ_0	Θ	**Member number**	$M'_{\Delta\rho}$	θ'_0	Θ'
1st	86	0.0059	0.2271	0.0178	62	0.4925	0.5370	0.0090
2nd	53	0.0052	0.2119	0.0022	16	0.4838	0.4594	6.8395×10^{-4}
3rd	47	0.0051	0.2368	0.0159	35	0.4625	0.5964	6.9589×10^{-4}
4th	88	0.0051	0.2260	0.0162	70	0.4542	0.3516	7.2113×10^{-4}
5th	13	0.0048	0.2222	0.0164	21	0.4523	0.3384	7.0453×10^{-4}

Table 3.4: *The best library members, in the sense that* $\rho(\theta_0;\text{PD})-\rho(\theta_0;\text{non-PD})$ *or* $\frac{\rho(\theta'_0;\text{PD})-\rho(\theta'_0;\text{non-PD})}{\rho(\theta'_0;\text{PD})+\rho(\theta'_0;\text{non-PD})}$ *is as large as possible (Coloured text corresponds to the disc colours in Figure 3.25)*

We see that using the $M'_{\Delta\rho}$ criterion pushes θ'_0 upwards in comparison to θ_0, which is what we might have expected, as higher θs cause $M_{\Delta\rho}$ to be the difference of smaller numbers, and therefore it naturally tends to be smaller, and consequently disadvantaged in the selection of the maximum.

To get some idea of the relation between $\rho(\theta;\text{PD})$ and $\rho(\theta,\text{non-PD})$ for our chosen θs and our selected library members, and the same quantities for the remaining members, we present scatter diagrams of $\rho(\theta,\text{PD})$ versus $\rho(\theta,\text{non-PD})$ in Figure 3.25.

We first note that, with the lower values of θ_0 that the first version of our algorithm produces, most of the crosses are on the "wrong" side of the dotted line, in that $\rho(\theta_0;\text{PD}) < \rho(\theta_0;\text{non-PD})$ for the corresponding library member. In particular, the crosses corresponding to the optimal members according to the second version are above the line. On the other hand, the θ_0s for the alternative version have almost all (for $\theta = 0.58M_A$) or all (for $\theta = 0.61M_A$) the crosses on the "right" side of the line.

We also note that the diagrams in the top row are very similar to each other, as are the ones in the bottom row, reflecting the fact that the θs are close to one another within a row.

Finally, in each case, the optimum library members are not too far from others, indicating that there may be other worthy candidates for distinguishing members.

In Figure 3.26 we show the empirical $\rho(\theta;\text{PD})$ and $\rho(\theta;\text{non-PD})$, their difference and their difference relative to their sum, for the most optimal and second most optimal library members for each criterion, members 86, 62, 53 and 16. We also show the median and first and third quartiles of these quantities over all the library members. For comparison, we also show the theoretical curve for ρ for a uniform distribution of activations over $[0, M_A]$.

The first thing we notice is that all the empirical curves for the library members drop much faster than the theoretical curve for a uniform distribution, on both the PD and non-

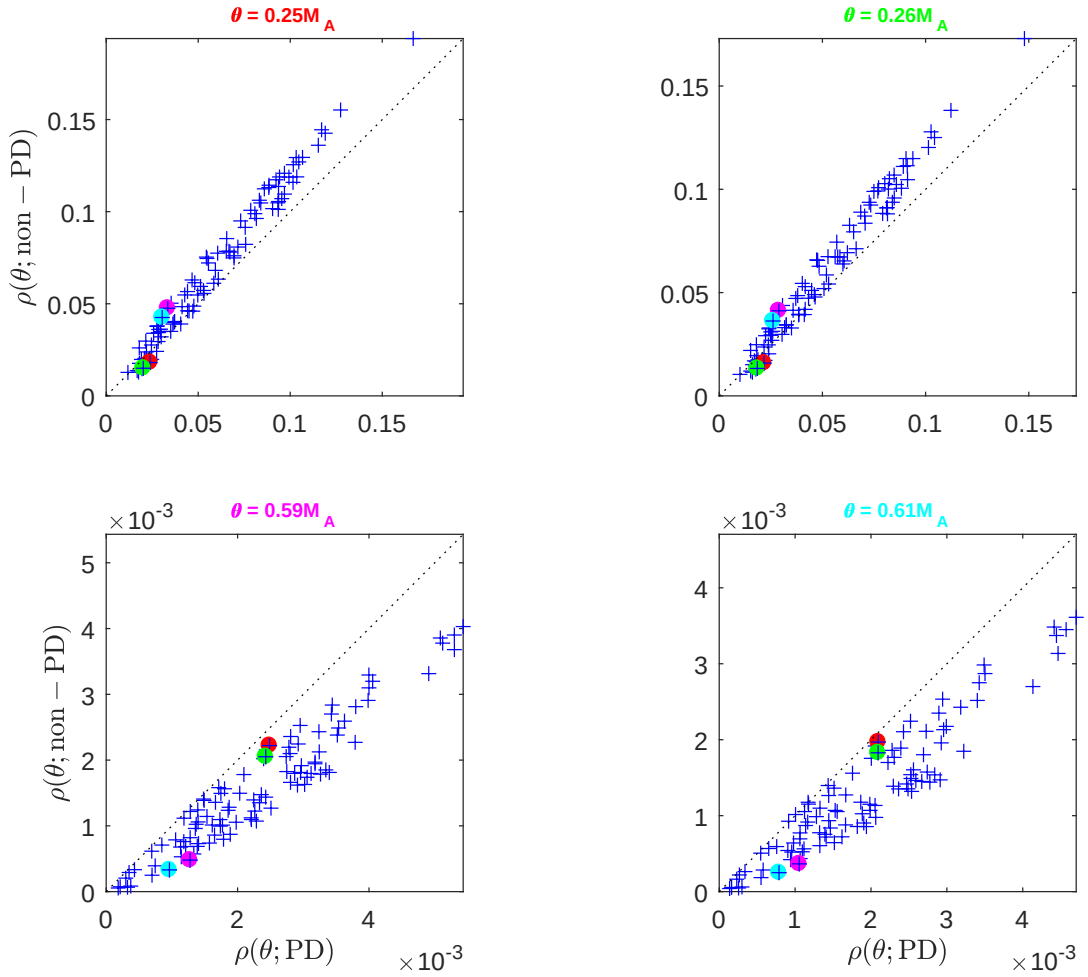


Figure 3.25: *Scatter diagrams for ρ for our values of θ. Each cross represents a library member, the crosses on discs correspond to members 86, 62, 53 and 16, with the colours of the discs matching the text in Table 3.4. The upper row corresponds to thresholds which optimise $\rho(\theta;\mathrm{PD}) - \rho(\theta;\text{non-PD})$ for the member of the same colour as the title, the lower to thresholds that optimise the relative version of these quantities for the relevant member. The dotted line separates, of course, the region where $\rho(\theta,\mathrm{PD}) > \rho(\theta;\text{non-PD})$ (below the line) from that where $\rho(\theta;\mathrm{PD}) < \rho(\theta;\text{non-PD})$*

PD data, as shown by the course of the dashed blue maximum line in the left-hand diagrams in comparison to the black curve. This is entirely in line with expectation if high activations correlate with particular patterns of movement, as there will be many such patterns (whether or not they are picked up by the techniques here), and individual patterns will be rare except in very pathological cases.

We also notice that all the curves corresponding to our selected library members (i.e., the red, green, magenta and cyan lines) are well below the maximum ρ curve in both left-hand diagrams, so none of them has a particularly large number of high activations on either data set.

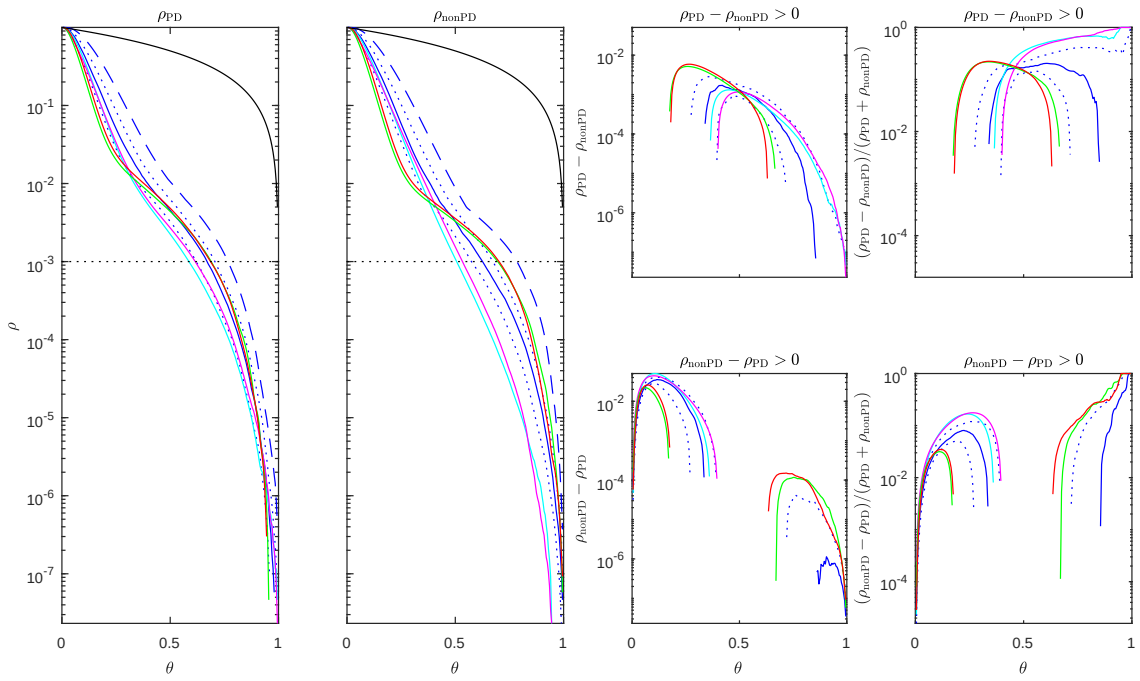


Figure 3.26: *Curves showing the empirical relation between the threshold* θ *and* $\rho(\theta;\mathrm{PD})$ *and* $\rho(\theta;\text{non-PD})$. *In the two left-hand diagrams,* $\rho(\theta;\mathrm{PD})$ *and* $\rho(\theta;\text{non-PD})$ *are shown for library members 86 (red), 62 (green), 53 (magenta) and 16 (cyan). Also shown (in blue) are the median (solid), the second and third quartiles (both dotted), and the maximum (dashed) of the* ρ*s for all the library members. The black curve corresponds to a uniform distribution of the* ρ*s (and would be a straight line if both axes were linear), and the dotted black line is* $\rho = \epsilon_\rho$. *The four right-hand diagrams show* $\rho(\theta;\mathrm{PD}) - \rho(\theta;\text{non-PD})$, $\rho(\theta;\text{non-PD}) - \rho(\theta;\mathrm{PD})$, $\frac{\rho(\theta;\mathrm{PD})-\rho(\theta;\text{non-PD})}{\rho(\theta;\mathrm{PD})+\rho(\theta;\text{non-PD})}$ *and* $\frac{\rho(\theta;\text{non-PD})-\rho(\theta;\mathrm{PD})}{\rho(\theta;\mathrm{PD})+\rho(\theta;\text{non-PD})}$ *as functions of* θ *(because of the log scale, negative values are not plotted)*

The red and green curves (corresponding to library members 86 and 62, selected according to the $M_{\Delta\rho}$ criterion) follow almost the same course through each of the two left-hand diagrams, and these courses are also similar across the diagrams. Initially, the courses fall below the third quartile line, and then recover to pass successively that line and the median. In the non-PD diagram, they even overtake the first quartile line, before falling back to approach it from above. In the PD diagram, they merely approach the first quartile line from below.

As can be seen from the upper left-hand diagram in the right-hand side of the Figure, these library members achieve the maximum of the criterion on which they were selected when the relevant values of $\rho(\theta;\mathrm{PD})$ are near to the PD first-quartile value, but the values of $\rho(\theta;\text{non-PD})$ are somewhat below the non-PD first-quartile value, i.e., when neither value of ρ is particularly large (in comparison to the values for other library members at the same values of θ).

The magenta and cyan traces (for the ρs for member 53 and for 16, selected according to the $M'_{\Delta\rho}$ criterion) also follow courses which are quite similar, between themselves and across the two right-hand diagrams, although these courses are quite different to ones followed by

the red and green curves.

As the $M'_{\Delta\rho}$ criterion excludes θs for a member for which both the $\rho(\theta;\text{PD})$ and $\rho(\theta;\text{non-PD})$ curves are below the dotted black line, we follow the magenta and cyan curves as θ increases only until they reach that line in both diagrams.

In both diagrams, the magenta and cyan start off above the 3rd quartile line, and then cross it, staying below it until they reach the dotted black line. In the $\rho(\theta;\text{PD})$ diagram, the divergence between them and the third quartile is not as great as it is in the $\rho(\theta;\text{non-PD})$ one. In the region just before the curves cross the dotted black line, $\frac{\rho(\theta;\text{PD})-\rho(\theta;\text{non-PD})}{\rho(\theta;\text{PD})+\rho(\theta;\text{non-PD})}$ reaches its maximum for both members 53 and 16, so again the maxima are reached where $\rho(\theta;\text{PD})$ and $\rho(\theta;\text{non-PD})$ are not that great for the the members concerned in comparison to their values for other members.

In Figure 3.27 we show the canonical representatives of the best two ECWTs under each criteria.

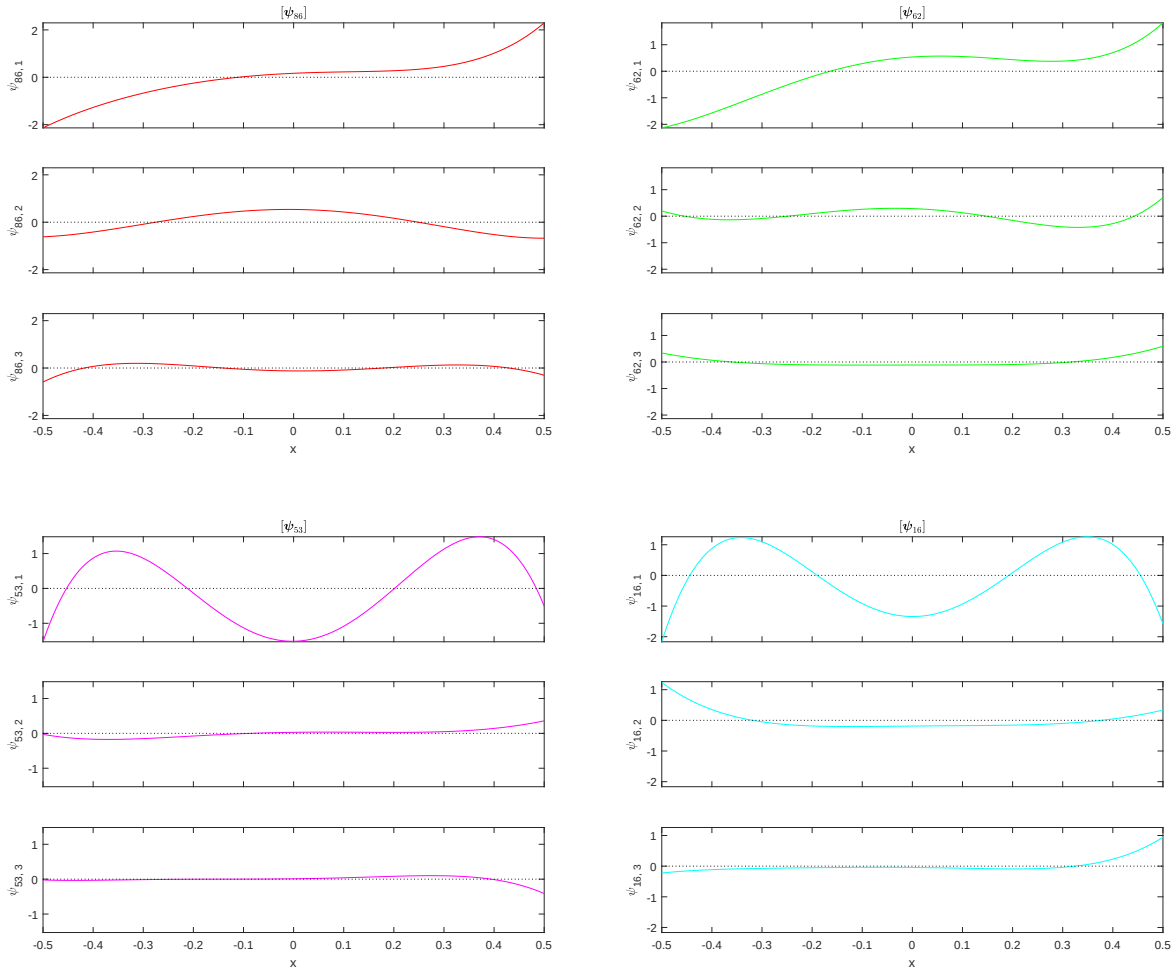


Figure 3.27: *The canonical representatives of the selected ECWTs, 86 (top left), 62 (top right), 53 (bottom left) and 16 (bottom right)*

Although member 86 and member 62 are fairly similar subjectively, they are far from

identical, and the same is true of member 53 and member 16. However, the pair selected according to the $M_{\Delta\rho}$ criterion are fairly different from the pair selected using the $M'_{\Delta\rho}$, again subjectively.

However, this subjective impression is confirmed by Table 3.5 (remember, the maximum squared distance between ECWTs is 2), which shows that the distances within the pairs 86 and 62, and 53 and 16, are much smaller than those between the pairs.

	Member number			
Member number	86	62	53	16
86	0	0.1633	1.1382	1.0387
62	0.1633	0	1.2971	1.0864
53	1.1382	1.2971	0	0.1851
16	1.0387	1.0864	0.1851	0

Table 3.5: *Squared distances between selected members*

Table 3.6 is the analogue of Table 3.2, but for $[\boldsymbol{\psi}_{86}]$, $[\boldsymbol{\psi}_{62}]$, $[\boldsymbol{\psi}_{53}]$ and $[\boldsymbol{\psi}_{16}]$.

	$[\boldsymbol{\psi}_j]$			
Property	$[\boldsymbol{\psi}_{86}]$	$[\boldsymbol{\psi}_{62}]$	$[\boldsymbol{\psi}_{53}]$	$[\boldsymbol{\psi}_{16}]$
epoch	25	24	1	17
window length	136	140	104	132
GoF	0.7855	0.7195	0.6954	0.6725
$\|\psi'_{j,1}\|^2$	0.8097	0.8983	0.9732	0.8671
$\|\psi'_{j,2}\|^2$	0.1690	0.0758	0.0164	0.0888
$\|\psi'_{j,3}\|^2$	0.0213	0.0259	0.0104	0.0440

Table 3.6: *Some properties of* $[\boldsymbol{\psi}_{86}]$, $[\boldsymbol{\psi}_{62}]$, $[\boldsymbol{\psi}_{53}]$ *and* $[\boldsymbol{\psi}_{16}]$

We see that $[\boldsymbol{\psi}_{86}]$ and $[\boldsymbol{\psi}_{62}]$ stem from data in neighbouring epochs of the library data, and so may relate to an episode of similar behaviour. The originating epochs of $[\boldsymbol{\psi}_{53}]$ and $[\boldsymbol{\psi}_{16}]$ are more widely separated, however.

The originating window lengths of $[\boldsymbol{\psi}_{86}]$ (136; corresponding to 3.68Hz), and $[\boldsymbol{\psi}_{62}]$ (140; 3.57Hz) are also quite similar, again pointing at the possibility that the same episode lies behind them, whereas the window lengths giving rise to $[\boldsymbol{\psi}_{53}]$ and $[\boldsymbol{\psi}_{16}]$ do not have this similarity.

From this point onwards we rebaptise $\rho(\theta_0;\text{PD})$, $\rho^{\text{select}}(\theta_0;\text{PD})$, and $\rho(\theta_0;\text{non-PD})$, $\rho^{\text{select}}(\theta_0;\text{non-PD})$.

Classification

We find the activation on windows into the two sets of test data of each of the top five library members according to one or other of the criteria of the previous section, and find the fraction

of windows where this activation exceeds the threshold, θ_0 or θ_0' found there. The ranges of window lengths considered are also the same as in the previous section.

That is, we calculate $\rho^{\text{test}}(\theta_0; \text{PD})$ and $\rho^{\text{test}}(\theta_0; \text{non-PD})$ in exactly the same way as we did $\rho^{\text{select}}(\theta_0; \text{PD})$ and $\rho^{\text{select}}(\theta_0; \text{non-PD})$, but we use the PD and non-PD test data instead of the PD and non-PD selection data.

The results are presented in Table 3.7.

	Fraction of activations above threshold for library member				
On data set	86	53	47	88	13
PD test	**0.0700**	**0.1754**	**0.0623**	**0.0682**	**0.0627**
Non-PD test	0.0299	0.0703	0.0262	0.0273	0.0280
	Fraction of activations above threshold for library member				
On data set	62	16	35	70	21
PD test	0.0032	**0.0185**	**0.0155**	**0.0186**	**0.0137**
Non-PD test	**0.0029**	0.0006	0.0006	0.0006	0.0005

Table 3.7: *Fraction of windows with exceeding the thresholds θ_0 or θ_0'. The values on the correct side of the initial decision boundaries Θ or Θ' are shown in bold*

As we hoped, we have $\rho^{\text{test}}(\theta_0; \text{PD}) > \rho^{\text{test}}(\theta_0; \text{non-PD})$ (or $\rho^{\text{test}}(\theta_0'; \text{PD}) > \rho^{\text{test}}(\theta_0'; \text{non-PD})$) for every library member we selected.

However, the use of Θ (or Θ') (see Table 3.4) as a decision boundary for the classification into PD and non-PD data is a complete failure — we do not have $\rho^{\text{test}}(\theta_0; \text{PD}) > \Theta > \rho^{\text{test}}(\theta_0; \text{non-PD})$ (or $\rho^{\text{test}}(\theta_0'; \text{PD}) > \Theta' > \rho^{\text{test}}(\theta_0';)$) in any case.

Although this example does not have any statistical significance in isolation from others, it does suggest that, given a better method of selecting decision boundaries (at any rate, choosing an analogue Θ closer to $\rho^{\text{select}}(\theta_0, \text{PD})$ than to $\rho^{\text{select}}(\theta_0; \text{non-PD})$), there is mileage in the method.

3.4 A better Θ?

In an attempt to find a better way of separating PD and non-PD data on the basis of a threshold Θ on the proportion of activations above a given θ, we will make an excursion into the statistics of confidence intervals on percentiles.

Given a sample of real numbers, $\mathcal{S}$, with N members drawn (with each member having the same probability of being drawn) from some population $\mathcal{P}$, a confidence interval at level γ for the $100p$th percentile is given by $[r_i, r_j]$, where r_i and r_j are the ith and jth ranked members of $\mathcal{S}$ arranged in increasing order, and i and j satisfy

$$B_{N,p}(j-1) - B_{N,p}(i-1) = \sum_{k=i}^{j-1} \binom{N}{k} p^k (1-p)^{n-k} \geq \gamma, \tag{3.45}$$

where $B_{N,p}$ is the cumulative binomial distribution function (Boudec[29]).

We assume that

$$\frac{i}{N} \leq p \leq \frac{j}{N}, \tag{3.46}$$

which will certainly be the case if γ is large enough.

Although symmetric confidence intervals are usually used, here we employ maximally asymmetric confidence intervals at level γ, given by $[r_1, r_{j_1}]$ and $[r_{i_N}, r_N]$, where

$$B_{N,p}(j_1 - 1) - B_{N,p}(0) = \sum_{k=1}^{j_1-1} \binom{N}{k} p^k (1-p)^{n-k} \geq \gamma \tag{3.47}$$

and

$$B_{N,p}(N-1) - B_{N,p}(i_N - 1) = \sum_{k=i_N}^{N-1} \binom{N}{k} p^k (1-p)^{n-k} \geq \gamma, \tag{3.48}$$

provided

$$\frac{1}{N}, \frac{i_N}{N} \leq p \leq \frac{j_1}{N}, 1 - \frac{1}{N}, \tag{3.49}$$

and these become the minimal lower maximally asymmetric confidence interval at level γ if

$$j_1 = \min\left\{ j \in \{\lceil Np \rceil, \lceil Np \rceil + 1, \ldots, N\} : \sum_{k=1}^{j-1} \binom{N}{k} p^k (1-p)^{N-k} \geq \gamma \right\} \tag{3.50}$$

and the minimal upper maximally asymmetric confidence interval at level γ if

$$i_N = \max\left\{ i \in \{1, 2, \ldots, \lfloor Np \rfloor\} : \sum_{k=i}^{N-1} \binom{N}{k} p^k (1-p)^{N-k} \geq \gamma \right\}, \tag{3.51}$$

as appropriate.

Given a fixed γ and samples of real numbers, $\mathcal{S}_1$, with N_1 members drawn from a population $\mathcal{P}_1$, and $\mathcal{S}_2$, with N_2 members drawn from a possibly different population $\mathcal{P}_2$, where the medians m_1 of $\mathcal{S}_1$ and m_2 of $\mathcal{S}_2$ obey $m_1 < m_2$, we would like to maximise $q \in \left[0, \frac{1}{2}\right]$ such that the minimal lower maximally asymmetric confidence interval at level γ for the $(50 + 100q)$th percentile of $\mathcal{P}_1$ has a upper endpoint equal to the lower endpoint of the minimal upper maximally asymmetric confidence interval at level γ for the $(50 - 100q)$th percentile of $\mathcal{P}_2$.

That is, we wish to find the largest q such that

$$j_1(q) = \min\left\{ j \in \left\{ \left\lceil N_1\left(q+\tfrac{1}{2}\right)\right\rceil, \left\lceil N_1\left(q+\tfrac{1}{2}\right)\right\rceil + 1, \ldots, N_1 \right\} : \sum_{k=1}^{j-1} \binom{N_1}{k} \left(\tfrac{1}{2}+q\right)^k \left(\tfrac{1}{2}-q\right)^{N_1-k} \geq \gamma \right\}, \quad (3.52)$$

$$i_{N_2}(q) = \max\left\{ i \in \left\{ 1, 2, \ldots, \left\lfloor N_2\left(q-\tfrac{1}{2}\right)\right\rfloor \right\} : \sum_{k=i}^{N_2-1} \binom{N_2}{k} \left(\tfrac{1}{2}-q\right)^k \left(\tfrac{1}{2}+q\right)^{N_2-k} \geq \gamma \right\}, \quad (3.53)$$

and $r^{(1)}_{j_1(q)} = r^{(2)}_{i_{N_1}(q)}$, where $r^{(1)}_j$ and $r^{(2)}_i$ are the jth and ith members of $\mathcal{S}_1$ and $\mathcal{S}_2$ ranked in increasing order.

However, there is no guarantee that $\mathcal{S}_1$ and $\mathcal{S}_2$ have any members in common, so the best we can do is maximise q subject to $r^{(1)}_{j_1(q)} \leq r^{(2)}_{i_{N_1}(q)}$.

Moreover, neither of our confidence intervals at level γ necessarily exist, although we assume they both do.

Consider $f_{N_1}(q,j) = \sum_{k=1}^{j-1} \binom{N_1}{k} \left(\frac{1}{2}+q\right)^k \left(\frac{1}{2}-q\right)^{N_1-k}$, $q \in \left[0, \frac{1}{2}\right]$, $j \in \{2,3,\ldots,N_1\}$. For all fixed q, $f_{N_1}(q,j)$ is obviously an increasing function of j.

We have $f_{N_1}(q,N_1) = 1 - \left(\frac{1}{2}+q\right)^{N_1} - \left(\frac{1}{2}-q\right)^{N_1}$, so $f'_{N_1}(q,N_1) = -N_1\left[\left(\frac{1}{2}+q\right)^{N_1-1} - \left(\frac{1}{2}-q\right)^{N_1-1}\right] < 0$, $\forall q \in \left(0, \frac{1}{2}\right]$, and, consequently, continuous $f_{N_1}(q,N_1)$ is a decreasing function of $q \in \left[0, \frac{1}{2}\right]$.

Also, $f'_{N_1}(q,j) = \sum_{k=1}^{j-1} \binom{N_1}{k} \left(\frac{1}{2}+q\right)^{k-1} \left(\frac{1}{2}-q\right)^{N_1-k-1} \left[k\left(\frac{1}{2}-q\right) - (N_1-k)\left(\frac{1}{2}+q\right)\right] = \sum_{k=1}^{j-1} \binom{N_1}{k} \left(\frac{1}{2}+q\right)^{k-1} \left(\frac{1}{2}-q\right)^{N_1-k-1} \left[k - \frac{1}{2}N_1 - N_1 q\right]$. If $j < \frac{1}{2}N_1 + 1$, or $q \in \left(\frac{j-1}{N_1} - \frac{1}{2}, \frac{1}{2}\right]$ and $j \geq \frac{1}{2}N_1 + 1$, all the terms in this sum are negative, and so the sum itself is negative. In the remaining case, $j \geq \frac{1}{2}N_1 + 1$ and $q \in \left[0, \frac{j-1}{N_1} - \frac{1}{2}\right]$, $f'_{N_1}(q,j) = f'_{N_1}(q,N_1) - \sum_{k=j}^{N_1-1} \binom{N_1}{k} \left(\frac{1}{2}+q\right)^{k-1} \left(\frac{1}{2}-q\right)^{N_1-k-1} \left[k - \frac{1}{2}N_1 - N_1 q\right] \leq f'_{N_1}(q,N_1) < 0$.

Consequently, $f_{N_1}(q,j)$ is a decreasing function of $q \in \left[0, \frac{1}{2}\right]$ for $j = 2,3\ldots,N_1$, so $j_1(q)$ is a nondecreasing function of q (where it exists).

Suppose there exists a least $j_0 \in \left\{ \left\lceil \frac{1}{2}N_1 \right\rceil, \left\lceil \frac{1}{2}N_1 \right\rceil + 1, \ldots, N_1 \right\}$ such that $\sum_{k=1}^{j_0-1} \binom{N_1}{k} \left(\frac{1}{2}+q\right)^k \left(\frac{1}{2}-q\right)^{N_1-k}\Big|_{q=0} = \frac{1}{2^{N_1}} \sum_{k=1}^{j_0-1} \binom{N_1}{k} \geq \gamma$. A necessary and sufficient condition for such a j_0 to exist is that $\sum_{k=1}^{N_1-1} \binom{N_1}{k} \left(\frac{1}{2}+q\right)^k \left(\frac{1}{2}-q\right)^{N_1-k}\Big|_{q=0} = 1 - \frac{1}{2^{N_1-1}} \geq \gamma$ or $N_1 \geq 1 - \log_2(1-\gamma)$.

If $\frac{1}{2^{N_1}} \sum_{k=1}^{j_0-1} \binom{N_1}{k} = \gamma$, then we can write $j_1^{-1}(j_0) = \{0\}$ and put $q_1^{(1)} = 0$, and, if $\frac{1}{2^{N_1}} \sum_{k=1}^{j_0-1} \binom{N_1}{k} > \gamma$, then, as $f_{N_1}(q,j_0)$ is a continuous decreasing function of q and $f_{N_1}\left(\frac{1}{2}, j_0\right) > 0$, there exists $q_1^{(1)} \in \left(0, \frac{1}{2}\right)$ such that $f_{N_1}\left(q_1^{(1)}, j_0\right) = \gamma$ and we can write $j_1^{-1}(j_0) = \left[0, q_1^{(1)}\right]$. In the former case, $\left[r_1^{(1)}, r_{j_0}^{(1)}\right]$ is a confidence interval at level γ for the median of $\mathcal{P}_1$, and, in the latter case, the same interval is a confidence interval at level γ for all the $(50+100q)$th percentiles of $\mathcal{P}_1$ for all $q \in \left[0, q_1^{(1)}\right]$.

If $j_0 = N_1$, we have finished, but otherwise we have $f_{N_1}\left(q_1^{(1)}, j_0+1\right) > \gamma$, and we can find $q_2^{(1)} \in \left(q_1^{(1)}, \frac{1}{2}\right)$ such that $j_1^{-1}(j_0+1) = \left(q_1^{(1)}, q_2^{(1)}\right]$, and then $\left[r_1^{(1)}, r_{j_0+1}^{(1)}\right]$ is a confidence interval at level γ for all the $(50+100q)$th percentiles of $\mathcal{P}_1$ for all $q \in \left[q_1^{(1)}, q_2^{(1)}\right]$.

Continuing in this fashion, we find that either

$$j_1^{-1}(j) = \begin{cases} \{0\}, & j = j_0; \\ \left(0, q_2^{(1)}\right]; & j = j_0+1; \\ \left(q_\ell^{(1)}, q_{\ell+1}^{(1)}\right]; & j = j_0+\ell, \quad \ell = 2, \ldots, N_1 - j_0; \\ \emptyset, & \text{otherwise,} \end{cases}$$

or

$$j_1^{-1}(j) = \begin{cases} \left[0, q_1^{(1)}\right]; & j = j_0; \\ \left(q_\ell^{(1)}, q_{\ell+1}^{(1)}\right]; & j = j_0+\ell, \quad \ell = 1, \ldots, N_1 - j_0; \\ \emptyset, & \text{otherwise,} \end{cases}$$

where $0 \le q_1^{(1)} < q_2^{(1)} < \cdots < q_{N_1-j_0+1}^{(1)} < \frac{1}{2}$, for some $j_0 \in \left\{\lceil \frac{1}{2}N_1 \rceil, \lceil \frac{1}{2}N_1 \rceil + 1, \ldots, N_1\right\}$, provided $N_1 \ge 1 - \log_2(1-\gamma)$, and dependent on whether $N_1 = 1 - \log_2(1-\gamma)$ or $N_1 > 1 - \log_2(1-\gamma)$. Obviously, $\left[r_1^{(1)}, r_{j_0+\ell-1}^{(1)}\right]$ will be a confidence interval at level γ for all the $(50+100q)$th percentiles of $\mathcal{P}_1$ for all $q \in \left[q_\ell^{(1)}, q_{\ell+1}^{(1)}\right]$.

Similarly, either

$$i_{N_2}^{-1}(i) = \begin{cases} \{0\}, & i = i_0; \\ \left(0, q_2^{(2)}\right]; & i = i_0-1; \\ \left(q_\ell^{(2)}, q_{\ell+1}^{(2)}\right]; & i = i_0-\ell, \quad \ell = 2, \ldots, i_0 - 1; \\ \emptyset, & \text{otherwise,} \end{cases}$$

or

$$i_{N_2}^{-1}(i) = \begin{cases} \left[0, q_1^{(2)}\right]; & i = i_0; \\ \left(q_\ell^{(2)}, q_{\ell+1}^{(2)}\right]; & i = i_0-\ell, \quad \ell = 2, \ldots, i_0 - 1; \\ \emptyset, & \text{otherwise,} \end{cases}$$

where $0 \le q_1^{(2)} < q_2^{(2)} < \cdots < q_{i_0}^{(2)} < \frac{1}{2}$, for some $i_0 \in \left\{1, 2, \ldots, \lfloor \frac{1}{2}N_2 \rfloor\right\}$, provided $N_2 \ge 1 - \log_2(1-\gamma)$, and dependent on whether $N_2 = 1 - \log_2(1-\gamma)$ or $N_2 > 1 - \log_2(1-\gamma)$, and $\left[r_{i_0-\ell+1}^{(2)}, r_{N_2}^{(2)}\right]$ will be a confidence interval at level γ for the $(50-100q)$th percentiles of $\mathcal{P}_2$ for all $q \in \left[q_\ell^{(2)}, q_{\ell+1}^{(2)}\right]$.

If $r_{j_0}^{(1)} > r_{i_0}^{(2)}$, all our confidence intervals at level γ for the $50+100q$th percentile of $\mathcal{P}_1$ will overlap with the corresponding confidence interval at level γ for the $50-100q$th percentile of $\mathcal{P}_2$ in more than a point, so we assume $r_{j_0}^{(1)} \le r_{i_0}^{(2)}$.

If $r_{j_0}^{(1)} = r_{i_0}^{(2)}$, the confidence interval at level γ for the median of $\mathcal{P}_1$ will overlap with the corresponding confidence interval at level γ for the median of $\mathcal{P}_2$ in a single point, but the

confidence intervals at level γ for positive q will overlap in more than a point, so we assume $r_{j_0}^{(1)} < r_{i_0}^{(2)}$.

For simplicity, we will also assume that $N_1, N_2 > \log_2(1-\gamma)$, the $q_1^{(1)}, q_2^{(1)}, \ldots, q_{N_1-j_0+1}^{(1)}, q_1^{(2)}, q_2^{(2)}, \ldots, q_{i_0}^{(2)}$ are all distinct, as are the $r_{j_0}^{(1)}, r_{j_0+1}^{(1)}, \ldots, r_{N_1}^{(1)}, r_{i_0}^{(2)}, r_{i_0-1}^{(2)}, \ldots, r_1^{(2)}$. If we do not make these latter assumptions, similar results will be obtained (differing mainly in the substitution of the word "disjoint" by the phrase "overlapping in at most a point").

As q increases from 0, initially we still have $\sum_{k=1}^{j_0-1}\binom{N_1}{k}\left(\frac{1}{2}+q\right)^k\left(\frac{1}{2}-q\right)^{N_1-k}$, $\sum_{k=i_0}^{N_2-1}\binom{N_2}{k}\left(\frac{1}{2}-q\right)^k\left(\frac{1}{2}+q\right)^{N_2-k} > \gamma$, so $\left[r_1^{(1)}, r_{j_0}^{(1)}\right]$ remains a confidence interval at level γ for the $(50+100q)$th percentiles of $\mathcal{P}_1$ and $\left[r_{i_0}^{(2)}, r_{N_2}^{(2)}\right]$ remains a confidence interval at level γ for the $(50-100q)$th percentiles of $\mathcal{P}_2$.

Eventually, however, q will reach either $q_1^{(1)}$, and then $\sum_{k=1}^{j_0-1}\binom{N_1}{k}\left(\frac{1}{2}+q\right)^k\left(\frac{1}{2}-q\right)^{N_1-k} = \sum_{k=1}^{j_0-1}\binom{N_1}{k}\left(\frac{1}{2}+q_1^{(1)}\right)^k\left(\frac{1}{2}-q_1^{(1)}\right)^{N_1-k} = \gamma$, or it will reach $q_1^{(2)}$, and then $\sum_{k=i_0}^{N_2-1}\binom{N_2}{k}\left(\frac{1}{2}-q\right)^k\left(\frac{1}{2}+q\right)^{N_2-k} = \sum_{k=i_0}^{N_2-1}\binom{N_2}{k}\left(\frac{1}{2}-q_1^{(2)}\right)^k\left(\frac{1}{2}+q_1^{(2)}\right)^{N_2-k} = \gamma$. For definiteness, assume that $q_1^{(1)} < q_1^{(2)}$. Then $\left[r_1^{(1)}, r_{j_0+1}^{(1)}\right]$ becomes a confidence interval at level γ for the $(50+100q)$th percentiles of $\mathcal{P}_1$ and, initially, $\left[r_{i_0}^{(2)}, r_{N_2}^{(2)}\right]$ remains a confidence interval at level γ for the $(50-100q)$th percentiles of $\mathcal{P}_2$, and this continues until either q exceeds $q_2^{(1)}$ or it exceeds $q_1^{(2)}$. If $r_{j_0+1}^{(1)} < r_{i_0}^{(2)}$, we continue to have confidence intervals which overlap in no more than a point.

Unless $r_{N_1}^{(1)} < r_1^{(2)}$, which would contradict our assumption that $m_1 < m_2$, we will reach a situation where, for some ℓ_1 and ℓ_2, $q \in \left[q_{\ell_1}^{(1)}, q_{\ell_1+1}^{(1)}\right] \cap \left[q_{\ell_2}^{(2)}, q_{\ell_2+1}^{(2)}\right]$, $\left[r_1^{(1)}, r_{j_0+\ell_1-1}^{(1)}\right]$ is a confidence interval at level γ for the $(50+100q)$th percentile of $\mathcal{P}_1$, $\left[r_{i_0-\ell_2+1}^{(2)}, r_{N_2}^{(2)}\right]$ is a confidence interval at level γ for the $(50-100q)$th percentile of $\mathcal{P}_2$, $\left[r_1^{(1)}, r_{j_0+\ell_1-1}^{(1)}\right] \cap \left[r_{i_0-\ell_2+1}^{(2)}, r_{N_2}^{(2)}\right] = \emptyset$ and, if $q_{\ell_1+1}^{(1)} < q_{\ell_2+1}^{(2)}$ (resp. $q_{\ell_1+1}^{(1)} > q_{\ell_2+1}^{(2)}$), $\left[r_1^{(1)}, r_{j_0+\ell_1}^{(1)}\right] \cap \left[r_{i_0-\ell_2+1}^{(2)}, r_{N_2}^{(2)}\right] \neq \emptyset$ (resp. $\left[r_1^{(1)}, r_{j_0+\ell_1-1}^{(1)}\right] \cap \left[r_{i_0-\ell_2}^{(2)}, r_{N_2}^{(2)}\right] \neq \emptyset$). Thus, $q = q^* = q_{\ell_1+1}^{(1)}$ (resp. $q = q^* = q_{\ell_2+1}^{(2)}$) is the largest q for which our confidence intervals are disjoint; these are $\left[r_1^{(1)}, r_{j_0+\ell_1-1}^{(1)}\right]$ and $\left[r_{i_0-\ell_2+1}^{(2)}, r_{N_2}^{(2)}\right]$.

We can now define Θ by

$$\Theta = \begin{cases} \frac{1}{2}\left(r_{N_1}^{(1)} + r_1^{(2)}\right), & r_{N_1}^{(1)} < r_1^{(2)}; \\ \frac{1}{2}\left(r_{j_0+\ell_1-1}^{(1)} + r_{i_0-\ell_2+1}^{(2)}\right), & r_{N_1}^{(1)} > r_1^{(2)}. \end{cases} \tag{3.54}$$

Θ is a threshold which separates intervals which contain the $(50+q^*)$th percentile of the population with the lower median with probability of at least γ, and the $(50-q^*)$th percentile of the population with the upper median with probability of at least γ. Bearing in mind that, as Boudec[29, page 36] says, "*it is the confidence interval that is random, not the unknown parameter*", if we assume that the samples of the two populations are independent, the probability that both these percentiles find themselves in their confidence intervals is at

least γ^2. In other words, we have a probability of γ^2 of having a classifier which labels at least $\frac{1}{2} + q^*$ of each class correctly.

The derivation of Θ is illustrated in Figure 3.28; the "particular percentiles" mentioned in the caption are the median, the $\lfloor 50 \pm \frac{100}{3}q^* + \frac{1}{2} \rfloor$th, the $\lfloor 50 \pm \frac{200}{3}q^* + \frac{1}{2} \rfloor$th and the $(50 \pm 100q^*)$th percentiles (where $\pm = +$ corresponds to the population $\mathcal{P}_1$ and its sample $\mathcal{S}_1$ and $\pm = -$ to the population $\mathcal{P}_2$ and its sample $\mathcal{S}_2$).

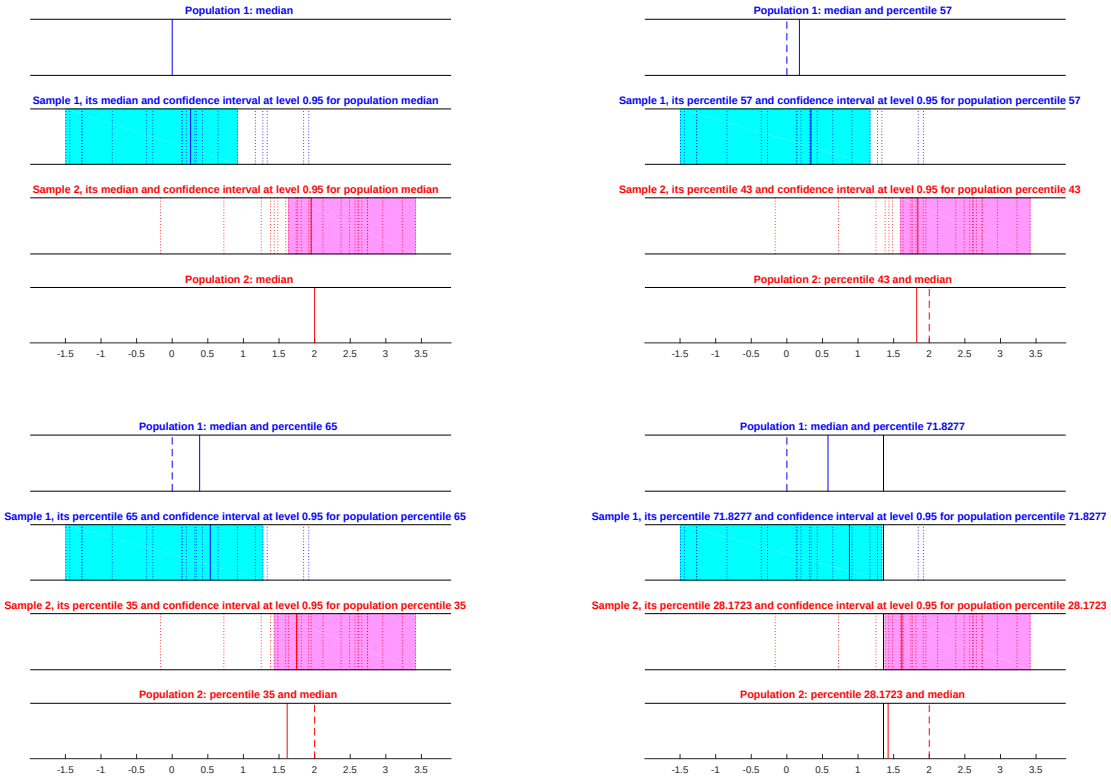


Figure 3.28: *The derivation of Θ when $\mathcal{S}_1$ consists of $N_1 = 20$ samples drawn from a normal population $\mathcal{P}_1$ with distribution $\mathcal{N}(0,1)$, $\mathcal{S}_2$ consists of $N_2 = 25$ samples drawn from a normal population $\mathcal{P}_2$ with distribution $\mathcal{N}(2,1)$ and $\gamma = 0.95$. In each of the four diagrams, the top subdiagram shows the population median and a particular percentile for $\mathcal{P}_1$ — the former is shown dashed except where it coincides with the latter, and the latter is solid. The bottom subdiagram is the same, but for $\mathcal{P}_2$, and the "reflected" percentile. The subdiagram which is second from the top (resp., bottom) shows the members of $\mathcal{S}_1$ (resp., $\mathcal{S}_2$) and, shaded in cyan (resp., magenta), the lower (resp., upper) confidence interval at level 0.99 for the particular percentile of $\mathcal{P}_1$ (resp., $\mathcal{P}_2$). Θ is represented by the vertical black line appearing in all the subdiagrams of the bottom right diagram*

3.4.1 Using the better Θ

We will use the material of the first part of this section to change the procedure of page 101 for using contrast data to select distinguishing ECWTs. Here, we will assume that the data for both the PD and control data is separated into epochs of contiguous data of fixed length, e.g., 5 minutes.

1 pick a window tolerance W;

2 for each ECWT $[\boldsymbol{\psi}]$ in the library, set $w_{[\boldsymbol{\psi}]}$ to the window length that gave rise to the data-fitted representative of $[\boldsymbol{\psi}]$ and set $\mathcal{S}_{[\boldsymbol{\psi}]} = \{w_{[\boldsymbol{\psi}]} - W, w_{[\boldsymbol{\psi}]} - W + 1, \ldots, w_{[\boldsymbol{\psi}]}, \ldots, w_{[\boldsymbol{\psi}]} + W - 1, w_{[\boldsymbol{\psi}]} + W\}$;

3 for each ECWT $[\boldsymbol{\psi}]$ and each $w \in \mathcal{S}_{[\boldsymbol{\psi}]}$, calculate $A_{w,[\boldsymbol{\psi}]}(x)$ for each x that is a centre of a (complete) window of length w into an epoch of the new PD data;

4 repeat 3 for the windows into the non-PD data;

5 find the maximum $M_A = A_{w,[\boldsymbol{\psi}]}(x)$ over $w \in \mathcal{S}_{[\boldsymbol{\psi}]}$ and x in both data sets;

6 for each $\theta \in T$, where $T \subset (0, M_A)$ is a discrete set of reasonably large cardinality, and k up to the number of epochs, find the number of (w, x) pairs with x in the kth epoch of the new PD data for which $A_{w,[\boldsymbol{\psi}]}(x) > \theta$ and find the ratio, $\rho(\theta, k; \text{PD})$ of this number to the total number of (w, x) pairs considered per epoch;

7 repeat 6 for the non-PD data, to find $\rho(\theta, k; \text{non-PD})$ analogous to $\rho(\theta, k; \text{PD})$;

8 for each fixed θ and library member m, attempt[17] to find $q^* = q^*(m, \theta)$, as above, using $\rho(\theta, k; \text{non-PD})$ in the role of $\mathcal{S}_1$ and $\rho(\theta, k; \text{PD})$ in the role of $\mathcal{S}_2$ (assuming $\text{median}(\rho(\theta, k; \text{non-PD})) < \text{median}(\rho(\theta, k; \text{PD}))$; if this is not the case, reverse the roles of $\rho(\theta, k; \text{non-PD})$ and $\mathcal{S}_1$ and $\rho(\theta, k; \text{PD})$). γ is fixed in advanced, and is a parameter of the method;

9 for each library member, find $q^{**}(m) = \max_\theta q^*(\theta, m)$, where the maximum is taken over θ such that $\text{median}(\rho(\theta, k; \text{non-PD})) < \text{median}(\rho(\theta, k; \text{PD}))$ and $q^*(m, \theta)$ is defined. If the set of such θ is empty, find the maximum over θ such that $\text{median}(\rho(\theta, k; \text{non-PD})) > \text{median}(\rho(\theta, k; \text{PD}))$ and $q^*(m, \theta)$ is defined. If this set is also empty, do not define $q^{**}(m)$;

10 if M library members are desired, we take the M members m of the subset of the library with $\{\theta \in \text{median}(\rho(\theta, k; \text{non-PD})) < \text{median}(\rho(\theta, k; \text{PD})), q^*(m, \theta) \text{ defined}\} \neq \emptyset$ with the greatest values of $q^{**}(m)$, and, if this yields insufficiently many members, we take the remaining library members such that $q^{**}(m)$ is defined and is as large as possible. Finally, if we still do not have enough library members, we run the procedure again, but with a reduced value of γ, for the remaining members of the library. We continue in this fashion, until we have enough members for our purposes, or we have reduced γ below a value we think is useful.

(In the case that the median of $\mathcal{S}_1$ exceeded that of $\mathcal{S}_2$, we exchanged the roles of $\mathcal{S}_1$ and $\mathcal{S}_2$. A provisional interpretation of this case is that a normal pattern of activity has been picked up from the PD data, but that this is actually attenuated with respect to its occurrence in the control data.)

[17] as noted above, it is not always possible to do this — when it is not, drop the particular θ from consideration

3.5 Summary

In this Chapter, we have developed the machinery for creating an enormous library of ECWTs, representing patterns of motion typical of PD subjects (not excluding normal patterns which may be present). This includes the definition of a distance between ECWTs which is not dependent on rotations or reflections of the data to which the original representatives of the ECWTs were fitted. This independence is obviously necessary, as the ECWTs are also independent of rotation and reflection. In addition, where there are large parts of pairs of ECWTs which would nearly coincide if one was shifted in time, this is taken into account and the distance is reduced[18].

Once we have this enormous library, we need to trim it down to a manageable size. We do not wish to have a library containing too many equivalence classes as representatives of patterns of motion which are nearly the same, so, ideally, we would obtain a smaller library with members which represented clusters[19] in the larger, original one (with respect to our distance).

Obtaining this directly would be computationally prohibitively expensive, requiring many evaluations of our distance function (which is not that cheap, for a fairly atomic function), so we tried to use the k-medoids[20] function built into Matlab®, letting the centres this function returns constitute the smaller library.

As this also turned out to be computationally too expensive, we devised an hierarchical version of k-medoids, which weights cluster centres returned at each stage according to the number of members of the original set in its cluster at the previous stage, before clustering these. (At every stage, a member of the original set is said to be in a cluster of clusters if it is in one of the lower level clusters.)

(In this process, we took no account of the quality of the clustering, and we set the number of clusters on an *ad hoc* basis, so as to produce roughly the required number of members of the final library, also chosen on an *ad hoc* basis. If the work of this Thesis is continued, it may be worthwhile to look at the cluster quality. If work on an hierarchical version of k-medoids is pursued, the same thing applies.)

Finally, we ranked the members of the smaller library according to their distances to a contrast set and to further data from the original set, using a criterion based on the p-level that would have been returned from a Wilcoxon ranked-sum test, producing a way to conduct classification experiments later in the Thesis.

[18] This is a slight simplification: the similarity concerns representatives of the equivalence classes, and an extremisation over choice of these representatives and the time shift

[19] In the sense that the intra-cluster distance sums were minimised

[20] Mediods, not means, as we wished the centres to be members of the larger library

Chapter 4

Effect of accelerometer bandwidth

As our accelerometer has a bandwidth of 80Hz, we wish to explore the effect of this on the task of distinguishing different continuous wavelets. We also propose a strategy for ameliorating this effect: as a lack of smoothness in a function implies that it has a higher high frequency content, we impose smoothness constraints on our wavelets.

We consider (and define) wec (wrapped endpoint continuity, given by equation 4.67) and wed (wrapped endpoint differentiability, wec and equation 4.68) constraints in detail below, but actually work with ec (endpoint continuity, equation 4.119) and ed (endpoint differentiability, ec and equation 4.120) in subsequent chapters. There are two chain of implication for these conditions: ed $\Rightarrow$ ec $\Rightarrow$ wec and ed $\Rightarrow$ wed $\Rightarrow$ wec.

We refer to the conditions ue and zi alone as the plain condition set.

As a subsidiary hypothesis, we propose:

increasing the strictness of the conditions set on smoothness improves the results of applying the method,

or, more specifically:

1 results under the ec condition set will be an improvement on those under the plain condition set;

2 results under the ed condition set will be an improvement on those under the ec condition set;

4.1 Model of bandwidth limitations

We will model the effect of using a limited bandwidth (2β) accelerometer on a one-component wavelet ψ by replacing ψ by ψ_β, where

$$\psi_\beta(x) = \frac{1}{\sqrt{2\pi}} \int_{-\beta}^{\beta} \mathrm{e}^{i\xi x} \hat{\psi}(y)\, \mathrm{d}\xi = \frac{1}{2\pi} \int_{-\beta}^{\beta} \mathrm{e}^{i\xi x} \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathrm{e}^{-i\xi y} \psi(y)\, \mathrm{d}y\, \mathrm{d}\xi, \tag{4.1}$$

that is, we simply truncate the Fourier transform of ψ at $\pm\beta$ before applying the inverse transform, to remove the higher frequency components.

If $\psi(x) = \sum_{k=0}^{n} a_k x^k$, $\psi'(x) = \sum_{k=0}^{n} a'_k x^k$ on $\left[-\frac{1}{2}, \frac{1}{2}\right)$ and ψ, ψ' are zero elsewhere, then

$$\begin{aligned}
\psi_\beta(x) &= \frac{1}{2\pi} \sum_{k=0}^{n} a_k \int_{-\beta}^{\beta} \mathrm{e}^{i\xi x} \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathrm{e}^{-i\xi y} y^k \, \mathrm{d}y \, \mathrm{d}\xi \\
&= \frac{1}{2\pi} \sum_{k=0}^{n} a_k \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\beta}^{\beta} \mathrm{e}^{i\xi(x-y)} y^k \, \mathrm{d}\xi \, \mathrm{d}y \\
&= \frac{1}{\pi} \sum_{k=0}^{n} a_k \int_{-\frac{1}{2}}^{\frac{1}{2}} y^k \frac{\sin \beta(x-y)}{x-y} \, \mathrm{d}y, \qquad (4.2)
\end{aligned}$$

where the interchange of the order of integration is justified by the uniform convergence of the double integral of the absolute value of the integrand over the given region of integration (see, e.g., Copson[44]).

Also

$$\begin{aligned}
\langle \psi_\beta, \psi'_\beta \rangle &= \langle \hat{\psi}_\beta, \hat{\psi}'_\beta \rangle = \int_{-\infty}^{\infty} \hat{\psi}_\beta(\xi) \hat{\psi}'_\beta(\xi) \mathrm{d}\xi = \int_{-\beta}^{\beta} \hat{\psi}_\beta(\xi) \hat{\psi}'_\beta(\xi) \mathrm{d}\xi = \int_{-\beta}^{\beta} \hat{\psi}_\beta(\xi) \hat{\psi}'(\xi) \mathrm{d}\xi \\
&= \langle \hat{\psi}_\beta, \hat{\psi}' \rangle = \langle \psi_\beta, \psi' \rangle \\
&= \frac{1}{\pi} \sum_{k=0}^{n} a_k \int_{-\frac{1}{2}}^{\frac{1}{2}} x^k \sum_{\ell=0}^{n} a'_\ell \int_{-\frac{1}{2}}^{\frac{1}{2}} u^\ell \frac{\sin \beta(x-u)}{x-u} \, \mathrm{d}u \, \mathrm{d}x \\
&= \frac{1}{\pi} \sum_{k=0}^{n} \sum_{\ell=0}^{n} a_k a'_\ell \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} x^k u^\ell \frac{\sin \beta(x-u)}{x-u} \, \mathrm{d}u \, \mathrm{d}x = a^{\mathrm{T}} P(\beta) a', \qquad (4.3)
\end{aligned}$$

where P is given by

$$P_{k\ell}(\beta) = \frac{1}{\pi} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} x^k u^\ell \frac{\sin \beta(x-u)}{x-u} \, \mathrm{d}u \, \mathrm{d}x, \qquad (4.4)$$

and $\langle \hat{\psi}, \hat{\psi}' \rangle = \langle \psi, \psi' \rangle$ is just Parseval's Theorem.

4.2 Value for small k, ℓ

Clearly, $P_{k\ell}(\beta) = P_{\ell k}(\beta)$ and $P_{k\ell}(\beta) = 0$ unless k and ℓ are either both even or are both odd, so we assume that $k + \ell$ is even.

By Maple®calculations

$$\begin{aligned}
P_{00}(\beta) &= \frac{2}{\pi}\left(\mathrm{Si}(\beta)-\frac{1}{\beta}+\frac{1}{\beta}\cos\beta\right),\\
P_{02}(\beta) &= \frac{1}{6\pi}\left[\mathrm{Si}(\beta)-\left(\frac{3}{\beta}-\frac{8}{\beta^3}\right)+\left(1-\frac{8}{\beta^2}\right)\frac{\cos\beta}{\beta}-\frac{2}{\beta}\frac{\sin\beta}{\beta}\right],\\
P_{11}(\beta) &= \frac{1}{6\pi}\left[\mathrm{Si}(\beta)-\left(\frac{3}{\beta}+\frac{4}{\beta^3}\right)+\left(1+\frac{4}{\beta^2}\right)\frac{\cos\beta}{\beta}+\frac{4}{\beta}\frac{\sin\beta}{\beta}\right],\\
P_{04}(\beta) &= \frac{1}{40\pi}\left[\mathrm{Si}(\beta)-\left(\frac{5}{\beta}-\frac{80}{\beta^3}+\frac{384}{\beta^5}\right)+\left(1-\frac{32}{\beta^2}+\frac{384}{\beta^4}\right)\frac{\cos\beta}{\beta}-\left(\frac{4}{\beta}-\frac{144}{\beta^3}\right)\frac{\sin\beta}{\beta}\right],\\
P_{13}(\beta) &= \frac{1}{40\pi}\left[\mathrm{Si}(\beta)-\left(\frac{5}{\beta}-\frac{20}{\beta^3}-\frac{96}{\beta^5}\right)+\left(1+\frac{28}{\beta^2}-\frac{96}{\beta^4}\right)\frac{\cos\beta}{\beta}+\left(\frac{6}{\beta}-\frac{96}{\beta^3}\right)\frac{\sin\beta}{\beta}\right],\\
P_{22}(\beta) &= \frac{1}{40\pi}\left[\mathrm{Si}(\beta)-\left(\frac{5}{\beta}+\frac{64}{\beta^5}\right)+\left(1-\frac{32}{\beta^2}+\frac{64}{\beta^4}\right)\frac{\cos\beta}{\beta}-\left(\frac{4}{\beta}-\frac{64}{\beta^3}\right)\frac{\sin\beta}{\beta}\right]. \qquad (4.5)
\end{aligned}$$

4.3 Iterative relation

As

$$\begin{aligned}
P_{k\ell}(\beta) &= \frac{1}{\pi}\int_{-\frac12}^{\frac12}\int_{-\frac12}^{\frac12} x^{k-1}(x-u+u)u^{\ell}\frac{\sin\beta(x-u)}{x-u}\,\mathrm{d}u\,\mathrm{d}x\\
&= P_{k-1,\ell+1}(\beta)+\frac{1}{\pi}\int_{-\frac12}^{\frac12}\int_{-\frac12}^{\frac12} x^{k-1}u^{\ell}\sin\beta(x-u)\,\mathrm{d}u\,\mathrm{d}x, \qquad (4.6)
\end{aligned}$$

evaluating $\Delta_{k\ell}(\beta)=\frac{1}{\pi}\int_{-\frac12}^{\frac12}\int_{-\frac12}^{\frac12} x^{k-1}u^{\ell}\sin\beta(x-u)\,\mathrm{d}u\,\mathrm{d}x$ will enable us to evaluate $P_{k\ell}(\beta)$ *via*

$$\begin{aligned}
P_{k\ell}(\beta) &= P_{k-1,\ell+1}(\beta)+\Delta_{k\ell}(\beta)\\
&= P_{k-2,\ell+2}(\beta)+\Delta_{k\ell}(\beta)+\Delta_{k-1,\ell+1}(\beta)\\
&= P_{0,k+\ell}(\beta)+\sum_{s=0}^{k-1}\Delta_{k-s,\ell+s}(\beta), \qquad (4.7)
\end{aligned}$$

provided we know $P_{0,k+\ell}(\beta)$.

But

$$\begin{aligned}
\Delta_{k\ell}(\beta) &= \frac{1}{\pi}\int_{-\frac12}^{\frac12} x^{k-1}\sin\beta x\,\mathrm{d}x\int_{-\frac12}^{\frac12} u^{\ell}\cos\beta u\,\mathrm{d}u-\frac{1}{\pi}\int_{-\frac12}^{\frac12} x^{k-1}\cos\beta x\,\mathrm{d}x\int_{-\frac12}^{\frac12} u^{\ell}\sin\beta u\,\mathrm{d}u\\
&= \begin{cases} \frac{1}{\pi}\int_{-\frac12}^{\frac12} x^{k-1}\sin\beta x\,\mathrm{d}x\int_{-\frac12}^{\frac12} u^{\ell}\cos\beta u\,\mathrm{d}u, & k,\,\ell \text{ both even};\\ -\frac{1}{\pi}\int_{-\frac12}^{\frac12} x^{k-1}\cos\beta x\,\mathrm{d}x\int_{-\frac12}^{\frac12} u^{\ell}\sin\beta u\,\mathrm{d}u, & k,\,\ell \text{ both odd}, \end{cases} \qquad (4.8)
\end{aligned}$$

where we assume that k and ℓ are both even or both odd, as $P_{k-s,\ell+s}$, $\Delta_{k-s,\ell+s}$ all vanish otherwise.

By Gradshteyn and Ryzhik[45],

$$\begin{aligned}
\int_{-\frac{1}{2}}^{\frac{1}{2}} u^m \sin\beta u \,\mathrm{d}u &= m!\left[-\cos\beta u \sum_{p=0}^{\frac{1}{2}(m-1)} \frac{(-1)^p u^{m-2p}}{(m-2p)!\beta^{2p+1}} + \right. \\
&\qquad \left. \sin\beta u \sum_{p=0}^{\frac{1}{2}(m-1)} \frac{(-1)^p u^{m-2p-1}}{(m-2p-1)!\beta^{2p+2}}\right]_{-\frac{1}{2}}^{\frac{1}{2}} \\
&= \frac{m!}{2^{m-1}}\left[-\cos\tfrac{1}{2}\beta \sum_{p=0}^{\frac{1}{2}(m-1)} \frac{2^{2p}(-1)^p}{(m-2p)!\beta^{2p+1}} + \right. \\
&\qquad \left. \sin\tfrac{1}{2}\beta \sum_{p=0}^{\frac{1}{2}(m-1)} \frac{2^{2p+1}(-1)^p}{(m-2p-1)!\beta^{2p+2}}\right] \\
&= C_m(\beta)\cos\beta + S_m(\beta)\sin\beta \qquad (4.9)
\end{aligned}$$

for m odd, and

$$\begin{aligned}
\int_{-\frac{1}{2}}^{\frac{1}{2}} u^m \cos\beta u \,\mathrm{d}u &= m!\left[\cos\beta u \sum_{p=0}^{\frac{1}{2}m-1} \frac{(-1)^p u^{m-2p-1}}{(m-2p-1)!\beta^{2p+2}} + \right. \\
&\qquad \left. \sin\beta u \sum_{p=0}^{\frac{1}{2}m} \frac{(-1)^p u^{m-2p}}{(m-2p)!\beta^{2p+1}}\right]_{-\frac{1}{2}}^{\frac{1}{2}} \\
&= \frac{m!}{2^{m-1}}\left[\cos\tfrac{1}{2}\beta \sum_{p=0}^{\frac{1}{2}m-1} \frac{2^{2p+1}(-1)^p}{(m-2p-1)!\beta^{2p+2}} + \right. \\
&\qquad \left. \sin\tfrac{1}{2}\beta \sum_{p=0}^{\frac{1}{2}m} \frac{2^{2p}(-1)^p}{(m-2p)!\beta^{2p+1}}\right] \\
&= C_m(\beta)\cos\beta + S_m(\beta)\sin\beta \qquad (4.10)
\end{aligned}$$

for $m > 0$ even, where

$$C_m(\beta) = \begin{cases} 0, & m = 0; \\ \frac{m!}{2^{m-1}} \sum_{p=0}^{\frac{1}{2}m-1} \frac{2^{2p+1}(-1)^p}{(m-2p-1)!\beta^{2p+2}}, & m > 0 \text{ even}; \\ -\frac{m!}{2^{m-1}} \sum_{p=0}^{\frac{1}{2}(m-1)} \frac{2^{2p}(-1)^p}{(m-2p)!\beta^{2p+1}}, & m \text{ odd}, \end{cases} \qquad (4.11)$$

and

$$S_m(\beta) = \begin{cases} \frac{m!}{2^{m-1}} \sum_{p=0}^{\frac{1}{2}m} \frac{2^{2p}(-1)^p}{(m-2p)!\beta^{2p+1}} & m \text{ even}; \\ \frac{m!}{2^{m-1}} \sum_{p=0}^{\frac{1}{2}(m-1)} \frac{2^{2p+1}(-1)^p}{(m-2p-1)!\beta^{2p+2}} & m \text{ odd}, \end{cases} \qquad (4.12)$$

where C_0 and S_0 are such that $\int_{-\frac{1}{2}}^{\frac{1}{2}} \cos\beta u \,\mathrm{d}u = C_0(\beta)\cos\beta + S_0(\beta)\sin\beta$.

Then, if k and ℓ are both even, equation (4.8) shows that

$$\begin{aligned}\Delta_{k\ell}(\beta) &= \tfrac{1}{\pi}[C_{k-1}(\beta)\cos\tfrac{1}{2}\beta + S_{k-1}(\beta)\sin\tfrac{1}{2}\beta][C_\ell(\beta)\cos\tfrac{1}{2}\beta + S_\ell(\beta)\sin\tfrac{1}{2}\beta] \\ &= \Delta^{(\mathrm{u})}_{k\ell}(\beta) + \Delta^{(\mathrm{c})}_{k\ell}(\beta)\cos\beta + \Delta^{(\mathrm{s})}_{k\ell}(\beta)\sin\beta \end{aligned} \tag{4.13}$$

where

$$\begin{aligned}\Delta^{(\mathrm{u})}_{k\ell}(\beta) &= \tfrac{1}{2\pi}[C_{k-1}(\beta)C_\ell(\beta) + S_{k-1}(\beta)S_\ell(\beta)], \\ \Delta^{(\mathrm{c})}_{k\ell}(\beta) &= \tfrac{1}{2\pi}[C_{k-1}(\beta)C_\ell(\beta) - S_{k-1}(\beta)S_\ell(\beta)], \\ \Delta^{(\mathrm{s})}_{k\ell}(\beta) &= \tfrac{1}{2\pi}[C_{k-1}(\beta)S_\ell(\beta) + S_{k-1}(\beta)C_\ell(\beta)]. \end{aligned} \tag{4.14}$$

If k and ℓ are both odd, we have

$$\Delta_{k\ell}(\beta) = -\Delta^{(\mathrm{u})}_{k\ell}(\beta) - \Delta^{(\mathrm{c})}_{k\ell}(\beta)\cos\beta - \Delta^{(\mathrm{s})}_{k\ell}(\beta)\sin\beta \tag{4.15}$$

instead.

We now need P_{k0} to use equation (4.7), and

$$\begin{aligned}P_{k0}(\beta) &= \frac{1}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} x^k \frac{\sin\beta(x-u)}{x-u}\,\mathrm{d}u\,\mathrm{d}x = \frac{1}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{x-\frac{1}{2}}^{x+\frac{1}{2}} x^k \frac{\sin\beta v}{v}\,\mathrm{d}v\,\mathrm{d}x \\ &= \frac{1}{\pi}\left(\int_{-1}^{0}\int_{-\frac{1}{2}}^{v+\frac{1}{2}} + \int_{0}^{1}\int_{v-\frac{1}{2}}^{\frac{1}{2}}\right) x^k \frac{\sin\beta v}{v}\,\mathrm{d}x\,\mathrm{d}v \\ &= \frac{1+(-1)^k}{\pi}\int_0^1\int_{v-\frac{1}{2}}^{\frac{1}{2}} x^k\frac{\sin\beta v}{v}\,\mathrm{d}x\,\mathrm{d}v \end{aligned} \tag{4.16}$$

where we have made the substitution $u = v - x$, changed the order of integration, and reflected x and v under one of the double integrals. Since $P_{k0}(\beta) = 0$ if k is odd, we assume that k is even. As we already have the expression in equation (4.5) for P_{00}, we may also assume that $k > 0$. Then, we have

$$\begin{aligned}P_{k0}(\beta) &= \frac{2}{(k+1)\pi}\int_0^1\left\{\left(\frac{1}{2}\right)^{k+1} - \left(v-\frac{1}{2}\right)^{k+1}\right\}\frac{\sin\beta v}{v}\,\mathrm{d}v \\ &= \frac{2}{(k+1)\pi}\int_0^1\left\{\left(\frac{1}{2}\right)^{k+1} - \sum_{s=0}^{k+1}\binom{k+1}{s}\frac{(-1)^{s+1}v^s}{2^{k+1-s}}\right\}\frac{\sin\beta v}{v}\,\mathrm{d}v \\ &= P^{(0)}_{k0}(\beta) + P^{(1)}_{k0}(\beta), \end{aligned} \tag{4.17}$$

where

$$P^{(0)}_{k0}(\beta) = \frac{1}{2^{k-1}(k+1)\pi}\int_0^1\frac{\sin\beta v}{v}\,\mathrm{d}v = \frac{\mathrm{Si}(\beta)}{2^{k-1}(k+1)\pi} \tag{4.18}$$

and

$$\begin{aligned} P_{k0}^{(1)}(\beta) &= -\frac{2}{(k+1)\pi}\sum_{s=0}^{k}\binom{k+1}{s+1}\frac{(-1)^s}{2^{k-s}}\int_0^1 v^s \sin\beta v \,\mathrm{d}v \\ &= \frac{2}{(k+1)\pi}\left[-\sum_{s=0}^{\frac{1}{2}k}\binom{k+1}{2s+1}\frac{1}{2^{k-2s}}\int_0^1 v^{2s}\sin\beta v\,\mathrm{d}v + \right. \\ &\qquad \left. \sum_{s=0}^{\frac{1}{2}k-1}\binom{k+1}{2s+2}\frac{1}{2^{k-2s-1}}\int_0^1 v^{2s+1}\sin\beta v\,\mathrm{d}v\right] \\ &= P_{k0}^{(\mathrm{e})}(\beta) + P_{k0}^{(\mathrm{o})}(\beta). \end{aligned} \tag{4.19}$$

The definition of $P_{k0}^{(\mathrm{o})}(\beta)$ is obvious given

$$\begin{aligned} P_{k0}^{(\mathrm{e})}(\beta) &= -\frac{2}{(k+1)\pi}\sum_{s=0}^{\frac{1}{2}k}\binom{k+1}{2s+1}\frac{1}{2^{k-2s}}\int_0^1 v^{2s}\sin\beta v\,\mathrm{d}v \\ &= -\frac{2}{(k+1)\pi}\sum_{s=0}^{\frac{1}{2}k}\binom{k+1}{2s+1}\frac{(2s)!}{2^{k-2s}}\left[\cos\beta v\sum_{p=0}^{s}\frac{(-1)^{p+1}v^{2(s-p)}}{\beta^{2p+1}(2s-2p)!} + \right. \\ &\qquad \left. \sin\beta v\sum_{p=0}^{s-1}\frac{(-1)^p v^{2(s-p)-1}}{\beta^{2p+2}(2s-2p-1)!}\right]_0^1 \\ &= -\frac{k!}{2^{k-1}\pi}\sum_{s=0}^{\frac{1}{2}k}\frac{2^{2s}}{(2s+1)(k-2s)!}\left[\cos\beta\sum_{p=0}^{s}\frac{(-1)^{p+1}}{\beta^{2p+1}(2s-2p)!} + \right. \\ &\qquad \left. \sin\beta\sum_{p=0}^{s-1}\frac{(-1)^p}{\beta^{2p+2}(2s-2p-1)!} + \frac{(-1)^s}{\beta^{2s+1}}\right] \\ &= P_{k0}^{(\mathrm{u})}(\beta) + P_{k0}^{(\mathrm{ec})}(\beta)\cos\beta + P_{k0}^{(\mathrm{es})}(\beta)\sin\beta, \end{aligned} \tag{4.20}$$

where we have used Gradshteyn and Ryzhik[45] for $\int v^{2s}\sin\beta v\,\mathrm{d}v$, $P_{k0}^{(1\mathrm{ec})}$ and $P_{k0}^{(1\mathrm{es})}$ will be defined in due course, and

$$P_{k0}^{(\mathrm{u})}(\beta) = -\frac{k!}{2^{k-1}\pi\beta}\sum_{s=0}^{\frac{1}{2}k}\frac{1}{(2s+1)(k-2s)!}\left(-\frac{4}{\beta^2}\right)^s. \tag{4.21}$$

Then,

$$\begin{aligned} P_{k0}^{(\mathrm{ec})}(\beta) &= \frac{k!}{2^{k-1}\pi}\sum_{s=0}^{\frac{1}{2}k}\sum_{p=0}^{s}\frac{2^{2s}(-1)^p}{(2s+1)(k-2s)!(2s-2p)!\beta^{2p+1}} \\ &= \frac{k!}{2^{k-1}\pi\beta}\sum_{p=0}^{\frac{1}{2}k}\left[\sum_{s=0}^{\frac{1}{2}k-p}\frac{2^{2s}}{(2s+2p+1)(k-2s-2p)!(2s)!}\right]\left(-\frac{4}{\beta^2}\right)^p \end{aligned} \tag{4.22}$$

and

$$\begin{aligned} P_{k0}^{(\mathrm{es})}(\beta) &= -\frac{k!}{2^{k-1}\pi}\sum_{s=0}^{\frac{1}{2}k}\sum_{p=0}^{s-1}\frac{2^{2s}(-1)^p}{(2s+1)(k-2s)!(2s-2p-1)!\beta^{2p+2}} \\ &= -\frac{k!}{2^{k-3}\pi\beta^2}\sum_{p=0}^{\frac{1}{2}k-1}\left[\sum_{s=0}^{\frac{1}{2}k-1-p}\frac{2^{2s}}{(2s+2p+3)(k-2s-2p-2)!(2s+1)!}\right]\left(-\frac{4}{\beta^2}\right)^p. \end{aligned} \tag{4.23}$$

Also

$$\begin{aligned} P_{k0}^{(\mathrm{o})}(\beta) &= \frac{2}{(k+1)\pi}\sum_{s=0}^{\frac{1}{2}k-1}\binom{k+1}{2s+2}\frac{1}{2^{k-2s-1}}\int_0^1 v^{2s+1}\sin\beta v\,\mathrm{d}v \\ &= \frac{2}{(k+1)\pi}\sum_{s=0}^{\frac{1}{2}k-1}\binom{k+1}{2s+2}\frac{(2s+1)!}{2^{k-2s-1}}\left[\cos\beta v\sum_{p=0}^{s}\frac{(-1)^{p+1}v^{2(s-p)+1}}{\beta^{2p+1}(2(s-p)+1)!}+\right. \\ &\qquad\left.\sin\beta v\sum_{p=0}^{s}\frac{(-1)^p v^{2(s-p)}}{\beta^{2p+2}(2s-2p)!}\right]_0^1 \\ &= \frac{2}{(k+1)\pi}\sum_{s=0}^{\frac{1}{2}k-1}\binom{k+1}{2s+2}\frac{(2s+1)!}{2^{k-2s-1}}\left[\cos\beta\sum_{p=0}^{s}\frac{(-1)^{p+1}}{\beta^{2p+1}(2(s-p)+1)!}+\right. \\ &\qquad\left.\sin\beta\sum_{p=0}^{s}\frac{(-1)^p}{\beta^{2p+2}(2s-2p)!}\right] \\ &= P_{k0}^{(\mathrm{oc})}(\beta)\cos\beta + P_{k0}^{(\mathrm{os})}(\beta)\sin\beta, \end{aligned} \tag{4.24}$$

where

$$\begin{aligned} P_{k0}^{(\mathrm{oc})}(\beta) &= -\frac{k!}{2^{k-2}\pi}\sum_{s=0}^{\frac{1}{2}k-1}\sum_{p=0}^{s}\frac{2^{2s}(-1)^p}{(2s+2)(k-2s-1)!(2(s-p)+1)!\beta^{2p+1}} \\ &= -\frac{k!}{2^{k-2}\pi\beta}\sum_{p=0}^{\frac{1}{2}k-1}\left[\sum_{s=0}^{\frac{1}{2}k-1-p}\frac{2^{2s}}{(2s+2p+2)(k-2s-2p-1)!(2s+1)!}\right]\left(-\frac{4}{\beta^2}\right)^p \end{aligned} \tag{4.25}$$

and

$$\begin{aligned} P_{k0}^{(\mathrm{os})}(\beta) &= \frac{k!}{2^{k-2}\pi}\sum_{s=0}^{\frac{1}{2}k-1}\sum_{p=0}^{s}\frac{2^{2s}(-1)^p}{(2s+2)(k-2s-1)!(2s-2p)!\beta^{2p+2}} \\ &= \frac{k!}{2^{k-2}\pi\beta^2}\sum_{p=0}^{\frac{1}{2}k-1}\left[\sum_{s=0}^{\frac{1}{2}k-1-p}\frac{2^{2s}}{(2s+2p+2)(k-2s-2p-1)!(2s)!}\right]\left(-\frac{4}{\beta^2}\right)^p. \end{aligned} \tag{4.26}$$

Now,

$$\begin{aligned} P_{k0}(\beta) &= P_{k0}^{(0)}(\beta)+P_{k0}^{(\mathrm{u})}(\beta)+\left[P_{k0}^{(\mathrm{ec})}(\beta)+P_{k0}^{(\mathrm{oc})}(\beta)\right]\cos\beta+\left[P_{k0}^{(\mathrm{es})}(\beta)+P_{k0}^{(\mathrm{os})}(\beta)\right]\sin\beta \\ &= P_{k0}^{(0)}(\beta)+P_{k0}^{(\mathrm{u})}(\beta)+P_{k0}^{(\mathrm{c})}(\beta)\cos\beta+P_{k0}^{(\mathrm{s})}(\beta)\sin\beta, \end{aligned} \tag{4.27}$$

where

$$\begin{aligned} &P_{k0}^{(\mathrm{c})}(\beta)=P_{k0}^{(\mathrm{ec})}(\beta)+P_{k0}^{(\mathrm{oc})}(\beta) \\ &= \frac{k!}{2^{k-1}\pi\beta}\sum_{p=0}^{\frac{1}{2}k}\left[\sum_{s=0}^{\frac{1}{2}k-p}\frac{2^{2s}}{(2s+2p+1)(k-2s-2p)!(2s)!}\right]\left(-\frac{4}{\beta^2}\right)^p- \\ &\qquad \frac{k!}{2^{k-2}\pi\beta}\sum_{p=0}^{\frac{1}{2}k-1}\left[\sum_{s=0}^{\frac{1}{2}k-1-p}\frac{2^{2s}}{(2s+2p+2)(k-2s-2p-1)!(2s+1)!}\right]\left(-\frac{4}{\beta^2}\right)^p \\ &= \begin{cases} \frac{1}{6\pi\beta}\left(1-\frac{8}{\beta^2}\right), & k=2 \\ \frac{1}{2^{k-1}(k+1)\pi\beta}+\frac{k!}{2^{k-1}\pi\beta}\left\{\sum_{p=1}^{\frac{1}{2}k}\frac{2^{k-2p}}{(k+1)(k-2p)!}\left(-\frac{4}{\beta^2}\right)^p+\right. & \\ \qquad \left.\sum_{p=1}^{\frac{1}{2}k-1}\sum_{s=0}^{\frac{1}{2}k-1-p}\frac{2^{2s}}{(2s+1)!(k-2s-2p)!}\times\right. & k>2, \\ \qquad \left.\left[\frac{2s+1}{2s+2p+1}-\frac{2(k-2s-2p)}{2s+2p+2}\right]\left(-\frac{4}{\beta^2}\right)^p\right\}, & \end{cases} \end{aligned} \tag{4.28}$$

and

$$\begin{aligned}
P_{k0}^{(\mathrm{s})}(\beta) &= P_{k0}^{(\mathrm{es})}(\beta) + P_{k0}^{(\mathrm{os})}(\beta) \\
&= -\frac{k!}{2^{k-3}\pi\beta^2}\sum_{p=0}^{\frac{1}{2}k-1}\left[\sum_{s=0}^{\frac{1}{2}k-1-p}\frac{2^{2s}}{(2s+2p+3)(k-2s-2p-2)!(2s+1)!}\right]\left(-\frac{4}{\beta^2}\right)^p + \\
&\qquad \frac{k!}{2^{k-2}\pi\beta^2}\sum_{p=0}^{\frac{1}{2}k-1}\left[\sum_{s=0}^{\frac{1}{2}k-1-p}\frac{2^{2s}}{(2s+2p+2)(k-2s-2p-1)!(2s)!}\right]\left(-\frac{4}{\beta^2}\right)^p \\
&= \frac{k!}{2^{k-2}\pi\beta^2}\sum_{p=0}^{\frac{1}{2}k-1}\sum_{s=0}^{\frac{1}{2}k-1-p}\frac{2^{2s}}{(2s+1)!(k-2s-2p-1)!}\left\{-\frac{2(k-2s-2p-1)}{2s+2p+3} + \right. \\
&\qquad \left.\frac{2s+1}{2s+2p+2}\right\}\left(-\frac{4}{\beta^2}\right)^p.
\end{aligned} \tag{4.29}$$

We can now use equations (4.18), (4.21), (4.28) and (4.29) in equation (4.27) to calculate P_{k0}, and then equation (4.7) to calculate $P_{k\ell}$, at least sufficiently far from $\beta = 0$ to avoid numerical problems stemming from the appearance of negative powers of β in our equations.

4.4 Bounds on $P_{k\ell}(\beta)$

Jordan's inequality,

$$\tfrac{2}{\pi}\theta \le \sin\theta \le \theta, \qquad \theta \in \left[0, \tfrac{\pi}{2}\right] \tag{4.30}$$

can be used to put bounds on $P_{k\ell}(\beta)$ for $\beta \in \left[0, \frac{1}{2}\pi\right]$.

Now

$$\begin{aligned}
P_{k\ell}(\beta) &= \frac{1}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{x-\frac{1}{2}}^{x+\frac{1}{2}} x^k(x-v)^\ell\frac{\sin\beta v}{v}\mathrm{d}v\mathrm{d}x \\
&= \frac{1}{\pi}\left(\int_{-1}^{0}\int_{-\frac{1}{2}}^{v+\frac{1}{2}} + \int_{0}^{1}\int_{v-\frac{1}{2}}^{\frac{1}{2}}\right) x^k(x-v)^\ell\frac{\sin\beta v}{v}\mathrm{d}x\mathrm{d}v \\
&= \frac{1+(-1)^{k+\ell}}{\pi}\int_0^1\int_{v-\frac{1}{2}}^{\frac{1}{2}} x^k(x-v)^\ell\frac{\sin\beta v}{v}\mathrm{d}x\mathrm{d}v.
\end{aligned} \tag{4.31}$$

If k and ℓ are both even and $\beta \in \left[0, \frac{\pi}{2}\right]$, then the integrand in equation (4.31) is non-

negative. This means that

$$\begin{aligned}
P_{k\ell}(\beta) &\geq \frac{\beta}{4\pi^2}\int_0^1\int_{v-\frac{1}{2}}^{\frac{1}{2}} x^k(x-v)^\ell \mathrm{d}x\mathrm{d}v \\
&= \frac{\beta}{4\pi^2}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_0^{x+\frac{1}{2}} x^k(x-v)^\ell \mathrm{d}v\mathrm{d}x \\
&= \frac{4\beta}{\pi^2(\ell+1)}\int_{-\frac{1}{2}}^{\frac{1}{2}} x^k\left[-(x-v)^{\ell+1}\right]_0^{x+\frac{1}{2}}\mathrm{d}x \\
&= \frac{4\beta}{\pi^2(\ell+1)}\int_{-\frac{1}{2}}^{\frac{1}{2}} x^k\left[\frac{1}{2^{\ell+1}}+x^{\ell+1}\right]\mathrm{d}x \\
&= \frac{4\beta}{\pi^2(\ell+1)}\left[\frac{x^{k+1}}{2^{\ell+1}(k+1)}\right]_{-\frac{1}{2}}^{\frac{1}{2}} = \frac{\beta}{2^{k+\ell-1}(\ell+1)(k+1)\pi^2}. \qquad (4.32)
\end{aligned}$$

By essentially the same argument, $P_{k\ell}(\beta) \leq \frac{\beta}{2^{k+\ell}(k+1)(\ell+1)\pi}$ and, so

$$\frac{\beta}{2^{k+\ell}(k+1)(\ell+1)\pi} \geq P_{k\ell}(\beta) \geq \frac{\beta}{2^{k+\ell-1}(k+1)(\ell+1)\pi^2}. \qquad (4.33)$$

On the other hand, if k and ℓ are both odd and $\beta \in \left[0, \frac{\pi}{2}\right]$, the integrand of (4.31) is non-negative on the region of integration only when $(x,v) \in \left\{v \in \left[0,\frac{1}{2}\right], v-\frac{1}{2} \leq x \leq 0\right\} \cup \left\{v \in \left[0,\frac{1}{2}\right], v \leq x \leq \frac{1}{2}\right\}$ and is non-positive when $(x,v) \in \left\{v \in \left[0,\frac{1}{2}\right], 0 \leq x \leq v\right\} \cup \left\{v \in \left[\frac{1}{2},1\right], v-\frac{1}{2} \leq x \leq \frac{1}{2}\right\}$.

Thus, in that case,

$$\begin{aligned}
P_{k\ell}(\beta) &\geq \beta\left\{\frac{\pi}{2}\left(\int_0^{\frac{1}{2}}\int_0^v + \int_{\frac{1}{2}}^1\int_{v-\frac{1}{2}}^{\frac{1}{2}}\right) + \int_0^{\frac{1}{2}}\int_{v-\frac{1}{2}}^0 + \int_0^{\frac{1}{2}}\int_v^{\frac{1}{2}}\right\} x^k(x-v)^\ell \mathrm{d}x\mathrm{d}v \\
&= \beta\left\{\frac{\pi}{2}\int_0^{\frac{1}{2}}\int_x^{x+\frac{1}{2}} + \int_{-\frac{1}{2}}^0\int_0^{x+\frac{1}{2}} + \int_0^{\frac{1}{2}}\int_0^x\right\} x^k(x-v)^\ell \mathrm{d}v\mathrm{d}x \\
&= \frac{\beta}{\ell+1}\left\{\frac{\pi}{2}\int_0^{\frac{1}{2}} x^k\left[-(x-v)^{\ell+1}\right]_x^{x+\frac{1}{2}}\mathrm{d}x + \right. \\
&\qquad \left. \int_{-\frac{1}{2}}^0 x^k\left[-(x-v)^{\ell+1}\right]_0^{x+\frac{1}{2}}\mathrm{d}x + \int_0^{\frac{1}{2}} x^k\left[-(x-v)^{\ell+1}\right]_0^x \mathrm{d}x\right\} \\
&= \frac{\beta}{\ell+1}\left\{-\frac{\pi}{2^{\ell+2}}\int_0^{\frac{1}{2}} x^k\mathrm{d}x + \right. \\
&\qquad \left. \int_{-\frac{1}{2}}^0 x^k\left[-\frac{1}{2^{\ell+1}}\right]\mathrm{d}x + \int_{-\frac{1}{2}}^{\frac{1}{2}} x^{k+\ell+1}\mathrm{d}x\right\} \qquad (4.34)
\end{aligned}$$

4.5 The Taylor-MacLaurin series

We have that $P_{k\ell}(\beta)$ is infinitely differentiable with

$$\begin{aligned}\frac{\mathrm{d}^{2m}}{\mathrm{d}\beta^{2m}}P_{k\ell}(\beta) = P_{k\ell}^{(2m)}(\beta) &= \frac{(-1)^m}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} x^k u^\ell (x-u)^{2m-1}\sin\beta(x-u)\mathrm{d}u\mathrm{d}x,\\ \frac{\mathrm{d}^{2m+1}}{\mathrm{d}\beta^{2m+1}}P_{k\ell}(\beta) = P_{k\ell}^{(2m+1)}(\beta) &= \frac{(-1)^m}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} x^k u^\ell (x-u)^{2m}\cos\beta(x-u)\mathrm{d}u\mathrm{d}x.\end{aligned}$$

Thus, $P_{k\ell}^{(2m)}(0) = 0$ and

$$\begin{aligned}P_{k\ell}^{(2m+1)}(0) &= \frac{(-1)^m}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} x^k u^\ell (x-u)^{2m}\mathrm{d}u\mathrm{d}x\\ &= \frac{(-1)^m}{\pi}\sum_{r=0}^{2m}\binom{2m}{r}(-1)^r\int_{-\frac{1}{2}}^{\frac{1}{2}} x^{k+2m-r}\mathrm{d}x\int_{-\frac{1}{2}}^{\frac{1}{2}} u^{\ell+r}\mathrm{d}u,\end{aligned}$$

so

$$\begin{aligned}P_{2k,2\ell}^{(2m+1)}(0) &= \frac{(-1)^m}{\pi}\sum_{r=0}^{2m}(-1)^r\binom{2m}{r}\int_{-\frac{1}{2}}^{\frac{1}{2}} x^{2k+2m-r}\mathrm{d}x\int_{-\frac{1}{2}}^{\frac{1}{2}} u^{2\ell+r}\mathrm{d}u\\ &= \frac{(-1)^m}{\pi}\sum_{r=0}^{2m}(-1)^r\binom{2m}{r}\frac{1+(-1)^r}{2^{2k+2m-r+1}(2k+2m-r+1)}\frac{1+(-1)^r}{2^{2\ell+r+1}(2\ell+r+1)}\\ &= \frac{(-1)^m}{2^{2k+2\ell+2m}\pi}\sum_{r=0}^{m}\binom{2m}{2r}\frac{1}{(2k+2m-2r+1)(2\ell+2r+1)}\\ &= \frac{(-1)^m}{2^{2k+2\ell+2m+1}(m+k+\ell+1)\pi}\sum_{r=0}^{m}\binom{2m}{2r}\left[\frac{1}{2\ell+2r+1}+\frac{1}{2k+2m-2r+1}\right]\\ &= \frac{(-1)^m}{2^{2k+2\ell+2m+1}(m+k+\ell+1)\pi}\sum_{r=0}^{m}\binom{2m}{2r}\left[\frac{1}{2\ell+2r+1}+\frac{1}{2k+2r+1}\right],\end{aligned}$$

and, by Taylor's Theorem,

$$P_{2k,2\ell}(\beta) = \sum_{m=0}^{M}\frac{1}{(2m+1)!}P_{2k,2\ell}^{(2m+1)}(0)\beta^{2m+1} + P_{R,2k,2\ell,M}(\beta) = P_{2k,2\ell,M}(\beta) + P_{R,2k,2\ell,M}(\beta), \tag{4.35}$$

where

$$
\begin{aligned}
&|P_{R,2k,2\ell,M}(\beta)| \qquad (4.36)\\
&\quad\le \frac{\beta^{2M+3}}{(2M+3)!\pi}\max_{\beta'\in[0,\beta]}\left|\int_{-\frac12}^{\frac12}\int_{-\frac12}^{\frac12} x^{2k}u^{2\ell}(x-u)^{2M+2}\cos\beta'(x-u)\mathrm{d}u\mathrm{d}x\right|\\
&\quad\le \frac{\beta^{2M+3}}{(2M+3)!\pi}\max_{\beta'\in[0,\beta]}\int_{-\frac12}^{\frac12}\int_{-\frac12}^{\frac12} |x^{2k}u^{2\ell}(x-u)^{2M+2}\cos\beta'(x-u)|\mathrm{d}u\mathrm{d}x\\
&\quad\le \frac{2\beta^{2M+3}}{(2M+3)!\pi}\int_{-\frac12}^{\frac12}\int_{u}^{\frac12} x^{2k}u^{2\ell}(x-u)^{2M+2}\mathrm{d}x\mathrm{d}u\\
&\quad\le \frac{1}{2^{2(k+\ell)+1}(2M+3)!\pi}\left[\frac{1}{(2k+1)(2\ell+1)}+\right.\\
&\qquad\left.\frac{1}{2^{2(M+1)}(2k+1)[2(M+\ell)+3]}+\frac{1}{2^{2(M+1)}[2(M+k)+3](2\ell+1)}\right]\beta^{2M+3} \qquad (4.37)
\end{aligned}
$$

(using the Lagrange form for the remainder in Taylor's Theorem), as

$$
\begin{aligned}
&\int_{-\frac12}^{\frac12}\int_{u}^{\frac12} x^{2k}u^{2\ell}(x-u)^{2M+2}\mathrm{d}x\mathrm{d}u\\
&\quad= \left(\int_{-\frac12}^{0}\int_{u}^{0}+\int_{-\frac12}^{0}\int_{0}^{\frac12}+\int_{0}^{\frac12}\int_{u}^{\frac12}\right)x^{2k}u^{2\ell}(x-u)^{2M+2}\mathrm{d}x\mathrm{d}u\\
&\quad= \int_0^{\frac12}\left[\int_0^{\frac12} x^{2k}u^{2\ell}(x+u)^{2M+2}\mathrm{d}x+\left(\int_0^{u}+\int_u^{\frac12}\right)x^{2k}u^{2\ell}(x-u)^{2M+2}\mathrm{d}x\right]\mathrm{d}u\\
&\quad\le \int_0^{\frac12}\left[\int_0^{\frac12} x^{2k}u^{2\ell}\mathrm{d}x+\int_0^{u} x^{2k}u^{2(M+\ell+1)}\mathrm{d}x+\int_u^{\frac12} x^{2(M+k+1)}u^{2\ell}\mathrm{d}x\right]\mathrm{d}u\\
&\quad\le \int_0^{\frac12}\left[\int_0^{\frac12} x^{2k}u^{2\ell}\mathrm{d}x+\int_0^{\frac12}[x^{2k}u^{2(M+\ell+1)}+x^{2(M+k+1)}u^{2\ell}]\mathrm{d}x\right]\mathrm{d}u\\
&\quad= \frac{1}{2^{2(k+\ell+1)}}\left[\frac{1}{(2k+1)(2\ell+1)}\frac{1}{2^{2(M+1)}(2k+1)[2(M+\ell)+3]}+\right.\\
&\qquad\left.\frac{1}{2^{2(M+1)}[2(M+k)+3](2\ell+1)}\right]. \qquad (4.38)
\end{aligned}
$$

As $\lim_{M\to\infty}|P_{R,2k,2\ell,M}(\beta)|=0$ for every fixed β, $P_{2k,2\ell,M}(\beta)$ converges for every fixed β. Inequality (4.37) implies that $|P_{R,2k,2\ell,M}|\le\frac{7}{12\pi}\frac{\beta^{2M+3}}{(2M+3)!}$ for every k,ℓ, so, if we wish the absolute error to be bounded by ϵ for fixed β, we choose M such that $\frac{7}{12\pi}\frac{\beta^{2M+3}}{(2M+3)!}<\epsilon$.

If $\beta\le 10$ and $\epsilon=4\times10^{-15}$, then $\frac{7}{12\pi}\frac{\beta^{2M+3}}{(2M+3)!}\le\frac{7}{12\pi}\frac{10^{2M+3}}{(2M+3)!}=\frac{7}{12\pi}\frac{10^{49}}{49!}\approx 3.0525\times10^{-15}<4\times10^{-15}=\epsilon$ for $M=23$.

Thus

$$
\left|P_{2k,2\ell}(\beta)-\sum_{m=0}^{23}\tilde{P}_{2k,2\ell,2m+1}\beta^{2m+1}\right|<4\times10^{-15} \qquad (4.39)
$$

for $\beta \in [0, 10]$, where

$$\tilde{P}_{2k,2\ell,2m+1} = \frac{(-1)^m}{2^{2(m+k+\ell)+1}(2m+1)(m+k+\ell+1)\pi} \sum_{r=0}^{m} \frac{1}{(2r)!(2m-2r)!}\left[\frac{1}{2k+2r+1}+\frac{1}{2\ell+2r+1}\right]. \quad (4.40)$$

Now, if $\beta \in [0, 10]$,

$$\begin{aligned} P_{2k,2\ell}(\beta) &\approx \sum_{r=0}^{M}\sum_{m=r}^{M} \frac{(-1)^m}{2^{2(m+k+\ell)+1}(2m+1)(m+k+\ell+1)(2r)!(2m-2r)!\pi} \times \\ &\qquad \left[\frac{1}{2k+2r+1}+\frac{1}{2\ell+2r+1}\right]\beta^{2m+1} \\ &= \frac{\beta}{2^{2(k+\ell)+1}\pi}\sum_{r=0}^{M}\frac{(-1)^r\beta^{2r}}{2^{2r}(2r)!}\left[\frac{1}{2k+2r+1}+\frac{1}{2\ell+2r+1}\right]\times \\ &\qquad \sum_{m=0}^{M-r}\frac{(-1)^m\beta^{2m}}{2^{2m}(2m+2r+1)(m+r+k+\ell+1)(2m)!} \end{aligned} \quad (4.41)$$

where we have ensured that the summations contain terms with expressions like $t\frac{\left(\frac{1}{2}\beta\right)^{2m}}{(2m)!}$ which can be written as $\exp\big(2m\ln\frac{1}{2}\beta - \ln\Gamma(2m+1)\big)$ so that the Matlab® function `gammaln` $(= \ln\Gamma(\cdot))$ can be used for accuracy of evaluation.

In order to get bounds on the relative absolute error, $\frac{|P_{R,2k,2\ell,M}(\beta)|}{P_{2k,2\ell}(\beta)}$, we set $M = 0$ in equation (4.35):

$$\begin{aligned} P_{2k,2\ell}(\beta) &= \tilde{P}_{2k,2\ell,1}\beta + P_{R,2k,2\ell,0}(\beta) \\ &\geq \frac{1}{2^{2(k+\ell)+1}(k+\ell+1)\pi}\left[\frac{1}{2k+1}+\frac{1}{2\ell+1}\right]\beta - \\ &\qquad \frac{1}{6\pi}\frac{1}{2^{2(k+\ell+1)}}\left[\frac{1}{(2k+1)(2\ell+3)}+\frac{1}{(2k+3)(2\ell+1)}\right]\beta^3 \\ &= \frac{1}{2^{2(k+\ell)}(2k+1)(2\ell+1)\pi}\left[1-\frac{1}{24}\left(\frac{2k+1}{2k+3}+\frac{2\ell+1}{2\ell+3}\right)\beta^2\right]\beta, \end{aligned} \quad (4.42)$$

where we have used equation (4.40) and, instead of inequality (4.37), a tighter version holding just for $M = 0$, $|P_{R,2k,2\ell,0}(\beta)| \leq \frac{1}{6\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} x^{2k}u^{2\ell}(x-u)^2\mathrm{d}x\mathrm{d}u = \frac{1}{6\pi}\left[\frac{1}{(2k+1)(2\ell+3)}+\frac{1}{(2k+3)(2\ell+1)}\right]$. The right-hand side of inequality (4.42) is positive if $\beta^2 < 24\left(\frac{2k+1}{2k+3}+\frac{2\ell+1}{2\ell+3}\right)^{-1} \in (12, 36]$, and, if $\beta_0 \in [0, 2\sqrt{3})$, we are guaranteed that $P_{2k,2\ell}(\beta) \geq \frac{1}{2^{2(k+\ell)}(2k+1)(2\ell+1)\pi}\left[1-\frac{1}{24}\left(\frac{2k+1}{2k+3}+\frac{2\ell+1}{2\ell+3}\right)\beta_0^2\right]\beta > 0$ for $\beta \in [0, \beta_0]$.

As $\pi \in [0, 2\sqrt{3})$, we must have

$$P_{2k,2\ell}(\beta) \geq \frac{1}{2^{2(k+\ell)}(2k+1)(2\ell+1)\pi}\left[1-\frac{1}{24}\left(\frac{2k+1}{2k+3}+\frac{2\ell+1}{2\ell+3}\right)\pi^2\right]\beta > 0, \beta \in [0, \pi]. \quad (4.43)$$

Consequently,

$$\begin{aligned}\frac{|P_{R,2k,2\ell,M}(\beta)|}{P_{2k,2\ell}(\beta)} &\leq \frac{1}{2(2k+1)(2\ell+1)(2M+3)!}\left[1-\frac{1}{24}\left(\frac{2k+1}{2k+3}+\frac{2\ell+1}{2\ell+3}\right)\pi^2\right]^{-1}\times \\ &\qquad \left[1+\frac{2\ell+1}{2^{2(M+1)}[2(M+\ell)+3]}+\frac{2k+1}{2^{2(M+1)}[2(M+k)+3]}\right]\beta^{2M+2} \\ &\leq \frac{1}{2(2k+1)(2\ell+1)}\left[1-\frac{1}{24}\left(\frac{2k+1}{2k+3}+\frac{2\ell+1}{2\ell+3}\right)\pi^2\right]^{-1}\times \\ &\qquad \left[1+\frac{2k+1}{4(2k+3)}+\frac{2\ell+1}{4(2\ell+3)}\right]\frac{\beta^{2M+2}}{(2M+3)!} \\ &\leq \frac{1}{2(2k+1)(2\ell+1)}\left[1-\frac{\pi^2}{12}\right]^{-1}\frac{3\beta^{2M+2}}{2(2M+3)!}=\frac{9}{(12-\pi^2)(2k+1)(2\ell+1)}\frac{\beta^{2M+2}}{(2M+3)!}.\end{aligned} \tag{4.44}$$

Hence, if we want to bound the relative absolute error in $P_{2k,2\ell}(\beta)$ (caused by taking the Taylor series on $\beta\in[0,\pi]$) by ϵ, we need to find M such that $\frac{9}{12-\pi^2}\frac{\pi^{2M+2}}{(2M+3)!}\leq\epsilon$. If we take $\epsilon=10^{-6}$, then a suitable M is 8, as $\frac{9\pi^{18}}{19!(12-\pi^2)}\approx 5.8633\times10^{-7}$.

If we may assume that $\min_{\beta\geq\pi,(k,\ell)\in\{(k',\ell')\in\{0,1,2,\ldots,10\}^2:k'+\ell' \text{ is even}\}} P_{k\ell}(\beta)\geq 4\times10^{-9}$ (which certainly appears to be the case from the evaluation of $P_{k\ell}(\beta)$ using Matlab®'s built-in functions), then, by inequalities (4.39) and (4.44),

$$\frac{|P_{R,2k,2\ell,23}(\beta)|}{P_{2k,2\ell}(\beta)}\leq 10^{-6},\beta\in[0,10]. \tag{4.45}$$

We now deal with the case where $P_{k\ell}$ has odd indices.

$$P^{(1)}_{2k+1,2\ell+1}(0) = \frac{1}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}x^{2k+1}\mathrm{d}x\int_{-\frac{1}{2}}^{\frac{1}{2}}u^{2\ell+1}\mathrm{d}u=0,$$

$$\begin{aligned}P^{(2m+1)}_{2k+1,2\ell+1}(0) &= \frac{(-1)^m}{\pi}\sum_{r=0}^{2m}\binom{2m}{r}(-1)^r\int_{-\frac{1}{2}}^{\frac{1}{2}}x^{2k+2m+1-r}\mathrm{d}x\int_{-\frac{1}{2}}^{\frac{1}{2}}u^{2\ell+1+r}\mathrm{d}u \\ &= \frac{(-1)^m}{\pi}\sum_{r=0}^{2m}\binom{2m}{r}(-1)^r\frac{1-(-1)^r}{2^{2k+2m+2-r}(2k+2m+2-r)}\frac{1-(-1)^r}{2^{2\ell+2+r}(2\ell+2+r)} \\ &= \frac{(-1)^{m+1}}{2^{2(m+k+\ell+1)}\pi}\sum_{r=0}^{m-1}\binom{2m}{2r+1}\frac{1}{(2k+2m+1-2r)(2\ell+3+2r)} \\ &= \frac{(-1)^{m+1}}{2^{2(m+k+\ell)+3}(m+k+\ell+2)\pi}\times \\ &\qquad \sum_{r=0}^{m-1}\binom{2m}{2r+1}\left[\frac{1}{2k+2m+1-2r}+\frac{1}{2\ell+3+2r}\right] \\ &= \frac{(-1)^{m+1}}{2^{2(m+k+\ell)+3}(m+k+\ell+2)\pi}\times \\ &\qquad \sum_{r=1}^{m}\binom{2m}{2r-1}\left[\frac{1}{2k+1+2r}+\frac{1}{2\ell+1+2r}\right],m>0,\end{aligned}$$

and, again by Taylor's Theorem,

$$\begin{aligned} P_{2k+1,2\ell+1}(\beta) &= \sum_{m=1}^{M} \frac{1}{(2m+1)!} P^{(2m+1)}_{2k+1,2\ell+1}(0)\beta^{2m+1} + P_{R,2k+1,2\ell+1,M}(\beta) \\ &= P_{2k+1,2\ell+1,M}(\beta) + P_{R,2k+1,2\ell+1,M}(\beta), \end{aligned} \tag{4.46}$$

where

$$\begin{aligned} &|P_{R,2k+1,2\ell+1,M}(\beta)| \\ &\quad\leq \frac{\beta^{2M+3}}{(2M+3)!\pi} \max_{\beta'\in[0,\beta]} \left| (-1)^{M+1} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} x^{2k+1} u^{2\ell+1} (x-u)^{2M+2} \right. \\ &\qquad\qquad \left. \cos\beta'(x-u)\mathrm{d}u\mathrm{d}x \right| \\ &\quad\leq \frac{\beta^{2M+3}}{(2M+3)!\pi} \max_{\beta'\in[0,\beta]} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} |x^{2k+1} u^{2\ell+1} (x-u)^{2M+2} \cos\beta'(x-u)| \mathrm{d}u\mathrm{d}x \\ &\quad\leq \frac{2\beta^{2M+3}}{(2M+3)!\pi} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{u}^{\frac{1}{2}} |x|^{2k+1} |u|^{2\ell+1} (x-u)^{2M+2} \mathrm{d}x\mathrm{d}u \\ &\quad= \frac{1}{2^{2(k+\ell)+5}\pi} \left[\frac{1}{(k+1)(\ell+1)} + \frac{1}{2^{2(M+1)}(k+1)(M+\ell+2)} + \frac{1}{2^{2(M+1)}(M+k+2)(\ell+1)} \right] \frac{\beta^{2M+3}}{(2M+3)!}, \end{aligned} \tag{4.47}$$

as

$$\begin{aligned} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{u}^{\frac{1}{2}} |x|^{2k+1} |u|^{2\ell+1} (x-u)^{2M+2} \mathrm{d}x\mathrm{d}u \;\leq\; & \frac{1}{2^{2(k+\ell+3)}} \left[\frac{1}{(k+1)(\ell+1)} + \frac{1}{2^{2(M+1)}(k+1)(M+\ell+2)} + \right. \\ & \left. \frac{1}{2^{2(M+1)}(M+k+2)(\ell+1)} \right], \end{aligned}$$

derived in the same way as inequality (4.38).

Inequality (4.47) implies that $|P_{R,2k+1,2\ell+1,M}(\beta)| \leq \frac{5}{128\pi}\frac{\beta^{2M+3}}{(2M+3)!}$ for all k, ℓ. Putting $M = 23$, $\beta = 10$ in this gives $|P_{R,2k+1,2\ell+1,23}(\beta)| \leq \frac{5\times 10^{49}}{128(49)!\pi} \approx 2.0441 \times 10^{-16} < 4 \times 10^{-15}$ for $\beta \in [0, 10]$.

Together with inequality (4.39), this means

$$\left| P_{k,\ell}(\beta) - \sum_{m=0}^{23} \tilde{P}_{k,\ell,2m+1} \beta^{2m+1} \right| < 4 \times 10^{-15}, \tag{4.48}$$

where

$$\begin{aligned}
\tilde{P}_{2k+1,2\ell+1,2m+1} &= \begin{cases} 0, & m=0; \\ \frac{(-1)^{m+1}}{2^{2(m+k+\ell)+3}(2m+1)(m+k+\ell+2)\pi}\times \\ \quad \sum_{r=1}^{m} \frac{1}{(2r-1)!(2m-2r+1)!}\left[\frac{1}{2k+1+2r}+\frac{1}{2\ell+1+2r}\right], & m>0, \end{cases} \\
\tilde{P}_{2k,2\ell,2m+1} &= \frac{(-1)^m}{2^{2(m+k+\ell)+1}(2m+1)(m+k+\ell+1)\pi}\times \\
&\qquad \sum_{r=0}^{m} \frac{1}{(2r)!(2m-2r)!}\left[\frac{1}{2\ell+2r+1}+\frac{1}{2m+2r+1}\right],
\end{aligned}$$

or, to summarise,

$$\tilde{P}_{k,\ell,2m+1} = \begin{cases} \frac{(-1)^m}{2^{2m+k+\ell}(2m+1)(2m+k+\ell+2)\pi}\times & \\ \quad \sum_{r=0}^{m} \frac{1}{(2r)!(2m-2r)!}\left[\frac{1}{k+2r+1}+\frac{1}{\ell+2r+1}\right], & k,\ell \text{ both even;} \\ \left.\begin{array}{ll} 0, & m=0; \\ \frac{(-1)^{m+1}}{2^{2m+k+\ell}(2m+1)(2m+k+\ell+2)\pi}\times & \\ \quad \sum_{r=0}^{m-1} \frac{1}{(2r+1)!(2m-2r-1)!}\times & m>0; \\ \quad \left[\frac{1}{k+2r+2}+\frac{1}{\ell+2r+2}\right], & \end{array}\right\} & k,\ell \text{ both odd;} \\ 0, & \text{otherwise.} \end{cases} \tag{4.49}$$

As

$$\begin{aligned}
\sum_{r=0}^{m} \frac{1}{(k+2r+1)(2r)!(2m-2r)!} &= \frac{1}{2(2m)!}\sum_{r=0}^{m}\binom{2m}{r}\frac{1+(-1)^r}{k+r+1} \\
&= \frac{1}{2(2m)!}\sum_{r=0}^{m}\binom{2m}{r}[1+(-1)^r]\int_0^1 x^{k+r}\mathrm{d}x \\
&= \frac{1}{2(2m)!}\int_0^1 x^k[(1-x)^{2m}+(1+x)^{2m}]\mathrm{d}x \\
&= \frac{1}{2(2m)!}\left[\frac{\Gamma(k+1)\Gamma(2m+1)}{\Gamma(2m+k+2)}+\int_1^2 (x-1)^k x^{2m}\mathrm{d}x\right] \\
&= \frac{k!}{2(2m+k+1)!}+\frac{1}{2(2m)!}\sum_{r=0}^{k}\binom{k}{r}(-1)^r\int_1^2 x^{2m+k-r}\mathrm{d}x \\
&= \frac{k!}{2(2m+k+1)!}+ \\
&\qquad \frac{1}{2(2m)!}\sum_{r=0}^{k}\binom{k}{r}\frac{(-1)^r}{2m+k-r+1}[2^{2m+k-r+1}-1] \\
&= \frac{k!}{2(2m+k+1)!}+\frac{(-1)^k}{2(2m)!}\sum_{r=0}^{k}\binom{k}{r}\frac{(-1)^r}{2m+r+1}[2^{2m+r+1}-1],
\end{aligned} \tag{4.50}$$

and, similarly,

$$\sum_{r=0}^{m-1} \frac{1}{(k+2r+2)(2r+1)!(2m-2r-1)!} \qquad \frac{1}{2(2m)!}\sum_{r=0}^{2m}\binom{2m}{r}\frac{1-(-1)^r}{k+r+1}$$
$$= -\frac{k!}{2(2m+k+1)!} + \frac{(-1)^k}{2(2m)!}\sum_{r=0}^{k}\binom{k}{r}\frac{(-1)^r}{2m+r+1}[2^{2m+r+1}-1], \tag{4.51}$$

we can rewrite equation (4.49) as

$$\tilde{P}_{k,\ell,2m+1} = \begin{cases} \frac{(-1)^m}{2^{2m+k+\ell+1}(2m+1)(2m+k+\ell+2)\pi}\times \\ \quad\left[\frac{k!}{(2m+k+1)!}+\frac{\ell!}{(2m+\ell+1)!}+\right. \\ \quad \left.\frac{1}{(2m)!}\left\{\sum_{r=0}^{k}\binom{k}{r}\frac{(-1)^r[2^{2m+r+1}-1]}{2m+r+1}+\right.\right. \\ \quad \left.\left.\sum_{r=0}^{\ell}\binom{\ell}{r}\frac{(-1)^r[2^{2m+r+1}-1]}{2m+r+1}\right\}\right], & k,\ell \text{ both even, or } k,\ell \text{ both odd and } m>0; \\ 0, & \text{otherwise.} \end{cases} \tag{4.52}$$

To determine how many terms are necessary to bound the relative absolute error when the indices of P are odd, we need to look at equation (4.46) with $M = 1$ (note that the leading term is in β^3, not β, as is the case for even indices):

$$\begin{aligned} P_{2k+1,2\ell+1}(\beta) &= P_{2k+1,2\ell+1,1}(\beta) + P_{R,2k+1,2\ell+1,1}(\beta) \\ &\geq \tilde{P}_{2k+1,2\ell+1,3}\beta^3 - |P_{R,2k+1,2\ell+1,1}(\beta)| \\ &= \frac{1}{3\times 2^{2(k+\ell)+5}(k+\ell+3)\pi}\left[\frac{1}{2k+3}+\frac{1}{2\ell+3}\right]\beta^3 - \\ &\qquad \frac{1}{2^{2(k+\ell+4)}5!\pi}\left[\frac{1}{(k+1)(\ell+3)}+\frac{6}{(k+2)(\ell+2)}+\frac{1}{(k+3)(\ell+1)}\right]\beta^5 \\ &= \frac{\beta^3}{3\times 2^{2(k+\ell+2)}(2k+3)(2\ell+3)\pi} - \frac{1}{2^{2(k+\ell+4)}5!\pi}\left[\frac{\beta^5}{(k+1)(\ell+3)}+\frac{6}{(k+2)(\ell+2)}+\frac{1}{(k+3)(\ell+1)}\right] \\ &= \frac{\beta^3}{2^{2(k+\ell+4)}\pi}\left[\frac{16}{3(2k+3)(2\ell+3)} - \frac{1}{5!}\left(\frac{1}{(k+1)(\ell+3)}+\frac{6}{(k+2)(\ell+2)}+\frac{1}{(k+3)(\ell+1)}\right)\beta^2\right], \end{aligned} \tag{4.53}$$

where we have used equation (4.49) and the stricter version of inequality (4.47) for $M = 1$

given by

$$\begin{aligned}
|P_{R,2k+1,2\ell+1,1}(\beta)| &\leq \frac{\beta^5}{5!\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}}|x|^{2k+1}|u|^{2\ell+1}(x-u)^4\mathrm{d}x\mathrm{d}u \\
&= \frac{\beta^5}{5!\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}}|x|^{2k+1}|u|^{2\ell+1}(x^4+6u^2x^2+u^4)\mathrm{d}x\mathrm{d}u \\
&= \frac{4\beta^5}{5!\pi}\int_{0}^{\frac{1}{2}}\int_{0}^{\frac{1}{2}}x^{2k+1}u^{2\ell+1}(x^4+6u^2x^2+u^4)\mathrm{d}x\mathrm{d}u \\
&= \frac{\beta^5}{2^{2(k+\ell+4)}5!\pi}\left[\frac{1}{(k+1)(\ell+3)}+\frac{6}{(k+2)(\ell+2)}+\frac{1}{(k+3)(\ell+1)}\right],
\end{aligned}$$

so, if $\beta_0^2 < \frac{640}{(2k+3)(2\ell+3)}\left(\frac{1}{(k+1)(\ell+3)}+\frac{6}{(k+2)(\ell+2)}+\frac{1}{(k+3)(\ell+1)}\right)^{-1}$, where the greatest lower bound of the right-hand side of this inequality is $\frac{1280}{39}$, we will have $P_{2k+1,2\ell+1}(\beta) \geq \frac{1}{2^{2(k+\ell+3)}\pi}\left[\frac{4}{3(2k+3)(2\ell+3)}-\frac{1}{5!}\left(\frac{1}{(k+1)(\ell+3)}+\frac{6}{(k+2)(\ell+2)}+\frac{1}{(k+3)(\ell+1)}\right)\beta_0^2\right]\beta^3$ for $\beta \in [0,\beta_0]$. But $\pi \in \left[0, 16\sqrt{\frac{5}{39}}\right]$, so

$$\begin{aligned}
P_{2k+1,2\ell+1}(\beta) \geq\ & \frac{1}{2^{2(k+\ell+3)}\pi}\left[\frac{16}{3(2k+4)(2\ell+3)}-\right. \\
& \left.\frac{1}{5!}\left(\frac{1}{(k+1)(\ell+3)}+\frac{6}{(k+2)(\ell+2)}+\frac{1}{(k+3)(\ell+1)}\right)\pi^2\right]\beta^3 > 0, \beta \in [0,\pi].
\end{aligned} \tag{4.54}$$

Thus

$$\begin{aligned}
\frac{|P_{R,2k+1,2\ell+1,M}(\beta)|}{P_{2k+1,2\ell+1}(\beta)} &\leq 2\left[\frac{16(k+1)(\ell+1)}{3(2k+3)(2\ell+3)}-\frac{1}{5!}\left(\frac{\ell+1}{\ell+3}+\frac{6(k+1)(\ell+1)}{(k+2)(\ell+2)}+\frac{k+1}{k+3}\right)\pi^2\right]^{-1}\times \\
&\quad \left[1+\frac{\ell+1}{2^{2(M+1)}(M+\ell+2)}+\frac{k+1}{2^{2(M+1)}(M+k+2)}\right]\frac{\beta^{2M}}{(2M+3)!} \\
&\leq 2\left[\frac{16}{27}-\frac{\pi^2}{18}\right]^{-1}\left[1+\frac{1}{2^{2M+1}}\right]\frac{\beta^{2M}}{(2M+3)!} \leq \frac{108}{32-3\pi^2}\frac{\beta^{2M}}{(2M+3)!},
\end{aligned} \tag{4.55}$$

as the turning points of $\frac{8(k+1)(\ell+1)}{3(2k+3)(2\ell+3)}-\frac{\ell+1}{120(\ell+3)}\pi^2 = \frac{4(k+1)}{3(2k+3)}-\frac{\pi^2}{120}-\frac{4(k+1)}{3(2k+3)(2\ell+3)}+\frac{1}{60(\ell+3)}\pi^2$ considered as a function of ℓ are at negative values of ℓ, so

$$\begin{aligned}
\frac{8(k+1)(\ell+1)}{3(2k+3)(2\ell+3)}-\frac{\ell+1}{120(\ell+3)}\pi^2 &\geq \inf_{\ell\in[0,\infty)}\left\{\frac{4(k+1)}{3(2k+3)}-\frac{\pi^2}{120}-\frac{4(k+1)}{3(2k+3)(2\ell+3)}+\frac{1}{60(\ell+3)}\pi^2\right\} \\
&= \min\left\{\frac{8(k+1)}{9(2k+3)}-\frac{\pi^2}{360}, \frac{4(k+1)}{3(2k+3)}-\frac{\pi^2}{120}\right\} = \frac{8(k+1)}{9(2k+3)}-\frac{\pi^2}{360} \geq \frac{8}{27}-\frac{\pi^2}{360}
\end{aligned}$$

and $\frac{16(k+1)(\ell+1)}{3(2k+3)(2\ell+3)}-\frac{1}{120}\left(\frac{\ell+1}{\ell+3}+\frac{6(k+1)(\ell+1)}{(k+2)(\ell+2)}+\frac{k+1}{k+3}\right)\pi^2 \geq \frac{16}{27}-\frac{\pi^2}{180}-\frac{(k+1)(\ell+1)}{20(k+2)(\ell+2)}\pi^2 \geq \frac{16}{27}-\frac{\pi^2}{18} > 0.$

Equation (4.55) means that the absolute relative error for $\beta \in [0,\pi]$ is bounded by $\epsilon = 10^{-6}$ if $\frac{108}{32-3\pi^2}\frac{\pi^{2M}}{(2M+3)!} < 10^{-6}$, which is satisfied if $M \geq 8$, as $\frac{108}{32-3\pi^2}\frac{\pi^{16}}{19!} \approx 6.3514\times 10^{-7}$.

If we again use $\min_{\beta\geq\pi,(k,\ell)\in\{(k',\ell')\in\{0,1,2,\ldots,10\}^2:k'+\ell' \text{ is even}\}} P_{k\ell}(\beta) \geq 4\times 10^{-9}$, inequalities (4.48) and (4.55) mean that inequality (4.45) also holds when P has odd indices:

$$\frac{|P_{R,2k+1,2\ell+1,23}(\beta)|}{P_{2k+1,2\ell+1}(\beta)} \leq 10^{-6}, \beta \in [0,10]. \tag{4.56}$$

Returning to equation (4.49), we have, in particular,

$$\tilde{P}_{k,\ell,1} = \begin{cases} \frac{1}{2^{k+\ell}(k+\ell+2)\pi}\left[\frac{1}{k+1}+\frac{1}{\ell+1}\right], & k,\ell \text{ both even;} \\ 0, & \text{otherwise.} \end{cases} \tag{4.57}$$

and

$$\tilde{P}_{k,\ell,3} = \begin{cases} -\frac{1}{3\cdot 2^{k+\ell+3}(k+\ell+4)\pi}\left[\frac{1}{k+1}+\frac{1}{\ell+1}+\frac{1}{k+3}+\frac{1}{\ell+3}\right], & k,\ell \text{ both even} \\ \frac{1}{3\cdot 2^{k+\ell+2}(k+\ell+4)\pi}\left[\frac{1}{k+2}+\frac{1}{\ell+2}\right], & k,\ell \text{ both odd;} \\ 0, & \text{otherwise,} \end{cases} \tag{4.58}$$

so

$$P = \begin{bmatrix} 1 & 0 & \frac{1}{12} & 0 & \frac{1}{80} & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ \frac{1}{12} & 0 & \frac{1}{144} & 0 & \frac{1}{960} & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ \frac{1}{80} & 0 & \frac{1}{960} & 0 & \frac{1}{6400} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \frac{\beta}{\pi} + \begin{bmatrix} -\frac{1}{36} & 0 & -\frac{7}{2160} & 0 & -\frac{11}{20160} & \cdots \\ 0 & \frac{1}{432} & 0 & \frac{1}{2880} & 0 & \cdots \\ -\frac{7}{2160} & 0 & -\frac{1}{2880} & 0 & -\frac{23}{403200} & \cdots \\ 0 & \frac{1}{2880} & 0 & \frac{1}{19200} & 0 & \cdots \\ -\frac{11}{20160} & 0 & -\frac{23}{403200} & 0 & -\frac{1}{107520} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \frac{\beta^3}{\pi} \tag{4.59}$$

to $\mathrm{O}(\beta^5)$.

The evaluation form similar to equation (4.41) for $\beta \in [0, 10]$, is

$$\begin{aligned} P_{2k+1,2\ell+1}(\beta) &\approx \sum_{m=1}^{M}\sum_{r=1}^{m} \frac{(-1)^{m+1}}{2^{2(m+k+\ell)+3}(2m+1)(m+k+\ell+2)\pi} \frac{1}{(2r-1)!(2m-2r+1)!} \times \\ &\qquad \left[\frac{1}{2k+1+2r}+\frac{1}{2\ell+1+2r}\right]\beta^{2m+1} \\ &= -\frac{\beta}{2^{2(k+\ell)+3}\pi}\sum_{r=1}^{M}\frac{(-1)^r}{2^{2r}(2r-1)!}\left[\frac{1}{2k+1+2r}+\frac{1}{2\ell+1+2r}\right]\beta^{2r} \times \\ &\qquad \sum_{m=0}^{M-r}\frac{(-1)^m}{2^{2m}(2m+2r+1)(m+r+k+\ell+2)(2m+1)!}\beta^{2m}, \end{aligned} \tag{4.60}$$

The results of using equations (4.41) and (4.60) with $M = 16$ to calculate $P_{00}(\beta)$, $P_{11}(\beta)$, $P_{22}(\beta)$, $P_{02}(\beta)$, $P_{13}(\beta)$ and $P_{04}(\beta)$ for $\beta \in [10^{-5}, 30]$ are shown in Figures 4.1 and 4.2, together with the Matlab®evaluation of equations (4.5).

In Figure 4.1, we see that the difference between the Taylor series and expressions of equations (4.5) appear to be extremely small well beyond the region $\beta \in (0, 10]$, where they are designed to be small by the choice $M = 16$. However, in the diagrams for $P_{11}(\beta)$, $P_{22}(\beta)$, $P_{13}(\beta)$ and $P_{04}(\beta)$ in Figure 4.1, the vertical axis is overlain by the curve corresponding to the evaluation of equation (4.5), indicating a difference between the Taylor series and that evaluation which is relatively large.

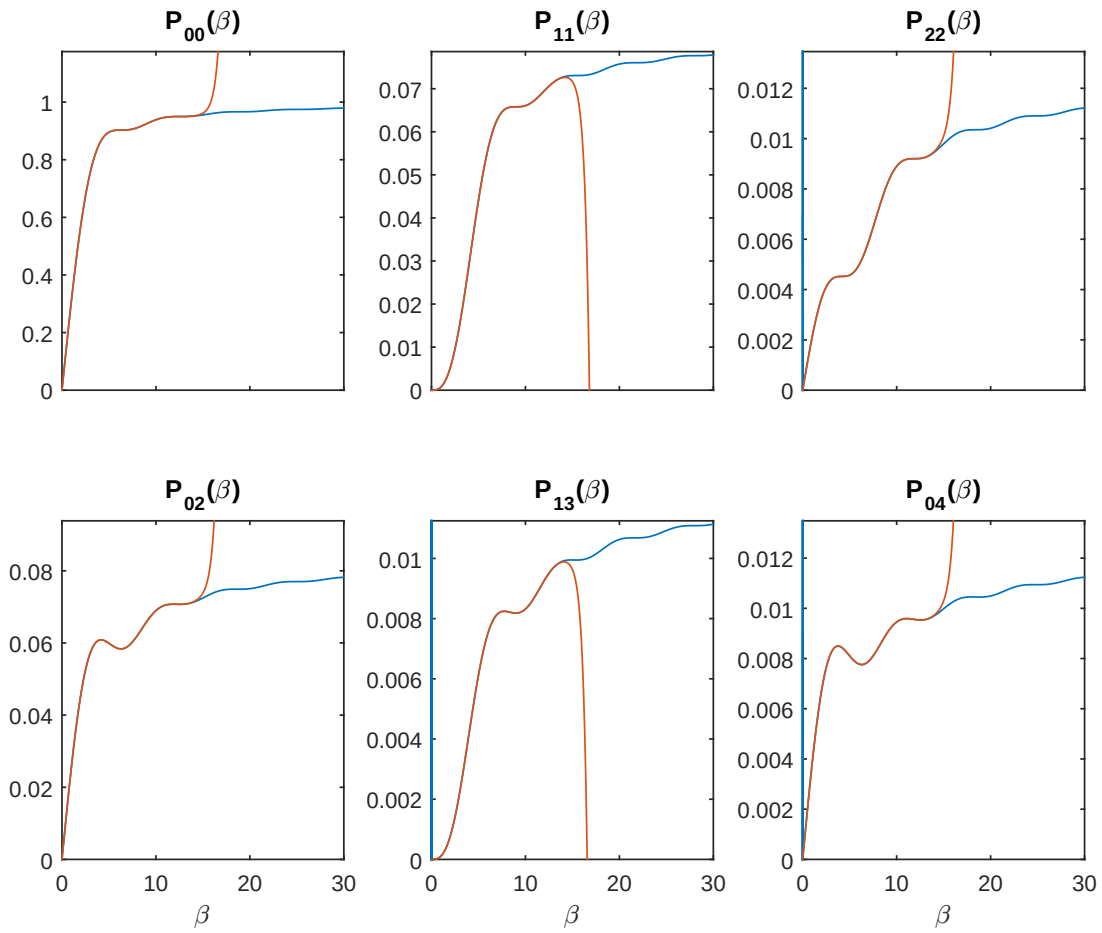


Figure 4.1: *Taylor series evaluation of $P_{k\ell}$ for small k and ℓ (red), compared with the result of using equations (4.5) (blue — or red when both curves are coincident)*

In Figure 4.2 we examine the curves of Figure 4.1 near $\beta = 0$. Except for the $P_{00}(\beta)$ curves, whose relative difference is bounded by 1.8×10^{-6}, there is a large, and oscillating, relative difference between the two ways of evaluating $P_{k\ell}$. This is due to the expression in equation (4.5) of pole-free quantities in terms of the sum of quantities which have poles at $\beta = 0$. Consequently, we will prefer to use equations (4.41) and (4.60) to evaluate $P_{k\ell}(\beta)$ for $\beta \in [0, 1]$.

This has little to do with the Matlab®evaluation of $\mathrm{Si}(\beta)$, which is compared with the evaluation of its Taylor series up to the 33rd term, $\sum_{m=0}^{16} \frac{(-1)^m \beta^{2m+1}}{(2m+1)(2m+1)!}$, in Figure 4.3. The relative error is bounded by 10^{-15}.

4.6 Expansion involving $\mathrm{Si}(\beta)$, $\sin\beta$ and $\cos\beta$

When we use expressions like (4.5) we want to be sure that they are evaluated accurately, so here we examine the asymptotic expansion for Si.

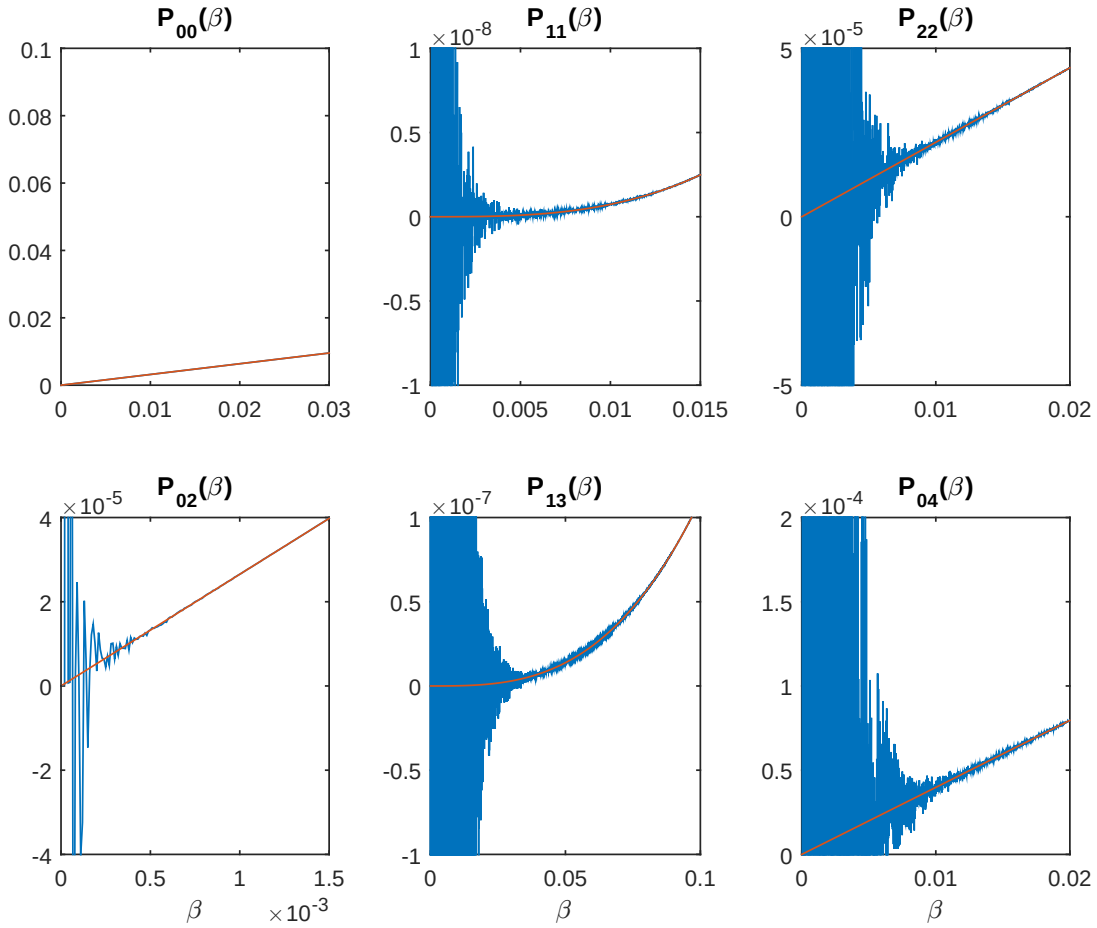


Figure 4.2: *Taylor series evaluation of $P_{k\ell}(\beta)$ near $\beta = 0$ for small k and ℓ, compared with the result of using equations (4.5)*

4.6.1 Asymptotic expansion for Si

Using $\mathrm{Si}(\beta) = \int_0^\beta \frac{\sin x}{x}\mathrm{d}x = \int_0^\infty \frac{\sin x}{x}\mathrm{d}x - \int_\beta^\infty \frac{\sin x}{x}\mathrm{d}x = \frac{\pi}{2} - \int_\beta^\infty \frac{\sin x}{x}\mathrm{d}x$ and integrating by parts, as in [44], gives

$$\begin{aligned}
\mathrm{Si}(\beta) &= \frac{\pi}{2} - \left[-\frac{\cos x}{x} - \int \frac{\cos x}{x^2}\mathrm{d}x\right]_\beta^\infty = \frac{\pi}{2} - \frac{\cos\beta}{\beta} + \int_\beta^\infty \frac{\cos x}{x^2}\mathrm{d}x \\
&= \frac{\pi}{2} - \frac{\cos\beta}{\beta} + \left[\frac{\sin x}{x^2} + 2\int \frac{\sin x}{x^3}\mathrm{d}x\right]_\beta^\infty = \frac{\pi}{2} - \frac{\cos\beta}{\beta} - \frac{\sin\beta}{\beta^2} + 2\int_\beta^\infty \frac{\sin x}{x^3}\mathrm{d}x \\
&= S_s(\beta) + R_s(\beta), \qquad (4.61)
\end{aligned}$$

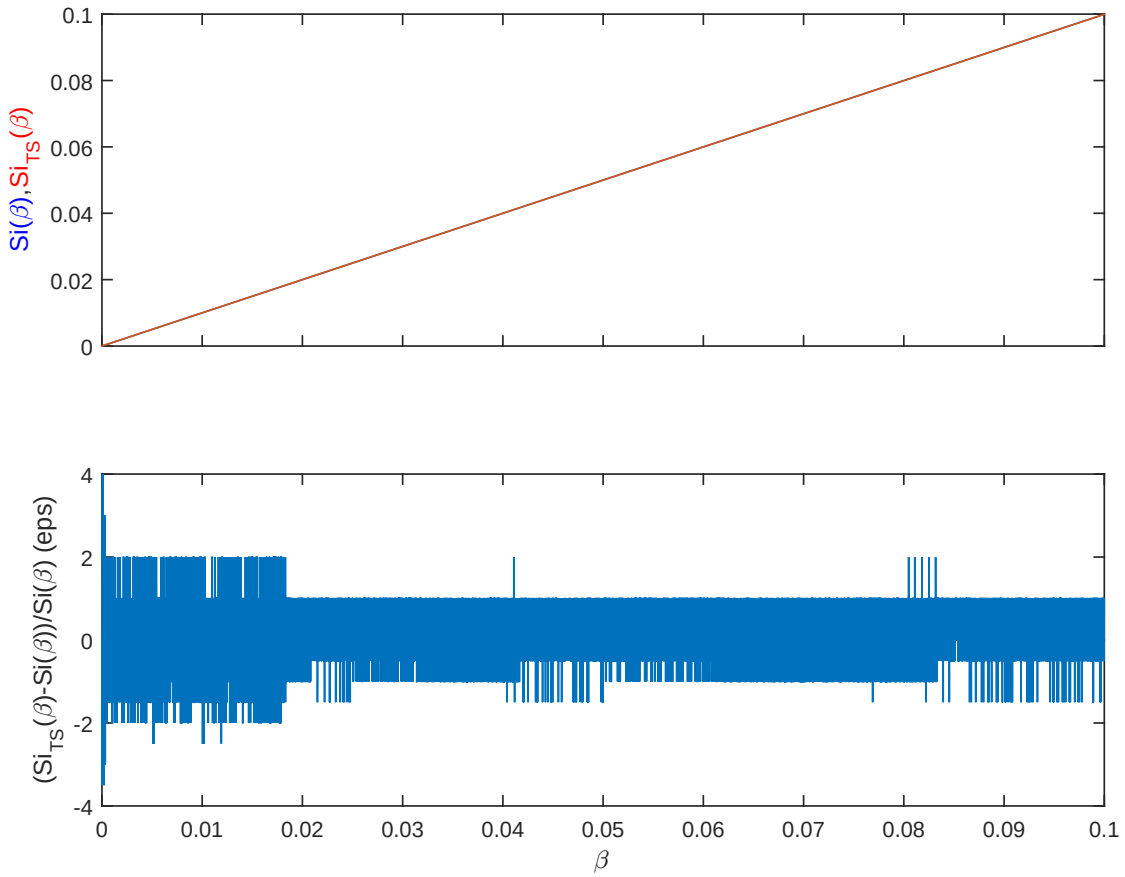


Figure 4.3: *Upper diagram: Taylor series evaluation of* $\mathrm{Si}(\beta)$ *near* $\beta = 0$ *(*$\mathrm{Si}_{TS}(\beta)$*, red), compared with the result of using the Matlab®-supplied routine (blue, but the curves are coincident to figure resolution); lower diagram: the relative error in terms of* `eps`*, the smallest value of* ϵ *such that* $1 + \epsilon \neq 1$ *to machine precision — for the machine used,* `eps` *is approximately* 2.2×10^{-16}

where

$$S_s(\beta) = \begin{cases} \frac{\pi}{2}, & s = 0; \\ \frac{\pi}{2} - \frac{\cos\beta}{\beta}, & s = 1; \\ \frac{\pi}{2} - \cos\beta \sum_{r=1}^{\frac{1}{2}s} \frac{(-1)^{r+1}(2r-2)!}{\beta^{2r-1}} - \sin\beta \sum_{r=1}^{\frac{1}{2}s} \frac{(-1)^{r+1}(2r-1)!}{\beta^{2r}}, & s \text{ even}, s > 0; \\ \frac{\pi}{2} - \cos\beta \sum_{r=1}^{\frac{1}{2}(s+1)} \frac{(-1)^{r+1}(2r-2)!}{\beta^{2r-1}} - \sin\beta \sum_{r=1}^{\frac{1}{2}(s-1)} \frac{(-1)^{r+1}(2r-1)!}{\beta^{2r}}, & s \text{ odd}, s > 1, \end{cases} \tag{4.62}$$

and

$$R_s(\beta) = \begin{cases} (-1)^{\frac{1}{2}s+1} s! \int_\beta^\infty \frac{\sin x}{x^{s+1}} \mathrm{d}x, & s \text{ even}; \\ (-1)^{\frac{1}{2}(s+1)} s! \int_\beta^\infty \frac{\cos x}{x^{s+1}} \mathrm{d}x, & s \text{ odd}. \end{cases} \tag{4.63}$$

$$\frac{\pi}{2} - \cos\beta \sum_{r=1}^{\infty} \frac{(-1)^{r+1}(2r-2)!}{\beta^{2r-1}} - \sin\beta \sum_{r=1}^{\infty} \frac{(-1)^{r+1}(2r-1)!}{\beta^{2r}} \tag{4.64}$$

is an asymptotic series for Si(β) (i.e. $S_s(\beta) = \mathrm{Si}(\beta) + \mathrm{O}\left(\frac{1}{\beta^s}\right)$, but, for fixed $\beta > 0$, $\lim_{s\to\infty} S_s(\beta)$ diverges), as

$$|R_s(\beta)| \leq s! \int_\beta^\infty \frac{\mathrm{d}x}{x^{s+1}} = \frac{(s-1)!}{\beta^s} = \tilde{R}_s(\beta). \tag{4.65}$$

Now, $\tilde{R}_{s_\beta}(\beta) < \tilde{R}_s(\beta)$ for $s \in \mathbb{N} \cup \{0\} - s_\beta$ if $\tilde{R}_{s_\beta}(\beta) < \tilde{R}_{s_\beta-1}(\beta), \tilde{R}_{s_\beta+1}(\beta) \Leftrightarrow s_\beta - 1 < \beta < s_\beta$, so the approximation of Si at β by a truncation of the asymptotic series (4.64) with the best bound of the form (4.65) is given by $S_{\lceil\beta\rceil}(\beta)$ and then the bound is given by

$$\tilde{R}_{\lceil\beta\rceil}(\beta) = \tfrac{(\lceil\beta\rceil-1)!}{\beta^{\lceil\beta\rceil}}. \tag{4.66}$$

	s			
β	$\lceil\beta\rceil$	$\min\{\lceil\beta\rceil, 30\}$	$\min\{\lceil\beta\rceil, 20\}$	$\min\{\lceil\beta\rceil, 10\}$
1	1			
2	0.25			
5	7.68×10^{-3}			
10	3.629×10^{-5}			
20	1.160×10^{-9}			3.544×10^{-8}
25	6.986×10^{-12}		1.338×10^{-11}	3.805×10^{-9}
30	4.294×10^{-14}		3.489×10^{-13}	6.145×10^{-10}
35	2.678×10^{-16}	4.212×10^{-16}	1.599×10^{-14}	1.315×10^{-10}
40	1.687×10^{-18}	7.669×10^{-18}	1.106×10^{-15}	3.461×10^{-11}

Table 4.1: *Bounds* $\tilde{R}_{\lceil\beta\rceil}(\beta)$, $\tilde{R}_{\min\{30,\lceil\beta\rceil\}}(\beta)$, $\tilde{R}_{\min\{20,\lceil\beta\rceil\}}(\beta)$ *and* $\tilde{R}_{\min\{10,\lceil\beta\rceil\}}(\beta)$ *for the absolute error caused by taking* s *terms of the asymptotic series for* Si(β)

This bound is given for various values of β in Table 4.1, together with similar bounds when the number of terms in the approximation is itself bounded.

As $\frac{\pi}{2} + \frac{2}{\beta} \geq \frac{\pi}{2} + \frac{1-\cos\beta}{\beta} \geq S_1(\beta) + |R_1(\beta)| \geq \mathrm{Si}(\beta) = S_1(\beta) + R_1(\beta) \geq S_1(\beta) - |R_1(\beta)| \geq \frac{\pi}{2} - \frac{1+\cos\beta}{\beta} \geq \frac{\pi}{2} - \frac{2}{\beta}$, we have $\frac{3\pi}{4} \geq \mathrm{Si}(\beta) \geq \frac{\pi}{4}$ for $\beta > \frac{8}{\pi}$, so the absolute relative error is of the same order of magnitude as the absolute error in Table 4.1, at least for the rows with $\beta \geq 5 > \frac{8}{\pi}$.

We will find that the quantities of interest to us depend on the eigenvalues and eigegenvectors of $H^{-1}P(\beta)$, or related matrices.

As H is a positive definite symmetric matrix, it has a decomposition $H = O_H D_H O_H^{\mathrm{T}}$, where O_H is an orthogonal matrix and D_H is a diagonal matrix with positive elements along the diagonal. Hence, $H^{-\frac{1}{2}} = O_H D_H^{-\frac{1}{2}} O_H^{\mathrm{T}}$, where $D_H^{-\frac{1}{2}}$ is the diagonal matrix whose diagonal elements are the positive square roots of the reciprocals of the corresponding elements of D_H, is a positive definite matrix such that $H^{-\frac{1}{2}} H^{-\frac{1}{2}} = H^{-1}$.

Then $H^{-\frac{1}{2}} P(\beta) H^{-\frac{1}{2}}$ is a positive definite symmetric matrix with the same eigenvalues as $H^{-1}P(\beta)$, and we also have $H^{-\frac{1}{2}} P(\beta) H^{-\frac{1}{2}} \to \mathrm{I} \Leftrightarrow H^{-1}P(\beta) \to \mathrm{I}$. The dependence of the

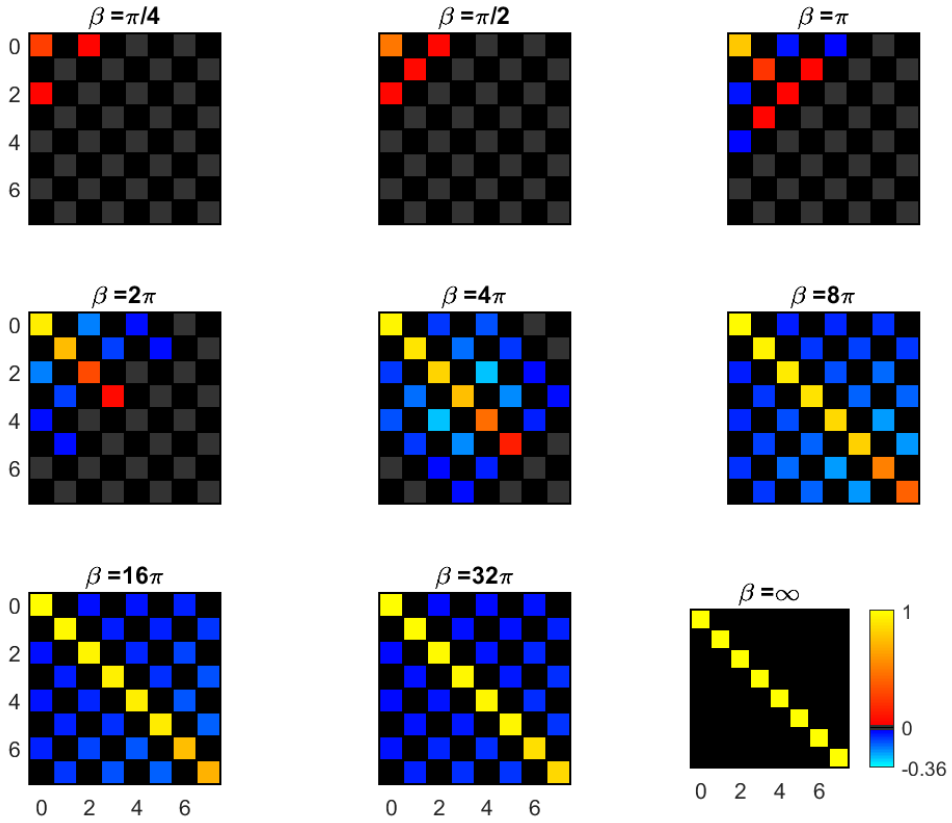


Figure 4.4: *Pseudocolour plot of the elements of* $H^{-\frac{1}{2}}P(\beta)H^{-\frac{1}{2}}$ *for* $n = 7$*. The colour scale given for the* $\beta = \infty$ *plot applies to all plots (exactly zero entries are shaded black and near-zero entries are shaded dark grey)*

entries of $H^{-\frac{1}{2}}P(\beta)H^{-\frac{1}{2}}$ on β for $n = 7$ is illustrated in Figures 4.4 and 4.5 (The elements of $H^{-\frac{1}{2}}P(\beta)H^{-\frac{1}{2}}$ fall into a narrower numerical range than those of $H^{-1}P(\beta)$, so it is easier to plot than $H^{-1}P(\beta)$; Figure 4.5 gives an idea of the magnitude of the values associated with the colours of Figure 4.4).

We can see that the convergence of the entries of $H^{-\frac{1}{2}}P(\beta)H^{-\frac{1}{2}}$ to those of I is fairly slow, especially for the off-diagonal elements.

4.7 Using P

As we shall see, the maximum L_2 distance between $\psi(x) = \begin{cases} \sum_{k=0}^{n} a_k x^k, & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise,} \end{cases}$

such that $\|\psi\| = 1$ and $\int_{-\frac{1}{2}}^{\frac{1}{2}} \psi(x)\mathrm{d}x = 0$, and the bandwidth limited version of ψ, ψ_β, is given by $1 - \lambda_0(\beta)$, where $\lambda_0(\beta)$ is the smallest eigenvalue of the generalised eigenvalue problem $Q_b P(\beta) Q_b a = \lambda Q_b H Q_b a$, where Q_b is the projection into the space orthogonal to b.

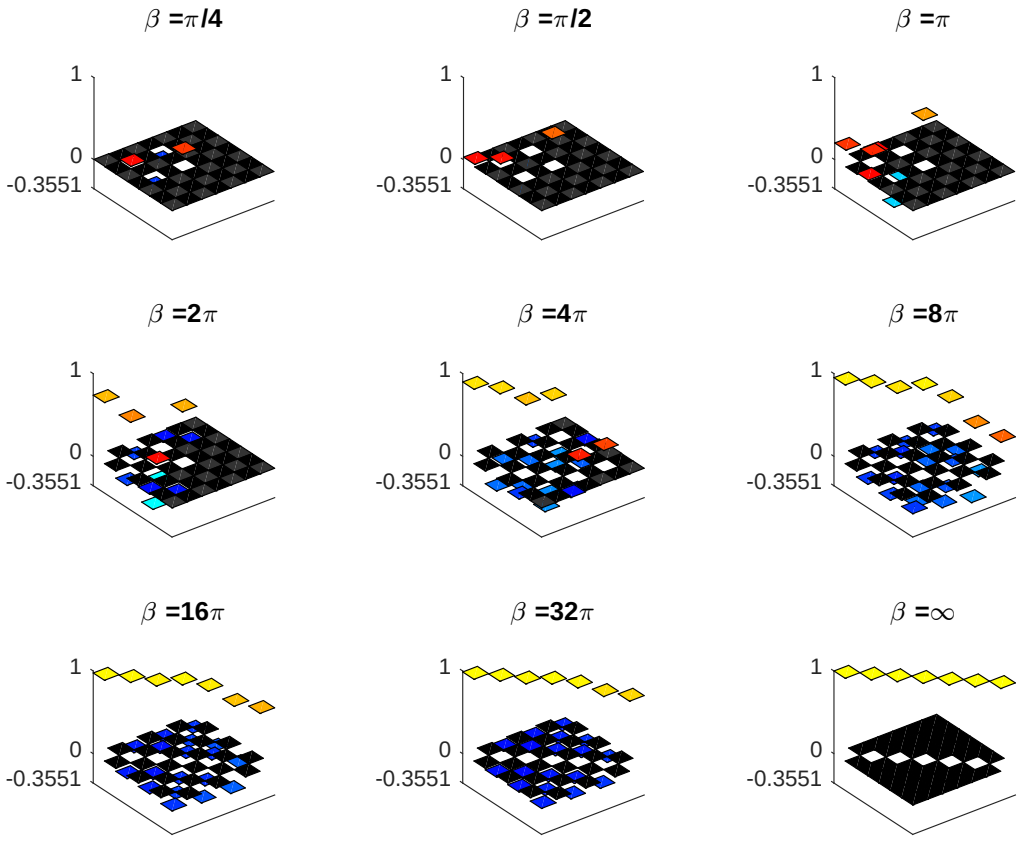


Figure 4.5: *3-d plot of the elements of* $\tilde{H}^{-\frac{1}{2}}\tilde{P}(\beta)\tilde{H}^{-\frac{1}{2}}$ *for* $n = 7$. *The colour scale given in Figure 4.4 for the* $\beta = \infty$ *plot applies to all plots here too*

This approach can be generalised to take account of any additional condition on ψ that can be expressed as $c^{\mathrm{T}}a = 0$. We will use the *wrapped endpoint* conditions *continuity* and *differentiability*.

Wrapped endpoint continuity, (*wec*),

$$\psi\left(-\frac{1}{2}\right) = \psi\left(\frac{1}{2}\right), \tag{4.67}$$

equivalent to $c^{\mathrm{T}}a = 0$ with $c = c_1 = \left[\begin{smallmatrix} 0 & 1 & 0 & \frac{1}{4} & 0 & \cdots & \frac{1+(-1)^n}{2^{n-1}} & \frac{1-(-1)^n}{2^n} \end{smallmatrix}\right]^{\mathrm{T}}$, ensures that the function $\sum_{k\in\mathbb{Z}}\psi(x-k)$ is continuous, and *wrapped endpoint differentiability*, (*wed*), wec together with

$$\left.\frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\right|_{x=-\frac{1}{2}} = \left.\frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\right|_{x=\frac{1}{2}}, \tag{4.68}$$

equivalent to wec plus the condition $c^{\mathrm{T}}a = 0$ with $c = c_2 = \left[\begin{smallmatrix} 0 & 0 & 2 & 0 & 1 & \cdots & \frac{1-(-1)^n}{2^{n-2}}(n-1) & \frac{1+(-1)^n}{2^{n-1}}n \end{smallmatrix}\right]^{\mathrm{T}}$, ensures that $\sum_{k\in\mathbb{Z}}\psi(x-k)$ is differentiable. Through their effect on $\sum_{k\in\mathbb{Z}}\psi(x-k)$, both of

these conditions reduce the presence of higher frequency components in $\hat{\psi}$.

When $n \geq 3$, $\det \begin{bmatrix} b^{(3)} & c_1^{(3)} & c_2^{(3)} \end{bmatrix} = 2 \neq 0$, where $b^{(3)} = \left[1, 0, \frac{1}{12}\right]^{\mathrm{T}}$, $c_1^{(3)} = [0, 1, 0]^{\mathrm{T}}$ and $c_2^{(3)} = [0, 0, 2]^{\mathrm{T}}$, so the wrapped endpoint conditions are independent of each other, and of the zi condition.

In order to evaluate the effect of the finite bandwidth of the NAT accelerometer, we wish to find the maximum L_2 distance between a pair of wavelets ψ and ψ' such that ψ' is the same distance from (or closer to) ψ_β than ψ is. (Clearly, such wavelets could be confused for each other if the replacements $\psi \to \psi'_\beta$, $\psi' \to \psi_\beta$ are made.)

That is, we wish to find the maximum value of $\|\psi' - \psi\|$ such that $\|\psi' - \psi_\beta\| \leq \|\psi - \psi_\beta\|$.

Since $\|\psi' - \psi\|^2 = 2(1 - \langle\psi', \psi\rangle)$, $\|\psi - \psi_\beta\|^2 = 1 - \|\psi_\beta\|^2$ and $\|\psi' - \psi_\beta\|^2 = 1 - 2\langle\psi', \psi_\beta\rangle + \|\psi_\beta\|^2$, an equivalent problem is that of minimising $\langle\psi', \psi\rangle$, subject to $\langle\psi', \psi_\beta\rangle \geq \|\psi_\beta\|^2$.

Writing $\psi'(x) = \begin{cases} \sum_{k=0}^{n} a'_k x^k, & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise,} \end{cases}$ our problem becomes that of minimising $a^{\mathrm{T}} H a'$, subject to $a^{\mathrm{T}} P(\beta) a' \geq a^{\mathrm{T}} P(\beta) a$, where, as ψ and ψ' obey the ue and zi conditions, we must also have $a^{\mathrm{T}} H a = a'^{\mathrm{T}} H a' = 1$ and $b^{\mathrm{T}} a = b^{\mathrm{T}} a' = 0$.

We will look at three cases:

1 where we impose no further conditions on ψ and ψ';

2 where we impose wrapped endpoint continuity, $\psi\left(-\frac{1}{2}\right) = \psi\left(\frac{1}{2}\right)$, $\psi'\left(-\frac{1}{2}\right) = \psi'\left(\frac{1}{2}\right)$, enforced by $c_1^{\mathrm{T}} a = c_1^{\mathrm{T}} a' = 0$;

3 where we impose $c_1^{\mathrm{T}} a = c_1^{\mathrm{T}} a' = 0$ and wrapped endpoint differentiability, $\frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\big|_{x=-\frac{1}{2}} = \frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\big|_{x=\frac{1}{2}}$, $\frac{\mathrm{d}}{\mathrm{d}x}\psi'(x)\big|_{x=-\frac{1}{2}} = \frac{\mathrm{d}}{\mathrm{d}x}\psi'(x)\big|_{x=\frac{1}{2}}$, enforced by $c_2^{\mathrm{T}} a = c_2^{\mathrm{T}} a' = 0$.

To take account of the requirement that $b^{\mathrm{T}} a = b^{\mathrm{T}} a' = 0$, we find an orthogonal matrix O such that $Ob = \|b\| e_0$, where $e_0 \in \mathbb{R}^{n+1}$ has a 1 as its first entry and zeros elsewhere.

Then $0 = b^{\mathrm{T}} a = \|b\| e_0^{\mathrm{T}} O a = \|b\| e_0^{\mathrm{T}} \tilde{a}$ implies $e_0^{\mathrm{T}} \tilde{a} = 0$, where $\tilde{a} = Oa$, so $\tilde{a} = \begin{bmatrix} 0 \\ \tilde{a}_1 \end{bmatrix}$. Similarly, $\tilde{a}' = Oa' = \begin{bmatrix} 0 \\ \tilde{a}'_1 \end{bmatrix}$, and our problem becomes that of minimising $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1$, subject to $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}'_1 \geq \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1$ and $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1 = \tilde{a}_1'^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1 = 1$, where $\tilde{H} = O^{\mathrm{T}} H O = \begin{bmatrix} \tilde{h}_0 & \tilde{h}_1^{\mathrm{T}} \\ \tilde{h}_1 & \tilde{H}_1 \end{bmatrix}$ and $\tilde{P}(\beta) = O^{\mathrm{T}} P(\beta) O = \begin{bmatrix} \tilde{p}_0(\beta) & \tilde{p}_1(\beta)^{\mathrm{T}} \\ \tilde{p}_1(\beta) & \tilde{P}_1(\beta) \end{bmatrix}$ for some $\tilde{h}_0, \tilde{p}_0(\beta) \in \mathbb{R}$ and $\tilde{h}_1, \tilde{p}_1(\beta) \in \mathbb{R}^n$.

Thus, we have projected case 1 into $\mathbb{R}^n$.

Analogously, we can project cases 2 and 3 into $\mathbb{R}^{n-1}$ and $\mathbb{R}^{n-2}$. As b and c_1 (resp. b, c_1 and c_2) are linearly independent if $n > 2$ (resp. $n > 3$), we can find an orthogonal matrix O such that $Ob = \|b\| e_0$ and $Oc_1 = \|c_1\| e_1$ (resp. $Ob = \|b\| e_0$, $Oc_1 = \|c_1\| e_1$ and $Oc_2 = \|c_2\| e_2$). In case 2 (resp. case 3), we can multiply by O as we did in case 1 and then delete the first two (resp three) elements of $\tilde{a}$ and $\tilde{a}'$, and the first two (resp. three) rows and columns of $\tilde{H}$ and $\tilde{P}(\beta)$ to obtain $\tilde{H}_1$ and $\tilde{P}_1(\beta)$.

Given $\mathcal{S} = \{u_0, u_1, u_2, u_3, \ldots, u_n\} = \begin{cases} \{b, e_1, e_2, e_3, \ldots, e_n\}, & \text{case 1;} \\ \{b, c_1, e_2, e_3, \ldots, e_n\}, & \text{case 2;} \\ \{b, c_1, c_2, e_3, \ldots, e_n\}, & \text{case 3} \end{cases}$ (where e_k is the vector with 1 in the kth position and zeros elsewhere) we can apply the well-known Gram-

Schmidt process to orthogonalise $\mathcal{S}$ and read off the rows of the orthogonal matrix O required for each of the three cases above:

$$\begin{array}{rcrcl} & & & e'_0 = \frac{u_0}{\|u_0\|}, & \text{so } O_{0k} = \frac{u_{0,k}}{\|u_0\|} \ (= \frac{b_{0,k}}{\|b\|} \text{ in all cases}), \\ e''_1 = & u_1 - ({e'_0}^{\mathrm{T}} u_1) e'_0, & e'_1 = \frac{e''_1}{\|e''_1\|}, & \text{so } O_{1k} = \frac{u_{1,k} - ({e'_0}^{\mathrm{T}} u_1) e'_{0,k}}{\sqrt{\|u_1\|^2 - ({e'_0}^{\mathrm{T}} u_1)^2}}, \\ e''_2 = & u_2 - ({e'_0}^{\mathrm{T}} u_2) e'_0 - ({e'_1}^{\mathrm{T}} u_2) e'_1, & e'_2 = \frac{e''_2}{\|e''_2\|}, & \\ \vdots = & \vdots & \vdots = & \vdots & \\ e''_n = & u_n - \sum_{k=0}^{n-1} ({e'_k}^{\mathrm{T}} u_n) e'_k, & e'_n = \frac{e''_n}{\|e''_n\|}, & \end{array}$$

and the rest of the $O_{k\ell}$ can be read off the equations $e''_k = \sum_{\ell=0}^{k-1} O_{k\ell} e_\ell$.

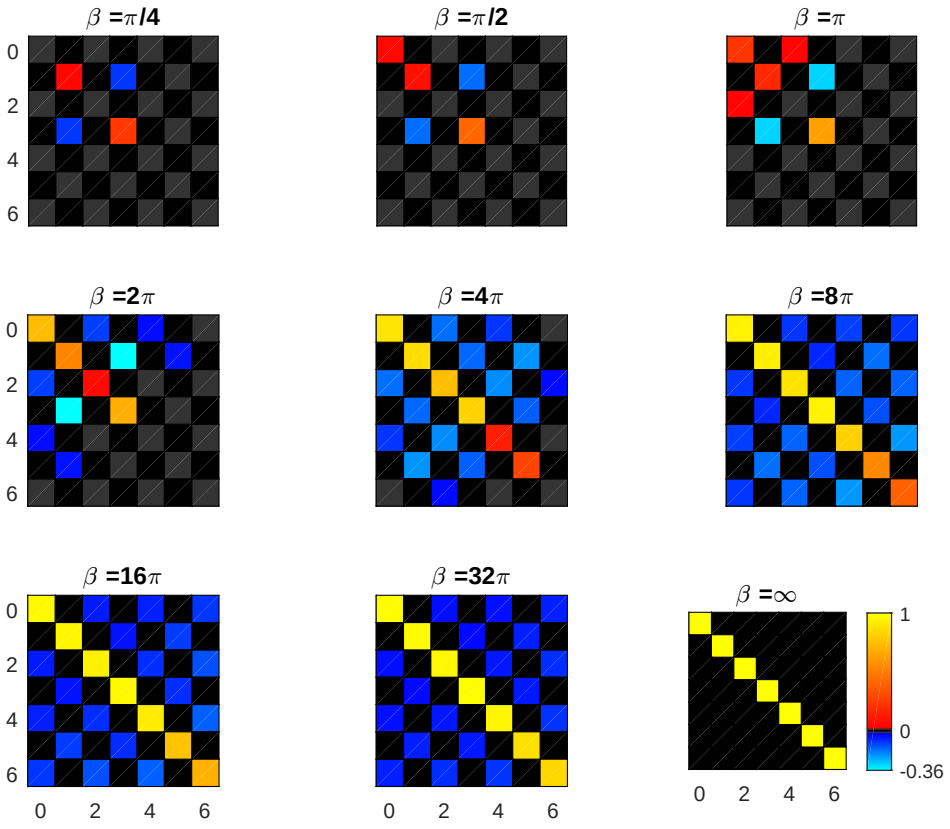


Figure 4.6: *Pseudocolour plot of the elements of* $\tilde{H}^{-\frac{1}{2}} \tilde{P}(\beta) \tilde{H}^{-\frac{1}{2}}$ *for* $n = 7$

Temporarily writing $\tilde{H}_{c_1}$ and $\tilde{P}_{c_1}(\beta)$ (resp. $\tilde{H}_{c_1 c_2}$ and $\tilde{P}_{c_1 c_2}(\beta)$) for $\tilde{H}_1$ and $\tilde{P}_1(\beta)$ in case 2 (resp. case 3) (but leaving $\tilde{H}_1$ and $\tilde{P}_1(\beta)$ unchanged in case 1), we can look at the behaviour of the projections of $H^{-\frac{1}{2}} P(\beta) H^{-\frac{1}{2}}$.

In Figures 4.7 and 4.6, we illustrate the behaviour of $\tilde{H}_1^{-\frac{1}{2}} \tilde{P}_1(\beta) \tilde{H}_1^{-\frac{1}{2}}$ as $\beta \to \infty$, for $n = 7$.

Although there is no clear apparent relation to the behaviour of the values of the elements of $H^{-\frac{1}{2}} P(\beta) H^{-\frac{1}{2}}$ (as shown in Figures 4.4 and 4.5) for small values of β, it seems that the

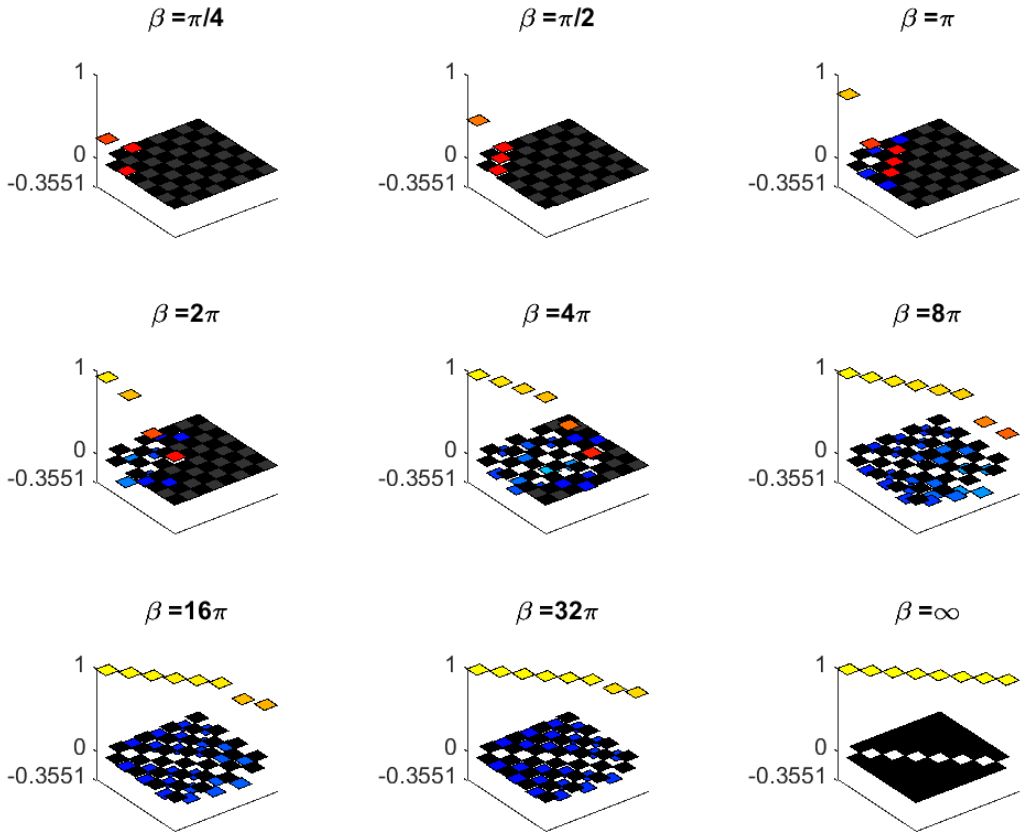


Figure 4.7: *3-d plot of the elements of* $H^{-\frac{1}{2}}P(\beta)H^{-\frac{1}{2}}$ *for* $n = 7$. *The colour scale given in Figure 4.4 for the* $\beta = \infty$ *plot applies to all plots here too*

convergence of $\tilde{H}_1^{-\frac{1}{2}}\tilde{P}_1(\beta)\tilde{H}_1^{-\frac{1}{2}}$ to $\mathrm{I}(\in \mathbb{R}^{7\times 7})$ as $\beta \to \infty$ is slower, especially for the diagonal elements of the matrices.

In Figures 4.9 and 4.8, we illustrate the behaviour of $\tilde{H}_{c_1}^{-\frac{1}{2}}\tilde{P}_{c_1}(\beta)\tilde{H}_{c_1}^{-\frac{1}{2}}$ as $\beta \to \infty$, for $n = 7$.

Although for the smallest value of β, $\tilde{H}_{c_1}^{-\frac{1}{2}}\tilde{P}_{c_1}(\beta)\tilde{H}_{c_1}^{-\frac{1}{2}}$ is quite similar to $\tilde{H}_1^{-\frac{1}{2}}\tilde{P}_1(\beta)\tilde{H}_1^{-\frac{1}{2}}$ (as shown in Figures 4.6 and 4.7) with the first row and column deleted, there is no clear pattern to the relationship $\tilde{H}_1^{-\frac{1}{2}}\tilde{P}_1(\beta)\tilde{H}_1^{-\frac{1}{2}}$ (as shown in Figures 4.6 and 4.7) and $\tilde{H}_{c_1}^{-\frac{1}{2}}\tilde{P}_{c_1}(\beta)\tilde{H}_{c_1}^{-\frac{1}{2}}$ for small and medium values of β, but it seems that the convergence of $\tilde{H}_{c_1}^{-\frac{1}{2}}\tilde{P}_{c_1}(\beta)\tilde{H}_{c_1}^{-\frac{1}{2}}$ to $\mathrm{I}(\in \mathbb{R}^{6\times 6})$ as $\beta \to \infty$ is slightly quicker for the off-diagonal elements, especially for the $(1,3)/(3,1)$ element. Also, positive off-diagonal elements persist longer in the $\tilde{H}_{c_1}^{-\frac{1}{2}}\tilde{P}_{c_1}(\beta)\tilde{H}_{c_1}^{-\frac{1}{2}}$ matrix.

In Figures 4.11 and 4.10, we illustrate the behaviour of $\tilde{H}_{c_1c_2}^{-\frac{1}{2}}\tilde{P}_{c_1c_2}(\beta)\tilde{H}_{c_1c_2}^{-\frac{1}{2}}$ as $\beta \to \infty$, for $n = 7$.

In a comparison between $\tilde{H}_{c_1c_2}^{-\frac{1}{2}}\tilde{P}_{c_1c_2}(\beta)\tilde{H}_{c_1c_2}^{-\frac{1}{2}}$ and $\tilde{H}_1^{-\frac{1}{2}}\tilde{P}_1(\beta)\tilde{H}_1^{-\frac{1}{2}}$ (Figures 4.6 and 4.7),

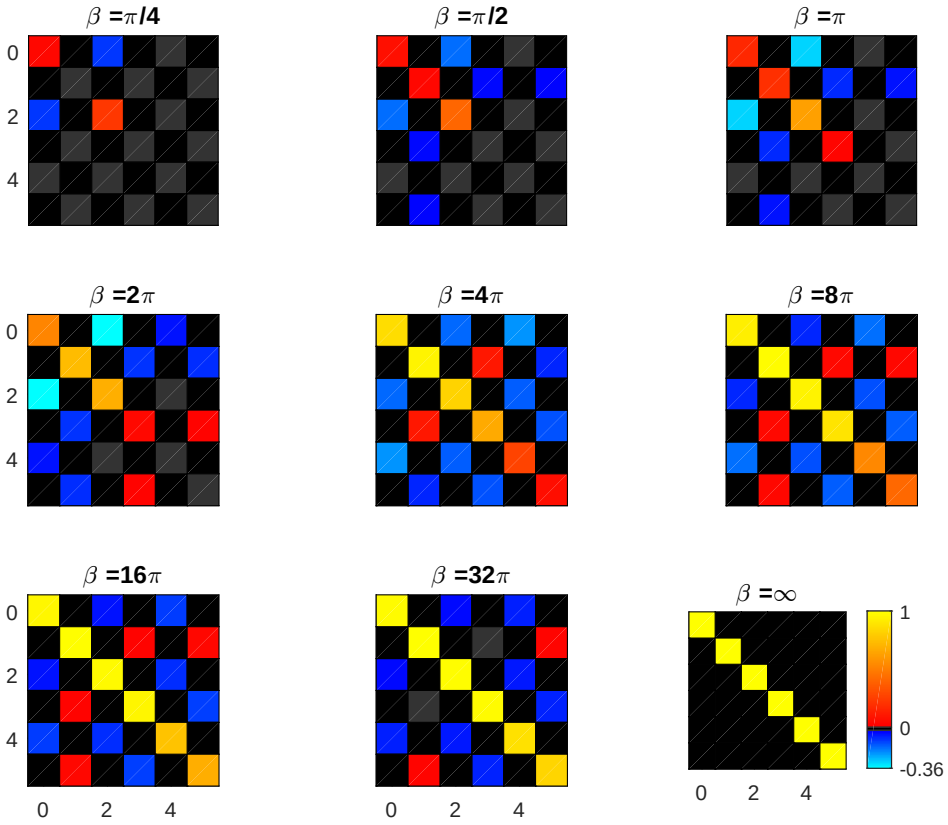


Figure 4.8: *Pseudocolour plot of the elements of* $\tilde{H}_{c_1}^{-\frac{1}{2}}\tilde{P}_{c_1}(\beta)\tilde{H}_{c_1}^{-\frac{1}{2}}$ *(wec) for* $n = 7$

the elements of $\tilde{H}_{c_1c_2}^{-\frac{1}{2}}\tilde{P}_{c_1c_2}(\beta)\tilde{H}_{c_1c_2}^{-\frac{1}{2}}$ are quite similar to those of $\tilde{H}_1^{-\frac{1}{2}}\tilde{P}_1(\beta)\tilde{H}_1^{-\frac{1}{2}}$ for small β, and the relationship becomes confused for medium β. As $\beta \to +\infty$, the diagonal elements of $\tilde{H}_{c_1c_2}^{-\frac{1}{2}}\tilde{P}_{c_1c_2}(\beta)\tilde{H}_{c_1c_2}^{-\frac{1}{2}}$ seem to converge more slowly to those of $\mathrm{I} \in \mathbb{R}^{5\times5}$ than do those of the top left-hand 5×5 corner of $\tilde{H}_1^{-\frac{1}{2}}\tilde{P}_1(\beta)\tilde{H}_1^{-\frac{1}{2}}$, whereas the reverse appears to be true for the off-diagonal elements.

When the comparison is between $\tilde{H}_{c_1c_2}^{-\frac{1}{2}}\tilde{P}_{c_1c_2}(\beta)\tilde{H}_{c_1c_2}^{-\frac{1}{2}}$ and $\tilde{H}_{c_1}^{-\frac{1}{2}}\tilde{P}_{c_1}(\beta)\tilde{H}_{c_1}^{-\frac{1}{2}}$ (Figures 4.8 and 4.9), the former seems to be similar to the lower right-hand 5×5 submatrix of the latter for the smallest value of β, and similar to the upper left-hand 5×5 submatrix of the latter as $\beta \to \infty$, with no clear pattern for intermediate values of β.

The eigenvalues of the matrices $\tilde{H}_1^{-1}\tilde{P}_1(\beta)$ $(n = 2, \ldots, 7)$, $\tilde{H}_{c_1}^{-1}\tilde{P}_{c_1}(\beta)$ $(n = 3, \ldots, 7)$ and $\tilde{H}_{c_1c_2}^{-1}\tilde{P}_{c_1c_2}(\beta)$ $(n = 4, \ldots, 7)$ are shown as functions of β in Figures 4.12 to 4.14.

Dropping our temporary notation, we can now treat all of our three cases in a similar fashion.

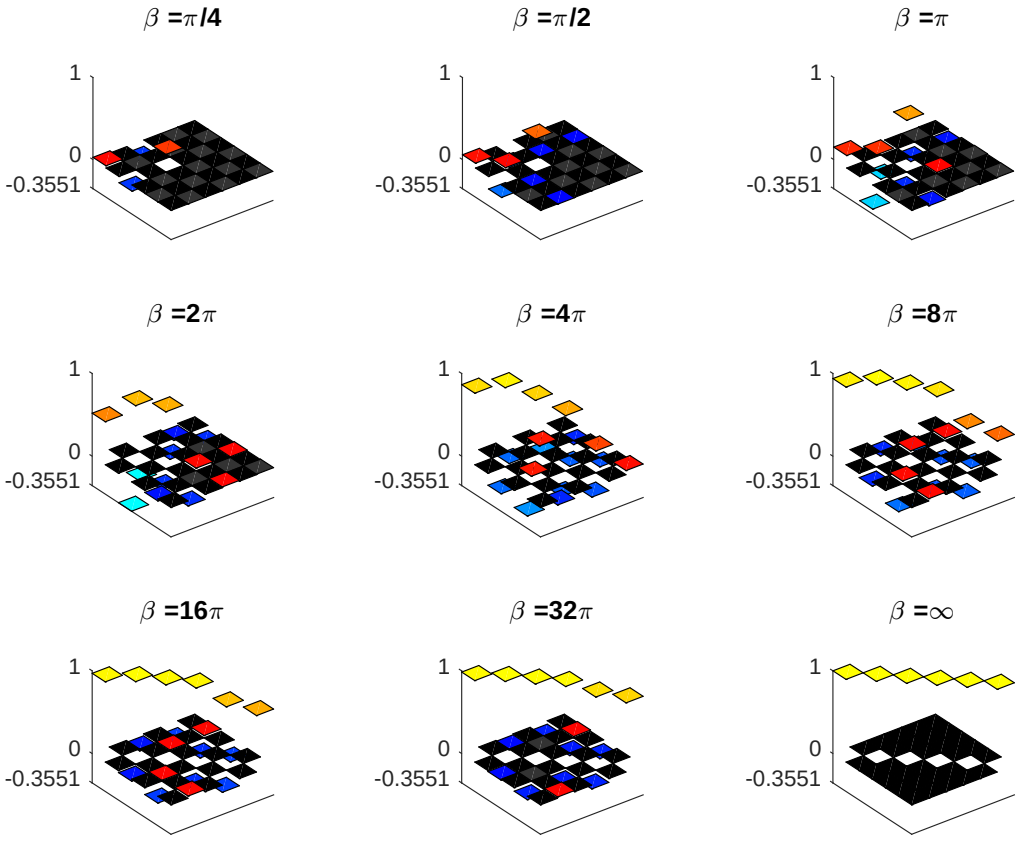


Figure 4.9: *3-d plot of the elements of* $\tilde{H}_{c_1}^{-\frac{1}{2}}\tilde{P}_{c_1}(\beta)\tilde{H}_{c_1}^{-\frac{1}{2}}$ *(wec) for* $n = 7$*. The colour scale given in Figure 4.4 for the* $\beta = \infty$ *plot applies to all plots here too*

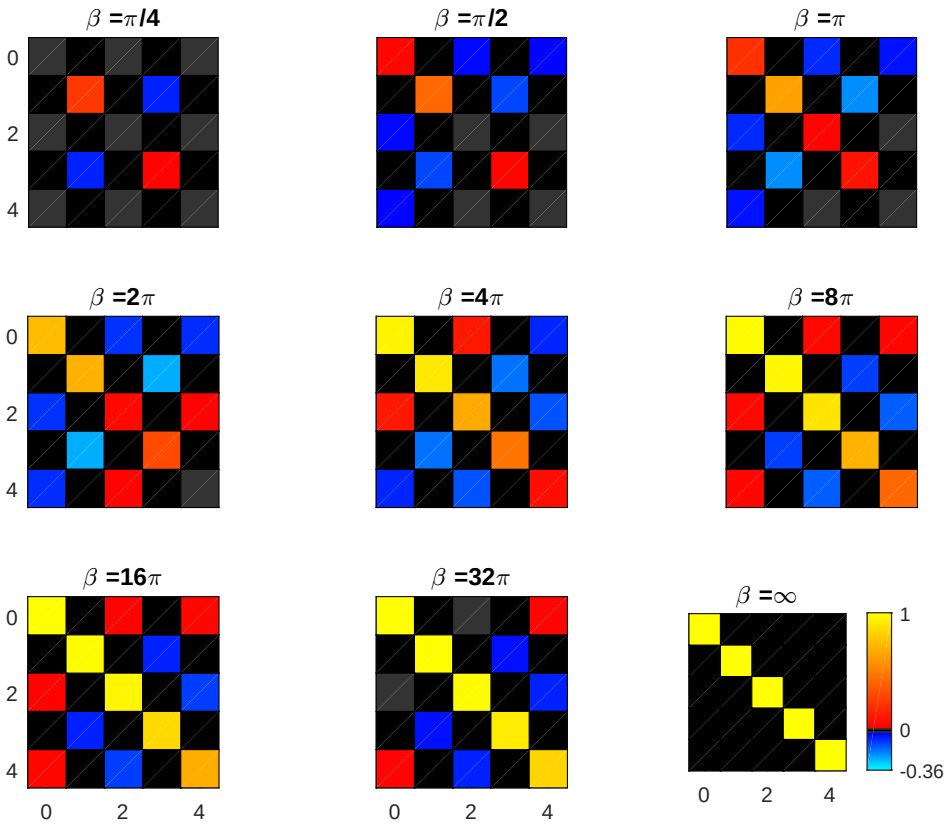


Figure 4.10: *Pseudocolour plot of the elements of* $\tilde{H}^{-\frac{1}{2}}_{c_1c_2}\tilde{P}_{c_1c_2}(\beta)\tilde{H}^{-\frac{1}{2}}_{c_1c_2}$ *(wed) for* $n = 7$

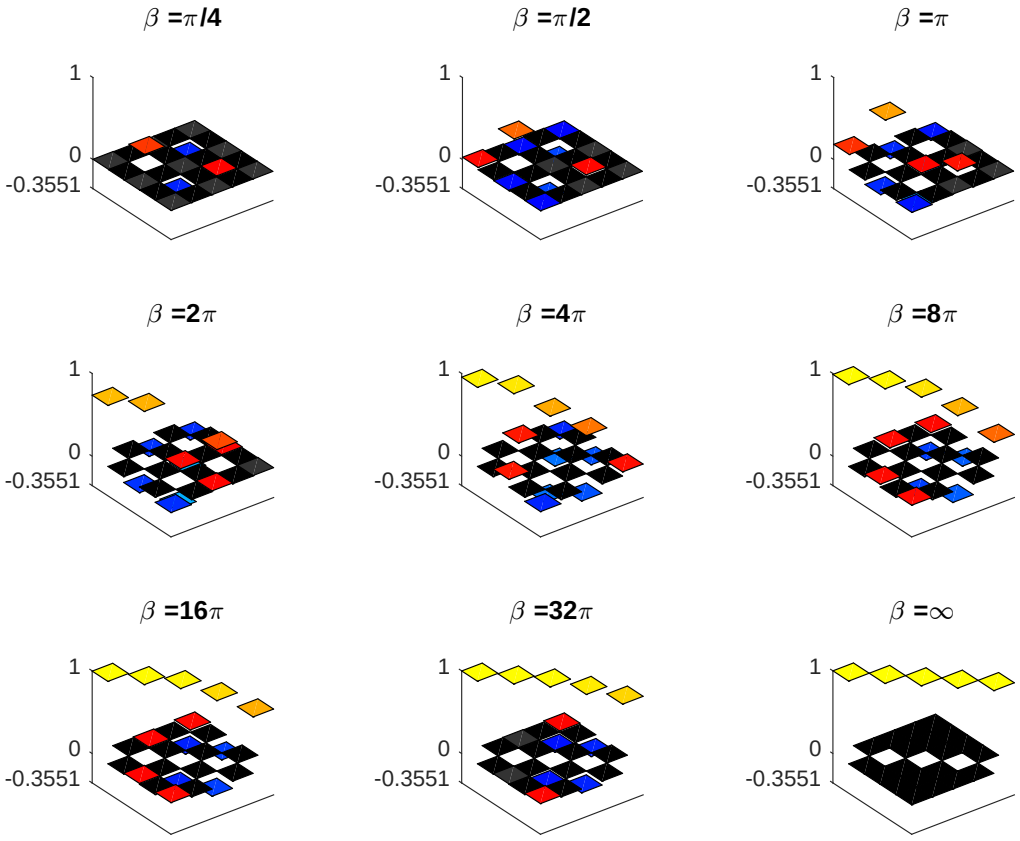


Figure 4.11: *3-d plot of the elements of* $\tilde{H}_{c_1c_2}^{-\frac{1}{2}}\tilde{P}_{c_1c_2}(\beta)\tilde{H}_{c_1c_2}^{-\frac{1}{2}}$ *(wed) for* $n = 7$*. The colour scale given in Figure 4.4 for the* $\beta = \infty$ *plot applies to all plots here too*

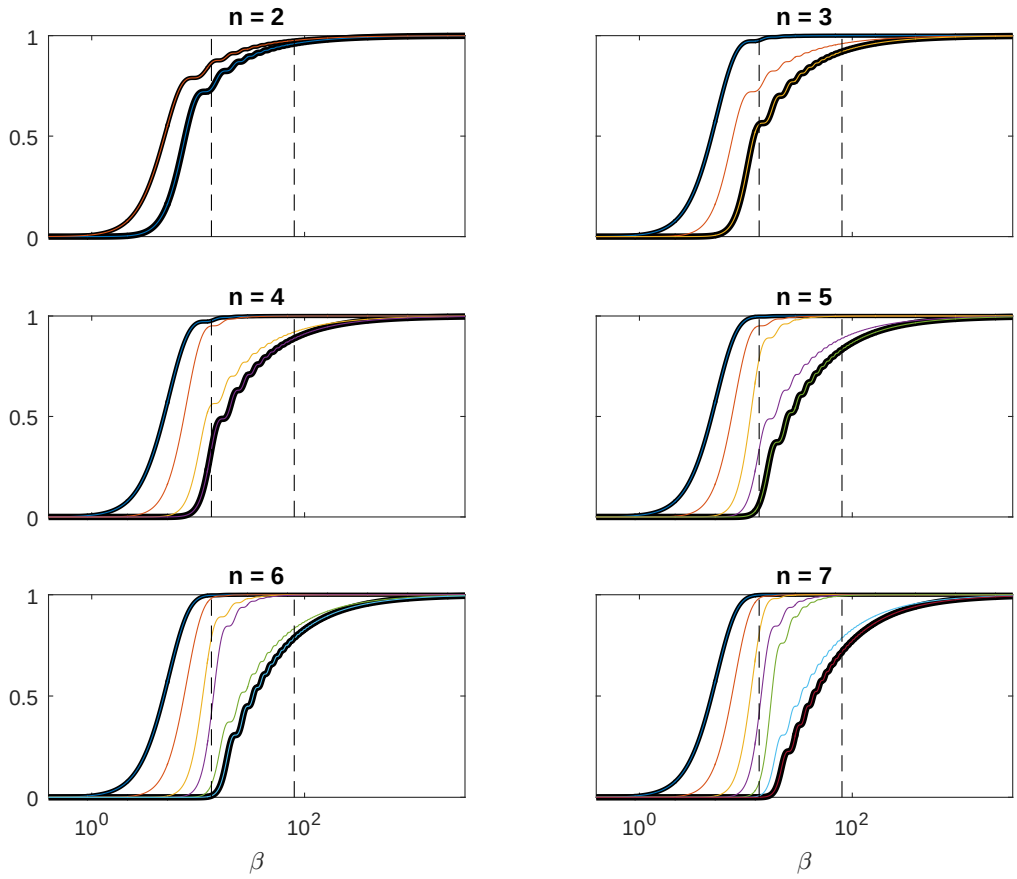


Figure 4.12: *Plots of the eigenvalues of* $\tilde{H}^{-1}\tilde{P}(\beta)$*, for* $n = 2, 3, 4, 5, 6, 7$*. The thicker black line gives the minimum of these eigenvalues, the thinner, the maximum. These eigenvalues correspond to the local extrema of the inner product* $\langle \psi, \psi_\beta \rangle$*. The dashed vertical lines correspond to the bandwidth of the NAT accelerometer (80Hz), and this value divided by the frequency most characteristic of PD movement disorders, 6Hz*

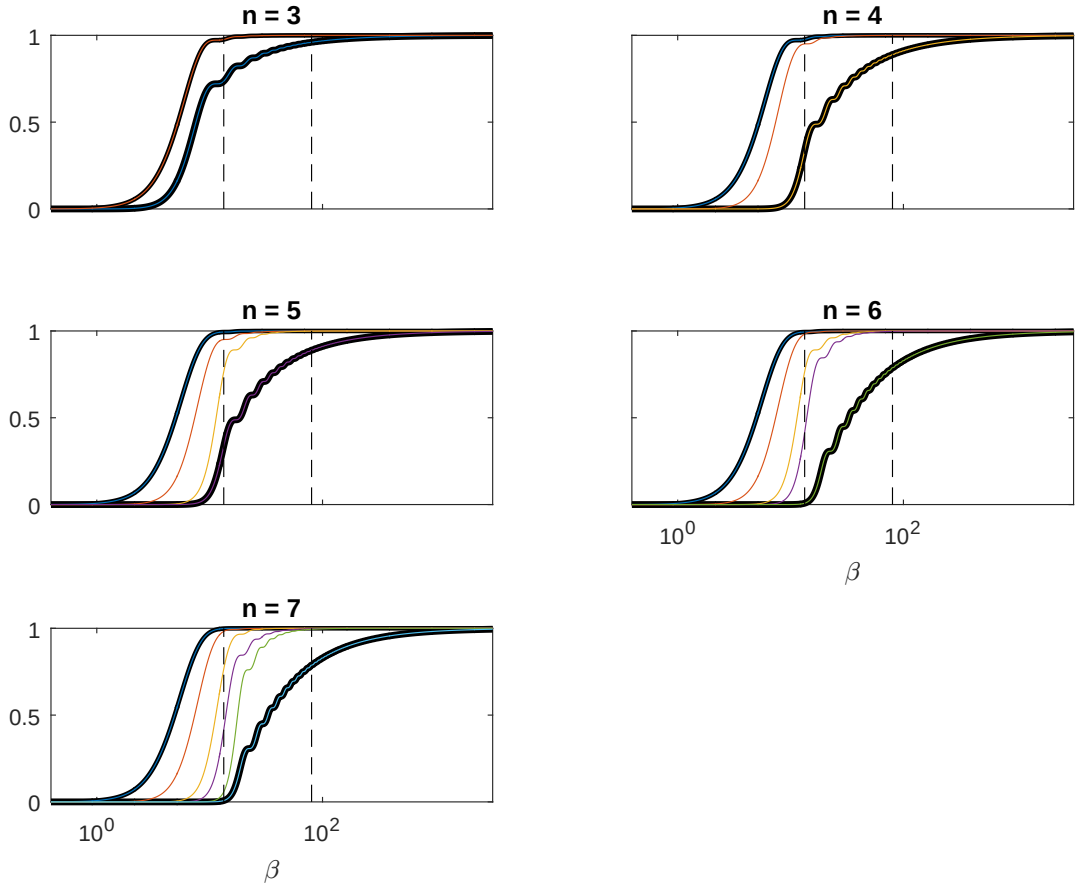


Figure 4.13: *Plots of the eigenvalues of* $\tilde{H}_{c_1}^{-1}\tilde{P}_{c_1}(\beta)$ *(wec), for* $n = 3, 4, 5, 6, 7$. *These eigenvalues correspond to the local extrema of the inner product* $\langle \psi, \psi_\beta \rangle$, *subject to the constraint that* $\psi\left(-\frac{1}{2}\right) = \psi\left(\frac{1}{2}\right)$

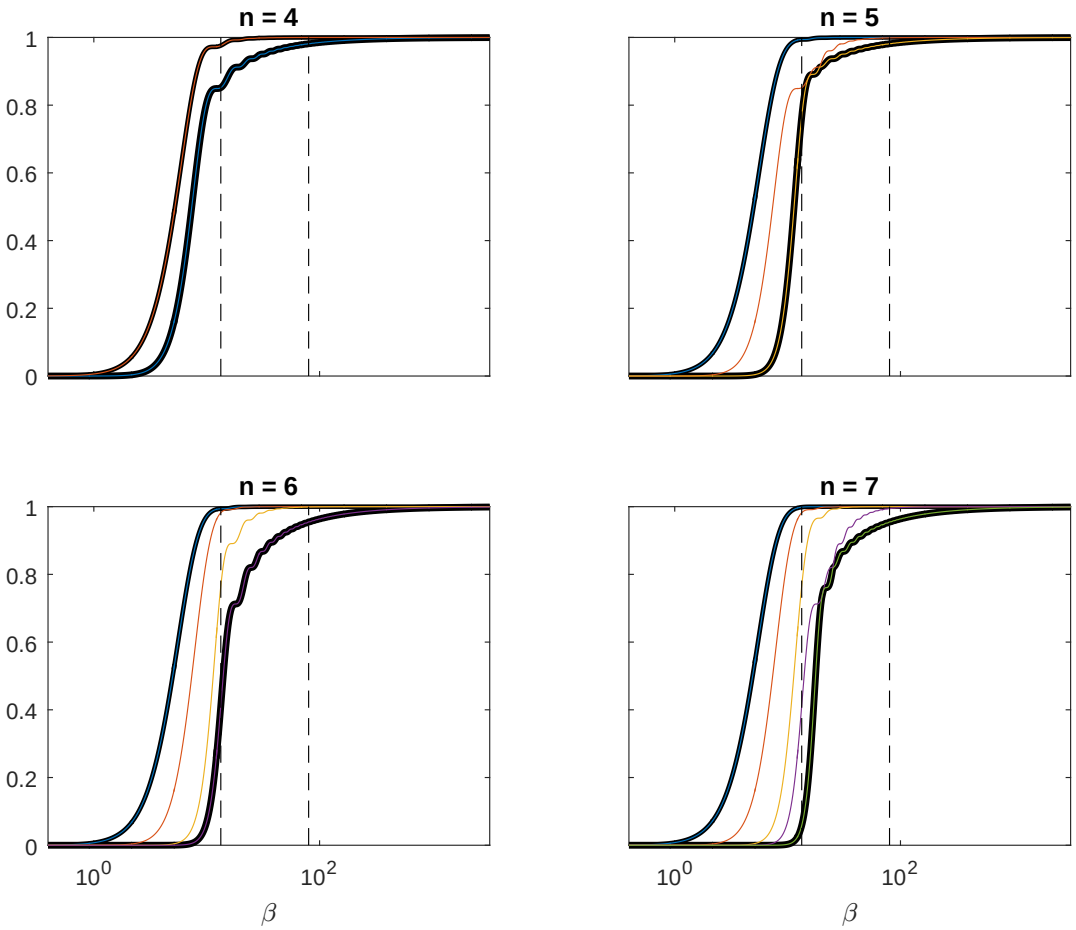


Figure 4.14: *Plots of the eigenvalues of* $\tilde{H}^{-1}_{c_1c_2}\tilde{P}_{c_1c_2}(\beta)$ *(wec), for* $n = 4, 5, 6, 7$*. These eigenvalues correspond to the local extrema of the inner product* $\langle \psi, \psi_\beta \rangle$*, subject to the constraints that* $\psi\left(-\frac{1}{2}\right) = \psi\left(\frac{1}{2}\right)$ *and* $\frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\big|_{x=-\frac{1}{2}} = \frac{\mathrm{d}}{\mathrm{d}x}\psi(x)\big|_{x=\frac{1}{2}}$

4.7.1 Maximum distance between ψ and ψ_β

Obviously, the L_2 distance between ψ and ψ_β is nonzero, because $\|\psi_\beta\|^2 = \|\hat{\psi}_\beta\|^2 = \int_{-\beta}^{\beta} |\hat{\psi}_\beta(\xi)|^2 \mathrm{d}\xi < \int_{-\infty}^{\infty} |\hat{\psi}_\beta(\xi)|^2 \mathrm{d}\xi = \int_{-\infty}^{\infty} |\hat{\psi}(\xi)|^2 \mathrm{d}\xi = \|\psi\|^2 = 1$, as a result of $|\hat{\psi}_\beta(\xi)|^2$ being bounded away from 0 on some nontrivial interval $I(\beta)$ contained in $(-\infty, -\beta] \cup [\beta, \infty)$. Consequently, $\psi \neq \psi_\beta$, and $\|\psi - \psi_\beta\|^2 \neq 0$.

As $\|\psi - \psi_\beta\|^2 = \|\psi\|^2 - 2\langle \psi, \psi_\beta \rangle + \|\psi_\beta\|^2 = 1 - \|\psi_\beta\|^2$ (because $\langle \psi, \psi_\beta \rangle = \langle \hat{\psi}, \hat{\psi}_\beta \rangle = \int_{-\infty}^{\infty} \hat{\psi}(\xi)\hat{\psi}_\beta(\xi)\mathrm{d}\xi = \int_{-\beta}^{\beta} \hat{\psi}(\xi)\hat{\psi}_\beta(\xi)\mathrm{d}\xi = \int_{-\beta}^{\beta} \hat{\psi}_\beta(\xi)\hat{\psi}_\beta(\xi)\mathrm{d}\xi = \|\hat{\psi}_\beta\|^2 = \|\psi_\beta\|^2$), we have $\|\psi - \psi_\beta\|^2 = 1 - a^{\mathrm{T}} P(\beta) a$, and maximising $\|\psi - \psi_\beta\|$ with respect to ψ subject to the zi and ui conditions (and possibly the wrapped endpoint conditions), is equivalent to minimising $a^{\mathrm{T}} P(\beta) a$ with respect to a, subject to $a^{\mathrm{T}} H a = 1$ and $b^{\mathrm{T}} a = 0$ (and possibly $c_1^{\mathrm{T}} a = 0$, or $c_1^{\mathrm{T}} a = c_2^{\mathrm{T}} a = 0$ as well).

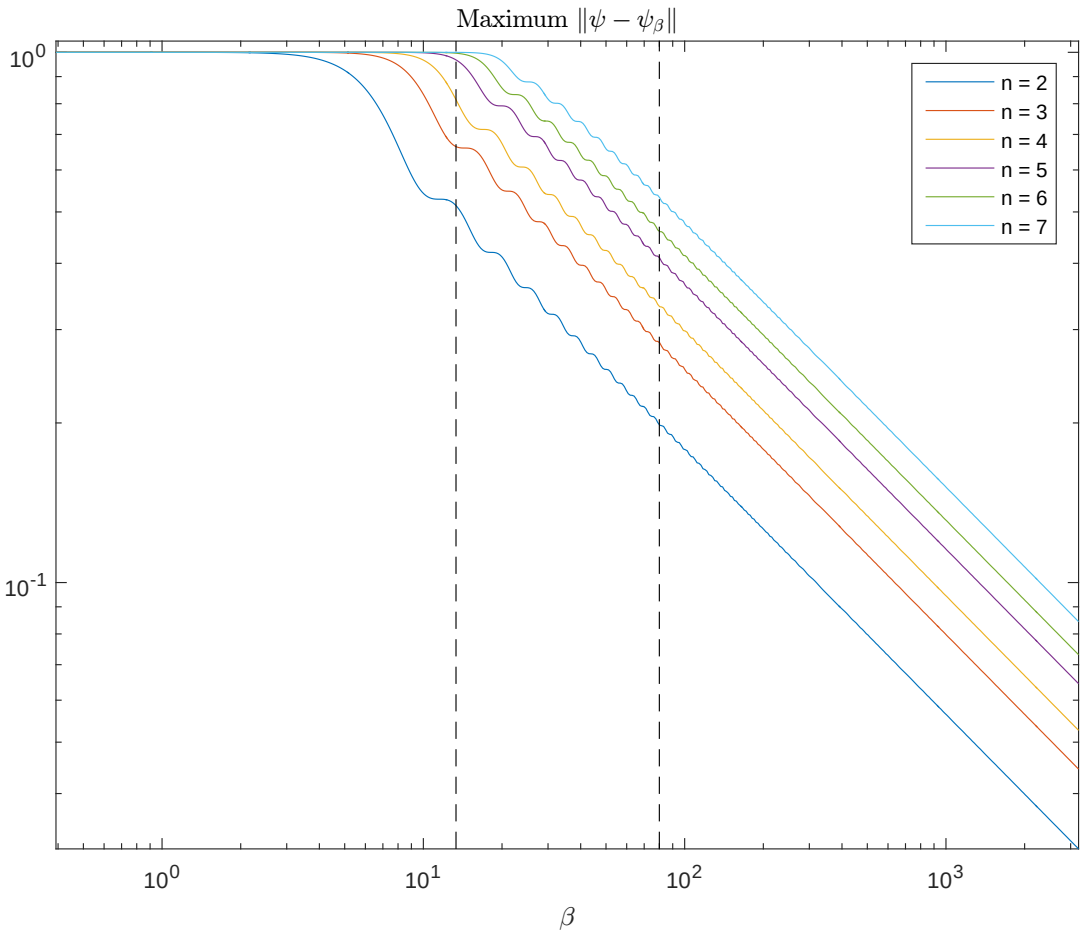


Figure 4.15: *Plots of the L_2 distance between ψ and ψ_β, maximised over wavelets ψ, against β and for $n = 2, 3, \ldots, 7$. The dashed vertical lines again correspond to the NAT accelerometer bandwidth (80Hz), and this value divided by 6Hz*

Using the projections above to take care of the conditions of the form $c^{\mathrm{T}} a = 0$, we wish to find $\tilde{a}_1$ minimising $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1$ subject to $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1 = 1$, where $\tilde{a}_1$ satisfies $Oa = \left[\begin{smallmatrix} 0 \\ \tilde{a}_1 \end{smallmatrix}\right]$. Then

the solution of the original problem is $a = O^{\mathrm{T}} \left[\begin{smallmatrix} 0 \\ \tilde{a}_1 \end{smallmatrix} \right]$.

To solve the projected problem, we simultaneously diagonalise the symmetric and positive definite matrices $\tilde{H}_1$ and $\tilde{P}_1(\beta)$.

Let $\tilde{H}_1 = O_H^{\mathrm{T}} D_H O_H$, where O_H is an orthogonal matrix and D_H is a diagonal matrix whose diagonal elements are the (positive) eigenvalues of $\tilde{H}_1$. As $\tilde{H}_1$ is positive definite and symmetric, such a decomposition exists. Then, if $D_H^{-\frac{1}{2}}$ is the diagonal matrix whose diagonal entries are the positive square roots of the reciprocals of the corresponding entries of D_H, $D_H^{-\frac{1}{2}} O_H \tilde{H}_1 O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} = I$ and $D_H^{-\frac{1}{2}} O_H \tilde{P}_1(\beta) O_H^{\mathrm{T}} D_H^{-\frac{1}{2}}$ is a positive definite symmetric matrix, so there exists a decomposition $O_P^{\mathrm{T}} D_P O_P = D_H^{-\frac{1}{2}} O_H \tilde{P}_1(\beta) O_H^{\mathrm{T}} D_H^{-\frac{1}{2}}$, where O_P is orthogonal and D_P is diagonal, with diagonal elements equal to the (positive) eigenvalues of $D_H^{-\frac{1}{2}} O_H \tilde{P}_1(\beta) O_H^{\mathrm{T}} D_H^{-\frac{1}{2}}$. Hence, $(O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}})^{\mathrm{T}} \tilde{H}_1 O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}} = I$ and $(O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}})^{\mathrm{T}} \tilde{P}_1(\beta) O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}} = D_P$.

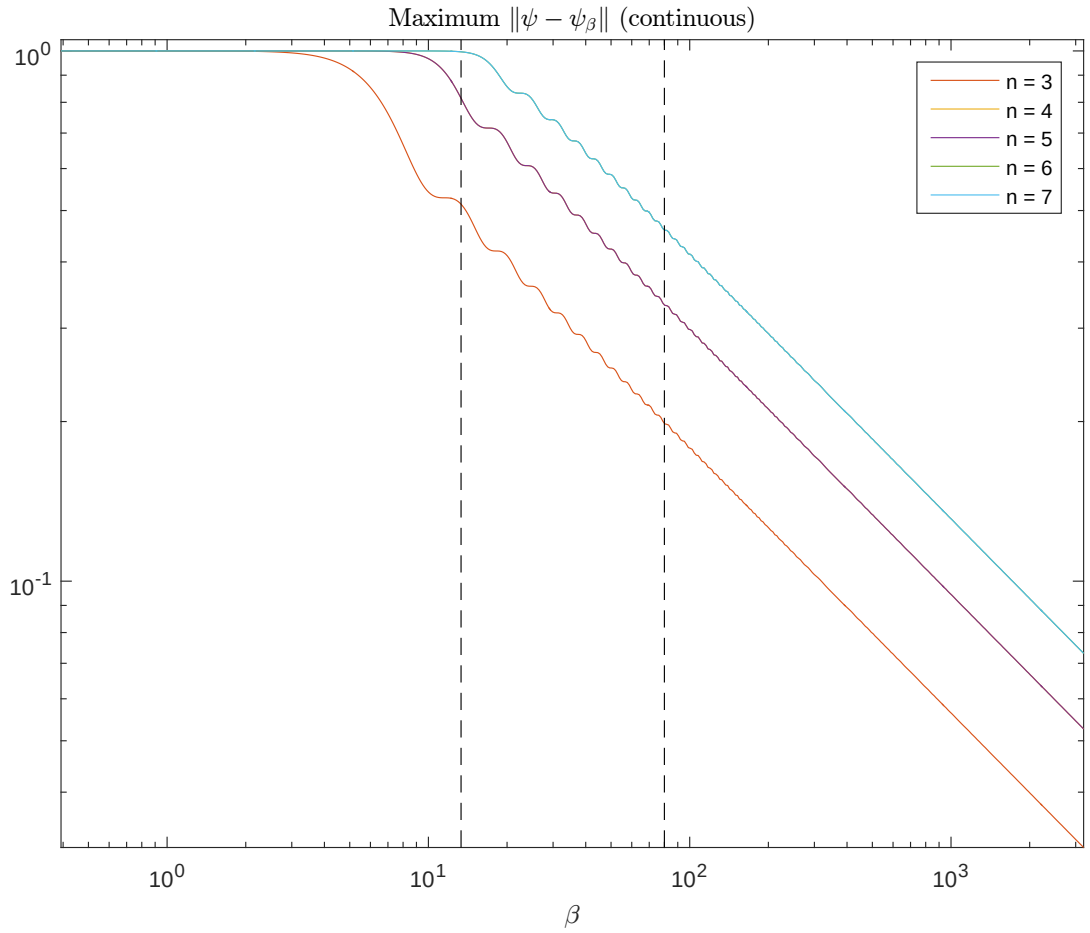


Figure 4.16: *Plots of the L_2 distance between ψ and ψ_β, maximised over wavelets ψ under wec, against β and for $n = 3, 4, \ldots, 7$ (note that the $n = 4$ curve is hidden by the $n = 5$ curve, and that the $n = 6$ curve is hidden by the $n = 7$ curve)*

Now put $\tilde{a}_1 = O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}} g$. Then $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1 = (O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}} g)^{\mathrm{T}} \tilde{H}_1 O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}} g = g^{\mathrm{T}} g$,

$\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1 = (O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}g)^{\mathrm{T}}\tilde{P}_1(\beta)O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}g = g^{\mathrm{T}}D_Pg$. Obviously, g minimising $g^{\mathrm{T}}D_Pg$, subject to $g^{\mathrm{T}}g = 1$, is a unit eigenvector of D_P corresponding to the least eigenvalue λ_0 of D_P: $D_Pg = \lambda_0 g$.

But then $D_PO_PD_H^{\frac{1}{2}}O_H\tilde{a}_1 = \lambda_0 O_PD_H^{\frac{1}{2}}O_H\tilde{a}_1$ as $g = O_PD_H^{\frac{1}{2}}O_H\tilde{a}_1$, and then, as $D_P = O_PD_H^{-\frac{1}{2}}O_H\tilde{P}_1(\beta)O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}$, $D_PO_PD_H^{\frac{1}{2}}O_H\tilde{a}_1 = O_PD_H^{-\frac{1}{2}}O_H\tilde{P}_1(\beta)\tilde{a}_1 = \lambda_0 O_PD_H^{\frac{1}{2}}O_H\tilde{a}_1$ or $\tilde{H}^{-1}\tilde{P}_1(\beta)\tilde{a}_1 = \lambda_0\tilde{a}_1$ on multiplying by $O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}$. Hence, $\tilde{a}_1$ is an eigenvector of $\tilde{H}_1^{-1}\tilde{P}_1(\beta)$ corresponding to the eigenvalue λ_0 (and it is easy to see that it is the least eigenvalue of $H_1^{-1}\tilde{P}_1(\beta)$).

However, $\tilde{a}_1$ is not, in general, a unit vector, as $\tilde{a}_1^{\mathrm{T}}\tilde{H}_1\tilde{a}_1 = 1$.

The value of $\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1$ at the minimum is simple to calculate: it is $\tilde{a}_1^{\mathrm{T}}\tilde{H}_1\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 = \lambda_0\tilde{a}_1^{\mathrm{T}}\tilde{H}_1\tilde{a}_1 = \lambda_0$. Consequently, the maximum L_2 distance between ψ and ψ_β is $\sqrt{1-\lambda_0} = \sqrt{1-\lambda_0(\beta)}$.

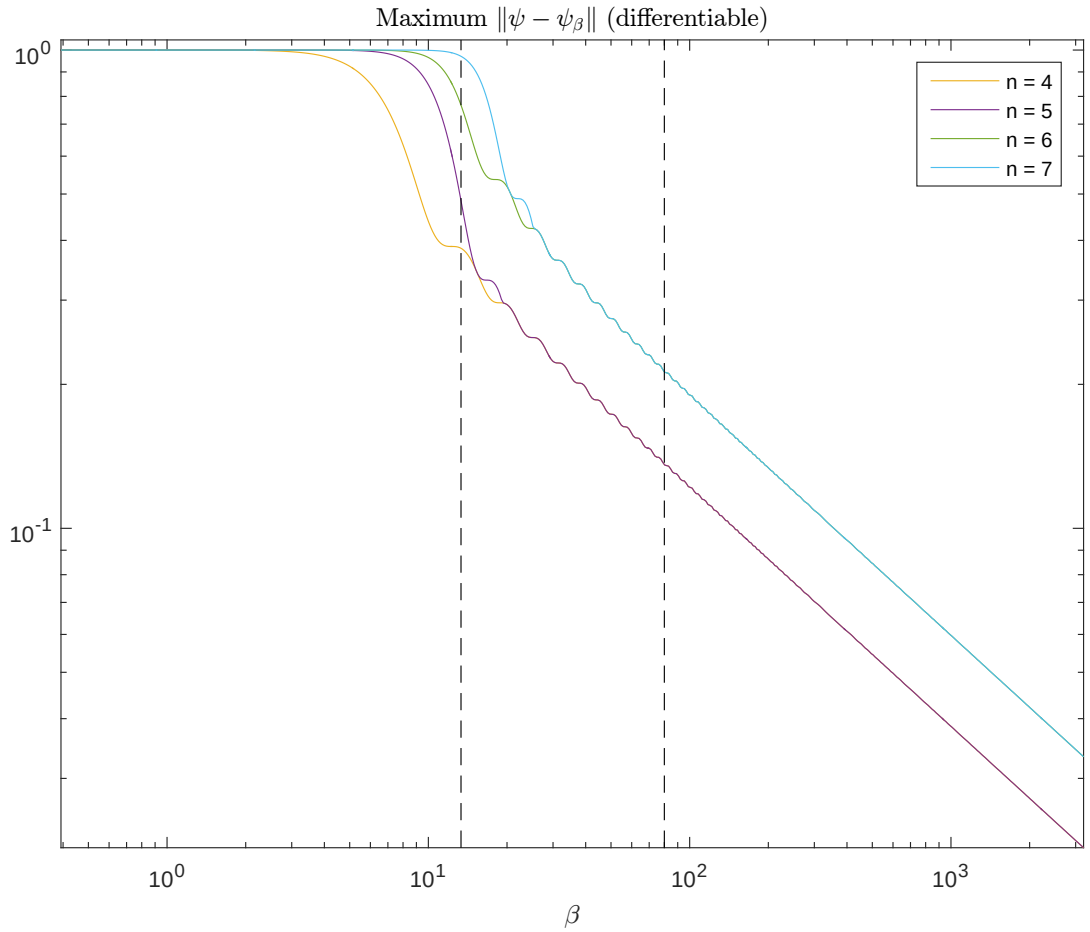


Figure 4.17: *Plots of the L_2 distance between ψ and ψ_β, maximised over wavelets ψ under wed, against β and for $n = 4, 5, 6, 7$*

The values of this for the case with no endpoint conditions are shown in Figure 4.15. In

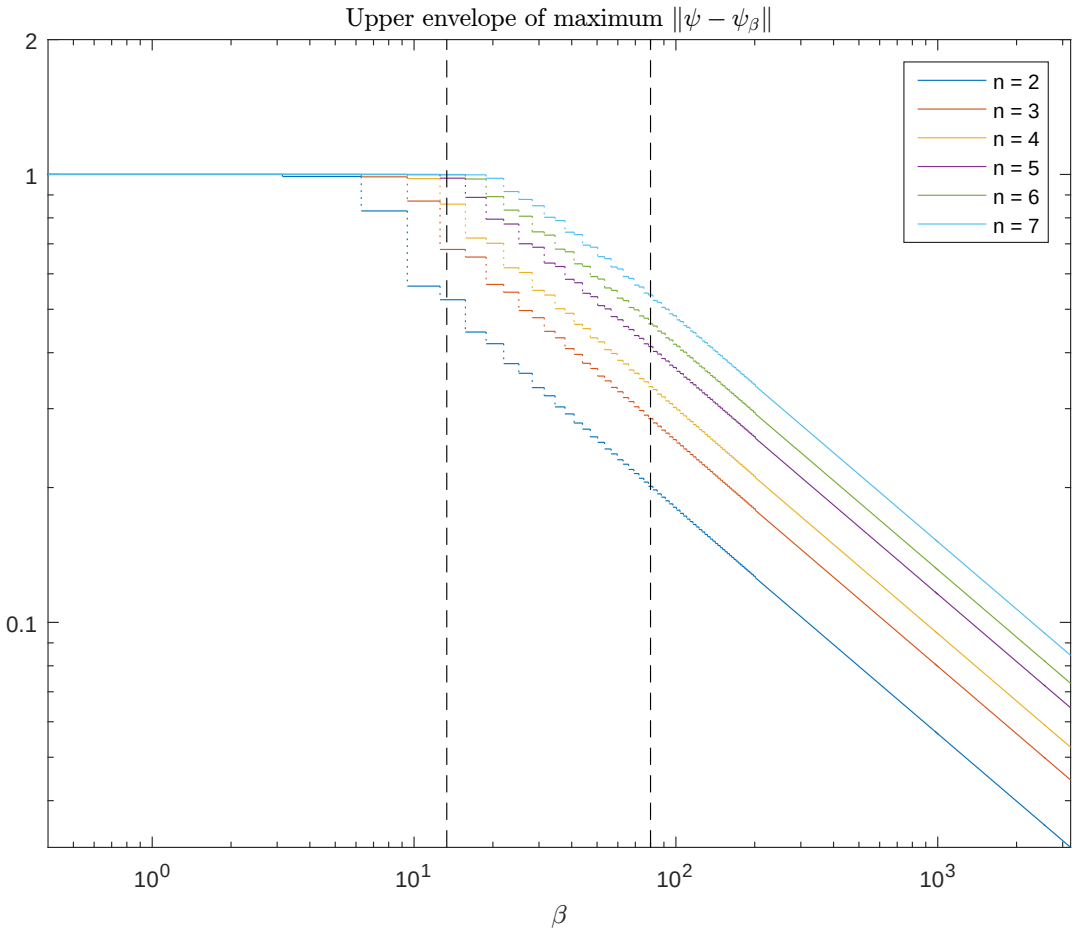


Figure 4.18: *Plots of the upper envelope of the distances of Figure 4.15*

the first two columns of Table 4.2[1] we show the values of the distances at $\beta = 80$Hz (the bandwidth of the NAT accelerometer) and $\beta = \frac{40}{3}$Hz (accelerometer bandwidth scaled by the typical signal frequency of Parkinson's movement disorders).

For general wavelets of our form, we see that the maximum of the distance $\|\psi - \psi_{80\text{Hz}}\|$ increases from 0.2012 to 0.5382 as n increases from 2 to 7. This increase is to be expected, as polynomials of degree n comprise a special case of the set of polynomials of degree $n+1$, so optimising over the latter is optimising over a larger set. Bearing in mind that the maximum L_2 distance between two functions ϕ, ϕ' on the unit sphere[2] ($\|\phi\|, \|\phi'\| = 1$) is only 2, and the

[1] Many of the entries in this Table are identical. In many cases, possibly all, this is not a coincidence. The matrices H and P can be transformed into block diagonal matrices, each with one $\lceil \frac{1}{2}(n+1) \rceil \times \lceil \frac{1}{2}(n+1) \rceil$ block (corresponding to the original even indices) and one $\lfloor \frac{1}{2}(n+1) \rfloor \times \lfloor \frac{1}{2}(n+1) \rfloor$ block (corresponding to the original odd indices), merely by permuting their indices. (Recall that our first index is 0, corresponding to x^0 in our polynomial.) But each of our projections either affects only the original even indices or only the original odd ones, so the eigenvalues of $\tilde{H}_1^{-1}\tilde{P}_1(\beta)$ fall into two groups, one of which is unaffected by a further projection. If the smallest eigenvalue of $\tilde{H}_1^{-1}\tilde{P}_1(\beta)$ before the additional projection is in the unaffected group, and no eigenvalue in the affected group becomes smaller, then the worst-case $\|\psi - \psi_\beta\|$ remains the same

[2] We are not claiming ψ_β is on the unit sphere — it is inside it — but only using such functions for a comparison

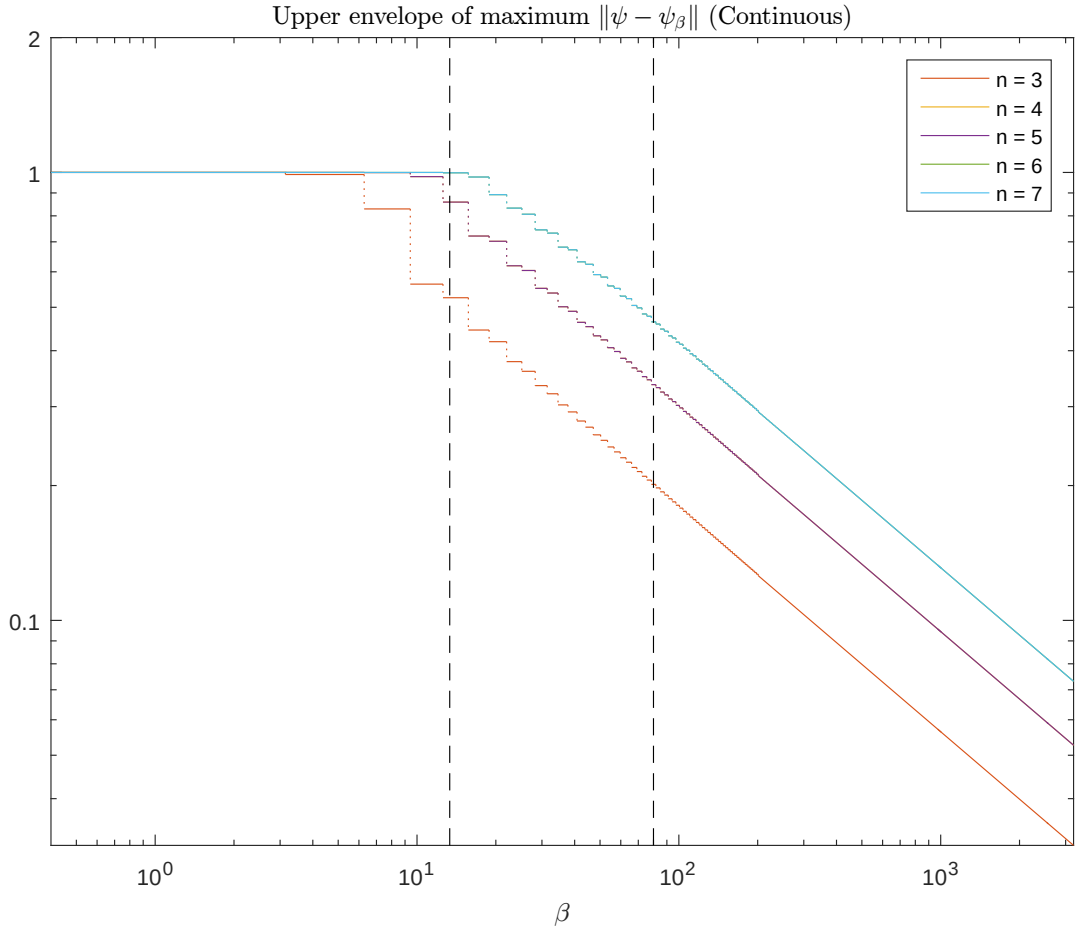


Figure 4.19: *Plots of the upper envelope of the distances of Figure 4.16*

maximum of $\min\{\|\phi - \phi'\|, \|\phi - (-\phi')\|\}$ is only $\sqrt{2}$ (achieved when ϕ and ϕ' are orthogonal, and more relevant here, as $\langle \psi, \psi_\beta \rangle > 0$), these values seem quite large.

The situation is worse (as might be expected) when we consider $\left\|\psi - \psi_{\frac{40}{3}\text{Hz}}\right\|$: this increases from 0.5252 to 0.9999 over the range of n considered.

For the maximum values of $\|\psi - \psi_\beta\|$ and, in particular, $\|\psi - \psi_{80\text{Hz}}\|$ and $\left\|\psi - \psi_{\frac{40}{3}\text{Hz}}\right\|$ when the wrapped endpoint continuity condition is imposed, see Figure 4.16 and the middle two columns of Table 4.2: the former range from 0.2012 to 0.4641, and the latter from 0.5252 to 0.9981, so the values remain large for the two special values of β.

When the wrapped endpoint differentiability conditions are imposed, $\|\psi - \psi_\beta\|$ is shown in Figure 4.17, and $\|\psi - \psi_\beta\|$ and, in particular, $\|\psi - \psi_{80\text{Hz}}\|$ and $\left\|\psi - \psi_{\frac{40}{3}\text{Hz}}\right\|$ in the last two columns of Table 4.2, and the values in the Table are still quite large, ranging from 0.1380 to 0.2158 and from 0.3883 to 0.9837.

As our general piecewise polynomial wavelets of nth degree have $n - 1$ free parameters, and the imposition of each of the wrapped endpoint conditions sacrifices one of these free parameters, if we wish to compare wavelets with the same number of parameters, we should

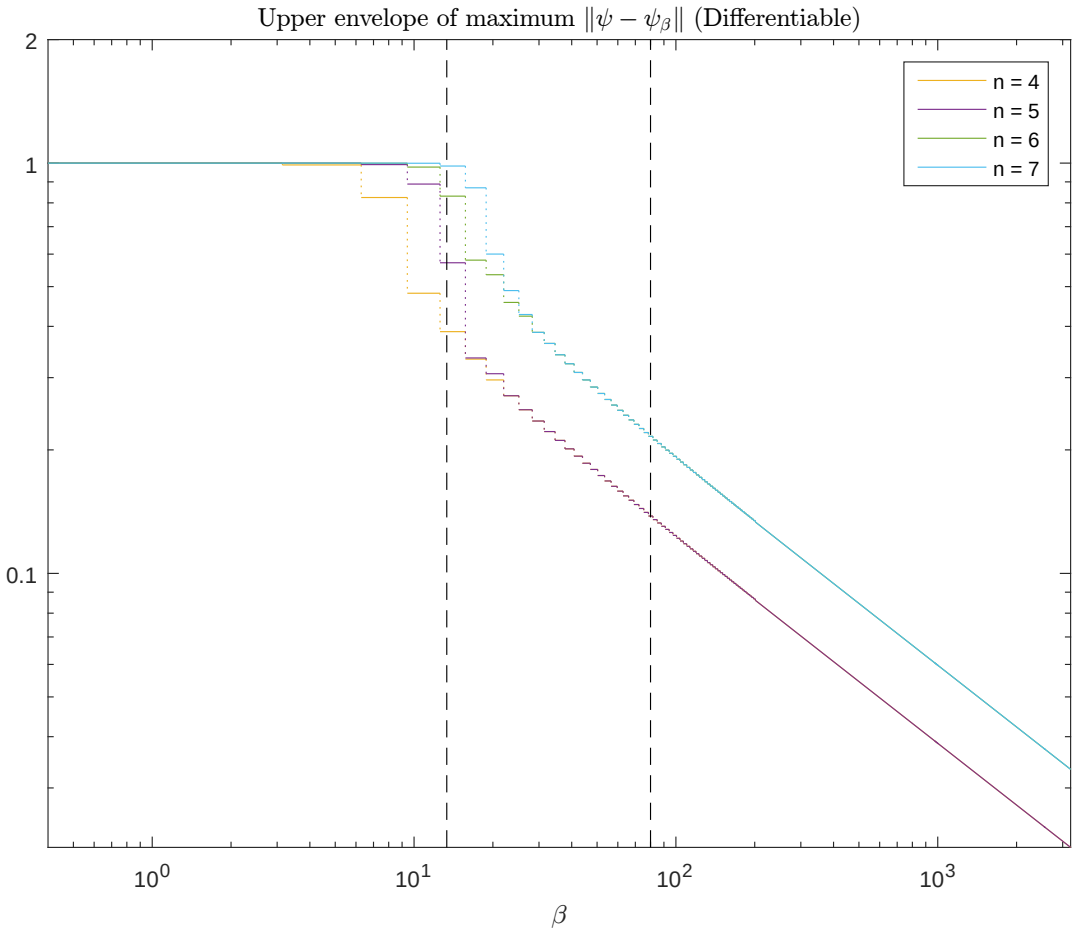


Figure 4.20: *Plots of the upper envelope of the distances of Figure 4.17*

	Wavelets			**Wavelets with wec**			**Wavelets with wed**		
	n	$\beta = \frac{40}{3}$Hz	$\beta = 80$Hz	n	$\beta = \frac{40}{3}$Hz	$\beta = 80$Hz	n	$\beta = \frac{40}{3}$Hz	$\beta = 80$Hz
p		$\|\psi - \psi_\beta\|$			$\|\psi - \psi_\beta\|$			$\|\psi - \psi_\beta\|$	
1	2	0.5252	0.2012	3	0.5252	0.2012	4	0.3883	0.1380
2	3	0.6800	0.2850	4	0.8587	0.3360	5	0.5716	0.1380
3	4	0.8587	0.3360	5	0.8587	0.3360	6	0.8308	0.2158
4	5	0.9810	0.4129	6	0.9981	0.4641	7	0.9837	0.2158
5	6	0.9981	0.4641	7	0.9981	0.4641	—	—	—
6	7	0.9999	0.5382	—	—	—	—	—	—

Table 4.2: *Values for the maxima of* $\left\|\psi - \psi_{\frac{40}{3}\text{Hz}}\right\|$ *and* $\|\psi - \psi_{80\text{Hz}}\|$ *over* ψ. n *is the degree of the polynomial on which the wavelets are based, and* p *is the number of free parameters. Values which are identical (to within the numerical accuracy) for the same* β *are given in coloured text of the same colour*

compare the nth degree general wavelets with the $(n+1)$th degree wavelets with the wrapped endpoint continuity condition imposed and the $(n+2)$th wavelets with the wrapped endpoint

differentiability conditions.

If we do this, we find that, for even n, the distances at $\beta = 80$Hz and $\frac{40}{3}$Hz are the same for general piecewise polynomial wavelets of degree n and for piecewise polynomial wavelets of degree $(n+1)$ with wrapped endpoint continuity, and for odd n, the distances for the latter actually increase, so, at first sight, there appears to be little point in imposing wrapped endpoint continuity on its own.

However, when we compare these distances for the general piecewise polynomial wavelets of degree n and for piecewise polynomial wavelets of degree $(n+2)$ with the wrapped endpoint differentiability conditions, we find that the latter are significantly smaller, except for the case $n = 5$ at $\beta = \frac{40}{3}$Hz, so there is an indication that imposing both conditions may improve performance without sacrificing the number of free parameters, at least for small n.

	Wavelets			**Wavelets with wec**			**Wavelets with wed**		
p	n	A	B	n	A	B	n	A	B
1	2	1.7847	-0.5000	3	1.7847	-0.5000	4	1.2223	-0.5004
2	3	2.5228	-0.5000	4	2.9846	-0.5000	5	1.2223	-0.5004
3	4	2.9846	-0.5000	5	2.9846	-0.5000	6	1.9102	-0.5015
4	5	3.6489	-0.4997	6	4.1315	-0.4995	7	1.9102	-0.5015
5	6	4.1315	-0.4995	7	4.1315	-0.4995	—	—	—
6	7	4.7499	-0.4989	—	—	—	—	—	—

Table 4.3: *Parameters for the approximate relation* $\|\psi - \psi_\beta\| = A\beta^B$ *which holds when* $\beta > 80$*Hz. Parameter pairs which are identical (to within the numerical accuracy) are given in coloured text of the same colour*

It will be noted that the curves in Figures 4.15 to 4.17 start off fairly flat, remaining very close to 1, before making a transition to being approximate straight lines with significant negative gradients to the right of $\beta = 80$Hz. Ironically, the transition commences around $\beta = \frac{40}{3}$Hz, between $\beta = \frac{1}{2}\frac{40}{3}$Hz and $\beta = 2\frac{40}{3}$Hz.

Thus, increasing the bandwidth of the accelerometer (equivalent to shifting the dashed lines in Figures 4.15 to 4.17 by the same distance to the right, but otherwise leaving the Figures unchanged), is likely to be rewarded by improved performance.

As both scales in these Figures are logarithmic, under the straight-line regime to the right of 80Hz, $\|\psi - \psi_\beta\| \approx A\beta^B$ for some A and B. By applying a least-squares regression to the pairs $(\ln\beta, \ln\|\psi - \psi_\beta\|)$ for $\beta \geq 80$Hz, we obtain the As and Bs collated in Table 4.3.

Further approximating, we can sum up Table 4.3 by saying that $\|\psi - \psi_\beta\| \approx \frac{A(n)}{\sqrt{\beta}}$, where $A(n) \in [2, 5]$ and $A(n)$ has a further dependence on the conditions imposed on ψ. Thus, replacing our accelerometers by ones with a bandwidth four times greater will reduce the worst-case $\|\psi - \psi_\beta\|$ by a factor of 2, when β is reasonably large.

Outside of this Subsection, we will refer to the wavelet ψ which maximises $\|\psi - \psi_\beta\|$ as $\psi_{(1)}$ (because $\|\psi - \psi_\beta\|$ involves one wavelet), so ψ_β becomes $\psi_{(1)\beta}$.

4.7.2 Maximum distance between ψ and ψ' such that $\|\psi' - \psi_\beta\| \leq \|\psi - \psi_\beta\|$

The use of the worst-case $\|\psi - \psi_\beta\|$ to assess the effect of limited accelerometer bandwidth on the possibility of confusing different wavelets suffers from the fact that ψ_β does not qualify as one of our wavelets.

Instead, we can use the worst-case $\|\psi - \psi'\|$, where ψ' is one of our wavelets, and is closer to ψ_β than, or at most is at the same distance as, ψ. In other words, $\|\psi' - \psi_\beta\| \leq \|\psi - \psi_\beta\|$.

That is, we seek to maximise $\|\psi - \psi'\|$ over wavelets ψ and ψ', subject to $\|\psi' - \psi_\beta\| \leq \|\psi - \psi_\beta\|$, which means that ψ and ψ' both obey the ue and zi conditions, and we may choose to impose either the wrapped endpoint continuity condition or the wrapped endpoint differentiability conditions on both of them.

To translate to "a" language, we set $\psi'(x) = \sum_{k=0}^{n} a'_k x^k$, and then we need to maximise $\|\psi - \psi'\|^2 = \|\psi\|^2 - 2\langle\psi, \psi'\rangle + \|\psi'\|^2 = 2(1 - \langle\psi, \psi'\rangle) = 2(1 - a^{\mathrm{T}} H a')$ with respect to a and a', subject to the conditions $a^{\mathrm{T}} H a = a'^{\mathrm{T}} H a' = 1$, $b^{\mathrm{T}} a = b^{\mathrm{T}} a' = 0$ and $\|\psi' - \psi_\beta\|^2 = \|\psi'\|^2 - 2\langle\psi', \psi_\beta\rangle + \|\psi_\beta\|^2 = 1 - 2a^{\mathrm{T}} P(\beta) a' + a^{\mathrm{T}} P(\beta) a \leq \|\psi - \psi_\beta\|^2 = 1 - a^{\mathrm{T}} P(\beta) a$. Optionally, we also impose $c_1^{\mathrm{T}} a = c_1^{\mathrm{T}} a' = 0$ or $c_1^{\mathrm{T}} a = c_1^{\mathrm{T}} a' = c_2^{\mathrm{T}} a = c_2^{\mathrm{T}} a' = 0$.

This is equivalent to minimising $a^{\mathrm{T}} H a'$ with respect to a and a', subject to $a^{\mathrm{T}} H a = a'^{\mathrm{T}} H a' = 1$, $b^{\mathrm{T}} a = b^{\mathrm{T}} a' = 0$ and $a^{\mathrm{T}} P(\beta) a' \geq a^{\mathrm{T}} P(\beta) a$, with or without one of the optional conditions.

If we make the projections of the previous section, we reduce this to the problem of minimising $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1$ with respect to $\tilde{a}_1$ and $\tilde{a}'_1$, subject to $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1 = \tilde{a}_1'^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1 = 1$ and $\tilde{a}^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}'_1 \geq \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1$, and we recover a and a' through $a = O^{\mathrm{T}} \left[\begin{smallmatrix} 0 \\ \tilde{a}_1 \end{smallmatrix}\right]$ and $a' = O^{\mathrm{T}} \left[\begin{smallmatrix} 0 \\ \tilde{a}'_1 \end{smallmatrix}\right]$.

We first fix a (and hence $\tilde{a}_1$), and seek to minimise $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1$, with respect to $\tilde{a}'_1$, subject to $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}'_1 \geq \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1$ and $\tilde{a}_1'^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1 = 1$, where it is given that $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1 = 1$.

Initially, we ignore the condition $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}'_1 \geq \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1$, and minimise $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1$, subject to just $\tilde{a}_1'^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1 = 1$.

Starting by using the method of Lagrange multipliers, we look for $\tilde{a}'_1$ satisfying the conditions for saddle points of

$$Q(\tilde{a}'_1, \lambda) = \tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1 + \lambda[\tilde{a}_1'^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1 - 1]. \tag{4.69}$$

We have that $\frac{\partial}{\partial \tilde{a}_1^{\mathrm{T}}} Q(\tilde{a}'_1, \lambda) = 2\lambda \tilde{H}_1 \tilde{a}'_1 + \tilde{H}_1 \tilde{a}_1$, so $\frac{\partial}{\partial \tilde{a}_1^{\mathrm{T}}} Q(\tilde{a}'_1, \lambda) = 0 \Rightarrow \tilde{a}'_1 = -\frac{1}{2\lambda} \tilde{a}_1$.

Rather than continuing down the Lagrangian path, we apply the condition $\tilde{a}_1'^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1 = 1$ to this partial solution directly, and soon find that $\lambda = \pm\frac{1}{2}$ and $\tilde{a}'_1 = \mp\tilde{a}_1$. The solution $\tilde{a}'_1 = \tilde{a}_1$ maximises $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}'_1$, and the solution $\tilde{a}'_1 = -\tilde{a}_1$ violates the hitherto ignored condition $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}'_1 \geq \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1 > 0$ (as $\tilde{P}_1(\beta)$ is positive definite and clearly $\tilde{a}_1 \neq 0$).

Since there are no valid solutions away from the "boundary" of the condition $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}'_1 \geq \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1$, we can replace it by the corresponding equality constraint, and, within the Lagrangian formulation, there is no constraint on the multiplier corresponding to this new constraint.

Thus, we find the condition on $\tilde{a}_1'$ for a saddlepoint of

$$\begin{aligned} Q(\tilde{a}_1', \lambda, \kappa) &= \tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1' + \lambda[\tilde{a}_1'^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1' - 1] + \kappa[\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1' - \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1] \\ &= \lambda \tilde{a}_1'^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1' + [\tilde{H}_1 \tilde{a}_1 + \kappa \tilde{P}_1(\beta) \tilde{a}_1]^{\mathrm{T}} \tilde{a}_1' - \lambda - \kappa \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1. \end{aligned} \tag{4.70}$$

This is $\frac{\partial}{\partial \tilde{a}_1^{\mathrm{T}}} Q(\tilde{a}_1', \lambda, \kappa) = 0$ or $2\lambda \tilde{H}_1 \tilde{a}_1' + [\tilde{H}_1 + \kappa \tilde{P}_1(\beta)] \tilde{a}_1 = 0$, so $\tilde{a}_1' = -\frac{1}{2\lambda}[\tilde{a}_1 + \kappa \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1]$. We again complete this partial solution by applying the conditions directly to it to find suitable values of λ and κ.

The constraint $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1' = \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1$ yields

$$-\tfrac{1}{2\lambda}[\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1 + \kappa \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1] = \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1 \tag{4.71}$$

or

$$\kappa = -\tfrac{(1+2\lambda) \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1}{\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1}. \tag{4.72}$$

Then

$$\tilde{a}_1' = -\tfrac{1}{2\lambda} \left[\tilde{a}_1 - \tfrac{(1+2\lambda) \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1}{\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1} \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1 \right] \tag{4.73}$$

and then the condition $\tilde{a}_1'^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1' = 1$ becomes

$$\tfrac{1}{4\lambda^2} \left[1 - \tfrac{2(1+2\lambda)(\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1)^2}{\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1} + \tfrac{(1+2\lambda)^2 (\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1)^2}{\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1} \right] = 1, \tag{4.74}$$

which, provided that $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1 \neq (\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1)^2$, has the solutions $\lambda = \pm \frac{1}{2}$ as before, and, also as before, the solution $\lambda = -\frac{1}{2}$ leads to the trivial, maximising, value of $\tilde{a}_1$ for $\tilde{a}_1'$. However, the solution $\lambda = \frac{1}{2}$ results in

$$\tilde{a}_1' = 2 \tfrac{\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1}{\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1} \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1 - \tilde{a}_1, \tag{4.75}$$

and then

$$\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1' = 2 \tfrac{(\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1)^2}{\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1} - 1. \tag{4.76}$$

If $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1 = (\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1)^2$, we have $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) [\tilde{H}_1^{-1} - \tilde{a}_1 \tilde{a}_1^{\mathrm{T}}] \tilde{P}_1(\beta) \tilde{a}_1 = 0$, which is certainly possible, as $\det(\tilde{H}_1^{-1} - \tilde{a}_1 \tilde{a}_1^{\mathrm{T}}) = \det(\tilde{H}_1^{-1}) \det(\mathrm{I} - \tilde{H}_1 \tilde{a}_1 \tilde{a}_1^{\mathrm{T}}) = \det(\tilde{H}_1^{-1})(1 - \tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1) = 0$. But equation (4.73) shows that $\tilde{a}_1'^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1' = \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1'$ and $\tilde{a}_1'^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1' = 1$, and $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1'$ has its maximum value of 1, no matter what the value of λ, under these circumstances. Since we wish to minimise $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1'$, we may rule out any $\tilde{a}_1$ such that $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1 = (\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1)^2$.

We now wish to minimise $\frac{(\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1)^2}{\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1}$ with respect to $\tilde{a}_1$, subject to $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1 = 1$, in order to make the left-hand side of equation (4.76) as small as possible.

Here we use the decompositions and notation of the previous section, $\tilde{H}_1 = O_H^{\mathrm{T}} D_H O_H$ and $D_H^{-\frac{1}{2}} O_H \tilde{P}_1(\beta) O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} = O_P^{\mathrm{T}} D_P O_P$, so $(O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}})^{\mathrm{T}} \tilde{H}_1 O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}} = I$ and $(O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}})^{\mathrm{T}} \tilde{P}_1(\beta) O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}} = D_P$.

Put $\tilde{a}_1 = O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}} g$ as before. Then $\tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1 = g^{\mathrm{T}} g$, $\tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1 = g^{\mathrm{T}} D_P g$.

Also, $\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 = (O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}g)^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}g = g^{\mathrm{T}}O_PD_H^{-\frac{1}{2}}O_HO_H^{\mathrm{T}}D_H^{\frac{1}{2}}O_P^{\mathrm{T}}D_PO_PD_H^{\frac{1}{2}}O_HO_H^{\mathrm{T}}D_H^{-1}O_HO_H^{\mathrm{T}}D_H^{\frac{1}{2}}O_P^{\mathrm{T}}D_PO_PD_H^{\frac{1}{2}}O_HO_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}g = g^{\mathrm{T}}D_P^2g$.

If $p = \begin{cases} n-1, & \text{case 1;} \\ n-2, & \text{case 2;} \\ n-3, & \text{case 3} \end{cases}$ and the eigenvalues of $D_H^{-\frac{1}{2}}O_H\tilde{P}_1(\beta)O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}$, $\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_p$, are arranged down the diagonal of D_P, then $g^{\mathrm{T}}g = \sum_{k=0}^{p} g_k^2$, $g^{\mathrm{T}}D_Pg = \sum_{k=0}^{p}\lambda_k g_k^2$ and $g^{\mathrm{T}}D_P^2g = \sum_{k=0}^{p}\lambda_k^2 g_k^2$, and, on putting $g_k^2 = h_k$, our problem becomes that of minimising $F(h) = \frac{\left[\sum_{k=0}^{p}\lambda_k h_k\right]^2}{\sum_{k=0}^{p}\lambda_k^2 h_k}$, subject to $\sum_{k=0}^{p} h_k = 1$ and $h_k \geq 0; \forall k$.

If we assume that precisely p_1+1 of the h_k are nonzero, and that the h_k and corresponding diagonal elements of D_P are relabelled so that the nonzero h_k become $h_0, h_1, h_2, \ldots, h_{p_1}$, with $\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{p_1}$, then $F(h) = \frac{\left[\sum_{k=0}^{p_1}\lambda_k h_k\right]^2}{\sum_{k=0}^{p_1}\lambda_k^2 h_k} = \frac{\left[\lambda_0+\sum_{k=1}^{p_1}(\lambda_k-\lambda_0)h_k\right]^2}{\lambda_0^2+\sum_{k=1}^{p_1}(\lambda_k^2-\lambda_0^2)h_k}$, when the condition $\sum_{k=0}^{p} h_k = \sum_{k=0}^{p_1} h_k = 1$ is taken into account.

Treating $F(h)$ as a function of the variables $h_1, h_2, \ldots, h_{p_1}$, the usual condition for an extremum of F, $\frac{\partial F}{\partial h_k} = 0, k = 1, 2, \ldots, p_1$, implies

$$(\lambda_\ell - \lambda_0)\left\{2\left[\lambda_0^2 + \sum_{k=1}^{p_1}(\lambda_k^2 - \lambda_0^2)h_k\right] - (\lambda_\ell + \lambda_0)\left[\lambda_0 + \sum_{k=1}^{p_1}(\lambda_k - \lambda_0)h_k\right]\right\} = 0, \qquad (4.77)$$

$\ell = 1, 2, \ldots, p_1$, and, if $\lambda_1 = \lambda_2 = \ldots = \lambda_{p_0-1} = \lambda_0$, we have

$$\begin{aligned} 2\left[\lambda_0^2 + \sum_{k=p_0}^{p_1}(\lambda_k^2 - \lambda_0^2)h_k\right] - (\lambda_\ell + \lambda_0)\left[\lambda_0 + \sum_{k=p_0}^{p_1}(\lambda_k - \lambda_0)h_k\right] &= \\ \lambda_0^2 + \sum_{k=p_0}^{p_1}(\lambda_k - \lambda_0)(2\lambda_k + \lambda_0)h_k - \lambda_\ell\left[\lambda_0 + \sum_{k=p_0}^{p_1}(\lambda_k - \lambda_0)h_k\right] &= 0, \end{aligned} \qquad (4.78)$$

$\ell = p_0, p_0+1, p_0+2, \ldots, p_1$, which means that either $\lambda_{p_0}, \lambda_{p_0+1}, \lambda_{p_0+2}, \ldots, \lambda_{p_1}$ have a common value, or $\lambda_0^2 + \sum_{k=p_0}^{p_1}(\lambda_k - \lambda_0)(2\lambda_k + \lambda_0)h_k = \lambda_0 + \sum_{k=p_0}^{p_1}(\lambda_k - \lambda_0)h_k = 0$. But the second alternative is impossible, as λ_k, $\lambda_k - \lambda_0$ and h_k are positive for $k = p_0, p_0+1, p_0+2, \ldots, p_1$, as is λ_0, so we must have $\lambda_{p_0} = \lambda_{p_0+1} = \lambda_{p_0+2} = \ldots = \lambda_{p_1} = \mu$, say.

Returning to our original labelling, this means that at an extremum of $F(h)$, the set $\{\lambda_k : h_k > 0\}$ contains either 1 or 2 values. If it contains one value, then $F(h) = 1$, a maximum. Suppose $\{\lambda_k : h_k > 0\} = \{\mu_0, \mu_1\}$, $\mu_0 < \mu_1$. Then $F(h) = \frac{\left[\mu_0+(\mu_1-\mu_0)\sum_{\{k:\lambda_k=\mu_1\}}h_k\right]^2}{\mu_0^2+(\mu_1^2-\mu_0^2)\sum_{\{k:\lambda_k=\mu_1\}}h_k} = \frac{[\mu_0+A(\mu_1-\mu_0)]^2}{\mu_0^2+A(\mu_1^2-\mu_0^2)}$ depends on h only via $A = \sum_{\{k:\lambda_k=\mu_1\}} h_k$.

But $\frac{\mathrm{d}F}{\mathrm{d}A} = 0$ implies $2(\mu_1-\mu_0)[\mu_0^2+A(\mu_1^2-\mu_0^2)]-(\mu_1^2-\mu_0^2)[\mu_0+A(\mu_1-\mu_0)] = 0$, which in turn implies $A = \frac{\mu_0}{\mu_0+\mu_1} \in (0,1)$, which corresponds to the local minimum $\frac{4\mu_0\mu_1}{(\mu_0+\mu_1)^2} = \frac{4}{\left(\frac{\mu_0}{\mu_1}\right)^2+\left(\frac{\mu_1}{\mu_0}\right)^2}$.

The global minimum is given by $\min_{(\mu_0,\mu_1)\in\{(\mu_0',\mu_1'):\mu_0',\mu_1'\in\{\lambda_0,\ldots,\lambda_p\},\mu_0<\mu_1\}}\left\{\frac{4}{\left(\frac{\mu_0}{\mu_1}\right)^2+\left(\frac{\mu_1}{\mu_0}\right)^2}\right\} =$

$\frac{4m_\lambda M_\lambda}{(m_\lambda+M_\lambda)^2}$, where $m_\lambda = \min\{\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_p\}$, $M_\lambda = \max\{\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_p\}$, and this global minimum will be less than 1 if $\{\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_p\}$ contains at least two distinct values.

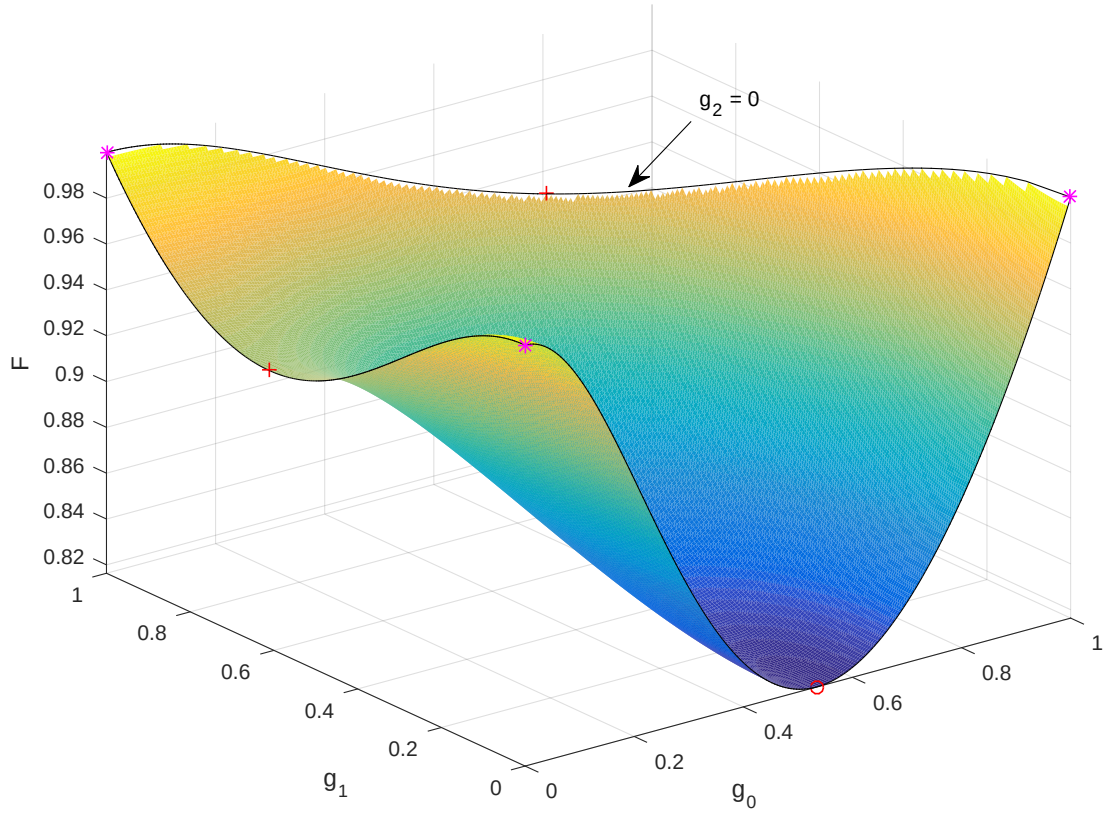


Figure 4.21: *Three dimensional plot of $F(h)$ against g_0 and g_1 for the case $p = 3$, with $\lambda_0 = \frac{1}{2}$, $\lambda_1 = \frac{1}{3}$ and $\lambda_2 = \frac{1}{5}$. The three maxima of value 1 at $g = [1, 0, 0]^{\mathrm{T}}$, $[0, 1, 0]^{\mathrm{T}}$ and $[0, 0, 1]^{\mathrm{T}}$ are marked with magenta stars, and the two merely local minima of $\frac{4\lambda_0\lambda_1}{(\lambda_0+\lambda_1)^2} = \frac{24}{25}$ at $g = \left[\sqrt{\frac{\lambda_1}{\lambda_0+\lambda_1}}, \sqrt{\frac{\lambda_0}{\lambda_0+\lambda_1}}, 0\right]^{\mathrm{T}} = \left[\sqrt{\frac{2}{5}}, \sqrt{\frac{3}{5}}, 0\right]^{\mathrm{T}}$ and $\frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2} = \frac{15}{16}$ at $g = \left[0, \sqrt{\frac{\lambda_2}{\lambda_1+\lambda_2}}, \sqrt{\frac{\lambda_1}{\lambda_1+\lambda_2}}\right]^{\mathrm{T}} = \left[0, \sqrt{\frac{5}{8}}, \sqrt{\frac{3}{8}}\right]^{\mathrm{T}}$ are shown by red crosses. The global minimum of $\frac{4\lambda_0\lambda_2}{(\lambda_0+\lambda_2)^2} = \frac{40}{49}$ at $g = \left[\sqrt{\frac{\lambda_2}{\lambda_0+\lambda_2}}, 0, \sqrt{\frac{\lambda_0}{\lambda_0+\lambda_2}}\right]^{\mathrm{T}} = \left[\sqrt{\frac{2}{7}}, 0, \sqrt{\frac{5}{7}}\right]^{\mathrm{T}}$ is given a red circle. The mismatch between the boundary of the coloured surface and the $g_2 = 0$ curve is due to the limited resolution of the plot*

An example of the relationship between F and its extrema is given in Figure 4.21.

If $h' \in \arg\min F(h)$, then $h'_k = 0$ if $\lambda_k \notin \{m_\lambda, M_\lambda\}$, $\sum_{k\in\{k':\lambda_{k'}=m_\lambda\}} h'_k = \frac{M_\lambda}{m_\lambda+M_\lambda}$ and $\sum_{k\in\{k':\lambda_{k'}=M_\lambda\}} h'_k = \frac{m_\lambda}{m_\lambda+M_\lambda}$, so h' is not uniquely determined unless both the maximum and minimum eigenvalue are of multiplicity 1.

Nevertheless, although the solution is not necessarily unique, it exists, and we can choose h, and thus g, minimising $\frac{g^{\mathrm{T}}D_P^2 g}{g^{\mathrm{T}}D_P g}$ subject to $g^{\mathrm{T}}g = 1$, so $\tilde{a}_1 = O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} O_P^{\mathrm{T}} g$ minimises $\frac{(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1)^2}{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1}$ subject to $\tilde{a}_1^{\mathrm{T}}\tilde{H}_1\tilde{a}_1 = 1$, and this minimum has the value $\frac{4m_\lambda M_\lambda}{(m_\lambda+M_\lambda)^2}$, where m_λ and M_λ are the minimum and maximum eigenvalues of D_P. If u is an eigen-

vector of D_P, with eigenvalue ν, then $D_P u = \nu u \Rightarrow D_H^{-\frac{1}{2}} O_H \tilde{P}_1(\beta) O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} u = \nu u \Rightarrow O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} D_H^{-\frac{1}{2}} O_H \tilde{P}_1(\beta) O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} D_H^{-\frac{1}{2}} O_H O_H^{\mathrm{T}} D_H^{\frac{1}{2}} u = \tilde{H}_1^{-1} \tilde{P}_1(\beta) O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} u = \nu O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} u$, so ν is an eigenvalue of $\tilde{H}_1^{-1} \tilde{P}_1(\beta)$ with eigenvector $O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} u$. Moreover,

$$
\begin{aligned}
(O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} u)^{\mathrm{T}} \tilde{H}_1 O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} u' &= u^{\mathrm{T}} D_H^{-\frac{1}{2}} O_H O_H^{\mathrm{T}} D_H O_H O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} u' = u^{\mathrm{T}} u' \\
&= \begin{cases} 0, & O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} u \text{ and } O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} u' \text{ are eigenvectors of } \tilde{H}_1^{-1} \tilde{P}_1(\beta) \text{ corresponding to distinct eigenvalues;} \\ 1, & O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} u' \text{ and } O_H^{\mathrm{T}} D_H^{-\frac{1}{2}} u \text{ are the same eigenvector of } \tilde{H}_1^{-1} \tilde{P}_1(\beta); \\ \text{some value in } [-1,1], & \text{otherwise.} \end{cases}
\end{aligned}
$$

Consequently, the minimum value of $\langle \psi, \psi' \rangle$ is given by $\frac{8 m_\lambda M_\lambda}{(m_\lambda + M_\lambda)^2} - 1$, where m_λ and M_λ are the minimum and maximum eigenvalues of $\tilde{H}_1^{-1} \tilde{P}_1(\beta)$, and the maximum distance is $\|\psi - \psi'\| = \sqrt{2\left[2 - \frac{8 m_\lambda M_\lambda}{(m_\lambda + M_\lambda)^2}\right]} = 2 \frac{M_\lambda - m_\lambda}{M_\lambda + m_\lambda}$.

In Figures 4.22 to 4.24 we show the worst-case $\|\psi - \psi'\|$ for general wavelets of our form, wavelets under the wrapped endpoint continuity condition, and wavelets under both the wrapped endpoint differentiability conditions. In and around the range $\beta \in [10, 100]$, the curves clearly have an important oscillatory component, so, in Figures 4.25 to 4.27, we plot their upper envelop[3], in order to avoid underestimating $\|\psi - \psi'\|$ because of minor inaccuracies in the effective[4] value of β.[5]

The plots in Figures 4.25 to 4.27 all show a rapid decline between $\beta = 4\pi$Hz ($\frac{40}{3} \in (4\pi, 5\pi)$) and 80Hz. In some cases, this decline commences earlier, and it is almost monotonic.

In Table 4.4, we collate the values of the plots in Figures 4.25 to 4.27 at the especially relevant values of β, $\frac{40}{3}$Hz and 80Hz. By this measure, too, the imposition of the wrapped endpoint continuity condition does not compensate for the loss of a free parameter, but imposing both such conditions more than compensates for the loss of two free parameters.

When we compare this Table with Table 4.2, we find[6], that $\|\psi - \psi'\|$ is much smaller than $\|\psi_{(1)} - \psi_{(1)\beta}\|$ when $\beta = 80$Hz, but, when $\beta = \frac{40}{3}$Hz, $\|\psi - \psi'\|$ is smaller than $\|\psi_{(1)} - \psi_{(1)\beta}\|$

[3]Defined here as follows — if f is a function, its upper envelope is $f_U(x) = \max_{x' \in \left[\pi \left\lfloor \frac{x}{\pi} \right\rfloor, \pi\left(\left\lfloor \frac{x}{\pi} \right\rfloor + 1\right)\right)} f(x')$. We use Matlab®'s `fminsearch` function to find the minimum (the code internal to this uses ordinary double arithmetic, but we use variable precision arithmetic in the functions it calls)

[4]We model the response of the accelerometer by a sharp cut-off. A more gentle decline in accelerometer response might mean the response is better modelled by a cut-off not at the specification value, but at a nearby point

[5]The bandwidth might be greater than stated, which could lead to $\|\psi - \psi'\|$ increasing rather than decreasing

[6]There are fewer coincident values here than in Table 4.2. Referring to the explanation of coincident values in footnote 1, a value in the Table will be unchanged if the both the maximum and minimum affected eigenvalues remain bracketted by the maximum and minimum eigenvalues in the unaffected group. The "near misses" in Table 4.4 may be genuinely different values, or may be due to arithmetic errors in calculating eigenvalues. At this juncture, this is a moot point

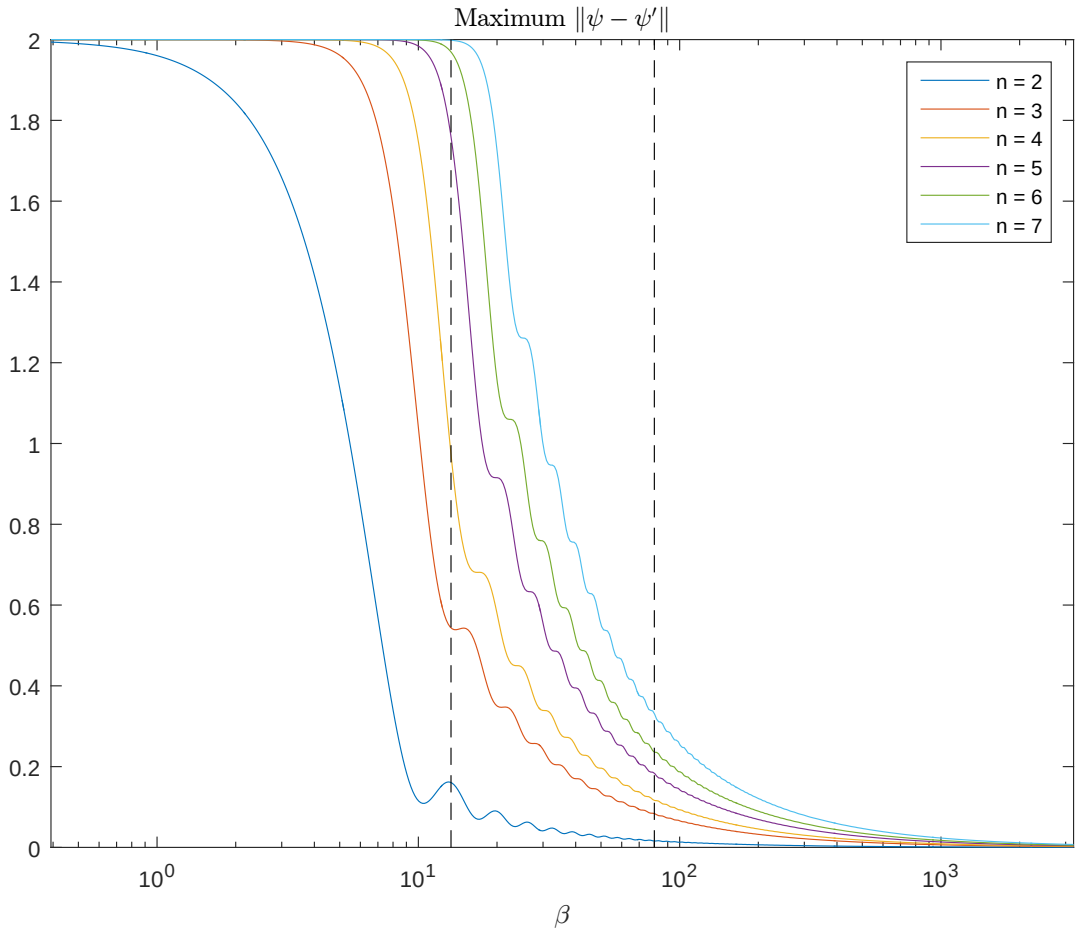


Figure 4.22: *Plots of the worst-case $\|\psi - \psi'\|$ corresponding to the eigenvalues of Figure 4.12 (i.e., for general wavelets of our form)*

only when the number of free parameters is 1 or 2 for the case without endpoint conditions, and only when $p = 1$ when there are such conditions. When p is larger, then $\|\psi - \psi'\|$ is much bigger than $\|\psi_{(1)} - \psi_{(1)\beta}\|$.

Returning to Figures 4.25 to 4.27 themselves, and comparing them to Figures 4.22 to 4.24, we see that the differences are small, and increase with bandwidth, but are still significant near the bandwidth of interest, represented by the leftmost dashed line.

In Figures 4.28 to 4.30, we show the wavelets ψ and ψ' which maximise $\|\psi - \psi'\|$ subject to $\|\psi' - \psi_\beta\| \leq \|\psi - \psi_\beta\|$ for the worst case ψ, together with ψ_β (the working which enables these plots to be drawn is in Appendix 7.7, which also contains Tables referring to values of $\|\psi - \psi_\beta\|$, etc., for special values of β).

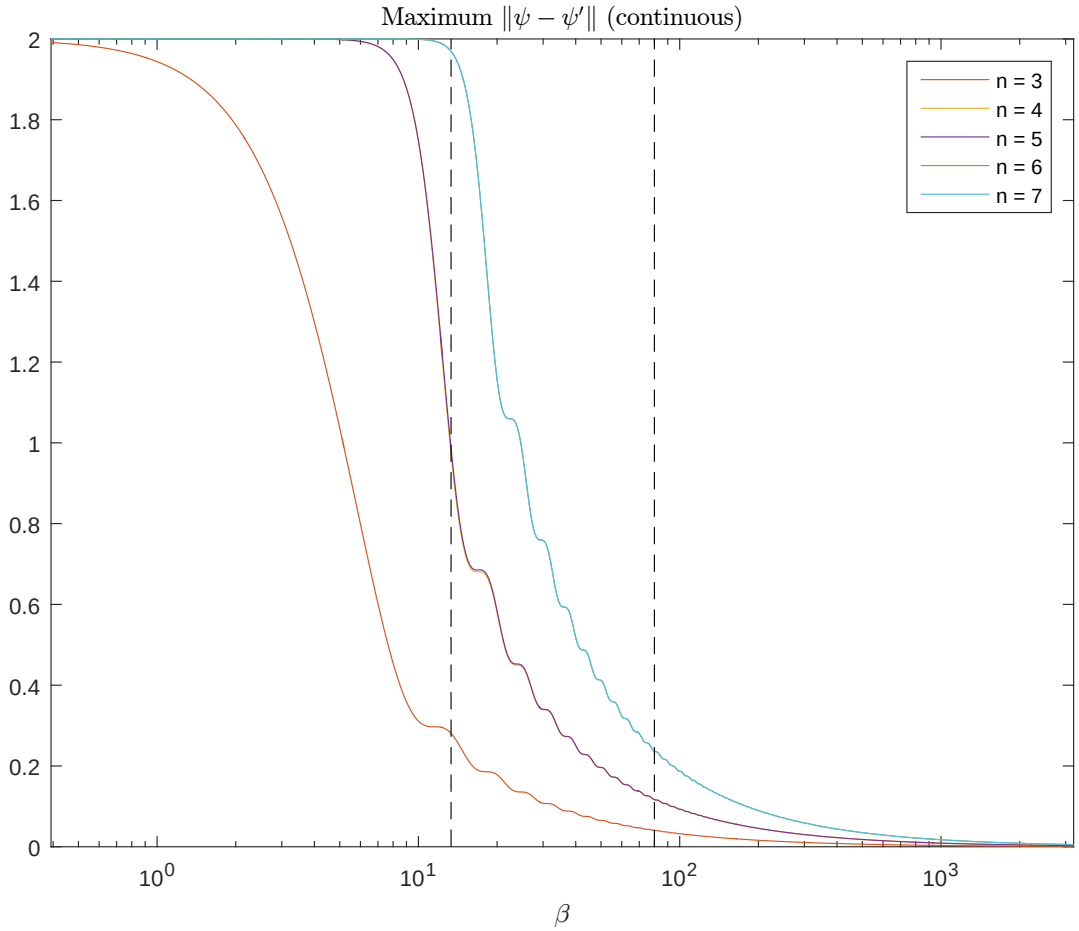


Figure 4.23: *Plots of the worst-case $\|\psi - \psi'\|$ corresponding to the eigenvalues of Figure 4.13 (i.e., for wavelets of our form subject to wec). Note that the $n = 4$ curve is coincident with, and obscured by, the $n = 5$ curve, and the same is true of the $n = 6$ and $n = 7$ curves*

	Wavelets			**Wavelets with wec**			**Wavelets with wed**		
	n	$\beta = \frac{40}{3}$Hz	$\beta = 80$Hz	n	$\beta = \frac{40}{3}$Hz	$\beta = 80$Hz	n	$\beta = \frac{40}{3}$Hz	$\beta = 80$Hz
p		$\|\psi - \psi'\|$			$\|\psi - \psi'\|$			$\|\psi - \psi'\|$	
1	2	0.1621	0.0162	3	0.2933	0.0412	4	0.1365	0.0191
2	3	0.5430	0.0846	4	1.1497	0.1195	5	0.3838	0.0192
3	4	1.1497	0.1195	5	1.1631	0.1196	6	1.0488	0.0477
4	5	1.8542	0.1864	6	1.9846	0.2413	7	1.8743	0.0477
5	6	1.9847	0.2413	7	1.9847	0.2413	—	—	—
6	7	1.9993	0.3387	—	—	—	—	—	—

Table 4.4: *Values of the upper envelopes of the worst-case $\|\psi - \psi'\|$ in Figures 4.25 to 4.27 at our special values of β*

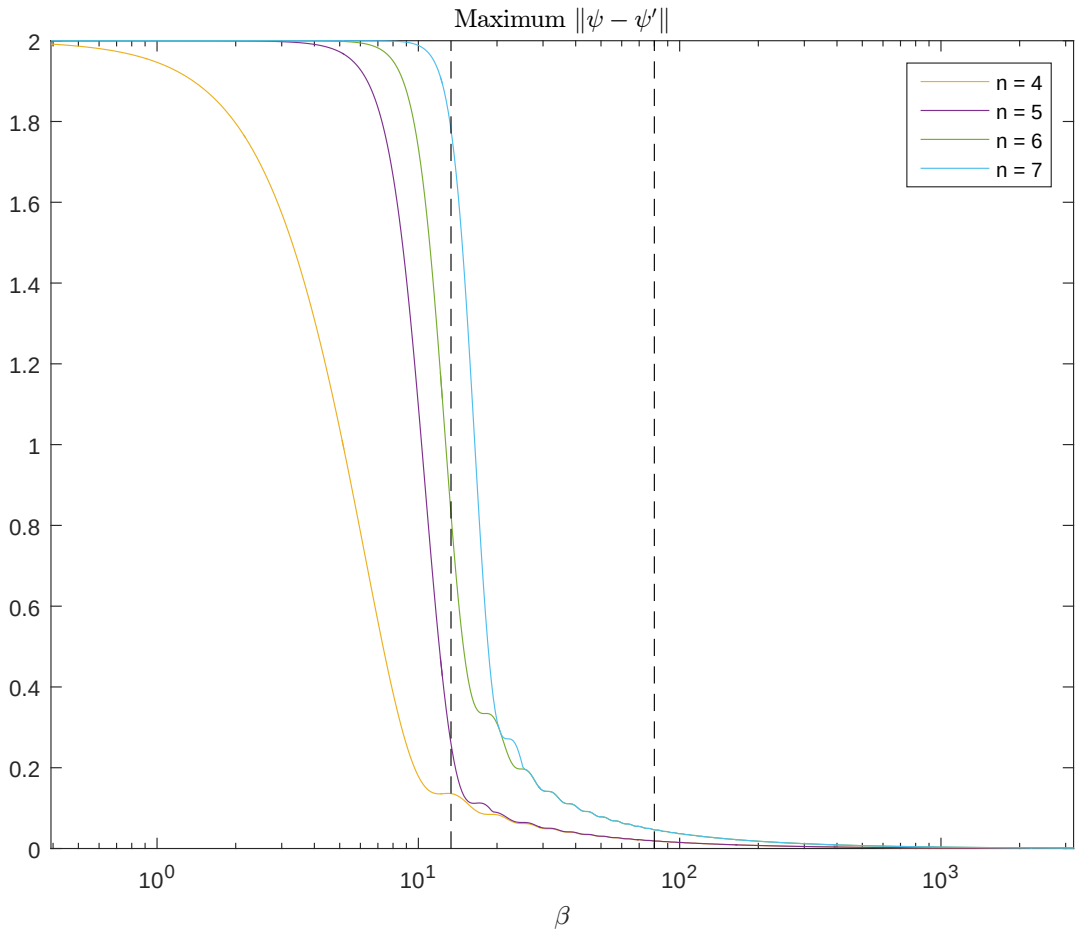


Figure 4.24: *Plots of the worst-case $\|\psi - \psi'\|$ corresponding to the eigenvalues of Figure 4.14 (i.e., for wavelets of our form subject to wed)*

	Wavelets			**Wavelets with wec**			**Wavelets with wed**		
p	n	A	B	n	A	B	n	A	B
1	2	1.3860	-1.0109	3	3.2743	-1.0037	4	1.5084	-1.0020
2	3	6.7484	-1.0076	4	9.6797	-1.0107	5	1.5151	-1.0026
3	4	9.6794	-1.0107	5	9.6865	-1.0108	6	3.7736	-1.0074
4	5	15.1124	-1.0160	6	20.1106	-1.0204	7	3.7735	-1.0074
5	6	20.1106	-1.0204	7	20.1106	-1.0204	—	—	—
6	7	28.1238	-1.0266	—	—	—	—	—	—

Table 4.5: *Parameters for the approximate relation $\|\psi - \psi'\| = A\beta^B$ which holds when $\beta > 80Hz$ (analogue of Table 4.3)*

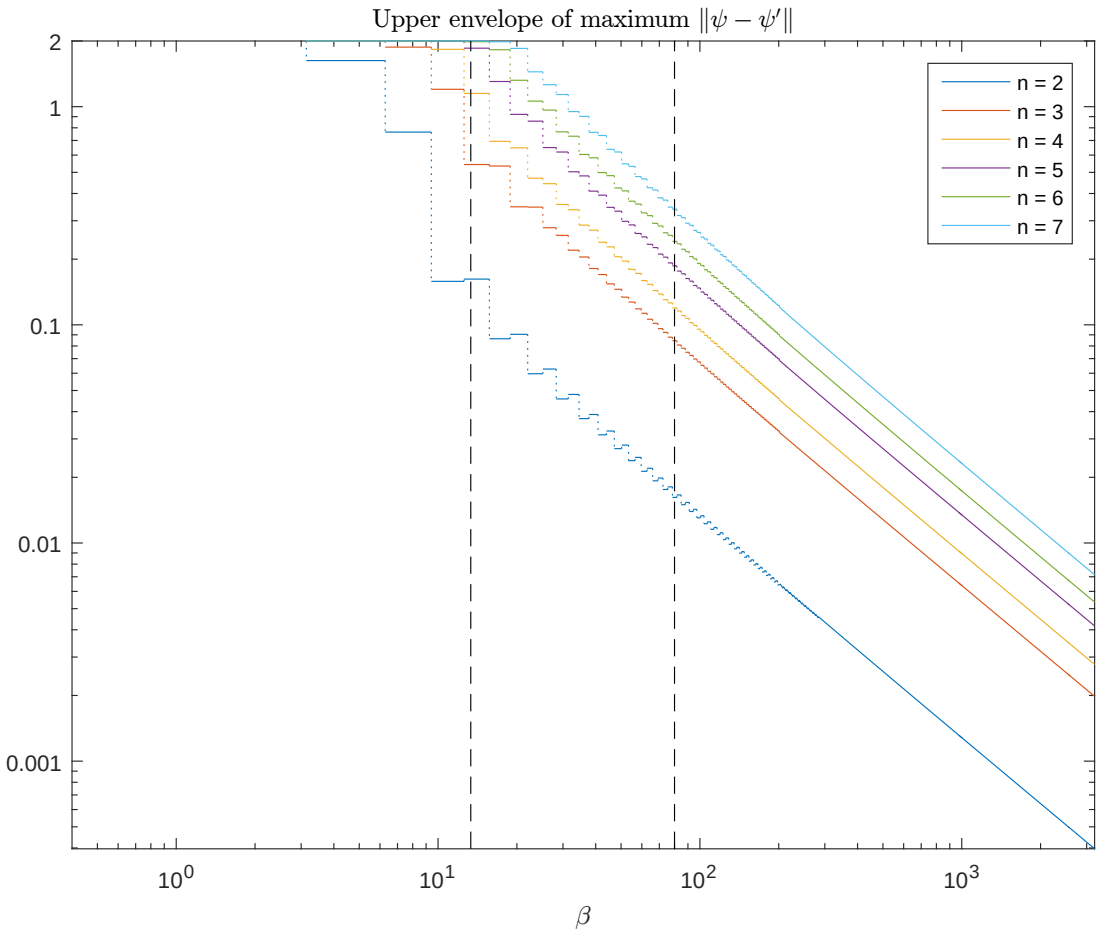


Figure 4.25: *Plots of the upper envelope of the distances of Figure 4.22*

	$\|\psi - \psi_\beta\|$		
β	—	**wec**	**wed**
5Hz	$\sim 1 - 9.98 \times 10^{-11}$	$\sim 1 - 9.66 \times 10^{-7}$	$\sim 1 - 9.66 \times 10^{-7}$
20Hz	0.9625	0.8853	0.8853
80Hz	0.5333	0.4757	0.4757
320Hz	0.2676	0.2379	0.2379

Table 4.6: *Maximum distance between ψ and ψ_β*

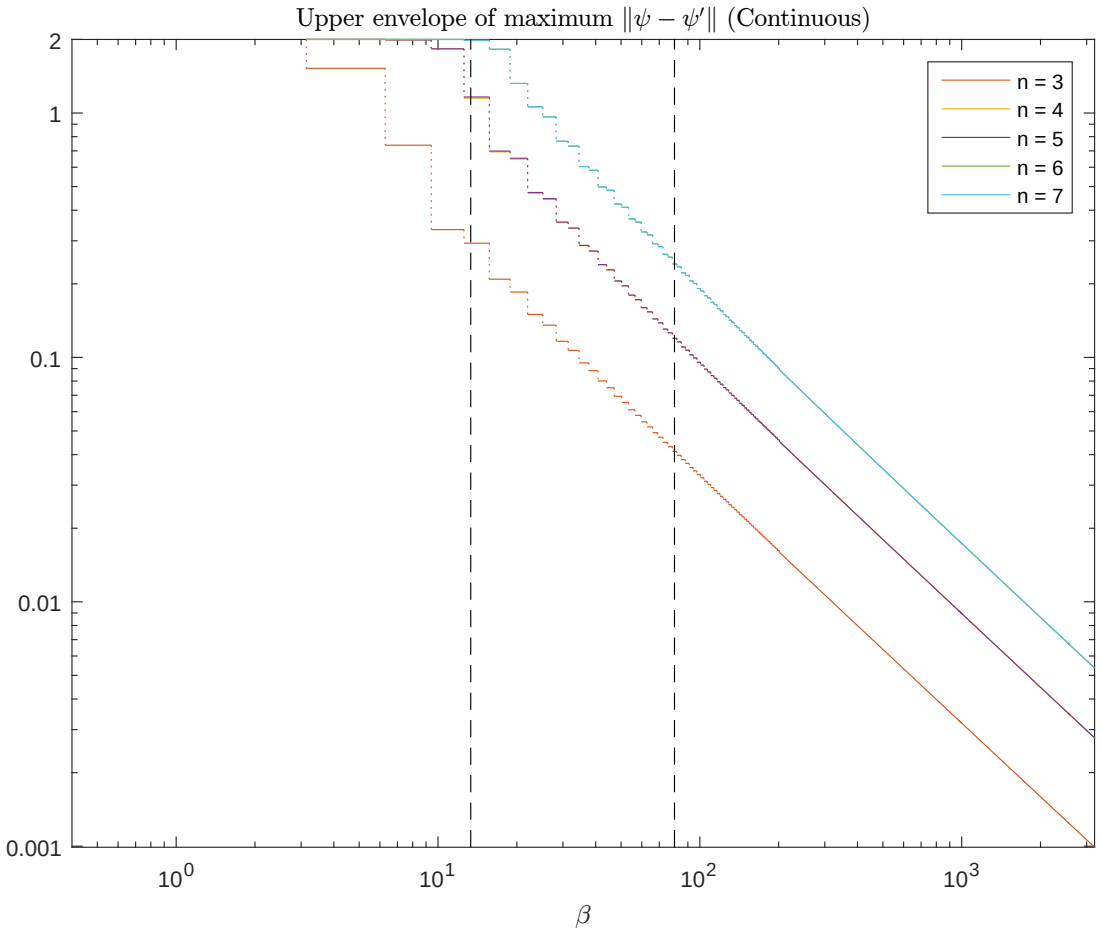


Figure 4.26: *Plots of the upper envelope of the distances of Figure 4.23. Note that the $n = 4$ plot is coincident with, and obscured by, the $n = 5$ plot, and the same is true of the $n = 6$ and $n = 7$ plots*

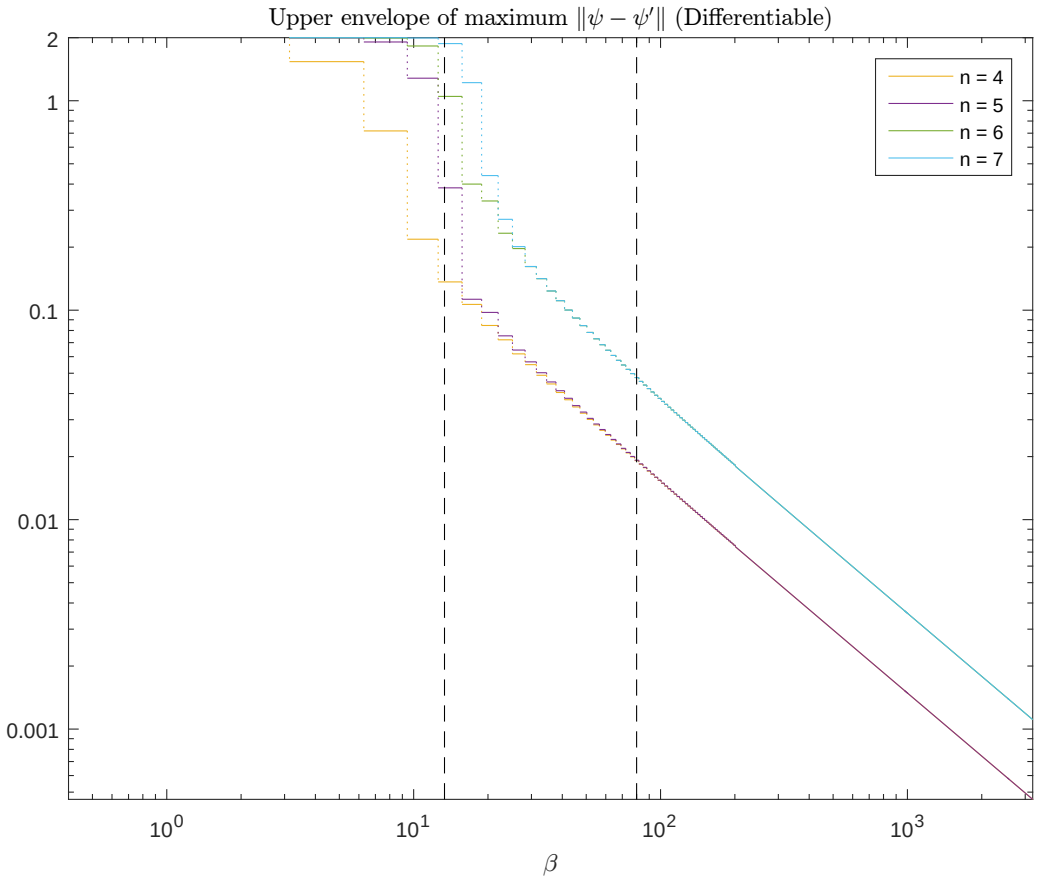


Figure 4.27: *Plots of the upper envelope of the distances of Figure 4.24*

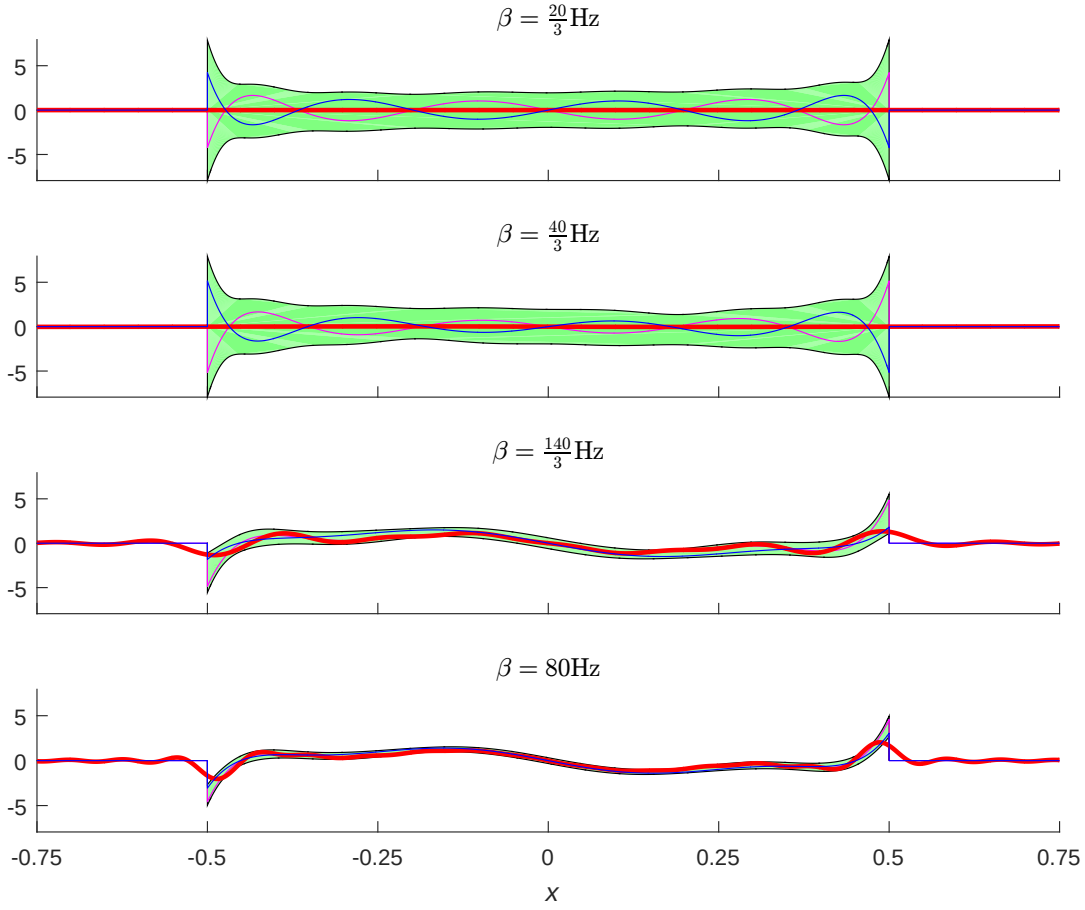


Figure 4.28: *Plots of the wavelets $\psi(x)$ (blue) and $\psi'(x)$ (magenta) which maximise $\|\psi - \psi'\|$ subject to $\|\psi' - \psi_\beta\| \leq \|\psi - \psi_\beta\|$, and of $\psi_\beta(x)$ (thick, red). The green area is the envelope of all ψ'' such that $\|\psi'' - \psi_\beta\| \leq \|\psi - \psi_\beta\|$*

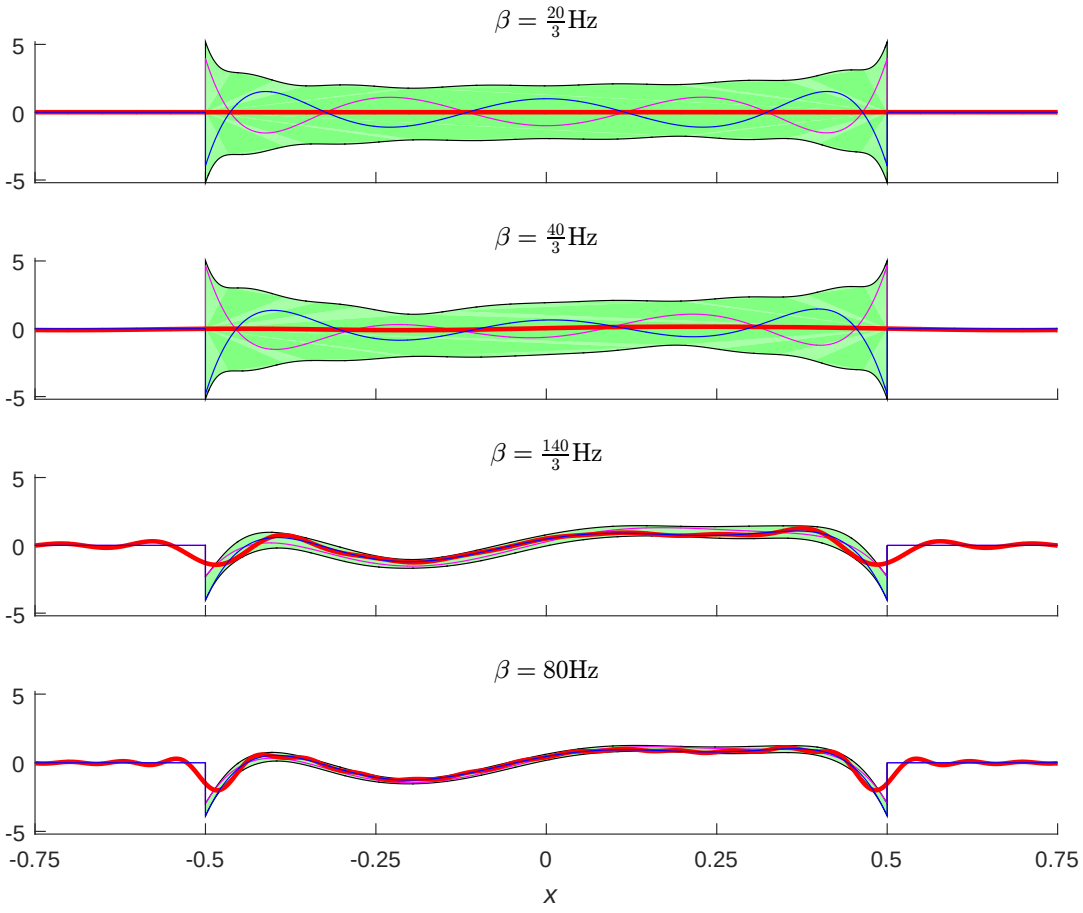


Figure 4.29: *Similar to Figure 4.28, but with all wavelets satisfying wec*

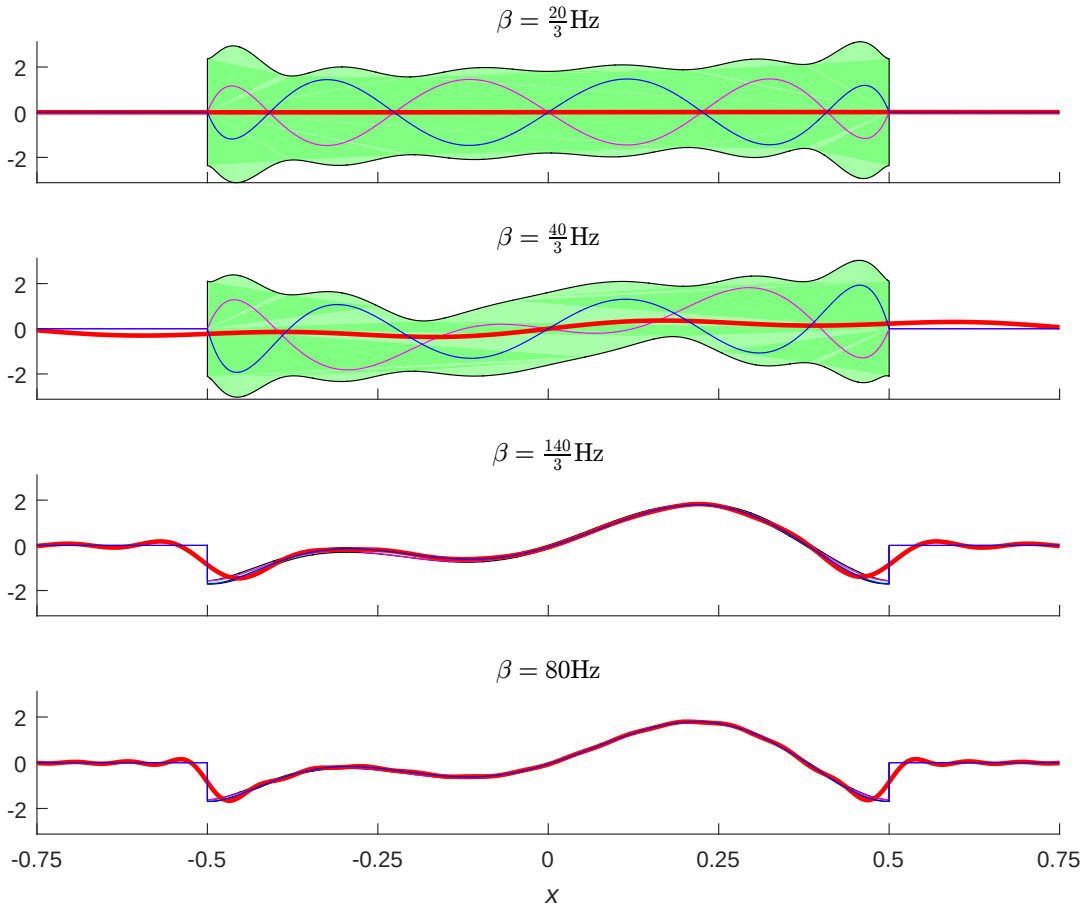


Figure 4.30: *Similar to Figure 4.28, but with all wavelets satisfying wed*

4.8 Analogous calculation for trigonometric polynomial wavelets

Let $\psi(x) = \begin{cases} \sum_{k=1}^{n}[a_k \sin 2k\pi x + c_k \cos 2k\pi x], & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise,} \end{cases}$ with the ue condition being $\int_{-\frac{1}{2}}^{\frac{1}{2}} \psi(x)^2 \mathrm{d}x = \frac{1}{2}\sum_{k=1}^{n}[a_k^2 + c_k^2] = 1$ or $\left[\begin{array}{cc} a^{\mathrm{T}} & c^{\mathrm{T}} \end{array}\right] H \left[\begin{array}{c} a \\ c \end{array}\right] = 1$, where $a = [a_1, \ldots, a_n]^{\mathrm{T}}, c = [c_1, \ldots, c_n]^{\mathrm{T}} \in \mathbb{R}^n$ and $H = \frac{1}{2}\mathrm{I} \in \mathbb{R}^{2n \times 2n}$.

The zi condition is automatically satisfied, as are the optional wrapped endpoint continuity or differentiability conditions.

Define the bandwidth limited version of ψ, $\psi_\beta(x) = [\hat{\psi}_\beta(\xi)]\check{}\,(x)$, where $\hat{\psi}_\beta(\xi) = \begin{cases} \hat{\psi}(\xi), & \xi \in [-\beta, \beta]; \\ 0, & \text{otherwise.} \end{cases}$

As before, $\|\psi - \psi_\beta\|^2 = \|\hat{\psi} - \hat{\psi}_\beta\|^2 = \|\hat{\psi}\| - 2\Re\langle\hat{\psi}, \hat{\psi}_\beta\rangle + \|\hat{\psi}_\beta\|^2 = 1 - \|\hat{\psi}_\beta\|^2$, and

$$\begin{aligned} \|\hat{\psi}_\beta\|^2 &= \sum_{k=1}^{n}\sum_{\ell=1}^{n}[a_k a_\ell \langle \hat{f}_{k,\beta}, \hat{f}_{\ell,\beta}\rangle + a_k c_\ell \langle \hat{f}_{k,\beta}, \hat{g}_{\ell,\beta}\rangle + c_k a_\ell \langle \hat{g}_{k,\beta}, \hat{f}_{\ell,\beta}\rangle + c_k c_\ell \langle \hat{g}_{k,\beta}, \hat{g}_{\ell,\beta}\rangle] \\ &= \sum_{k=1}^{n}[a_k^2 \|\hat{f}_{k,\beta}\|^2 + 2a_k c_k \Re\langle \hat{f}_{k,\beta}, \hat{g}_{k,\beta}\rangle + c_k^2 \|\hat{g}_{k,\beta}\|^2] + \\ &\qquad 2\Re\sum_{k=1}^{n-1}\sum_{\ell=k+1}^{n}[a_k a_\ell \langle \hat{f}_{k,\beta}, \hat{f}_{\ell,\beta}\rangle + a_k c_\ell \langle \hat{f}_{k,\beta}, \hat{g}_{\ell,\beta}\rangle + c_k c_\ell \langle \hat{g}_{k,\beta}, \hat{g}_{\ell,\beta}\rangle], \end{aligned} \tag{4.79}$$

where

$$f_k(x) = \begin{cases} \sin 2k\pi x, & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise,} \end{cases} \qquad g_k(x) = \begin{cases} \cos 2k\pi x, & x \in \left[-\frac{1}{2}, \frac{1}{2}\right); \\ 0, & \text{otherwise,} \end{cases} \tag{4.80}$$

and

$$\hat{f}_{k,\beta}(\xi) = \begin{cases} \hat{f}_k(\xi), & \xi \in [-\beta, \beta]; \\ 0, & \text{otherwise,} \end{cases} \qquad \hat{g}_{k,\beta}(\xi) = \begin{cases} \hat{g}_k(\xi), & \xi \in [-\beta, \beta]; \\ 0, & \text{otherwise.} \end{cases} \tag{4.81}$$

But f_k is real, so $\hat{f}_k(-\xi) = \overline{\hat{f}_k(\xi)}$, and f_k is odd, so $\hat{f}_k$ is purely imaginary. Together, these facts imply that $\hat{f}_{k,\beta}$ is i times a real odd function. Similarly, g_k real and even $\Rightarrow \hat{g}_k$ is real and $\hat{g}_k(-\xi) = \overline{\hat{g}_k(\xi)}$, so $\hat{g}_{k,\beta}$ is a real even function, for all k. Hence, $\hat{f}_{k,\beta}\hat{g}_{\ell,\beta}$ is i times a real odd function and $\langle \hat{f}_{k,\beta}, \hat{g}_{\ell,\beta}\rangle = 0$, for all k and ℓ, is an easy consequence of this. The same argument leads to the conclusion that $\langle \hat{f}_{k,\beta}, \hat{f}_{\ell,\beta}\rangle$ and $\langle \hat{g}_{k,\beta}, \hat{g}_{\ell,\beta}\rangle$ are real.

Thus, equation (4.79) becomes

$$\begin{aligned} \|\hat{\psi}_\beta\|^2 &= \sum_{k=1}^{n}[a_k^2 \|\hat{f}_{k,\beta}\|^2 + c_k^2 \|\hat{g}_{k,\beta}\|^2] + 2\sum_{k=1}^{n-1}\sum_{\ell=k+1}^{n}[a_k a_\ell \langle \hat{f}_{k,\beta}, \hat{f}_{\ell,\beta}\rangle + c_k c_\ell \langle \hat{g}_{k,\beta}, \hat{g}_{\ell,\beta}\rangle] \\ &= \left[\begin{array}{cc} a^{\mathrm{T}} & c^{\mathrm{T}} \end{array}\right] \left[\begin{array}{cc} P_f(\beta) & 0 \\ 0 & P_g(\beta) \end{array}\right] \left[\begin{array}{c} a \\ c \end{array}\right] = \left[\begin{array}{cc} a^{\mathrm{T}} & c^{\mathrm{T}} \end{array}\right] P \left[\begin{array}{c} a \\ c \end{array}\right], \end{aligned} \tag{4.82}$$

where

$$
\begin{aligned}
P_f(\beta) &= \begin{bmatrix} \|\hat{f}_{1,\beta}\|^2 & \langle \hat{f}_{1,\beta}, \hat{f}_{2,\beta}\rangle & \cdots & \langle \hat{f}_{1,\beta}, \hat{f}_{n,\beta}\rangle \\ \langle \hat{f}_{1,\beta}, \hat{f}_{2,\beta}\rangle & \|\hat{f}_{2,\beta}\|^2 & \cdots & \langle \hat{f}_{2,\beta}, \hat{f}_{n,\beta}\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \hat{f}_{1,\beta}, \hat{f}_{n,\beta}\rangle & \langle \hat{f}_{2,\beta}, \hat{f}_{n,\beta}\rangle & \cdots & \|\hat{f}_{n,\beta}\|^2 \end{bmatrix}, \\
P_g(\beta) &= \begin{bmatrix} \|\hat{g}_{1,\beta}\|^2 & \langle \hat{g}_{1,\beta}, \hat{g}_{2,\beta}\rangle & \cdots & \langle \hat{g}_{1,\beta}, \hat{g}_{n,\beta}\rangle \\ \langle \hat{g}_{1,\beta}, \hat{g}_{2,\beta}\rangle & \|\hat{g}_{2,\beta}\|^2 & \cdots & \langle \hat{g}_{2,\beta}, \hat{g}_{n,\beta}\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \hat{g}_{1,\beta}, \hat{g}_{n,\beta}\rangle & \langle \hat{g}_{2,\beta}, \hat{g}_{n,\beta}\rangle & \cdots & \|\hat{g}_{n,\beta}\|^2 \end{bmatrix}, \\
P &= \begin{bmatrix} P_f & 0 \\ 0 & P_g \end{bmatrix}. \qquad (4.83)
\end{aligned}
$$

Now

$$
\begin{aligned}
\hat{f}_k(\xi) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f_k(x) \mathrm{e}^{-i\xi x} \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathrm{e}^{-i\xi x} \sin 2k\pi x \mathrm{d}x \\
&= \frac{1}{2i\sqrt{2\pi}} \int_{-\frac{1}{2}}^{\frac{1}{2}} [\mathrm{e}^{-i(\xi-2k\pi)x} - \mathrm{e}^{-i(\xi+2k\pi)x}] \mathrm{d}x = \tfrac{1}{2i\sqrt{2\pi}} \left[\tfrac{\mathrm{e}^{-i(\xi-2k\pi)x}}{-i(\xi-2k\pi)} - \tfrac{\mathrm{e}^{-i(\xi+2k\pi)x}}{-i(\xi+2k\pi)} \right]_{-\frac{1}{2}}^{\frac{1}{2}} \\
&= \tfrac{(-1)^{k+1} i}{\sqrt{2\pi}} \left[\tfrac{1}{\xi-2k\pi} - \tfrac{1}{\xi+2k\pi} \right] \sin \tfrac{1}{2}\xi, \qquad (4.84)
\end{aligned}
$$

and similarly,

$$
\hat{g}_k(\xi) = \tfrac{(-1)^k}{\sqrt{2\pi}} \left[\tfrac{1}{\xi-2k\pi} + \tfrac{1}{\xi+2k\pi} \right] \sin \tfrac{1}{2}\xi. \qquad (4.85)
$$

Hence,

$$
\begin{aligned}
\|\hat{f}_{k,\beta}\|^2 &= \frac{1}{2\pi} \int_{-\beta}^{\beta} \left[\frac{1}{\xi - 2k\pi} - \frac{1}{\xi + 2k\pi} \right]^2 \sin^2 \tfrac{1}{2}\xi \mathrm{d}\xi \\
&= \frac{1}{4\pi} \int_{-\beta}^{\beta} \left[\frac{1}{(\xi - 2k\pi)^2} + \frac{1}{(\xi + 2k\pi)^2} - \frac{1}{2k\pi} \left(\frac{1}{\xi - 2k\pi} - \frac{1}{\xi + 2k\pi} \right) \right] (1 - \cos\xi) \mathrm{d}\xi \\
&= \frac{1}{2\pi} \int_{-\beta}^{\beta} \left[\frac{1}{(\xi - 2k\pi)^2} - \frac{1}{2k\pi} \frac{1}{\xi - 2k\pi} \right] (1 - \cos\xi) \mathrm{d}\xi \\
&= \tfrac{1}{2\pi} \left[S'_k(\beta) - \tfrac{1}{2k\pi} C'_k(\beta) \right], k \geq 1, \qquad (4.86)
\end{aligned}
$$

where

$$\begin{aligned} S'_k(\beta) &= \int_{-\beta}^{\beta} \frac{1-\cos\xi}{(\xi-2k\pi)^2}\mathrm{d}\xi = \int_{-\beta-2k\pi}^{\beta-2k\pi} \frac{1-\cos\xi}{\xi^2}\mathrm{d}\xi \\ &= \left[-\frac{1-\cos\xi}{\xi} + \int \frac{\sin\xi}{\xi}\mathrm{d}\xi\right]_{-\beta-2k\pi}^{\beta-2k\pi} \\ &= \mathrm{Si}(\beta-2k\pi) - \frac{1-\cos\beta}{\beta-2k\pi} + \mathrm{Si}(\beta+2k\pi) - \frac{1+\cos\beta}{\beta+2k\pi} \\ &= S_k(\beta) + S_{-k}(\beta) \end{aligned} \tag{4.87}$$

and

$$\begin{aligned} C'_k(\beta) &= \int_{-\beta}^{\beta} \frac{1-\cos\xi}{\xi-2k\pi}\mathrm{d}\xi = \int_{-\beta-2k\pi}^{\beta-2k\pi} \frac{1-\cos\xi}{\xi}\mathrm{d}\xi \\ &= \begin{cases} \int_{-\beta+2k\pi}^{\beta+2k\pi} \frac{\cos\xi-1}{\xi}\mathrm{d}\xi, & \beta \in (0, 2k\pi]; \\ \left(-\int_{-\beta-2k\pi}^{0} - \int_{0}^{\beta-2k\pi}\right) \frac{\cos\xi-1}{\xi}\mathrm{d}\xi, & \beta \in (2k\pi, \infty) \end{cases} \\ &= \left(\int_0^{\beta+2k\pi} - \int_0^{|\beta-2k\pi|}\right) \frac{\cos\xi-1}{\xi}\mathrm{d}\xi \\ &= \mathrm{Ci}(\beta+2k\pi) - \gamma - \ln(\beta+2k\pi) - \mathrm{Ci}(|\beta-2k\pi|) + \gamma + \ln|\beta-2k\pi| \\ &= C_k(\beta) - C_{-k}(\beta) \end{aligned} \tag{4.88}$$

for the sine and cosine integrals Si and Ci, where $\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \ln n\right)$ is the Euler-Mascheroni constant, for

$$\begin{aligned} S_k(\beta) &= \mathrm{Si}(\beta+2k\pi) - \frac{1+\cos\beta}{\beta+2k\pi}, \\ C_k(\beta) &= \begin{cases} \gamma, & \beta = -2k\pi; \\ \mathrm{Ci}(|\beta+2k\pi|) - \ln|\beta+2k\pi|, & \text{otherwise.} \end{cases} \end{aligned} \tag{4.89}$$

We consequently have

$$\|\hat{f}_{k,\beta}\|^2 = \tfrac{1}{2\pi}\left[S_k(\beta) + S_{-k}(\beta) - \tfrac{1}{2k\pi}\{C_k(\beta) - C_{-k}(\beta)\}\right], k \geq 1. \tag{4.90}$$

Similar to the result for $\|\hat{f}_{k,\beta}\|^2$, we have

$$\|\hat{g}_{k,\beta}\|^2 = \tfrac{1}{2\pi}\left[S_k(\beta) + S_{-k}(\beta) + \tfrac{1}{2k\pi}\{C_k(\beta) - C_{-k}(\beta)\}\right], k \geq 1. \tag{4.91}$$

If $\ell \neq k$,

$$\begin{aligned}
\langle \hat{f}_{k,\beta}, \hat{f}_{\ell,\beta} \rangle &= \frac{1}{2\pi} \int_{-\beta}^{\beta} \left[\frac{1}{\xi - 2k\pi} - \frac{1}{\xi + 2k\pi} \right] \left[\frac{1}{\xi - 2\ell\pi} - \frac{1}{\xi + 2\ell\pi} \right] \sin^2 \tfrac{1}{2}\xi \mathrm{d}\xi \\
&= \frac{1}{\pi} \int_{-\beta}^{\beta} \frac{1}{\xi - 2k\pi} \left[\frac{1}{\xi - 2\ell\pi} - \frac{1}{\xi + 2\ell\pi} \right] \sin^2 \tfrac{1}{2}\xi \mathrm{d}\xi \\
&= \frac{1}{4\pi^2} \int_{-\beta}^{\beta} \left[\frac{1}{k - \ell} \left(\frac{1}{\xi - 2k\pi} - \frac{1}{\xi - 2\ell\pi} \right) - \frac{1}{k + \ell} \left(\frac{1}{\xi - 2k\pi} - \frac{1}{\xi + 2\ell\pi} \right) \right] \times \\
&\qquad (1 - \cos\xi) \mathrm{d}\xi \\
&= \frac{1}{2(k^2 - \ell^2)\pi^2} \int_{-\beta}^{\beta} \left[\frac{\ell}{\xi - 2k\pi} - \frac{k}{\xi - 2\ell\pi} \right] (1 - \cos\xi) \mathrm{d}\xi \\
&= \tfrac{1}{2(k^2 - \ell^2)\pi^2} [\ell \{ C_k(\beta) - C_{-k}(\beta) \} - k \{ C_\ell(\beta) - C_{-\ell}(\beta) \}], \qquad (4.92)
\end{aligned}$$

and

$$\begin{aligned}
\langle \hat{g}_{k,\beta}, \hat{f}_{\ell,\beta} \rangle &= \frac{1}{4\pi^2} \int_{-\beta}^{\beta} \left[\frac{1}{k - \ell} \left(\frac{1}{\xi - 2k\pi} - \frac{1}{\xi - 2\ell\pi} \right) + \frac{1}{k + \ell} \left(\frac{1}{\xi - 2k\pi} - \frac{1}{\xi + 2\ell\pi} \right) \right] \times \\
&\qquad (1 - \cos\xi) \mathrm{d}\xi \\
&= \frac{1}{2(k^2 - \ell^2)\pi^2} \int_{-\beta}^{\beta} \left[\frac{k}{\xi - 2k\pi} - \frac{\ell}{\xi - 2\ell\pi} \right] (1 - \cos\xi) \mathrm{d}\xi \\
&= \tfrac{1}{2(k^2 - \ell^2)\pi^2} [k \{ C_k(\beta) - C_{-k}(\beta) \} - \ell \{ C_\ell(\beta) - C_{-\ell}(\beta) \}]. \qquad (4.93)
\end{aligned}$$

To find the wavelets ψ, ψ' which maximise $\|\psi - \psi'\|$ subject to $\|\psi' - \psi_\beta\| \leq \|\psi - \psi_\beta\|$, we follow the argument of Section 4.7, with the considerable simplifications that we need not make any projections (as the zi and wrapped endpoint conditions are satisfied), $X_0^{\mathrm{T}} X_0 = n$ and $H = \frac{1}{2}\mathrm{I}$.

We then find that, given $[a^{\mathrm{T}}, c^{\mathrm{T}}]^{\mathrm{T}}$, $[a'^{\mathrm{T}}, c'^{\mathrm{T}}]^{\mathrm{T}}$, with the obvious meaning, is given by

$$\tilde{a}' = 2 \frac{\tilde{a}^{\mathrm{T}} P(\beta) \tilde{a}}{\tilde{a}^{\mathrm{T}} P(\beta)^2 \tilde{a}} P(\beta) \tilde{a} - \tilde{a}, \qquad (4.94)$$

and then

$$\tilde{a}^{\mathrm{T}} H \tilde{a}' = \frac{(\tilde{a}^{\mathrm{T}} P(\beta) \tilde{a})^2}{\tilde{a}^{\mathrm{T}} P(\beta)^2 \tilde{a}} - 1, \qquad (4.95)$$

where we are now using the tilde to denote quantities like $\tilde{a} = [a^{\mathrm{T}}, c^{\mathrm{T}}]^{\mathrm{T}}$, $\tilde{a}' = [a'^{\mathrm{T}}, c'^{\mathrm{T}}]^{\mathrm{T}}$.

Continuing with the argument, we find that the minimum value of $\tilde{a}^{\mathrm{T}} H \tilde{a}'$, as given by equation (4.95), is $\frac{8\lambda_1\lambda_{2n}}{(\lambda_1 + \lambda_{2n})^2} - 1$, where λ_1 is the minimum eigenvalue of $H^{-1}P(\beta) = 2P(\beta)$ and λ_{2n} is the maximum such eigenvalue. Then, the maximum value of $\|\psi - \psi'\|$ subject to

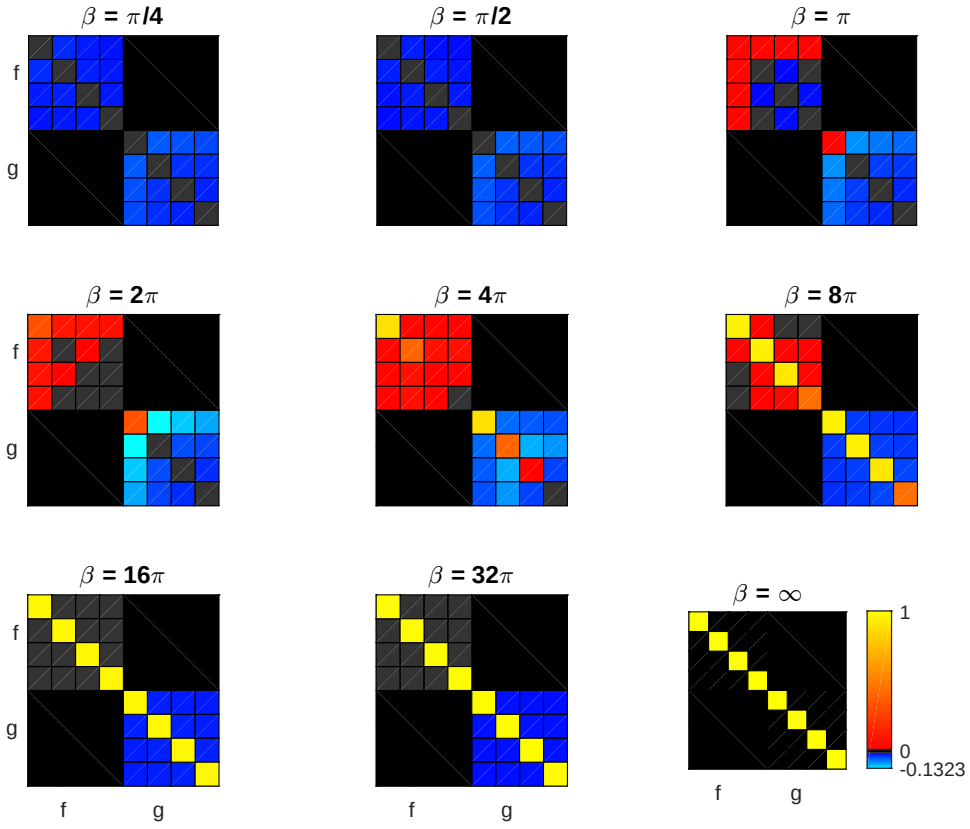


Figure 4.31: *Pseudocolour plot of the elements of* $2P(\beta)(= H^{-\frac{1}{2}}P(\beta)H^{-\frac{1}{2}})$ *for* $n = 4$. *The colour scale given for the* $\beta = \infty$ *plot applies to all plots (exactly zero entries are shaded black and near-zero entries are shaded dark grey). (This figure is analogous to Figure 4.4)*

our conditions is $2\frac{\lambda_{2n}-\lambda_1}{\lambda_1+\lambda_{2n}}$, and (non-unique) values for $\tilde{a}$ and $\tilde{a}'$ are

$$\begin{aligned} \tilde{a} &= \sqrt{\tfrac{\lambda_{2n}}{\lambda_1+\lambda_{2n}}}u_1 + \sqrt{\tfrac{\lambda_1}{\lambda_1+\lambda_{2n}}}u_{2n}, \\ \tilde{a}' &= \tfrac{3\lambda_1-\lambda_{2n}}{\lambda_1+\lambda_{2n}}\sqrt{\tfrac{\lambda_{2n}}{\lambda_1+\lambda_{2n}}}u_1 + \tfrac{3\lambda_{2n}-\lambda_1}{\lambda_1+\lambda_{2n}}\sqrt{\tfrac{\lambda_1}{\lambda_1+\lambda_{2n}}}u_{2n}, \end{aligned} \tag{4.96}$$

where u_1 and u_{2n} are unit eigenvectors of $2P(\beta)$ corresponding to the eigenvalues λ_1 and λ_{2n} respectively.

Again, we may find the envelope of ψ'' such that $\|\psi'' - \psi_\beta\| \leq \|\psi - \psi_\beta\|$ by defining $X_0 = [\sin 2\pi x_0, \sin 4\pi x_0, \ldots, \sin 2n\pi x_0, \cos 2\pi x_0, \cos 4\pi x_0, \ldots, \cos 2n\pi x_0]^{\mathrm{T}}$, and then substituting the resulting equality $X_0^{\mathrm{T}}X_0 = n$ and $\tilde{H}_1 = \frac{1}{2}\mathrm{I}$ into equation (8), along with the

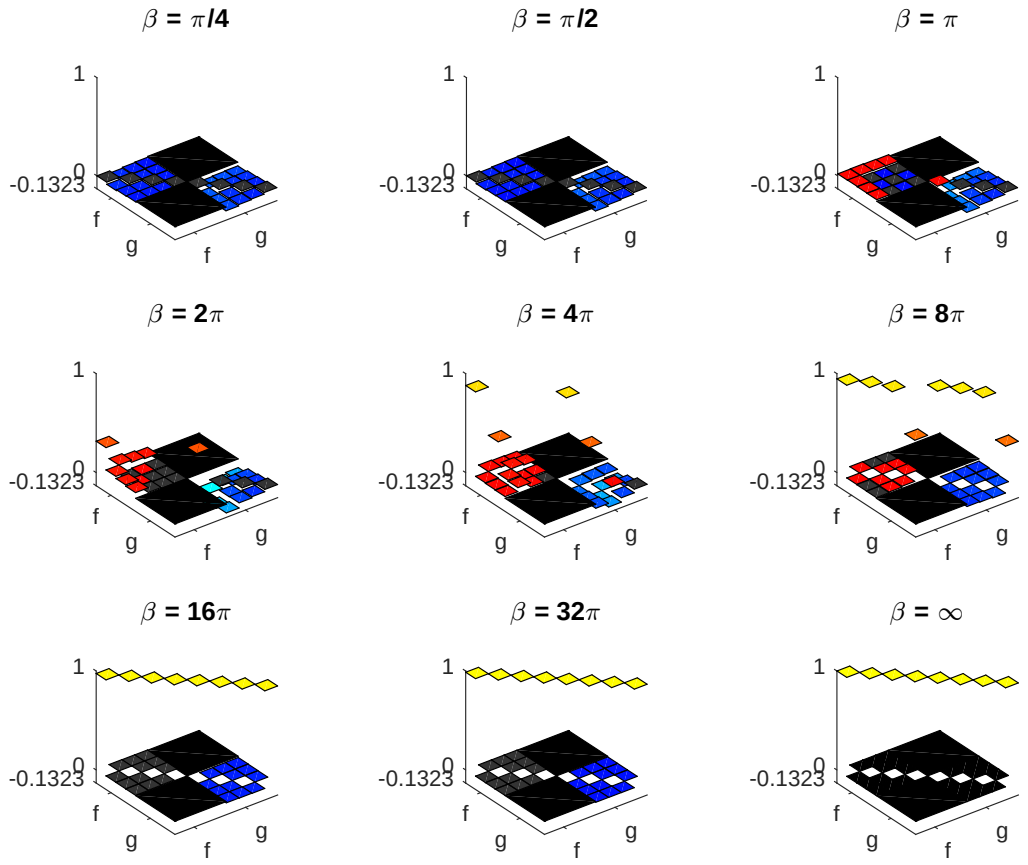


Figure 4.32: *3d plot of the elements of* $2P(\beta)$ *for* $n = 4$. *(This figure is analogous to Figure 4.5)*

relevant notational changes:

$$\Psi'_{\pm}(x_0) = X_0^{\mathrm{T}}\tilde{a}''(\lambda_{\mp}, \kappa(\lambda_{\mp})) \quad = \quad \frac{(\tilde{a}^{\mathrm{T}}P(\beta)\tilde{a})(X_0^{\mathrm{T}}P(\beta)\tilde{a})}{\tilde{a}^{\mathrm{T}}P(\beta)^2\tilde{a}} \pm \frac{\sqrt{n(\tilde{a}^{\mathrm{T}}P(\beta)^2\tilde{a}) - (X_0^{\mathrm{T}}P(\beta)\tilde{a})^2}}{\tilde{a}^{\mathrm{T}}P(\beta)^2\tilde{a}} \times \sqrt{2\tilde{a}^{\mathrm{T}}P(\beta)^2\tilde{a} - (\tilde{a}^{\mathrm{T}}P(\beta)\tilde{a})^2}. \tag{4.97}$$

We will also need

$$\Psi''_{\pm}(x_0) = \pm\sqrt{2n}, \tag{4.98}$$

and then the upper boundary of the envelope is given by

$$\Psi_{+}(x_0) = \begin{cases} \max\{\Psi'_{+}(x_0), \Psi''_{+}(x_0)\}, & \text{if } \sqrt{2}\tilde{a}^{\mathrm{T}}P(\beta)X_0 \geq \sqrt{n}\tilde{a}^{\mathrm{T}}P(\beta)\tilde{a}; \\ \Psi'_{+}(x_0), & \text{otherwise;} \end{cases} \tag{4.99}$$

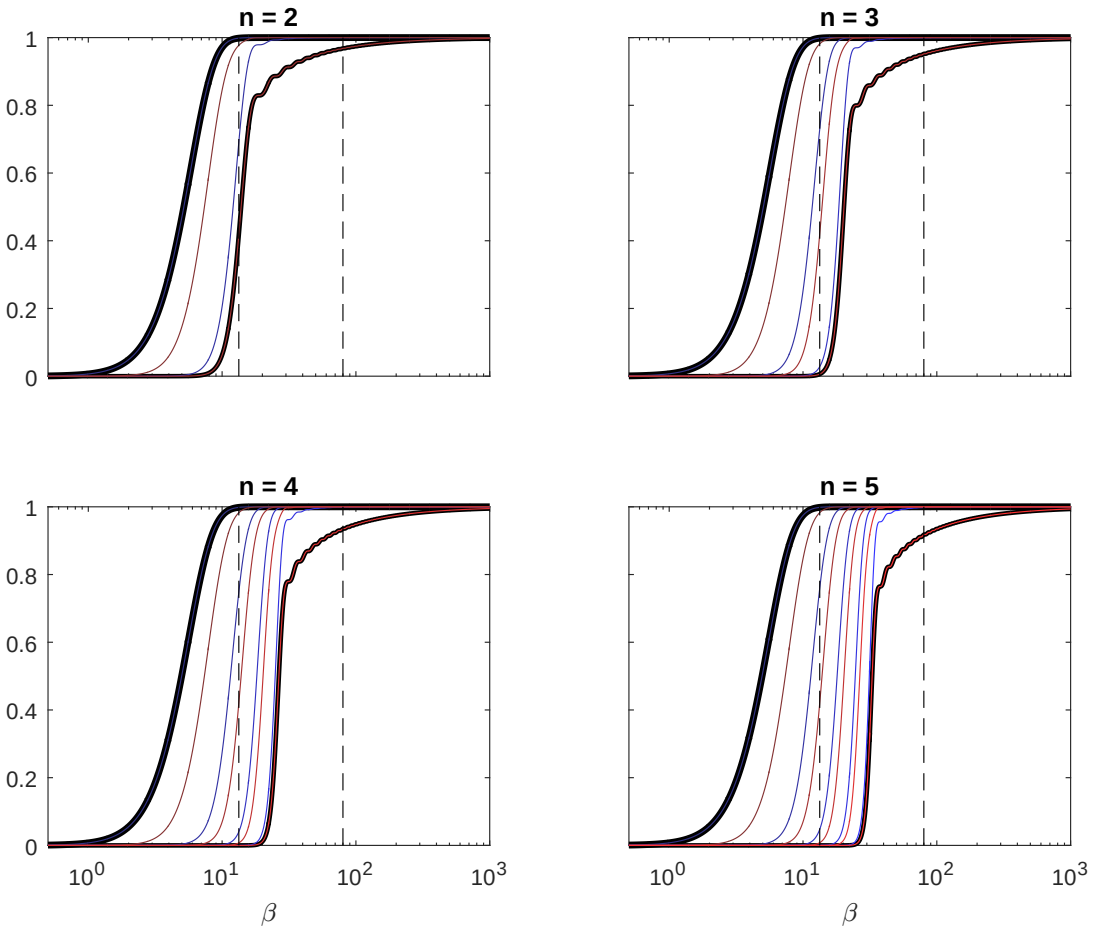


Figure 4.33: *Plots of the eigenvalues of $H^{-1}P(\beta) = 2P(\beta)$, for $n = 2, 3, 4, 5$. The thicker black line gives the minimum of these eigenvalues, the thinner, the maximum. The eigenvalues of $P_f(\beta)$ are shown by the blue curves, those of $P_g(\beta)$ by the red ones. The dashed vertical lines correspond to the bandwidth of the NAT accelerometer (80Hz), and this value divided by the frequency most characteristic of PD movement disorders, 6Hz*

similarly, the lower boundary is given by

$$\Psi_-(x_0) = \begin{cases} \min\{\Psi'_-(x_0), \Psi''_-(x_0)\}, & \text{if } \sqrt{2}\tilde{a}^{\mathrm{T}} P(\beta) X_0 \leq -\sqrt{n}\tilde{a}^{\mathrm{T}} P(\beta)\tilde{a}; \\ \Psi'_-(x_0), & \text{otherwise.} \end{cases} \tag{4.100}$$

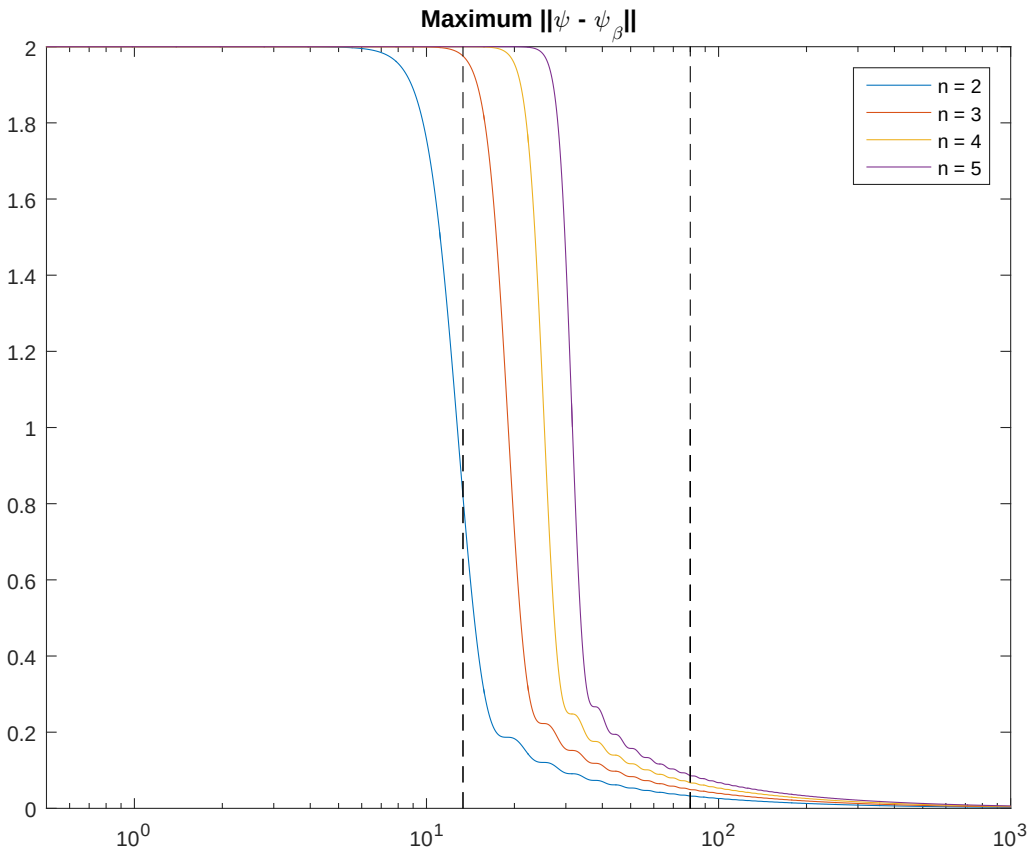


Figure 4.34: *Plots of the maximum distances corresponding to the eigenvalues of Figure 4.33*

4.9 Windowed polynomial wavelets

If we fix β, we can use the frequency-windowed functions

$$f_k(x) = \frac{1}{2\pi}\int_{-\beta}^{\beta}\int_{-\frac{1}{2}}^{\frac{1}{2}} y^k \mathrm{e}^{-i\xi y}\mathrm{d}y\mathrm{e}^{ix\xi}\mathrm{d}\xi = \frac{1}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}} y^k \frac{\sin\beta(y-x)}{y-x}\mathrm{d}y \tag{4.101}$$

as a basis for wavelets $\psi(x) = \sum_{k=0}^{n} a_k f_k(x)$.

The zi condition then becomes $\bar{b}^{\mathrm{T}}a = 0$, where

$$\begin{aligned}
\bar{b}_k &= \int_{-\infty}^{\infty} f_k(x)\mathrm{d}x \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\frac{1}{2}}^{\frac{1}{2}} y^k \frac{\sin\beta(y-x)}{y-x}\mathrm{d}y\mathrm{d}x \\
&= \frac{1}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}} y^k \int_{-\infty}^{\infty} \frac{\sin\beta(y-x)}{y-x}\mathrm{d}x\mathrm{d}y = \int_{-\frac{1}{2}}^{\frac{1}{2}} y^k \mathrm{d}y = \begin{cases} \frac{1}{2^k(k+1)}, & k \text{ even;} \\ 0, & k \text{ odd} \end{cases} = b_k,
\end{aligned} \tag{4.102}$$

and the ue condition $a^{\mathrm{T}}\bar{H}a = 1$, where

$$\begin{aligned}
\bar{H}_{k\ell} &= \int_{-\infty}^{\infty} f_k(x) f_\ell(x)\mathrm{d}x \\
&= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\frac{1}{2}}^{\frac{1}{2}} y^k \frac{\sin\beta(y-x)}{y-x}\mathrm{d}y \int_{-\frac{1}{2}}^{\frac{1}{2}} w^\ell \frac{\sin\beta(w-x)}{w-x}\mathrm{d}w\mathrm{d}x \\
&= \frac{1}{\pi^2}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} y^k w^\ell \int_{-\infty}^{\infty} \frac{\sin\beta(y-x)}{y-x}\frac{\sin\beta(w-x)}{w-x}\mathrm{d}x\mathrm{d}y\mathrm{d}w \\
&= \frac{1}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} y^k w^\ell \frac{\sin\beta(w-y)}{w-y}\mathrm{d}y\mathrm{d}w = P_{k\ell}(\beta);
\end{aligned} \tag{4.103}$$

that is, $\bar{b} = b$ and $\bar{H} = P(\beta)$, where b and $P(\beta)$ are the quantities with the same roles as encountered when discussing piecewise polynomial wavelets.

4.10 Windowed trigonometric polynomial wavelets

If we fix β, we can use the frequency-windowed functions

$$\begin{aligned}
f_k(x) &= \frac{1}{2\pi}\int_{-\beta}^{\beta}\int_{-\frac{1}{2}}^{\frac{1}{2}} \mathrm{e}^{-i\xi y}\sin 2\pi k y\ \mathrm{d}y\ \mathrm{e}^{ix\xi}\ \mathrm{d}\xi = \frac{1}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\sin\beta(y-x)}{y-x}\sin 2\pi k y\ \mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\beta}^{\beta} \mathrm{e}^{ix\xi}\hat{f}_k(\xi)\ \mathrm{d}\xi = \frac{(-1)^{k+1}i}{2\pi}\int_{-\beta}^{\beta} \mathrm{e}^{ix\xi}\left(\frac{1}{\xi-2k\pi} - \frac{1}{\xi+2k\pi}\right)\sin\tfrac{1}{2}\xi\ \mathrm{d}\xi \\
&= \frac{(-1)^k}{2\pi}\int_{-\beta}^{\beta}\left(\frac{1}{\xi-2k\pi} - \frac{1}{\xi+2k\pi}\right)\sin\tfrac{1}{2}\xi\sin x\xi\ \mathrm{d}\xi = \frac{(-1)^k}{\pi}\int_{-\beta}^{\beta}\frac{\sin\frac{1}{2}\xi\sin x\xi}{\xi-2k\pi}\mathrm{d}\xi
\end{aligned} \tag{4.104}$$

(as $\hat{f}_k(\xi)$ is given by equation (4.84) on its support, and we have used the symmetries of the integrands), and

$$\begin{aligned} g_k(x) &= \frac{1}{2\pi}\int_{-\beta}^{\beta}\int_{-\frac{1}{2}}^{\frac{1}{2}} \mathrm{e}^{-i\xi y}\cos 2\pi k y \,\mathrm{d}y\, \mathrm{e}^{ix\xi}\,\mathrm{d}\xi = \frac{1}{\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\sin\beta(y-x)}{y-x}\cos 2\pi k y\,\mathrm{d}y \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\beta}^{\beta}\mathrm{e}^{ix\xi}\hat{g}_k(\xi)\,\mathrm{d}\xi = \frac{(-1)^k}{2\pi}\int_{-\beta}^{\beta}\mathrm{e}^{ix\xi}\left(\frac{1}{\xi-2k\pi}+\frac{1}{\xi+2k\pi}\right)\sin\tfrac{1}{2}\xi\,\mathrm{d}\xi \\ &= \frac{(-1)^k}{2\pi}\int_{-\beta}^{\beta}\left(\frac{1}{\xi-2k\pi}+\frac{1}{\xi+2k\pi}\right)\sin\tfrac{1}{2}\xi\cos x\xi\,\mathrm{d}\xi = \frac{(-1)^k}{\pi}\int_{-\beta}^{\beta}\frac{\sin\frac{1}{2}\xi\cos x\xi}{\xi-2k\pi}\mathrm{d}\xi \end{aligned} \tag{4.105}$$

($\hat{g}_k(\xi)$ is given by equation (4.85) on its support) as a basis for the wavelets $\psi(x) = \sum_{k=1}^{n}[a_k f_k(x) + c_k g_k(x)]$.

By the substitution equivalent to the translation $\xi \to \xi + 2k\pi$ and using the argument addition properties of trigonometric functions,

$$f_k(x) = \tfrac{1}{2\pi}\left[\left\{G_k\left(x+\tfrac{1}{2}\right)-G_k\left(x-\tfrac{1}{2}\right)\right\}\cos 2k\pi x + \left\{F_k\left(x+\tfrac{1}{2}\right)-F_k\left(x-\tfrac{1}{2}\right)\right\}\sin 2k\pi x\right], \tag{4.106}$$

$$g_k(x) = \tfrac{1}{2\pi}\left[\left\{F_k\left(x+\tfrac{1}{2}\right)-F_k\left(x-\tfrac{1}{2}\right)\right\}\cos 2k\pi x - \left\{G_k\left(x+\tfrac{1}{2}\right)-G_k\left(x-\tfrac{1}{2}\right)\right\}\sin 2k\pi x\right], \tag{4.107}$$

where

$$F_k(x) = \mathrm{Si}((\beta+2k\pi)x) + \mathrm{Si}((\beta-2k\pi)x), \tag{4.108}$$

$$G_k(x) = \begin{cases} 0, & \beta = 2k\pi, x = 0; \\ \mathrm{Ci}(4k\pi|x|) - \ln(4k\pi|x|), & \beta = 2k\pi, x \neq 0; \\ \ln\left|\frac{\beta+2k\pi}{\beta-2k\pi}\right|, & \beta \neq 2k\pi, x = 0; \\ \mathrm{Ci}((\beta+2k\pi)|x|) - \mathrm{Ci}(|\beta-2k\pi||x|), & \text{otherwise.} \end{cases} \tag{4.109}$$

The functions given by equations (4.106) and (4.107) are shown in Figure 4.35 for $\beta =$ 80Hz and $k = 1$, 2, 3 and 4. Unsurprisingly, $f_k(x)$ is mostly close to $\sin 2k\pi x$ on $\left[-\frac{1}{2}, \frac{1}{2}\right]$ and mostly close to zero elsewhere, and $g_k(x)$ is mostly close to $\cos 2k\pi x$ on $\left[-\frac{1}{2}, \frac{1}{2}\right]$ and mostly close to zero elsewhere, although there is a certain amount of "ringing" in both these functions.

The zi condition becomes

$$\begin{aligned} \int_{-\infty}^{\infty}\psi(x)\mathrm{d}x &= \int_{-\infty}^{\infty}\sum_{k=1}^{n}[a_k f_k(x) + c_k g_k(x)]\mathrm{d}x \\ &= \sum_{k=1}^{n}\left[a_k\int_{-\frac{1}{2}}^{\frac{1}{2}}\sin 2k\pi x\mathrm{d}x + c_k\int_{-\frac{1}{2}}^{\frac{1}{2}}\cos 2k\pi x\mathrm{d}x\right] = 0, \end{aligned} \tag{4.110}$$

which is automatically satisfied (as for piecewise trigonometric wavelets), and the ue condition

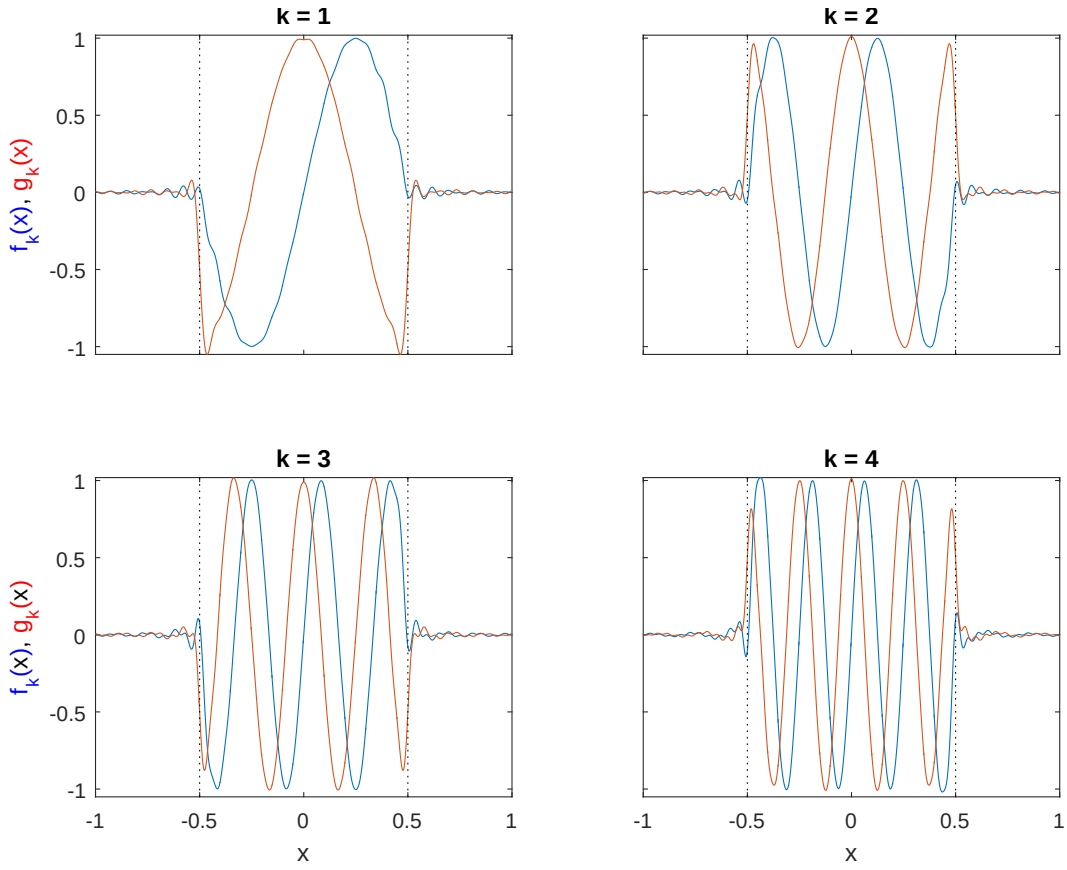


Figure 4.35: $f_1(x)$, $f_2(x)$, $f_3(x)$ *and* $f_4(x)$ *(blue), and* $g_1(x)$, $g_2(x)$, $g_3(x)$ *and* $g_4(x)$ *(red)*

is

$$\begin{aligned}
\int_{-\infty}^{\infty} \psi(x)^2 \mathrm{d}x &= \int_{-\infty}^{\infty} \sum_{k=1}^{n} \sum_{\ell=1}^{n} [a_k f_k(x) + c_k g_k(x)][a_\ell f_\ell(x) + c_\ell g_\ell(x)] \mathrm{d}x \\
&= \sum_{k=1}^{n} \sum_{\ell=1}^{n} \left[a_k a_\ell \int_{-\infty}^{\infty} f_k(x) f_\ell(x) \mathrm{d}x + c_k c_\ell \int_{-\infty}^{\infty} g_k(x) g_\ell(x) \mathrm{d}x \right] \\
&= \frac{1}{4\pi} \sum_{k=1}^{n} \sum_{\ell=1}^{n} \left[a_k a_\ell \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \sin 2k\pi w \sin 2\ell\pi y \frac{\sin \beta(w-y)}{w-y} \mathrm{d}y \mathrm{d}w + \right. \\
&\qquad \left. c_k c_\ell \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos 2k\pi w \cos 2\ell\pi y \frac{\sin \beta(w-y)}{w-y} \mathrm{d}y \mathrm{d}w \right] \\
&= \begin{bmatrix} a^{\mathrm{T}} & c^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} P_f & 0 \\ 0 & P_g \end{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix} = 1, \qquad (4.111)
\end{aligned}$$

where we have exploited the fact that the f_k are even functions and the g_k odd, and changed the order of integration to integrate with respect to x first. In the final line, we have used equation (4.82), where P_f and P_g are given by substituting equations (4.86) to (4.8) into equations (4.83).

We would also like to find

$$\int_{-x_0}^{x_0} \psi(x)^2 \mathrm{d}x = \begin{bmatrix} a^{\mathrm{T}} & c^{\mathrm{T}} \end{bmatrix} T \begin{bmatrix} a \\ c \end{bmatrix},$$

as our new wavelets no longer have compact support and we wish to determine the region where they may be assumed zero without too much loss of accuracy.

The matrix T has the form $T = \begin{bmatrix} T_f(x_0) & 0 \\ 0 & T_g(x_0) \end{bmatrix}$ (again because of the parities of the fs and gs), where

$$T_{f,k\ell}(x_0) = \int_{-x_0}^{x_0} f_k(x) f_\ell(x) \mathrm{d}x, T_{g,k\ell}(x_0) = \int_{-x_0}^{x_0} g_k(x) g_\ell(x) \mathrm{d}x$$

but evaluating these integrals in terms of standard special functions is not easy.

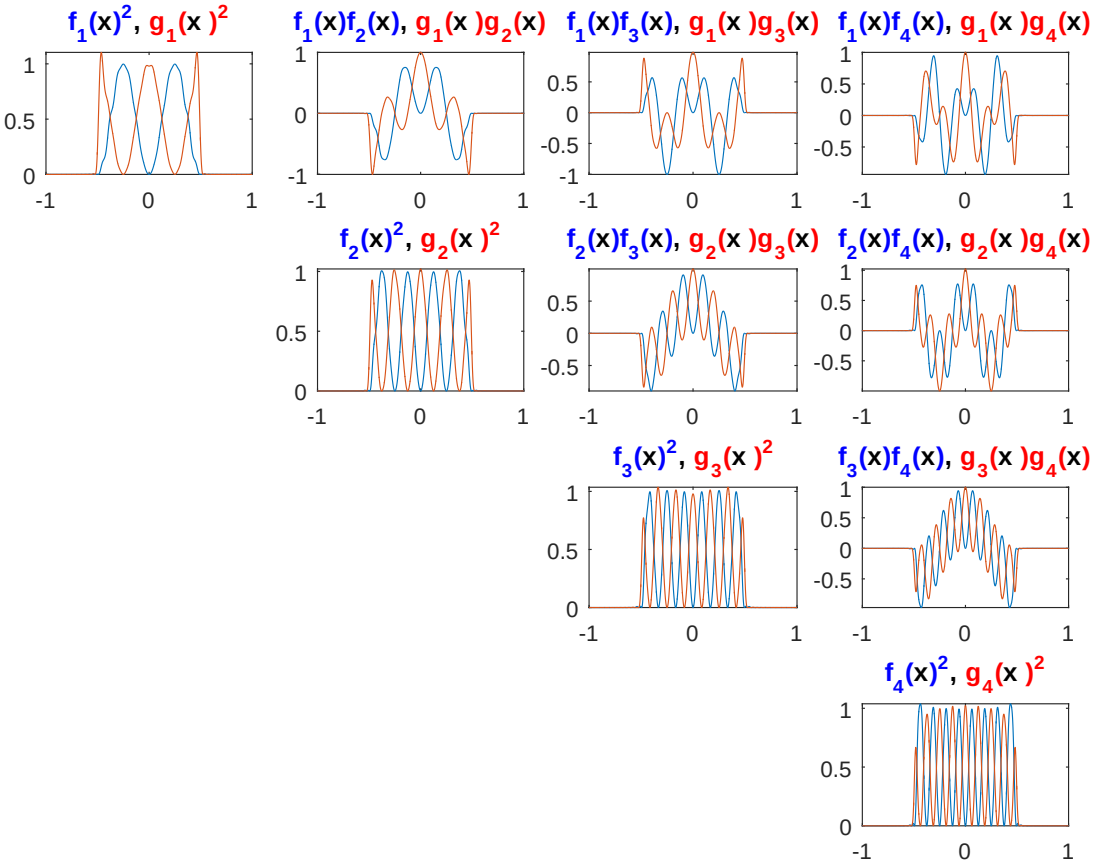


Figure 4.36: *$f_k(x)f_\ell(x)$ (blue) and $g_k(x)g_\ell(x)$ (red) for $k, \ell = 1, 2, 3$ and 4, $(k < \ell)$*

Nevertheless, we can find T numerically, by using Matlab®'s function `integral` to integrate the quantities[7] under the integral sign in the right-handmost members of equa-

[7] As these quantities are sensitive to arithmetical error, we use Matlab®'s `vpa` (variable precision arithmetic)

tions (4.104) and (4.105).

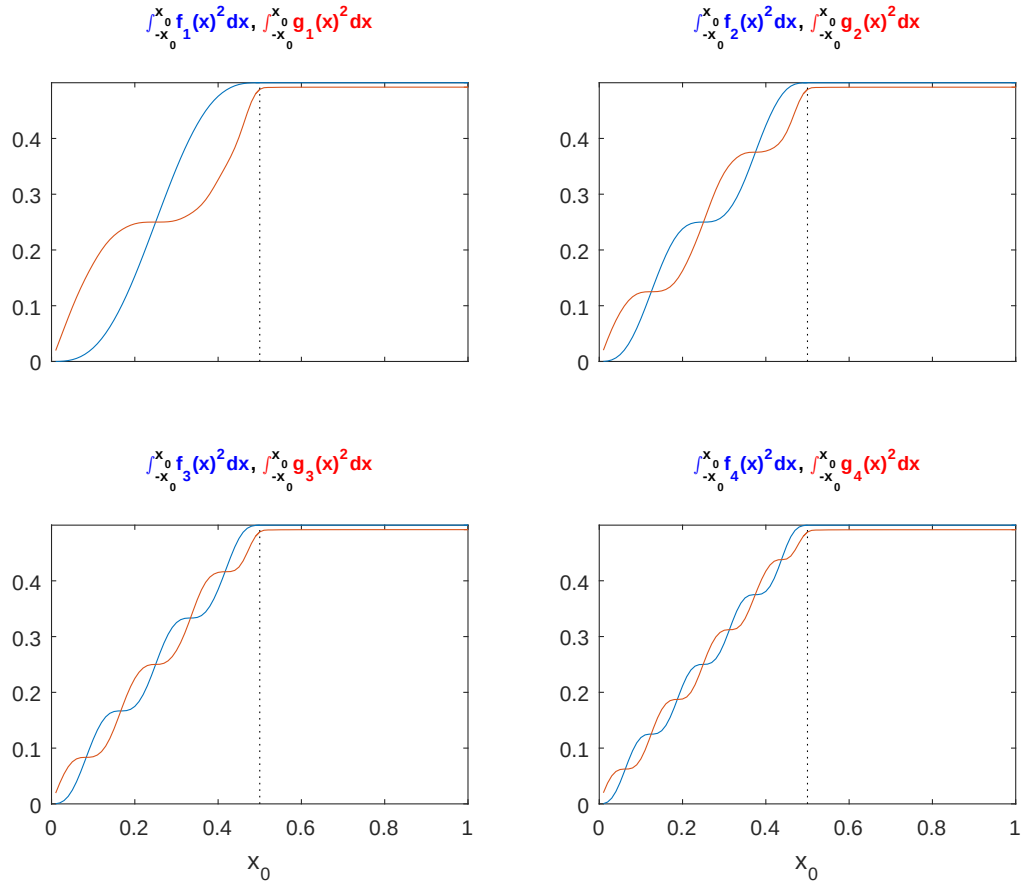


Figure 4.37: *The diagonal elements of T_f (blue) and T_g (red) for $k = 1, 2, 3$ and 4. The vertical dotted line is at $x_0 = \frac{1}{2}$*

The products $f_k(x)f_\ell(x)$ and $g_k(x)g_\ell(x)$ for $\beta = 80$Hz and $k, \ell = 1, 2, 3$ and 4 are shown in Figure 4.36, and the integrals of these quantities (i.e., the elements of T_f and T_g) are shown in Figures 4.37 and 4.38.

In Figures 4.39 and 4.40 we show the elements of $P^{-\frac{1}{2}}T(x_0)P^{-\frac{1}{2}}$, which is a symmetric matrix with the same eigenvalues as $P^{-1}T(x_0)$, and these Figures confirm that $T(x_0) \to P$ rather slowly as $x_0 \to \infty$.

We now wish to find $\min \int_{-x_0}^{x_0} \psi(x)^2 \mathrm{d}x$ over our wavelets for each value of $x_0 \in (0, \infty)$, that is we wish to find $\min \left\{ \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]^{\mathrm{T}} \left[\begin{smallmatrix} T_f(x_0) & 0 \\ 0 & T_g(x_0) \end{smallmatrix}\right] \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right] \right\}$, subject to the condition of the last line of equation (4.111).

This is similar to other extremisation problems we have considered, and we have $\min \left\{ \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]^{\mathrm{T}} T(x_0) \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right] : \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]^{\mathrm{T}} P \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right] = 1 \right\} = \lambda$, where $\lambda = \lambda(x_0)$ is the smallest eigenvalue of $P^{-1}T(x_0)$.

We now find $\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta} f_k(x)\mathrm{d}x$ and $\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta} g_k(x)\mathrm{d}x$ and use these to find filter coefficients,

to evaluate them

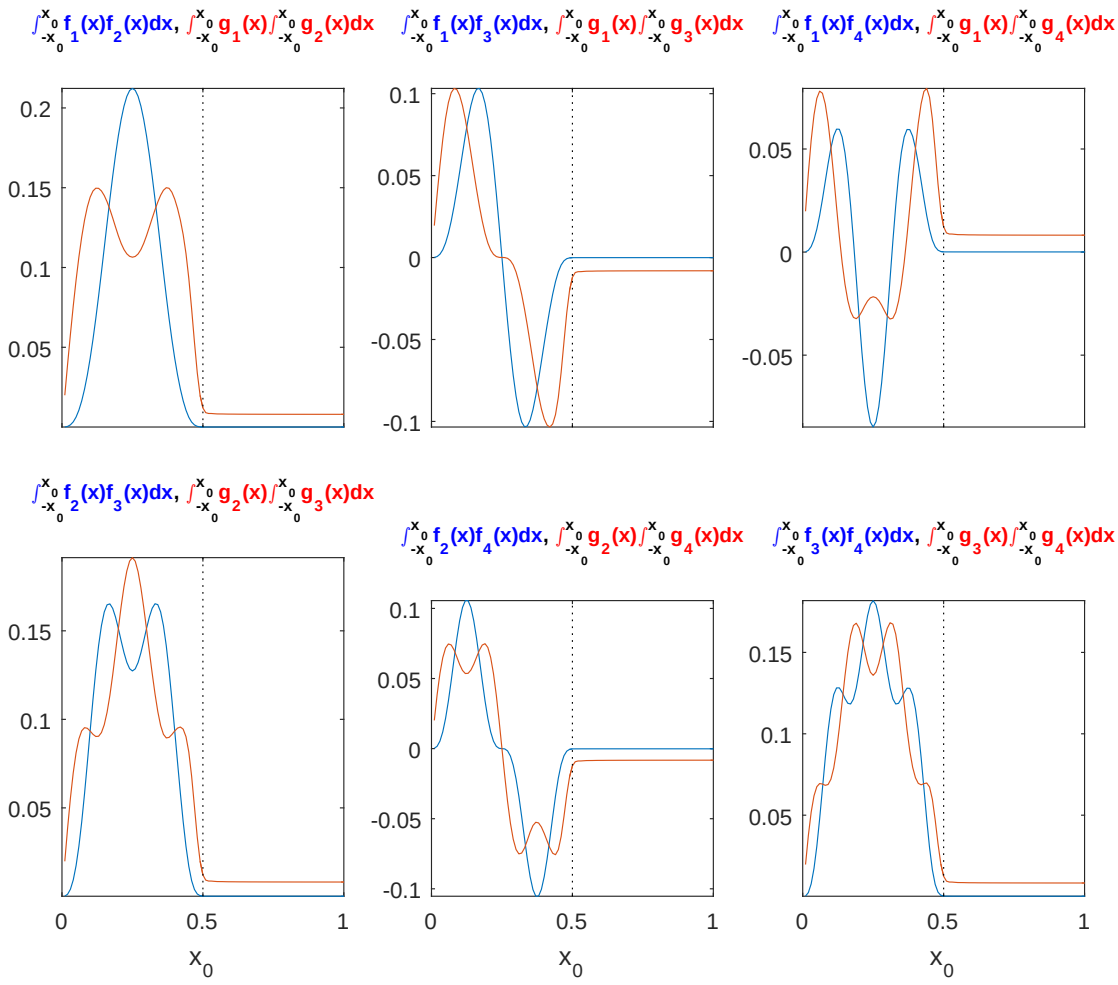


Figure 4.38: *The off-diagonal elements of T_f (blue) and T_g (red) for $k, \ell = 1, 2, 3$ and 4 ($k < \ell$)*

on setting Δ to the gap between sampling points, and x_0 to integer multiples of Δ.

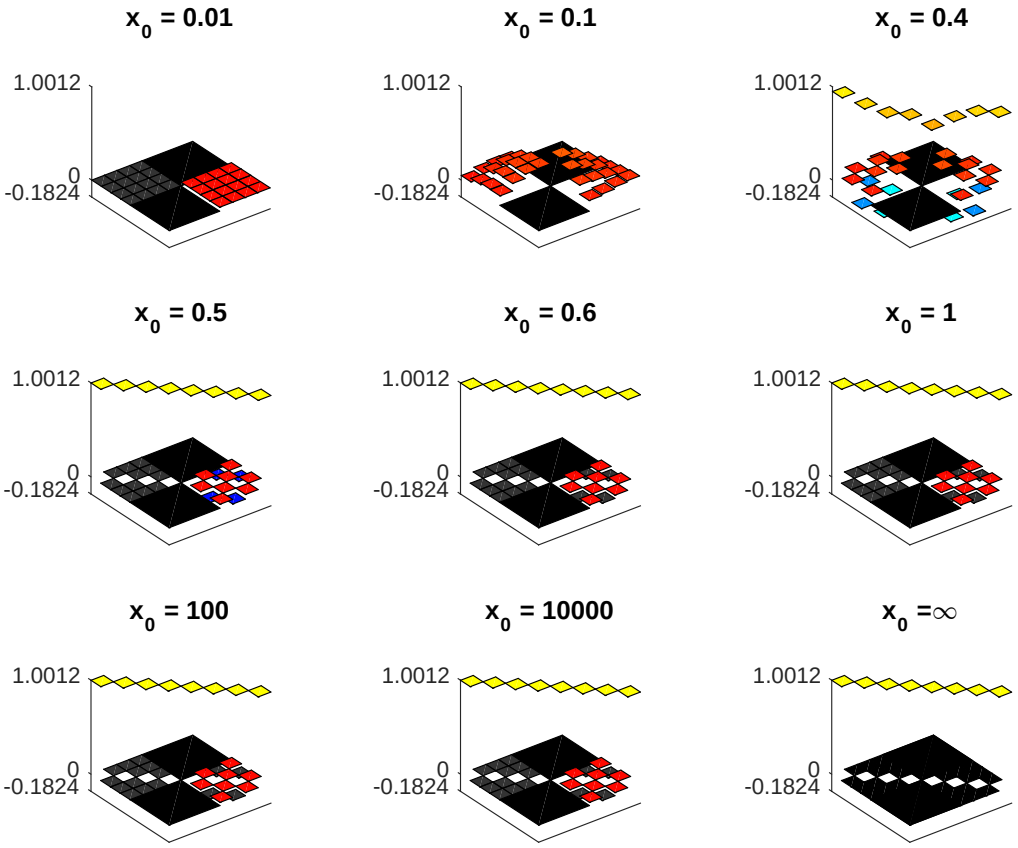


Figure 4.39: *3-d plots of the elements of* $P^{-\frac{1}{2}}T(x_0)P^{-\frac{1}{2}}$ *(*$P_f^{-\frac{1}{2}}T_f(x_0)P_f^{-\frac{1}{2}}$ *to the left corner of the plots, and* $P_g^{-\frac{1}{2}}T_g(x_0)P_g^{-\frac{1}{2}}$ *to the right) for various values of* x_0*. The colour map is given in Figure 4.40 below*

Now,

$$\begin{aligned}
\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta} f_k(x)\mathrm{d}x &= \frac{1}{2\pi}\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\beta}^{\beta} \mathrm{e}^{-i\xi(y-x)}\sin 2k\pi y \,\mathrm{d}\xi\mathrm{d}y\mathrm{d}x \\
&= -\frac{i}{2\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\beta}^{\beta} \mathrm{e}^{-i\xi y}\frac{\mathrm{e}^{i\xi\left(x_0+\frac{1}{2}\Delta\right)}-\mathrm{e}^{i\xi\left(x_0-\frac{1}{2}\Delta\right)}}{\xi}\sin 2k\pi y \,\mathrm{d}\xi\mathrm{d}y \\
&= -\frac{1}{4\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\beta}^{\beta} \mathrm{e}^{ix_0\xi}\frac{\mathrm{e}^{\frac{1}{2}i\Delta\xi}-\mathrm{e}^{-\frac{1}{2}i\Delta\xi}}{\xi}\mathrm{e}^{-i\xi y}\left[\mathrm{e}^{2ik\pi y}-\mathrm{e}^{-2ik\pi y}\right]\mathrm{d}\xi\mathrm{d}y \\
&= \frac{(-1)^{k+1}i}{4\pi}\int_{-\beta}^{\beta} \mathrm{e}^{ix_0\xi}\frac{\mathrm{e}^{\frac{1}{2}i\Delta\xi}-\mathrm{e}^{-\frac{1}{2}i\Delta\xi}}{\xi}\left[\mathrm{e}^{-\frac{1}{2}i\xi}-\mathrm{e}^{\frac{1}{2}i\xi}\right]\times \\
&\qquad\qquad \left[\frac{1}{\xi-2k\pi}-\frac{1}{\xi+2k\pi}\right]\mathrm{d}\xi \\
&= \frac{(-1)^{k+1}i}{\pi}\int_{-\beta}^{\beta} \mathrm{e}^{ix_0\xi}\frac{\sin\frac{1}{2}\Delta\xi\sin\frac{1}{2}\xi}{\xi}\left[\frac{1}{\xi-2k\pi}-\frac{1}{\xi+2k\pi}\right]\mathrm{d}\xi \\
&= \frac{(-1)^{k}}{\pi}\int_{-\beta}^{\beta} \frac{\sin x_0\xi\sin\frac{1}{2}\Delta\xi\sin\frac{1}{2}\xi}{\xi}\left[\frac{1}{\xi-2k\pi}-\frac{1}{\xi+2k\pi}\right]\mathrm{d}\xi
\end{aligned}$$

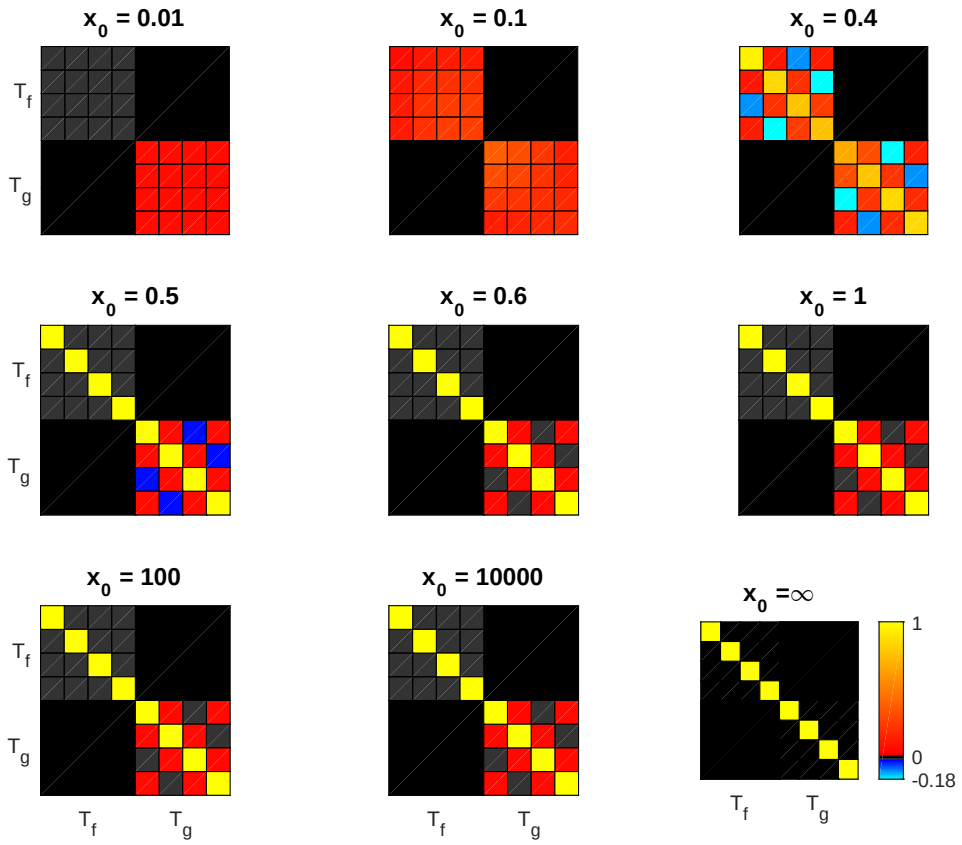


Figure 4.40: *Pseudocolour plots of the elements of* $P^{-\frac{1}{2}}T(x_0)P^{-\frac{1}{2}}$ *(*$P_f^{-\frac{1}{2}}T_f(x_0)P_f^{-\frac{1}{2}}$ *to the upper left corner of the plots, and* $P_g^{-\frac{1}{2}}T_g(x_0)P_g^{-\frac{1}{2}}$ *to the lower right)*

where the last line follows from the fact that the integrand of the previous line without the factor $\mathrm{e}^{ix_0\xi}$ is an odd function of ξ, so the even part of $\mathrm{e}^{ix_0\xi}$ may be ignored.

Hence,

$$\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta} f_k(x)\mathrm{d}x = \frac{2(-1)^k}{\pi}\int_{-\beta}^{\beta}\frac{\sin x_0\xi\sin\frac{1}{2}\Delta\xi\sin\frac{1}{2}\xi}{\xi(\xi-2k\pi)}\mathrm{d}\xi,$$

by the evenness of $\frac{1}{\xi}\sin x_0\xi\sin\frac{1}{2}\Delta\xi\sin\frac{1}{2}\xi$, and so

$$\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta} f_k(x)\mathrm{d}x = \frac{(-1)^k}{k\pi^2}\int_{-\beta}^{\beta}\left[\frac{1}{\xi-2k\pi}-\frac{1}{\xi}\right]\sin x_0\xi\sin\tfrac{1}{2}\Delta\xi\sin\tfrac{1}{2}\xi\mathrm{d}\xi.$$

As $\sin x_0\xi\sin\frac{1}{2}\Delta\xi\sin\frac{1}{2}\xi = -\frac{1}{4}\left\{\sin\left(x_0+\frac{1}{2}+\frac{1}{2}\Delta\right)\xi+\ \sin\left(x_0-\frac{1}{2}-\frac{1}{2}\Delta\right)\xi-\sin\left(x_0+\frac{1}{2}-\frac{1}{2}\Delta\right)\xi-\sin\left(x_0-\frac{1}{2}+\frac{1}{2}\Delta\right)\xi\right\}$ (easily derivable by writing $\sin x_0\xi$ etc. as $\frac{1}{2i}(\mathrm{e}^{ix_0\xi}-\mathrm{e}^{-ix_0\xi})$ etc.), we

have

$$\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta} f_k(x)\mathrm{d}x = J_{k,+,+}(x_0,\Delta) + J_{k,-,-}(x_0,\Delta) - J_{k,+,-}(x_0,\Delta) - J_{k,-,+}(x_0,\Delta), \quad (4.112)$$

where

$$\begin{aligned} J_{k,\pm,\pm'}(x_0,\Delta) &= \frac{(-1)^{k+1}}{4k\pi^2}\int_{-\beta}^{\beta}\left[\frac{1}{\xi-2k\pi}-\frac{1}{\xi}\right]\sin\left(x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right)\xi\mathrm{d}\xi \\ &= \frac{(-1)^{k+1}}{4k\pi^2}\int_{-\beta-2k\pi}^{\beta-2k\pi}\frac{\sin\left(x_0 \pm \frac{1}{2} \pm' \frac{1}{2}\Delta\right)(\xi+2k\pi)}{\xi}\mathrm{d}\xi - \\ &\qquad \frac{(-1)^{k+1}}{4k\pi^2}\int_{-\beta}^{\beta}\frac{\sin\left(x_0 \pm \frac{1}{2} \pm' \frac{1}{2}\Delta\right)\xi}{\xi}\mathrm{d}\xi \\ &= -\frac{1}{4k\pi^2}\int_{-\beta-2k\pi}^{\beta-2k\pi}\frac{\sin\left(\left[x_0 \pm \frac{1}{2} \pm' \frac{1}{2}\Delta\right]\xi + 2k\pi\left[x_0 \pm' \frac{1}{2}\Delta\right]\right)}{\xi}\mathrm{d}\xi + \\ &\qquad \tfrac{(-1)^k}{2k\pi^2}\mathrm{Si}\left(\beta\left[x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right]\right) \end{aligned}$$

or

$$\begin{aligned} J_{k,\pm,\pm'}(x_0,\Delta) &= -\frac{1}{4k\pi^2}\cos 2k\pi\left(x_0 \pm' \tfrac{1}{2}\Delta\right)\int_{-\beta-2k\pi}^{\beta-2k\pi}\frac{\sin\left(x_0 \pm \frac{1}{2} \pm' \frac{1}{2}\Delta\right)\xi}{\xi}\mathrm{d}\xi - \\ &\qquad \frac{1}{4k\pi^2}\sin 2k\pi\left(x_0 \pm' \tfrac{1}{2}\Delta\right)\int_{-\beta-2k\pi}^{\beta-2k\pi}\frac{\cos\left(x_0 \pm \frac{1}{2} \pm' \frac{1}{2}\Delta\right)\xi}{\xi}\mathrm{d}\xi + \\ &\qquad \tfrac{(-1)^k}{2k\pi^2}\mathrm{Si}\left(\beta\left[x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right]\right) \\ &= -\tfrac{1}{4k\pi^2}\cos 2k\pi\left(x_0 \pm' \tfrac{1}{2}\Delta\right)\times \\ &\quad \left\{\mathrm{Si}\big((\beta-2k\pi)\left(x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right)\big) - \mathrm{Si}\big((-\beta-2k\pi)\left(x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right)\big)\right\} - \\ &\quad \tfrac{1}{4k\pi^2}\sin 2k\pi\left(x_0 \pm' \tfrac{1}{2}\Delta\right)\times \\ &\quad \left\{\mathrm{Ci}\big((\beta-2k\pi)\left(x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right)\big) - \mathrm{Ci}\big((-\beta-2k\pi)\left(x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right)\big)\right\} + \\ &\quad \tfrac{(-1)^k}{2k\pi^2}\mathrm{Si}\left(\beta\left[x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right]\right) \\ &= -\tfrac{1}{4k\pi^2}\cos 2k\pi\left(x_0 \pm' \tfrac{1}{2}\Delta\right)\times \\ &\quad \left\{\mathrm{Si}\big((\beta-2k\pi)\left(x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right)\big) + \mathrm{Si}\big((\beta+2k\pi)\left(x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right)\big)\right\} - \\ &\quad \tfrac{1}{4k\pi^2}\sin 2k\pi\left(x_0 \pm' \tfrac{1}{2}\Delta\right)\times \\ &\quad \left\{\mathrm{Ci}\big(|\beta-2k\pi|\left|x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right|\big) - \mathrm{Ci}\big((\beta+2k\pi)\left|x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right|\big) + ui\pi\right\} + \\ &\quad \tfrac{(-1)^k}{2k\pi^2}\mathrm{Si}\left(\beta\left[x_0 \pm \tfrac{1}{2} \pm' \tfrac{1}{2}\Delta\right]\right), \end{aligned} \quad (4.113)$$

where $u = \begin{cases} -1, & \beta < 2k\pi; \\ \text{undetermined}, & \beta = 2k\pi; \\ 0, & \beta > 2k\pi \end{cases}$ (when u is non-zero, it is so as a consequence of using the Cauchy Principal Value of the integral in the definition of Ci).

As we may drop the part of $J_{k,\pm,\pm'}(x_0,\Delta)$ in equation (4.113) which is unchanged under

the interchange $\pm \leftrightarrow \mp$ (or $\pm' \leftrightarrow \mp'$), we have

$$\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta} f_k(x)\mathrm{d}x = J'_{k,+,+}(x_0,\Delta) + J'_{k,-,-}(x_0,\Delta) - J'_{k,+,-}(x_0,\Delta) - J'_{k,-,+}(x_0,\Delta), \quad (4.114)$$

where

$$\begin{aligned} J'_{k,\pm,\pm'}(x_0,\Delta) &= -\tfrac{1}{4k\pi^2}\cos 2k\pi\left(x_0 \pm' \tfrac{1}{2}\Delta\right)\times \\ &\quad \left\{\mathrm{Si}\big((\beta-2k\pi)\left(x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right)\big)+\mathrm{Si}\big((\beta+2k\pi)\left(x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right)\big)\right\} - \\ &\quad \tfrac{1}{4k\pi^2}\sin 2k\pi\left(x_0 \pm' \tfrac{1}{2}\Delta\right)\times \\ &\quad \left\{\mathrm{Ci}\big(|\beta-2k\pi|\,\big|x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\big|\big)-\mathrm{Ci}\big((\beta+2k\pi)\,\big|x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\big|\big)\right\} + \\ &\quad \tfrac{(-1)^k}{2k\pi^2}\mathrm{Si}\left(\beta\left[x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right]\right). \end{aligned} \quad (4.115)$$

Similarly,

$$\begin{aligned} \int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta} g_k(x)\mathrm{d}x &= \frac{1}{2\pi}\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\beta}^{\beta} \mathrm{e}^{-i\xi(y-x)}\cos 2k\pi y\,\mathrm{d}\xi\mathrm{d}y\mathrm{d}x \\ &= -\frac{i}{4\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\beta}^{\beta}\mathrm{e}^{ix_0\xi}\frac{\mathrm{e}^{\frac{1}{2}i\Delta\xi}-\mathrm{e}^{-\frac{1}{2}i\Delta\xi}}{\xi}\mathrm{e}^{-i\xi y}\left[\mathrm{e}^{2ik\pi y}+\mathrm{e}^{-2ik\pi y}\right]\mathrm{d}\xi\mathrm{d}y \\ &= \frac{(-1)^k}{4\pi}\int_{-\beta}^{\beta}\mathrm{e}^{ix_0\xi}\frac{\mathrm{e}^{\frac{1}{2}i\Delta\xi}-\mathrm{e}^{-\frac{1}{2}i\Delta\xi}}{\xi}\left[\mathrm{e}^{-\frac{1}{2}i\xi}-\mathrm{e}^{\frac{1}{2}i\xi}\right]\left[\frac{1}{\xi-2k\pi}+\frac{1}{\xi+2k\pi}\right]\mathrm{d}\xi \\ &= \frac{(-1)^k}{\pi}\int_{-\beta}^{\beta}\mathrm{e}^{ix_0\xi}\frac{\sin\frac{1}{2}\Delta\xi\sin\frac{1}{2}\xi}{\xi}\left[\frac{1}{\xi-2k\pi}+\frac{1}{\xi+2k\pi}\right]\mathrm{d}\xi \\ &= \frac{(-1)^k}{\pi}\int_{-\beta}^{\beta}\frac{\cos x_0\xi\sin\frac{1}{2}\Delta\xi\sin\frac{1}{2}\xi}{\xi}\left[\frac{1}{\xi-2k\pi}+\frac{1}{\xi+2k\pi}\right]\mathrm{d}\xi \end{aligned}$$

where the last line follows from the fact that the integrand of the previous line without the factor $\mathrm{e}^{ix_0\xi}$ is an even function of ξ, so the odd part of $\mathrm{e}^{ix_0\xi}$ may be ignored.

Hence,

$$\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta} f_k(x)\mathrm{d}x = \frac{(-1)^k}{k\pi^2}\int_{-\beta}^{\beta}\left[\frac{1}{\xi-2k\pi}-\frac{1}{\xi}\right]\cos x_0\xi\sin\tfrac{1}{2}\Delta\xi\sin\tfrac{1}{2}\xi\mathrm{d}\xi,$$

by the oddness of $\frac{1}{\xi}\cos x_0\xi\sin\frac{1}{2}\Delta\xi\sin\frac{1}{2}\xi$.

As $\cos x_0\xi\cos\frac{1}{2}\Delta\xi\cos\frac{1}{2}\xi = -\frac{1}{4}\left\{\cos\left(x_0+\frac{1}{2}+\frac{1}{2}\Delta\right)\xi + \cos\left(x_0-\frac{1}{2}-\frac{1}{2}\Delta\right)\xi - \cos\left(x_0+\frac{1}{2}-\frac{1}{2}\Delta\right)\xi - \cos\left(x_0-\frac{1}{2}+\frac{1}{2}\Delta\right)\xi\right\}$ (easily derivable by writing $\cos x_0\xi$ as $\frac{1}{2}(\mathrm{e}^{ix_0\xi}+\mathrm{e}^{-ix_0\xi})$ etc. and $\sin\frac{1}{2}\xi$ etc. as $\frac{1}{2i}(\mathrm{e}^{\frac{1}{2}i\xi}-\mathrm{e}^{-\frac{1}{2}i\xi})$ etc.), we have

$$\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta} f_k(x)\mathrm{d}x = K_{k,+,+}(x_0,\Delta)+K_{k,-,-}(x_0,\Delta)-K_{k,+,-}(x_0,\Delta)-K_{k,-,+}(x_0,\Delta), \quad (4.116)$$

where

$$
\begin{aligned}
K_{k,\pm,\pm'}(x_0,\Delta) &= \frac{(-1)^{k+1}}{4k\pi^2}\int_{-\beta}^{\beta}\left[\frac{1}{\xi-2k\pi}-\frac{1}{\xi}\right]\cos\left(x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right)\xi\mathrm{d}\xi \\
&= \frac{(-1)^{k+1}}{4k\pi^2}\int_{-\beta-2k\pi}^{\beta-2k\pi}\frac{\cos\left(x_0\pm\frac{1}{2}\pm'\frac{1}{2}\Delta\right)(\xi+2k\pi)}{\xi}\mathrm{d}\xi - \\
&\qquad \frac{(-1)^{k+1}}{4k\pi^2}\int_{-\beta}^{\beta}\frac{\cos\left(x_0\pm\frac{1}{2}\pm'\frac{1}{2}\Delta\right)\xi}{\xi}\mathrm{d}\xi \\
&= -\frac{1}{4k\pi^2}\int_{-\beta-2k\pi}^{\beta-2k\pi}\frac{\cos\left(\left[x_0\pm\frac{1}{2}\pm'\frac{1}{2}\Delta\right]\xi+2k\pi\left[x_0\pm'\frac{1}{2}\Delta\right]\right)}{\xi}\mathrm{d}\xi \\
&= -\frac{1}{4k\pi^2}\cos 2k\pi\left(x_0\pm'\tfrac{1}{2}\Delta\right)\int_{-\beta-2k\pi}^{\beta-2k\pi}\frac{\cos\left(x_0\pm\frac{1}{2}\pm'\frac{1}{2}\Delta\right)\xi}{\xi}\mathrm{d}\xi + \\
&\qquad \frac{1}{4k\pi^2}\sin 2k\pi\left(x_0\pm'\tfrac{1}{2}\Delta\right)\int_{-\beta-2k\pi}^{\beta-2k\pi}\frac{\sin\left(x_0\pm\frac{1}{2}\pm'\frac{1}{2}\Delta\right)\xi}{\xi}\mathrm{d}\xi \\
&= -\tfrac{1}{4k\pi^2}\cos 2k\pi\left(x_0\pm'\tfrac{1}{2}\Delta\right)\times \\
&\quad \left\{\mathrm{Ci}\big((\beta-2k\pi)\left(x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right)\big)-\mathrm{Ci}\big((-\beta-2k\pi)\left(x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right)\big)\right\}+ \\
&\quad \tfrac{1}{4k\pi^2}\sin 2k\pi\left(x_0\pm'\tfrac{1}{2}\Delta\right)\times \\
&\quad \left\{\mathrm{Si}\big((\beta-2k\pi)\left(x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right)\big)-\mathrm{Si}\big((-\beta-2k\pi)\left(x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right)\big)\right\} \\
&= -\tfrac{1}{4k\pi^2}\cos 2k\pi\left(x_0\pm'\tfrac{1}{2}\Delta\right)\times \\
&\quad \left\{\mathrm{Ci}\big(|\beta-2k\pi|\left|x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right|\big)-\mathrm{Ci}\big(|\beta+2k\pi|\left|x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right|\big)+u\pi i\right\}+ \\
&\quad \tfrac{1}{4k\pi^2}\sin 2k\pi\left(x_0\pm'\tfrac{1}{2}\Delta\right)\times \\
&\quad \left\{\mathrm{Si}\big((\beta-2k\pi)\left(x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right)\big)+\mathrm{Si}\big((\beta+2k\pi)\left(x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right)\big)\right\}.
\end{aligned}
\tag{4.117}
$$

As we may drop the part of $K_{k,\pm,\pm'}(x_0,\Delta)$ in equation (4.117) which is unchanged under $\pm\leftrightarrow\mp$ (or $\pm'\leftrightarrow\mp'$), we have

$$
\int_{x_0-\frac{1}{2}\Delta}^{x_0+\frac{1}{2}\Delta} f_k(x)\mathrm{d}x = K'_{k,+,+}(x_0,\Delta)+K'_{k,-,-}(x_0,\Delta)-K'_{k,+,-}(x_0,\Delta)-K'_{k,-,+}(x_0,\Delta),
$$

where

$$
\begin{aligned}
K'_{k,\pm,\pm'}(x_0,\Delta) &= -\tfrac{1}{4k\pi^2}\cos 2k\pi\left(x_0\pm'\tfrac{1}{2}\Delta\right)\times \\
&\qquad \left\{\mathrm{Ci}\big(|\beta-2k\pi|\left|x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right|\big)-\mathrm{Ci}\big(|\beta+2k\pi|\left|x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right|\big)\right\}+ \\
&\quad \tfrac{1}{4k\pi^2}\sin 2k\pi\left(x_0\pm'\tfrac{1}{2}\Delta\right)\times \\
&\qquad \left\{\mathrm{Si}\big((\beta-2k\pi)\left(x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right)\big)+\mathrm{Si}\big((\beta+2k\pi)\left(x_0\pm\tfrac{1}{2}\pm'\tfrac{1}{2}\Delta\right)\big)\right\}.
\end{aligned}
\tag{4.118}
$$

4.11 The subsidiary hypothesis

As the results of this section have shown that wavelets obeying the wec and wed conditions are successively less affected by bandwidth limitations, we hypothesis that wavelets obeying the endpoint continuity condition, ec,

$$\psi\left(-\frac{1}{2}\right) = \psi\left(\frac{1}{2}\right) = 0 \tag{4.119}$$

(which implies wec) and endpoint differentiability conditions, ed,

$$\lim_{x\to-\frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}x}\psi(x) = \lim_{x\to\frac{1}{2}} \frac{\mathrm{d}}{\mathrm{d}x}\psi(x) = 0 \tag{4.120}$$

(which implies wed) will be less affected by these limitations than wec and wed respectively.

Numerical problems have made an explicit calculation of the effect of ec and ed difficult, but we believe that these difficulties can be overcome.

Nevertheless, if this hypothesis holds, then it is likely that the further hypothesis mentioned at the start of this Chapter will also hold: i.e., the performance of the method with wavelets with ec or ed will be better than that without.

4.12 Summary

The most important result of this chapter is that, according to a worst-case analysis, the bandwidth of the NAT accelerometer is on the boundary of what is needed to obtain data of the quality to make the method work, but that the effects of the bandwidth are ameliorated by the imposition of stricter smoothness conditions, at least as far as the wec and wed conditions are concerned. The quickest way to see this is to examine Figures 4.18 to 4.20 and 4.25 to 4.27.

Although this is a worst-case analysis, and we might be "lucky" and find ourselves working in a region of the function spaces concerned which is away from the worst part of that space, we cannot rely on this.

In addition, the amelioration through smoothness conditions has enabled us to erect a subsidiary hypothesis.

Chapter 5

Experiments and results

Our aim here is to evaluate the efficacy of classification of unseen data for our method, when it is trained on data from known sources.

We wish to recognise unseen data from the same sources as was used in training, but also to see how the method generalises across wrist of attachment, and across subjects. As the number of our subjects is very limited, we can only look at generalisation across control subjects.

We will also compare changes in performance as we introduce the ec and ed conditions (defined on p192) on our ECWTs.

A major consideration is the question of *diversity* — we do not wish to measure the same characteristic in many ways, but to measure *different* characteristics.

5.1 Introduction and method

The method utilised here is adapted from that discussed in Kohavi[46], and found to have good results in estimating the performance of a wide range of algorithms. Following that reference, we basically divide our data into ten "plies", train a classifier on nine of these and test on the tenth. This procedure is repeated which each of the ten plies playing the role of test ply in turn. This procedure is believed to yield a good approximation to performance statistics generated on larger amounts of data without re-use.

Here, we use seven sets of triaxial data: left hand and right hand data from subjects PD1, who has Parkinson's Disease, C2 and C4, and left hand data from subject C3. Subjects C2 and C3 have no known underlying health problems, and C4 is a diabetic.

All the data was generated by NATs operating at a sampling rate of 500Hz with the range set to ±2g, with a right-hand or left-hand wrist attachment.

The data was divided up into 5 minute epochs of consecutive samples, with any epoch containing saturated data (any of the three channels registering $\pm 2\frac{5}{18}$g) discarded.

The dimensions of the data and the uses we put each set to are given in Table 5.1.

For each data set which has a "Y" in the "Library extraction and member selection" column of Table 5.1, we uniformly randomly select 10 non-intersecting "plies" of $\lfloor N_{\text{data set}}/10 \rfloor$ epochs, where, of course, $N_{\text{data set}}$ is the number of epochs in the data set.

Subject	Hand	Number of epochs	Number of epochs in a ply	Library extraction and member selection	Testing
PD1	left	385	38	Y	Y
	right	527	52	Y	Y
C2	left	350	35	Y	Y
	right	441	44	Y	Y
C3	left	296	29	N	Y
C4	left	105	10	N	Y
	right	477	47	N	Y

Table 5.1: *Size and use of each data set*

We will specify a *basic test suite* by giving three data sources $(\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3)$, where $\mathcal{S}_1, \mathcal{S}_2 \in \{(\text{PD1}, \text{LH}), (\text{PD1}, \text{RH}), (\text{C2}, \text{LH}), (\text{C2}, \text{RH})\}$, $\mathcal{S}_2 \neq \mathcal{S}_1$, and $\mathcal{S}_3 \in \{(\text{PD1}, \text{LH}), (\text{PD1}, \text{RH}), (\text{C2}, \text{LH}), (\text{C2}, \text{RH}), (\text{C3}, \text{LH}), (\text{C4}, \text{LH}), (\text{C4}, \text{RH})\} - \{\mathcal{S}_1\}$. The first two elements of a test suite, $(\mathcal{S}_1, \mathcal{S}_2)$ will be called a *training suite*. A full *test suite*, $(\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4)$ is the combination of the two basic test suites $(\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3)$ and $(\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_4)$. We will label some of the sections of our text by the relevant suites.

Test suite	Experiment type
(PD1, RH, C2, RH, PD1, RH, C2, RH)	Recognition
(PD1, LH, C2, LH, PD1, LH, C2, LH)	Recognition
(C2, RH, PD1, RH, C2, RH, PD1, RH)	Recognition
(C2, LH, PD1, LH, C2, LH, PD1, LH)	Recognition
(PD1, RH, C2, RH, PD1, LH, C2, LH)	Generalisation: attachment
(PD1, LH, C2, LH, PD1, RH, C2, RH)	Generalisation: attachment
(PD1, RH, C2, RH, PD1, RH, C4, RH)	Generalisation: control subject
(PD1, LH, C2, LH, PD1, LH, C3, LH)	Generalisation: control subject
(PD1, LH, C2, LH, PD1, LH, C4, LH)	Generalisation: control subject

Table 5.2: *Test suites and experiment types*

In Table 5.2, we list our test suites and the type of experiment they are involved in: a *recognition* experiment is when an attempt is made to classify different data from the same sources as the training data, and a *generalisation* experiment is when an attempt is made to classify data from (at least partially) different sources to the training data. (If some of the data in a generalisation experiment is not from a new source, we ensure that it is not the same as the training data.)

For each ply k of $\mathcal{S}_1$, $\mathcal{F}_{1k}$, we extract a library of ECWTs, $\mathcal{L}^{\mathcal{S}_1 k}$, from data comprised of a uniformly randomly selected subset $\mathcal{F}'_{1k}$ containing $\frac{2}{3}$ of the members of $\cup_{k' \neq k} \mathcal{F}_{1k'}$. We then attempt to select ten library members, $\mathcal{M}^{\mathcal{S}_1 \mathcal{S}_2 k m}$, indexed by m, from $\mathcal{L}^{\mathcal{S}_1 k}$ by using

$\cup_{k' \neq k} \mathcal{F}_{1k'} - \mathcal{F}'_{1k}$ and $\cup_{k' \neq k} \mathcal{F}_{2k'}$ as a contrast set, where the $\mathcal{F}_{2k'}$ are the plies of $\mathcal{S}_2$,[1] by using the q^{**} method at the end of the previous chapter, which, of course, also delivers a $\Theta^{\mathcal{S}_1\mathcal{S}_2 km}$ and $\theta^{\mathcal{S}_1\mathcal{S}_2 km}$ for each library member m successfully selected (and an indication of whether the proportion of PD activations exceeding $\theta^{\mathcal{S}_1\mathcal{S}_2 km}$ "should" be greater than $\Theta^{\mathcal{S}_1\mathcal{S}_2 km}$ and the proportion of non-PD activations exceeding $\theta^{\mathcal{S}_1\mathcal{S}_2 km}$ less than $\Theta^{\mathcal{S}_1\mathcal{S}_2 km}$ — the "normal" situation — or *vice versa*).

We then evaluate the activations of each window of each of our chosen lengths that fits into each of the the epochs of $\mathcal{F}_{1k}$ with respect to $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 km}$, to find the ratio of the number of the activations exceeding $\theta^{\mathcal{S}_1\mathcal{S}_2 km}$ to the total number for each epoch, $\rho^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_3 \ell}$. The first superscript of ρ here is, of course, the data source for the extraction of the library, the second is the data source which serves as contrast to the extraction source in the selection of the library members, the third is the ply index and the fourth is the library member. The first subscript is the source of the data being tested and the second numbers the epoch of the test data.

If $\mathcal{S}_3 = \mathcal{S}_2$, we find $\rho^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_2 \ell}$ only for the epochs ℓ of $\mathcal{F}_{2k}$, to avoid using epochs of $\mathcal{S}_2$ in both test and training (contrast) roles.

On the other hand, if $\mathcal{S}_3 \neq \mathcal{S}_2$, we find $\rho^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_3 \ell}$ for each epoch ℓ of $\mathcal{S}_3$.

As the libraries $\mathcal{L}^{\mathcal{S}_1 k}$ and $\mathcal{L}^{\mathcal{S}_1 k'}$ are extracted from different data sets, unless $k' = k$, they are very unlikely to contain the same ECWTs, and, if even they did, the ordered sets of selected members $(\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 k1}, \mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 k2}, \mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 k3}, \ldots)$ and $(\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 k'1}, \mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 k'2}, \mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 k'3}, \ldots)$ are likely to differ, as selection also depends on different sets. Consequently, the ρs with respect to the best-ranked $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 km}$ and $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 k'm'}$, $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 k1}$ and $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 k'1}$, i.e., $\rho^{\mathcal{S}_1\mathcal{S}_2 k1}_{\mathcal{S}_3 \ell}$ and $\rho^{\mathcal{S}_1\mathcal{S}_2 k'1}_{\mathcal{S}_3 \ell}$, are likely to be with respect to different ECWTs, and, of course, the same applies to equal-ranked ECWTs of lower rank, m, and their corresponding ρs, $\rho^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_3 \ell}$ and $\rho^{\mathcal{S}_1\mathcal{S}_2 k'm}_{\mathcal{S}_3 \ell}$. This fact is important when inspecting some of the later diagrams.

We evaluate our results in three ways:

1 we classify each ply with respect to the libraries generated for it;

2 we classify each epoch of each ply with respect to the libraries generated for its ply;

3 we look at pairs of value lists (repetitions allowed) $\left\{\rho^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_2 1}, \rho^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_2 2}, \rho^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_2 3}, \ldots\right\}$ and $\left\{\rho^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_3 1}, \rho^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_3 2}, \rho^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_3 3}, \ldots\right\}$ to see if they have significant differences.

Before doing so, set the variable $\iota^{\mathcal{S}_1\mathcal{S}_2 km}$ to 1 if $\text{median}_\ell\left(\tilde{\rho}^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_1 \ell}\right) > \text{median}_\ell\left(\tilde{\rho}^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_2 \ell}\right)$, to -1 if the inequality is reversed, and to 0 if neither inequality holds, where $\tilde{\rho}^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_3 \ell}$ is the equivalent of $\rho^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_1 \ell}$ calculated from the part of $\mathcal{S}_1$ (when $\mathcal{S}_3 = \mathcal{S}_1$) used in the selection of $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2 m\ell}$ from $\mathcal{L}^{\mathcal{S}_1 m}$, or the part of $\mathcal{S}_2$ (when $\mathcal{S}_3 = \mathcal{S}_2$) used in the same process. When $\iota^{\mathcal{S}_1\mathcal{S}_2 km} = 1$, what "should" happen happens, and the tail (as defined by $\theta^{\mathcal{S}_1\mathcal{S}_2 km}$) of the empirical distribution of $\tilde{\rho}^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_1 \ell}$ is fatter than that of the empirical distribution of $\tilde{\rho}^{\mathcal{S}_1\mathcal{S}_2 km}_{\mathcal{S}_2 \ell}$,

[1] We somewhat arbitrarily pair $\mathcal{F}_{1k'}$ and $\mathcal{F}_{2k'}$ for each k', but this should not matter, as the plies of $\mathcal{S}_1$ and $\mathcal{S}_2$ are selected independently

when $\iota^{\mathcal{S}_1\mathcal{S}_2km} = -1$, the relationship between the tails of these two distributions is reversed, and when $\iota^{\mathcal{S}_1\mathcal{S}_2km} = 0$, the tails are equally fat, and the selection of $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2m\ell}$ fails, anyway.

We now work on the assumption that $\mathcal{S}_1$ is a PD data set, and $\mathcal{S}_2$ and $\mathcal{S}_3$ are not PD data sets, or *vice versa*.

(a) If $\iota^{\mathcal{S}_1\mathcal{S}_2km} = 1$

1. find $\text{median}_\ell\left(\rho_{\mathcal{S}_1\ell}^{\mathcal{S}_1\mathcal{S}_2km}\right)$ and $\text{median}_\ell\left(\rho_{\mathcal{S}_3\ell}^{\mathcal{S}_1\mathcal{S}_2km}\right)$.
   i. If $\text{median}_\ell\left(\rho_{\mathcal{S}_1\ell}^{\mathcal{S}_1\mathcal{S}_2km}\right) > \Theta^{\mathcal{S}_1\mathcal{S}_2km}$, label $\mathcal{S}_1$ as PD with respect to $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2m\ell}$ if $\mathcal{S}_1$ is PD, and as non-PD with respect to $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2m\ell}$ if $\mathcal{S}_1$ is non-PD — that is label $\mathcal{S}_1$ with its own class with respect to $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2m\ell}$;
   ii. If $\text{median}_\ell\left(\rho_{\mathcal{S}_1\ell}^{\mathcal{S}_1\mathcal{S}_2km}\right) \leq \Theta^{\mathcal{S}_1\mathcal{S}_2km}$, label $\mathcal{S}_1$ with $\mathcal{S}_2$'s class with respect to $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2m\ell}$;
   iii. If $\text{median}_\ell\left(\rho_{\mathcal{S}_3\ell}^{\mathcal{S}_1\mathcal{S}_2km}\right) > \Theta^{\mathcal{S}_1\mathcal{S}_2km}$, label $\mathcal{S}_3$ with $\mathcal{S}_1$'s class with respect to $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2m\ell}$;
   iv. If $\text{median}_\ell\left(\rho_{\mathcal{S}_3\ell}^{\mathcal{S}_1\mathcal{S}_2km}\right) \leq \Theta^{\mathcal{S}_1\mathcal{S}_2km}$, label $\mathcal{S}_3$ with $\mathcal{S}_2$'s class with respect to $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2m\ell}$;
2. find $\rho_{\mathcal{S}_1\ell}^{\mathcal{S}_1\mathcal{S}_2km}$ and $\rho_{\mathcal{S}_3\ell}^{\mathcal{S}_1\mathcal{S}_2km}$.
   i. If $\rho_{\mathcal{S}_1\ell}^{\mathcal{S}_1\mathcal{S}_2km} > \Theta^{\mathcal{S}_1\mathcal{S}_2km}$, label $\mathcal{F}_{1\ell}$ with $\mathcal{S}_1$'s class with respect to $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2m\ell}$;
   ii. If $\rho_{\mathcal{S}_1\ell}^{\mathcal{S}_1\mathcal{S}_2km} \leq \Theta^{\mathcal{S}_1\mathcal{S}_2km}$, label $\mathcal{F}_{1\ell}$ with $\mathcal{S}_2$'s class with respect to $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2m\ell}$;
   iii. If $\rho_{\mathcal{S}_3\ell}^{\mathcal{S}_1\mathcal{S}_2km} > \Theta^{\mathcal{S}_1\mathcal{S}_2km}$, label $\mathcal{F}_{3\ell}$ with $\mathcal{S}_1$'s class with respect to $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2m\ell}$;
   iv. If $\rho_{\mathcal{S}_3\ell}^{\mathcal{S}_1\mathcal{S}_2km} \leq \Theta^{\mathcal{S}_1\mathcal{S}_2km}$, label $\mathcal{F}_{3\ell}$ with $\mathcal{S}_2$'s class with respect to $\mathcal{M}^{\mathcal{S}_1\mathcal{S}_2m\ell}$;
3. perform a one-sided Wilcoxon test on $\left\{\rho_{\mathcal{S}_1\ell}^{\mathcal{S}_1\mathcal{S}_2km}\right\}$ and $\left\{\rho_{\mathcal{S}_3\ell}^{\mathcal{S}_1\mathcal{S}_2km}\right\}$ found above, to test the null hypothesis that the latter has a median greater than or equal to the median of the former, and to find out at what level, if any, this hypothesis is rejected;

(b) $\iota^{\mathcal{S}_1\mathcal{S}_2km} = -1$

1. replace $>$ by $<$ and $\leq$ by $>$ in this item under (a);
2. replace $>$ by $<$ and $\leq$ by $>$ in this item under (a);
3. change the "greater than" to "less than" in the null hypothesis in this item under (a).

(We use the Matlab® built-in `ranksum` for the Wilcoxon test.)

5.2 Training — diversity, thick tails

Here, we examine the diversity of the library members selected in the training phase, and also whether the distribution of the activations has a thicker tail for the extraction subject (as expected) or for the contrast subject.

As can be seen from Table 5.2, we only have four training suites to examine here: (PD1, RH, C2, RH), (PD1, LH, C2, LH), (C2, RH, PD1, RH) and (C2, LH, PD1, LH).

5.2.1 Diversity

In Figure 5.1 we display the distances between the ten selected members[2] from the libraries extracted from the first ply of the right-hand PD1 data, using right-hand C2 data as a contrast set, where the extraction was under our three sets of conditions — "plain" (only ue and zi), ec and ed. The underlying polynomial for the plain results was of degree 4, that for ec, 6 and that for ed, 8, so there are $p = 2$ degrees of freedom in each case ($n + 1$ parameters for an nth degree polynomial, minus 2 conditions for the plain condition set — ue and zi – 2 more for ec and a further 2 for ed).

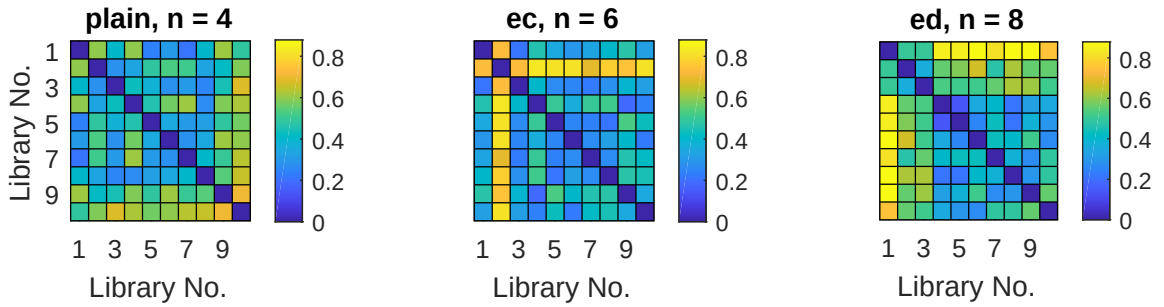


Figure 5.1: *Example distance squares, for the selected members of the library $\mathcal{L}$ extracted from the first ply of the right-hand PD1 data, using right-hand C2 data as a contrast set*

As the colour scale is the same for each diagram, the yellower the squares, the more widely separated are the corresponding library members.

We first note that the maximum distance between members is $\sqrt{2}$, as they can be considered as equivalence classes of unit vectors, and the distance between these classes is analogous to $\min\{\|\mathbf{v}_1 - \mathbf{v}_2\|, \|\mathbf{v}_1 + \mathbf{v}_2\|\}$ rather than simply the Euclidean distance $\|\mathbf{v}_1 - \mathbf{v}_2\|$. Hence, there are many pairs with a separation greater than 25% of the maximum, and several with one greater than 50%, shown across our diagrams.

For this ply of the (PD1, RH, C2, RH) training suite, at least, it seems that the imposition of stronger condition sets on the ECWTs does not force the selected members to be closer together, if the degree of freedom is the same. In fact, the ec diagram shows that one member is strongly separated from the others, and the ed diagram has one member even more strongly separated from most of the others. Any such features of the plain diagram are certainly less marked.

In Figure 5.2, we show the diagram corresponding to Figure 5.1, but for ten random members of $\mathcal{L}$ for the first ply. The colour scale is the same as Figure 5.1.

[2]It will be recalled that ten members will *usually* be selected. In this case, 11 members are selected under the ed condition set, but we only display the square for the first ten

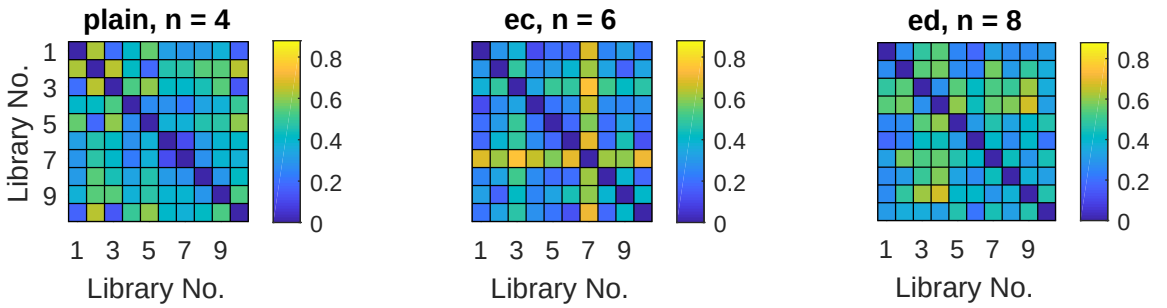


Figure 5.2: *Example distance squares, for random members of the library $\mathcal{L}$ extracted from the first ply of the right-hand PD1 data*

The ec and ed diagrams in Figure 5.2 are clearly bluer than the corresponding diagrams in Figure 5.1, and it is arguable that the same is true of the plain diagrams. In other words, there is evidence that the selection procedure does *not* favour less distant members, in this particular instance, anyway.

In order to see if these observations hold more generally, we assemble sets consisting of all the *intra*-library distances for the (distinct) selected members for all the plies, for each training suite. We repeat this for each available degree of the underlying polynomials for the ECWTs. Statistics relating to these distances are displayed in the box plots of Figure 5.3.

Note that:

1. it is possible for (lower quartile - interquartile range[3]) to be negative for a distribution of positive values — we set the lower limit of the vertical axis to be 0, cutting off some of the "whiskers";

2. in some cases the notch is very narrow — this will happen when there are a lot of values quite close to the median value.

We also do the same for the intra-library distances between sets of ten randomly chosen members in Figure 5.5.

Across all the diagrams of the Figure for the selected members, Figure 5.3, the lowest median ("waist") appearing is 0.3175, roughly midway between a $\frac{1}{5}$ and a $\frac{1}{4}$ of the maximum possible, and the lowest lower quartile (bottom of the box) is 0.2234, just under a $\frac{1}{6}$ of the maximum, so, even at the lower end, the separations are considerable.

In Figure 5.4, we plot histograms of the distances within 223 randomly generated[4] sets of ten ECWTs (45 distances per set, so roughly 10^4 distances in all).

We can see that

[3] Abbreviated as iqr from here onwards

[4] Details of the generation: for each component of the individual ECWTs, we produce a random $(n+1)$ dimensional vector from the normal distribution $\mathcal{N}(0,1)$ and multiply it by $H^{-\frac{1}{2}}$, producing vectors whose directions are uniformly distributed with respect to the metric H. These vectors are then projected into the space orthogonal to that spanned by the vectors corresponding to the condition sets, and the resulting sets of three vectors are normalised to give an ECWT

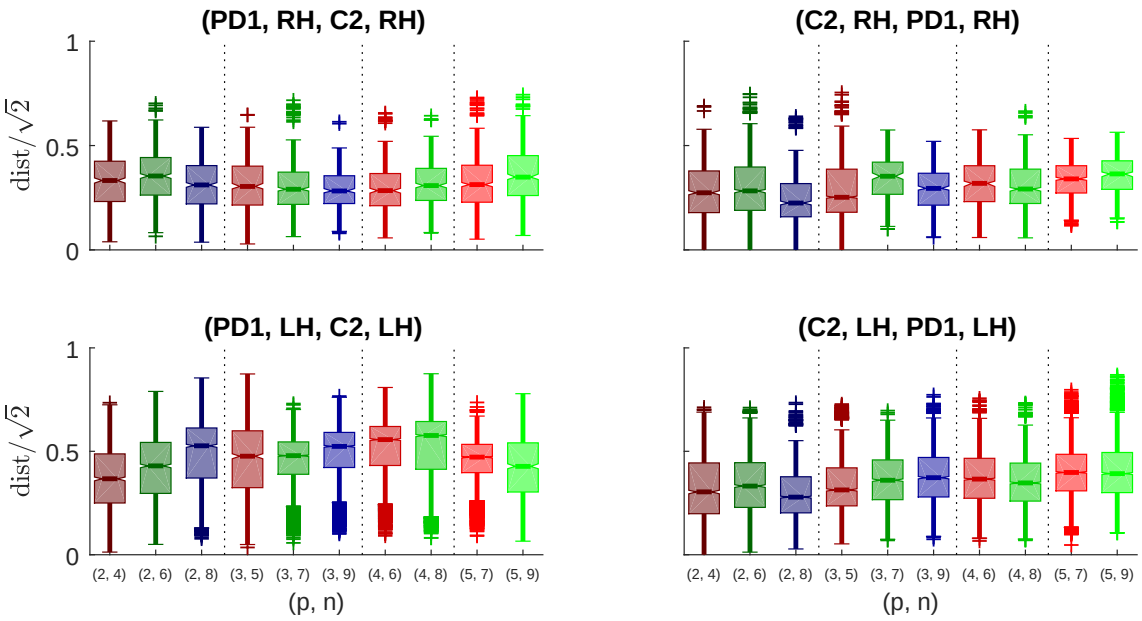


Figure 5.3: *Notched box plots for intra-library distances (selected members). The horizontal line at the "waist" of each box is the median of the intra-library separations, the boxes themselves stretch from the 1st to the 3rd quartile of the separations. The "notch" begins at the lower bound for the 95% confidence interval*[a] *for the median and ends at the upper bound. The vertical lines, or "whiskers" stretch from the lower quartile downwards a distance of one iqr, and from the upper quartile upwards a distance of one iqr. The crosses represent the distances not in the range spanning 1.5 iqrs either side of the median. The red boxes represent the plain results, the green ones the ec ones and the red ones the ed results. These colours become lighter as the degrees of freedom increase*

[a]Actually, there are many such confidence intervals. The particular one shown is symmetric, in the sense that the ratio between the estimated probability that the actual population median lies in the part of the confidence interval above the sample median and the estimated probability that the actual population median lies in the part of the confidence interval below the sample median is as close as possible to 1

1. the empirical distribution of the distances appears to point at an underlying unimodal distribution in each case;
2. the position of the mode varies from about $0.05\sqrt{2}$ to $0.7\sqrt{2}$;
3. the mode moves to the right as p increases;
4. the mode moves to the left as the stringency of the condition set increases.

Thus, it is confirmed that the distances giving rise to Figure 5.3 are still considerable, taking into account that those shown in Figure 5.4 are between ECWTs not constrained to match any data.

There is an individual diagram in Figure 5.3 for each training suite we consider, and, within each of these diagrams, the sections demarcated by the dotted lines contain box plots for different condition sets, but for the same degree of freedom, p.

In very few of the sections with all three condition sets, i.e., the two left sections, does the sequence determined by increasing median correspond to the sequence of condition sets (plain,

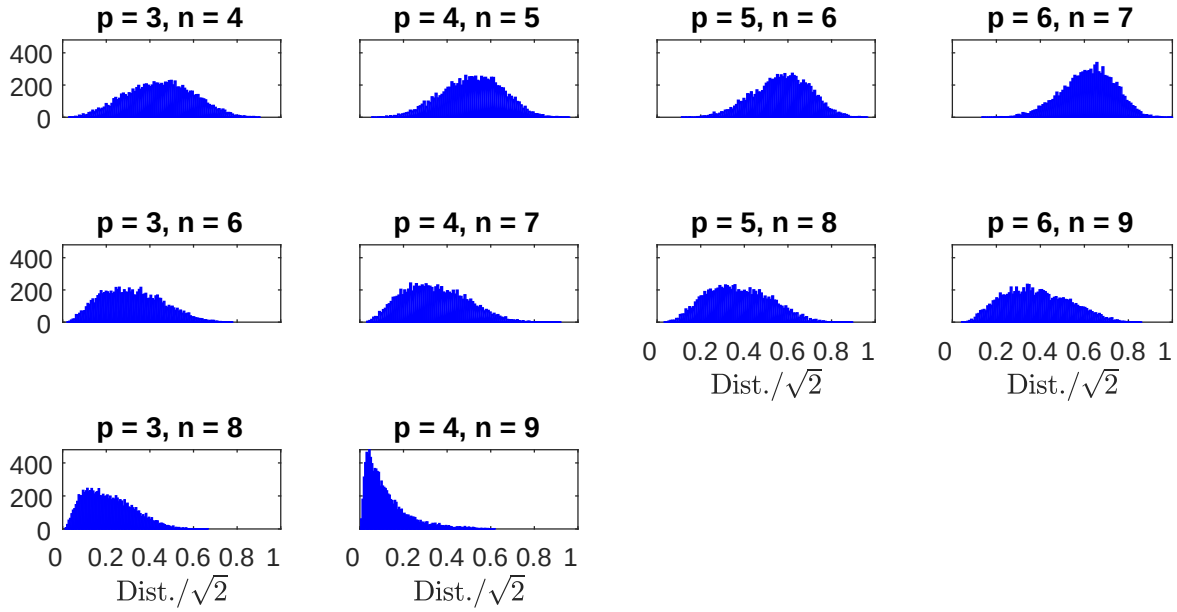


Figure 5.4: *Histograms for randomly generated ECWTs (the top row is for the plain condition set, the middle one for ec, the bottom for ed)*

ec, ed). Even when there are just two conditions for constant p, the sequence corresponding to increasing medians is (plain, ec) less often than the reverse, so there is little evidence that increasing the severity of the conditions on the libraries pushes the selected members together. Moreover, the differences between medians are quite small within the results for each training set, both absolutely and relative to the iqr. Nevertheless, as the (projections of) notches (onto the vertical axis) rarely overlap, these differences are *statistically* significant in most cases.

Making a comparison of the green box from the first section and the red from the third ($n = 6$), the green box from the second section and the red from the fourth ($n = 7$), the blue box from the first section and the green from the third ($n = 8$), and the blue box from the second section and the green from the fourth ($n = 9$), for each given training suite, we see that the evidence that increasing the rigour of the condition set automatically pushes the selected members together is also very weak when only n, the degree of the underlying polynomials, is held fixed.

Making comparisons between training suites, i.e., across, rather than within the diagrams of Figure 5.3, we see that there *are* systematic differences for the separations for different training suites — for instance, the range of the medians for (PD1, LH, C2, LH) is $0.3672\sqrt{2}$–$0.5765\sqrt{2}$, which is entirely above the corresponding ranges for (PD1, RH, C2, RH), $0.2824\sqrt{2}$–$0.3542\sqrt{2}$ and (C2, RH, PD1, RH), $0.2245\sqrt{2}$–$0.3641\sqrt{2}$, and mostly above that for (C2, LH, PD1, LH), $0.2785\sqrt{2}$–$0.3536\sqrt{2}$. In terms of decreasing median of the medians, the training suites have the following order: (PD1, LH, C2, LH) ($0.4777\sqrt{2}$); (C2, LH,

PD1, LH), $(0.3536\sqrt{2})$; (PD1, RH, C2, RH) $(0.3097\sqrt{2})$; (C2, RH, PD1, RH), $(0.2932\sqrt{2})$; so, the separation of the selected members seems to be greater for the left-hand data, and, within that ordering, for the libraries generated using the PD1 data.

We now look at, in Figure 5.5, the distribution of the distance for random, rather than selected, members of the libraries.

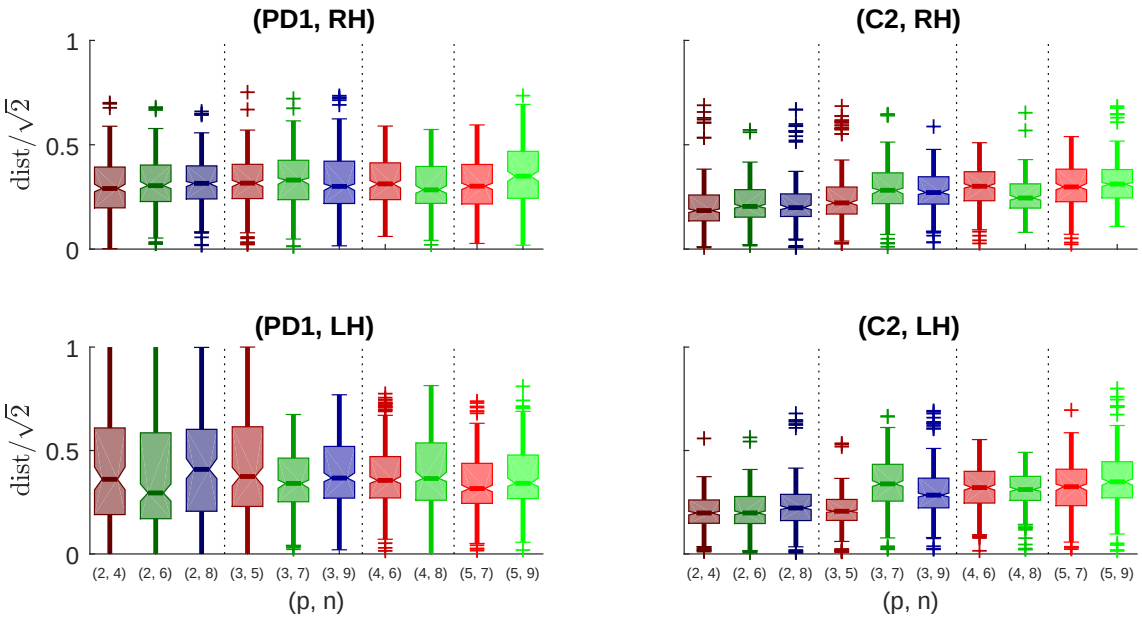


Figure 5.5: *Box plots for intra-library distances (random members). As the second subject-hand pair of the training suite designation plays a role only in member selection, we drop it from the labelling of the plots*

The major systematic differences with Figure 5.3 are that the notches are broader, and corresponding medians are usually lower. The first difference means that the concentration of the most central part of the distribution of distances for the randomly chosen members is lower. However, the iqr is, if anything, smaller, except for the (PD1, LH) libraries, so there is a slightly greater concentration over a more broadly defined centre of the distribution in the non-exceptional cases. Of course, the (PD1, LH) libraries are less concentrated over the broadly defined centre, too.

In 34 out of 40 cases, the median separation of the selected members is greater than the corresponding median for random members. Moreover, in the sense that the respective notches do not overlap, this difference is significant in 30 of these cases, so we can safely say that the selection process tends to increase the median distance, by selecting some of the more widely separated members.

As in Figure 5.3, there is no evidence for a decrease in intra-library distances with stricter conditions, when p or even n is held constant.

5.2.2 Thick tails

As we extract libraries from one set of data and select members on the basis of the difference between the distributions of the activations with respect to these members of windows in similar and supposedly different data, where the difference is measured in terms of the relative thickness of the tails of these distributions, we would expect that the tails of the distributions of the activations related to the first set would be thicker. As we shall see, this is often *not* the case. When it is not the case, we say that the tails are *reversed*.

(PD1, RH, C2, RH)

When we extract $\mathcal{L}$ from the (PD1, RH) data and use the (C2, RH) as a contrast set to select members of $\mathcal{L}$, the situation is simple — no tails reversed for any of the selected members for any combination of n and the condition sets we have examined.

(PD1, LH, C2, LH)

When we extract $\mathcal{L}$ from the (PD1, LH) data and use the (C2, LH) as a contrast, the situation is very different, as shown by Figures 5.6 (plain condition set), 5.7 (ec) and 5.8 (ed). Apart from half the libraries selected without using the ply pair 9, all of the tails are reversed.

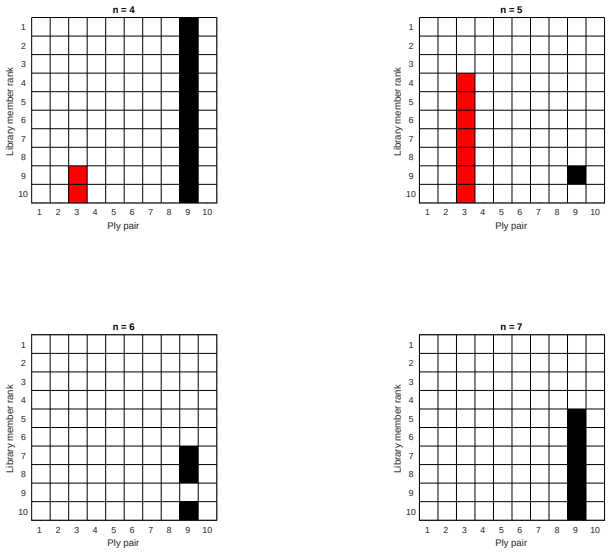


Figure 5.6: *Fat tails diagram for (PD1, LH, C2, LH), with the plain condition set. The white squares correspond to reversed tails, the black ones to the expected fatness of the tails and the red ones to missing libraries*

A possible explanation for the generally higher relative number of activations of windows in the (C2, LH) data with respected to the selected libraries extracted from (PD1, LH) data

than is the case with unseen[5] (PD1, LH) is that PD1's left hand is his least affected one. If at this stage of development of PD, the motion of his left hand is characterised primarily by attenuation of normal patterns, then the library members extracted might mainly correspond to normal patterns which are more strongly present in C2's data.

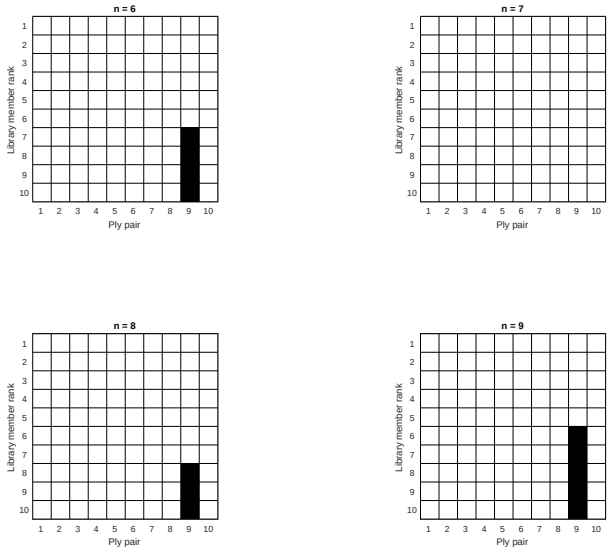


Figure 5.7: *Fat tails diagram for (PD1, LH, C2, LH), with the ec condition set*

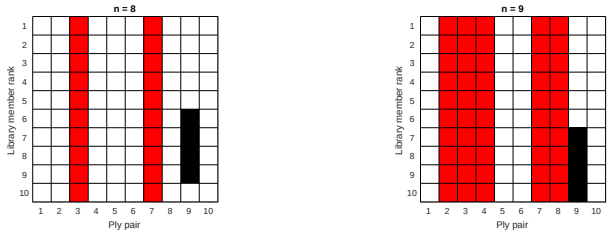


Figure 5.8: *Fat tails diagram for (PD1, LH, C2, LH), with the ed condition set*

(C2, RH, PD1, RH)

The results here are simple: all tails reversed for any combination of n and the condition sets.

This result is slightly problematic, as it seems to imply that motion patterns present in the (C2, RH) data and picked up by our method, are present to an exaggerated extent in the (PD1, RH) data. However, this is certainly a possibility.

[5]Before the selection stage, naturally

(C2, LH, PD1, LH)

Here, the results for the plain condition set, $n = 4$–7 are given in Figure 5.9, and we can see that the tails are reversed in just one case. The results for the ec ($n = 6$–9) and the ed ($n = 8$–9) condition sets are that no tails are reversed (a result too simple to need a picture).

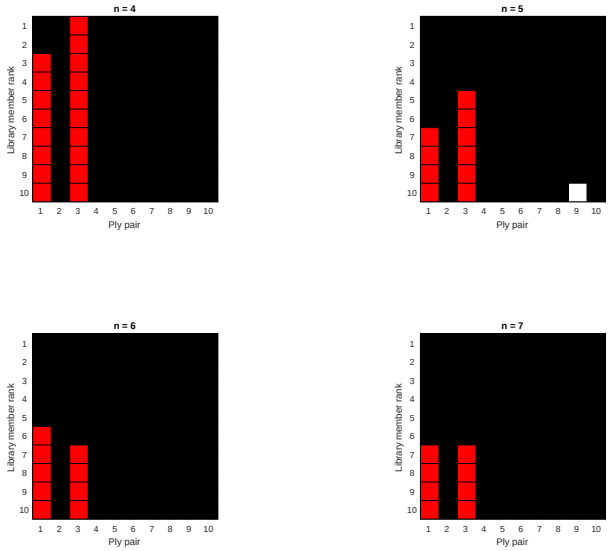


Figure 5.9: *Fat tails diagram for (C2, LH, PD1, LH), with the plain condition set*

Not only is this result what we broadly expect, it also does not contradict the idea used to explain the results above on page 202, under the heading (PD1, LH, C2, LH), that attenuation of normal patterns of movement may be a feature of the PD1 data for the left-hand attachment.

5.2.3 Summary

We have found that the median intra-library separation between our selected ECWTs is a substantial fraction of the maximum possible, and compares well with the median separations of groups of ten randomly chosen library members, and those of groups of ten randomly selected members of the entire set of ECWTs. Moreover, the median separation of the selected ECWTs is not greatly or systematically affected by the imposition of our condition sets.

Contrary to expectation, our tails are frequently reversed, with the tail of the contrast set activation distribution being thicker than that of the extraction set in a substantial number of cases.

But if

1. the PD1 left-hand data contains attenuated, but normal, patterns of motion present in the C2 left-hand data;

2. the C2 right-hand data contains patterns of motion present to an exaggerated extent in the PD1 right-hand data

this reversal of the tails can be explained.

5.3 Recognition

In this section, we first examine the ability of our methods to correctly classify data stemming from previously unseen PD1 and C2 data, when the libraries have been extracted from PD1 data and C2 data has been used as a contrast.

We then do the same, but with libraries extracted from C2 data and using PD1 data as contrast.

In both cases, we first look at right-hand and then left-hand data.

5.3.1 Test suite (PD1, RH, C2, RH, PD1, RH, C2, RH)

Plain

We first examine the results under the plain condition set, for $n = 4$–7.

In Table 5.3, we show the classifications of the PD1 and C2 right-hand plies with respect to the library members ranked one to ten (it should be emphasised that, although the plies are constant down the columns, it is only the rank of the library members with respect to the plies that is constant across the rows — even the libraries from which these members come are different for different partitions).

	ply																			
Mem. rank	**1**		**2**		**3**		**4**		**5**		**6**		**7**		**8**		**9**		**10**	
	P	**C**	**P**	**C**	**P**	**C**	**P**	**C**	**P**	**C**	**P**	**C**	**P**	**C**	**P**	**C**	**P**	**C**	**P**	**C**
1	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	P	C	C	P	C
2	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	P	P	C	P	C
3	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	P	C	C	P	C
4	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	P	C	C	P	C
5	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	P	C	C	P	C
6	P	C	P	C	P	C	P	C	C	P	P	C	P	C	P	P	C	C	P	C
7	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	P	C	C	P	C
8	P	C	P	C	P	C	P	C	C	C	P	C	P	C	P	P	C	C	P	C
9	P	C	P	C	P	C	P	C	P	C	P	C	P	C	P	P	C	C	P	C
10	P	C	P	C	P	C	P	C	C	C	P	C	P	C	P	P	P	C	P	C

Table 5.3: *Classification of the test data for each partition (i.e., the data* not *in the training data for that partition): $n = 4$*

The information of Table 5.3 can be represented more compactly graphically, and this graphical representation is given in Figure 5.10.

We can see from Figure 5.10 that PC1 and C2 plies 1-4, 7, 8 and 10 are classified perfectly, for $n = 4$–7, and that the number of (ply pair number, library member) pairs where the PC1

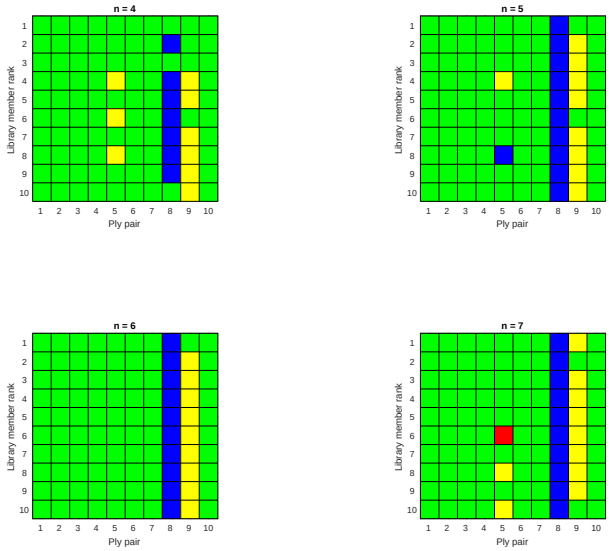


Figure 5.10: *Graphical version of Table 5.3, but with* $n = 5, 6$ *and 7 as well as 4. Green squares indicate* $(P, C) \to (P, C)$*, or that Parkinson's data is classified as Parkinson's data and control data is classified as control data, blue squares indicate* $(P, C) \to (P, P)$*, or that both Parkinson's data and control data are classified as Parkinson's data, yellow squares indicate* $(P, C) \to (C, C)$ *and red squares indicate* $(P, C) \to (C, P)$*, both with the obvious meanings*

and C2 plies are both classified correctly varies from 79% ($n = 7$) to 84% ($n = 4$), while the number of pairs where both plies are classified incorrectly varies from 0% ($n = 4, 5, 6$) to 1% ($n = 7$).

Both plies in the pairs 1–4, 6, 7 and 10 are always classified correctly, but there is some difficulty in classifying the pairs 8 and 9, and, to a lesser extent 5. The difficulty is in classifying the PD1 ply for pair 8, and the C2 ply for pairs 5 and 9 (except when $n = 7$, where neither ply of pair 8 is classified correctly by using the activations of the 6th ranked library for that ply pair).

In Table 5.4 we present the p values for the Wilcoxon for the hypotheses mentioned above. Although there is some interest in the details of p, its major value is in deciding hypotheses at standard levels, and it is easier to see the p-values of several hypotheses at once *via* a diagram like Figure 5.11.

From Figure 5.11, we can see that the null hypotheses can mostly be rejected, and quite strongly, for plies 1-3, 6, 7 and 10, and also sporadically elsewhere. In other words, we can be quite certain that the median proportion of activations exceeding θ is greater for PD1 data than for C2 in the majority of cases.

We would expect that the grey squares of Figure 5.11 to correspond to green squares in Figure 5.10, and this is mostly the case. The only real exception is column 4, corresponding to a ply pair which is correctly classified in all cases, but which has higher p-values, indicating

Mem. rank	plies				
	1	**2**	**3**	**4**	**5**
1	***5.12×10⁻⁴***	***1.64×10⁻⁴***	***1.19×10⁻⁵***	*0.065*	*0.091*
2	***1.78×10⁻³***	***1.36×10⁻⁴***	***3.69×10⁻⁵***	0.111	*0.085*
3	***6.47×10⁻⁴***	***2.61×10⁻⁴***	***1.05×10⁻⁵***	0.145	*0.070*
4	**1.23×10⁻²**	***2.04×10⁻⁴***	***2.58×10⁻⁵***	0.355	0.380
5	***9.93×10⁻⁴***	***2.10×10⁻⁴***	***8.97×10⁻⁵***	0.230	0.437
6	***2.75×10⁻³***	***4.68×10⁻⁴***	***1.07×10⁻⁴***	0.361	0.487
7	***1.99×10⁻³***	***2.70×10⁻⁴***	***1.27×10⁻⁴***	0.307	0.528
8	***2.75×10⁻³***	***2.41×10⁻⁴***	***1.27×10⁻⁴***	0.366	0.545
9	**1.04×10⁻²**	***4.10×10⁻⁴***	***1.19×10⁻⁵***	0.145	0.504
10	**1.41×10⁻²**	***4.01×10⁻⁴***	***1.05×10⁻⁵***	0.307	0.583
Mem. rank	**plies**				
	6	**7**	**8**	**9**	**10**
1	***1.82×10⁻³***	***3.75×10⁻⁵***	***3.07×10⁻³***	***7.22×10⁻³***	***5.01×10⁻⁵***
2	***5.02×10⁻³***	***5.72×10⁻⁵***	**1.16×10⁻²**	**1.53×10⁻²**	***1.25×10⁻⁴***
3	***4.80×10⁻³***	***6.55×10⁻⁴***	***7.65×10⁻³***	***9.49×10⁻³***	***1.68×10⁻⁴***
4	***7.66×10⁻³***	***1.52×10⁻⁴***	**1.75×10⁻²**	**2.68×10⁻²**	***1.83×10⁻⁴***
5	**2.27×10⁻²**	***3.31×10⁻⁴***	**3.38×10⁻²**	**1.39×10⁻²**	***3.78×10⁻⁴***
6	**1.28×10⁻²**	***1.23×10⁻⁴***	*0.059*	***9.15×10⁻³***	***2.28×10⁻⁴***
7	***9.29×10⁻³***	***2.58×10⁻⁴***	**3.16×10⁻²**	**3.72×10⁻²**	***2.46×10⁻⁴***
8	***8.43×10⁻³***	***3.62×10⁻⁴***	0.104	0.109	***5.27×10⁻⁴***
9	***5.02×10⁻³***	***1.38×10⁻⁴***	**1.16×10⁻²**	0.106	***7.26×10⁻⁴***
10	***1.82×10⁻³***	***5.72×10⁻⁵***	***3.07×10⁻³***	0.106	***1.54×10⁻³***

Table 5.4: *p-values of the one-sided Wilcoxon of the hypothesis that the median fraction of the activations of epochs of Parkinson's data exceeding θ is less than or equal to the median of such a fraction calculated on epochs of control data. Here, $n = 4$, the bold italic font indicates that the hypothesis is rejected at the 1% level, a bold roman font rejection at 5% and non-bold italic one at 10%. A plain font indicates that the hypothesis is not rejected at these levels.*

that the relative disposition of the $\left\{\rho_{\mathcal{S}_1\ell}^{\mathcal{S}_1\mathcal{S}_2 km}\right\}$ and $\left\{\rho_{\mathcal{S}_3\ell}^{\mathcal{S}_1\mathcal{S}_2 km}\right\}$ is not *very* unlikely if the null hypothesis holds.

We note that

1. "not very unlikely" is not the same as "very likely": from Table 5.4, we can see that the most likely disposition of the $\left\{\rho_{\mathcal{S}_1\ell}^{\mathcal{S}_1\mathcal{S}_2 4m}\right\}$ and $\left\{\rho_{\mathcal{S}_3\ell}^{\mathcal{S}_1\mathcal{S}_2 4m}\right\}$, $m = 1, 2, \ldots, 10$ is that of $\left\{\rho_{\mathcal{S}_1\ell}^{\mathcal{S}_1\mathcal{S}_2 48}\right\}$ and $\left\{\rho_{\mathcal{S}_3\ell}^{\mathcal{S}_1\mathcal{S}_2 48}\right\}$, and the probability of this disposition (or worse, from the point of view of the null hypothesis) is 0.361;

2. the tests represented by a column of a Figure like 5.11 are *not* independent — at the very least, the quantities we use are generated from the same test data, and, even if we succeed in picking up different characteristics with different library members, these

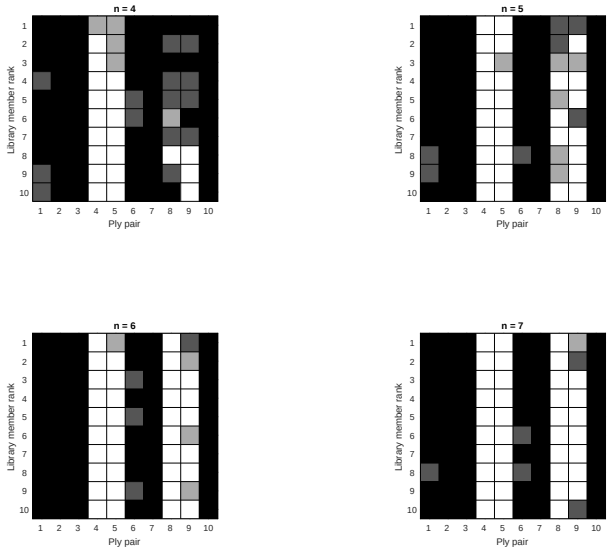


Figure 5.11: *Graphical version of Table 5.4, but with $n = 5, 6, 7$ as well as 4. Black squares indicate that the hypothesis is rejected at the 1% level, dark gray ones at 5% and light gray ones at 10%. White squares indicate that the hypothesis is not rejected at any of those levels*

characteristics are likely to be correlated in time;

3. although we do not have enough evidence to reject the null hypothesis at a desirable significance level, we may not have *collected* enough evidence — we are dealing with the disposition of $\underset{\text{total}}{96} = \underset{\text{PD1}}{52} + \underset{\text{C2}}{44}$ real numbers.

Obviously, as well as using the median of the fractions of tail activations of epochs in a ply in comparison to the appropriate Θ to classify the ply as a whole, the tail activations of an individual epoch can be used to classify the epoch itself. Each of the library members (together with their respective Θs and θs) implies a different classification method, so the number of these classification methods which result in a correct classification can be considered a function of the epoch, and we can show this in the form of a relative-frequency histogram, as in Figure 5.12. If a bar is labelled k, then its height represents the proportion of plies which are correctly classified by exactly k of the relevant selected library members.

The profiles of the bars in Figures 5.12 all have a rough "U" shape, as most epochs are either correctly classified by all the selected library members for its ply, or are incorrectly classified by all of them, with substantially more in the former category than the latter. This result indicates that either the motion patterns picked up by the library members are similar, or that they occur at around the same time. We believe the results on diversity rule out the former explanation.

There is no easily discernible systematic change with n in any of the Figures 5.10-5.12.

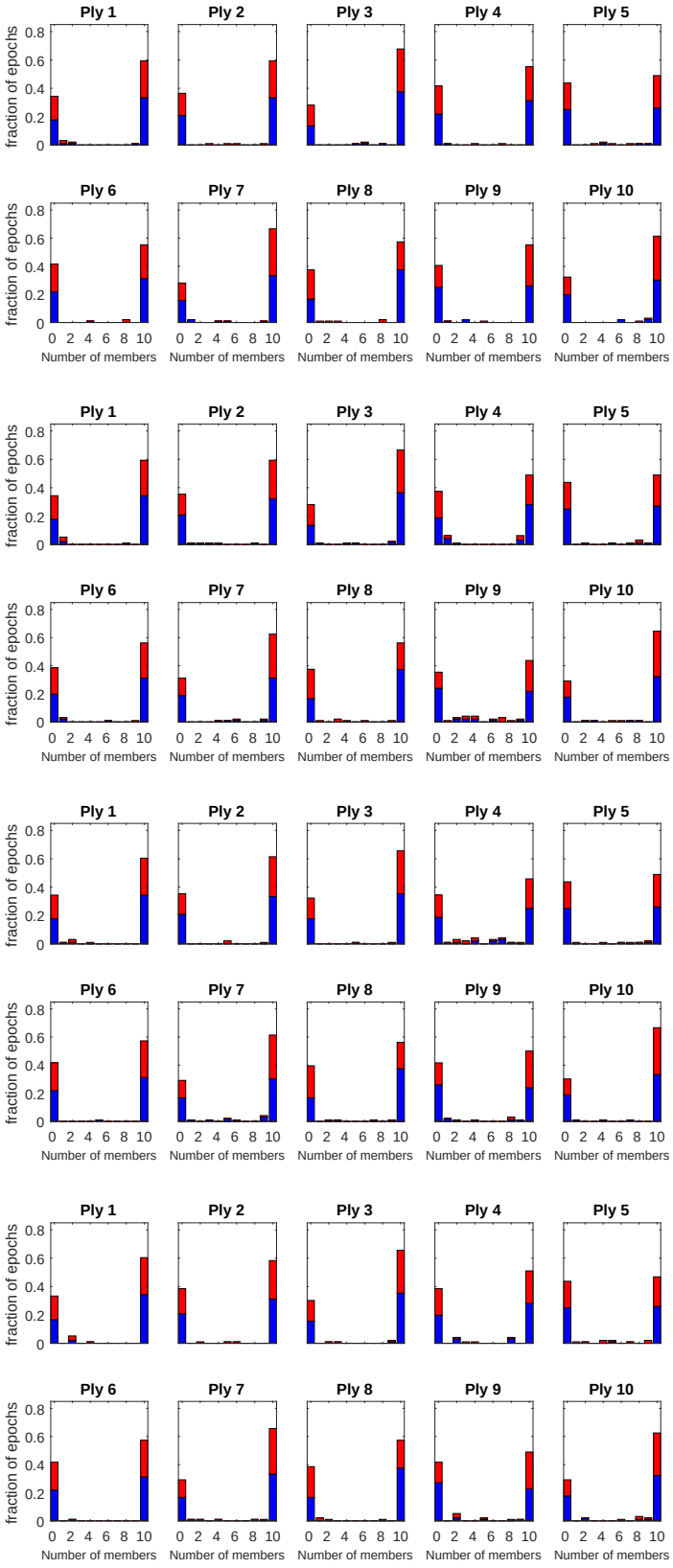


Figure 5.12: *Epoch-level classification diagram for (PD1, RH, C2, RH, PD1, RH, C2, RH), with the plain condition set. The portion of each bar coloured blue represents epochs in the PD1 plies, that coloured red those in the C2 plies ($n = 4$–7, top to bottom)*

ec

In Figures 5.13 to 5.15 we present the ply classification, Wilcoxon and epoch classification diagrams analogous to Figures 5.10 to 5.12, but for the ec condition set.

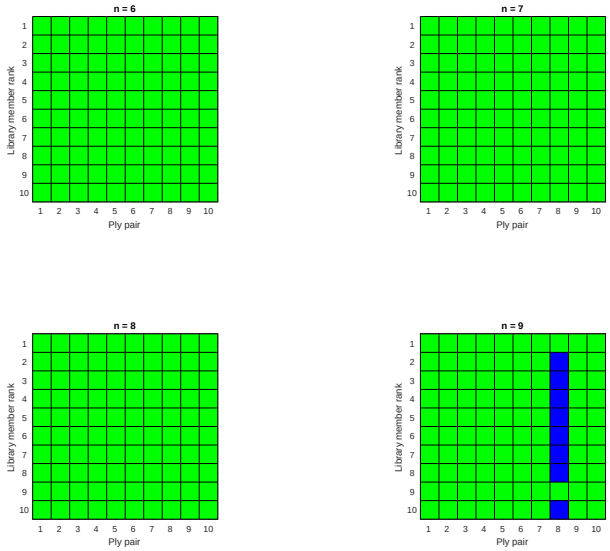


Figure 5.13: *Diagram like Figure 5.10, but for (PD1, RH, C2, RH, PD1, RH, C2, RH), with ec condition set*

The ply classification (Figure 5.13) is perfect for $n = 4$–6, and only fails for 8 of the 10 selected members for the single ply pair 8 for $n = 7$, and only on the C2 member of the pair. The Wilcoxon tests (Figure 5.14) show that the null hypotheses are mostly strongly rejected (i.e., at low p-levels) where the classification is correct, with the only slight exception being the squares for ply 5 and a few of those for ply 4. The "difficult" ply for $n = 7$, ply 8, corresponds with the null hypothesis being "only" rejected at a 5% level for 6 of the 8 selected members where the classification fails on the C2 ply (and being rejected at the same level for 1 of the 2 where classification was correct).

The epoch-level classification (Figure 5.15) again shows the rough "U" distribution.

There does appear to be some dependence on n in this case, with performance being roughly the same for $n = 4$–6, but falling off slightly for $n = 7$. This is visible in both Figures 5.13 (perfect classification everywhere for $n = 4$–6, classification failure on ply 8 (C2) in 8 of 10 cases for $n = 7$) and 5.14 ($n = 4$-6: rejection of null hypothesis at the 1% level for 86-88% of the squares in the diagrams, failure to reject at any of the given levels for 0-1%; $n = 7$: rejection of null hypothesis at the 1% level for 78% of the squares in the diagrams, failure to reject at any of the given levels for 5%).

When we compare the classification and Wilcoxon results with the ec condition set to those with the more relaxed plain condition set, we see a clear improvement with the ec condition set, both for constant degrees of freedom (diagrams in the same position in Figures 5.10

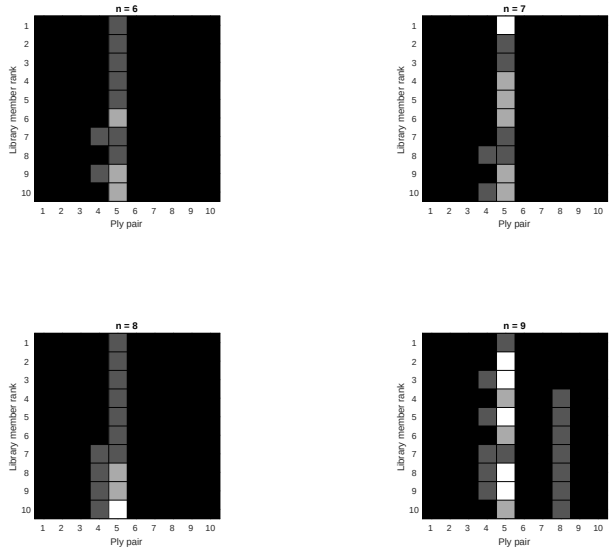


Figure 5.14: *Diagram like Figure 5.11, but for (PD1, RH, C2, RH, PD1, RH, C2, RH), with ec condition set*

and 5.13, and in Figures 5.11 and 5.14), and for constant n (diagrams in the bottom row of the plain Figures and top row of the ec Figures).

The epoch classification diagrams, Figures 5.12 and 5.15, for the plain and ec condition set, appear to be qualitatively the same.

ed

Figures 5.16 to 5.18 display our results when the ed condition set is imposed.

We first note that there are missing library selections for certain ply pairs (white squares in the ply-classification Figure, red squares in the Wilcoxon Figure and missing plots in the epoch-classification Figure). Indeed, there are no selections for plies 4 and 5 when $n = 8$, and none for plies 1 and 3 when $n = 9$. This due to our algorithm being unable to detect a difference in the tails of the relevant activation distributions over the PD1 and C2 contrast sets.

However, where library members *have* been selected, the ply classification is perfect (Figure 5.16), and the hypotheses of equal or "wrongly-ordered" medians are all, mostly strongly, rejected (Figure 5.17).

The epoch-level classification diagrams (Figure 5.18) are once again "U"-shaped.

In this case, we might co-opt the library members selected for $n = 8$ for plies 1 and 3 to serve alongside the library members selected for $n = 9$ for plies 2 and 4–10. Although the general 9th degree polynomial cannot be considered an 8th degree polynomial, this is no reason, in principle, why the reverse cannot be done, using members selected for $n = 9$ to cover the gaps in the members for $n = 8$.

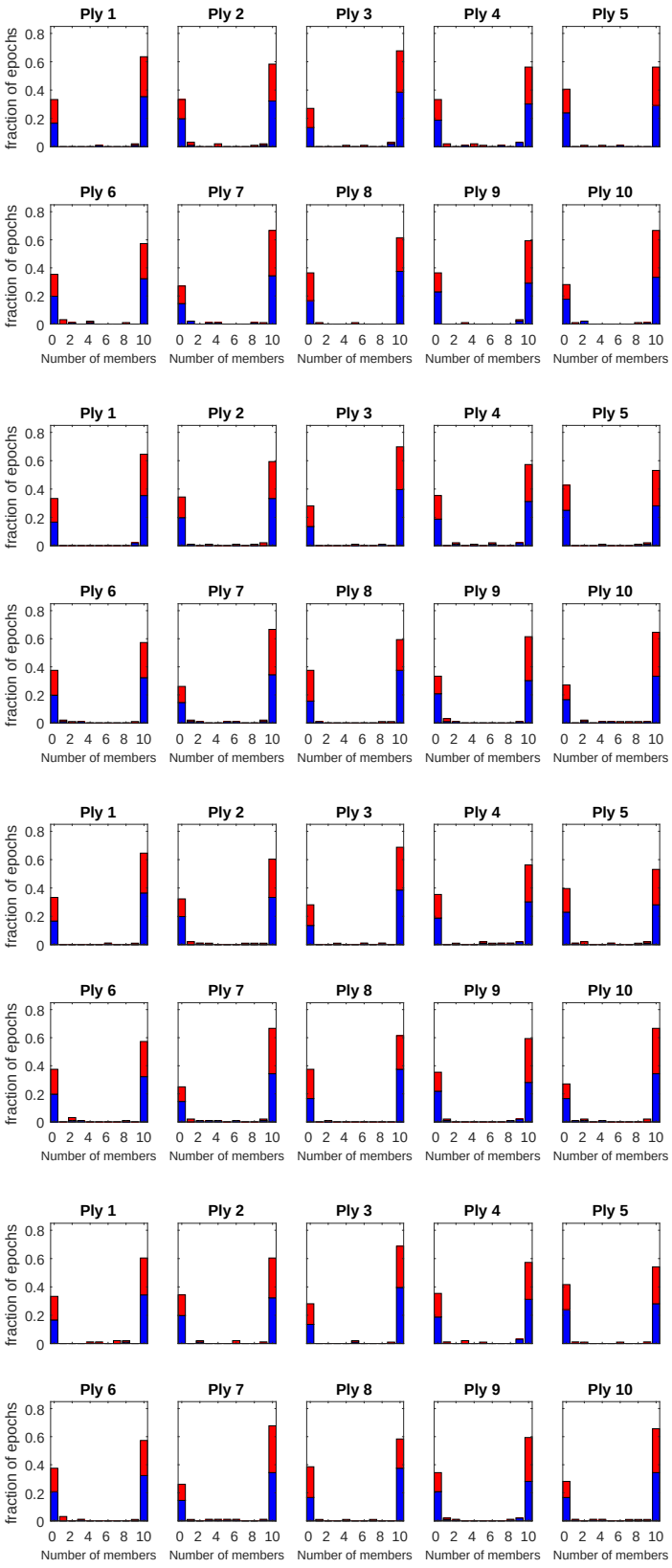


Figure 5.15: *Diagram like Figure 5.12, but for (PD1, RH, C2, RH, PD1, RH, C2, RH), with ec condition set ($n = 6$–9, top to bottom)*

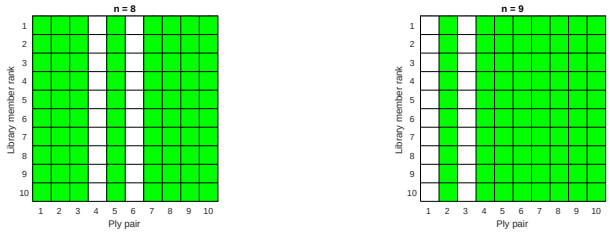


Figure 5.16: *Diagram like Figure 5.10, but for (PD1, RH, C2, RH, PD1, RH, C2, RH), with ed condition set. The white squares indicate that the corresponding library member is missing for the corresponding ply pair*

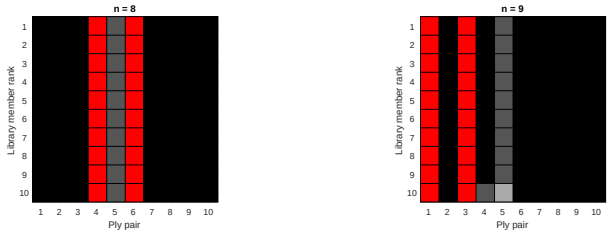


Figure 5.17: *Diagram like Figure 5.11, but for (PD1, RH, C2, RH, PD1, RH, C2, RH), with ed condition set (red squares for missing library members)*

However, we are unlikely to always have this option, as the gaps for one value of n need not be covered by the members selected for another.

If we now compare the ply-level classification under the ec and ed condition sets, Figures 5.13 and 5.16, we can only claim that the ed results are better than the ec ones when $n = 9$, for the 8th ply pair; but then it is difficult to improve on perfection!

Nor can we claim the missing selections as an improvement.

It is a similar story for the Wilcoxon tests in Figures 5.14 and 5.17: we can only claim that the ed results are better than the ec on a handful of squares, when we exclude those that are red in the ed Figure.

Apart from the fact that it only shows two values of n, the epoch-level classification Figure 5.17 is little different from Figure 5.14, the corresponding Figure for the ec condition set.

Summary

The method appears to have a good ability to distinguish right-hand PD1 plies from C2 ones, and this improves with the imposition of the ec condition set. The further imposition of the ed condition set means that the method fails to produce a member selection in 20% of

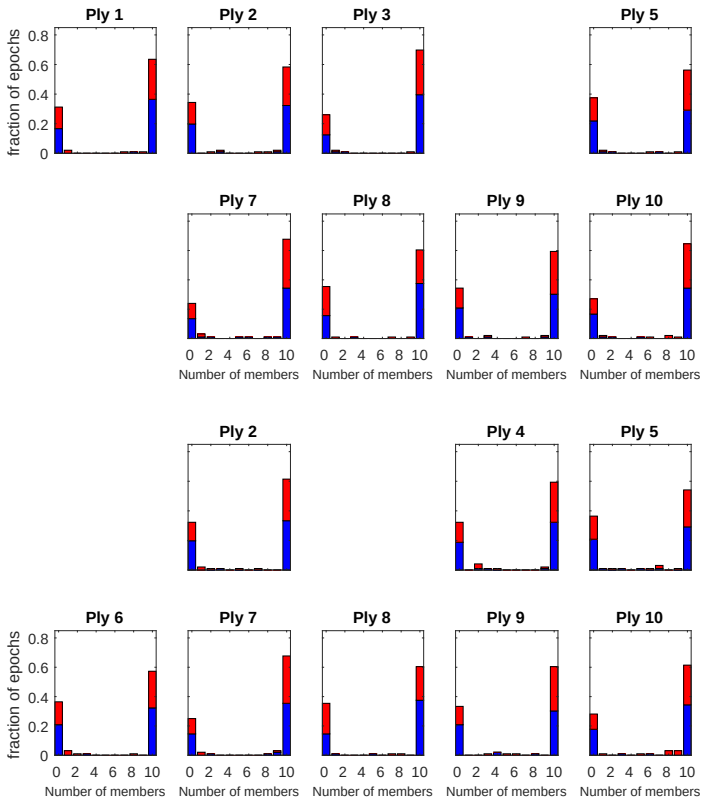


Figure 5.18: *Diagram like Figure 5.12, but for (PD1, RH, C2, RH, PD1, RH, C2, RH), with ed condition set ($n = 8$–9, top to bottom)*

cases, and cannot simply be used to yield an improvement. Combining results from running the algorithms under the ed conditions could enable a full suite of selected members to be produced, but the improvement on the ec results is ambiguous and anyway minor.

		Difficulty			
Cond. Set	n	**PD1** (yellow)	**C2** (blue)	**Both** (red)	**Not selected** (white)
plain	4	5, 9	8	—	—
	5	5, 9	8	—	—
	6	9	8	—	—
	7	5, 9	8	5	—
ec	9	—	8	—	—
ed	8	—	—	—	4, 6
	9	—	—	—	1, 3

Table 5.5: *Plies which pose particular problems for the method (colours refer to the ply-level classification diagrams)*

In Table 5.5 we collate the plies which cause difficulties for the method, so that it has a non-green square in its column, either corresponding to at least one misclassification or at at

least one failure to distinguish the tails of the activation distributions. There appears to be a high degree of consistency for the difficult plies (both in which ply pair they are found in, and which of the two plies they are) when the difficulty is in classification, but no evidence for any consistency when the difficulty is in library selection.

The Wilcoxon tests show when the difference in the tails of the activations is significant, and therefore where we should be able to obtain a good performance by a good choice of Θ. In most cases, this coincides with the combination of library member and ply pair where we have actually classified correctly, so, on this data, our policy for the choice of Θ has been vindicated, but there are a few cases where the Wilcoxon appears to show that the difference between the tails is not significant (see the squares for ply 4 in Figures 5.10 and 5.11), yet our classification appears to work. However, it is not necessary for a threshold to work for the differences in the medians to be statistically significant. And it is possible for the differences to be significant over the data used in the selection phase, which is three times larger, and not over the test data.

5.3.2 Test suite (PD1, LH, C2, LH, PD1, LH, C2, LH)

We now look at the recognition of PD1 and C2 left-hand data when our method is used with the training suite (PD1, LH, C2, LH).

Plain

With the minimal plain conditions set, the ply-level classification, Wilcoxon and epoch-level classification Figures for this test suite are 5.19–5.24.

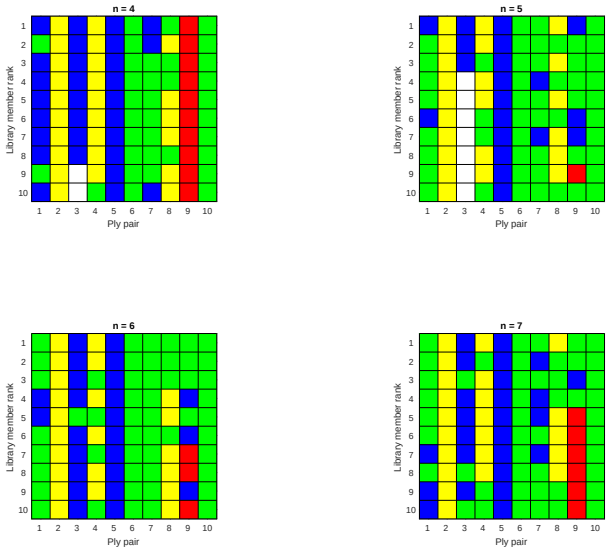


Figure 5.19: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, LH, C2, LH), with the plain condition set*

The classifications in Figure 5.22 are somewhat disappointing, in that the green squares (double classification successes) vary from 34% to 51% of the total. This may given a pessimistic view of the efficacy of classification, as the overall classification success rates can be read off the diagrams by using the formula $\frac{2G+B+Y}{200}$, where G is the number of green squares, B the number of blue and Y the number of yellow ones (we count a failure to obtain a selected member as a double classification failure). These rates are 61%, 71%, 74% and 70.5% for $n = 4, 5, 6$ and 7 respectively, but these are still not very good values.

The Wilcoxon Figure is also disappointing, as few of the null hypotheses we would like to reject can be rejected at any reasonable level. It is only for ply pairs 1, 3 and 10 that there are a good number of grey squares across all values of n considered, indicating that the tails of the activation distributions are sufficiently distinct to have much hope of improving the results by adjusting thresholds, and ply 10 is already one of the two plies where classification is already perfect. When $n \neq 4$, there are more grey squares for ply 9, indicating that there is possible scope for improvement here, too.

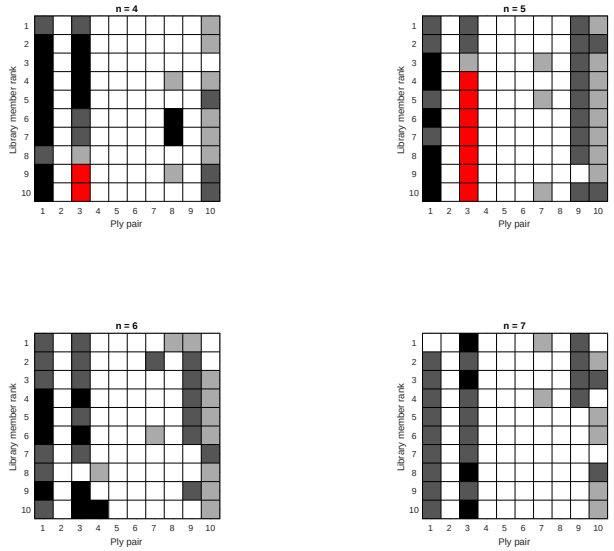


Figure 5.20: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, LH, C2, LH), with the plain condition set*

Obviously, these results are especially disappointing when compared to the similar results for the corresponding test suite for the right-hand attachment.

When we come to Figure 5.24, we see that some of the diagrams show a different pattern to the 'U"-shaped pattern of corresponding diagrams for the right-hand attachment. A much more significant proportion of epochs is correctly classified by other than 0 or 10 of the selected members. This shows that the selected members are classifying different things. We recall that the "U"-shape *does not* show the reverse, as these different things may occur at roughly the same time, i.e., in the same epoch. (However, it should be noted that the distributions of distances between the selected members in a library are displayed in Figure 5.3, and this clearly shows that these distances tend to be greater in the (PD1, LH, C2, LH) case than in the (PD1, RH, C2, RH) one, i.e., the "different things" are more "different" in the (PD1, LH, C2, LH) case than in the (PD1, RH, C2, RH) one.)

In the current case, some of the central bars of the non-"U" histograms are mainly red, which means that some of the motion patterns extracted from the (PD1, LH) training data and chosen for their relative absence from the (C2, LH) training data are also relatively absent from the (C2, LH) test data, whereas other motion patterns extracted and selected in the same way are *not* absent from the (C2, LH) test data.

Nevertheless, most of the larger central bars are more evenly split between red and blue, which means that the above paragraph continues to hold, but, in addition, there are some motion patterns derived from the same training data as above, and in the same way, which are only weakly present in the (PD1, LH) test data.

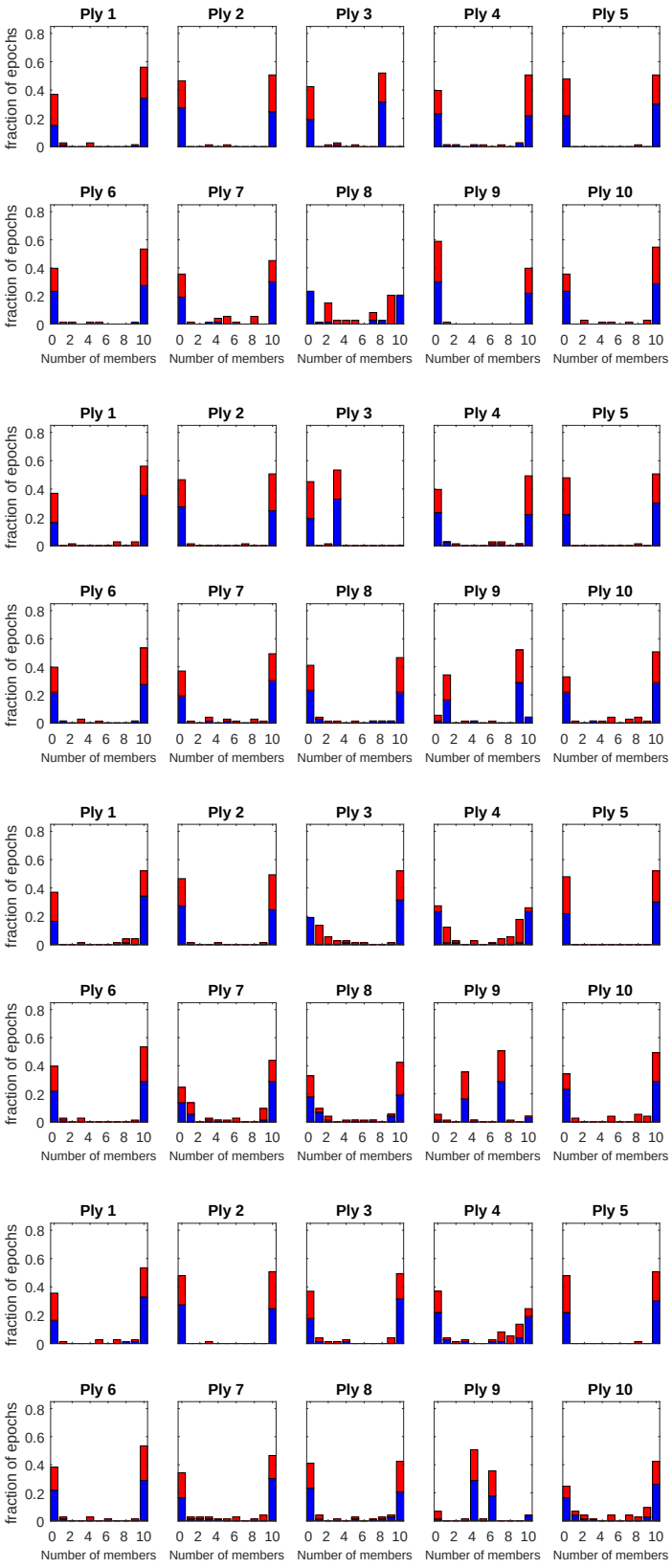


Figure 5.21: *Diagram like Figure 5.12, but for (PD1, LH, C2, LH, PD1, LH, C2, LH), with the plain condition set ($n = 4$–7, top to bottom)*

ec

We now move onto the results for this test suite when the ec condition set is imposed. These are shown in Figure 5.22 to 5.23.

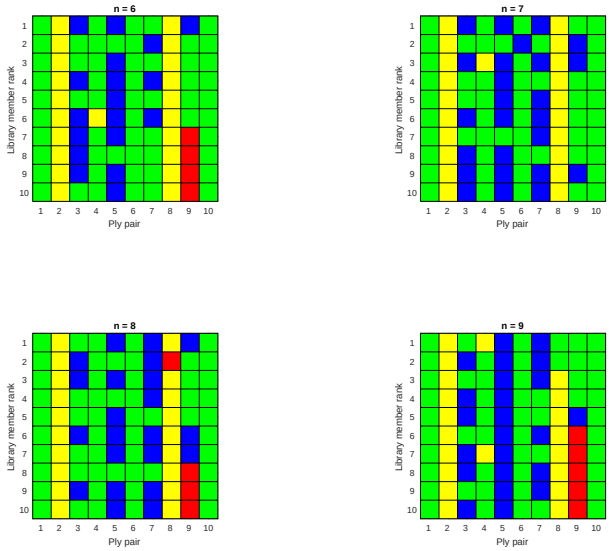


Figure 5.22: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, LH, C2, LH), with ec condition set*

The ply-level classification Figure, 5.22, has no missing selected members, and it has 57% green squares with an overall success rate of 76.5% for $n = 6$, with corresponding values of 54% and 77% for $n = 7$, 55% and 75.5% for $n = 8$ and 52% and 73.5% for $n = 9$.

In the Wilcoxon Figure, 5.23, aside from certain plies, we have a lack of the desirable grey squares, and in the epoch-level classification Figure, 5.24, there is a mixture of "U"-shaped and other profiles, with the former predominating.

If we compare the plain condition set results with the ec condition set results, we find that the percentages of both plies being correctly classified (green squares in the diagrams) increases from 34% to 57% when the degrees of freedom, p, are 3, from 46% to 54% ($p = 4$), from 51% to 55% ($p = 5$) and from 47% to 50%. The same comparison with n fixed is from 51% to 57% ($n = 6$) and from 47% to 54% ($n = 7$).

For the overall correct classification rate, we also find increases in going from the plain condition set to the ec condition set: for constant p, these increases are: 60.5% to 76.5%, 66% to 77%, 74% to 75.5% and 69.5% to 72.5%; for constant n, they are: 74% to 76.5% and 69.5% to 77%.

Although these increases do not result in very good rates, they are substantial.

When we compare the epoch-level classification diagrams for the plain condition set, Figure 5.21 and those for the ec condition set, Figure 5.24, we see that the two diagrams for each p and ply are very similar, so the comments made above, in the plain condition set

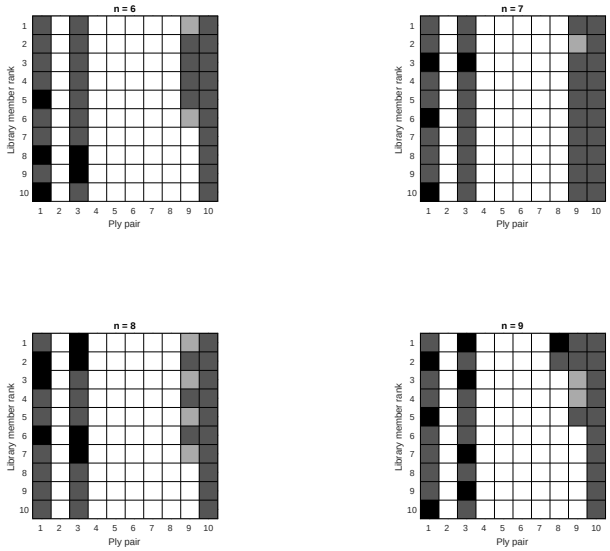


Figure 5.23: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, LH, C2, LH), with ec condition set*

section, about these diagrams apply here, too.

ed

We now look at the processing of left-hand data under the ed condition set: the results for this are summarised in Figures 5.27 to 5.27. We can immediately see that there are no library members selected for ply pairs 3 and 7 when $n = 8$ and for ply pairs 2–4 and 7–8 when $n = 9$.

These missing members are the reason why the percentages of green squares (double successes) in the ply-level classification diagrams of Figure 5.27 are 45% ($n = 8$, $p = 3$) and 34% ($n = 9$, $p = 4$), and the overall classification rates are 60% and 40%, despite the green squares being the majority among the coloured squares, i.e. excluding selection failures, for both values of n.

The Wilcoxon Figure for the ed condition set, 5.26, is similarly marked by selection failures, but, even where library members are selected, the white squares are in the majority (just, for $n = 9$, at least).

The epoch-level classification diagrams of Figure 5.26 again show a mixture of "U" and "non-U" profiles, with central bars either mostly red or evenly split between red and blue, so we can draw the same conclusions as under the plain condition set for this test suite — in particular, the different squares in a ply-pair column of the ply-level classification diagrams represent the present or absence of different things in the data.

If we use the raw double successes and overall classification rates, then clearly introducing the ed condition set in the place of the ec one leads to a reduction in performance. We might, however, take selected members from elsewhere to play the role of the missing selections.

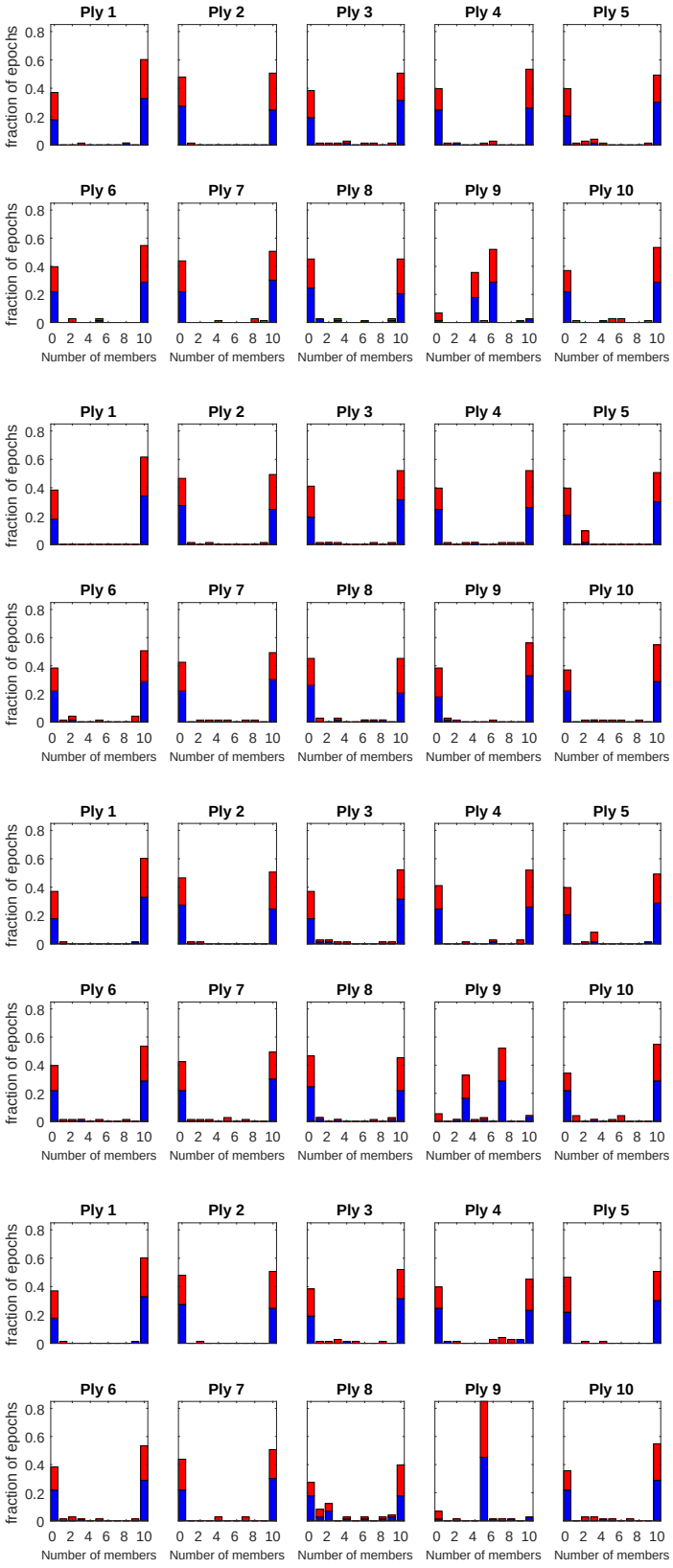


Figure 5.24: *Diagram like Figure 5.12, but for (PD1, LH, C2, LH, PD1, LH, C2, LH), with ec condition set ($n = 6$–9, top to bottom)*

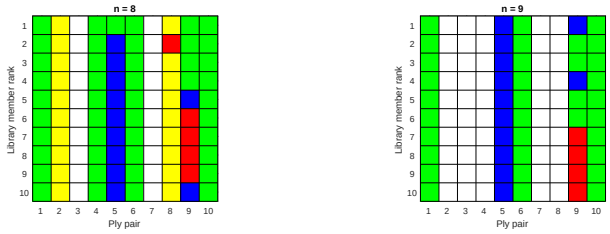


Figure 5.25: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, LH, C2, LH), with ed condition set*

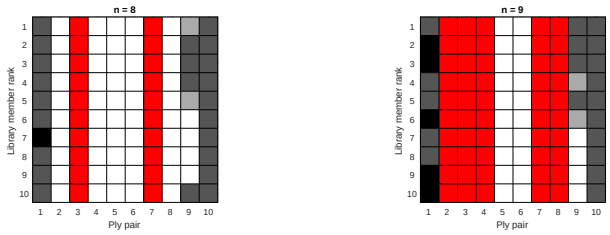


Figure 5.26: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, LH, C2, LH), with ed condition set*

If we use the library members selected for plies 2, 4 and 8 when $n = 8$ alongside those selected for plies 1, 5, 6, 9 and 10 when $n = 9$, we would increase the percentage of double successes to 44%, and the overall classification success rate to 49.5% for the $n = 9$ ply-level classification diagram: these values are slightly below those for the unaltered $n = 8$ diagram.

Alternatively, we might use selected members under the ec condition set to stand in for missing members. If we do this keeping p constant, we will take the members for plies 3 and 7 from the ec results for $n = 6$ to play the roles of the missing members for the same plies for the ed condition set and $n = 8$, and do the analogous thing using ec, $n = 7$ members to play the roles of the missing ed, $n = 9$ selections.

If we do this, the double success/overall success rates become 56%/62.5% for ed $n = 8$ and 50%/56.5% for ed $n = 9$.

We can do the same thing, but keeping n constant between the ed and ec libraries, resulting in 53%/62% ($n = 8$) and 47%/56% ($n = 9$).

Although these results substituting ec library members for missing selections in the ed outcomes are close to those for the pure ec procedure above in terms of the double success rate, they are quite a bit below in terms of the overall success rate.

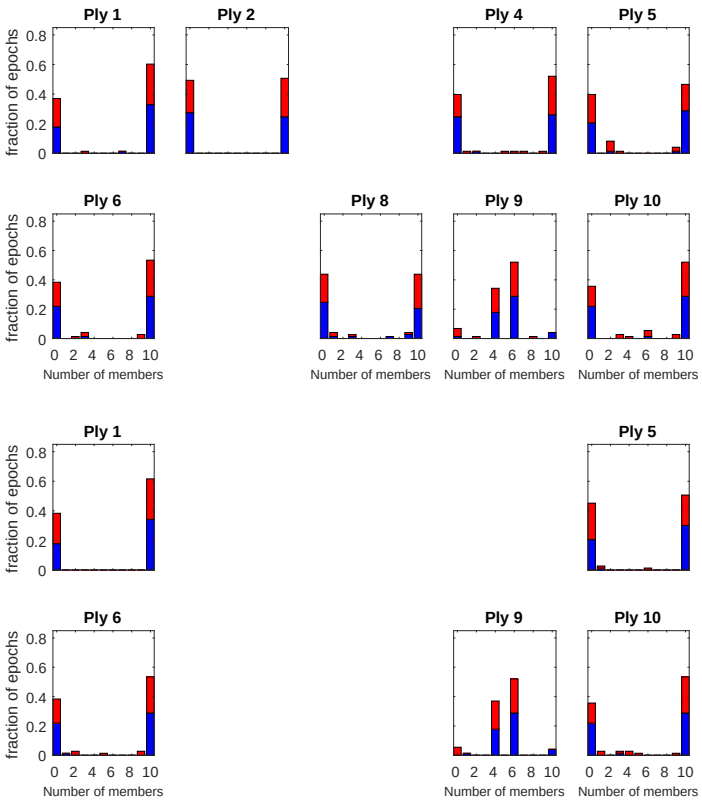


Figure 5.27: *Diagram like Figure 5.12, but for (PD1, LH, C2, LH, PD1, LH, C2, LH), with ed condition set ($n = 8$–9, top to bottom)*

Summary

The most striking thing about the performance of our algorithms as classifiers for this test suite, (PD1, LH, C2, LH, PD1, LH, C2, LH), is that, although it is certainly far better than random, it is not anywhere near as good as it is on the suite (PD1, RH, C2, RH, PD1, RH, C2, RH).

One thing is constant over the two test suites: as shown in Table 5.6, the plies that are difficult for one combination of n and condition set to classify, tend to be difficult for other combinations. For this test suite, however, the stability extends to the difficulty in selecting members, at least as far as concerns ply 3.

We also have evidence that, for at least some plies, the method is picking up characteristics which do not occur together in time, i.e., the non-"U" shape occurring in some of the epoch-level classification diagrams.

		Difficulty			
Cond. Set	n	**PD1** (yellow)	**C2** (blue)	**Both** (red)	**Not selected** (white)
plain	4	2, 4, 8	1, 3, 5, 7	9	3
	5	2, 4, 8	1, 3, 5, 7, 9	9	3
	6	2, 4, 8	1, 3, 5, 9	9	—
	7	2, 4, 8	1, 3, 5, 7, 9	9	—
ec	6	2, 4, 8	3, 5, 7, 9	9	—
	7	2, 4, 8	3, 5, 7, 9	—	—
	8	2, 8	3, 5, 7, 9	8, 9	—
	9	2, 4, 8	3, 5, 7, 9	9	—
ed	8	2, 8	5, 9	8, 9	3, 7
	9	—	5, 9	9	2, 3, 4, 7, 8

Table 5.6: *Plies which pose particular problems for the method*

5.3.3 Test suite (C2, RH, PD1, RH, C2, RH, PD1, RH)

Our main aim in considering libraries of ECWTs extracted from C2 data, with members selected using PD1 data as a contrast set, is to determine whether classification for left-hand data based on these members can make good the deficiency of the classification based on extracting libraries from PD1 data using C2 data as a contrast set.

However, we will stick to our order of considering right-hand data first, and look at the left-hand data later.

Plain

For the (C2, RH, PD1, RH, C2, RH, PD1, RH) test suite under the plain condition set, we present the ply-level classification results in Figure 5.28, the Wilcoxon results in Figure 5.29 and the epoch-level results in Figure 5.30.

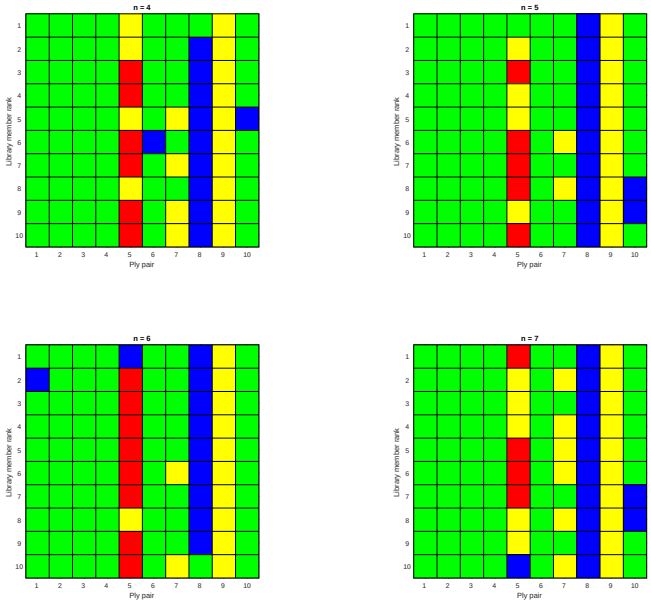


Figure 5.28: *Diagram like Figure 5.10, but for (C2, RH, PD1, RH, C2, RH, PD1, RH), with the plain condition set*

The ply-level classification Figure 5.28 suggests a partial return to form — the double success/overall success rates are 65%/79.5%, 67%/81%, 68%/80% and 64%/81% as n increases from 4 to 6.

Figure 5.29 is not quite so good news, as the non-rejection (at any of the p-levels we consider) rates of the null hypotheses (which we would like to be rejected!) are 56%, 53%, 49% and 53%.

Only a few of the profiles of the bar charts in Figure 5.30 are non-"U"-shaped, i.e., plies 1 and 6 for $n = 4$, ply 1 for $n = 5$, plies 5, 6 and 10 for $n = 6$ and ply 1 for $n = 7$.

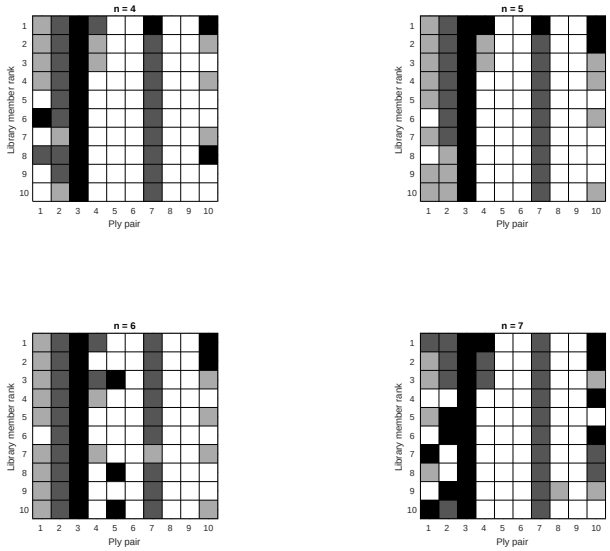


Figure 5.29: *Diagram like Figure 5.11, but for (C2, RH, PD1, RH, C2, RH, PD1, RH), with the plain condition set*

We note that, in the current case, the meaning of the central bars is somewhat reversed with regard to the discussion on page 217: the blue part of the central bars of the non-"U" histograms means that some of the motion patterns extracted from the (C2, RH) training data and chosen for their relative absence from the (PD1, RH) training data are also relatively absent from the (PD1, RH) test data, whereas other motion patterns extracted and selected in the same way are *not* absent from the (PD1, RH) test data. The interpretation of the red part here relates to the presence and absence of the same motion patterns from the (C2, RH) test data.

ec

The Figures summarising the results for the ec conditions set are 5.31, 5.32 and 5.33.

The ply-level classification, Figure 5.31 shows a good performance except where selections are missing (ply 4 for $n = 4$, and plies 4 and 5 for $n = 5$).

Again, except where selections are missing, the Wilcoxon Figure, 5.32, also shows good performance, with black squares dominating the non-red squares for the $n = 4$ and $n = 6$ cases, and grey squares dominating the $n = 5$ and $n = 7$ ones.

There are few non-"U" profiles in Figure 5.33, so there is little evidence *here* of different motion patterns being picked up.

When we compare these results with those for the plain condition set, we see that for ply-level classification, the double success/overall success rates increase for $p = 3, 5$ and 6, to 78%/84%, 85%/92% and 79%/89.5%, while, for $p = 3$, the decreases to 64%/71.5% are primarily due to the missing selections. If we co-opt the corresponding plain condition set

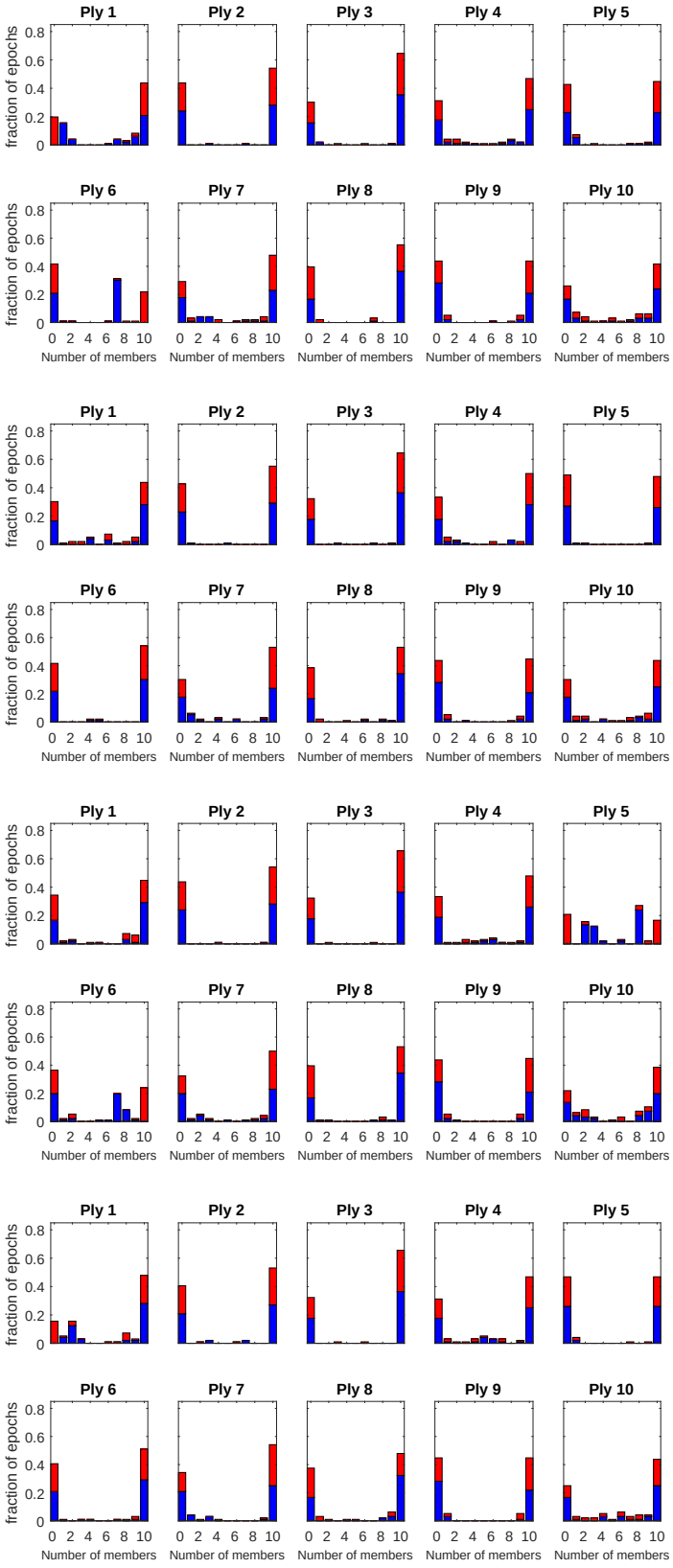


Figure 5.30: *Diagram like Figure 5.12, but for (C2, RH, PD1, RH, C2, RH, PD1, RH), with the plain condition set ($n = 4$–7, top to bottom)*

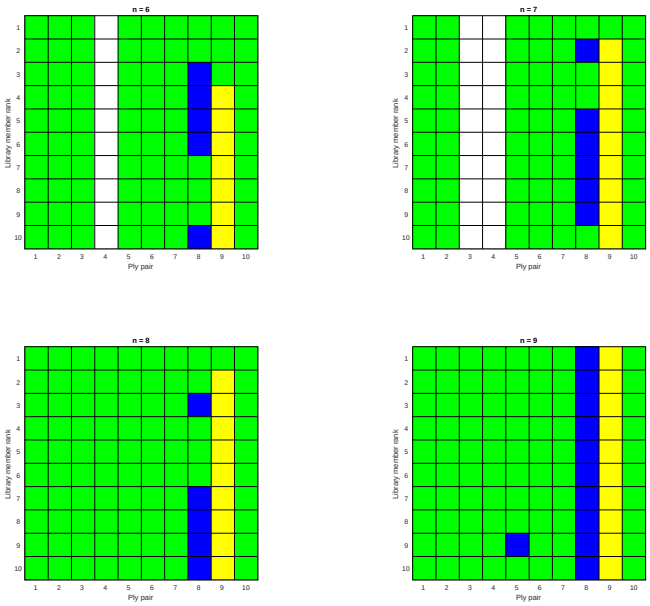


Figure 5.31: *Diagram like Figure 5.10, but for (C2, RH, PD1, RH, C2, RH, PD1, RH), with ec condition set*

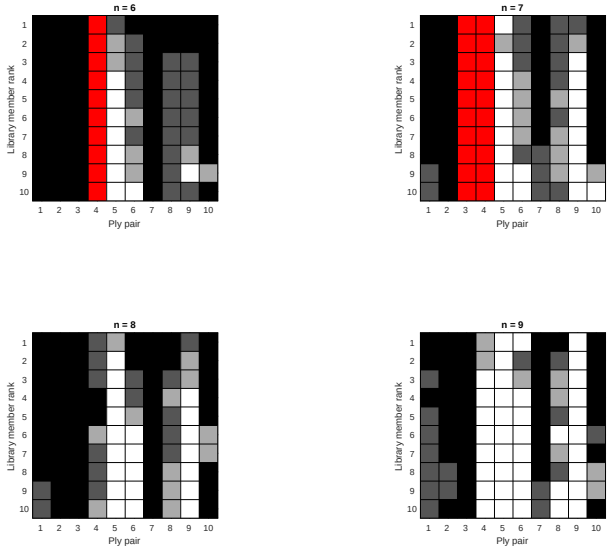


Figure 5.32: *Diagram like Figure 5.11, but for (C2, RH, PD1, RH, C2, RH, PD1, RH), with ec condition set*

selections to play the role of the missing ec selections, then there are increases across the board: to 88%/94% for $p = 3$ and 74%/81.5% for $p = 4$. Similarly, the Wilcoxon Figure for the ec condition set, Figure 5.32, is much darker than that for the plain condition set,

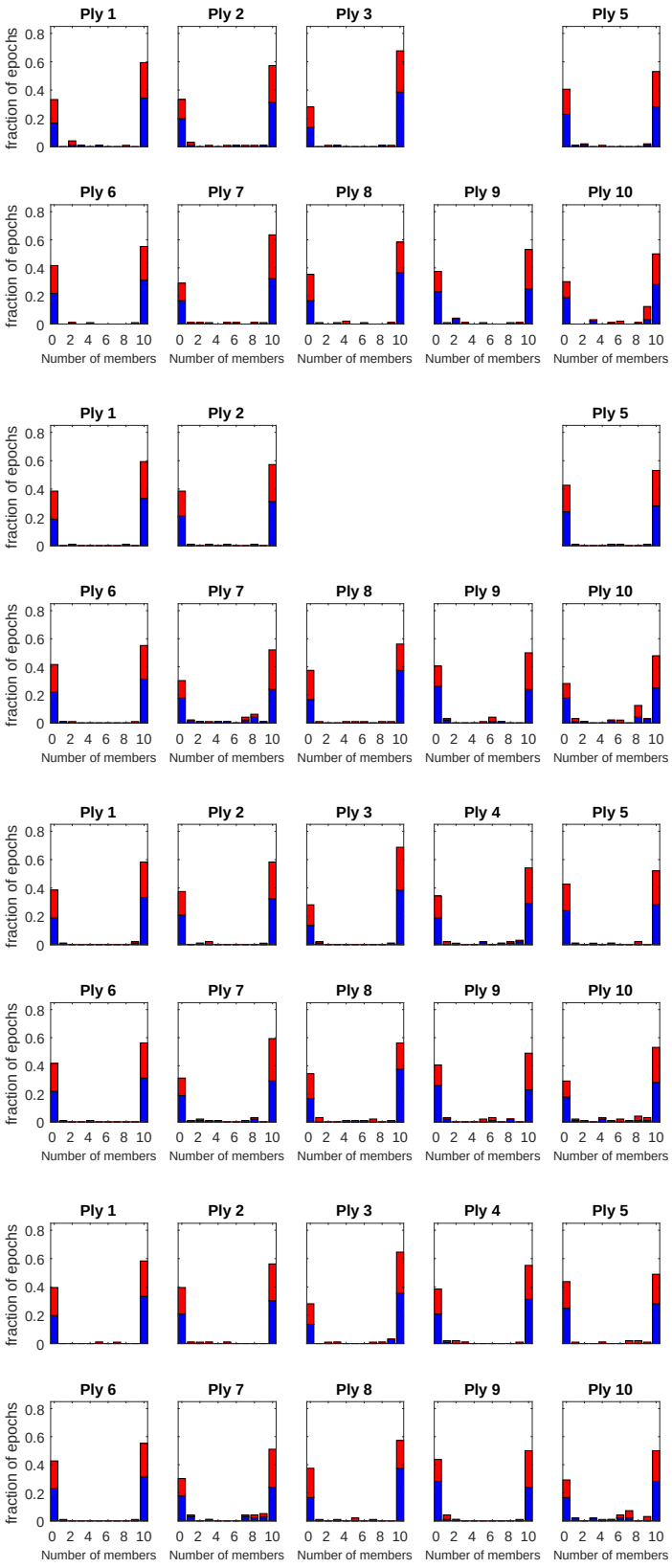


Figure 5.33: *Diagram like Figure 5.12, but for (C2, RH, PD1, RH, C2, RH, PD1, RH), with ec condition set*

Figure 5.29, a clear improvement.

For some unknown reason, the number of non-"U" epoch-level classification diagrams is reduced here in comparison to the case for the plain condition set.

ed

The ply-level classification here (Figure 5.34) shows near-perfect performance for $n = 8/p = 3$, and almost as good a performance for $n = 9/p = 4$.

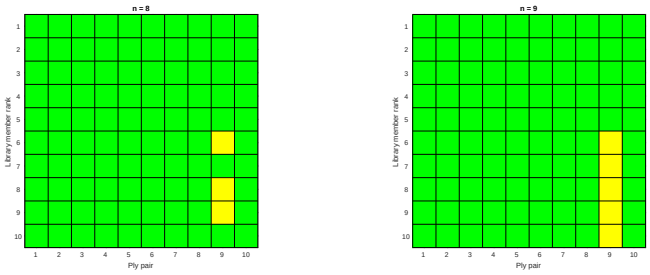


Figure 5.34: *Diagram like Figure 5.10, but for (C2, RH, PD1, RH, C2, RH, PD1, RH), with ed condition set*

Similarly, the Wilcoxon diagrams (Figure 5.35) are close to being solid black, apart from the ply 5 column, indicating that the tails of the activation distributions for PD1 and C2 are very different here.

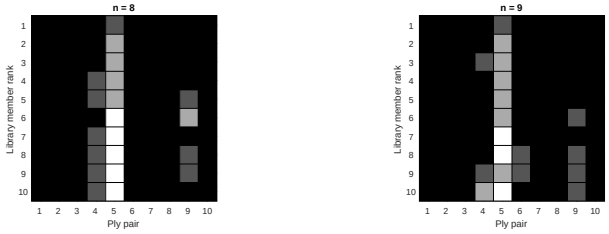


Figure 5.35: *Diagram like Figure 5.11, but for (C2, RH, PD1, RH, C2, RH, PD1, RH), with ed condition set*

The trend (for this test suite) to return to "U"-shaped profiles of the epoch-level classification as the severity of the imposed conditions increases continues here, as Figure 5.36 shows — almost all epochs are either classified correctly by all the selected members, or by none of them.

In comparison with the results for the ec condition set, we see a general improvement of classification performance and the significance of the difference in the tails of the activation distributions.

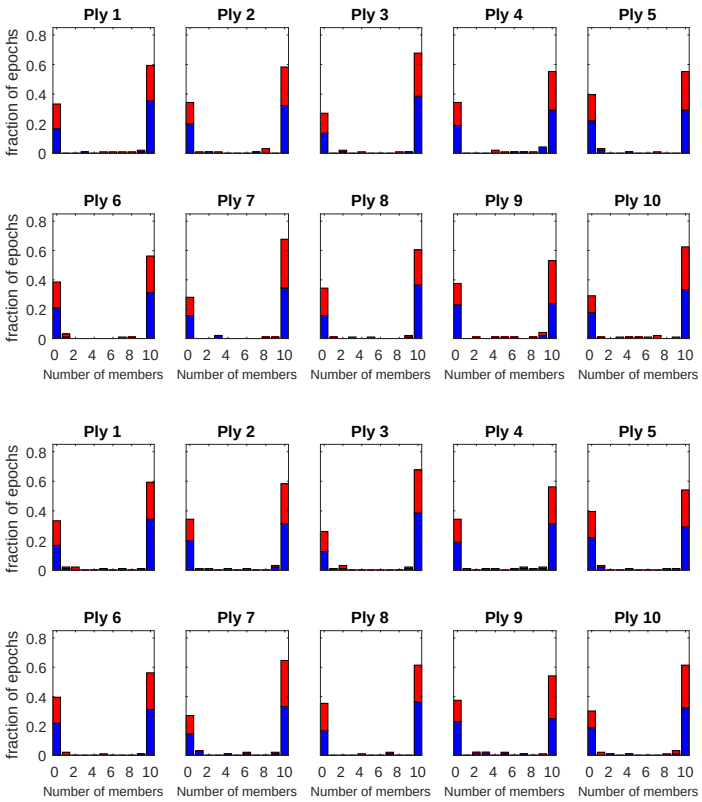


Figure 5.36: *Diagram like Figure 5.12, but for (C2, RH, PD1, RH, C2, RH, PD1, RH), with ed condition set ($n = 8$–9, top to bottom)*

Summary

As we have increased the severity of the conditions on the ECWTs involved, the ply-level classification has improved from mediocre (plain conditions) to rather good (ed conditions), although there were some selection failures for the ec condition set.

In Table 5.7 we list the problematic ply pairs. We see that there is some degree of stability in the plies which are difficult for the algorithm. Ply pairs 1, 2 and 1–8 never experience selection failure (however, remember that the selection happens on data *excluding* these plies), there is difficulty in classifying the PD1 member of pair 8 (except under the ed condition set), the C2 member of pair 9 and both members of ply 5 (but only under the plain condition set).

Because test suite (PD1, RH, C2, RH, PD1, RH, C2, RH) involves the classification of exactly the same data as the current suite, we compare the results there with those here.

Comparing the corresponding ply-level classification Figures, Figures 5.10 and 5.28, 5.13 and 5.31 and 5.16 and 5.34 immediately reveals, without any need to count squares, that the members selected using C2 data as a contrast set from libraries extracted from PD1 data do much better than members selected using PD1 data as a contrast set from libraries extracted from C2 data, except under the ed condition set, where the (PD1, RH, C2, RH) performance

		Difficulty			
Cond. Set	n	**PD1** (yellow)	**C2** (blue)	**Both** (red)	**Not selected** (white)
plain	4	5, 7, 9	6, 8, 10	5	—
	5	5, 7, 9	8, 10	5	—
	6	5, 7, 9	1, 5, 8	5	—
	7	5, 7, 9	5, 8, 10	5	—
ec	6	9	8	—	4
	7	9	8	—	3, 4
	8	9	8	—	—
	9	9	8	—	—
ed	8	9	—	—	—
	9	9	—	—	—

Table 5.7: *Plies which pose particular problems for the method*

is affected by selection failure.

It is exactly the same story for the Wilcoxon Figures — many more black squares, and much fewer white ones, for the (PD1, RH, C2, RH, PD1, RH, C2, RH) test suite.

However, the result of this comparison is what we expected: more motion patterns extracted from PD data are absent in control data, than *vice versa*.

Finally, we compare Tables 5.5 and 5.7: ply pair 5, both of whose plies are problematic here for the plain condition set, is also problematic under the same circumstances for (PD1, RH, C2, RH) training; the PD1 ply of pair 9 is difficult everywhere here, and under the plain condition set there; the C2 ply of pair 8 is difficult for all the classifiers produced under the plain and ec condition set here, and also the classifiers produced there, under the same condition sets (excluding ec, $n = 8$); and it is the C2 ply of pair 9 that is difficult everywhere here, but only under the plain conditions set.

Ply pairs 5, 8 and 9 are the only ones which cause any classification problems for the training suite (PD1, RH, C2, RH), and only 9 does so for any set of selected members under the ec and ed condition set (and just for one value of the ordered pair $(n, \text{condition set})$) where as several others do for the suite (C2, RH, PD1, RH), so, in this respect too, the former training suite appears superior.

In fact, it is only from the point of view of selection failure that (C2, RH, PD1, RH) appears superior, and this is an empty superiority, as perfect selection is possible using (PD1, RH, C2, RH): i.e., under the ec condition set for $n = 6$–8.

5.3.4 Test suite (C2, LH, PD1, LH, C2, LH, PD1, LH)

Because of the relative failure of the training suite (PD1, LH, C2, LH) to classify (PD1, LH) and (C2, LH) data (at least compared to the success of the training suite (PD1, RH, C2, RH) in classifying (PD1, RH) and (C2, RH) data)), we wish to examine the efficacy of the outcome of the training suite (C2, LH, PD1, LH) in classifying (PD1, LH) and (C2, LH) data.

Plain

In Figures 5.37 (ply-level classification), 5.38 (Wilcoxon) and 5.39 (epoch-level classification), we present the results for this test suite under the plain condition set.

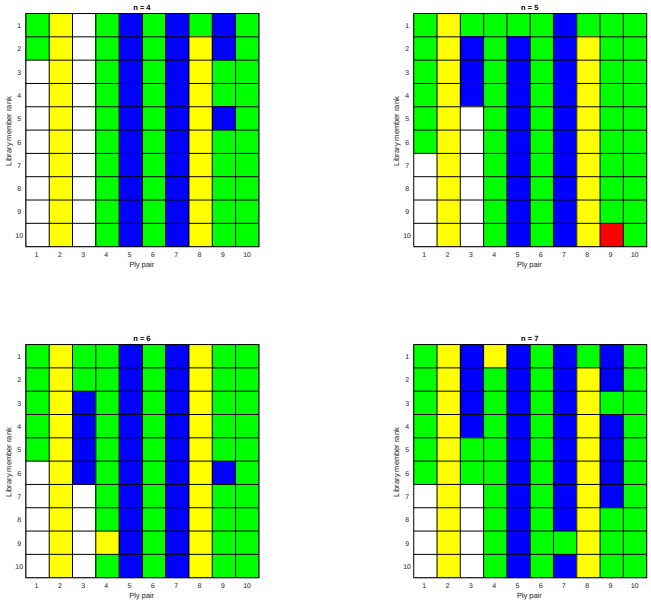


Figure 5.37: *Diagram like Figure 5.10, but for (C2, LH, PD1, LH, C2, LH, PD1, LH), with the plain condition set*

Handicapped by a number of selection failures ranging from 8% to 18%, we can see from the ply-level classification Figure, 5.37, that the double classification/overall classification success rates are 40%/61%, 48%/68.5%, 45%/68% and 43%/67.5%, for $n = 4, 5, 6, 7$ respectively.

Also, in the Wilcoxon Figure, 5.38, apart from the columns for ply pairs 9 and 10, and the mostly red columns for pairs 1 and 3, only columns 8 and 9 are dominated by rejections of the null hypotheses (and none of these is at the 1% p-level), so it is unlikely to be easy to improve the ply-level classification results by a better choice of thresholds.

Turning to the epoch-level classification diagrams of Figure 5.39, we see that roughly 25% are non-"U", so there is evidence here for the activations with respect to the selected members reflecting different motion patterns in the data.

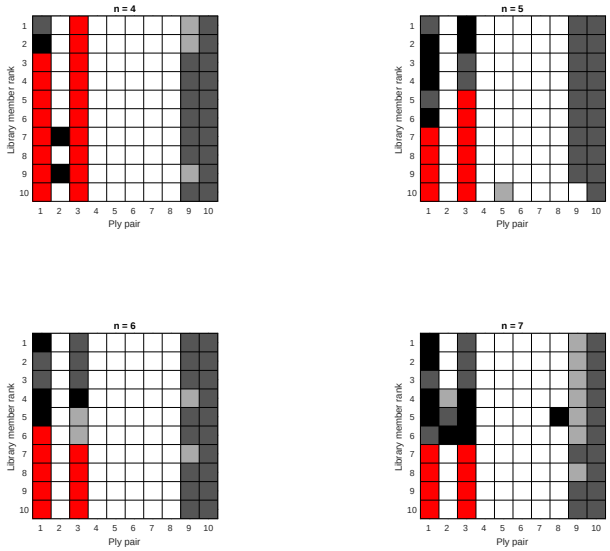


Figure 5.38: *Diagram like Figure 5.11, but for (C2, LH, PD1, LH, C2, LH, PD1, LH), with the plain condition set*

ec

Our summary diagrams in this section are contained in Figures 5.40 to 5.42.

The ply-level classification diagrams in Figure 5.40 display double/overall classification success rates of 56%/78%, 51%/78%, 54%/77% and 54%/77% (again), as n runs from 4 to 7, and the Wilcoxon diagrams of Figure 5.41 are dominated by white squares.

We do not have any markedly non-"U" profiles among the epoch-level classification diagrams of Figure 5.41, so here there is no further evidence of the selected members leading to measurement of different motion patterns.

When we compare these results with those for the plain condition set of the previous section, we find:

1. the number of selection failures has dropped;
2. primarily as a result of this, the double/overall success rate increases substantially;
3. although the ec Wilcoxon diagrams remain dominated by white squares, everywhere where there was a selection failure under the plain condition set, the null hypothesis is rejected, while elsewhere the rejections (at one or other of our levels) in diagrams representing results at the same number of degrees of freedom are in the same place. This naturally results in an increase of the rate of rejection of the null hypotheses;
4. the number of non-"U" profiles in the epoch-level classification diagrams is reduced in the ec Figure.

We can sum up by saying that the ec results are substantially better than the plain ones.

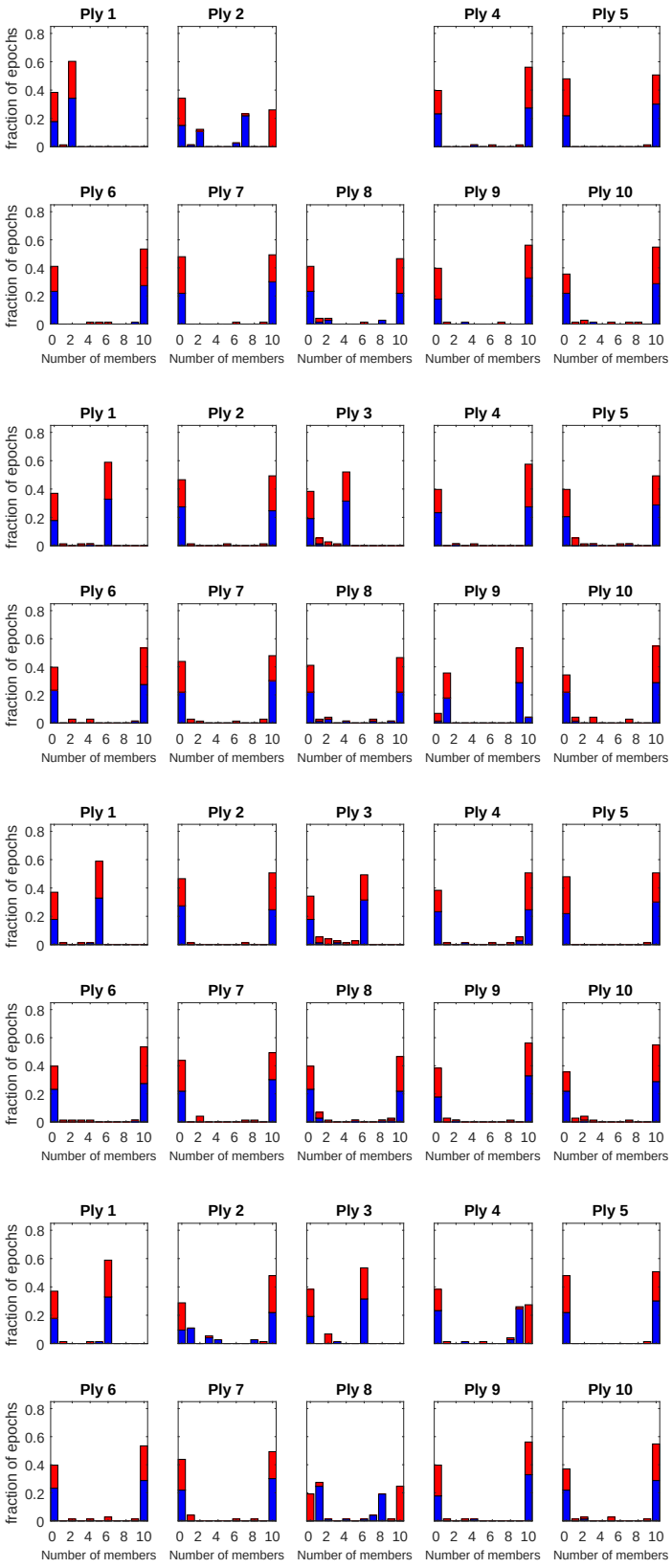


Figure 5.39: *Diagram like Figure 5.12, but for (C2, LH, PD1, LH, C2, LH, PD1, LH), with the plain condition set ($n = 4$–7, top to bottom)*

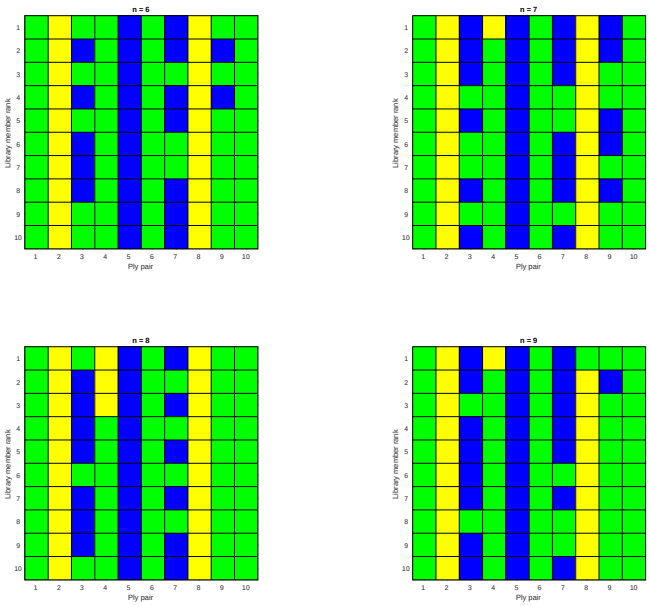


Figure 5.40: *Diagram like Figure 5.10, but for (C2, LH, PD1, LH, C2, LH, PD1, LH), with ec condition set*

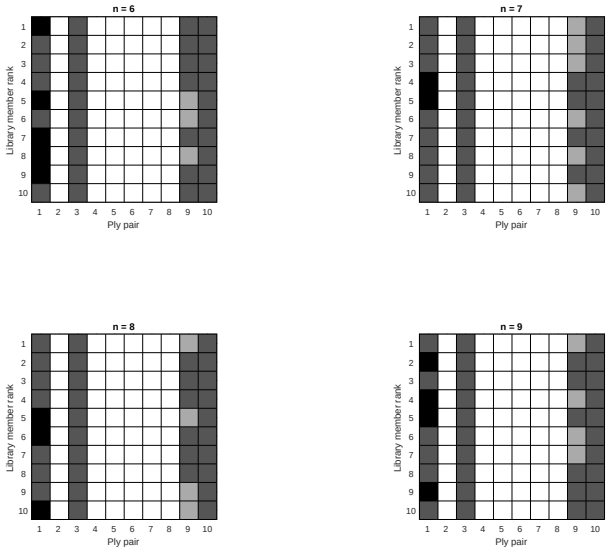


Figure 5.41: *Diagram like Figure 5.11, but for (C2, LH, PD1, LH, C2, LH, PD1, LH), with ec condition set*

ed

In the ply-level classification diagrams for the ed condition set, Figure 5.43, there are identical double/overall classification rates: 58%/79%, which are not marvellously good, but approach

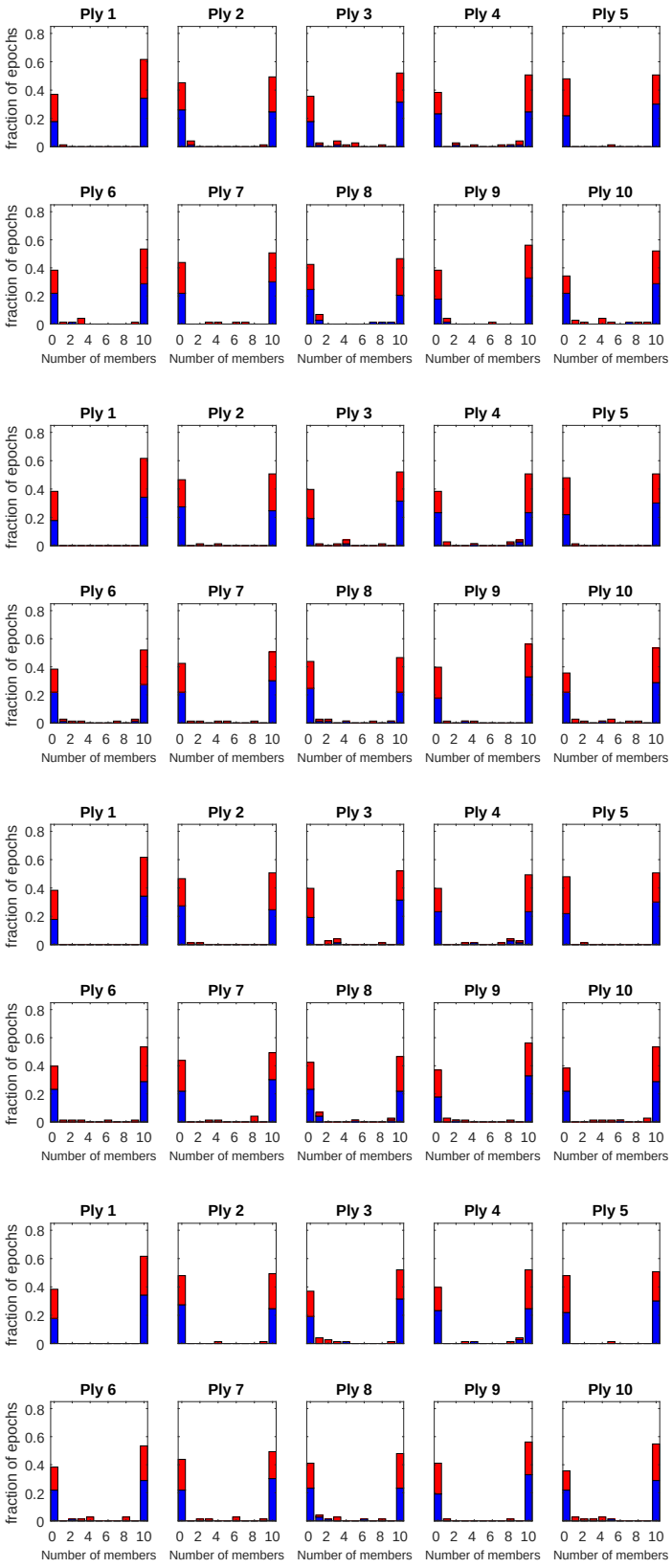


Figure 5.42: *Diagram like Figure 5.12, but for (C2, LH, PD1, LH, C2, LH, PD1, LH), with ec condition set ($n = 6$–9, top to bottom)*

the useful.

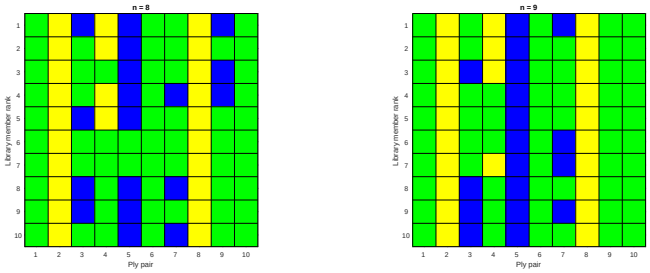


Figure 5.43: *Diagram like Figure 5.10, but for (C2, LH, PD1, LH, C2, LH, PD1, LH), with ed condition set*

There is no great indication that these results could be improved by a different choice of thresholds from the Wilcoxon diagrams in Figure 5.44.

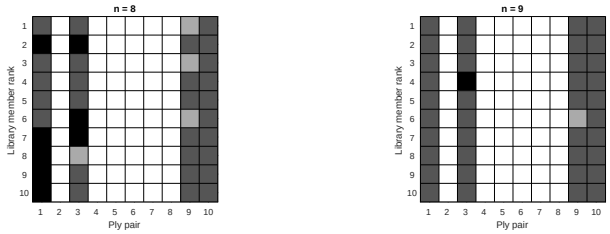


Figure 5.44: *Diagram like Figure 5.11, but for (C2, LH, PD1, LH, C2, LH, PD1, LH), with ed condition set*

As far as the epoch-level classification diagrams of Figure 5.45 are concerned, there is no evidence for different motion patterns being found in different epochs, as there are no non-"U" profiles.

When we compare the ed results to the ec ones, we find that they are slightly better (ec: double/overall classification success rates: 51-56%/75.5-78%; ed: 58%/79%), but the Wilcoxon diagrams are close to being identical across the two condition sets and the values of n considered. There is no striking difference in the epoch-level classification diagrams, either.

Summary

In Table 5.8 we list the ply pairs which were particularly difficult to classify. We see that there is a lot of stability in which plies are difficult to classify — PD1 plies 3, 5, 7 and 9 and C2 plies 2, 4 and 8. There was also some difficulty in selecting members for ply pairs 1 and 3, but only under the plain condition.

When we compare the results for this test suite with those for (PD1, RH, C2, RH, PD1,

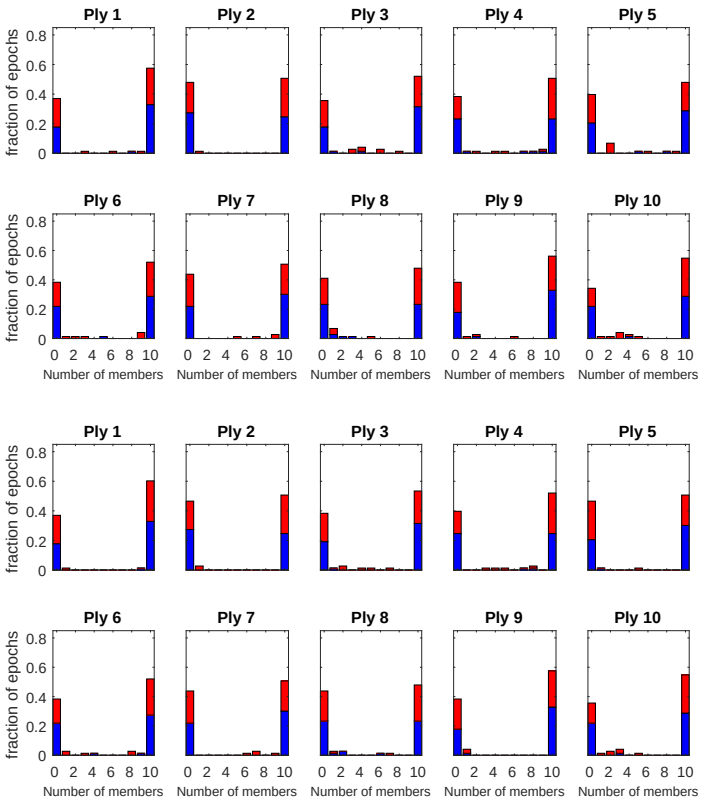


Figure 5.45: *Diagram like Figure 5.12, but for (C2, LH, PD1, LH, C2, LH, PD1, LH), with ed condition set ($n = 8$–9, top to bottom)*

		Difficulty			
Cond. Set	n	**PD1** (yellow)	**C2** (blue)	**Both** (red)	**Not selected** (white)
plain	4	2, 8	5, 7, 9	—	1, 3
	5	2, 8	3, 5, 7	9	1, 3
	6	2, 4, 8	3, 5, 7	—	1, 3
	7	2, 4, 8	3, 5, 7, 9	—	1, 3
ec	6	2, 8	3, 5, 7, 9	—	—
	7	2, 4, 8	3, 5, 7, 9	—	—
	8	2, 4, 8	3, 5, 7	—	—
	9	2, 4, 8	3, 5, 7, 9	—	—
ed	8	2, 4, 8	3, 5, 7, 9	—	—
	9	2, 4, 8	3, 5, 7	—	—

Table 5.8: *Plies which pose particular problems for the method*

RH, C2, RH), we find that the latter are far better in terms of double/overall classification success rates.

However, the motivation for considering the current suite, (C2, LH, PD1, LH, C2, LH, PD1, LH) was to see if we could improve over the results for the suite (PD1, LH, C2, LH,

PD1, LH, C2, LH), i.e., to see if we could find better sets of classifiers for the unseen parts of the data sets (PD1, LH) and (C2, LH).

By comparing Tables 5.6 and 5.8, we see that the same ply pairs are difficult for both training suites: 2, 4 and 8 for their PD1 member, and 3, 5, 7 and 9 for their C2 plies.

When we collate the double and overall success rates for the two suites in Table 5.9, we can easily see that the best results come for (C2, LH, PD1, LH, C2, LH, PD1, LH), but that the improvement is slight, and that frequently, for the same condition set and n, the larger value is for the other suite.

		(PD1, LH, C2, LH, PD1, LH, C2, LH)		**(C2, LH, PD1, LH, C2, LH, PD1, LH)**	
Cond. set	n	**Double success**	**Overall success**	**Double success**	**Overall success**
plain	4	34%	61%	40%	61%
	5	50%	71%	48%	68.5%
	6	51%	74%	45%	68%
	7	47%	70.5%	43%	67.5%
ec	6	**57%**	76.5%	56%	78%
	7	54%	**77%**	51%	75.5%
	8	55%	75.5%	54%	77%
	9	52%	73.5%	54%	77%
ed	8	45%	60%	**58%**	**79%**
	9	34%	40%	**58%**	**79%**

Table 5.9: *Double and overall success rates for (PD1, LH, C2, LH, PD1, LH, C2, LH) and (C2, LH, PD1, LH, C2, LH, PD1, LH) (the largest values in each rate column are in a bold font)*

In relative terms, there is a great improvement in the number of cases in which the null hypotheses about the tails of the distributions which we would like to reject, *are* rejected for the (C2, LH, PD1, LH, C2, LH, PD1, LH) test suite, as we can see by comparing the Wilcoxon Figures 5.20 ((PD1, LH, C2, LH, PD1, LH, C2, LH), plain condition set) and 5.38 ((C2, LH, PD1, LH, C2, LH, PD1, LH), plain condition set), Figures 5.23 ((PD1, LH, C2, LH, PD1, LH, C2, LH), ec condition set) and 5.41 ((C2, LH, PD1, LH, C2, LH, PD1, LH), ec condition set) and Figures 5.23 ((PD1, LH, C2, LH, PD1, LH, C2, LH), ed condition set) and 5.44 ((C2, LH, PD1, LH, C2, LH, PD1, LH), ed condition set). Nevertheless, this improvement still leaves the Wilcoxon diagrams dominated by white, non-rejection, squares.

5.4 Generalisation across attachment

In this section, we determine whether there is any apparent generalisation across the attachment point of the NAT; we attempt to use library members extracted and selected using right-hand data when classifying left-hand data, and *vice versa*.

5.4.1 Test suite (PD1, RH, C2, RH, PD1, LH, C2, LH)

Here we attempt to classify left-hand data using library members resulting from the training suite (PD1, RH, C2, RH).

Plain

Under the plain condition set, we obtain the ply-level classifications shown in Figure 5.46.

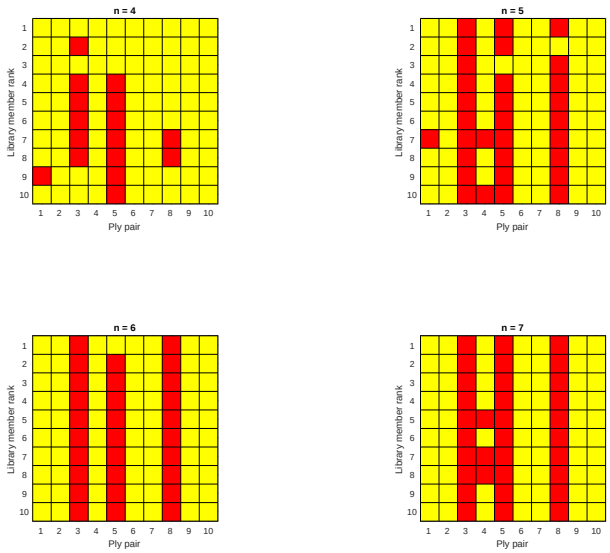


Figure 5.46: *Diagram like Figure 5.10, but for (PD1, RH, C2, RH, PD1, LH, C2, LH), with the plain condition set*

We can immediately see that the PD ply for every ply pair is misclassified for every library member selected for the pair, and that the misclassification rate of the control plies varies from 16% to 33%, so we can also immediately say that the generalisation across attachment under the plain condition set is terrible, at least with the given training suite.

We do not present Wilcoxon diagrams like those of Figure 5.11 for (PD1, RH, C2, RH, PD1, LH, C2, LH), with the plain condition set, as the hypothesis is not rejected at any of our levels for any of the member-ply pair combinations which would be represented on such diagrams In other words, the diagrams would consist solely of white squares. Of course, this fact also points to a very poor generalisation performance.

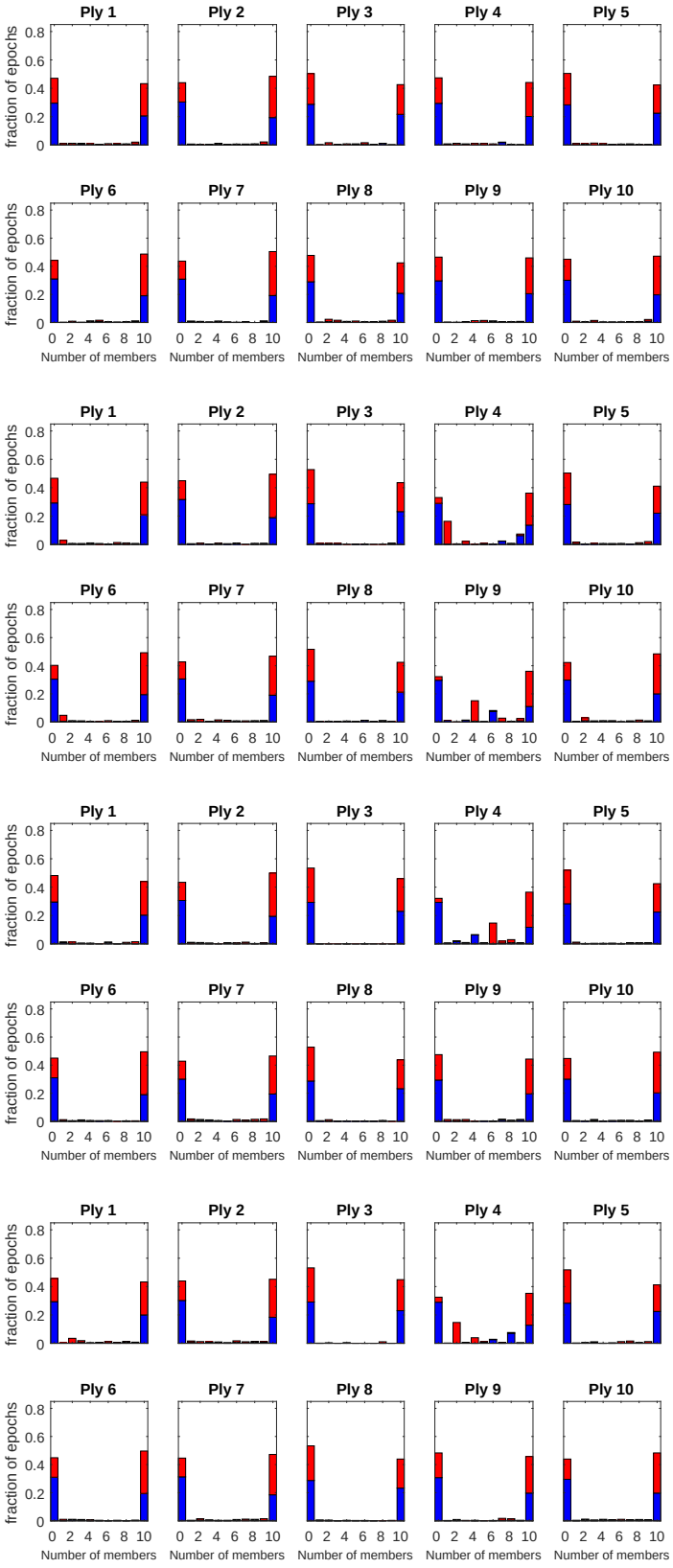


Figure 5.47: *Diagram like Figure 5.12, but for (PD1, RH, C2, RH, PD1, LH, C2, LH), with the plain condition set ($n = 4$–7, top to bottom)*

As regards the epoch-level performance depicted in Figure 5.47, we note the presence of several non-"U" profiles, but we can easily see that the left-hand upright in the "U"-profiles is larger than the right-hand one, which is the reverse of the case for our more successful classifications — there are more epochs which are incorrectly classified by all the relevant library members than are correctly classified by all the relevant members.

ec and ed

The story for the ec and ed condition sets is mostly the same as for the plain condition set — the classification performance is very poor, although there are slightly fewer misclassifications of the C2 plies; on the other hand, there are some selection failures under the ed condition set.

For completeness, we present the ply-level classification diagrams in Figures 5.48 and 5.49, and the epoch-level classifications in Figures 5.50 and 5.51.

We do not present Wilcox diagrams like those of Figure 5.11 for (PD1, RH, C2, RH, PD1, RH, C2, RH) for (PD1, RH, C2, RH, PD1, LH, C2, LH), with the ec and ed condition sets, as, again, the diagrams would be all white, or red and white in the case of the ed set.

Summary

The generalisation performance of the classification based on libraries extracted and selected from the right-hand datasets we used to left-hand data under the ec and ed condition sets is virtually non-existent, and is even worse under the plain condition set.

The large amounts of yellow and red across the ply-level classification diagrams in Figures 5.46, 5.48 and 5.49 mean that, for these classifiers, left-hand PD1 data resembles right-hand C2 data more than it does right-hand PD1 data.

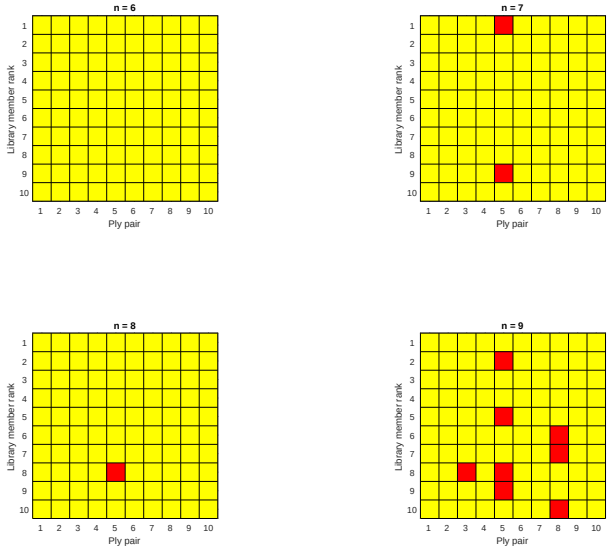


Figure 5.48: *Diagram like Figure 5.10, but for (PD1, RH, C2, RH, PD1, LH, C2, LH), with ec condition set*

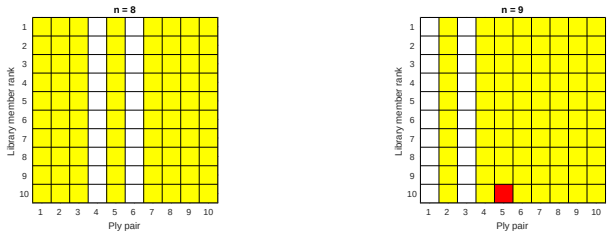


Figure 5.49: *Diagram like Figure 5.10, but for (PD1, RH, C2, RH, PD1, LH, C2, LH), with ed condition set*

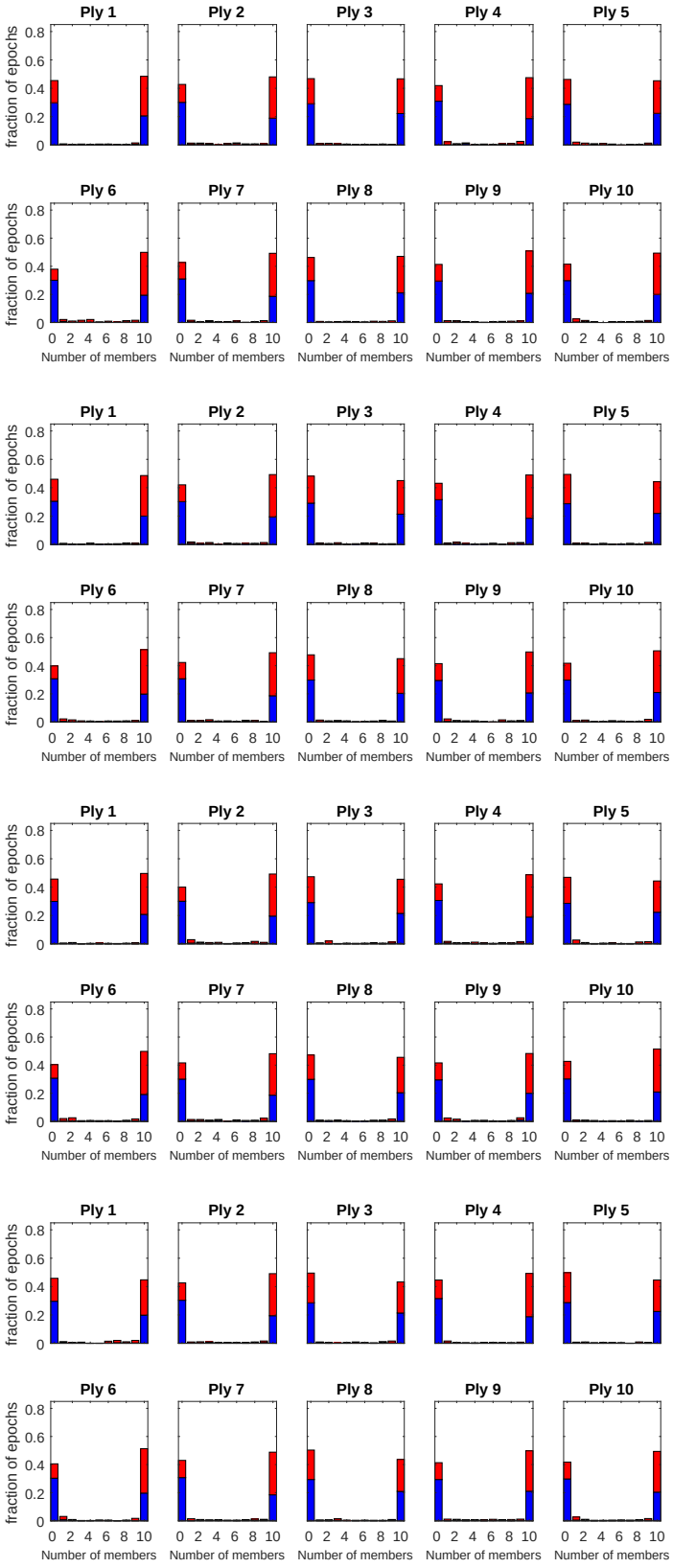


Figure 5.50: *Diagram like Figure 5.12, but for (PD1, RH, C2, RH, PD1, LH, C2, LH), with ec condition set ($n = 6$–9, top to bottom)*

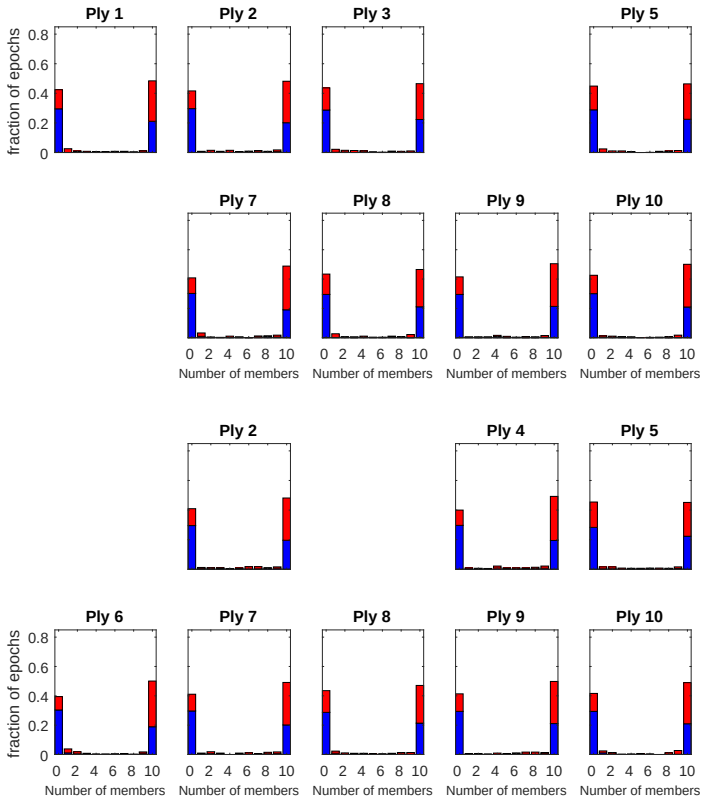


Figure 5.51: *Diagram like Figure 5.12, but for (PD1, RH, C2, RH, PD1, LH, C2, LH), with ed condition set ($n = 8$–9, top to bottom)*

5.4.2 Test suite (PD1, LH, C2, LH, PD1, RH, C2, RH)

Here we examine the generalisation from left-hand data to right-hand data.

Plain

Ply pair 9 aside, every PD1 member of a ply pair is misclassified in the plain condition set ply-level classification diagrams of Figure 5.52, and, even for ply pair 9, the number of selected members for fixed n which correctly classify the PD1 member (green or blue squares) varies from 10% to 80%, with more in the lower part of the range. On the other hand, the C2 members are mostly correctly classified (and the C2 member of ply pair 9 is always correctly classified). This combination means that the overall success rate is greater than 50% only for $n = 7$ (and there it is only 51%).

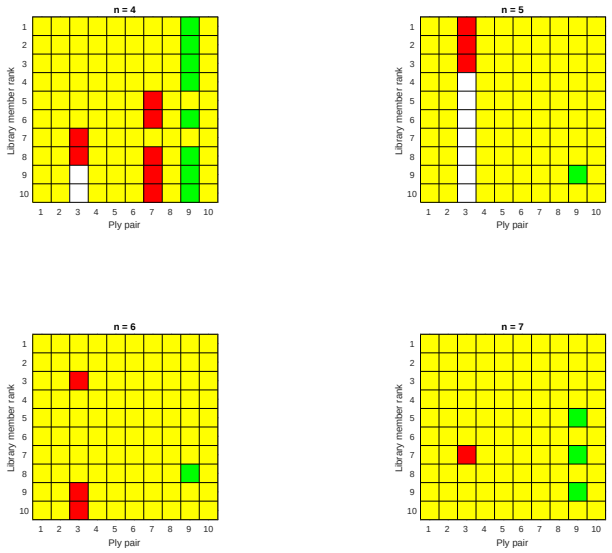


Figure 5.52: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, RH, C2, RH), with the plain condition set*

This picture is confirmed by the diagrams of the Wilcoxon Figure 5.53, where the desired rejections of the null hypotheses are concentrated in the column for ply pair 9, and then only for appreciable numbers when $n = 4$.

In Figure 5.54, we see that several of the epoch-level classification diagrams are non-"U" (including the diagrams for plies 8 and 9 for $n = 5$, which, although strictly speaking, *are* "U"-shaped, have doubled left and right vertical strokes), with the usual implication that different selected library members measure different things.

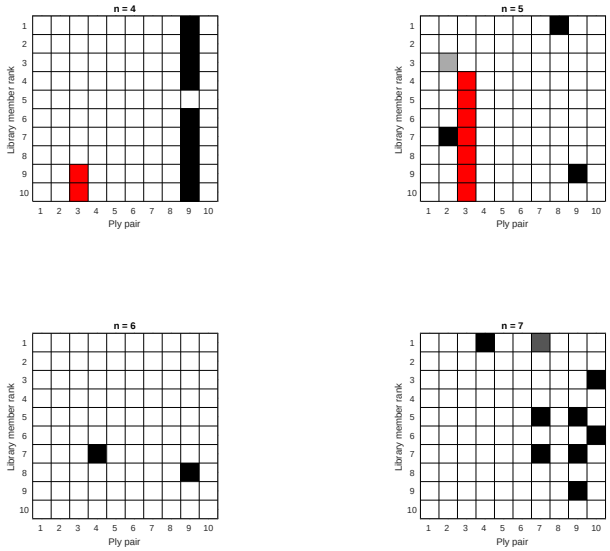


Figure 5.53: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, RH, C2, RH), with the plain condition set*

ec

Although the ec ply-level classification diagrams of Figure 5.55 look a little different to those of Figure 5.52, in that blue squares make their appearance, reflecting an increase in the number of times that the PD1 member of a ply is correctly classified, the fact that they also represent a decrease in the number of times the C2 member is correctly classified, and that there is also an increase in the number of red squares, means that the overall success rate is reduced with respect to the plain condition set. The even poorer quality of the ec classification is also reflected in the low number of grey squares in the Wilcoxon Figure, Figure 5.56.

For some reason, most of the extra red squares occur in ply-pair columns 1, and, where the corresponding squares were not already red in Figure 5.52, 7. That is, it seems particularly difficult to classify the C2 ply of these pairs under the ec condition set.

The ply pair 9 remains an exception, in that green squares also occur in its column of the ply-level classification diagrams under the ec condition set, and also grey squares in the Wilcoxon diagram.

In the epoch-level classification diagrams, Figure 5.57, we again have several non-"U" profiles, indicating diversity in the selected members.

ed

Because of the 7 columns of selection failures over the twenty columns, two per ply pair, over the two ply-level classification diagrams of Figure 5.58, it is difficult to tell whether the

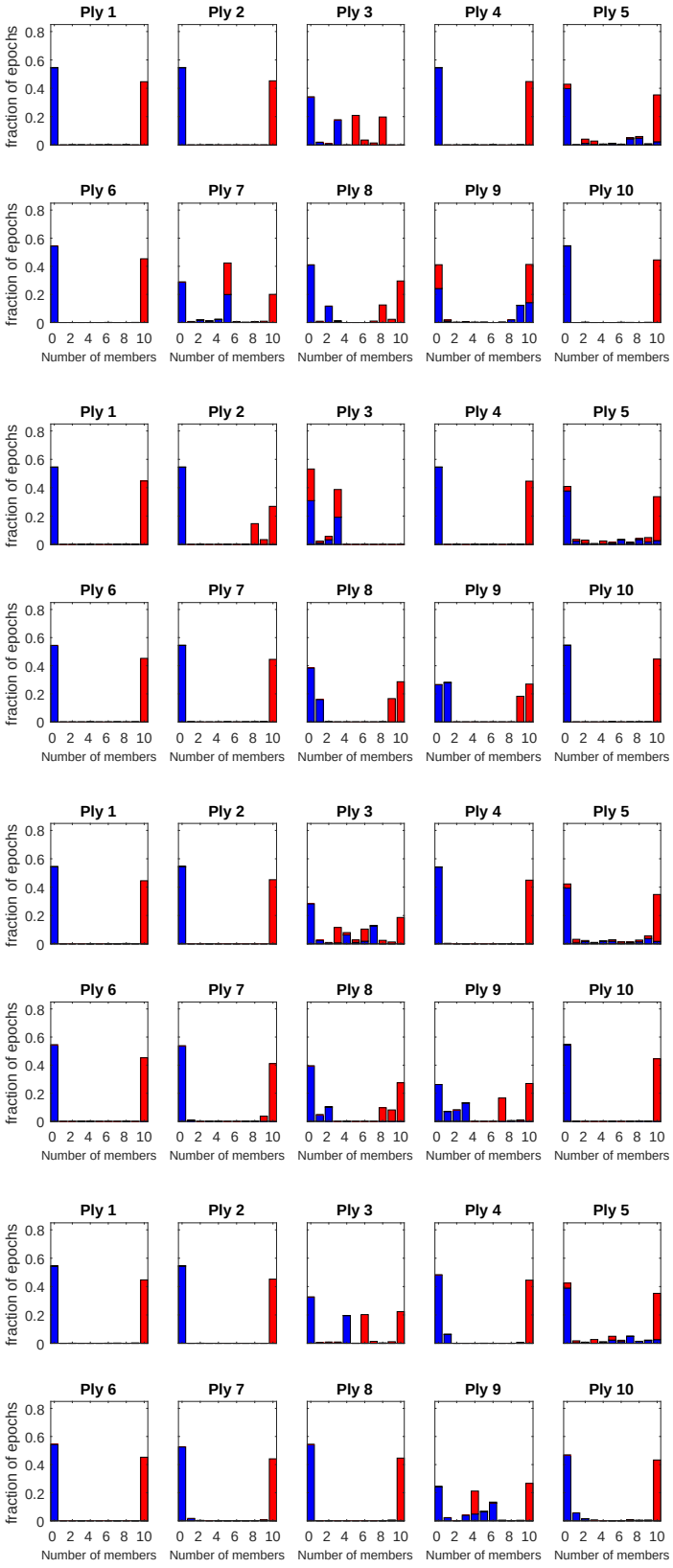


Figure 5.54: *Diagram like Figure 5.12, but for (PD1, LH, C2, LH, PD1, RH, C2, RH), with the plain condition set (n = 4–7, top to bottom)*

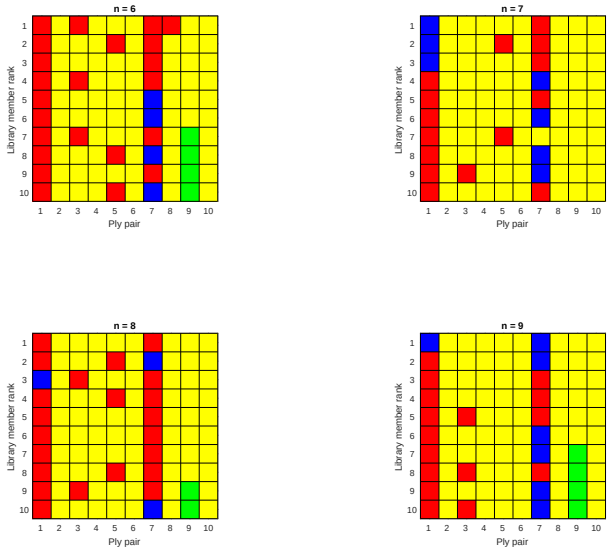


Figure 5.55: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, RH, C2, RH), with ec condition set*

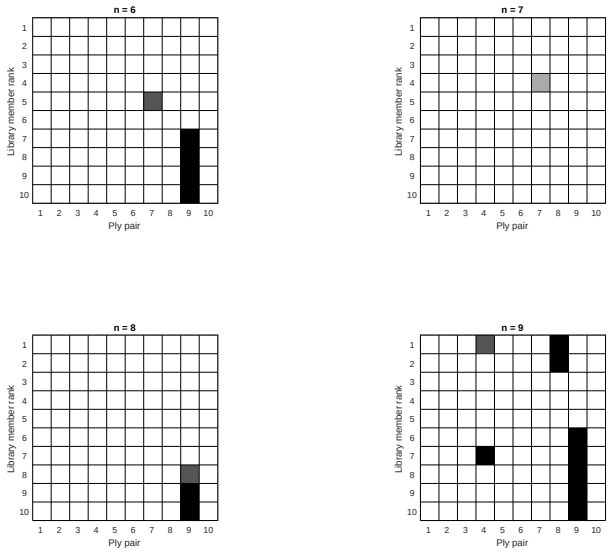


Figure 5.56: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, RH, C2, RH), with ec condition set*

classification results for the ed condition set more closely resemble those for the plain or ec conditions sets, but they are similar to both in that they document a poor generalisation from the right-hand data to the left-hand data. That said, the particular difficulty that the classification under the ec set has with the C2 ply of pair 1 extends to the ed set.

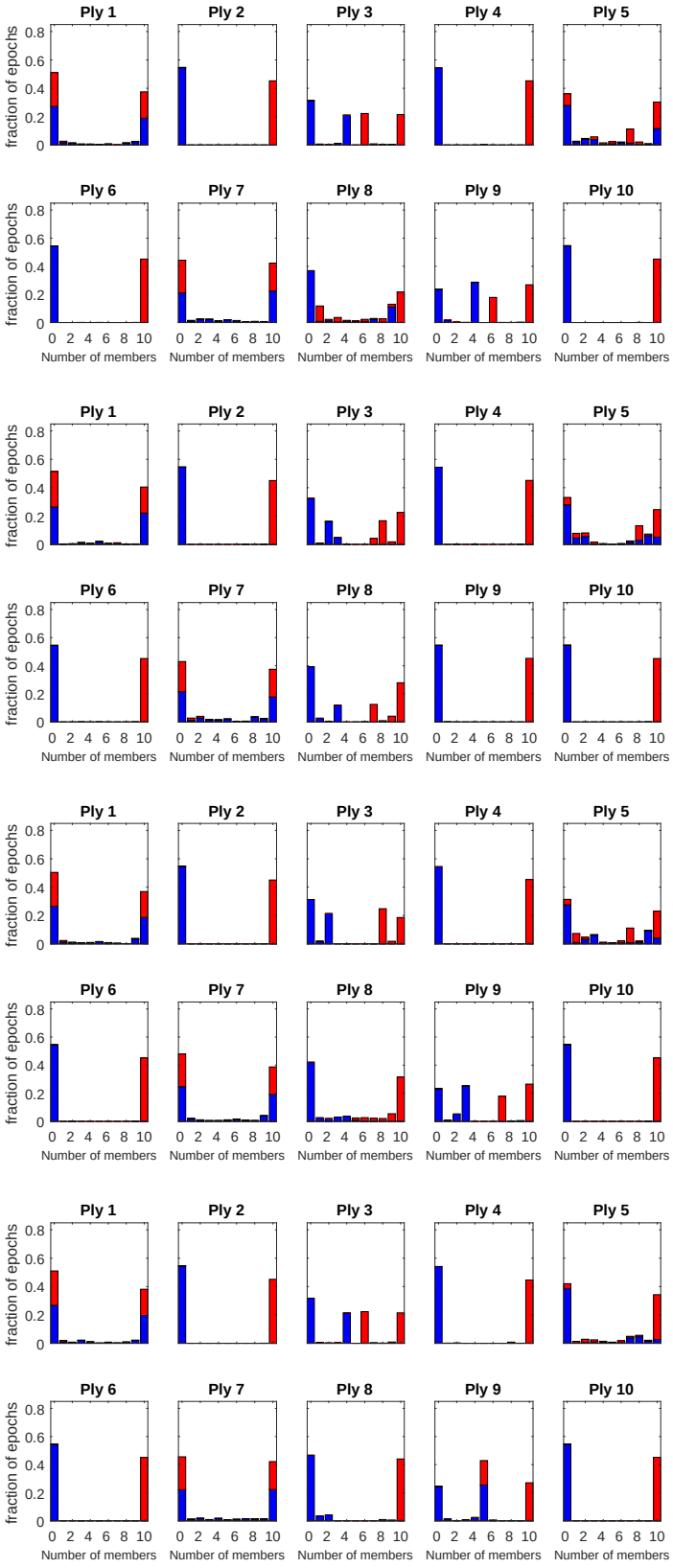


Figure 5.57: *Diagram like Figure 5.12, but for (PD1, LH, C2, LH, PD1, RH, C2, RH), with ec condition set ($n = 6$–9, top to bottom)*

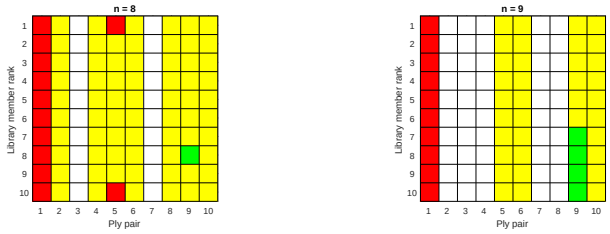


Figure 5.58: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, RH, C2, RH), with ed condition set*

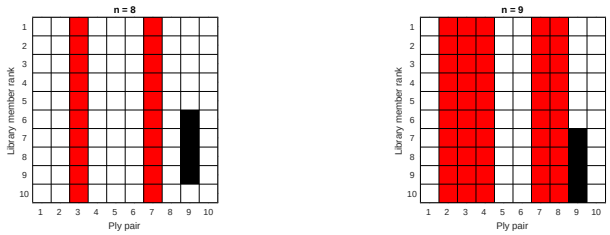


Figure 5.59: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, RH, C2, RH), with ed condition set*

Summary

Apart from ply pair 9, under the plain and ed condition sets, the story for the generalisation from left-hand to right-hand data is the same as the reverse generalisation — it is poor.

For the ec condition set, the generalisation is also poor apart from ply pair 9, but there are some blue squares, meaning that there are some library members selected for some ply pairs which correctly classify PD1 data, but not C2 data. On balance, the performance with the ec condition set is worse than under the plain condition set, and probably only better than the performance under the ed condition set because of the missing selections under the latter set.

The large amounts of yellow and red in the ply-level classification diagrams here mean that, for these classifiers, right-hand PD1 data resembles left-hand C2 data more than it does left-hand PD1 data.

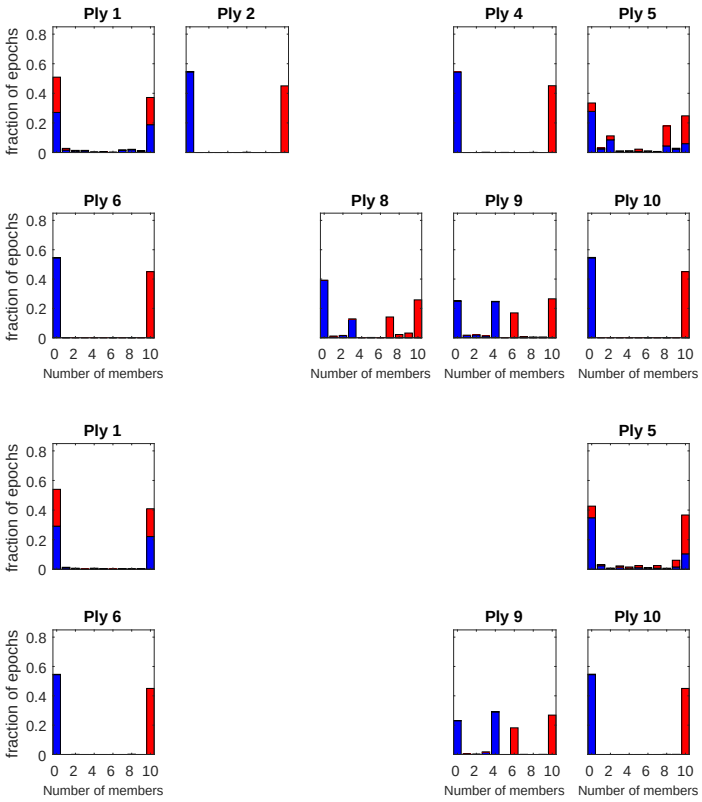


Figure 5.60: *Diagram like Figure 5.12, but for (PD1, LH, C2, LH, PD1, RH, C2, RH), with ed condition set ($n = 8$–9, top to bottom)*

5.5 Generalisation across control subjects

In this section, we replace the control subject C2 by the subjects C3 and C4 in the *test* data.

The same plies of PD1 data (left or right) are used in these experiments, and so the classification of the PD1 half of the ply pairs here is the same as when they were classified alongside C2 data. Consequently, the map which takes squares of, say, Figure 5.10, to the corresponding squares of the corresponding Figure here (Figure 5.61 below), and which is determined by the classification of the other half of the ply pair, is constrained to take the green squares (representing (P, C) → (P, C)) to green or blue squares ((P, C) → (P, P)), the blue squares to green or blue squares, the yellow squares ((P, C) → (C, C)) to yellow or red squares ((P, C) → (C, P)), and the red squares to yellow or red squares. Of course, as the training suite is unchanged, white squares are taken to white squares.

As the results for the PD1 epochs are unchanged from earlier work, the epoch-level classification diagrams here are, like the ply-level classification diagrams, a kind of blend of new results and ones seen earlier. However, they add little to the discussion from this point on, so we refrain from displaying them.

5.5.1 Test suite (PD1, RH, C2, RH, PD1, RH, C4, RH)

Here we examine the effect of replacing the C2 test data by C4 data, in the tests for the right-hand attachment.

Plain

Figure 5.61 contains the ply-level classification diagrams for this test suite under the plain condition set. The results are disappointing, because they show a large amount of misclassification of the C4 plies (blue and red squares).

However, if we look at the Wilcoxon Figure, 5.62, we see large areas of overlap between the black squares and the blue squares of Figure 5.61, suggesting that the null hypotheses that the tails of the PD1 activation distributions are *not* thicker than the C4 activation distribution tails can comfortably be rejected. Cutting through the mess of multiple negatives here, we can say that the distributions are statistically significantly different, and that adjusting thresholds can improve the classification performance.

This point is reinforced if the grey squares are counted alongside the black ones.

ec

If we now look at the ply-level classifications for the ec condition set shown in Figure 5.63, we see a clear improvement over the results for the plain condition set: the double/overall success rates increase from 20%/55.5% to 51%/75.5% ($p = 3$ degrees of freedom); 17%/54% to 46%/73% ($p = 4$); 17%/54% to 44%/72% ($p = 5$); and 16%/52.5% to 48%/74% ($p = 6$)

The almost total domination of the Wilcoxon Figure 5.64 by black squares suggests that the removal of most of the red and blue squares from Figure 5.63, and their replacement by

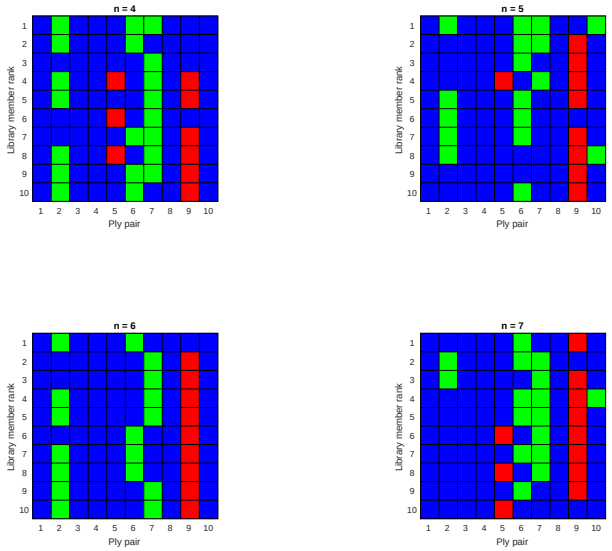


Figure 5.61: *Diagram like Figure 5.10, but for (PD1, RH, C2, RH, PD1, RH, C4, RH), with the plain condition set*

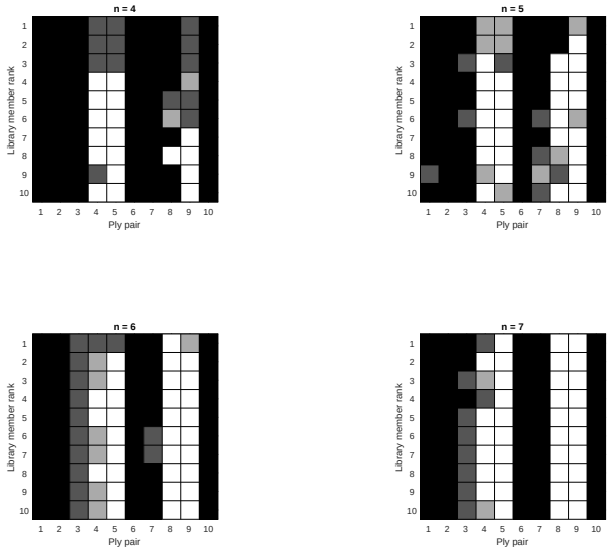


Figure 5.62: *Diagram like Figure 5.11, but for (PD1, RH, C2, RH, PD1, RH, C4, RH), with the plain condition set*

green ones could be achieved by altering thresholds.

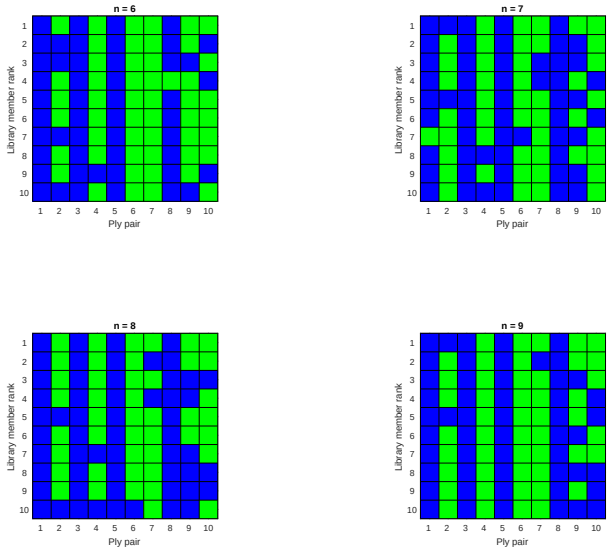


Figure 5.63: *Diagram like Figure 5.10, but for (PD1, RH, C2, RH, PD1, RH, C4, RH), with ec condition set*

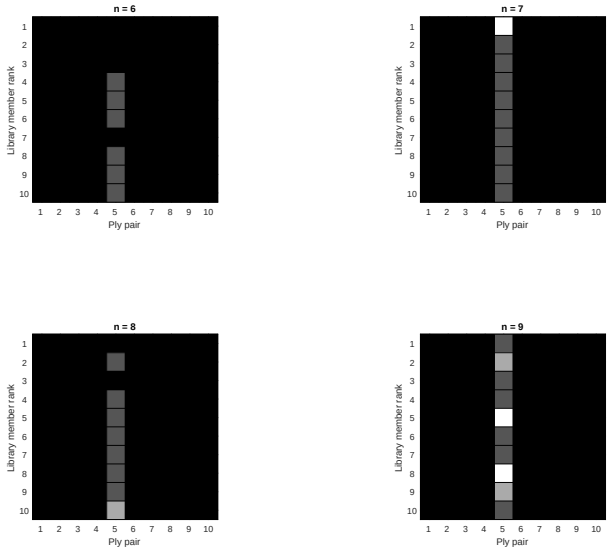


Figure 5.64: *Diagram like Figure 5.11, but for (PD1, RH, C2, RH, PD1, RH, C4, RH), with ec condition set*

ed

The double/overall success rates under the ed condition set are 36%/58% ($n = 8$) and 51%/65.5% ($n = 9$). However, if we ignore ply pairs 4 and 6 when $n = 8$, which are columns with selection failures throughout, the rates for $n = 8$ become 45% and 72.5%. If we also ignore

these columns in the ec diagrams, the resultant double/overall success rates become 40%/70% ($n = 6$), 36.25%/68.125% ($n = 7$), 33.75%/66.875% ($n = 8$), and 35%/67.5% ($n = 9$).

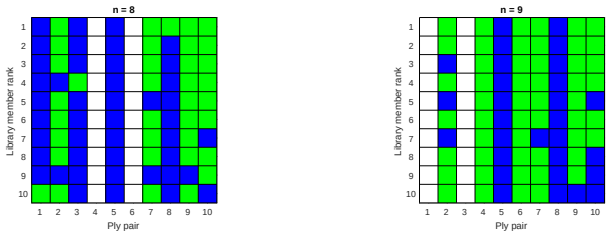


Figure 5.65: *Diagram like Figure 5.10, but for (PD1, RH, C2, RH, PD1, RH, C4, RH), with ed condition set*

Similarly, if we ignore ply pairs 1 and 3 when $n = 9$, which are all-white columns, the rates for this n become 63.75% and 81.875%. If we also ignore these columns in the ec diagrams, the resultant double/overall success rates become 63.75%/81.875% ($n = 6$), 56.75%/78.625% ($n = 7$), 55%/77.5% ($n = 8$), and 60%/80% ($n = 9$). Hence, if we ignore the columns which are all white in the ed, $n = 8$ case, both the double and overall success rates for the ed, $n = 8$ case beat the same rates for all the ec cases, and, if we do the same for the white columns in the ed, $n = 9$, the double and overall success rates for the ed, $n = 9$ case beat the same rates for all the ec cases except one, and, in the exception, the ed and ec rates are the same.

If we look at the Wilcoxon diagrams in Figure 5.66 and disregard the red columns, we again see lots of black — every square apart from those of column 5. This means that readjusting thresholds might improve classification rates.

Ply pair 5 is also exceptional for the ec method, in that the undesirable null hypotheses regarding the relative magnitudes of the tail medians are less strongly rejected for this pair than any other (excluding the columns of selection failure in the ed case).

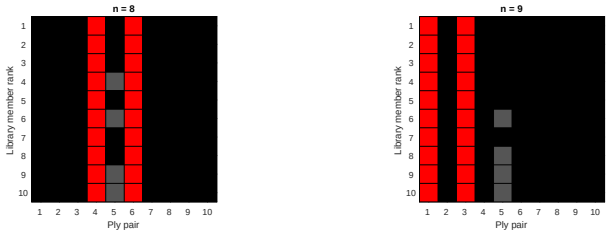


Figure 5.66: *Diagram like Figure 5.11, but for (PD1, RH, C2, RH, PD1, RH, C4, RH), with ed condition set*

Summary

Although the good performance on the (PD1, RH, C2, RH, PD1, RH, C2, RH) test suite does not fully generalise to the test suite here, (PD1, RH, C2, RH, PD1, RH, C4, RH), we do get moderately good results under the ec condition set, which are an improvement on those under the plain condition set. A further improvement under the ed condition set appears to be prevented by the high rates of selection failure (20%) under the latter set.

5.5.2 Test suite (PD1, LH, C2, LH, PD1, LH, C3, LH)

In this and the next section we will examine the generalisation of the classification based on the training suite (PD1, LH, C2, LH) to the classification of the left-hand data of PD1 and of control subjects other than C2. We start here with C3.

Of course, the maps ϕ which take square S_k of the ply-level classification diagrams of Figures 5.19, 5.22, and 5.25 to the corresponding squares of Figures 5.67, 5.69 and 5.71 below, have the same characteristics as those mentioned at the start of the previous section: if C is the map which takes a square to its colour, coded in the obvious way by W, G, B, Y and R, we have $C(\phi(\{S_k : C(S_k) = W\})) = \{W\}$, $C(\phi(\{S_k : C(S_k) = G\})) \subset \{G, B\}$, $C(\phi(\{S_k : C(S_k) = B\})) \subset \{G, B\}$, $C(\phi(\{S_k : C(S_k) = Y\})) \subset \{Y, R\}$ and $C(\phi(\{S_k : C(S_k) = R\})) \subset \{Y, R\}$.

Plain

We can immediately see from the ply-level classification diagrams for this condition set that the double success rate is 0% for all the values of n we consider.

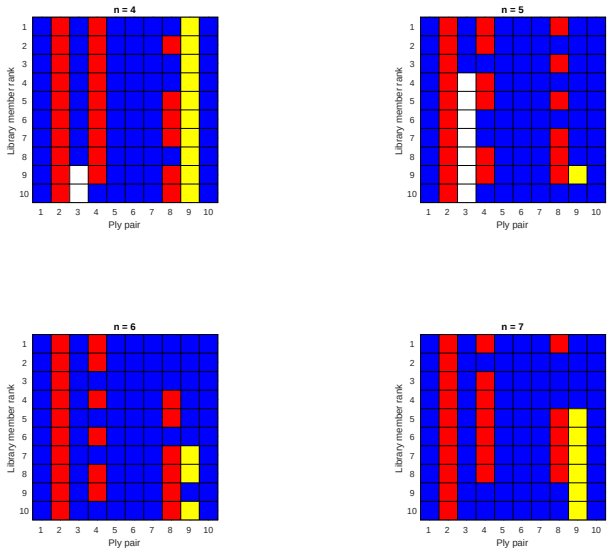


Figure 5.67: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, LH, C3, LH), with the plain condition set*

The overall classification rates are all at or below 50% as a consequence of the zero double success rate, and they are all in fact well below this level.

The paucity of grey and black squares in the Wilcoxon Figure 5.68 shows that the poor success rates are not merely the result of a poor choice of thresholds, but that the expected differences in the tails of the activation distributions for the PD1 and C3 data do not materialise, except for ply pairs 1 and 3 and a few other scattered cases.

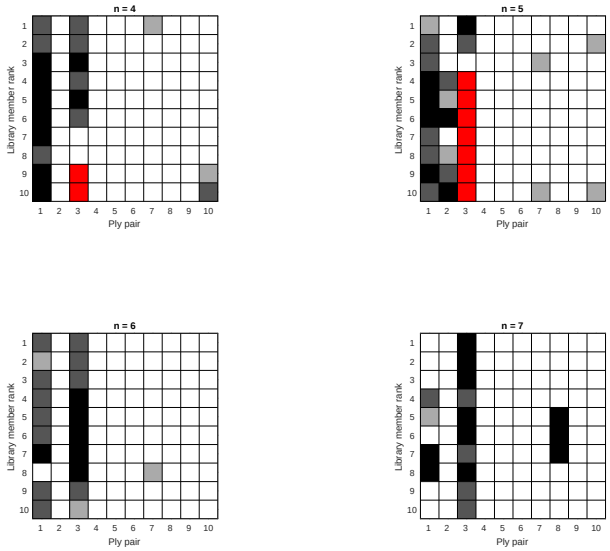


Figure 5.68: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, LH, C3, LH), with the plain condition set*

ec

As Figures 5.69 and 5.70 tell much the same story as Figures 5.67 and 5.68, despite a 1% rise in classification rates where there are no selection failures under the plain condition set, and a slightly greater rise elsewhere, the classification is poor and the difference in the tails is not great, except for a few ply pairs (with ply pair 10 being added to the list of these pair).

ed

In Figures 5.71 another pair of poor classification rates is demonstrated, made worse by 7 columns of selection failures, and in Figure 5.72 we see that this poor performance is mainly due to the distribution tails not being sufficiently distinct — with the exception of ply pairs 1 and 10.

Summary

Although there is a slight improvement in moving from the plain condition set to the ec condition set, the overall classification rate is very poor here.

This is a result

1. the classification of the PD1 members of the ply pairs not being outstanding, as it is unchanged from the test suite (PD1, LH, C2, LH, PD1, LH, C2, LH);

2. the green squares of the (PD1, LH, C2, LH, PD1, LH, C2, LH) ply-level classification Figures 5.19 (plain), 5.22 (ec) and 5.25 (ed) being mostly mapped to blue squares of

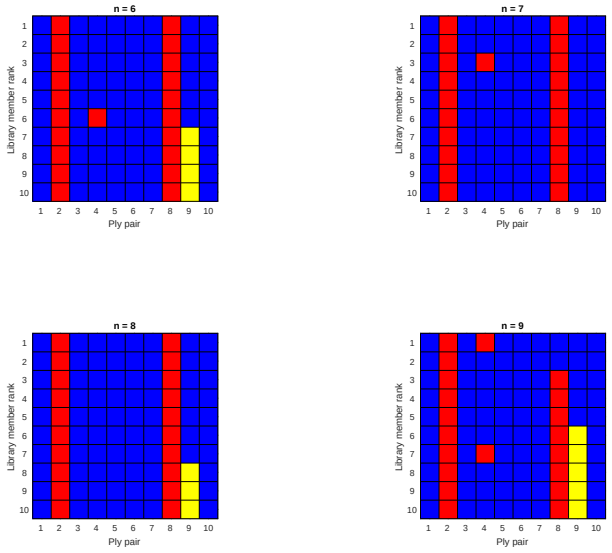


Figure 5.69: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, LH, C3, LH), with ec condition set*

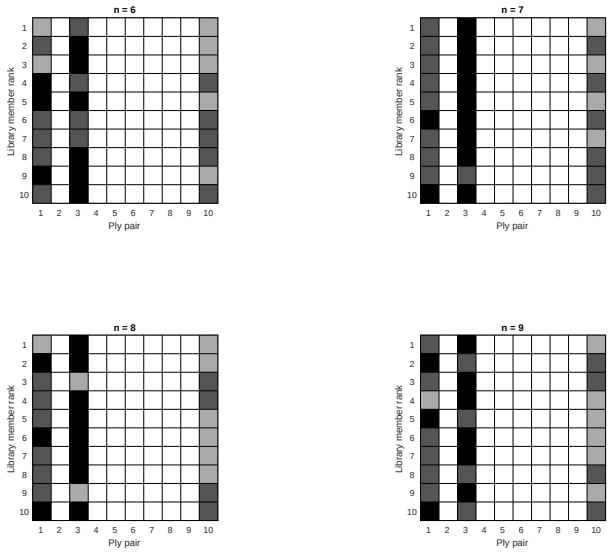


Figure 5.70: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, LH, C3, LH), with ec condition set*

the (PD1, LH, C2, LH, PD1, LH, C3, LH) ply-level classification Figures 5.67, 5.69 and 5.71 (i.e., correct C2 classifications being mapped to incorrect C3 classifications);

3. the yellow squares of Figures 5.19, 5.22 and 5.25 being mostly mapped to red squares of Figures 5.67, 5.69 and 5.71 (i.e., correct C2 classifications being mapped to incorrect

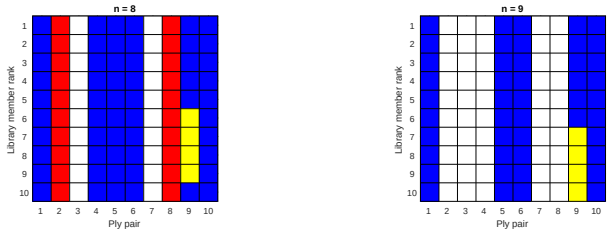


Figure 5.71: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, LH, C3, LH), with ed condition set*

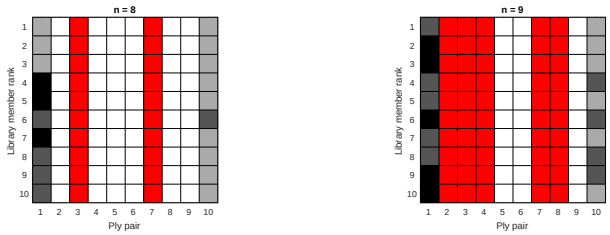


Figure 5.72: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, LH, C3, LH), with ed condition set*

C3 classifications again).

(There are a few exceptions to the above: e.g. for the plain condition set and $n = 4$, the red column for ply pair 9 is mapped to a yellow column.)

In other words, from the point of view of most of the classifiers based on the selected members, the C3 plies are closer to the PD1 training data than to the C2 training data.

5.5.3 Test suite (PD1, LH, C2, LH, PD1, LH, C4, LH)

The ply-level classification diagrams for the test suite (PD1, LH, C2, LH, PD1, LH, C4, LH), Figures 5.73 (plain), 5.74 (ec) and 5.75 (ed) are extremely similar to those for (PD1, LH, C2, LH, PD1, LH, C3, LH), Figures 5.67, 5.69 and 5.71, so that everything that was said about the classification above can also be said here, despite the fact that the Wilcoxon diagrams here, Figures 5.76 (plain), 5.77 (ec) and 5.78 (ed) are somewhat less similar to those for (PD1, LH, C2, LH, PD1, LH, C3, LH), Figures 5.68, 5.70 and 5.72. This is because the latter diagrams differ mainly from the former in having even fewer grey squares, and so even less chance of mitigating the poverty of the classification by choosing better thresholds.

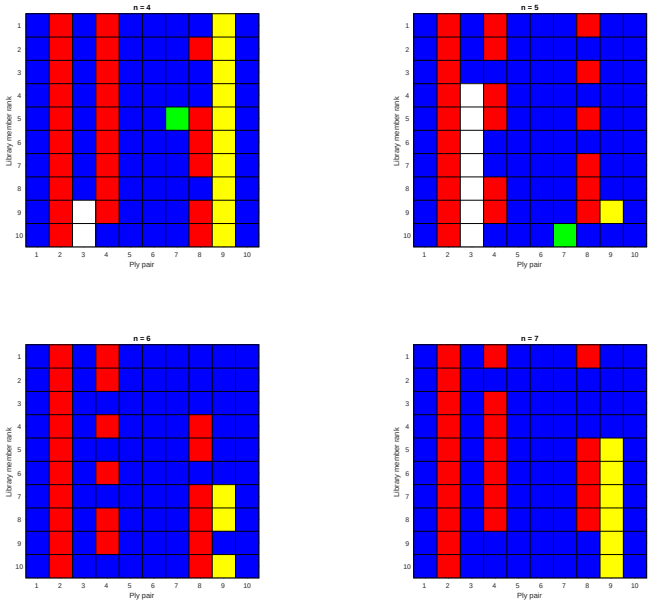


Figure 5.73: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, LH, C4, LH), with the plain condition set*

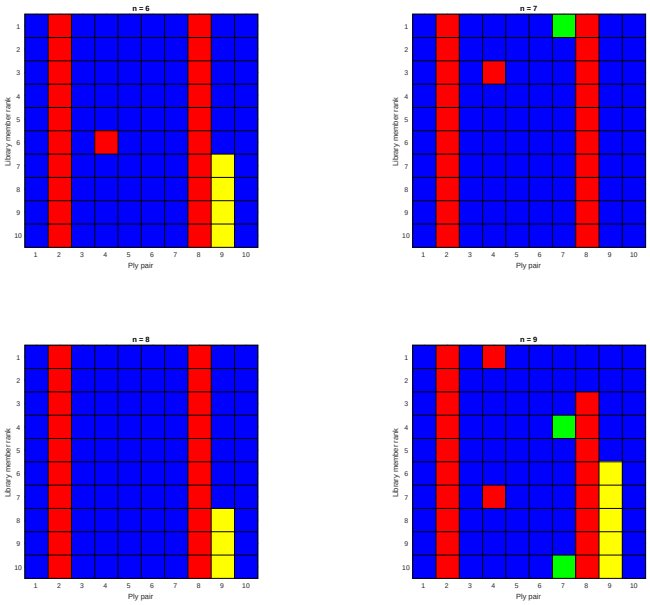


Figure 5.74: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, LH, C4, LH), with ec condition set*

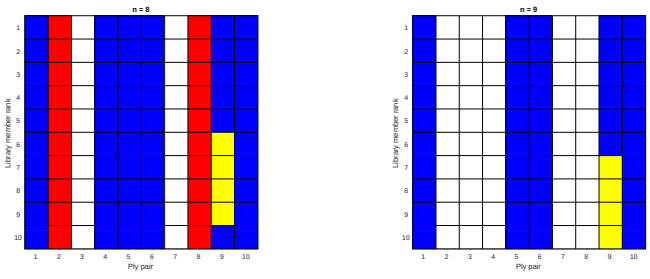


Figure 5.75: *Diagram like Figure 5.10, but for (PD1, LH, C2, LH, PD1, LH, C4, LH), with ed condition set*

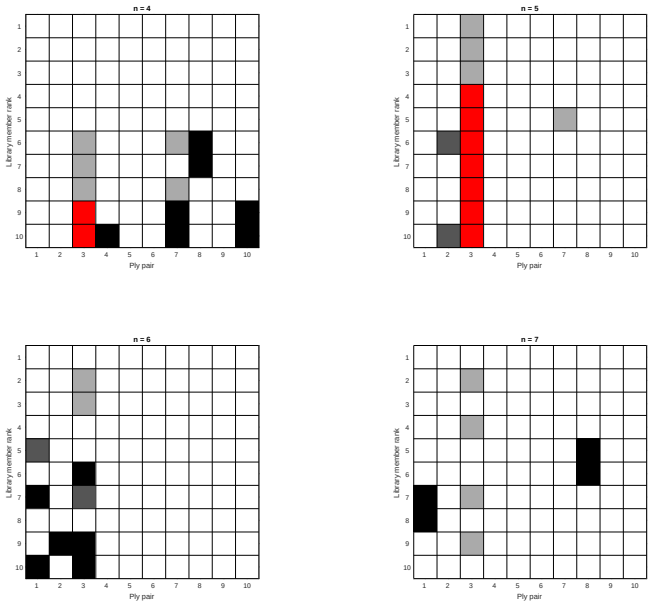


Figure 5.76: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, LH, C4, LH), with the plain condition set*

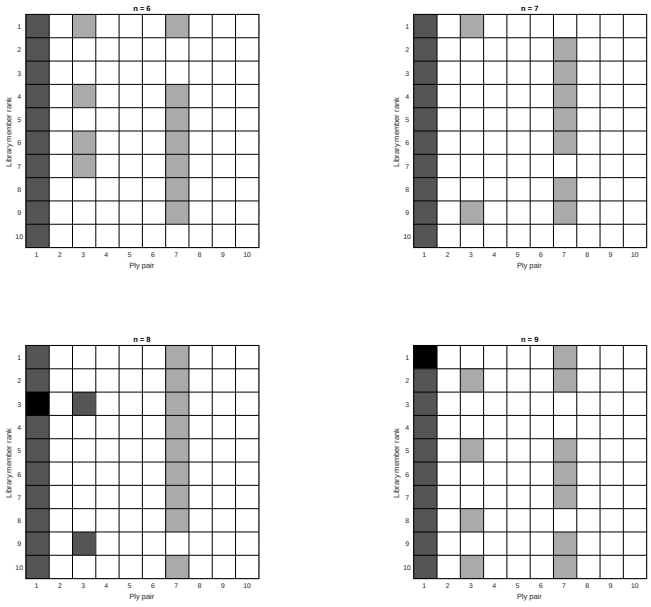


Figure 5.77: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, LH, C4, LH), with ec condition set*

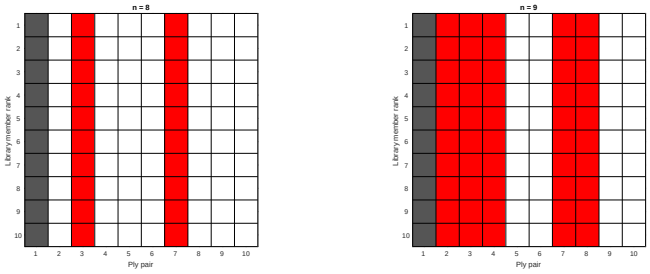


Figure 5.78: *Diagram like Figure 5.11, but for (PD1, LH, C2, LH, PD1, LH, C4, LH), with ed condition set*

5.6 Benchmark

We describe a benchmark in this section, and train it on (PD1, RH, C2, RH) data to classify unseen (PD1, RH) and (C2, RH) data, and on (PD1, LH, C2, LH) data to classify unseen (PD1, LH) and (C2, LH) data.

In order to obtain this benchmark for the classification performance, we exploit the excess activity with a frequency of around 6Hz which is typical of the motion of some PD patients.

To make a classifier, we automate the process of interpreting "about 6Hz" and "excess activity" — we automatically find a frequency band and a threshold on the proportion of the total power which is in the frequency band, where the "activity" in the band is excessive if its proportion of the total power exceeds the threshold.

For each of our epochs of the right-hand data, we take the raw single-sided Fourier transform of each channel of our acceleration, square its magnitude and sum it over the three channels, to yield a traces like the top diagram of Figure 5.79.[6] We then take this power spectrum, truncate it to run between 0.1 and 15Hz, and then divide by the sum of its values over this truncated range, to yield the normalised power spectra of the middle diagram. We call these $P_S(\omega; k, \mathrm{PD1})$ and $P_S(\omega; k', \mathrm{C2})$, where k and k' label the particular epoch.

Then, for each pair of numbers $\omega_L, \omega_U \in (0.1, 15)$ such that $\omega_U > \omega_L + 1$,[7] $R_S(\omega_L, \omega_U; k, \mathrm{PD1}) = \int_{\omega_L}^{\omega_U} P_S(\omega; k, \mathrm{PD1})\mathrm{d}\omega$, $R_S(\omega_L, \omega_U; k', \mathrm{C2}) = \int_{\omega_L}^{\omega_U} P_S(\omega; k', \mathrm{C2})\mathrm{d}\omega \in [0, 1]$.[8]

Consequently, for the ℓth ply pair (maintaining the same division of epochs into associated PD1 and C2 plies as before), we have the sets of real numbers in $[0, 1]$, $S(\ell; \omega_L, \omega_U, \mathrm{PD1}) = \{R_S(\omega_L, \omega_U; k, \mathrm{PD1}) : \text{epoch } k \text{ is a PD1 epoch not in the PD1 ply of pair } \ell\}$ and $S(\ell; \omega_L, \omega_U, \mathrm{C2}) = \{R_S(\omega_L, \omega_U; k, \mathrm{C2}) : \text{epoch } k \text{ is a C2 epoch not in the C2 ply of pair } \ell\}$.

For each pair (ω_L, ω_U) which coincides with the discrete points of the Fourier transform, we carry out a Wilcoxon test on the hypothesis that the median of the underlying distribution of the $S(\ell; \omega_L, \omega_U, \mathrm{C2})$ points is greater than or equal to the median of the $S(\ell; \omega_L, \omega_U, \mathrm{PD1})$, and then choose the pair $(\omega_L^*(\ell), \omega_U^*(\ell))$ which results in the rejection of this null hypothesis at the lowest p-level.

To obtain a threshold $\Theta(\ell)$ for the classification of the unseen data in the ℓth ply pair, we find the greatest integer $j \leq 50$ such that the $(50 + j)$th percentile of $S(\ell; \omega_L, \omega_U, \mathrm{C2})$ does not exceed the $(50 - j)$th percentile of $S(\ell; \omega_L, \omega_U, \mathrm{PD1})$ and set $\Theta(\ell)$ to the mean of these two percentiles.

Table 5.10 is the result of doing all this with the right-hand data, and the bottom diagram of Figure 5.79 shows our example epochs and the appropriate band for them.

We note that the results are fairly consistent, no matter which ply pair is excluded to be used as test data: 5.1–55.4Hz for ω_L^*, 6.15–6.55 for ω_U^*, band widths of 1 (the minimum we allowed)–1.15, and Θ from 0.042–0.048. We note that 6Hz is always found within the band,

[6] We have taken real but extreme data for illustrative purposes

[7] We add 1 to ensure that the resultant frequency band is not too narrow

[8] For speed of execution, as well as speed of development, we use the simplest possible numerical integration, one based on the cumulative sum

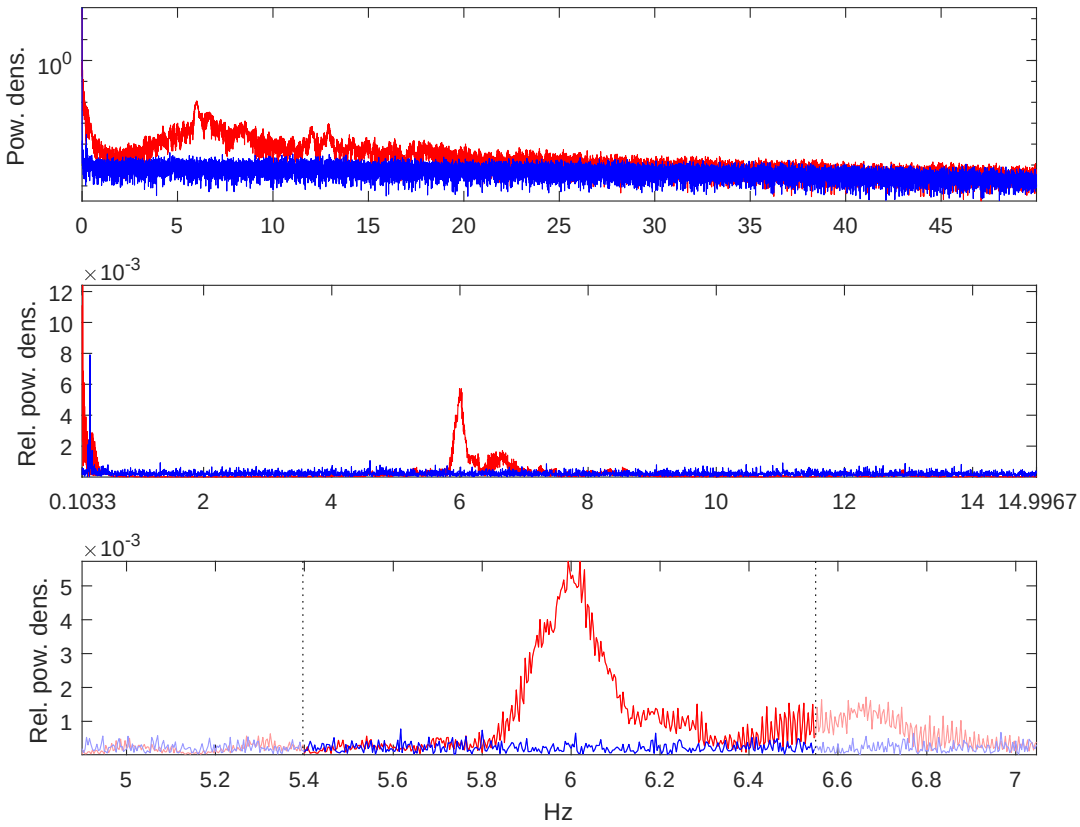


Figure 5.79: *Power spectra over sample epochs for benchmark processing. The blue trace is for a control epoch, the red for a PD one. In the lowest diagram, the dotted lines are for the frequency band* (ω_L^*, ω_R^*) *defined in the text. Note that the vertical scale in the top diagram is logarithmic, but the lower two are linear*

Excluded ply pair	ω_L^*	ω_U^*	Θ
1	5.15	6.15	0.044
2	5.15	6.15	0.044
3	5.43	6.55	0.046
4	5.43	6.55	0.046
5	5.43	6.55	0.046
6	5.40	6.55	0.048
7	5.41	6.41	0.042
8	5.43	6.55	0.047
9	5.15	6.15	0.044
10	5.15	6.15	0.043

Table 5.10: *Frequency bands and thresholds for benchmark — right hand*

and that epochs will be classified as PD if and only if the signal energy contained in a band of width roughly 1Hz amounts to more than about 4.5-5% of the total between 0.1 and 15Hz.

We now classify the kth PD1 (resp., C2) epoch of the right-hand ply pair ℓ as PD1 if $R_S(\omega_L^*(\ell), \omega_U^*(\ell); k, \text{PD1}) > \Theta(\ell)$ (resp., $R_S(\omega_L^*(\ell), \omega_U^*(\ell); k, \text{PD1}) > \Theta(\ell)$) and as C2 otherwise, and we classify the PD1 (resp., C2) member of the ℓth ply as PD1 if the median of the set $S(\ell; \omega_L^*(\ell), \omega_U^*(\ell), \text{PD1})$ (resp. $S(\ell; \omega_L^*(\ell), \omega_U^*(\ell), \text{C2})$) exceeds $\Theta(\ell)$, and as C2 otherwise.

When the ply-level classifications are carried out here, the result is that both the PD1 members and the C2 members of all the ply pairs are correctly classified, equivalent to a green row in the ply-level classification diagrams with respect to ECWTs of the rest of this chapter. The epoch-level classification rates are shown in Figure 5.80 below.

Excluded ply pair	ω_L^*	ω_U^*	Θ
1	7.19	8.29	0.016
2	7.18	8.28	0.016
3	7.19	8.28	0.015
4	7.72	8.74	0.014
5	7.19	8.28	0.015
6	7.19	8.29	0.016
7	7.19	8.28	0.015
8	7.18	8.31	0.016
9	7.05	8.27	0.017
10	7.19	8.28	0.015

Table 5.11: *Frequency bands and thresholds for benchmark — left hand*

Table 5.11 collates the same information for the left-hand data as Table 5.10 does for the right-hand data. Here, the lower boundary of the band is between 7.05 and 7.72HZ, the upper boundary is between 8.27 and 8.74Hz, and the width varies between 1.02 and 1.22Hz, with the threshold proportion of signal energy in the band being 1.4%-1.7%. We can see that the band is always above 6Hz.

As with the right-hand data, both the PD1 and C2 ply of every ply pair is correctly classified.

However, as the scatter diagram for the epoch-level success rates, Figure 5.80, shows, in neither the right- or left-hand case does the ply-level success rest on a very high rate of correct epoch-level classifications, even though these rates must be over 50% to account for the ply-level succeses. In fact, no epoch-level classification rate exceeds 80%, with some below 60%.

Although some of the combinations of n and condition sets for our algorithms found greater difficulty in classifying the PD1 member of ply pairs 5 and 9 within the (PD1, RH, C2, RH, PD1, RH, C2, RH) test suite, this is only reflected here by ply pair 9, which has a relatively low epoch-level classification rate in Figure 5.80 (9 towards the left among the blue numbers). Similarly, the difficulty found for the C2 member of ply 8 is not reflected here.

The difficulties found in the (PD1, LH, C2, LH, PD1, LH, C2, LH) find a greater echo here, as the difficult PD1 members of plies 2, 4 and 8 have their (red) numbers towards

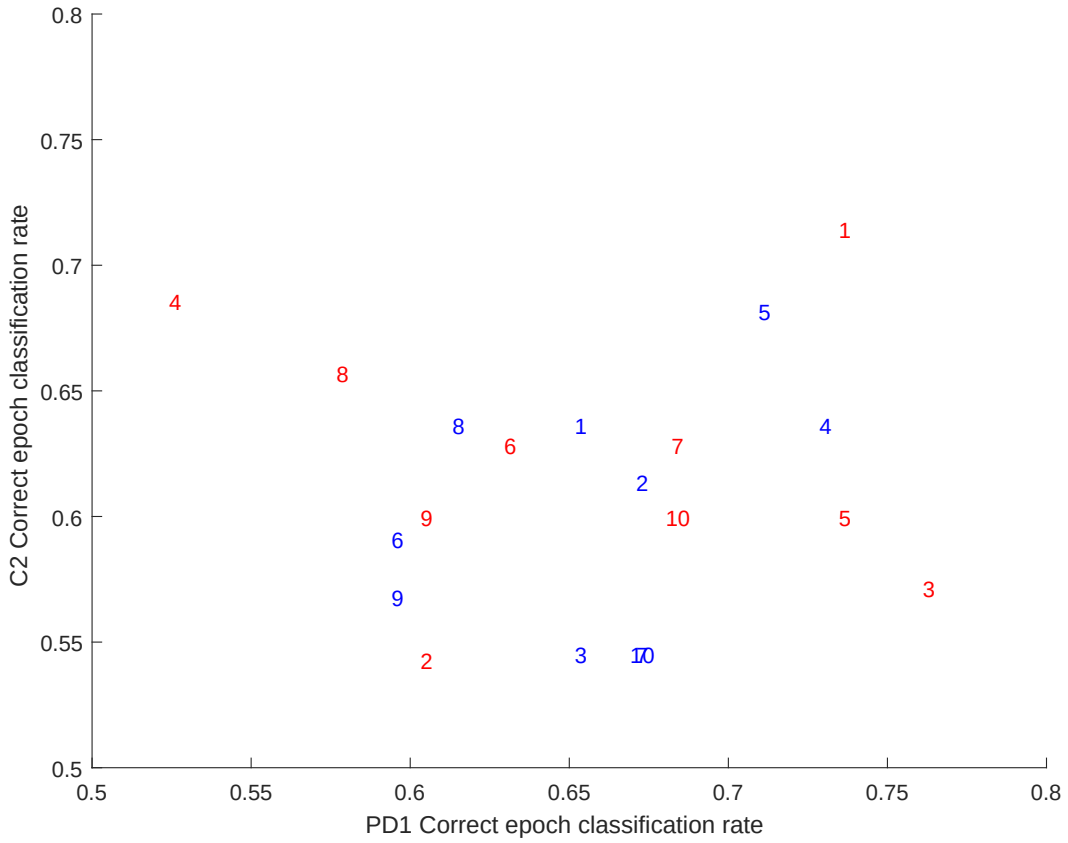


Figure 5.80: *Scatter diagram with the epoch-level classification rate for PD1 epochs in a ply pair being plotted against that for C2 epochs in the same ply pair. Each number represents its ply pair, the blue ones being for the right-hand data*

the left, and difficult C2 members of plies 3, and 9 have their numbers towards the bottom. However, the difficult C2 members of the plies 5 and 7 have higher epoch-level classification rates.

Also, for example, the left position of blue 6 in Figure 5.80 or the low position of red 2, does not reflect any difficulty in classifying the C2 member of ply 6 for the right-hand test suite, or of classifying the PD1 member of ply 2 for the left-hand test suite.

5.7 Summary

In this Chapter, we documented several experiments using our methods to extract libraries and to employ them to classify unseen data.

5.7.1 Member separation and tail thickness

We first noted that the library members extracted and selected were fairly well separated for all the training suites, degrees of freedom and condition sets we chose, and that there

was little systematic variation in this separation with degrees of freedom and condition sets. Despite the fact that there were statistically significant variations with training set, these variations were not large in absolute terms.

Support for the idea that the selected members are well separated comes from the epoch-level classification diagrams. When these are "U"-shaped, then either what is picked up by the high activations corresponding to different members is very similar, or frequently occurs in the same epoch. When the diagram is *not* "U"-shaped, neither is the case, so the different activations reflect different things. But across our experiments, there are often non-"U" diagrams.

When we looked at the relative thickness of the tails of the distributions of activations of the windows of the PD1 and C2 selection data with respect to the library members which were selected for each training suite, what we found was that, when this suite was (PD1, RH, C2, RH) or (C2, LH, PD1, LH), all the results were as expected (with one exception): where there were no selection failures, the tails of the distribution derived from the extraction data set were thickest.

However, this expectation was not fulfilled when the suite was (PD1, LH, C2, LH) (respectively (C2, RH, PD1, RH)): nearly all (respectively all) the contrast set distribution tails were thicker.

A possible explanation for these thwarted expectations is that the motion patterns which are present in the movement of C2's left hand are attenuated in the movement of PD1's left hand, and that patterns present in the movement of C2's right hand are exaggerated in the movement of PD1's right hand, reflecting the fact that PD1's left hand was the least affected at the time the data was collected. (NB, it is not suggested that the movement of either subject's right or left hand is dominated by these motion patterns, merely that the extent to which they are present differs the most between subjects.)

5.7.2 Recognition

We obtained a good classification performance for most of the library members selected for the (PD1, RH, C2, RH) training set when they were confronted with unseen (PD1, RH) and (C2, RH) data — in the case of the ec condition set, the classification of the unseen data was perfect for $n = 6, 7$ and 8. Selection failures prevented the ed selected members from achieving similar results, but co-opting ec results to cover the missing selections would have achieved these. Although the plain condition set results were not as good, their overall classification rates were 89%+. A possible improvement with the strictness of the imposed condition set may be hidden by the facts that the ed results are brought down by selection failures, and it is not possible to improve on the perfection of the ec results. However, this perfection is only through the prism of the of the ply-level classification. If we look at the Wilcox diagrams, there are more black squares for ply pair 5 (the only ply pair which has grey or white squares in the ec diagrams) in the ed diagrams.

For the (PD1, LH, C2, LH) training set operating on the unseen (PD1, LH) and (C2, LH)

data, the classification results were not as good, although they did improve as the condition set was changed from plain to ec (this improvement may have been exaggerated by selection failures in the case of the plain condition set with $n = 4$ and 5, but there were no other such failures, so the increase with a constant number of degrees of freedom — from plain, $n = 6$ or 7, to ec, $n = 8$ or 9 — or constant n — i.e., $n = 6$ or 7, was real). Any further improvement under the ed condition set was hidden by a large number of selection failures there. Moreover, the picture is not much changed by looking at the Wilcoxon diagrams.

As we had begun to suspect that the relative failure of the results of (PD1, LH, C2, LH) training was due to attenuation of normal modes of movement in the LH PD1 data, rather than the strong presence of abnormal patterns in it, we decided to investigate the training suite (C2, LH, PD1, LH) and its ability to classify unseen (PD1, LH) and (C2, LH) data.

Although the ply-level classification results were slightly better, in that the best double and overall classification rates were 58% and 79% for $n = 8$ and 9 under the ed condition set (cf best double rate of 57% for ec, $n = 6$ and best overall rate of 77%, for ec, $n = 7$ for the (PD1, LH, C2, LH) training suite), the clarity of the result is muddied by the fact that, of the 4 combinations of condition set and n where both the (PD1, LH, C2, LH) and (C2, LH, PD1, LH) training suites are unaffected by selection failures (they happen to be the ec combinations), the former yields better results for overall success rates in two cases, and for double success rates in one.

If we wish to restrict ourselves to looking at cases without selection failures within the (C2, LH, PD1, LH, C2, LH, PD1, LH) test suite, we can only compare the different condition sets plain and ec: plain, $n = 6$ and $n = 7$ with ec, $n = 6$ and $n = 7$ for the comparison with constant n, and with ec, $n = 8$ and $n = 9$ for the comparison with constant degrees of freedom ($p = 5$ and $p = 6$). In both comparisons, there is an increase in performance in both double success rates and overall ones in going from the plain to the ec condition set.

The Wilcoxon diagrams differ very little from the corresponding diagrams for the (C2, LH, PD1, LH) training suite.

For completeness, we looked at the (C2, RH, PD1, RH, C2, RH, PD1, RH) test suite. The results are, as expected, not quite as good as those for the (PD1, RH, C2, RH, PD1, RH, C2, RH) suite, presumably because motion patterns characteristic of PD are excluded from the outset. Nevertheless, the results are better than those of both the test suites for LH data.

5.7.3 Generalisation

Across attachment

When we attempted generalisation across attachment, we found that the generalisation was poor, without much scope for correction by choosing better thresholds[9], and that this was mostly due to the misclassification of the PD1 members of the ply pairs.

In other words, from the point of view of classifiers built on our selected members,

[9] As evidenced by the relevant Wilcoxon diagrams

1. LH PD1 data is closer to RH C2 data than it is to RH PD1 data;
2. but LH C2 data is closer to RH C2 data than it is to RH PD1 data,

in terms of characteristics extracted from RH PD1 data and chosen to be infrequent in RH C2 data, and

1. RH PD1 data is closer to LH C2 data than it is to LH PD1 data;
2. but RH C2 data is closer to LH C2 data than it is to LH PD1 data

in terms of characteristics extracted from LH PD1 data and chosen to be infrequent in LH C2 data. The second pair of conclusions is less significant here, as

1. the classifiers resulting from the (PD1, LH, C2, LH) training suite are not as good in classifying LH PD1 and LH C2 data as are those from (PD1, LH, C2, LH) training suite in classifying even RH PD1 and RH C2 data;
2. there are more exceptions to the inability of the classifiers resulting from the (PD1, LH, C2, LH) training suite to classify PD1 RH and C2 RH data.

Across control subjects

When we examine the generalisation from classifiers built on right-hand PD1 and C2 data to the task of distinguishing unseen right-hand PD1 and *C4* data, we find that the generalisation is quite poor under the plain condition set, but that there is quite an improvement in moving to the ec set, and that this improvement could well be maintained for the ed set, were it not for the number of selection failures under the latter. We have also seen that the Wilcoxon diagrams point towards some of the deficiencies being remediable by changing thresholds.

It is a different story with the generalisation of left-hand PD1 and C2 data to the job of distinguishing left-hand PD1 and C3 or C4 data. In addition to the relatively poor performance in classifying the LH PD1 data in the test suite (PD1, LH, C2, LH, PD1, LH, C2, LH) compared to its right-hand counterpart, we now have a very poor performance in classifying the LH C3 or C4 data. Nor is there much evidence that this is due merely to a poor choice of thresholds for separating different activation distributions.

It seems that, from the perspective of the training suite (PD1, LH, C2, LH), the C3 and C4 left-hand data is more similar to the LH PD1 data than to the LH C2 data.

5.7.4 Improvement of performance with strictness of condition sets

In Chapter 3, we examined the blurring effect that the limited bandwidth of the NAT accelerometer has on the ECWTs we are employing, and discovered that this effect reduces as we impose stricter conditions on the ECWTs (because these force a reduction in the higher frequency content of the polynomials underlying the ECWTs).

We now briefly examine the practical outcome of diminishing this effect through continuity and differentiability conditions. In Table 5.12 we collate the double classification success

excluding the selection failures[10], and, for each section of constant degrees of freedom and test suite where the rates increase, at least monotonically, we have chosen a bold font.

			Test suite								
p	**Cond. set**	n	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
3	**plain**	4	**84**	*35.7*	**65**	***48.8***	0	*8.7*	20	*0*	*19*
	ec	6	**100**	57	***86.7***	**56**	0	4	51	0	51
	ed	8	***100***	*43.8*	**97**	**58**	*0*	*1.2*	*45*	*0*	*45*
4	**plain**	5	**80**	*58.1*	**67**	*53.3*	0	*1*	**17**	*0*	***17***
	ec	7	**100**	54	***80***	51	0	0	**46**	0	**44**
	ed	9	***100***	*68*	**95**	58	*0*	*8*	***63.8***	*0*	***63.8***
5	**plain**	6	**81**	**51**	**68**	***49.4***	0	**1**	**17**	0	**17**
	ec	8	**100**	**55**	**85**	**54**	0	**2**	**44**	0	**44**
6	**plain**	7	**79**	**47**	**64**	***46.7***	0	**3**	**16**	0	**16**
	ec	9	**92**	**52**	**79**	**54**	0	**4**	**48**	0	**48**

Table 5.12: *Double success percentages for test suites 1) (PD1, RH, C2, RH, PD1, RH, C2, RH), 2) (PD1, LH, C2, LH, PD1, LH, C2, LH), 3) (C2, RH, PD1, RH, PD1, RH, C2, RH), 4) (C2, LH, PD1, LH, PD1, LH, C2, LH), 5) (PD1, RH, C2, RH, PD1, LH, C2, LH), 6) (PD1, LH, C2, LH, PD1, RH, C2, RH), 7) (PD1, RH, C2, RH, PD1, RH, C4, RH), 8) (PD1, LH, C2, LH, PD1, LH, C3, LH) and 9) (PD1, LH, C2, LH, PD1, LH, C4, LH), relative to number of selection successes per combination of test suite, condition set and n. Values affected by selection failures are in italics, and sections where the test suite and p are constant and the rate monotonically increases with strictness of condition set are in bold — although we exclude constant zero cases*

Over all the sections, the number of cases where we find monotonic increases is 55.6% of the total of 36. If we restrict attention to the training suites 1–4, corresponding to recognition, the monotonically increasing cases are 81.2%, and, if we further restrict attention to the sections with three condition sets, which obviously constitute a stricter test, we find monotonic increases in 62% of the cases.

When we do the same thing, but defining sections with constant n rather than p,[11] as in Table 5.13, we find that that the increasing cases are 69.4% of the total 36, and looking at training suites 1–4 again, they are 93.8% of the total of 16 (obviously, all sections have two condition sets here!)

We can repeat this exercise with the overall success rate rather than the double, and this is what we do in Tables 5.14 and 5.15.

The percentages of the sections with constant p (in Table 5.14) where the overall success rate monotonically increases with the strictness of the conditions are 66.7% (all sections), 93.8% (sections with test suites 1–4) and 87.5% (three-condition sections with test suites 1–4).

[10] This seems more apposite when we are examining the possible improvement of the results with increasing condition strictness, rather than claiming success

[11] We drop the results for the plain condition set with $n = 4$ or 5, as they have no ec or ed partner with the same n

		Test suite								
n	**Cond. set**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
6	**plain**	**81**	**51**	**68**	***49.4***	0	**1**	**17**	0	**17**
	ec	**100**	**57**	***86.7***	**56**	0	**4**	**51**	0	**51**
7	**plain**	**79**	**47**	**64**	***46.7***	0	3	**16**	0	**16**
	ec	**100**	**54**	***80***	**51**	0	0	**46**	0	**44**
8	**ec**	**100**	55	**85**	**54**	0	2	**44**	0	**44**
	ed	***100***	*43.8*	**97**	**58**	*0*	*1.2*	***45***	*0*	***45***
9	**ec**	**92**	**52**	**79**	**54**	0	**4**	**48**	0	**48**
	ed	***100***	***68***	**95**	**58**	*0*	***8***	***63.8***	*0*	***63.8***

Table 5.13: *Re-ordered version of Table 5.12*

			Test suite								
p	**Cond. set**	n	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
3	**plain**	4	**92**	***62.2***	**79.5**	***74.3***	**42**	*42.3*	55.5	*37.2*	*55*
	ec	6	**100**	**76.5**	***93.3***	**78**	**50**	40.5	75.5	39.5	75.5
	ed	8	***100***	***78.1***	**98.5**	**79**	***50***	*43.1*	*72.5*	*37.5*	*72.5*
4	**plain**	5	**90**	***76.3***	**81**	*76.1*	**34.5**	*48.9*	**54**	***38.2***	***54.5***
	ec	7	**100**	**77**	***79.4***	75.5	**49**	42.5	**73**	**39.5**	**72**
	ed	9	***100***	***80***	**97.5**	79	***49.4***	*44*	***81.9***	***50***	***81.9***
5	**plain**	6	**90.5**	**74**	**80**	***74.7***	**35.5**	49	**54**	**39**	**54**
	ec	8	**100**	**75.5**	**92**	**77**	**49.5**	40	**72**	**40**	**72**
6	**plain**	7	**89**	**52**	**81**	***73.4***	**33.5**	51	**52.5**	**39**	**52.5**
	ec	9	**96**	**70.5**	**89.5**	**77**	**46**	44	**74**	**40**	**74**

Table 5.14: *Overall percentage success rates for test suites labelled as in Table 5.12, relative to number of selection successes per combination of test suite, condition set and n*

When we re-order the data of Table 5.14 by n in Table 5.15, we obtain the following percentages for the monotonically increasing sections: 88.9% (all sections) and 93.8% (sections for test suites 1–4).

We think all this constitutes evidence for a strong tendency for the efficacy of the method improving as the strictness of the condition set increases.

		Test suite								
n	**Cond. set**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
6	**plain**	**90.5**	**74**	**80**	***74.7***	**35.5**	49	**54**	**39**	**54**
	ec	**100**	**76.5**	***93.3***	**78**	**50**	40.5	**75.5**	**39.5**	**75.5**
7	**plain**	**89**	**52**	81	***73.4***	**33.5**	51	**52.5**	**39**	**52.5**
	ec	**100**	**77**	*79.4*	**75.5**	**49**	42.5	**73**	**39.5**	**72**
8	**ec**	**100**	**75.5**	**92**	**77**	**49.5**	**40**	**72**	40	**72**
	ed	***100***	***78.1***	**98.5**	**79**	***50***	***43.1***	***72.5***	*37.5*	***72.5***
9	**ec**	**96**	**70.5**	**89.5**	**77**	**46**	**44**	**74**	**40**	**74**
	ed	***100***	***80***	**97.5**	**79**	***49.4***	***44***	***81.9***	***50***	***81.9***

Table 5.15: *Re-ordered version of Table 5.14*

5.7.5 The benchmark: classification or characterisation?

The results for the benchmark set a very high bar for the performance of a classifier, at least on this very limited data.

Although we generate a different set of up to ten classifiers for each ply pair, if the Kohavi[46] methodology is correct, we should get very similar results if we repeated the same experiment with ten sets of entirely independently collected training and test data of similar size (from the same sources).

What we have done is produce (up to 10) classifiers $\mathcal{C}_{k,\mathcal{S}_{R,\ell}}$, based on tRaining data $\mathcal{S}_{R,\ell} = \cup_{r\in[\{1,2,...,10\}-\{\ell\}]}\mathcal{S}_{E,r}$, where k indexes the individual classifier, for each r, and applied them to the tEst data $\mathcal{S}_{E,\ell}$ (we abstract from the internal organisation of these collections of data here). Assuming the Kohavi method can be applied here, we should obtain results statistically similar to producing classifiers $\mathcal{C}_{k,\mathcal{S}_R}$ in the same way from a set of training data $\mathcal{S}_R$, of the same size as the $\mathcal{S}_{R,\ell}$, but collected independently from the $\mathcal{S}_{E,\ell}$, and applying these to the $\mathcal{S}_{E,\ell}$.

If such classifiers $\mathcal{C}_{k,\mathcal{S}_R}$ exist for some set of training data $\mathcal{S}_R$ and our results can be applied to them, then, what we have shown with the benchmark is that they can only compete (as least as far as data from our sources is concerned) in the case of the training suite (PD1, RH, C2, RH) being applied to the task of classifying unseen (PD1, RH) and (C2, RH) data.

But should our method primarily be used as a classifier in competition with other approaches? It is designed to pick out different patterns of motion which are characteristic of PD, but need not be present at all times — indeed, it would be better if they were not present at all times, but reflected the development of the disease and/or the place in the medication cycle.

Moreover, it could be used in conjunction with other methods to diagnose PD, rather than in competition with them.

Any success it has in classification should perhaps best be seen as supporting evidence for its ability to pick out different patterns of motion which are indeed characteristic of PD.

5.7.6 Composite classifiers

However, if we insist on building a tool based on the method to diagnose PD in the first place, the combination of results from more than one library member may provide the best results.

We can regard each of the ECWTs determined by a row and column of our ply-level classification diagrams as part of a composite classifier. For example, a ply can be classified as PD if at least one of the ECWTs represented in its column classifies the ply as PD, and as a control otherwise (except when there are *no* selected members for that ply, when we will regard the composite classifier as undefined).[12]

[12]Here, we do not wish to argue too much with the benefit of hindsight — we could have chosen this example in advance, as it is sensible to at least consider classifying a subject as PD if they display any one of a set of motion patterns characteristic of PD

In terms of the colours of the squares in the ply-level classification, this composite classifier will correctly classify the PD member of the ply pair if there is a green or blue square in the pair's column, and will classify the control member correctly if there are no blue or red squares in the pair's column (and at least one square which is not white).

If the result of this composite classification is represented in a kind of summary row for the ply pairs, whose squares are colour-coded in the same way as the original classification diagrams, then

summary square white $\Leftrightarrow$ entirely white column;

summary square green $\Leftrightarrow$ at least one correct PD1 classification in column
and no incorrect C2 classifications in column;
$\Leftrightarrow$ at least one green or blue square in column
and no blue or red squares in column;
$\Leftrightarrow$ at least one green square in column containing only green, yellow or white squares;

summary square blue $\Leftrightarrow$ at least one correct PD1 classification in column
and at least one incorrect C2 classifications in column;
$\Leftrightarrow$ at least one green or blue square in column
and at least one blue or red square in column;
$\Leftrightarrow$ at least one blue square in column
or at least one green and at least one red square in column;

summary square yellow $\Leftrightarrow$ no correct PD1 classification in column
and no incorrect C2 classifications in column;
$\Leftrightarrow$ no green or blue square in column
and at no blue or red square in column;
$\Leftrightarrow$ at least one yellow square in column consisting of yellow and white squares;

summary square red summary square not white, green, blue or yellow.

For each test suite, we will have one of these summary rows for each combination of n and condition set, and we assemble these 10 combinations for each of our recognition test suites in Figure 5.81 (note that we have *not* built a combination classifier for each row of the diagrams in Figure 5.81, but rather have a recipe for building a classifier from data excluding that contained in the ply pair corresponding to each square in the row, which results in the classification of the ply pair illustrated. It is the success rate — or otherwise — of the recipe that is recorded in the diagrams.)

As we reverse the roles of the PD1 and C2 data in the two test suites (C2, RH/LH, PD1, RH/LH, C2, RH/LH, PD1, RH/LH), we could equally well argue that the combination policy should be inverted for these test suites — a ply is classified as PD if all of the ECWTs

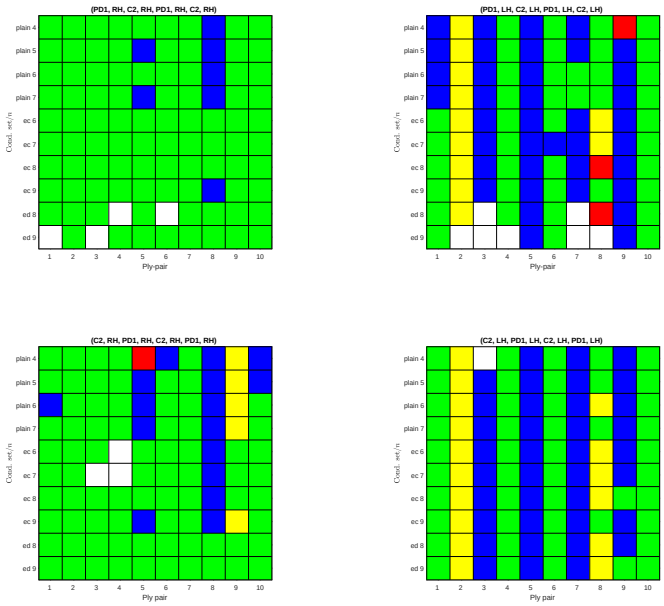


Figure 5.81: *Ply-level classification diagrams for combination classifiers*

represented in its column classify the ply as PD, and as a control otherwise (with obvious modifications if there are no coloured squares in the column).

Then the map from column to summary square satisfies

summary square white $\Leftrightarrow$ entirely white column;

summary square green $\Leftrightarrow$ at least one green square in a green, blue and white column;

summary square blue $\Leftrightarrow$ blue square in blue, green and white column;

summary square yellow $\Leftrightarrow$ at least one yellow square in column
or at least one red square and at least one green square in column;

summary square red summary square not white, green, blue or yellow.

The results for applying this "inverse" method to the two test suites (C2, RH/LH, PD1, RH/LH, C2, RH/LH, PD1, RH/LH) are depicted in Figure 5.82, and the double/overall success rates are collated in Table 5.16.

As we already had perfect classification results for right-hand data, we focus on the left-hand data results here: under the "inverse" version of the combination, with libraries extracted from C2 data and using PD1 data as a contrast, we obtain double/overall success rates of 70%/85% with the ec condition set and $n = 6$, and with the ed condition set and $n = 8$ (which means the degrees of freedom are the same: 3). However, this is something of a "cherry-picked" result: with so little data behind it, it should be treated as the basis for a hypothesis that such results are reliably derivable.

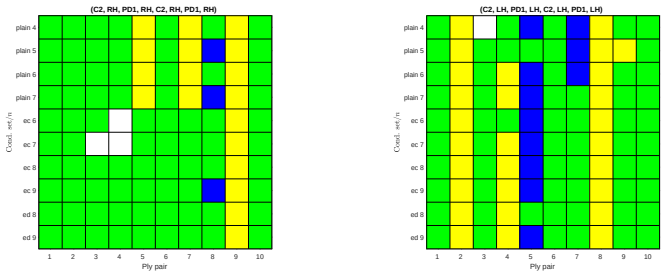


Figure 5.82: *Ply-level classification diagrams for "inverse" combination classifiers*

Nevertheless, it is certainly possible that a combined method could compete with the benchmark for left-hand data of the kind we have collected.

	Double success rate			Overall success rate		
Test suite	**min.**	**max.**	**Best methods**	**min.**	**max.**	**Best methods**
1	80%	100%	ec, $n = 6, 7, 8$	80%	100%	ec, $n = 6, 7, 8$
2	30%	50%	plain, $n = 6, 7$; ec, $n = 9$	40%	75%	plain, $n = 6, 7$; ec, $n = 9$
3	50%	100%	ed, $n = 8, 9$	70%	100%	ed, $n = 8, 9$
4	40%	50%	plain, $n = 4, 5, 7$; ec, $n = 9$; ed, $n = 9$	70%	75%	plain, $n = 5, 7$; ec, $n = 9$; ed, $n = 9$
5	60%	90%	ec, $n = 8$; ed, $n = 8, 9$	75%	95%	ec, $n = 8$; ed, $n = 8, 9$
6	50%	70%	ec, $n = 6$; ed, $n = 8$	70%	85%	ec, $n = 6$; ed, $n = 8$

Table 5.16: *Combined classifier success rates. Test suites are: 1) (PD1, RH, C2, RH, PD1, RH, C2, RH); 2) (PD1, LH, C2, LH, PD1, LH, C2, LH); 3) (C2, RH, PD1, RH, PD1, RH, C2, RH); 4) (C2, LH, PD1, LH, PD1, LH, C2, LH); 5) (C2, RH, PD1, RH, PD1, RH, C2, RH), (inverse); 6) (C2, LH, PD1, LH, PD1, LH, C2, LH) (inverse)*

We expect our individual classifiers to fail when the pattern of motion they are modelled on is not present in the data, and we also expect that these patterns are not ever-present — the point of these classification experiments, despite their form, is to establish that the patterns are frequently in the data. For this reason, we have avoided giving precision and sensitivity figures for these individual classifiers.

However, for the combined classifiers, which are actually supposed to act as classifiers, we will not be so shy.

We will use conservative definitions for precision and recall, defining a false positive as a failure to find a positive, even if this is the result of selection failure leading to a lack of a classifier, and similarly, a false negative is a failure to find a negative.

Then precision is

$$P = \frac{t_p}{t_p + f_p} = \frac{G + B}{G + B + Y + R + W} = \frac{G + B}{10} \tag{5.1}$$

and sensitivity

$$S = \frac{t_p}{t_p + f_n} = \frac{G + B}{G + 2B + R + W}, \tag{5.2}$$

where t_p is the number of correctly identified PD1 plies, t_n is the number of correctly identified C2 plies, f_p is the number of incorrectly identified PD1 plies and f_n is the number of incorrectly identified C2 plies, G is the number of green squares, and similarly for the other colours. These formulæ work because each green square represents a contribution of 1 to t_p and 1 to t_n, etc. As each row represents a combined classifier, the formulæ are applied across the rows.

We then find that the triplet $(\min\{P\}, \text{median}\{P\}, \max\{P\})$ is (0.8, 1, 1) for the test suite (PD1, RH, C2, RH, PD1, RH, C2, RH), (0.5, 0.8, 0.9) for (PD1, LH, C2, LH, PD1, LH, C2, LH), (0.8, 0.9, 1) for (C2, RH, PD1, RH, PD1, RH, C2, RH) and (0.8, 0.8, 0.9) for (C2, LH, PD1, LH, PD1, LH, C2, LH), for the combined classifier of Figure 5.81.

The triplet $(\min\{S\}, \text{median}\{S\}, \max\{S\})$ is (0.8, 0.909, 1) for (PD1, RH, C2, RH, PD1, RH, C2, RH), (0.417, 0.655, 0.692) for (PD1, LH, C2, LH, PD1, LH, C2, LH), (0.667, 0.784, 1) for (C2, RH, PD1, RH, PD1, RH, C2, RH) and (0.667, 0.680, 0.727) for (C2, LH, PD1, LH, PD1, LH, C2, LH), for the same combined classifier.

For the inverse combined classifier, we have precisions (0.7, 0.75, 0.9) for (C2, RH, PD1, RH, PD1, RH, C2, RH) and (0.6, 0.7, 0.7) for (C2, LH, PD1, LH, PD1, LH, C2, LH), and sensitivities (0.778, 0.95, 1) for (C2, RH, PD1, RH, PD1, RH, C2, RH) and (0.7, 0.875, 1) for (C2, LH, PD1, LH, PD1, LH, C2, LH).

These are not bad numbers, but clearly the perfect benchmark beats most of them.

Chapter 6

Conclusions

It is important to recognise that, given the limited amount of data available to us, any evidence presented here can only be towards a proof of concept.

6.1 Hypotheses

We restate our main hypothesis here for convenience:

1. it is possible, using the techniques to be developed, to distinguish between data stemming from PD patients and controls on the basis of patterns of motion;

2. these patterns of motion can be captured sufficiently well by the NAT and CWT analysis of the resultant data;

3. these patterns of motion are diverse enough to be easily distinguished from each other, and frequently occur in the data.

Obviously, our experiments test items 1 and 2 together.

The classification results in the previous chapter show that it is indeed possible to distinguish between unseen data stemming from one PD control and one control subject, and in the case of data stemming from the right-hand attachment, to do this easily, provided that our method is trained on data stemming from the same subjects. Within the limitations of this proviso, this is evidence for items 1 and 2.

Further evidence for these items comes from the experiment examining the generalisation from the right-hand data of PD1 and C2 to the classification of unseen PD1 and C4 right-hand data. Although the classification is poor here, compared to the recognition experiment carried out with the same training suite, the Wilcoxon diagrams for the generalisation experiment are dominated by black squares, indicating that the tails of the distributions are dissimilar, and that the thresholds used to separate them could be adjusted to produce a considerably better classification performance.

The results for the same items from the experiments involving left-hand data are not as good, but we have a tentative explanation, in terms of a characterisation of PD1, for this, which we shall give below.

As for item 3, we have two main pieces of evidence that it holds — the actual separation of the selected library members for each combination of training suite, n and condition set, which is reasonably large, and documented in section 5.2.1; and the existence of non-"U" profiles for the classification diagrams throughout the previous chapter, at least for some of the experiments documented there.

We also drew up a subsidiary hypothesis, also given here:

> increasing the strictness of the conditions set on smoothness improves the results of applying the method.

It would have been nice to carry out the same experiments for n = 8–11 for the ed condition set as we did for $n = 4$–7 for the plain set and $n = 6$–9 for the ec set, so we could more thoroughly examine this hypothesis, but time and resource pressures meant that we could only do this for ed $n = 8$ and 9.

However, we do have some good evidence that this hypothesis holds, at least as a tendency — taking the three-condition set sections of Tables 5.12 and 5.14 (where the sections are defined by constant p), restricted to the recognition test suites (which have comparatively good results, which presumably means that results here are less distorted by interference from any other factor which might stop the method from working), we find that the double (respectively, overall) success rates increase with the strictness of the condition in 62.5% (respectively 87.5%) of cases.

From the sections of Tables 5.13 and 5.15 (where the sections are defined by constant N), again restricted to the recognition test suites, we find that the double (respectively, overall) success rates increase with the strictness of the condition in 93.8% of cases, for both double and overall success rates. Obviously, all the sections of these Tables are two-condition set sections, so the test is less vigorous here, but the indication is that the method improves with the strictness of the smoothness requirement, even at the expense of loss of degrees of freedom.

6.1.1 A particular weakness of the method

As the method is tuned to picking up patterns of motion largely disregarding their speed of execution (but not entirely — the windowing of the data is within limits on the window length), the method may entirely miss "pure" bradykinesis, that is, bradykinesis unaccompanied by abnormal shapes of motion. A possible remedy for this is to store (ECWT, scale) pairs in the libraries, and to modify the distances we have used to measure the separation between ECWTs to take into account the scale member of these pairs.

It should be noted that there is no reason why the method should not pick up hypokinesis, as a reduction in the *amplitude* of movement results in a reduction in the activations with respect to all ECWTs.

6.2 Other conclusions

A large number of our selected library members result in a classification indicative of a performance of our method equal to the benchmark (green rows in many of the ply-level classification diagrams), but the success of the benchmark in classifying *all* the plies thrown at it, and not just the right-hand ones, makes it seem like the methods here are not competitive as classifiers, but this, perhaps, should be no surprise, as the library/CWT methods are more *atomic*, focussed on details. If these details are sometimes there, and sometimes absent, classification failure of classifiers based on a single member is to be expected.

We have seen that a combined classifier is better than a single one on the left-hand plies, although it does not approach the benchmark. Perhaps the fairest comparison between the benchmark and classification based on the methods of this thesis would involve the best combined classifier on a set of training data, either a subset of the data used to derive the classifiers, or, preferably, entirely training new data (from the same sources). Of course, this best classifier would have to be derived *before* applying it to unseen data.

Even on the left-hand data, the classification performance is not so bad that these methods necessarily bring nothing to the table. In combination with classifiers based on other methods, including the benchmark used here, they may be useful.

On the subject of classification, it should also be noted that the paucity of data also affects the negative conclusions drawn from it.

However, although the association with PD of the characteristics picked up by the activation tails of the ECWTs may not be strong enough for them to always operate as individual *classifiers*, if they measure different things, the balance of these things may characterise the development of the disease, and/or the position in the treatment cycle. Beyond classifying the library members into linear, planar and full-dimensional, we have done little work in interpreting these members and the motion patterns they capture. (With good reason — a much larger amount of data seems to be necessary to make any conclusions about these details.)

A possible characterisation of PD1 (at the time of the data collection) has emerged from this work.

When we extracted the libraries from the RH PD1 data and selected library members such that the activation tails were, according to our criterion, as different as possible for the activations of the windows into the RH PD1 data and into the RH C2 data, the tails of the former were thicker than those of the latter, as expected. We interpret this as meaning that there are patterns of motion which are strong in the PD1 data but weak in the C2 data.

When we extracted the libraries from the LH PD1 data and selected library members such that the tails were as different as possible for the activations of the windows into the LH PD1 data and into the LH C2 data, the tails of the latter were thicker than those of the former, contrary to expectation. We interpret this as meaning that there are no patterns of motion which are strong in the PD1 data but weak in the C2 data, although patterns which are strong in the C2 data but weak in the PD1.

We interpret these facts together as meaning that patterns of motion characteristic of PD were a feature of the movement of PD1's right hand, but the most prominent feature of the motion of PD1's left hand was the attenuation of normal movement.

At the time the measurements were taken, PD1's left hand was the least affected by the progression of the disease.

Although the analysis in Chapter 4 is a worst-case analysis, its conclusion that, with a NAT accelerometer bandwidth of 80Hz, we are on the edge of these methods being unworkable is supported by the fact that the tactic we used to ameliorate the problem actually appears to be effective. Consequently, a more direct way of making the problem go away would be to use accelerometers with higher bandwidths. Given the steepness of the drop-off of the blurring caused by limited bandwidth, a modest increase to, say, 100Hz would suffice to remove most of the problem.

6.3 Novelty

We list here the more important contributions made in this Theses which we believe to be novel:

Libraries of motion shapes: at least within the field of the analysis of the motion of PD subjects;

Representing motion shapes: by equivalence classes of piecewise polynomial wavelet triplets, which are invariant under rotations and reflections;

Distance functions: working out the form of a distance based on the L_2 distance between individual functions, but which works on the equivalence classes mentioned above;

hierarchical k-medoids : an approximation to k-medoids, using the clustering of cluster centres, which runs faster than k-medoids.

The value of the last item in finding clusters of clusters which are meaningful rather than just a step towards a clustering of individual points, should, perhaps, be investigated separately.

Chapter 7

Future directions

Here we discuss what work can be done to complete and extend the work done here, and also other research avenues.

7.1 Work that could have been done within the context of the present thesis

Some of the results of our experiments point towards further experiments.

For example, all of the available left-hand or right-hand control data could be amalgamated into one collection of epochs, which could then be split into plies and the recognition experiments carried out on these plies alongside the PD1 plies. This would be especially interesting for the classification of the left-hand plies, as the method currently works better with right-hand ones.

With more control data, we would expect the selected library members to be more distant from common control characteristics, and less affected by C2's idiosyncrasies (if any!).

As our methods are based on equivalence classes which are rotation- and reflection-invariant, *all* of the control data could be amalgamated and then divided randomly into plies for further experiments, where the classification task would be to separate PD1 unseen right- or left-hand data from unseen general control data.

Of course, it is also possible to amalgamate the right- and left-hand PD1 data, but, as we believe that these data sets are somewhat different from one another in rotation- and reflection-invariant ways, and probably to a far greater extent than are the right- and left-hand data sets stemming from the control subjects, there is the possibility of the characteristics of the left-hand data sets being hidden by this amalgamation. Nevertheless, the amalgamation of this data could be investigated.

Experiments for the generalisation across attachment or control subject which involve the amalgamation of the control and/or the PD1 data to some extent can easily be devised, e.g., use the right-hand PD1 data and all of the right-hand control data for training, and divide the left-hand control data into plies to be used alongside the left-hand PD1 for testing.

As mentioned before, the experiments with the ed condition set and $n = 10$ and 11 could

not be carried out because of time and resource constraints. These could now be carried out on York University's new cluster, the Viking Cluster[47], or perhaps by modifying the software.

Much of the work that has been done here can be extended to ECWTs based on trigonometric polynomials, requiring only the tweaking of the fundamental equations without altering the basic framework. If we had had the time, we would have done this.

7.2 Further work on the CWT

7.2.1 PD

Several of the animal behaviour papers considered above use smoothing (Wilson, Shepard and Liebsch[11], Shepard *et al.*[12], Pastell *et al.*[13], Yoda *et al.*[6]) or explicit filtering (Guo *et al.*[9], Watanuki *et al.*[8]) to obtain the static acceleration. There is no reason why these techniques cannot also be applied to PD data.

Although it is perfectly possible that the wavelet techniques considered above may produce results which classify movement patterns, it is also likely that some improvement may be obtained by removing the static acceleration for the classifiers, and then using either the dynamic acceleration alone, or the dynamic acceleration together with the static acceleration as input to wavelet-based pattern matching. When this is done, it would be useful to bear in mind that a careful consideration of the pass and stop frequencies, in the manner of Watanuki *et al.*[8], could be the best way to proceed.

Another approach would be to set $\psi(x) = \begin{cases} P(x)\sum_{\ell=1}^{n+1} a_\ell x^{\ell-1}, & x \in [0,1); \\ 0, & \text{otherwise} \end{cases}$ for some fixed polynomial P (e.g., $P(x) = x(1-x)$, which has the additional property of enforcing $\psi(0) = \psi(1) = 0$, which in turn compels a degree of smoothness on $\langle \psi, f \rangle$ as the data window represented by f changes, and this may be an advantage), and analyse, or at least observe, the effects on the well-conditionedness of the matrix playing the role of H in the equation for the norm, $\|\psi\|^2 = a^{\mathrm{T}} H a$.

We may also exploit the repetitive nature of tremor by looking for mother wavelets of the form $\psi(x) = \begin{cases} \sum_{\ell=1}^{n} a_\ell x^{\ell-1}, & x \in \left[0, \frac{1}{q}\right); \\ \sum_{\ell=1}^{n} a_\ell \left(x - \frac{1}{q}\right)^{\ell-1}, & x \in \left[\frac{1}{q}, \frac{2}{q}\right); \\ \quad\vdots & \quad\vdots \\ \sum_{\ell=1}^{n} a_\ell \left(x - \frac{q-1}{q}\right)^{\ell-1}, & x \in \left[\frac{q-1}{q}, 1\right); \\ 0, & \text{otherwise,} \end{cases}$ with windows q times as long as the ones considered here (assuming the data remains sampled at 500Hz). Once derived, the "$\frac{1}{q}$ wavelet", $\psi_{\frac{1}{q}}(x) = \begin{cases} \sum_{\ell=1}^{n} a_\ell \left(\frac{x}{q}\right)^{\ell-1}, & x \in [0,1); \\ 0, & \text{otherwise} \end{cases}$ can be used.

Obviously, this combining this approach with wavelets of the trigonometric polynomial form will be simpler, as then the relevant mother wavelet is $\psi(x) =$

$$\begin{cases} \sum_{\ell=-n}^{n} c_\ell \mathrm{e}^{2\pi i q \ell x}, & x \in [0,1); \\ 0, & \text{otherwise.} \end{cases}$$

7.2.2 Animal behaviour

Although the emphasis of the work has shifted towards PD, the vectorisation of the algorithms means that there should be ample opportunity to complete the analysis of the animal data that is to hand.

7.3 The DWT and accelerometer data

With the switch of emphasis to PD, the use of multiresolution analyses with scales related by factors of two will probably become less relevant, as important features are concentrated around 4-6Hz (corresponding to scales $\frac{1}{6}$-$\frac{1}{4}$s, i.e., a range of scales whose lower limit is greater than half its upper limit). However, as noted by Daubechies[48], there are multiresolution analyses with scales related by factors q for any rational number q (except 0 and 1). This corresponds to replacing condition II on page 31 by $f(x) \in V_j \Leftrightarrow f(qx) \in V_{j+1}\ \forall j \in \mathbb{Z}$. This is an additional complicating factor in the work described below.

To fit the DWT into the methodology used for the CWT, it is necessary to find families of wavelets satisfying conditions I and II on page 31, which can be fitted to the data in a similar fashion to the functions we have used for the CWT.

Since the fast algorithms for calculating $\{\langle f, \psi_{j,k}\rangle : k \in \mathbb{Z}\}$, mentioned on p32

i) are accurate to within machine accuracy if the wavelets have compact support;

ii) are subject to greater approximation the more slowly the wavelets decay;

iii) are simpler if the wavelets are orthonormal rather than biorthonormal

(Daubechies[48]), we wish to find, in decreasing order of desirability:

a: a family of orthonormal, compactly-supported wavelets with several parameters whose waveform varies considerably;

b: a family of biorthonormal, compactly-supported wavelets with several parameters whose waveform varies considerably;

c: a family of rapidly-decaying orthonormal wavelets with several parameters whose waveform varies considerably;

d: a family of rapidly-decaying biorthonormal wavelets with several parameters whose waveform varies considerably.

The work of investigating wavelets defined by equation (2.1) with $\nu(\xi)$ other than $\nu(\xi) = 1$ has already begun, with the idea of fitting ν to the data.

In the case that $\phi = \phi_H$, $\hat{\phi}_H(\xi) = \frac{i}{\sqrt{2\pi}\xi}(1 - \mathrm{e}^{-i\xi})$, $m_0(\xi) = \frac{1+\mathrm{e}^{-i\xi}}{2}$, so wavelets corresponding to the MRA associated with ϕ_H are given by $\hat{\psi}(\xi) = -\frac{i}{\sqrt{2\pi}\xi}(1 - \mathrm{e}^{\frac{1}{2}i\xi})^2\nu(\xi)$, where ν is an arbitrary unit-modulus 2π-periodic measurable function.

Given a function f, where it is desired to maximise $|\langle f, \psi\rangle|$, we can make use of $|\langle f, \psi\rangle|^2 = |\langle \hat{f}, \hat{\psi}\rangle|^2 = \frac{1}{2\pi}\left|\int_{-\infty}^{\infty} \frac{\hat{f}(\xi)}{\xi}(1 - \mathrm{e}^{-\frac{1}{2}i\xi})^2\overline{\nu(\xi)}\mathrm{d}\xi\right|^2 = \frac{1}{2\pi}\left|\int_0^{2\pi} \overline{\nu(\xi)} \sum_{\ell\in\mathbb{Z}} \frac{\hat{f}(\xi+2\ell\pi)}{\xi+2\ell\pi}[1 - (-1)^\ell \mathrm{e}^{-\frac{1}{2}i\xi}]^2\mathrm{d}\xi\right|^2$, which, if the sum converges, means that we need only do our optimisations for a finite integral.

7.4 Non-human species

If the work on animal behaviour is resumed, it might be extended to monitor the behaviour of rats, which would most likely be mostly a question of scaling the accelerometer ranges.

7.5 Combination with other approaches

There has been much useful work in using accelerometry to diagnose PD and classify the stages of its progression. ABNs have been used to handle fairly long time series as a unit, successfully classifying these as stemming from PD patients and controls at a sub-expert, but useful, level. It is possible that a more analytic, wavelet-based approach, detecting short-lived patterns in the acceleration traces may complete with or complement the ABN and other approaches (possibly as preprocessing, transforming the input of ABNs to represent wavelet content at various scales).

7.6 Other data

As the NAT can be attached to a range of devices to log their output, there is a vast range of data which could be analysed in conjunction with accelerometry data, some of which could be analysed in the same way — e.g., in PD, EMG data is likely to have content indicating the progress of the disease, and this content may be accessible after a suitable wavelet transform and may supplement that of acceleration data; in animal experiments, the DWTs/CWTs[1] of EEG traces may have different relationships to the DWTs/CWTs[2] of accelerometry data under different experimental conditions.

Monitoring and diagnosis depends on the detection of changes in features of the object of these activities, and the interpretation of these changes. It is hoped that picking out "behaviours" in biological signals and associating them with wavelets tailored to fit them will enable them to be detected again, even if their scale changes. The wavelet content, as manifested in the scaleogram, counts as a feature, whose changes (absolute or in relation to

[1] With respect to wavelets picking out specific components present in the trace

[2] Picking out different specific components present in *these* traces

other features) may or may not be useful for monitoring and diagnosis. This needs to be investigated.

7.7 Further work on the subject of the Thesis, narrowly defined

We close the Thesis with a discuss of further work which could not have been done within its scope, but which continue its start.

The biggest impediment to a more thorough investigation of the power of the methods of the Thesis was the chronic lack of data, or the narrow range of subjects available, to say the same thing in a different way. Not only was the work reduced to a feasibility study, it was close to the minimum feasibility study.

A second impediment was the limited bandwidth of The NAT accelerometer. Although it is not possible to more precisely determine its impact in the light of the first impediment, it does appear to have affected the results. Moreover, some of the work done here indicates that only a modest increase in bandwidth might be sufficient to improve performance.

So the first suggestion for further work is: do it again, but with more subjects (especially PD subjects), and with slightly better equipment. At the same time, the weakness in detecting bradykinesis noted in the previous Chapter could be addressed by changing the contents of the libraries and altering the distance functions, as suggested there.

As the major result of this Thesis is to show the feasibility of a characterisation of PD by

showing, *via* an examination of the ability to the methods to distinguish between PD and control data, that the methods can find patterns of motion more strongly present in PD data than control data;

showing that these patterns of motion are reasonably distinct from one another,

an exploration of the patterns themselves and their clinical meaning is necessary.

A modest beginning has been made in this Thesis to the first part of this exploration, in the dimensionality analysis of some of our ECWTs, and even to the second part, in the tentative explanation of some of our results being due to attenuation of normal patterns of motion in the movement of PD1's left hand.

Appendix: Working behind Figures 4.28–4.30

Although we do not need a_1 and a_1' explicitly to calculate the maximum distance $\|\psi - \psi'\|$, we wish to find ψ and ψ' for illustrative purposes, and for this we *do* need a_1 and a_1'.

	$\beta = 5$Hz			$\beta = 20$Hz		
n	$\|\psi - \psi'\|$	$\|\psi - \psi_\beta\|$	$\|\psi' - \psi'_\beta\|$	$\|\psi - \psi'\|$	$\|\psi - \psi_\beta\|$	$\|\psi' - \psi'_\beta\|$
2	1.1378	0.8787	0.6907	0.0895	0.3632	0.3535
3	1.9614	0.9947	0.9742	0.3501	0.4247	0.2827
4	1.9985	0.9998	0.9990	0.5792	0.5424	0.2396
5	$\sim$2 - 1.64$\times10^{-5}$	$\sim$1 - 2.24$\times10^{-6}$	$\sim$1 - 1.12$\times10^{-5}$	0.9154	0.6767	0.0607
6	$\sim$2 - 2.42$\times10^{-7}$	$\sim$1 - 3.29$\times10^{-8}$	$\sim$1 - 1.64$\times10^{-7}$	1.1568	0.7606	0.1209
7	$\sim$2 - 1.47$\times10^{-9}$	$\sim$1 - 2.00$\times10^{-10}$	$\sim$1 - 9.98$\times10^{-10}$	1.7260	0.9290	0.6745
	$\beta = 80$Hz			$\beta = 320$Hz		
n	$\|\psi - \psi'\|$	$\|\psi - \psi_\beta\|$	$\|\psi' - \psi'_\beta\|$	$\|\psi - \psi'\|$	$\|\psi - \psi_\beta\|$	$\|\psi' - \psi'_\beta\|$
2	0.0156	0.1783	0.1777	0.0040	0.0892	0.0891
3	0.0839	0.2050	0.1878	0.0201	0.1003	0.0983
4	0.1163	0.2413	0.2133	0.0282	0.1188	0.1154
5	0.1838	0.3032	0.2474	0.0427	0.1461	0.1399
6	0.2358	0.3434	0.2624	0.0551	0.1660	0.1569
7	0.3315	0.4071	0.2722	0.0743	0.1927	0.1784

Table 1: $\|\psi - \psi_\beta\|$ *and* $\|\psi' - \psi'_\beta\|$ *for wavelets without the wec or wed conditions*

If we order the eigenvalues of D_P in increasing order along its diagonal, relabelling if necessary, so that $m_\lambda = \lambda_0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_p = M_\lambda$ (which involves a choice if either λ_0 or λ_p has multiplicity greater than 1), then we can choose $g = \left[\sqrt{\frac{\lambda_p}{\lambda_0+\lambda_p}}, 0, 0, \ldots, 0, \sqrt{\frac{\lambda_0}{\lambda_0+\lambda_p}}\right]^{\mathrm{T}}$, and then $a = O[0, \tilde{a}_1]^{\mathrm{T}}$ and $a' = O[0, \tilde{a}_1']^{\mathrm{T}}$, where $\tilde{a}_1'$ is given by equation (4.75) and the 0 in these equations is a scalar in case 1, is in $\mathbb{R}^{1\times 2}$ in case 2 and is in $\mathbb{R}^{1\times 3}$ in case 3.

We have

$$\tilde{a}_1 = \sqrt{\tfrac{\lambda_p}{\lambda_0+\lambda_p}}u_0 + \sqrt{\tfrac{\lambda_0}{\lambda_0+\lambda_p}}u_p, \tag{1}$$

where u_0 and u_p are eigenvectors of $\tilde{H}_1^{-1}\tilde{P}_1(\beta)$, corresponding to the eigenvalues λ_0 and λ_p, and normalised such that $u_0^{\mathrm{T}}\tilde{H}_1 u_0 = u_p^{\mathrm{T}}\tilde{H}_1 u_p = 1$. If the value set of the eigenvalues of $\tilde{H}_1^{-1}\tilde{P}_1(\beta)$ contains at least two values, we also have $u_0^{\mathrm{T}}\tilde{H}_1 u_p = 0$ (if this is not the case,

we can deduce[3] that $\tilde{H}_1^{-1}\tilde{P}_1(\beta) = \lambda_0 \mathrm{I}$; but, then the minimisation over $\tilde{a}_1$ and $\tilde{a}_1'$ becomes trivial, with the solution $\tilde{a}_1' = \tilde{a}_1$, for any $\tilde{a}_1$ such that $\tilde{a}_1^{\mathrm{T}}\tilde{H}_1\tilde{a}_1 = 1$ and minimum value 0, so we henceforth assume that the value set of the eigenvalues of $\tilde{H}_1^{-1}\tilde{P}_1(\beta)$ has cardinality greater than one).

Then $\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 = \lambda_0\sqrt{\frac{\lambda_p}{\lambda_0+\lambda_p}}u_0 + \lambda_p\sqrt{\frac{\lambda_0}{\lambda_0+\lambda_p}}u_p$, $\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1 = \tilde{a}_1^{\mathrm{T}}\tilde{H}_1\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 = \left[\sqrt{\frac{\lambda_p}{\lambda_0+\lambda_p}}u_0 + \sqrt{\frac{\lambda_0}{\lambda_0+\lambda_p}}u_p\right]^{\mathrm{T}}\tilde{H}_1\left[\lambda_0\sqrt{\frac{\lambda_p}{\lambda_0+\lambda_p}}u_0 + \lambda_p\sqrt{\frac{\lambda_0}{\lambda_0+\lambda_p}}u_p\right] = \frac{2\lambda_0\lambda_p}{\lambda_0+\lambda_p}$ and $\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 = \tilde{a}_1^{\mathrm{T}}\tilde{H}_1\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 = \left[\sqrt{\frac{\lambda_p}{\lambda_0+\lambda_p}}u_0 + \sqrt{\frac{\lambda_0}{\lambda_0+\lambda_p}}u_p\right]^{\mathrm{T}}\tilde{H}_1\left[\lambda_0^2\sqrt{\frac{\lambda_p}{\lambda_0+\lambda_p}}u_0 + \lambda_p^2\sqrt{\frac{\lambda_0}{\lambda_0+\lambda_p}}u_p\right] = \lambda_0\lambda_p$, so equation (4.75) becomes

$$\begin{aligned}\tilde{a}_1' &= \tfrac{4}{\lambda_0+\lambda_p}\left[\lambda_0\sqrt{\tfrac{\lambda_p}{\lambda_0+\lambda_p}}u_0 + \lambda_p\sqrt{\tfrac{\lambda_0}{\lambda_0+\lambda_p}}u_p\right] - \sqrt{\tfrac{\lambda_p}{\lambda_0+\lambda_p}}u_0 - \sqrt{\tfrac{\lambda_0}{\lambda_0+\lambda_p}}u_p \\ &= \tfrac{3\lambda_0-\lambda_p}{\lambda_0+\lambda_p}\sqrt{\tfrac{\lambda_p}{\lambda_0+\lambda_p}}u_0 + \tfrac{3\lambda_p-\lambda_0}{\lambda_0+\lambda_p}\sqrt{\tfrac{\lambda_0}{\lambda_0+\lambda_p}}u_p. \end{aligned} \tag{2}$$

	$\beta = 5$Hz			$\beta = 20$Hz		
n	$\|\psi-\psi'\|$	$\|\psi-\psi_\beta\|$	$\|\psi'-\psi'_\beta\|$	$\|\psi-\psi'\|$	$\|\psi-\psi_\beta\|$	$\|\psi'-\psi'_\beta\|$
3	1.0363	0.8829	0.7367	0.1782	0.3083	0.2575
4	1.9982	0.9998	0.9990	0.5792	0.5424	0.2398
5	1.9984	0.9998	0.9990	0.5843	0.5411	0.2265
6	$\sim$2 - 2.63$\times10^{-7}$	$\sim$1 - 3.29$\times10^{-8}$	$\sim$1 - 1.64$\times10^{-7}$	1.1564	0.7607	0.1226
7	$\sim$2 - 2.54$\times10^{-7}$	$\sim$1 - 3.29$\times10^{-8}$	$\sim$1 - 1.64$\times10^{-7}$	1.1570	0.7606	0.1198
	$\beta = 80$Hz			$\beta = 320$Hz		
n	$\|\psi-\psi'\|$	$\|\psi-\psi_\beta\|$	$\|\psi'-\psi'_\beta\|$	$\|\psi-\psi'\|$	$\|\psi-\psi_\beta\|$	$\|\psi'-\psi'_\beta\|$
3	0.0400	0.1417	0.1361	0.0100	0.0707	0.0700
4	0.1163	0.2413	0.2133	0.0282	0.1188	0.1154
5	0.1164	0.2412	0.2131	0.0282	0.1188	0.1154
6	0.2358	0.3434	0.2624	0.0551	0.1660	0.1569
7	0.2358	0.3434	0.2624	0.0551	0.1660	0.1569

Table 2: $\|\psi-\psi_\beta\|$ *and* $\|\psi'-\psi'_\beta\|$ *for wavelets with wec*

We also have

$$\begin{aligned}\|\psi-\psi_\beta\|^2 &= 1-\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1 = 1-\tfrac{2\lambda_0\lambda_p}{\lambda_0+\lambda_p}; \\ \|\psi'-\psi'_\beta\|^2 &= 1-\tilde{a}_1'^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1' = 1-\tfrac{2\lambda_0\lambda_p(5\lambda_0^2-6\lambda_0\lambda_p+5\lambda_p^2)}{(\lambda_0+\lambda_p)^3}, \end{aligned} \tag{3}$$

and we already know that $\|\psi'-\psi_\beta\| = \|\psi-\psi_\beta\|$.

To complete our evaluation of the effect of a limited bandwidth on our wavelets, at each point $x_0 \in \left[-\frac{1}{2},\frac{1}{2}\right)$ we wish to extremise $\psi''(x_0) - \psi(x_0)$ over all wavelets ψ'' such that

[3] $\tilde{H}_1^{-1}\tilde{P}_1(\beta) = O_H^{\mathrm{T}}D_H^{-1}O_HO_H^{\mathrm{T}}D_H^{\frac{1}{2}}O_P^{\mathrm{T}}D_PO_PD_H^{\frac{1}{2}}O_H = O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}D_PO_PD_H^{\frac{1}{2}}O_H = \lambda_0O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}O_PD_H^{\frac{1}{2}}O_H = \lambda_0\mathrm{I}$

	$\beta = 5$Hz			$\beta = 20$Hz		
n	$\|\psi - \psi'\|$	$\|\psi - \psi_\beta\|$	$\|\psi' - \psi'_\beta\|$	$\|\psi - \psi'\|$	$\|\psi - \psi_\beta\|$	$\|\psi' - \psi'_\beta\|$
4	1.0392	0.8833	0.7367	0.0836	0.2192	0.2035
5	1.9724	0.9965	0.9830	0.0892	0.2132	0.1946
6	1.9985	0.9998	0.9991	0.3093	0.3942	0.2731
7	$\sim$2 - 1.04$\times 10^{-5}$	$\sim$1 - 1.35$\times 10^{-6}$	$\sim$1 - 6.76$\times 10^{-6}$	0.3158	0.3974	0.2720
	$\beta = 80$Hz			$\beta = 320$Hz		
n	$\|\psi - \psi'\|$	$\|\psi - \psi_\beta\|$	$\|\psi' - \psi'_\beta\|$	$\|\psi - \psi'\|$	$\|\psi - \psi_\beta\|$	$\|\psi' - \psi'_\beta\|$
4	0.0185	0.0967	0.0949	0.0046	0.0482	0.0480
5	0.0186	0.0965	0.0947	0.0046	0.0482	0.0480
6	0.0461	0.1519	0.1449	0.0112	0.0749	0.0740
7	0.0461	0.1519	0.1449	0.0112	0.0749	0.0740

Table 3: $\|\psi - \psi_\beta\|$ *and* $\|\psi' - \psi'_\beta\|$ *for wavelets with wed*

$\|\psi'' - \psi_\beta\| \leq \|\psi - \psi_\beta\|$. As $\psi(x_0) = a^{\mathrm{T}} X_0$ and $\psi''(x_0) = a''^{\mathrm{T}} X_0$ (where a'' has the same relation to ψ'' as a has to ψ) for $X_0 = [1, x_0, x_0^2, \ldots, x_0^n]^{\mathrm{T}}$, our problem is that of extremising $a''^{\mathrm{T}} X_0$ with respect to a'', subject to $a''^{\mathrm{T}} P(\beta) a \geq a^{\mathrm{T}} P(\beta) a \Leftrightarrow \|\psi'' - \psi_\beta\| \leq \|\psi - \psi_\beta\|$, the wavelet conditions $a''^{\mathrm{T}} H a'' = 1$ and $b^{\mathrm{T}} a'' = 0$, and, optionally, $c_1^{\mathrm{T}} a'' = 0$ and $c_2^{\mathrm{T}} a'' = 0$.

As above, we use the orthogonal matrix O to reduce these problems, to that of extremising $\tilde{a}_1''^{\mathrm{T}} \tilde{X}_{01}$ with respect to $\tilde{a}_1''$, subject to $\tilde{a}_1''^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1 \geq \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1$ and $\tilde{a}_1''^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1'' = 1$, where, of course, $\tilde{a}'' = Oa'' = \begin{bmatrix} 0 \\ \tilde{a}_1'' \end{bmatrix}$, where 0 is a scalar in case 1, is in $\mathbb{R}^2$ in case 2 and is in $\mathbb{R}^3$ in case 3, and $\tilde{X}_0 = OX_0 = \begin{bmatrix} \tilde{X}_{00} \\ \tilde{X}_{01} \end{bmatrix}$, where $\tilde{X}_{00} \in \mathbb{R}, \mathbb{R}^2, \mathbb{R}^3$ in case 1 (resp. case 2, case 3).

We use the Lagrangian

$$Q(\tilde{a}_1'', \lambda) = \tilde{a}_1''^{\mathrm{T}} \tilde{X}_{01} + \lambda[\tilde{a}_1''^{\mathrm{T}} \tilde{H}_1 \tilde{a}_1'' - 1], \tag{4}$$

for the problem *without* the condition $\tilde{a}_1''^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1 \geq \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1$. This has, by analogy with the working above, turning points when $\tilde{a}_1'' = -\frac{1}{2\lambda} \tilde{H}_1^{-1} \tilde{X}_{01}$ with $\frac{1}{4\lambda^2} \tilde{X}_{01}^{\mathrm{T}} \tilde{H}_1^{-1} \tilde{X}_{01} = 1$, i.e. $\lambda = \pm \frac{1}{2} \sqrt{\tilde{X}_{01}^{\mathrm{T}} \tilde{H}_1^{-1} \tilde{X}_{01}}$, $\tilde{a}_1'' = \mp \frac{1}{\sqrt{\tilde{X}_{01}^{\mathrm{T}} \tilde{H}_1^{-1} \tilde{X}_{01}}} \tilde{H}_1^{-1} \tilde{X}_{01}$. These solutions can only be solutions of the problem *with* $\tilde{a}_1''^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1 \geq \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1$ if $\mp \frac{1}{\sqrt{\tilde{X}_{01}^{\mathrm{T}} \tilde{H}_1^{-1} \tilde{X}_{01}}} \tilde{X}_{01}^{\mathrm{T}} \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1 \geq \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1$. As the right-hand side of this inequality is positive, at most one of these solutions is valid for the more rigorous problem (i.e., the one with the extra condition). The extremal values of the laxer problem are $\mp \sqrt{\tilde{X}_{01}^{\mathrm{T}} \tilde{H}_1^{-1} \tilde{X}_{01}}$.

Using the expression (1) for $\tilde{a}_1$, we have

$$\begin{aligned} \tilde{a}_1^{\mathrm{T}} \tilde{P}_1(\beta) \tilde{a}_1 &= \tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \tilde{H}_1^{-1} \tilde{P}_1(\beta) \tilde{a}_1 \\ &= \tilde{a}_1^{\mathrm{T}} \tilde{H}_1 \left[\lambda_0 \sqrt{\tfrac{\lambda_p}{\lambda_0 + \lambda_p}} u_0 + \lambda_p \sqrt{\tfrac{\lambda_0}{\lambda_0 + \lambda_p}} u_p \right] \\ &= \tfrac{2\lambda_0 \lambda_p}{\lambda_0 + \lambda_p} \end{aligned} \tag{5}$$

and

$$\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 \;=\; \sqrt{\tfrac{\lambda_0\lambda_p}{\lambda_0+\lambda_p}}\tilde{X}_{01}^{\mathrm{T}}\left[\sqrt{\lambda_0}u_0+\sqrt{\lambda_p}u_p\right], \tag{6}$$

so

1. the maximum $\sqrt{\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01}}$ of $\tilde{a}_1''^{\mathrm{T}}\tilde{X}_{01}$ at $\tilde{a}_1'' = \frac{\tilde{H}_1^{-1}\tilde{X}_{01}}{\sqrt{\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01}}}$ is valid if $\tilde{X}_{01}^{\mathrm{T}}\left[\sqrt{\lambda_0}u_0+\sqrt{\lambda_p}u_p\right] \geq 2\sqrt{\frac{\lambda_0\lambda_p}{\lambda_0+\lambda_p}}\sqrt{\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01}}$;
2. the minimum $-\sqrt{\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01}}$ of $\tilde{a}_1''^{\mathrm{T}}\tilde{X}_{01}$ at $\tilde{a}_1'' = -\frac{\tilde{H}_1^{-1}\tilde{X}_{01}}{\sqrt{\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01}}}$ is valid if $\tilde{X}_{01}^{\mathrm{T}}\left[\sqrt{\lambda_0}u_0+\sqrt{\lambda_p}u_p\right] \leq -2\sqrt{\frac{\lambda_0\lambda_p}{\lambda_0+\lambda_p}}\sqrt{\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01}}$.

Further solutions of the more rigorous problem are obtained by replacing $\tilde{a}_1''^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1 \geq \tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1$ by $\tilde{a}_1''^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1 = \tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1$, and using the Lagrangian

$$\begin{aligned} Q(\tilde{a}_1'',\lambda,\kappa) &= \tilde{a}_1''^{\mathrm{T}}\tilde{X}_{01} + \lambda[\tilde{a}_1''^{\mathrm{T}}\tilde{H}_1\tilde{a}_1'' - 1] + \kappa[\tilde{a}_1''^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1 - \tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1] \\ &= \lambda\tilde{a}_1''^{\mathrm{T}}\tilde{H}_1\tilde{a}_1'' + [\tilde{X}_{01}+\kappa\tilde{P}_1(\beta)\tilde{a}_1]^{\mathrm{T}}\tilde{a}_1'' - \kappa\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1 - \lambda. \end{aligned} \tag{7}$$

This has turning points when

$$\begin{aligned} \tilde{a}_1'' = \tilde{a}_1''(\lambda,\kappa) &= -\tfrac{1}{2\lambda}\tilde{H}_1^{-1}[\tilde{X}_{01}+\kappa\tilde{P}_1(\beta)\tilde{a}_1], \\ \tilde{a}_1''^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{a}_1'' &= \tfrac{1}{4\lambda^2}[\tilde{X}_{01}+\kappa\tilde{P}_1(\beta)\tilde{a}_1]^{\mathrm{T}}\tilde{H}_1^{-1}[\tilde{X}_{01}+\kappa\tilde{P}_1(\beta)\tilde{a}_1] \\ &= \tfrac{1}{4\lambda^2}[\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01} + 2\kappa\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{X}_{01} + \kappa^2\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1] = 1 \text{ and} \\ \tilde{a}_1''^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1 &= -\tfrac{1}{2\lambda}\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}[\tilde{X}_{01}+\kappa\tilde{P}_1(\beta)\tilde{a}_1] \\ &= -\tfrac{1}{2\lambda}[\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{X}_{01} + \kappa\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1] = \tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1. \end{aligned}$$

Thus, $\kappa = \kappa(\lambda) = -\frac{2\lambda\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1 + \tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{X}_{01}}{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1}$ and $\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01} + \frac{4\lambda^2(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1)^2 - (\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{X}_{01})^2}{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1} = 4\lambda^2 \Leftrightarrow \lambda = \lambda_\pm = \pm\Lambda = \pm\frac{1}{2}\sqrt{\frac{(\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01})(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1) - (\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{X}_{01})^2}{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 - (\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1)^2}}$, provided $\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 - (\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1)^2 = \tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)[\tilde{H}_1^{-1} - \tilde{a}_1\tilde{a}_1^{\mathrm{T}}]\tilde{P}_1(\beta)\tilde{a}_1 \neq 0$ and the quantity under the square root sign is non-negative.

But we have already ruled out any $\tilde{a}_1$ such that $\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 - (\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1)^2 = 0$, and, using O_H, O_P, D_H and g as defined above, $\tilde{H}_1^{-1} - \tilde{a}_1\tilde{a}_1^{\mathrm{T}} = O_H^{\mathrm{T}}D_H^{-1}O_H - O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}gg^{\mathrm{T}}O_PD_H^{-\frac{1}{2}}O_H = O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}[\mathrm{I} - gg^{\mathrm{T}}]O_PD_H^{-\frac{1}{2}}O_H$. As we have that $g^{\mathrm{T}}g = 1$, O_H and O_P are orthogonal and D_H is positive definite, this means that $\tilde{H}_1^{-1} - \tilde{a}_1\tilde{a}_1^{\mathrm{T}}$ is positive semi-definite and $\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 - (\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1)^2 > 0$.

Thus, we require that $(\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01})(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1) - (\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{X}_{01})^2 = \tilde{X}_{01}^{\mathrm{T}}[(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1)\tilde{H}_1^{-1} - \tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}]\tilde{X}_{01} \geq 0$. But $\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{a}_1 = g^{\mathrm{T}}D_P^2g$ and $\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 = (O_H^{\mathrm{T}}D_H^{-1}O_H)(O_H^{\mathrm{T}}D_H^{\frac{1}{2}}O_P^{\mathrm{T}}D_PO_PD_H^{\frac{1}{2}}O_H)(O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}g) = O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}D_Pg$, so $(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1)\tilde{H}_1^{-1} - \tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1} =$

$(g^{\mathrm{T}}D_P^2 g)O_H^{\mathrm{T}}D_H^{-1}O_H - O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}O_P^{\mathrm{T}}D_P g g^{\mathrm{T}}D_P O_P D_H^{-\frac{1}{2}}O_H = O_H^{\mathrm{T}}D_H^{-\frac{1}{2}}[(g^{\mathrm{T}}D_P^2 g)\mathrm{I} - O_P^{\mathrm{T}}D_P g g^{\mathrm{T}}D_P O_P]D_H^{-\frac{1}{2}}O_H$, which is positive semi-definite, as $(g^{\mathrm{T}}D_P^2 g)\mathrm{I} - O_P^{\mathrm{T}}D_P g g^{\mathrm{T}}D_P O_P$ is positive semi-definite, because $g^{\mathrm{T}}D_P O_P O_P^{\mathrm{T}}D_P g = g^{\mathrm{T}}D_P^2 g$. Hence, we do have $(\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01})(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1) - (\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{X}_{01})^2 \geq 0$.

This means that the values $\tilde{X}_{01}^{\mathrm{T}}\tilde{a}_1''(\lambda_\pm, \kappa(\lambda_\pm))$ of $X_0^{\mathrm{T}}a''$ are at least locally extremal, where

$$\begin{aligned}
\tilde{X}_{01}^{\mathrm{T}}\tilde{a}_1''(\lambda_\pm, \kappa(\lambda_\pm)) &= \mp\tfrac{1}{2\Lambda}[\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01} + \kappa(\pm\Lambda)\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1] \\
&= \frac{(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1)(\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1)}{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1} \mp \\
&\qquad \tfrac{1}{2\Lambda}\frac{(\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01})(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1) - (\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1)^2}{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1} \\
&= \frac{(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1)(\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1)}{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1} \mp \\
&\qquad \frac{\sqrt{(\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01})(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1) - (\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1)^2}}{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1} \times \\
&\qquad \sqrt{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 - (\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1)^2}, \qquad (8)
\end{aligned}$$

and

$$\begin{aligned}
\tilde{a}_1''(\lambda_\pm, \kappa(\lambda_\pm)) &= \frac{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1}{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1}\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 \mp \\
&\qquad \sqrt{\frac{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 - (\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{a}_1)^2}{(\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01})(\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1) - (\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{X}_{01})^2}} \times \\
&\qquad \left[\tilde{H}_1^{-1}\tilde{X}_{01} - \frac{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{X}_{01}}{\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1}\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1\right]. \qquad (9)
\end{aligned}$$

Using equations (5) and (6) and

$$\begin{aligned}
\tilde{a}_1^{\mathrm{T}}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 &= \tilde{a}_1^{\mathrm{T}}\tilde{H}_1\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{H}_1^{-1}\tilde{P}_1(\beta)\tilde{a}_1 \\
&= \tilde{a}_1^{\mathrm{T}}\tilde{H}_1\left[\lambda_0^2\sqrt{\tfrac{\lambda_p}{\lambda_0+\lambda_p}}u_0 + \lambda_p^2\sqrt{\tfrac{\lambda_0}{\lambda_0+\lambda_p}}u_p\right] = \lambda_0\lambda_p, \qquad (10)
\end{aligned}$$

we have

$$\begin{aligned}
\tilde{X}_{01}^{\mathrm{T}}\tilde{a}_1''(\lambda_\pm, \kappa(\lambda_\pm)) &= \tfrac{2}{\lambda_0+\lambda_p}\sqrt{\tfrac{\lambda_0\lambda_p}{\lambda_0+\lambda_p}}\tilde{X}_{01}^{\mathrm{T}}\left[\sqrt{\lambda_0}u_0 + \sqrt{\lambda_p}u_p\right] \mp \\
&\qquad \tfrac{\lambda_p-\lambda_0}{\lambda_0+\lambda_p}\sqrt{\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01} - \tfrac{1}{\lambda_0+\lambda_p}\left[\tilde{X}_{01}^{\mathrm{T}}\left(\sqrt{\lambda_0}u_0 + \sqrt{\lambda_p}u_p\right)\right]^2}, \\
&\qquad (11)
\end{aligned}$$

and

$$\begin{aligned}\tilde{a}_1''(\lambda_\pm,\kappa(\lambda_\pm)) &= 2\sqrt{\tfrac{\lambda_0\lambda_p}{(\lambda_0+\lambda_p)^3}}\left[\sqrt{\lambda_0}u_0+\sqrt{\lambda_p}u_p\right]\mp \\ &\quad \tfrac{\lambda_p-\lambda_0}{\lambda_0+\lambda_p}\sqrt{\frac{\lambda_0+\lambda_p}{(\lambda_0+\lambda_p)\tilde{X}_{01}^{\mathrm{T}}\tilde{H}_1^{-1}\tilde{X}_{01}-\left[\tilde{X}_{01}^{\mathrm{T}}\left(\sqrt{\lambda_0}u_0+\sqrt{\lambda_p}u_p\right)\right]^2}}\times \\ &\quad \left\{\tilde{H}_1^{-1}\tilde{X}_{01}-\tfrac{\tilde{X}_{01}^{\mathrm{T}}\left[\sqrt{\lambda_0}u_0+\sqrt{\lambda_p}u_p\right]}{\lambda_0+\lambda_p}\left[\sqrt{\lambda_0}u_0+\sqrt{\lambda_p}u_p\right]\right\}. \end{aligned} \tag{12}$$

Bibliography

[1] Wikipedia, "Parkinson's disease," http://en.wikipedia.org/wiki/Parkinson's_disease, last modified, 11th July 2014, accessed 19th July 2014.

[2] ——, "Management of Parkinson's disease," http://en.wikipedia.org/wiki/Management_of_Parkinson%27s_disease, last modified, 13th July 2014, accessed 21st July 2014.

[3] P. Hagell and L. Swinn, *Diagnosing Parkinson's Disease*, 1st ed. Whurr, 2005.

[4] M. A. Lones, S. L. Smith, A. M. Tyrrell, J. E. Alty, and D. R. S. Jamieson, "Characterising neurological time series data using biologically motivated networks of coupled discrete maps," *Biosytems*, vol. 112, pp. 94–101, 2013.

[5] J. Austin, C. Bailey, A. Moulds, G. Hollier, M. Freeman, A. Fargas, T. Lampert, B. Platt, and G. Riedel, "A miniaturised 4-channel, 2kHz biosignal data recorder with 3-axis accelerometer and infra-red timestamp function," in *Proceedings of Sensorcom2013, Barcelona, August, 2013*, 2013, pp. 213–219.

[6] K. Yoda, Y. Naito, A. Takahashi, J. Nishikawa, Y. Ropert-Coudert, M. Kurita, and Y. le Maho, "A new technique for monitoring the behaviour of free-ranging Adélie penguins," *Journal of Experimental Biology*, vol. 204, pp. 685–690, 2001.

[7] H. Tanaka, Y. Takagi, and Y. Naito, "Swimming speeds and buoyancy compensation of migrating adult chum salmon *Oncorhyncus keta* revealed by speed/depth/acceleration data logger," *Journal of Experimental Biology*, pp. 3895–3904, 2001.

[8] Y. Watanuki, A. Takahashi, F. Daunt, S. Wanless, M. Harris, K. Sato, and Y. Naito, "Regulation of stroke and glide in an foot-propelled avian glider," *Journal of Experimental Biology*, vol. 208, pp. 2207–2216, 2005.

[9] Y. Guo, P. Corke, T. Wark, G. Bishop-Hurley, and D. Swain, "Animal behaviour understanding using wireless sensor networks," in *31st IEEE conference on local computer networks*, 2006.

[10] M. Moreau, S. Siebert, A. Buerkert, and E. Schlecht, "Use of a tri-axial accelerometer for automated recording and classification of goats' grazing behaviour," *Applied Animal Behaviour Science*, vol. 119, pp. 158–170, 2009.

[11] R. P. Wilson, E. Shepard, and N. Liebsch, "Prying into the intimate details of animal lives: use of a daily diary on animals," *Endangered Species Research*, vol. 4, pp. 123–137, 2008.

[12] E. L. C. Shepard, R. P. Wilson, F. Quintana, A. G. Laich, N. Liebsch, D. A. Albareda, L. G. Halsey, A. Gleiss, D. T. Morgan, A. E. Myers, C. Newman, and D. W. Macdonald, "Identification of animal movement patterns using tri-axial accelerometry," *Endangered Species Research*, Mar. 2008, published online.

[13] M. Pastell, J. Tiusanen, and L. Hänninen, "A wireless accelerometer system with wavelet analysis for assessing lameness in cattle," *Biosystems Engineering*, vol. 104, pp. 545–551, 2009.

[14] H. Terashi, H. Utsumi, Y. Ishimura, T. Takazawa, Y. Okuma, M. Yoneyama, and H. Mitoma, "Deficits in scaling of gait force and cycle in Parkinsonian gait identified by long-term monitoring of acceleration with the portable gait rhythmogram," *ISRN Neurology*, 2012, article id.: 306816.

[15] "FreeScale mma7260q ±1.5-6g three axis low-g micromachined accelerometer (rev. 1)," Spark fun Electronics — https://www.sparkfun.com/datasheets/Accelerometers/MMA7260Q-Rev1.pdf, Jun. 2005, accessed 21st June 2016.

[16] R. M. Rao and A. S. Bopardikar, *Wavelet Transforms: Introduction to Theory and Applications*. Prentice Hall, 1998.

[17] E. Hernández and G. Weiss, *A First Course on Wavelets*, ser. Studies in Advanced Mathematics. CRC Press, 1996.

[18] S. L. Smith, "Cartesian genetic programming and its application to medical diagnosis," *IEEE Computational Intelligence Magazine*, pp. 56–67, Nov. 2011.

[19] M. A. Lones, S. L. Smith, A. M. Tyrrell, J. E. Alty, and D. R. S. Jamieson, "Evolving computational dynamic systems to recognise abnormal human motor function," in *Information Processing in Cells and Tissues*, ser. Lecture Notes in Computer Science. Springer, 2012, vol. 7223.

[20] M. A. Lones, S. L. Smith, J. E. Alty, , S. E. Lacy, K. L. Possin, D. R. S. Jamieson, and A. M. Tyrrell, "Evolving classifiers to recognise the movement characteristics of Parkinson's Disease patients," accepted for publication in IEEE Transactions on Evolutionary Computation, available through ieeexplore.ieee.org prior to publication (subject to further editing prior to publication), 2013.

[21] S. M. Rissanen, M. Kankaanpää, M. P. Tarvainen, J. Nuutinen, O. Airaksinen, and P. A. Karjalainen, "EMG and acceleration signal analysis for quantifying the effects of medication in Parkinson's disease," in *33rd International Conference of the IEEE EMBS*, 2011.

[22] Wikipedia, "Kurtosis," http://en.wikipedia.org/wiki/Kurtosis, Nov. 2013, accessed 28th November 2013.

[23] ——, "Unified Parkinson's Disease Rating Scale," http://en.wikipedia.org/wiki/UPDRS, Nov. 2013, accessed 28th November 2013.

[24] European Parkinson's Disease Association, "Unified Parkinson's Disease Rating Scale (UPDRS)," http://www.epda.eu.com/en/parkinsons/in-depth/parkinsonsdisease/rating-scales/updrs/, 2013, accessed 28th November 2013.

[25] C. Bailey, J. Austin, A. Fargus, T. Lampert, G. Hollier, A. Moulds, and M. Freeman, "A miniature multisensor biosignal data recorder and its evaluation for unsupervised Parkinson's disease data collection," 2014, submitted to Sensor Devices 2014 (IARIA).

[26] N. P. Foundation, "Tremor," http://www.toolkit.parkinson.org/ content/tremor, 2014, accessed 25th August 2014.

[27] Wikipedia, "Stationary wavelet transform," http://en.wikipedia.org/wiki/Stationary_wavelet_transform#Synonyms, Jun. 2013, accessed 26th November 2013.

[28] J. E. Fowler, "The redundant discrete wavelet transform and additive noise," *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 629–632, Sep. 2005.

[29] J.-Y. L. Boudec, *Performance Evaluation of Computer and Communication Systems*. Presse polytechniques et universitaires romandes, EPFL-Rolex Learning Center, Post office box 119, CH-1015, Luasanne, Switzerland: EPFL Press, 2010.

[30] P. Martiskainen, M. Järvinen, J.-P. Skön, J. Tiirikainen, M. Kolehmainen, and J. Mononen, "Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines," *Applied Animal Behaviour Science*, vol. 119, pp. 32–38, 2009.

[31] M. I. Jordan, "The kernel trick," http://www.cs.berkeley.edu/j̃ordan/courses/281B-spring04/lectures/lec3.pdf, accessed 2nd December 2013.

[32] Wikipedia, "Support vector machine," http://en.wikipedia.org/wiki/Support_vector_machine, Nov. 2013, accessed 2nd December 2013.

[33] M. Muthuraman, A. Galka, G. Deuschl, U. Heute, and J. Raethjen, "Dynamical correlation of non-stationary signals in time domain -— a comparative study," *Biomedical Signal Processing and Control*, vol. 5, 2010.

[34] S. Sello, S. kyung Strambi, Gennaro De Michele, and N. Ambrosino, "Respiratory sound analysis in healthy and pathological subjects: A wavelet approach," *Biomedical Signal Processing and Control*, vol. 5, 2008.

[35] S. kyung Strambi, B. Rossi, Gennaro De Michele, and S. Sello, "Effect of medication in parkinson's disease: a wavelet analysis of emg signals," *Medical Engineering and Physics*, vol. 26, 2004.

[36] S. Rezvanian and T. Lockhart, "Towards real-time detection of freezing of gait using wavelet transform on wireless accelerometer data," in *Sensors (Basel)*, April 2016.

[37] "Trademark Image Recognition," https://www.cs.york.ac.uk/arch/ neural-networks/technologies/trademark-image-recognition.html, 2011.

[38] H. Goldstein, *Classical Mechanics*, 1951.

[39] R. L. Burden and J. D. Faires, *Numerical Analysis*, 7th ed. Brooks/Cole, 2001.

[40] G. Bachman and L. Narici, *Functional Analysis.* Academic Press, London, 1966.

[41] X. Jin and J. Han, *K-Medoids Clustering.* Boston, MA: Springer US, 2010, pp. 564–565. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-30164-8_426

[42] L. Jaulin and E. Walter, "Set inversion via interval analysis for nonlinear bounded-error estimation," *Automatica*, vol. 29, no. 4, 1993.

[43] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2007, p. 1027–1035.

[44] E. COPSON, *AN INTRODUCTION TO THE THEORY OF FUNCTIONS OF A COMPLEX VARIABLE*, 1960. [Online]. Available: https://books.google.co.uk/books?id=OwCf9ke0urYC

[45] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed. Elsevier/Academic Press, Amsterdam, 2007.

[46] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 1995, pp. 1137–1143.

[47] L. Moor, "The new York Advanced Research Computing Cluster - the Viking cluster," https://wiki.york.ac.uk/display/RHPC/ The+new+York+Advanced+Research+Computing+Cluster++-+the+Viking+cluster, Dec. 2018, modified by Andrew Smith.

[48] I. Daubechies, *Ten Lectures on Wavelets.* Society for Industrial and Applied Mathematics, 1992.