UNIVERSITY OF CANTERBURY

Department of Mechanical Engineering
Christchurch New Zealand

Rowing Performance Monitoring System Development

by

Thomas Williams

A thesis submitted in fulfilment of the
requirements for the Degree
of
Master of Engineering
at the University of Canterbury

June 2001

Acknowledgements

A number of people deserve thanks for helping directly and indirectly in the creation of this thesis.

First of all many thanks are due to my supervisor, Dr David Aitchison without whom the opportunity to undertake this project would not have existed. Thanks for the encouragement, knowledgeable support and Yorkshire puddings!

Thanks to the workshop staff: the ever-tolerant (and sarcastic) Otto Bolt, and the ever-cheerful Ken Brown. Thanks also to Eric Cox.

A big thank you to the 'Julians': Julian Phillips and Julian Murphy, the electronics technicians who helped me so much over my undergraduate and graduate years.

Thanks to Richard Parr and all at Kerrs Reach for your help and patience.

Thank you to my fellow postgraduate students, who made life more enjoyable with lunchtime discussions or squash duels.

Thanks to Sports Science New Zealand and the Brian Mason Trust for providing funding for this work.

Lastly, and by all means most, thanks to my family.

Abstract

The aim of this work was to develop sensory devices and data acquisition system to facilitate investigations into the mechanics of the rowing system, comprising the rower(s), boat and oars. As such, the parameters to be measured were: boat and seat position, velocity and acceleration; oar force; foot force; oar angle and rower heart rate.

An oar force sensor was designed that fitted into the cavity of a modified oarlock. This sensor design is cheap, yields sound results and its presence is almost not noticeable to the rower. A review of previously applied methods of oar force measurement, predating 1900, is included.

Foot force is of interest to many different fields of research, thus there is a large amount of literature on the subject of foot force measurement. A comprehensive review of this literature is used to aid in the design of the required sensor. The combination of a non-simple dynamic loading (i.e. time varying spatially distributed normal and shear forces), with static foot position distinguishes the problem of measuring the force under the feet during rowing from most previously considered cases. A strain gauge-based force sensing plate was designed to measure both normal force distribution and unidirectional shear force under the feet. Sample results are presented from a study with international class New Zealand rowers on a rowing ergometer. The sensor, performs well under normal force loadings, but needs modification to measure shear accurately. Possible modifications are suggested.

While only a single oar angle, known as the sweep angle, was required to be measured, a sensor combination capable of measuring the spatial orientation of the oar relative to the boat was conceived. A new method of relative orientation estimation, via approximation of the Rodrigues' vector, which allows relative weighting of sensory data, was derived. Unfortunately, calibration issues prevented the gathering of meaningful data in the time available. A full theoretical development, including a new calibration scheme, which should alleviate the encountered problems, is included.

While the motion of the rower within the boat is an important consideration in the dynamics of the rowing system, few previous researchers have measured it. These previous methods are briefly described, before the sensor used in this study, the optical rotary encoder, is detailed. Differentiation of the encoder signal to obtain seat velocity and acceleration relative to the boat was achieved using a purpose designed simple Kalman filter.

The kinematic parameters of the boat, i.e. position, velocity and acceleration were measured using a combination of accelerometer and submerged impeller. The information from these two sensors was combined using a variant of the Kalman filter used in the differentiation of the encoder signal. The combination of the seat and boat kinematics allows study of the motion of the system centre of mass.

Supplying power to, and collecting data from the above sensory devices was a purpose built data acquisition system dubbed ORAC (On-the-water Rowing Acquisition Computer). ORAC was designed to transmit the collected information, in real-time, to a remote laptop computer via wireless LAN, but the system used proved to have insufficient range, and hence ORAC was used as a standalone computer.

Table of Contents

List of Figures

List of Tables

Chapter 1

The aim of this research was to design sensors capable of measuring the dynamic and kinematic parameters of the rowing system, comprising rower, boat and oars, during on-the-water rowing. As such, the logical starting point is to describe the basics of the dynamics of this system, both to illuminate the topic for people unfamiliar with rowing and to specify the requirements of the instrumentation to monitor the dynamics; a qualitative and conservative 'observability analysis'.

This chapter first informally develops crude 'one-dimensional' equations of motion of a rowing system. This model is then used to make decisions about which parameters need to be measured, for a full description of the rowing system to be provided. The chosen parameters are:

- force at the oarlock
- force at the feet
- oar angle
- boat displacement, velocity and acceleration
- seat displacement, velocity and acceleration relative to the boat

In addition to these kinematic and dynamic parameters, the rower's heart rate is also measured.

After formally stating the research aims in Section 1.6, Section 1.7 briefly describes the implemented sensors, and the rest of the thesis.

1.1 Description of Rowing Cycle

For those who have not observed rowing, a brief description of the rowing motion is probably required. The rowing stroke can be decomposed into four stages: catch, drive, release and recovery. The two main phases are the drive and the recovery, with the catch and the release being transitional. During the drive the rower pulls on the oar while the blade is submerged, at the same time pushing on the foot-stretcher and straightening his legs, moving on a sliding seat to the bow. While the blade of the oar remains essentially motionless, the passage of the boat means that the oar, which started near the bow of the boat, ends the power phase near the stern. During recovery the oar is extracted from the water (the release) and moved back towards the bow of the boat. Simultaneously the rower draws himself towards the stern by pulling on the foot-stretchers (although rowing coaches will tell you that the rowers simply let the boat slide under them). As the oar is moved through the air it is rotated about the loom (shaft) so that the frontal area of the blade is minimised. This is known as feathering. The rower then submerges the oar again (the catch) and the cycle is completed.

1.2 Assumptions and Conventions

In the following development, the oar is represented as a lever that has its fulcrum at the blade. This follows the approach of Brearley and de Mestre [13]. An alternative approach is to consider the oar as a lever with the fulcrum at the oarlock. Both approaches have previously been used in models of rowing dynamics. Dudhia [23] suggests that the fulcrum at the oarlock is more believable for rowers, since, to them, this point seems fixed while the blade appears to move, while that considering the fulcrum at the blade is natural to stationary observers as the blade moves very little through the water. Naturally either convention results in the same equations of motion.

For simplification, all drag forces on the boat, hydrodynamic and aerodynamic, are combined into a single quantity, D, that acts to oppose the motion of the boat. Previous researchers have stated that the majority of the hydrodynamic drag is due to viscous effects, Pope [56], Wellicome [67]. As a consequence a suitable model for D is some quadratic function of the velocity of the boat relative to the water. If it is assumed that the water is motionless, then D obviously becomes a function of the instantaneous hull velocity.

In models of rowing dynamics it is normal to ignore the mass of the oar, the exception being the comprehensive models developed by Rose [57] and Zatsiorsky [70]. There is very little to be gained in understanding through including the inertial effects of the oar, so the additional complication is not warranted. The only time that one may want to include the inertial effects of the oar in a model is if a new type of oar with radically different mass distribution characteristics was proposed.

1.3 System Components

1.3.1 Oar

The rower pulls on the oar with force, F_{hand}. Fluid forces exert a force F_{blade} at the centre of the oar blade, and the reaction force due to the boat is F_{lock}, transmitted to the oar through the rigger and oarlock. As shown it is assumed that all oar forces are normal to the shaft of the oar and in the plane of the page.

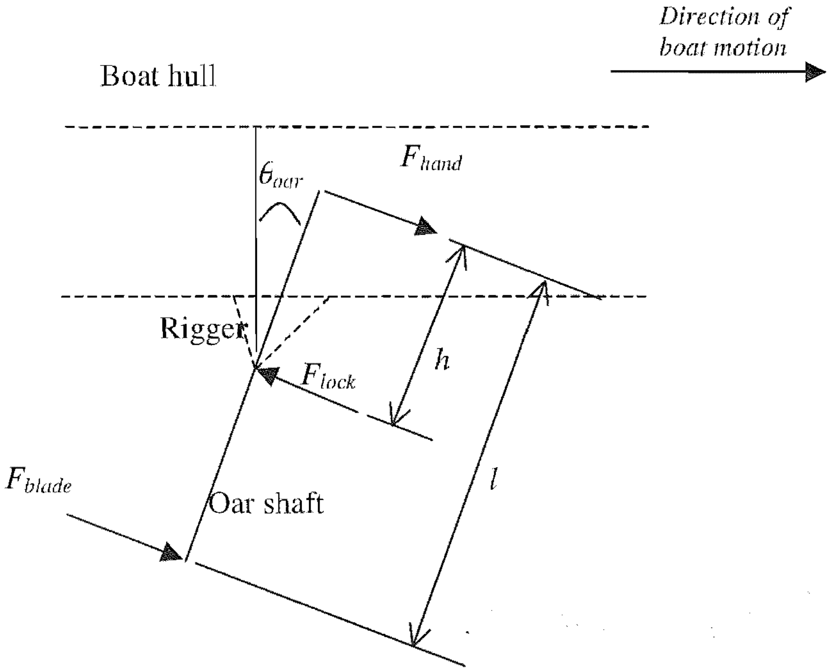


Figure 1.1 Free body diagram of the oar. Forces positive to the right, moments positive clockwise.

Since the oar is assumed massless the forces and moments are summed to zero

$$\Sigma F=0: \qquad F_{hand} + F_{blade} = F_{lock} \tag{1.1}$$

$$\Sigma M=0: \qquad F_{hand}\, l = F_{lock}(l - h) \tag{1.2}$$

yielding the two relations

$$F_{hand} = F_{lock}(l - h)/l \qquad \text{(note } F_{hand} < F_{lock}\text{)} \tag{1.3}$$

$$F_{blade} = F_{lock} - F_{hand} = F_{lock}(1 - (l - h)/l) = F_{lock}(h/l) \tag{1.4}$$

1.3.2 Boat

The force that the oar imparts to the rigger is equal and opposite to the force of the boat on the oar, F_{lock}. This force (again assumed to act in the horizontal plane) is at angle θ_{oar} to the direction of motion, thus the force in the direction of travel is $F_{lock}\cos\theta_{oar}$. Of course during rowing there are an even number of oars on either side of the boat ensuring that so long as the forces and angles are equal the boat should not turn. The components of force perpendicular to the direction of travel tend to deform the sides of the boat and are considered to be lost as dissipated elastic energy. The remaining forces on the boat in the direction of travel are the force exerted on the foot-stretchers by the rower, F_{foot}, and the drag on the boat and rower, D, which always opposes the motion of the boat. Forces not in the direction of travel, such as the weight force of the rower and the buoyancy force on the boat are not shown. It is assumed that the rollers of the seat move with no friction.

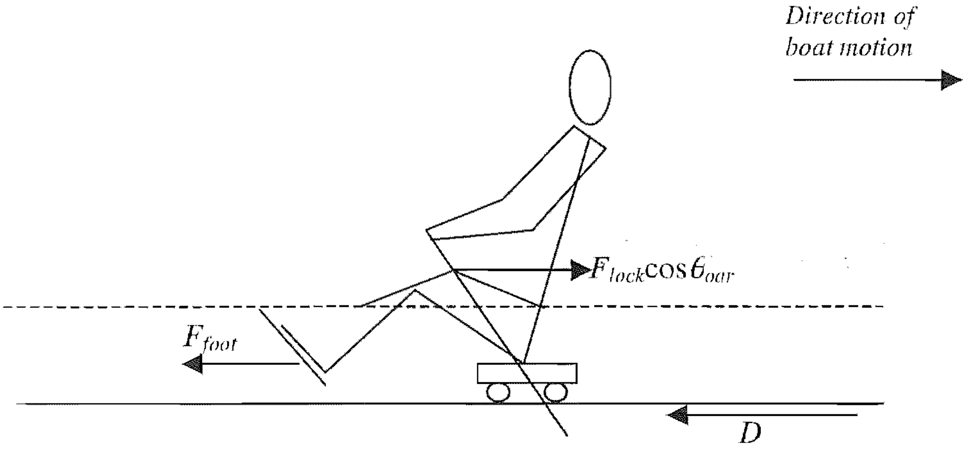


Figure 1.2 Forces on the boat in the direction of motion. Force is positive to the right.

Summing the forces in the direction of motion gives the equation of motion of the boat:

$$\Sigma F = m_{boat}a_{boat}: \qquad F_{lock}\cos\theta_{oar} - F_{foot} - D = m_{boat}a_{boat} \qquad (1.5)$$

where m_{boat} is the mass of the boat, and a_{boat} is the acceleration of the boat.

1.3.3 Rower

The two forces that are on the rower in the direction of motion, neglecting aerodynamic drag, are F_{foot} and F_{hand} both of which are of course equal and opposite to the forces that the rower applied to the stretchers and oar respectively. Recalling that the direction of F_{hand} is assumed to be normal to the oar, the force in the direction of motion of the rower is $F_{hand}\cos\theta_{oar}$.

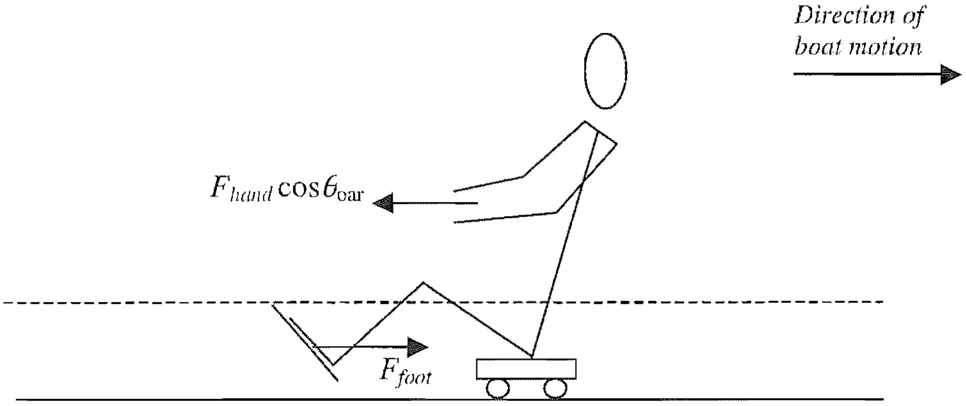


Figure 1.3 Forces on the rower in the direction of motion. Force is positive to the right.

The acceleration of the rower with respect to the boat is denoted by a_{rower}. Since Newton's laws only hold in an inertial frame, the sum of the forces is equal to the mass of the rower multiplied by the absolute acceleration of the rower, that is, the sum of the acceleration of the rower relative to the boat and the acceleration of the boat relative to an inertial frame

$$\Sigma F = m_{rower}(a_{rower} + a_{boat}): \qquad -F_{hand}\cos\theta_{oar} + F_{foot} = m_{rower}(a_{rower} + a_{boat}) \qquad (1.6)$$

1.4 Model Summary

1.4.1 Summary of Equations

The main equations developed are

$$\text{Oar:} \qquad F_{blade} = F_{lock} - F_{hand} = F_{lock}(1 - (l - h)/l) = F_{lock}(h/l) \tag{1.4}$$

$$\text{Boat:} \qquad F_{lock}\cos\theta_{oar} - F_{foot} - D = m_{boat}a_{boat} \tag{1.5}$$

$$\text{Rower:} \qquad F_{foot} - F_{hand}\cos\theta_{oar} = m_{rower}(a_{rower} + a_{boat}) \tag{1.6}$$

The boat equation (1.5) may be used as it is, or the rower equation (1.6) may be used to substitute a value for F_{foot} as follows:

$$F_{lock}\cos\theta_{oar} - [F_{hand}\cos\theta_{oar} + m_{rower}(a_{rower} + a_{boat})] - D = m_{boat}a_{boat}$$

$$(F_{lock} - F_{hand})\cos\theta_{oar} - m_{rower}(a_{rower} + a_{boat}) - D = m_{boat}a_{boat} \tag{1.7}$$

Using the oar equation, $F_{lock} - F_{hand} = F_{blade} = F_{lock}(h/l)$, gives two more forms of the boat equation

$$F_{blade}\cos\theta_{oar} - m_{rower}a_{rower} - D = (m_{boat} + m_{rower})\, a_{boat} \tag{1.8}$$

$$F_{lock}(h/l)\cos\theta_{oar} - m_{rower}a_{rower} - D = (m_{boat} + m_{rower})a_{boat} \tag{1.9}$$

When considering (1.8) and (1.9), it is seen that a positive relative acceleration of the rower acts as a force against the motion of the boat. Clearly then the rower must consider more in his technique than the method by which he can impart the maximum oar force; his motion within the boat is also important. The sequencing of the rower's motion and the application of force is also significant.

1.4.2 Limitations of Model

The developed model incorporates the salient points of rowing dynamics. Limitations of the model, excluding the obvious assumed one-dimensional aspect of motion, are mainly to do with propulsion. For example, the force on the blade of the oar is generated as a consequence of the oar-blade velocity the relative to the water. Thus, in reality the oar must move through the water to create force, meaning that it is incorrect to place the fulcrum at the blade, the basis of this model. Previous researchers have on occasion suggested that the fulcrum is slightly inside the blade [6]. This assumption however does not radically alter the form of the equations.

1.5 Parameters to Measure

The equations (1.5), (1.8) and (1.9) describe a model of sufficient accuracy to indicate the approximate relationships between system variables. Given that the aim of this work is to measure the dynamics of rowing, the instrumentation must be able to quantify each of the parameters involved in these equations of motion.

The ultimate aim of competitive rowing is to beat all other boats to the finish line. Without consideration of the dynamics it is obvious that the boat's instantaneous velocity and distance travelled need to be measured.

Unless the drag coefficients of the boat are known, in which case the system drag, D, can be approximated as a function of instantaneous velocity, it is impossible to measure D. Since D appears in the equations of motion, and it is an unknown in this case, all other variables must be measured. From the various forms of the equation this means that the required measurands are:

(1.5) F_{lock} θ_{oar} F_{foot} a_{boat} with known parameter m_{boat}
(1.8): F_{blade} θ_{oar} a_{rower} a_{boat} with known parameters m_{boat}, m_{rower}
(1.9): F_{lock} θ_{oar} a_{rower} a_{boat} with known parameters m_{boat}, m_{rower}, h, l

The two main options are therefore seen to be measuring oar force at the oarlock or the blade, and measuring the acceleration of the rower relative to the boat or foot force. The next section details the aims of the research, while Section 1.7 is a summary of the designed instrumentation and the thesis.

1.6 Research Aims

This research was not intended to answer specific questions about rowing, i.e. the determination of optimal oar force curve, or the best way in which to sequence the events of rowing; rather it was intended that a tool that could be used to answer such questions be constructed.

Section 1.5 outlined the parameters of interest in a one-dimensional study of the mechanics of rowing. As such, these are the baseline parameters that need to be measured in the study. Beyond these parameters, there are others of obvious interest, such as the velocity and displacement of the boat as a function of time. To be specific, it was required that sensors be designed or acquired to measure:

- The force at the oarlock
- The normal and shear forces at the rower's feet
- The oar angle
- The displacement, velocity and acceleration of the boat
- The displacement, velocity and acceleration of the seat relative to the boat
- The rower's heart rate

In addition to these sensors, it was required that a compact, lightweight data acquisition system be constructed to fit onto a boat, supply power to the various sensors, store their output, and also transmit the data to a remote computer.

1.7 Instrumentation and Thesis Summary

Referring to Section 1.5, there were three sets of parameters that could be exclusively measured to provide a full description of the one-dimensional dynamics of rowing. This section outlines the reasons for the choices, which were outlined in Section 1.6, the form the sensors took, and refers to the parts of the thesis in which the full descriptions of the sensors may be found.

It was chosen to measure the force at the oarlock rather than the blade because it is easier and because it made for a more generally applicable system, i.e. oarlocks are a fairly standard piece of equipment, while rowers are likely to have their personal oars. The oarlock force sensor, a small cylindrical aluminium insert is detailed in Chapter 2.

Rather than make a choice between measuring F_{foot} and a_{rower}, both were measured. This was because coaches had expressed an interest in measuring the precise timing of rowing events such as the pull on the oar and the push on the foot-stretcher during the drive. The foot force sensor was designed to measure centre of applied normal force, magnitude of normal force and magnitude of shear force. The first two objectives were achieved, but the design needs slight modification to reliably measure shear. The design and function of the foot force sensor is covered in Chapter 3. Suggestions for modifications are also included.

While it is strictly required to measure only the angle θ_{oar} that the projection of the oar makes with a perpendicular to the boat in the horizontal plane, it was decided that an attempt would be made to measure the instantaneous spatial orientation of the oar. In Chapter 4, the method by which the earth's magnetic and gravitational fields can be used to measure the relative orientation of non-ferrous bodies, such as the oar and the boat, is described. Also included in this chapter is a review of methods by which relative orientations can be discerned through vector observations from two bodies. After this review and comments on efficacy, a new 'least squares' method, which has a number of benefits over the reviewed methods, is derived. While this method shows great theoretical promise, annoying hardware problems prevented the gathering of sound results. An investigation into the problem and methods by which it can be overcome are included in this chapter.

The acceleration of the rower relative to the boat is estimated by double differentiation of the seat position signal yielded from an incremental rotary encoder. Differentiation of a quantized signal can cause significant noise. The novel method that was used to avoid this is described in Chapter 5. Also in Chapter 5 is the discussion of the method by which boat motion is measured. The acceleration of the boat is measured using an accelerometer, while the distance travelled is estimated using a commercially available submerged magnetic impeller. While it may be seen as a redundancy to measure parameters that are related by integration/differentiation, the error characteristics of the sensors mean that each measurement is only an approximation to reality, thus the outputs are combined using a simple sensor fusion technique (Kalman filtering).

Appendix A1, describes a geometrical method of optimisation that was initially used in the determination of foot force, from the voltage outputs of the foot force sensor.

Included in Appendix A2 are full derivations, and comments on strengths and weaknesses for a number of methods of attitude estimation.

A derivation of the Kalman filter algorithm, used in Chapter 5, can be found in Appendix A3. This derivation works from the general topic of minimum mean square estimation, then introduces the discrete time state space model, to arrive at the final algorithm.

The data acquisition hardware and software, and the methods of data capture are described in Appendix A4.

Chapter 2

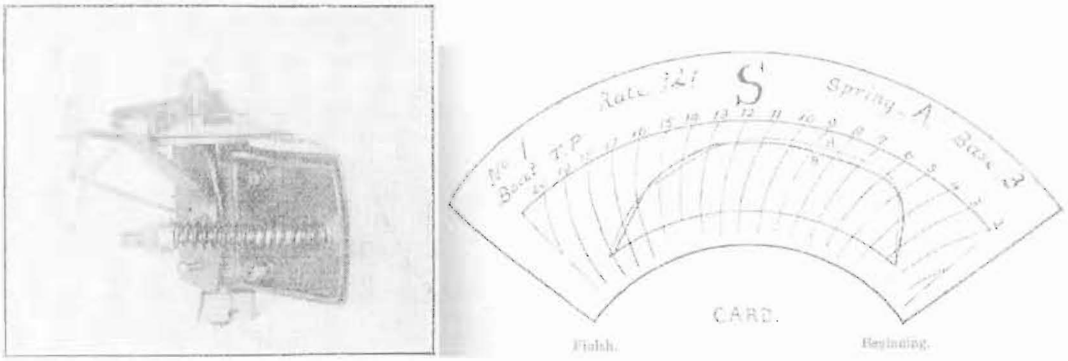
When investigating rowing, one of the main parameters of interest is the propulsive force. As shown in the introductory chapter, the forces that drive the boat are a weighted combination of the force at the oarlock and the acceleration of the rower relative to the boat, thus the force generated at the oarlock is an integral part of rowing.

Measurement of oar force has historically been achieved in three ways. In this chapter these three methods are summarised and then a new method is proposed. This new method is convenient, cheap and yields sound results. The only problem identified with the sensor is a very slow drift in the offset voltage. While a simple method can be used to overcome this drift, it is preferable that it be understood and eliminated.

2.1 Review of Oar Force Measuring Methodology

As discussed in the first chapter, the term 'oar force' can mean the force at the handle, the oarlock or the blade. This is a brief chronology of methods that have been applied to measure oar force.

In 1896 in volume 8 of Natural Science, nestled between 'The Pigments of Animals', and 'Dispersal of Seeds by Birds', is an article that may be regarded as the beginning of rowing instrumentation. 'A Rowing Indicator', by Atkinson [5] details the design of a modified oarlock that includes a spring loaded scribing arrangement that traces the rowing force against rowing angle on a removable plate. In a later issue [6], Atkinson also published another article 'Some More Rowing Experiments', in which he explains the method by which his original indicator has been improved so that it records data from every fifth stroke for a duration of up to 500 strokes! Also in this article he approaches topics such as stroke efficiency and estimates the location of the turning point of the oar (see Chapter 1), which he approximated to be "3 inches above the top of the blade". While Atkinson's method of instrumentation would be unacceptable today due to its modification of the oarlock and the large displacements experienced during operation, it was certainly a great start and it was sadly a long while before such an inspired effort was forthcoming.



(a) (b)

Figure 2.1 (a) Atkinson's original oar force transducer, and (b) an example of the transducer's output [5].

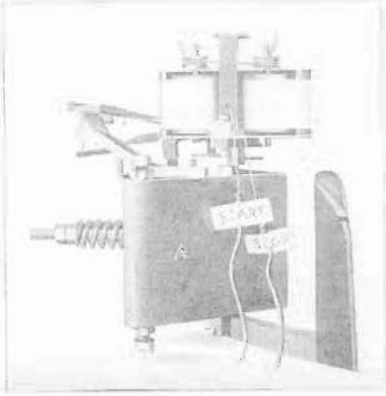


Figure 2.2 Atkinson's modified transducer capable of recording data from up to 500 strokes [6].

Cameron [67] describes a method of estimating the force experienced by the oar via photographic methods. First he subjected an oar constrained from displacement by knife-edges at the handle and button (the part normally engaged with the oarlock) to known loads by hanging masses near the blade of the oar. Adding 7lb masses up to a maximum of 56lb he found that 'the tip deflected 1in for every 7lb tied on the neck'. Supposedly from atop a bridge he then took bird's-eye photographs of the rowing action. Knowing the length of the oar he was able to scale the deflection of the oar from the photograph, and hence estimate the force on the blade. Cameron states that Group Captain H. R. A. Edwards in his book 'The Way of Man with a Blade' had measured force by 'putting strain gauges on the oar' but decided that his own method was more convenient as it did not require additional apparatus within the boat.

It seems that the next generation (1970-1980) of rowing investigators did not share Cameron's views, and the method of choice of oar force measurement was bonding strain gauges to the oar. A representative of this 'school' is Bompa [10]. Bompa's 'measurement oars' had 'four strain gauges placed on the flat side of the oar, 8 cm above the collar...(a thin disc, normal to the shaft of the oar that prevents longitudinal translation of the oar)... and covered with epoxy to make them water proof'.

Since oars are not of standard stiffness, the method of bonding strain gauges to the oar requires that individual oars be calibrated. This is not accommodating for rowers, who like any sportsmen, have their own favourite equipment. This may have been the motivation for Nolte who designed an instrumented oarlock in 1980. A schematic of this is shown below.

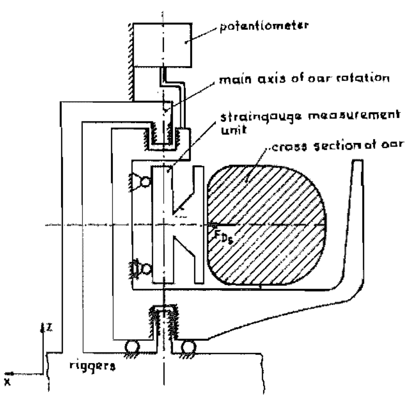


Figure 2.3 Nolte's measurement oarlock [19]

No trace of Nolte's method could be found in subsequent years. Perhaps his oarlock was too complicated or altered the feel of rowing too much. The next method of force measurement, which appears to have been first suggested by Gerber [27] in 1987, is still very much the most popular choice among investigators Teague [64], Kleshnev [39]. Gerber fixed a metal plate and inductive proximity sensor to an oar so that when the oar deflected the distance between the plate and sensor varied (see Figure 2.4). A basic variation on this theme involves the replacement of the inductive sensor and plate with a Hall-effect sensor and magnet respectively. The amount of bending within the oar, and therefore the force causing the deflection is measured by monitoring the distance between the sensor and the plate/magnet. Again, the disadvantage of this measuring system is that each different oar requires a separate calibration, as the stiffness of the oar is the factor relating the force to the measured deflection.

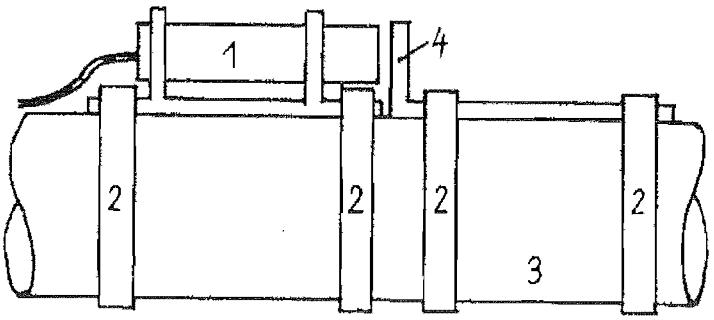


Figure 1 Force-measuring cell for the oar. 1 = inductive sensor, 2 = steel band, 3 = oar, 4 = metal plate. Weight of whole device 80 g, length 10 cm.

Figure 2.4 Gerber's inductive oar force measuring device [27]

A design team at Virginia Tech [65] employed a new method of measuring the force applied by the oar so that they could optimise the design of the rigger. The group 'strain-gauged' an existing rigger using the format shown in Fig 2.5.

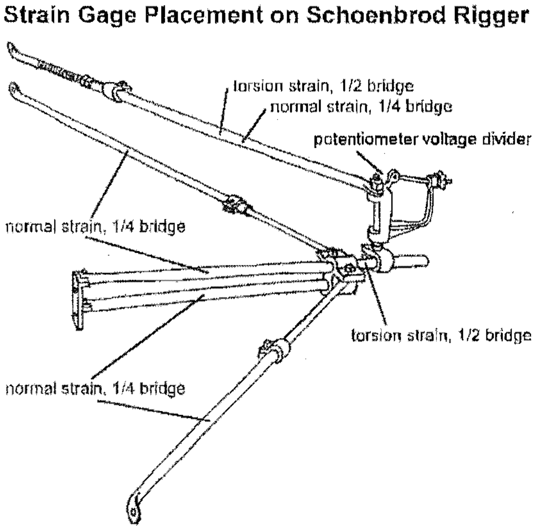


Figure 2.5 Strain gauged rigger used by Virginia Tech [65]

This method of measuring oar force is not suitable for our application, as it necessitates a large number of channels, and is rigger specific.

2.2 Oar Force Measurement within the Department of Mechanical Engineering at the University of Canterbury

Oar force measurement, which has a short history within the Department of Mechanical Engineering (under the direction of Dr Aitchison), has been attempted using a variety of strain gauge based techniques. Figure 2.6 shows strain gauges applied to the pin (on which the oarlock rotates). Theoretically, this method measures the bending stress in the pin caused by a force orthogonal to the plane in which the gauges are attached. In operation the gauged pin is oriented so that forces in the direction of travel are measured. There are a number of problems with this approach, including the fact that variation in the height at which the oarlock is mounted on the pin effects the bending moment. (In operation the oarlock is often propped above the collar of the pin by washers.) Also, while the instrumented pin should theoretically indicate the driving force, it gives no indication of the wasted force, i.e. forces that act perpendicular to the direction of motion.

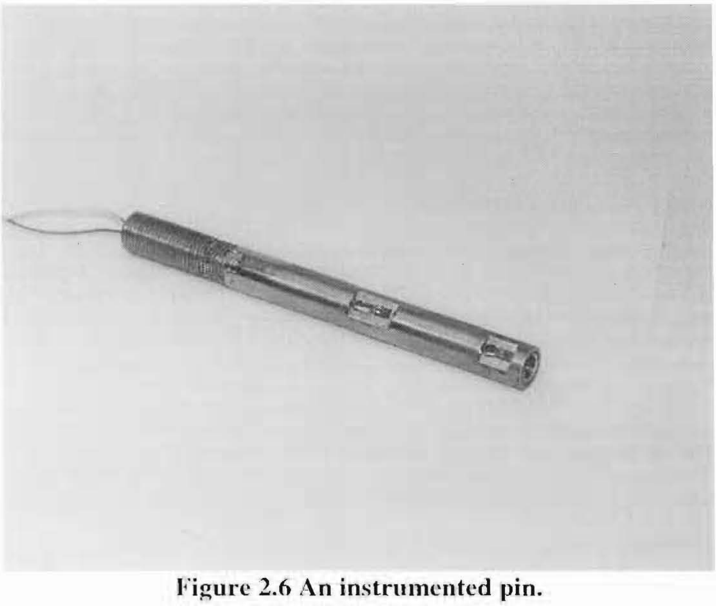


Figure 2.6 An instrumented pin.

Shown in Figure 2.7 is a sophisticated sensor in which a fairly standard design load cell is incorporated into the rear of a Concept II oarlock. Bronze inserts that screw into the top and bottom of the load cell allow for adjustment in the pitch of the oarlock. Stress is induced in the load cell when force is applied to the face of the oarlock since the bronze inserts prevent translation.

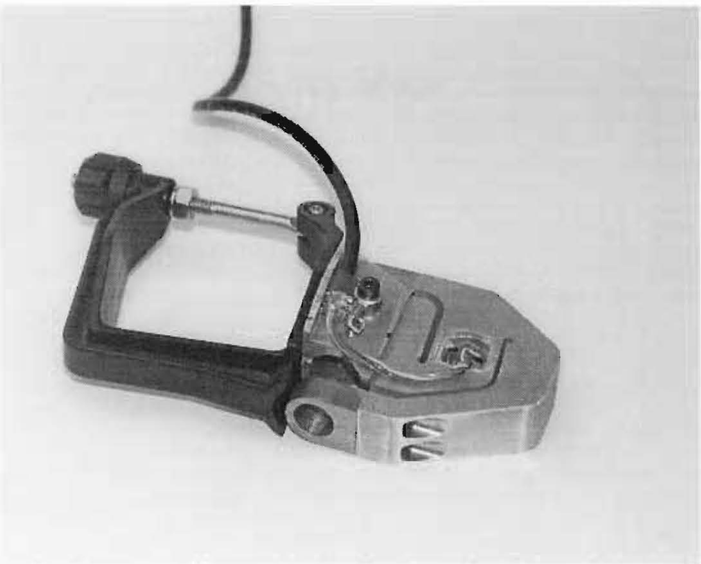


Figure 2.7 An instrumented Concept II oarlock previously made in the Department.

During the design of this sensor, a major issue was designing the protruding load cell to satisfy tight spatial constraints enforced by some types of riggers. While the sensor performed well, it had the disadvantages of relatively high cost and a slow method of changing the oarlock orientation. Also, a rower's performance might be affected by the different frictional characteristics associated with the bronze bushes.

2.3 Requirements

2.3.1 General sensor requirements

Since instrumentation systems now typically comprise sensor, signal conditioning and computer, rather than just sensor and display, the requirements of sensors have changed. Where linearity was once important, it is now important that the sensor have an output that can be 'well approximated' as a function of the measurand. Non-linearities are permissible in a sensor's characteristics so long as they are predictable. Thus the equivalent of linearity is reproducibility and 'identifiability', i.e. a sensor should always have the same response to the same conditions and it should be possible to model the response, so that knowledge of the sensor output is equivalent to knowing the condition of the measurand. Clearly, if one is to work from knowledge of the system's output to an estimate of the measurand, the approximating function must be invertible, i.e. one-to-one. Additional requirements of general sensors are that they do not alter the measurand through their presence, have good quality signals (large range and good signal to noise ratio) and are reliable.

2.3.2 Requirements Peculiar to Measuring Oar Force

The requirements of the sensor can be divided into those imposed by the environment and those imposed by function.

The oar has a high chance of getting wet during rowing (certainly at least some of it must) thus any sensor employed to measure the oar force must either be enclosed so that is 'splash-proof' or designed so that its function is not affected by water. Additionally, the long-term effects of moisture must be considered, i.e. the sensor must be designed so that corrosion cannot take place.

By functional requirements it is meant that the sensor must be suitable for use in its specified role. The role of this sensor is to measure oar force, not for a particular oar or rigging, but for a large range of combinations of these two. As such, the sensor must require little or no adaptation when equipment is changed.

2.4 Concept Development

The first stage in the design of a generally applicable oar force sensor is investigation of riggings and identification of any common component that the sensor must accommodate. Before this was undertaken it had been decided that 'instrumenting' an oar, using either the strain gauge or inductive approach was unacceptable due to the required individual calibrations.

The riggers showed a large range of variability, but one component was found to be almost standard, the Concept II oarlock. It was realized that for the system to be adaptive it must either be compatible with the oarlock or modify the oarlock itself in a way that would not jeopardize its function or general applicability.

2.4 1 Description of the Concept II Oarlock.

The almost omnipresent Concept II oarlock comes in sculling and rowing varieties, the only difference being the scale (see Figure 2.8). Inserting plastic bush plugs into either end of the oarlock cavity controls the included angle that the axis of the pin-cavity makes with the axis of the pin (pitch). The plugs come in pairs, are numbered 'x,y', where x+y=8, and the difference of x and 4 indicates the magnitude of the angle caused by the plug, e.g. 4,4 is a straight plug and 7,1 is the most extreme plug. Generally top-level rowers row with either 4,4 or 5,3 plugs, the higher angle plugs are generally used to correct for a novice's poor technique.

Although, as stated, elite athletes tend to row with a vertical oarlock face, it was decided that the sensor must allow the same degree of adaptability as the oarlock itself. Moreover it was considered to be attractive that the method of adjusting the angle be exactly the same as that used in the oarlock.

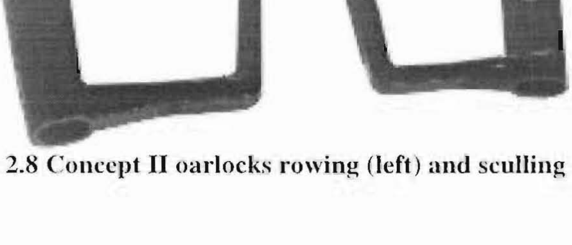


Figure 2.8 Concept II oarlocks rowing (left) and sculling (right)

2.4.2 Considered Concepts

Several concepts considered included modification of the plugs to incorporate measurement facility. Manufacturing the instrumented plugs from bronze would allow for both good 'running' and strain measurement. One of the concepts is shown in Figure 2.9.

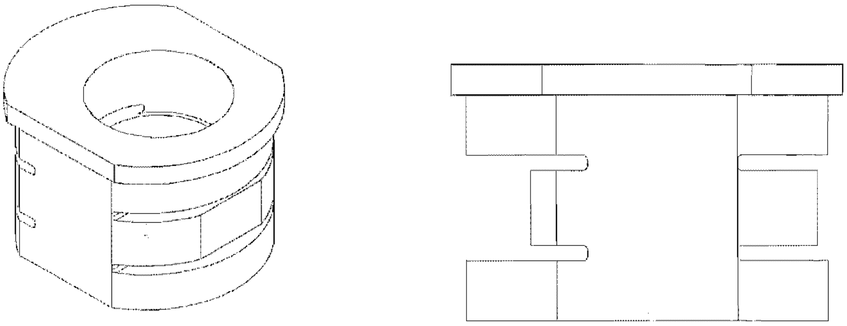


Figure 2.9 Original measuring plug concept

The upper and lower parts of the plug fit tightly into the oarlock cavity, with the central part free. Only the central part of the plugs has contact with the pin, thus when force is applied to the oarlock the 'beams' are stressed. Strain gauges were to be attached across the flats of the beams. The low level of strain in these beams would have necessitated very high gain gauges, which are renowned for poor signal integrity. The plug concept was abandoned for this reason, and also because of high manufacturing costs and the requirement that not just one pair, but a variety of angled plugs be constructed.

While the idea of instrument plugs was abandoned, the idea of working within the cavity of the oarlock was seen to be attractive, as there would be no way that the instrumentation would foul on the rigging. Rather than working strictly to the spatial constraints imposed by the dimensions of the cavity, it was considered that enlarging the hole and making it circular would allow for simpler designs without jeopardising the integrity of the oarlock. One of the ideas is shown below in Figure 2.10. The upper and lower parts fit tightly into the bored cavity of the oarlock. The central piece does not contact the walls of the cavity. Angle plugs are inserted into either side of the central piece so that contact is made with the pin. The columns connecting the extreme and central parts are subject to strain when force is applied to the face of the oarlock.

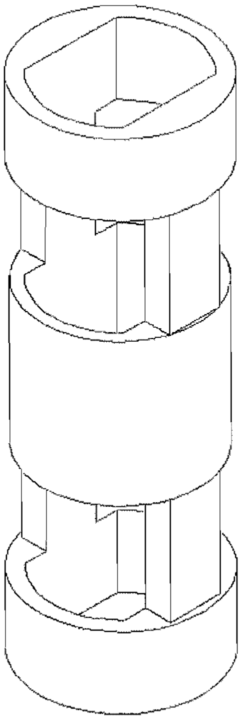


Figure 2.10 Four beam instrumented cavity concept

There were many problems with this design, including dubious angular stability caused by the central placement of the plugs and poor candidate locations for applications of strain gauges. There were two places in which strain could be measured in this design, either the front and rear faces of the beam, measuring strain due to bending, or on the external flats of the beams, measuring shear strain. Both of these sites had problems. Firstly the beams were very thin, allowing very little space for strain gauge placement. Also since bending strain is proportional to the distance from the point of application of force, the strain 'seen' by the gauge would vary greatly over its length. In normal applications, the length of the gauge is small in comparison to the distance from the application of force, so this is not a problem. Measuring shear strain is a way in which this problem can be avoided, since shear strain is not a function of distance for a beam in bending. The problem with measuring shear strain in this design, however, is that the levels would be very low, due to the thickness of the beams required for rigidity of the sensor.

The final concept, and the one eventually used, was to instrument the cavity itself and continue to use plastic plugs inserted in the top and bottom of the cavity to modify the oarlock angle. The contacting and free areas were reversed from the previous design, i.e. the central part fits the cavity while the upper and lower parts fit (with plugs) onto the pin. The idea of measuring shear strain was maintained. The outside of the sensor is simply a stepped cylinder; the inside has a constant internal diameter, making manufacturing easy. The gauged sites are the sections with the smallest wall thickness (0.75mm). Rather than performing elaborate calculations to determine this thickness, it was checked that this would produce a measurable level of strain without causing fabrication concerns for the workshop technicians. The external diameter was chosen by a combination of the latter of the factors and what could reasonably be removed from the oarlock wall without ruining its structural integrity. The height of the gauged sections was determined by the dimensions required for the comfortable placement of a strain gauge rosette with its two gauges at 45° to the longitudinal axis of the cylinder. Placing rosettes on points directly opposite each other on one of the thin sections allows for a full bridge configuration. (Strain-gauge bridges are described more fully in Chapter 4.)

In summary, the sensor was designed using shear stress/strain calculations to ensure that yielding would not take place and that shear strain would be of a measurable level. Simple beam approximations were used to estimate the expected deflection of the upper and lower parts of the sensor. Combining what was desirable and what was easily manufactured determined the final dimensions. The manufacturing drawing for the sensor (generated in Solidworks™) is shown in Figure 2.12. Initially, pitch adjustment plugs were entirely machined, but it was found to be easier to 'turn down' existing plugs so that they fit the modified cavity. This approach also has the benefit that the coefficient of friction between the pin and the plugs will remain unchanged, and hence the 'feel' of rowing affected only minimally.

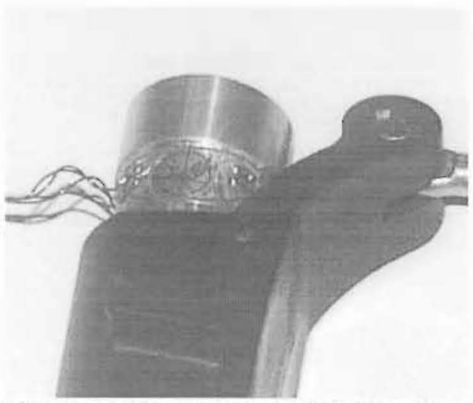


Figure 2.11 Sensor inserted part way into the oarlock, with one strain gauge rosette exposed

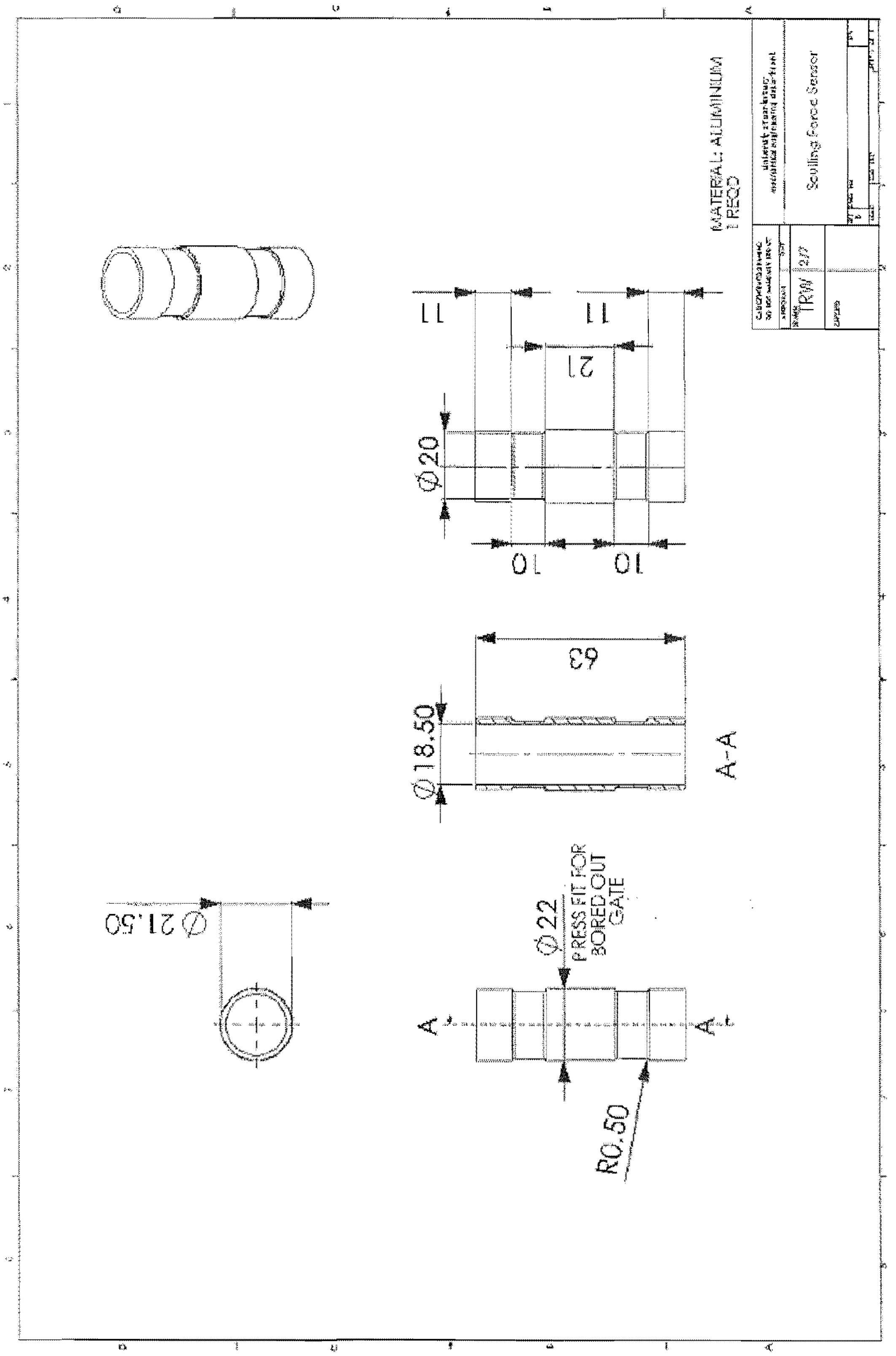


Figure 2.12 Oarlock force sensor manufacturing drawing.

2.5 Sensor Characteristics

The sensor was calibrated in a way similar to the method of Cameron [67], described in the introductory review. A coach-bolt acting as pin was mounted horizontally. The oar handle was prevented from translation using a G-clamp. Masses, suspended from the oar near the blade were added to a hanger of 0.7kg mass up to a total of 27.7kg. The longitudinal dimensions of the oar were measured and moments were balanced about the pin to show that the force at the oarlock was almost exactly three times that suspended for near the blade. Data was collected at 25Hz and averaged over approximately 10 seconds.

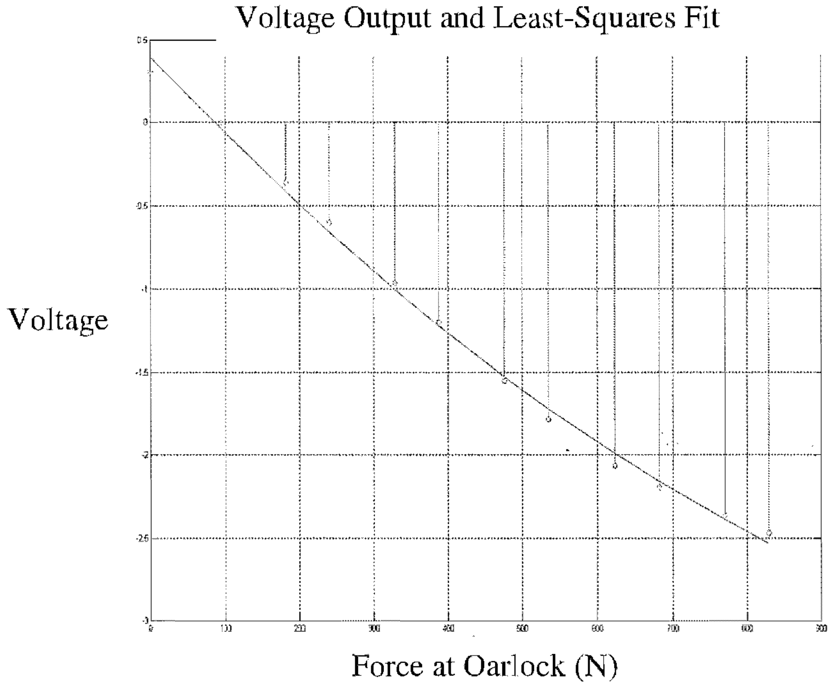


Figure 2.13 Calibration curve for the oar force sensor

It was ensured in all cases that the substantial oscillations caused by the addition of the mass had subsided before measurements were taken. At low levels of force, the response of the sensor was found to be close to linear, however, above 600N (at the oarlock) the output becomes slightly non-linear. A quadratic approximation to the sensor's force response was calculated using least squares. A plot of the sensor outputs and the calibration curve are shown in Figure 2.13. At an amplification of 100 the amplitude and signal to noise ratio were satisfactory.

As mentioned previously, reproducibility is the key to instrumentation. To check the repeatability of the measurements, the sensor was loaded to 27.7kg and unloaded, with the outputs at each level compared. An example of the output of such a repeatability test is shown in Figure 2.14.

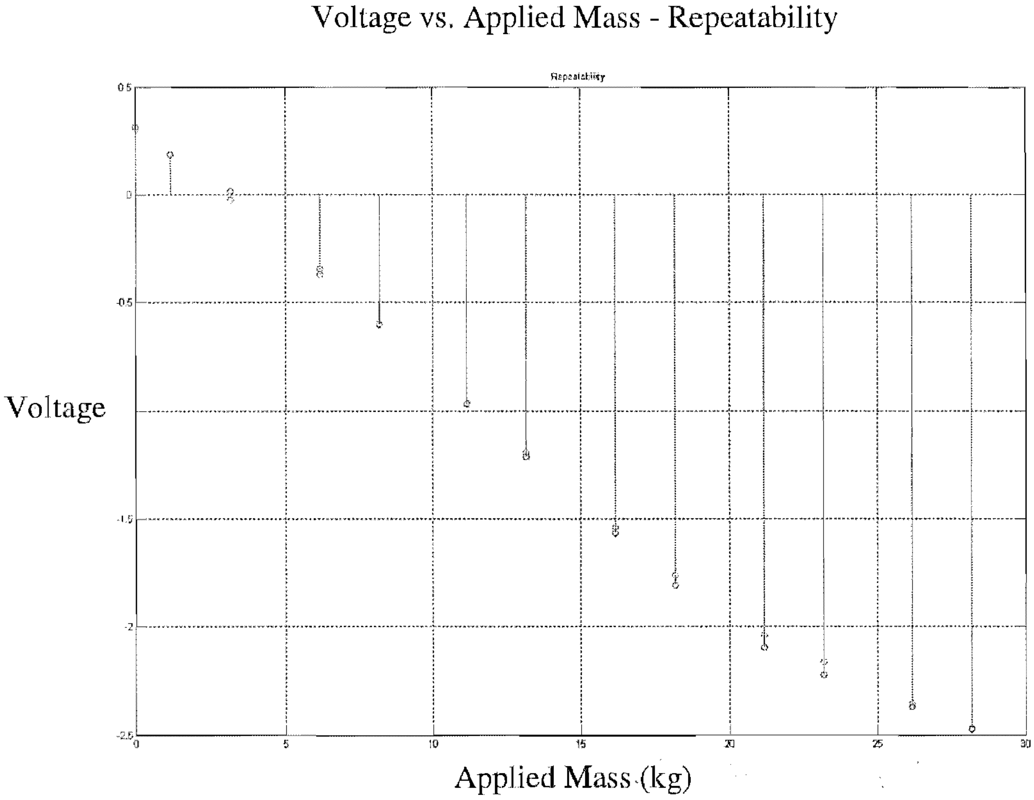


Figure 2.14 Oar force sensor repeatability.

An unexpected characteristic of the sensor was that over long periods of time, e.g. days, the offset value (output value at zero load) drifted slightly. Normally one could attribute this to a thermal problem, but the symmetry of the strain gauge placement combined with the properties of the full bridge in which the strain gauges are combined make this unlikely. The only other possibility is some inconsistency in the amplifier, which, incidentally did exhibit a number of other problems. While the problem should be further investigated, it can be factored out by taking a reading from the sensor when no load is applied and using this to correct the offset value.

2.6 Results

In possession of the three calibration factors C_0, C_1 and C_2 (the offset, linear and quadratic terms respectively), it is simple to convert measured voltages into the force applied at the oarlock, f. Since the voltage is approximated by a quadratic function of force

$$V = C_0 + C_1 f + C_2 f^2 \tag{2.1}$$

The force, f, is a solution to the equation

$$C_2 f^2 + C_1 f + C_0 - V = 0 \tag{2.2}$$

both of which can be found using the quadratic formula:

$$f = \frac{-C_1 \pm \sqrt{\left(C_1^{\,2} - 4(C_0 - V)C_2\right)}}{2C_2} \tag{2.3}$$

In all cases the 'correct' solution is given by assigning a positive sign to the square root term. While the force was determined during post-processing, this simple method of calculation could obviously be implemented in real-time at the currently used sampling rate of 25Hz. Shown below in Figure 2.15(a) is a graphical example of the sensor's output. The rower in this case is a heavyweight experienced female. The shape and magnitude indicated is consistent with intuition and previous research.

The only filtering used on the oar force sensor output was inside the amplifier and in the kilohertz range. This is clearly not suitable as an anti-aliasing filter when the sampling frequency is only 25Hz, however, performing fast Fourier transforms (FFTs) on the data showed that the spectrum of the signal was almost entirely below 4.5 Hz, meaning that aliasing is not a potential problem. An FFT of the signal in Figure 2.15(a) is shown in Figure 2.15(b). The large 'spike' at just below 0.5Hz is the base frequency of rowing, indicating that the rower was performing just under 30 strokes per minute.

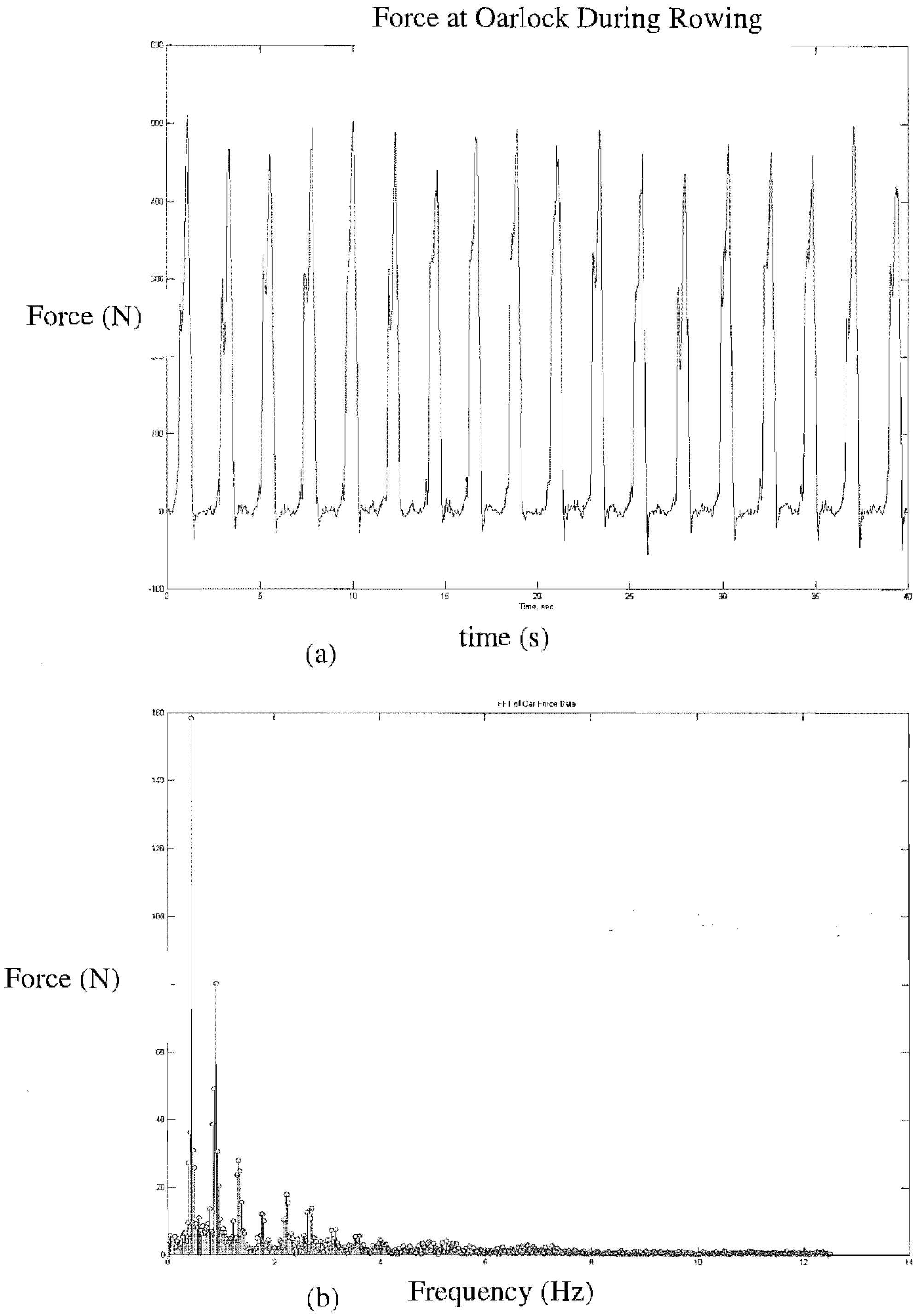


Figure 2.15 (a) Typical oar force data, and (b) the associated frequency spectrum

2.7 Discussion

The designed oar force sensor yields sound results and has a number of advantages of previous methodologies. Firstly, apart from the wire protruding from the rear of the oarlock, the rigging is externally identical to that normally used. The rower may personalise their set-up in the normal fashion, by changing the pitch plugs and stacking washers beneath the oarlock without affecting the sensor characteristics. Another important advantage of the sensor is that the characteristics of the oar are immaterial to its functioning. This is beneficial because the rower gets to use his own oar, and hence row to the best of his abilities, without necessitating a recalibration.

Additional to the functional benefits of the sensor, it is also very easy to manufacture and uses inexpensive materials.

The drift is occurring at such a low rate that it is negligible over the testing period; however, over hours or days it becomes noticeable. Assuming that the other coefficients relating the force to voltage of the sensor do not similarly vary, a simple method of accounting for the drift is to take a zero load recording prior to performing a run and using this value as the new offset.

Based on the aforementioned positive aspects of the sensor, it is thought that further development is warranted. This development might include more detailed design and calibration. An investigation into the drift of the offset value should also be undertaken.

Chapter 3

The application of force at the feet is an important issue in rowing performance. The timing, magnitude and direction of applied force are all variables of interest.

Foot force measurement in general is a topic that has generated much literature over the last century. The following review investigates the methodology that has been applied to the problem over this period, focussing particularly on approaches that could conceivably be used to measure foot force during rowing.

Factors that make the measurement of force at the foot stretcher unusual are the combination of static foot position; the generation of dynamic bi-directional normal forces and the additional requirement of measurement of shear force.

Like all interface measurement problems the foot-force sensor must measure the applied force with as little modification to the interface as possible.

The designed transducer measured normal force and centre of force to a good degree of accuracy but the shear output was very poor.

This section discusses the design of the transducer and the ideal response. The method of data extraction is then detailed. Sample results are shown, the reasons for poor shear measurement performance discussed, and the way in which these problems can be overcome in design and analysis is described. In particular an in-depth discussion of error propagation in least squares estimation is included.

3.1 Foot Force Measurement Review

3.1.1 Introduction

The force at the feet has been measured for many reasons including diagnosis of gait disorders and investigation of pressure ulcer causation. Since 1882, when Beely stood subjects on a sack containing Plaster of Paris [42], state of the art technology has been applied to the problem of measuring foot force. Many approaches have been attempted, including optical, strain gauge, magnetoresistive, piezoelectric and capacitive methods. Not all these methods are applicable to the problem at hand, that is measuring the force at the feet during rowing, and hence only the relevant methods are reviewed. Readers are referred to two comprehensive reviews [17], [42] if a full development of the foot force measuring 'scene' is desired.

In rowing, the oarsman's feet are secured, via rowing shoes, onto the foot stretcher, which is in turn attached to the boat or rigger. The toe of the shoes are prohibited from motion, while the heels rise and fall during the rowing cycle. The possible sites for the measurement of foot force are seen to be within the shoe (between sole of foot and shoe), between the shoe and the stretcher or within the stretcher itself (modification of stretcher). Bearing this in mind, technologies that could potentially be used in one of these sites were researched.

Measuring foot force during rowing is a peculiar problem since the feet are essentially static while dynamic forces are produced. The requirements of the sensor are increased by the fact that normal force is generated in both directions during the rowing stroke – the rower pulls on the stretcher to bring himself forward during the recovery and pushes during the drive. Added to this bi-directional normal force is the presence of a shear force, since the rower does not exert force exactly normal to the stretcher surface. There will also be a lateral shear force, although it is expected that this component will be very small, and thus is ignored in this development.

The potential methods of foot force measurement are significantly narrowed by these requirements. Most methods considered below are not capable of measuring all of these parameters in isolation, but were considered worth reviewing since it would be possible to use a combination of sensors to fulfil the requirements. Also, while some methods seem on the surface to be unsuitable due to their physical size it was thought that the concept of the sensor could be captured and altered to more amenable dimensions.

The following review of foot force measurement methodology roughly divides work into sensors capable of measuring only normal force, those that measure shear force, and multi-component transducers that can measure both shear and normal forces. Within each category is a range of sensing strategies. Following a brief description of each sensor is a discussion of the applicability of the method.

3.1.2 Normal Force Sensors

3.1.2.1 Capacitive Sensors

Miyazaki [52], [53] describes an insole shaped capacitive sensor that is attached to the bottom of the shoe. The sensor is only a few millimetres in thickness, with the change in capacitance caused by the variation in separation between copper foil sheets separated by a foam rubber layer. The sensor was divided into two sensing areas, the heel and the forefoot, with the voltage measured across each of the capacitors related to the force depressing the associated rubber layer. Miyazaki [53] details an error analysis to investigate the effects of a uniformly distributed force versus the same magnitude load applied at a single point. This type of sensor could conceivably be sandwiched between the rowing shoe and foot stretcher plate.

The potential problems concerning such a sensor are the robustness of the design and the coarse information regarding force distribution. Increasing the number of sensing areas could yield more precise information. Some considerable time was spent trying to conceive of a sensor that would reliably measure both shear and normal forces using a capacitive approach. Consider, for example, a capacitive sensor with a large number of pairs of plates. Some of the pairs have exactly the same size and are oriented above one another, separated by an appropriate layer of foam. Other pairs have one plate that is significantly larger than its mate. Since capacitance is proportional to the effective plate area and inversely proportional to the distance between the plates, the pairs that are the same size will be affected by both shear and normal force due to the relative motion of the top plate in the direction of shear. If the plates of the pair of different sized plates are arranged so that any expected shear force will not change the effective area then this pair is insensitive to shear. This is illustrated in Figure 3.1. The thick lines represent the upper and lower capacitor plates as viewed in profile. The rectangle indicates the effective area of the plate pairs. In the left hand side case, in which no shear is applied, both capacitors have the same effective area. When shear is applied (right), the upper plates are displaced, reducing the effective area of the capacitor formed by the plates of similar size, while the effective area of the other capacitor remains unaffected. Note that it is assumed that the size of the overall sensor is much larger than the dimensions of the plates, so that the rotation of the upper plates caused by deformation of the rubber layer during shear loading is minimised. In possession of the outputs of the various plate pairs it should be possible to separate the effects of normal and shear force. Adding complexity to the required analysis is the fact that the force is not uniformly distributed. A secondary effect that would also cause difficulties is the existence of a coupling between shear and vertical displacements, i.e. displacements of the upper plates due purely to shear will reduce the vertical plate spacing. The magnitude of this effect would be dependent on the properties of the material between the plates. While a stiff material would reduce this cross-coupling effect, it would also decrease the sensor's sensitivity.

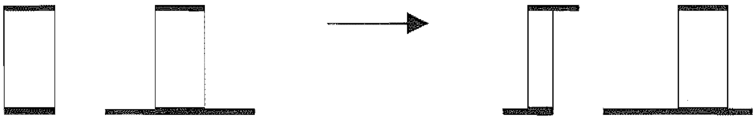


Figure 3.1 Shear and normal force sensing capacitive sensor concept.

Comments on the Applicability of Capacitive Sensors

The idea of a capacitive sensor was abandoned due to the potential robustness problems (separation of glued layers), difficult analysis and warnings from technicians about poor signal integrity. Also note that a capacitive sensor capable of measuring a tensile (pulling) force is difficult to conceive.

3.1.2.2 Force Sensing Resistors

The Force Sensing Resistor (FSR) is a relatively new type of sensor that is being used to measure interface pressures in a range of applications. They are approximately 0.5mm in thickness and come in a range of sensing area sizes and shapes. Their basic construction and method of operation is as follows. Two plastic film discs are separated by a circular ring at their perimeters. The upper disc has interdigited electrodes printed upon it and the lower is coated with a conductive polymer. There is very little contact between the electrodes and polymer when no force is applied to the FSR, but when a compressive force is applied to the FSR the area of contact increases and the resistance decreases, as more current flow occurs through the conductive polymer. This type of technology has been applied to the measurement of foot force in two different manifestations, discrete sensors and matrix insoles. Zhu et al [71] and Abu-Faraj et al [1] describe the development of an FSR measurement system based on the former approach. Using discrete sensors at the interface requires the identification of anatomical sites involved in the transmission of force, a non-trivial task. Two commercially available matrix insole type systems employ technology that is essentially that described above. The Musgrave Footprint system is a matrix of 2048 3x3 mm FSRs [17]. It was not possible to find any information on this product. The second product is the F-Scan system produced by TekScan, shown in Figure 3.2. The F-Scan is a very thin (0.1mm) insole comprising 960 sensors that are formed using conductive and resistive inks [17]. With the insole, which may be cut to size to fit the shoe, comes proprietary Windows software, and acquisition card [www.tekscan.com]. It has been reported that 'calibration between sensors was found to be poor and the sensors showed significant wear with use' [17]. Personal experience with FSR type sensors has shown the output to be heavily dependent upon the operating temperature. It is not known, but considered unlikely that thermal compensation for each of the sensors forming the matrix is included.

Comments on the Applicability of FSR type sensors

It was considered that discrete FSRs within the shoe would be too much trouble to locate on the foot. A reliable, robust and temperature compensated insole system, if such a product exists, would be almost ideal if a full investigation of the plantar pressure distribution were being made, but the expected cost and volume of data yielded by the sensor make it inappropriate for our purpose. Note also that FSRs and similar sensors could not be used to measure the pulling force during recovery. Thus if an insole system were used an additional sensor would be required to measure this force.

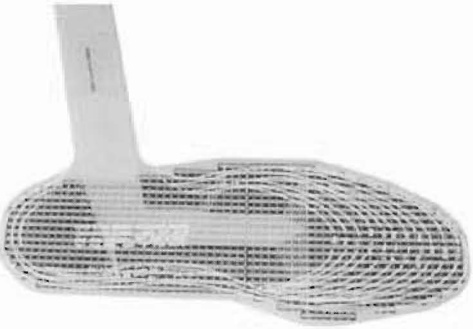


Figure 3.2 The F-Scan Insole, by Tekscan

3.1.2.3 Piezoelectric Sensors

Sensors that exploit piezoelectric materials have also been designed in discrete and insole matrix formats. An example of the former is the transducer developed by Gross and Bunch [30]. In these sensors, copper tabs were soldered to either side of small piezoelectric ceramic squares. The sensor was constructed as shown in Figure 3.3. Eight sensors were positioned under the insole of a shoe at prescribed anatomical sites. More recently Nevill used a piezoelectric film to design discrete normal force sensors [17]. Hennig et al [33] developed a matrix insole of 499 4.78mm square, 1.2 mm thick piezoelectric ceramic sensors, also shown in Figure 3.3. While this sensor had excellent sensing characteristics, Nevill notes that such arrays 'can be difficult to construct and can be subject to rapid mechanical fatigue'. He also suggests that 'both problems can be reduced by using piezoelectric polymer film.' [17].

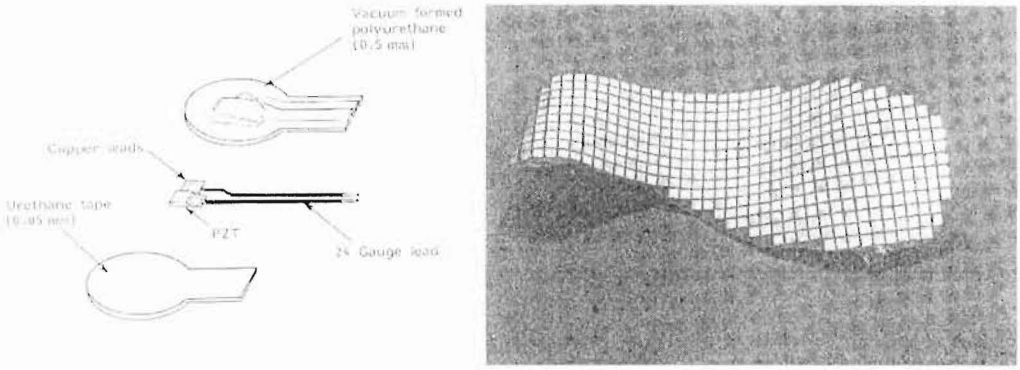


Figure 3.3 Two types of piezoelectric sensors that have been used. (Left) The discrete sensor used by Gross et al [30], and (Right) the matrix insole of Hennig et al [33].

Comments on the Applicability of Piezoelectric sensors

While the piezoelectric effect has been known for some time, it seems that the effective use of piezoelectric materials is still best left to those in commercial operations. It would have been interesting to research piezoelectric materials further, but it was thought that the design of a reliable piezoelectric sensor would constitute several years' work in itself. The sensors, if available commercially are subject to the same criticisms as the FSR type transducers, namely discrete sensor location would be problematic, the insole would yield too much information and neither yield information on the tensile force. Incidentally piezoelectric film can be sensitive to both shear and normal stress [22], meaning that they are potentially the ultimate sensor type for measurement of multi-component interface forces.

3.1.1.4 Strain Gauge

The strain gauge, as in most areas of force measurement, has been used to great extent in the measurement of foot force. The following is a very brief review of the applied methods.

Beam Type Sensors

Stott et al [62] describe an apparatus consisting of twelve beams, each 1.4cm wide and 25 cm long, connected by pin joint at each end to a strain gauge load cell that is in turn pin-jointed to a support frame. The load cell pairs are used in such a way that only the longitudinal tension of each of the beams is measured. An ink imprint is made as the subject walks on top of the beams so that it is known exactly where the foot is oriented with respect to the beams.

Comments on the Applicability of Beam Type Strain Gauge Sensors

By using discrete beams, the sensor allows for one 'degree of resolution' in the calculation of the instantaneous centre of force. It should be possible to use the bending stress in a beam to also approximate the centre of application of force for each beam. Combining the data would allow for at least a rough approximation of the centre of force. It may also be possible to gauge the fore and aft faces of the beams so that shear forces could be similarly estimated. These two suggested modifications would require modification of the beam supports and would also add to the number of channels required to record the data. While the idea is intuitively appealing, the number of channels required would be prohibitive in our application. It is also doubtful that a sensor of this type could be 'downsized' to an extent acceptable for use on a rowing boat. While this sensor could be used to measure tensile force, it would not give information regarding the distribution, since the shoes would be likely to be attached to only a small number of the beams, and torsion of the beams is not being measured.

Proving Ring Type Sensors

Strain gauge proving ring type load cells have been used in the measurement of foot force by both Dhanendran [21], [18] and Arvikar [3]. Dhanendran created a close packed matrix of load cells suitable for use in a walkway (see Figure 3.4). Arvikar, on the other hand, expected his patients to balance on six proving rings for what was obviously a static measurement!

Comments on the Applicability of Proving Ring type Strain Gauge Sensors

The Dhandendran force plate gives information on the total normal force and could be used to accurately estimate the centre of force. The problem with this sensor in our application is its size, which could not be reduced if a good response was required of the proving rings. A large number of channels would also be required and extra sensors would be required for the measurement of shear.

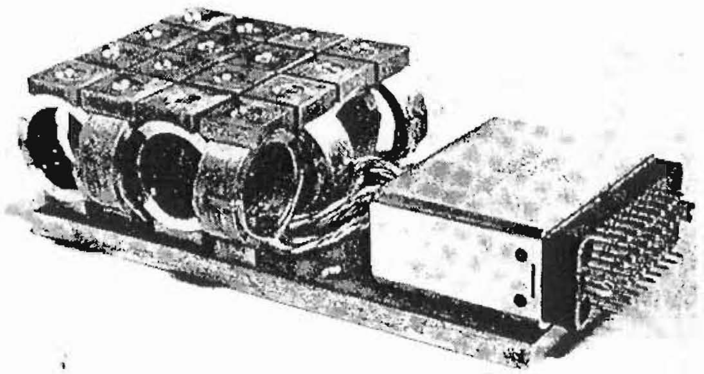


Figure 3.4 A quarter of Dhanendran's proving ring matrix [21].

Miniature Cantilever Strain Gauge Sensors

Soames [61] designed a very compact beryllium copper sensor, shown in Figure 3.5. A semiconductor strain gauge was used, presumably because of their high gain. It was noted that 'To give an accurate measurement of pressure requires that the soft tissues of the sole of the foot are sufficiently compliant to distort the cantilever without significant change in the pressure exerted, that the surface beneath the transducer does not deform to an extent that it obliterates the recess, and that the load is evenly distributed on the cantilever.' Fifteen of these sensors were located on one foot of the subject.

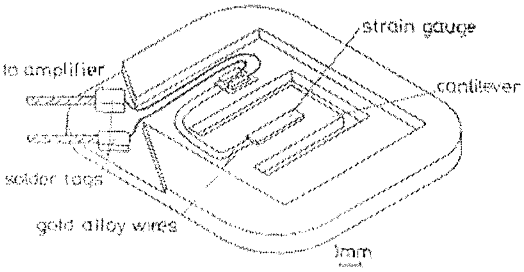


Figure 3.5 The sensor used by Soames et al [61]. This view is from below, showing the recess into which the central cantilever is expected to deflect.

Comments on the Applicability of Miniature Cantilever Strain Gauge Sensors

While the design and manufacture of these sensors is admirable they suffer from the same problem as all previously mentioned discrete interface sensors.

3.1.3 Shear Sensors

A number of investigators have designed discrete shear force sensors that are placed at the interface [17],[40],[55],[68]. All use essentially the same principle. The basic sensor, which appears to have first been used by Pollard and Le Quesne [55], consists of two thin metal discs, one with a groove, and the other with a matching ridge. The upper disc has a magnet mounted in the ridge and the second has a magnetoresistive (MR) sensor located centrally in the groove. The two discs are separated by a rubber element to oppose relative motion of the discs. Shear force causes the upper disc, and therefore the magnet, to be displaced by an amount proportional to the applied load and the MR sensor gives an output related to the displacement of the magnet. Williams [68] designed a sensor based on this approach capable of biaxial shear and normal force measurement. The purpose of the sensor was to investigate the forces involved at prosthetic limb interfaces. The shear force part of the sensor is essentially two of the aforementioned sensors stacked on top of each other, as shown in Figure 3.6. The normal force part of the sensor consisted of a strain-gauged circular diaphragm that was forced onto a central indenter.

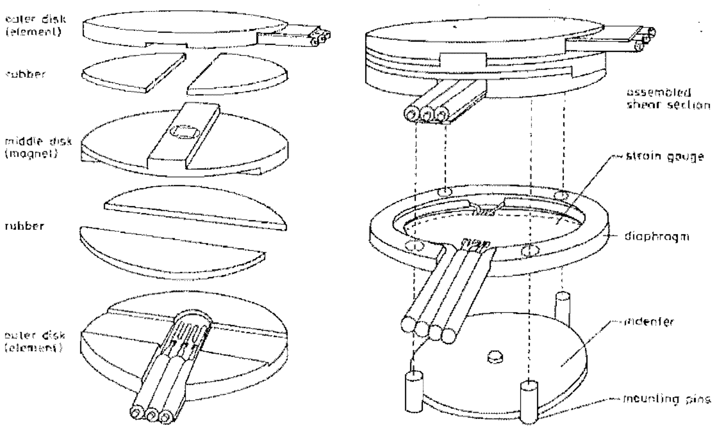


Figure 3.6 The multi-component force sensor used by Williams [68] to measure stump-socket interface forces.

A similar method of shear measurement was applied by Lebar et al [40]. An LED and solar cell were located opposite one another in a circular bronze housing. Located at the intersection of the diameter perpendicular to that connecting the LED and solar cell are two beam spring elements. A disc with a ridge that has a central triangular notch is inserted into the first part so that the ends of the ridge deflect the two springs. In this position light passes through the notch unimpeded. When a shear force is applied to the upper plate the springs allow the disc, and therefore the notch, limited progress in the direction of shear. As the notch displaces the amount of light intersected by the ridge increases and hence the motion is sensed by the solar cell. The sensor components are each 15mm in diameter and 3.8mm in thickness.

Comments on the Applicability of Discrete Shear Force sensors

Again, these sensors, since discrete, would require to be fixed to the inner sole of the shoe at sites where load transfer was expected. The position of the sensors would have to be altered for each rower. Also, placing sensors within the shoe can modify the way in which the force is transferred. It was thought however, that a similar method of shear sensing could be used exterior to the shoe. Consider, for example, a normal force sensor that is securely fixed to the top plate of a shear force sensor, the bottom plate of which is fastened to the foot stretcher. A possible difficulty with this type of sensor would be finding an adhesive strong enough to keep the layers of the sensor bonded together during the phases of the rowing cycle when tensile forces would tend to pull them apart.

3.1.4 Multi-Component Sensors

3.1.4.1 Force Plate

A common use of piezoelectric materials in the measurement of foot force is in load cells supporting the corners of a plate, an arrangement is known as a force plate. While some researchers have designed their own force plates using strain gauge load cells [31], the commercially available Kistler force plate is almost omnipresent in the field of biomechanical measurement. The Kistler force plate is supported by four tri-axial piezoelectric load cells. Since the reactions at each corner of the plate are known, it is possible to estimate the total force in each direction (normal force and both longitudinal and lateral shear forces) and the centre of applied force.

Comments on the Applicability of Force Plates

The force plate yields the data required, but commercially available units have the disadvantages of high cost (tens of thousands of dollars) and bulk. The principle of operation is attractive and it is easy, in theory, to see that a sensor using a similar approach could be designed to be of a more convenient size.

3.1.4.2 Pressure Platform

Giacomozzi and Macellari made an interesting and useful sensor by placing a pressure sensitive mat on top of a standard force platform [28], [49]. Their pressure sensitive mat used a principle similar to the FSR described above. The upper and lower layers of the sensor are a flexible Kapton (a polyimide film manufactured by Dupont) sheet and printed circuit board respectively. Parallel conductive tracks are printed on each of the layers, with the layers oriented so that the tracks are orthogonal. A layer of conductive polyethylene separates the tracks. Knowing which part of the foot is in contact with the ground (from the pressure mat) and the total shear force (from the force plate), they were able to investigate the shear stress acting on small areas of the foot during gait. Redundant data from the compound sensor was also used to investigate other areas of interest.

Comments on the Applicability of the Pressure Platform

While it was not seriously proposed that this sensor be used in the application at hand, it is obviously subject to the combined criticisms of the force plate and matrix type sensor, the way in which the sensors are combined to form a transducer with excellent capabilities is of interest. A comparable method would see some normal force distribution sensor mounted on a rigid plate and supported by multi-component load cells.

3.1.4.3 Strain Gauged, Column Mounted, Cantilever Sensor

Recently Davis et al [20] described their 'Device for Simultaneous Measurement of Pressure and Shear Force Distribution on the Plantar Surface of the Foot.' This device consists of an array of strain gauge sensors that are each composed of two parts.

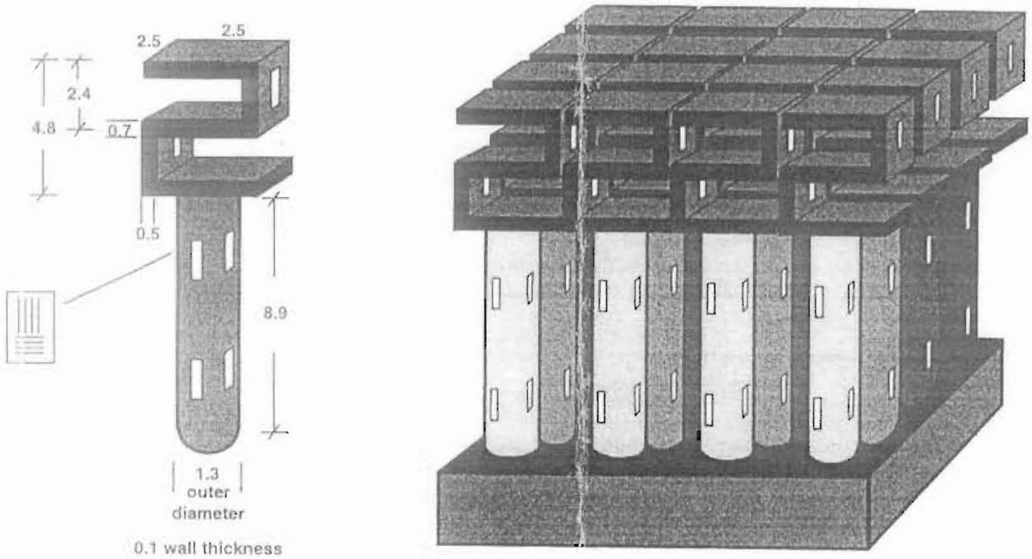


Figure 3.7 A single element and packed matrix of sensors developed by Davis [20]. The exploded rectangle on the left indicates a strain gauge T-rosette.

The upper part, designed to measure the compressive force is an 'S-shaped cantilever'. The shear-sensing element is an aluminium tube on which the upper part is located, see Figure 3.7.

Comments on the Applicability of the Strain Gauged, Column Mounted Cantilever Sensor

The dimensions of this device are such that it is unsuitable for our purpose, further it is not easy to see how such a sensor could be reduced in size. However, the method by which shear is isolated from the normal component, and the idea of a two-component sensor are interesting. Note also that this sensor would be incapable of measuring the distribution of force during the recovery since the shoes would have to be connected to a finite number of the sensing elements, which are incapable of measuring bending moment.

Note: The interface sensor designed by Williams and detailed in the Shear Sensors section is truly a multi-component sensor, but was included in the previous section due to the fact that two of its axes used exactly the same technique as the dedicated shear sensors.

3.2 Summary of Review

The only sensors in the preceding review that could be implemented to measure normal force in both directions and longitudinal shear are the force plate and Davis's device. This is ignoring geometric constraints, which rules out the use of the latter (assuming that a force plate could be miniaturised to a sufficient extent). Davis's sensor is also unacceptable in the number of channels involved and since it would only measure the load distribution when normal force was compressive, i.e. pushing down on the sensor.

All discrete sensors that are placed within the shoe are undesirable due to the time it would take to locate anatomical sites of load transfer, also no discrete 'in-shoe' sensor can conceivably be used to measure force as the rower pulls on the stretcher.

A possibility however is the use of discrete sensors in an external role, say as the support of a normal force sensor. ('Support' entails both physical constraint and the addition of information.) An example of this would be mounting a rigid sensor capable of measuring normal forces on supports that incorporated shear sensors. Mounting a normal force sensor that had some cross-axis effect on shear sensors that were relatively free of this contamination would enable shear and normal forces to be accurately evaluated. In an even more ideal situation a multi-component sensor (shear and normal) could be mounted on shear sensors. If this approach were taken, the shear force could be estimated using some optimal combination of the sensor outputs. The cross-axis effect of the multi-component sensor could be well approximated, and knowing the shear to an accurate level, the normal force and distribution could be well estimated. These comments also apply to any combination of complementary or redundant sensors such as a shear sensor supported by normal force sensors or a multi-component sensor and normal force sensors. Alternatively, a multi-component discrete sensor could be used to support a rigid plate, essentially creating a force plate.

It was stated in the introduction of the review that the sites available for instrumentation were within the shoe, between the shoe and stretcher, or the stretcher itself. The first option has been shown to be undesirable and most of the discussion has centred on placing a sensor between the shoe and stretcher. To show that this is the most sensible option requires the elimination of the third site – modification of the foot stretcher structure. While this seems an attractive option, since it would be possible to directly relate the strain in the stretcher support to the applied load, it is unachievable due to the large range of foot stretchers that are currently in use. If it were only required that a single boat be instrumented then it may be possible to design a specially gauged support structure, but the aim of this work was to design the instrumentation to be as generally applicable as possible.

The required measurands, spatial constraints and inter-boat variability lead to the conclusion that the sensor must be at least in part rigid (this is imposed by the bi-directional normal force), and fit in between the shoe and the stretcher surface. The following is a design that was intended to fulfil these criteria.

3.3 Sensor Description

The foot-force transducer was designed to measure bi-directional normal and longitudinal shear forces, as well as the approximate centre of applied normal force.

A rigid rectangular plate is supported at each corner by slender beams that are cantilevered from a support central to the longitudinal axis of the plate. Thin slots cut through the plate form the beams. Generous radii blend the cantilever supports to feet at either end. The feet fix by means of socket head cap screws to a location plate. Milled recesses in the support plate prevent the motion of the upper plate in the longitudinal direction. Beyond the recesses, a small clearance allows limited vertical deflection of the upper plate but prevents potential overloading of the support beams. Both the upper and support plates are monolithic, CNC machined from Aluminium alloy.

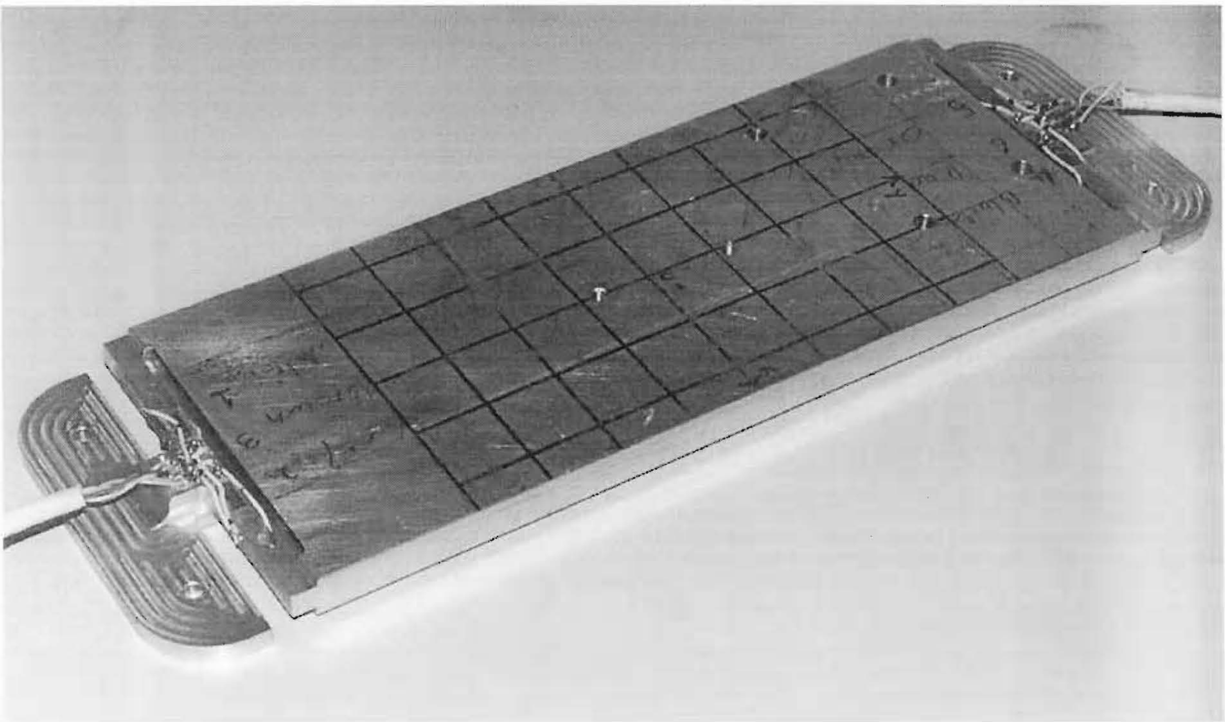


Figure 3.8 The designed foot force sensor, without base plate.

Twelve strain gauges are bonded to the cantilever beams, two on the upper and lower surfaces of each beam, and one on the exposed vertical face of each beam. The gauges on the upper and lower surfaces are placed ostensibly to monitor the normal force (tensile and compressive) while the gauges on the fore and aft faces are placed to measure shear. The upper and lower gauges on each of the beams form a half bridge, as do the 'shear' gauges at each end of the sensor. The bridges are completed by gauges that are bonded to an aluminium block so that thermal effects are minimised.

Not knowing the approximate magnitudes of the force components experienced during rowing meant that the sensor design had to be conservative. The overall dimensions of the sensor were defined by what could be fit onto a reasonably representative foot-stretcher. The beams were sized using simple beam theory, to ensure that yielding would not occur, but that a measurable value of strain would be induced.

In Figure 3.9, the sensor with main dimensions is shown with its location plate.

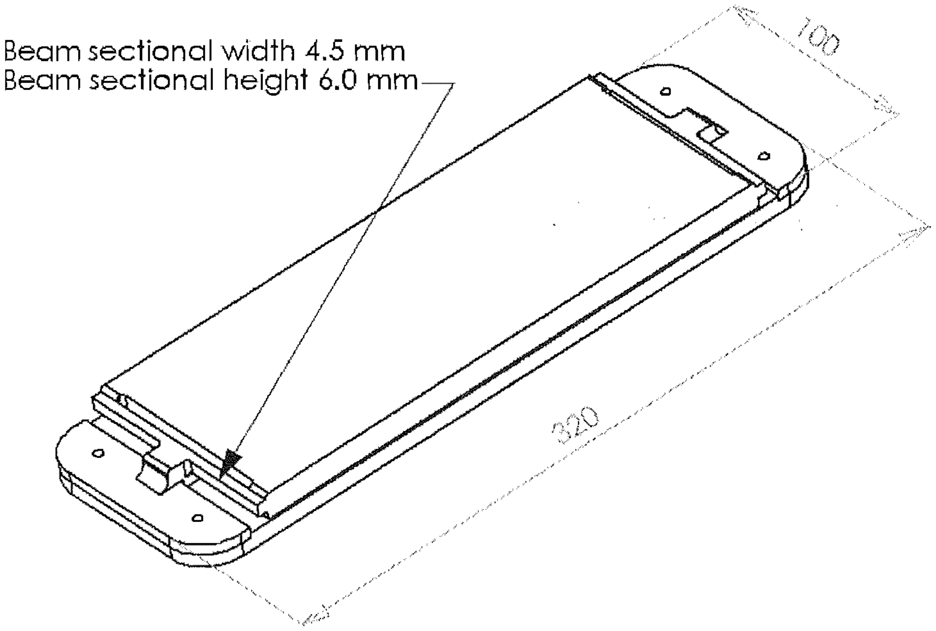


Figure 3.9 Main dimensions of the foot force sensor.

The next section describes the theoretical outputs of the strain-gauged beams in response to general loadings, and the way in which the loading condition can be implied from the data.

3.4 Ideal Sensor Theory

3.4.1 Sensor Loading

Consider an arbitrary 2D distributed loading over a plate of length, L_p, and width, w. The loading can be reduced to a central normal force, two orthogonal moments and a shear force using simple statics.

Consider first the case when all loading is normal to the sensor plate. Define the normal arbitrary load as being made up of discrete loads, Δf_i, acting at coordinate (x_i,y_i) with the origin of the coordinate system at the plate's centre. The long axis of the plate is Y the short axis is X. Shown in Figure 3.10 is the sensor plate geometry and the application of a single discrete load, Δf_i. The letters at the corners of the plate, *F*, *A*, *L*, *R*, denote fore, aft, left and right respectively. The reaction at the rear of the plate is denoted R_A i.e. Reaction Aft.

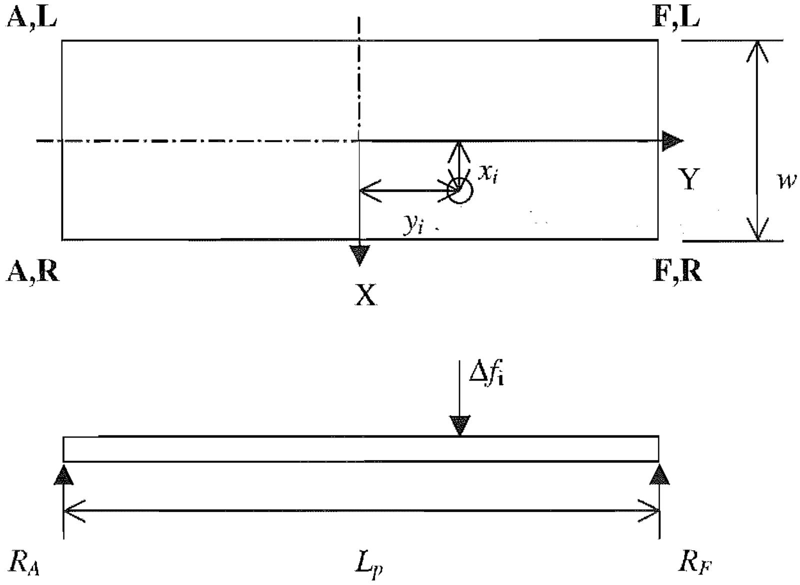


Figure 3.10 Theoretical normal sensor loading. Diagram showing position of discrete load Δf_i, dimensions of plate and labelling of corners.

The net normal load on the plate is the sum of all discrete loads

$$F_N = \sum_i \Delta f_i \tag{3.1}$$

The moments about the X and Y-axes of the plate at the origin are given by

$$\begin{aligned} M_X &= \sum_i \Delta f_i y_i = y_c F_N \\ M_Y &= \sum_i \Delta f_i x_i = x_c F_N \end{aligned} \tag{3.2}$$

where (x_c,y_c) is the coordinate of the point load which is equivalent to the distributed load. Assuming that the plate can be adequately modelled as being simply supported the reaction to the normal loading and M_X at each end of the plate is calculated to be:

$$R_{F,A} = \left(\frac{F_N}{2} \pm \frac{M_X}{L_p}\right) = F_N\left(\frac{1}{2} \pm \frac{y_c}{L_p}\right) \tag{3.3}$$

where subscripts F,A denote fore and aft. Similarly the load supported at each side, due to the moment M_Y is:

$$R_{L,R} = \frac{\pm M_Y}{w} = \frac{\pm F_N x_c}{w} \tag{3.4}$$

where L,R denote left and right. Adding together the contributions due to the vertical loading, the following four vertical reactions result:

$$\begin{aligned} R_{LFn} &= F_N\left(\frac{1}{4} + \frac{y_c}{2L_p} - \frac{x_c}{2w}\right) \\ R_{RFn} &= \mathrm{F}_N\left(\frac{1}{4} + \frac{y_c}{2L_p} + \frac{x_c}{2w}\right) \\ R_{LRn} &= \mathrm{F}_N\left(\frac{1}{4} - \frac{y_c}{2L_p} - \frac{x_c}{2w}\right) \\ R_{RRn} &= \mathrm{F}_N\left(\frac{1}{4} - \frac{y_c}{2L_p} + \frac{x_c}{2w}\right) \end{aligned} \tag{3.5}$$

where R_{LFn} indicates the reaction at the Left Front corner due to Normal loading etc. The reactions show a 'complementary nature' in the signs associated with the centre of force coordinates. This is an important feature that is referred to in later sections.

If a uniform shear loading, S, is introduced, the reactions due to this force at each of the corners of the plate are simply

$$R_s = \frac{S}{4} \tag{3.6}$$

where the subscript s denotes reaction due to shear. Shear is assumed to be positive when in the positive Y direction.

3.4.2 Beam Strains

The beams that are loaded at each corner by the normal and shear reaction forces are of length l, breadth b and depth d. The material of which they are constructed is of Young's modulus, E.

The strain developed in the beams and measured by the gauges placed on the upper and lower surfaces, at distance l_s from the external ends of the beams are given by:

$$\varepsilon_{LFn} = \pm \frac{R_{LFn} l_s d}{2EI_n} \tag{3.7}$$

where $I_n = \dfrac{bd^3}{12}$

where it is assumed that the previously calculated reaction forces act at a point at the end of the beams. The positive/negative signs in front of the expressions indicate that when subject to a particular vertical loading, one gauge of each pair is in tension (+), while the other is in compression (-).

The strain due to the shear loading is measured by gauges mounted on the external faces of the beams and are

$$\varepsilon_s = \pm \frac{S l_s b}{2EI_s} \tag{3.8}$$

where $I_s = \dfrac{db^3}{12}$

The pairs formed by gauges measuring shear at the same end of the plate will be in the same stress state (compression/tension) at all times (assuming uniform shear distribution).

3.4.3 Transducer Outputs (Ideal)

While the simple approach given above suggests a model for the strains at each of the gauged sites in terms of the plate geometry and applied loading, it would be naïve to believe that the coefficients of proportionality were exactly defined by the physical measurements of the plate. Among other factors, nonlinearities in the stress state of the transducer at the gauged sites and inaccuracies in strain gauge placement will lead to deviations from the ideal case. This section considers the ideal outputs of each of the half-bridges. In the next section, possible deviations from the ideal are considered and a method in which to use the sensor in the presence of uncertainties is developed.

Each strain gauge pair measuring normal force is connected in a half bridge configuration. The strain gauges are of gauge factor g, i.e.

$$\frac{\Delta R}{R} = g\varepsilon \tag{3.9}$$

where R is the nominal resistance of the gauge, and ΔR, is the change in resistance due to strain.

The gauge in tension has resistance $R_T = R + \Delta R$, while the gauge in compression has resistance $R_C = R - \Delta R$.

The gauges are arranged in the bridge as shown in Figure 3.11, with the outputs calculated as follows

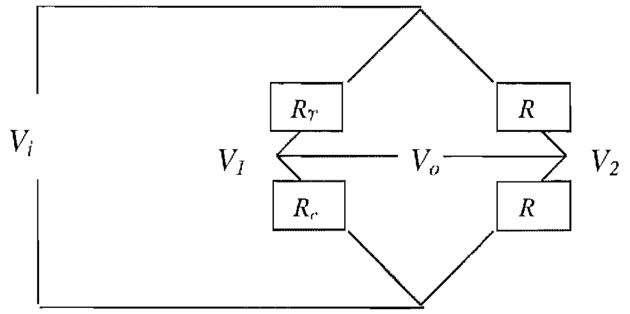


Figure 3.11 The half-bridge configuration used for the normal channels of the foot force sensor. V_i is the input voltage, and $V_o = V_2 - V_1$ is the measured output.

$$
\begin{aligned}
V_o &= V_2 - V_1 \\
V_1 &= \frac{R_C}{R_T + R_C} V_i = \frac{R - \Delta R}{2R} V_i \\
V_2 &= \frac{R}{2R} V_i \\
V_o &= \left(\frac{R}{2R} - \left(\frac{R - \Delta R}{2R} \right) \right) V_i = \frac{\Delta R}{2R} V_i = \frac{g\varepsilon}{2} V_i
\end{aligned}
\tag{3.10}
$$

Thus, each normal strain gauge pair yields an output that is directly proportional to the strain at the site of the pair. The shear gauges, which have the active gauges mounted diagonally opposite each other in a bridge format yields the same result.

Substituting (3.7) and (3.8) into (3.10) gives expressions of the form

$$
\begin{aligned}
V_n &= \frac{g l_s d}{4EI_n} F_N \left(\frac{1}{4} \pm \frac{y_c}{2L_p} \pm \frac{x_c}{2w} \right) V_i \\
V_s &= \frac{g l_s b}{4EI_s} S V_i
\end{aligned}
\tag{3.11}
$$

where V_n is the form of the voltage output of a normal channel, and V_s is the general form of the shear bridges output. Recall the complementary nature of the reaction forces due to normal loadings. This is reflected in the output channel associated with each corner of the plate.

3.5 Deviation from the Ideal Transducer Model

3.5.1 Causes and Effects of Deviations

The basic development assumes that the system behaves in a completely linear manner, obeying the simply supported assumption and that the strain gauges are mounted precisely (collinear with the neutral axis of the strain element) and equivalently (all strain gauges mounted the same distance along the strain element from the point of application of load). The linearity of the actual load cell (independent of the gauges) depends on its degree of symmetry.

While it is naïve to think that all the above assumptions will be totally fulfilled, the development does lead to a simple model that can be extended further through testing and parameter estimation.

Quick consideration of the errors that are likely to cause the actuator's behaviour to deviate from that of the model are:

(i) Orientation of strain gauge
(ii) Eccentricity of strain gauge (parallel to axis of sensing element)
(iii) Inaccuracy of the simply supported assumption for the plate
(iv) Inaccuracy of simple beam bending model for the sensing beams.

The first two factors will cause the strain gauges to be subject to strains due to shear stress as well as the orthogonal force (i.e. the gauges measuring vertical force may be affected shear force and vice versa). This will be known as the cross-coupling effect.

If the sensor plate is more 'built in' than simply supported, there will be moments at either end that are functions of both the force and the eccentricity of the centre of force. These moments will be transmitted to the sensing beams.

The ideal strain of the sensing beams was developed using the assumptions that they behaved as simple beams built in at one end, and loaded at a point at the other. The reality is more complicated. Both ends of the beam are effectively built in, preventing rotations. When the plate is translated downwards, a tensile stress caused by these built-in conditions accompanies the bending stress of the beams (when a beam is in simple bending, the neutral axis remains the same length, since this is prevented by the physical constraints, a stress results). Thus the strain at the gauged sites is due to a superposition of the effects of bending and tensile forces. Clearly these tensile forces do not change sense when the plate is translated upwards, and thus a translation dependent offset is introduced. The translation is of course a function of the magnitude and centre of the normal force.

The actual output of each of the channels should be approximately linearly related to each of the quantities involved in the ideal output, with the addition of cross-coupling effects (shear effects normal channels and vice versa). The constants relating the parameters to the output cannot be expected to be the physical characteristics of the force plate, but the complementary nature of the outputs should be preserved.

Given parameters that the output of the bridges should be proportional to, the problem of system identification, i.e. relating input to output, is reduced to that of finding coefficients that relate the parameters of interest to the output; parameter estimation. The simplest methods of parameter estimation are the least squares techniques, which aim to minimise the mean-squared error between the actual output and the output of the estimated model. The application of least squares parameter estimation to the problem will be discussed in the following section.

3.5.2 Least Squares Parameter Estimation

As discussed, it was assumed that the output voltage of each channel is linearly related to F (previously denoted F_N), Fx_c, Fy_c and S by a series of constants. In the ideal case, these constants would be simple expressions containing the physical dimensions of the sensor. Due to the potential inaccuracies previously discussed, these constants will be much more 'arbitrary'.

To estimate the actual constants involved in the expressions, least squares estimation was used. (Appendix A1 describes another method that was used). Consider a single channel, j. Its output voltage, for a constant input voltage, is approximated as follows:

$$V_j = F.C_{fj} + F.x_c.C_{xj} + F.y_c.C_{yj} + S.C_{sj} \qquad (3.12)$$

where F is the normal force, (x_c,y_c) are the coordinates of the equivalent normal point load, S, is the shear load, and $(C_{fj}, C_{xj}, C_{yj}, C_{sj})$ are coefficients relating the loading condtion to the output voltage of the j^{th} channel, V_j. If the output voltage of channel j is recorded for N different values of F, x_c, y_c and S, the equations can be written as follows:

$$V(1)_j = F(1).C_{fj} + F\,(1).x_c(1).\ C_{xj} + F(1).y_c(1).\ Cyj + S(1).\ C_{sj}$$

$$V\,(2)_j = F\,(2).\ C_{fj} + F\,(2).x_c(2).C_{xj} + F(2).y_c(2).\ Cyj + S(2).\ C_{sj}$$

$$\ldots$$

$$V\,(N)_j = F\,(N).\ C_{fj} + F\,(N).x_c(N).C_{xj} + F(N).y_c(N).Cyj + S(N).C_{sj}$$

$$\begin{bmatrix} V(1)_j \\ V(2)_j \\ \vdots \\ V(N)_j \end{bmatrix} = \begin{bmatrix} F(1) & F(1)x_c(1) & F(1)y_c(1) & S(1) \\ F(2) & F(2)x_c(2) & F(2)y_c(2) & S(2) \\ \vdots & \vdots & \vdots & \vdots \\ F(N) & F(N)x_c(N) & F(N)y_c(N) & S(N) \end{bmatrix} \begin{bmatrix} C_{fj} \\ C_{xj} \\ C_{yj} \\ C_{sj} \end{bmatrix} \qquad (3.13)$$

Equation (3.13) can be concisely written

$$\mathbf{V}_j = \mathbf{FC}_j. \qquad (3.14)$$

Least squares parameter estimation involves finding the parameters $\mathbf{C}_i$, such that the cost function,

$$J = (\mathbf{V}_i - \mathbf{F}\mathbf{C}_i)^{\mathrm{T}}(\mathbf{V}_i - \mathbf{F}\mathbf{C}_i) = \varepsilon^{\mathrm{T}}\varepsilon \tag{3.15}$$

is minimised, where ε is the 'error vector'. It can be shown (see Sensor Analysis and Improvement) that the least squares estimate of the constant vector $\mathbf{C}_i$ is given by

$$\mathbf{C}_i = (\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{V}_i. \tag{3.16}$$

Applying parameter estimation techniques to each of the channels then yields a set of constants that relates the voltage output to the loading conditions.

The method of calibration is to subject the sensor to a large range of known loading conditions, recording the data for each channel, and then applying least squares estimation to determine the best linear relationship between the variables and the output voltages.

3.6 Calibration

The requirement of calibration was that known loading conditions be created and the outputs recorded. A simple test-rig, shown in Fig. 3.12, was manufactured. During the application of normal force, the frame is horizontal, with each end supported on a desk. Known 'compressive' forces are applied by hanging masses below the sensor via a cross bar arrangement. The cross bar is balanced on a small circular plug that is placed on the sensor surface to localise the loading. Forces of the opposite sense are applied by inverting the entire rig, during which the sensor was both screwed and clamped to the rig frame. A 20mmx20mm grid was drawn on both sides of the sensor plate, with an origin placed at the centre of the plate (this is partly visible in Figure 3.8) to act as a guide for the load placement.

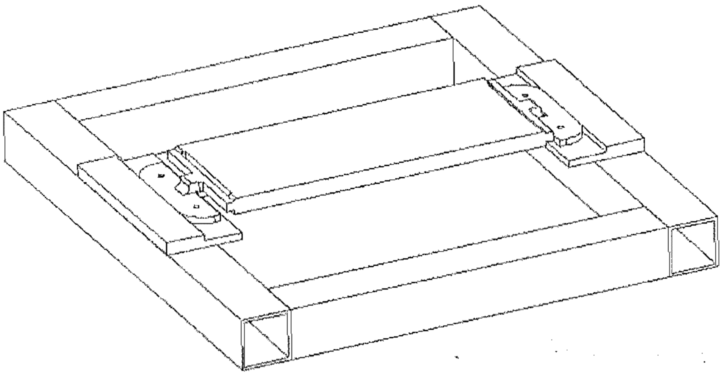


Figure 3.12 The foot force calibration test-rig

The initial rig (shown above) did not allow for simultaneous application of shear and normal forces. To apply shear, the rig was clamped upright (long axis of sensor vertical) and weights were applied to the sensor via rods that extended from plates that were screwed and clamped to the sensor. Later the rig was modified by suspending a bearing on a horizontal shaft, projected out from one of the ends of the rig. A cable that was attached by a hook to the sensor plate ran over the bearing and was used to suspend masses. The bending moment caused by the cable was minimised by attaching it to the hook at close to the level of the plate, and ensuring it was as horizontal as possible. This modification allowed the application of a limited amount of combined loadings (normal and shear). The amount of testing was limited because only compressive forces could be applied concurrently and also since the presence of the hook and cable limited the positions at which the mass causing the normal load could be placed.

The graphical results of a relatively coarse (25 points per mass) normal force calibration are shown in Figure 3.13. These show three normalised performance surfaces for each channel created by dividing the voltage yielded at each location by the applied force.

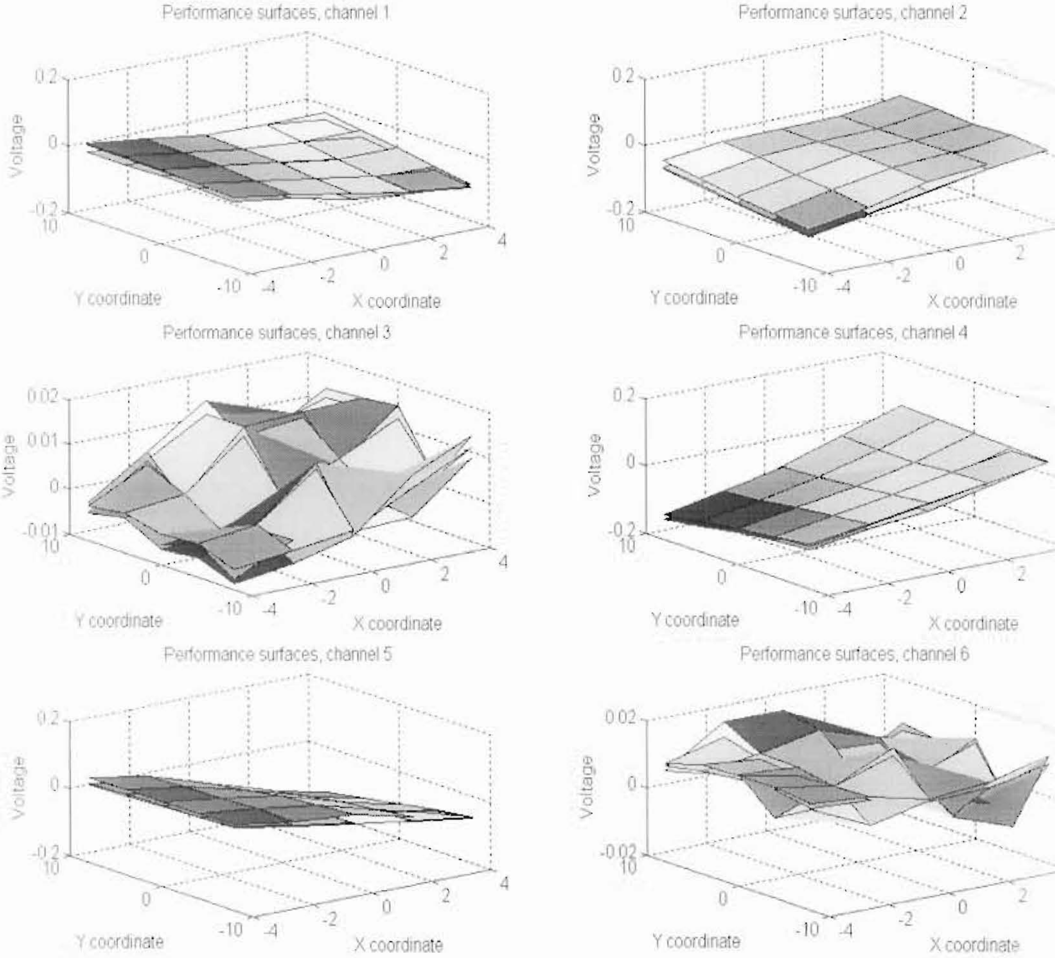


Figure 3.13 Normalised performance surfaces for response to compressive normal forces of 5, 15 and 25 kg.

In Fig. 3.13, in which a number of important phenomena are visible, channels 3 and 6 are the shear channels. Note the complementary nature of the surface gradients of the normal channels (1,2,4&5) that was predicted earlier. Other important phenomena are

- the good linearity and low variability with increasing mass of the normal channels
- the arbitrary appearance (but apparent repeatability) of the response of the shear channels to normal loading
- the shallow gradient of the plane that passes through the shear surfaces

Separate shear calibration results are also shown below in Fig 3.14. These results were the effective downfall of the sensor. It was hoped that the shear channels would offer much higher sensitivity to shear loading than the normal channels. Observation of the results shows that this was not the case.

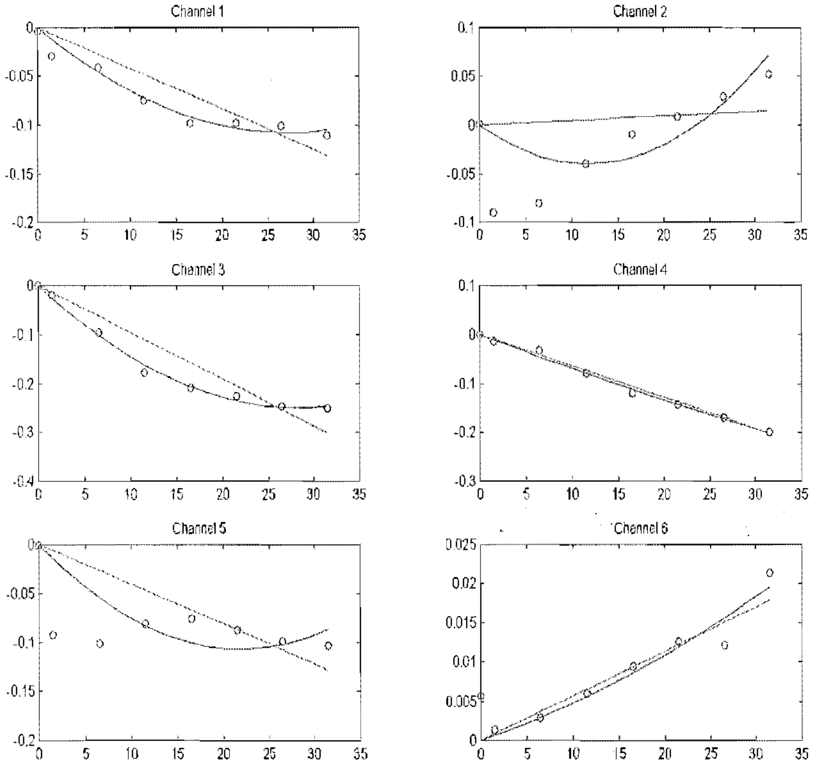


Figure 3.14 Results of a shear calibration with linear and quadratic lines of best fit forced to pass through zero at zero load.

In addition to the poor sensitivity of the 'shear channels', it was found that their response was dependent on the direction of application of the shear force, i.e. channel 3 may be linear when shear force acts towards the front of the sensor, but highly erratic with force in the other sense. A method was conceived to alleviate this problem that is detailed later.

The coefficients obtained by normal force application are shown below in Table 3.1. The columns represent the change in voltage associated with: (i) unit normal load at the centre of the plate (ii) unit normal load at unit distance in positive direction along X-axis (iii) unit normal load at unit distance in positive direction along Y-axis

C_f	C_{fx}	C_{fy}
-0.0357	**0.0154**	**0.0029**
-0.0409	**-0.0152**	**0.0027**
0.0047	0.0014	-0.0000
-0.0415	**-0.0195**	**-0.0028**
-0.0386	**0.0192**	**-0.0024**
0.0034	-0.0020	-0.0002

Table 3.1 Coefficients generated by normal force calibration.

The bold entries correspond to the 'normal force' channels (1,2,4,5); the two remaining are 'shear'. Ignoring the shear channels, note the complementary nature of the signs associated with each channel, i.e. channel 1 (-,+,+), channel 2 (-,-,+), channel 4 (-,-,-) and channel 5 (-,+,-). The effect of these patterns is discussed in Section 3.9. Another important feature of the coefficients is that the elements relating the response of the shear channels to normal loading are generally an order of magnitude smaller than the other entries. Physically, this means that normal loading has very little effect on the shear channels.

The coefficients in Table 3.2 resulted from combined loadings, applied as discussed previously. The fourth column represents the increase in each channel due to a unit shear load in the positive direction.

C_f	C_{fx}	C_{fy}	C_s
-0.0414	**0.0211**	**0.0029**	**-0.0203**
-0.0421	**-0.0211**	**0.0032**	**0.0250**
0.0026	0.0013	0.0001	-0.0047
-0.0443	**-0.0158**	**-0.0033**	**-0.0129**
-0.0397	**0.0156**	**-0.0026**	**0.0009**
0.0017	-0.0019	-0.0000	0.0010

Table 3.2 Coefficients generated by combined loading calibration.

Note that the magnitudes and signs associated with the first three columns remain almost unchanged and that channels 3&6, that were designed to respond to shear, have very low magnitude response.

3.7 Sensor Function

In possession of the six constant vectors from the calibration, the next task is the inverse of calibration, i.e. given a certain voltage vector, determining the (most likely) loading condition, i.e. combination of normal force, normal force coordinates and shear force.

This problem was originally approached using what appeared to be a novel method. The normal force and coordinate estimation problem was reduced to that of minimising the diagonal of a quadrilateral with sides that have fixed gradient and intercept inversely proportional to the applied force. After solving the problem and applying the results, it became obvious that this problem is a restatement of 'least squares' in a very particular application, which as previously shown has a very concise solution. Thus, rather than inserting the derivation of this 'novel' method of finding the solution to a group of approximate equations, it is relegated to the appendices (see Appendix A1). The fact that it is included at all is due to the geometric interpretation that the method gives to this problem. It is also conceivable that a similar method may be useful in problems where least squares estimation is not applicable.

At each instant the output of the sensors channels is $\mathbf{V} = [V_1\ V_2 \ldots V_6]$. Recall that each channel's voltage is approximately related to the loading condition by the relation

$$V_j \approx F.C_{\mathrm{f}j} + F.x_{\mathrm{c}}.C_{\mathrm{x}j} + F.y_{\mathrm{c}}.C_{\mathrm{y}j} + S.C_{\mathrm{s}j} \tag{3.17}$$

which, defining the loading condition vector $\mathbf{F} = [F\ \ F.x_{\mathrm{c}}\ \ F.y_{\mathrm{c}}\ \ S]$ and constant vector $\mathbf{C}_j = [C_{\mathrm{f}j}\ C_{\mathrm{x}j}\ C_{\mathrm{y}j}\ C_{\mathrm{s}}]^{\mathrm{T}}$, can be rewritten

$$V_j \approx \mathbf{F}\mathbf{C}_j = \mathbf{C}_j^{\mathrm{T}}\mathbf{F}^{\mathrm{T}} \tag{3.18}$$

for j=1,2,..,6. Using the measurement (voltage) vector, **V**, and constant vectors $\mathbf{C}_j$ j=1,2,..,6 we must solve for **F**. As a first step each of the scalar relations are stacked to form the matrix equation

$$\mathbf{V} \approx \mathbf{C}\mathbf{F}^T \tag{3.19}$$

where **V** and **F** are as previously defined and $\mathbf{C}=[\mathbf{C}_1{}^T\ \mathbf{C}_2{}^T\ \mathbf{C}_3{}^T\ \mathbf{C}_4{}^T\ \mathbf{C}_5{}^T\ \mathbf{C}_6{}^T]^T$, which is a 6x4 matrix, i.e. there are six equations in four unknowns. An alternative, and ultimately more useful notation is $\mathbf{C}=[\mathbf{C}_f\ \mathbf{C}_{fx}\ \mathbf{C}_{fy}\ \mathbf{C}_s]$, where the (6x1) column vectors of the matrix are as previously described. Each of these equations is approximate, thus we are trying to find the best solution, **F**, given **V**. The method of solving such an over-determined system of equations is to use the matrix pseudo-inverse, which is the least squares solution, thus

$$\mathbf{F}^T = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{V}. \tag{3.20}$$

The loading conditions are included within the vector, **F**, and are extracted by simple division, ($x_c \approx F(2)/F(1)$, $y_c \approx F(3)/F(1)$). These expressions become undefined when the applied force is calculated to be zero. Also the accuracy of the centre of force estimate becomes low as $F \rightarrow 0$.

Shown in Fig. 3.15 are the graphical results obtained when this method of determining the loading conditions is applied to calibration data. From top left, they show the accuracy in determination of load coordinates (in cm), accuracy of load determination, norm of coordinate error, approximated load against X – coordinate and coordinate error norm against X – coordinate. All plots are for a nominal compressive load of 20kg, but the hanging apparatus (chains, beams and hanger) add 3kg to this value. The characteristics of the sensor were plotted against X – coordinate to discern if there were points that were not as reliable as others. The results are pleasing. The norm of the coordinate error is less than 1cm in all cases, and the approximated load is within 0.5kg of the true value.

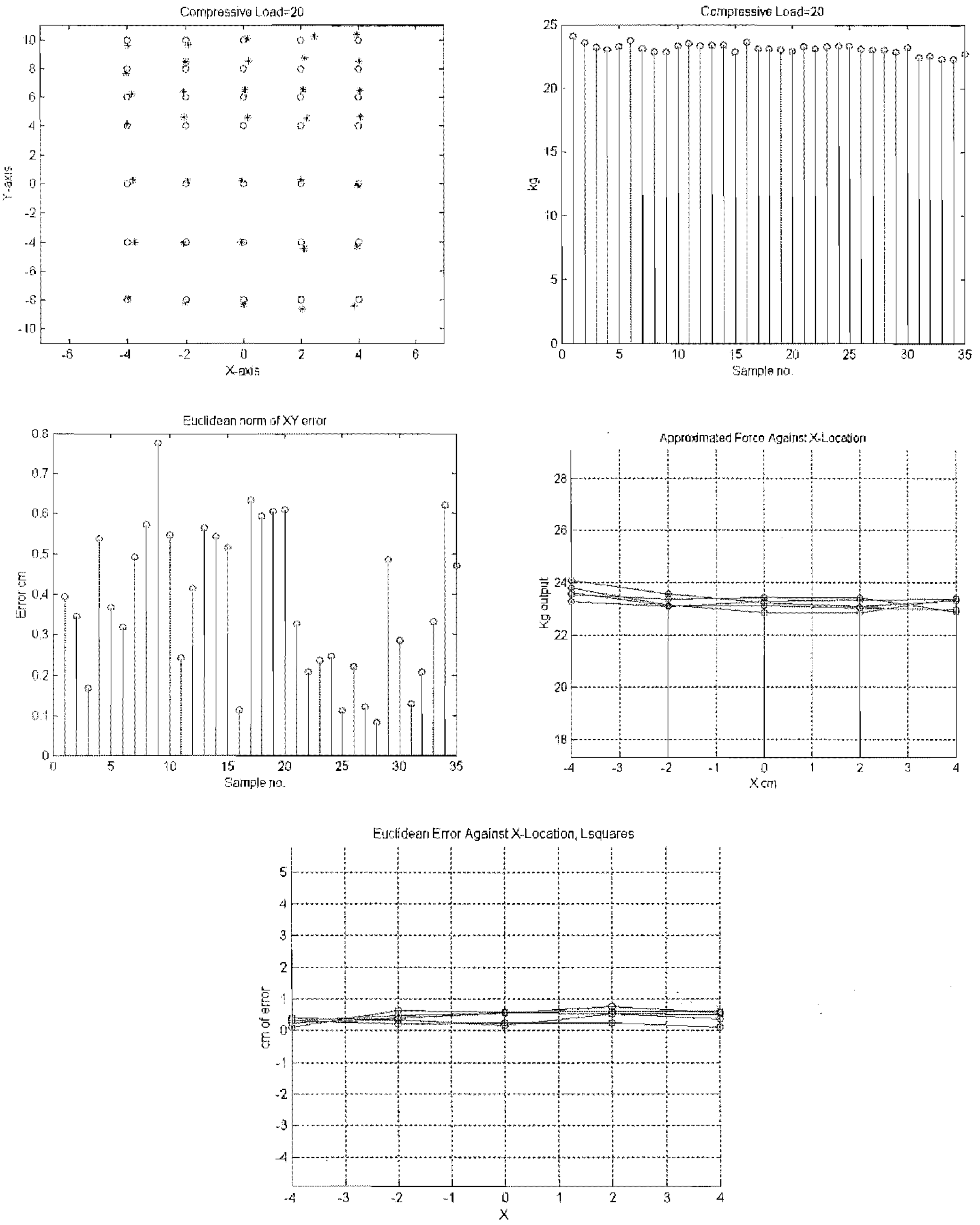


Figure 3.15 Characteristics of sensor for normal loading of 20kg.

As can be seen in the approximated load against X – coordinate plot, the sensor tends to slightly overestimate the load at the left edge of the plate. Methods to overcome this are described in later sections.

3.8 Results

Results included in this section are from ergometer trials of a number of the New Zealand 1999 rowing squad. Patria Hume, a biomechanist at the University of Auckland, was conducting a study into the effect of foot stretcher angle on ergometer rowing and required that foot force be measured [36]. In this study, the sampling frequency was 1kHz, principally because EMG data was also being recorded and this is the recommended rate. Each subject rowed for one minute, trying to maintain a constant rating (strokes/minute) and power for five different foot stretcher angles. The rowers were given a two-minute break between each minute of rowing while the angle was adjusted. A photograph of the experimental setup is shown in Figure 3.16.

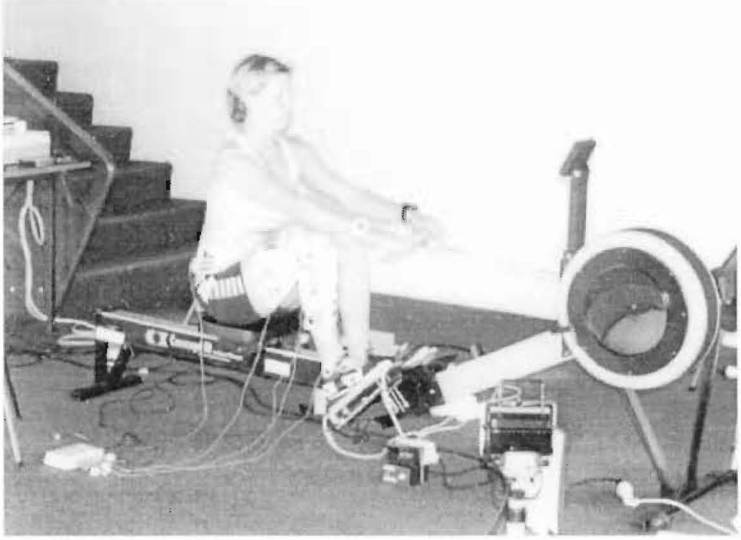


Figure 3.16 A national level rower using the foot force sensor on an ergometer

While it is impossible to quantify the accuracy of these results, since it is unknown to what extent shear is affecting the data, the normal force plots, showed what are intuitively believable results.

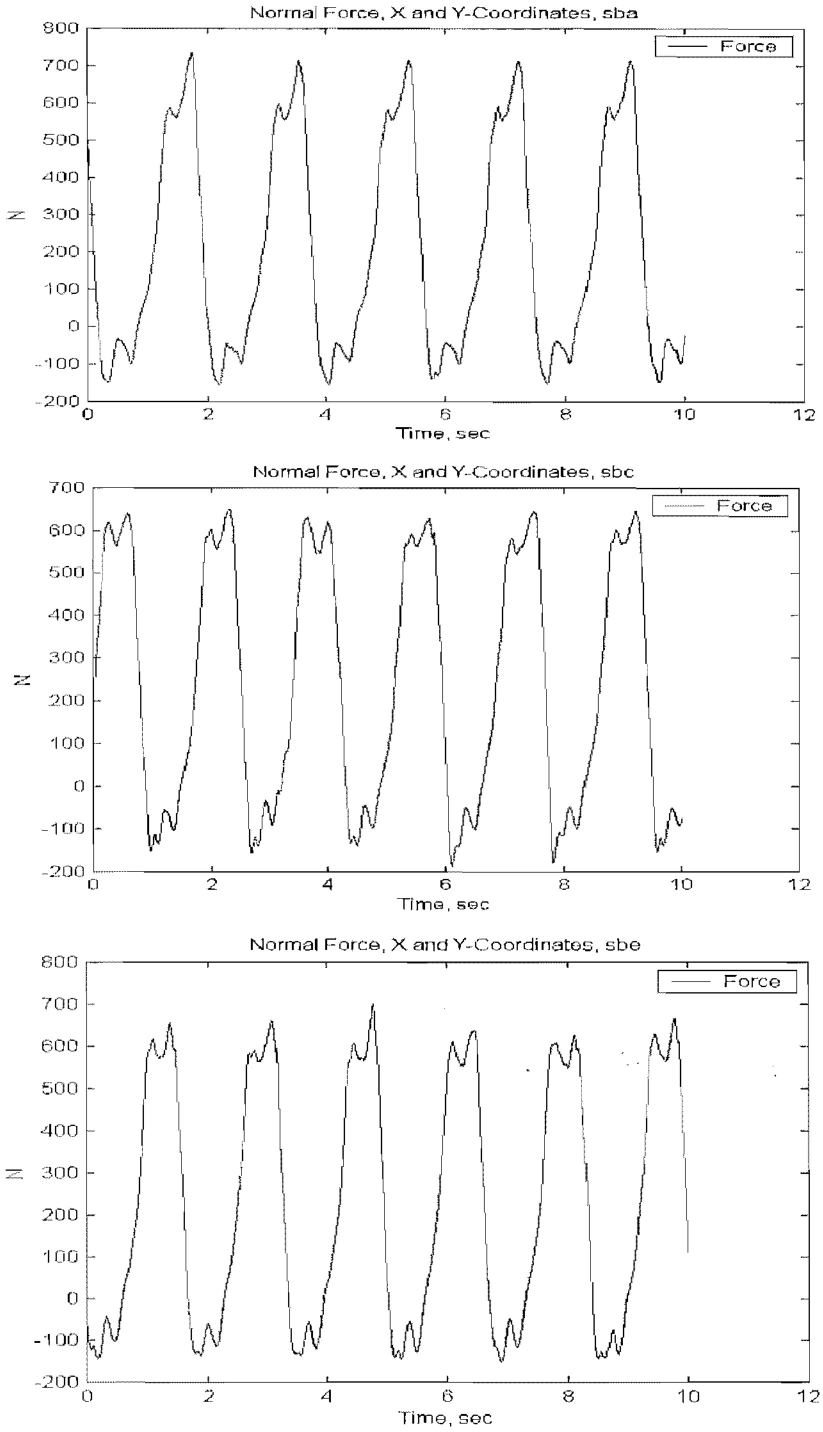


Figure 3.17 Normal force generated by subject 'sb' for three separate foot stretcher angles

It is evident from Figure 3.17 that changing the foot stretcher angle effects not only the peak force attained during ergometer rowing, but the profile of the force.

The X and Y coordinates of force were also computed. Interesting results were found by plotting the Y-coordinates and X-coordinates against the applied force.

While the X-coordinate of the centre of force was not seen to vary greatly, indicating that the normal force is applied towards the centreline of the force plate, the Y-coordinate varied in a way that is consistent with intuition (see Figure 3.18). Consider the cycle of foot force from the end of the drive onwards. At first the rower's foot is flat on the plate and the force is practically zero. As he pulls himself forward the force becomes negative and a moment about the X-axis is created that puts the centre of force significantly passed the area of the plate. (This situation is caused by a small net vertical force, F, and moment $F\,y_c$ such that $y_c > ½L_p$.) An interesting circling effect occurs as the rower begins to slow his progress and the force becomes positive (pushing on the plate). Now at the drive, the rower's heels are well off the plate. While there is no easy way to indicate elapsed time on plots, the 'comet' function on MATLAB showed that the centre of force quickly tended towards the centre of the plate as the compressive force increased suggesting that the full foot quickly comes into contact with the plate. The trace shows that the rower gets his feet flat to the plate and applies the largest magnitude force with a centre of load at around 60-70mm above the centre of the plate.

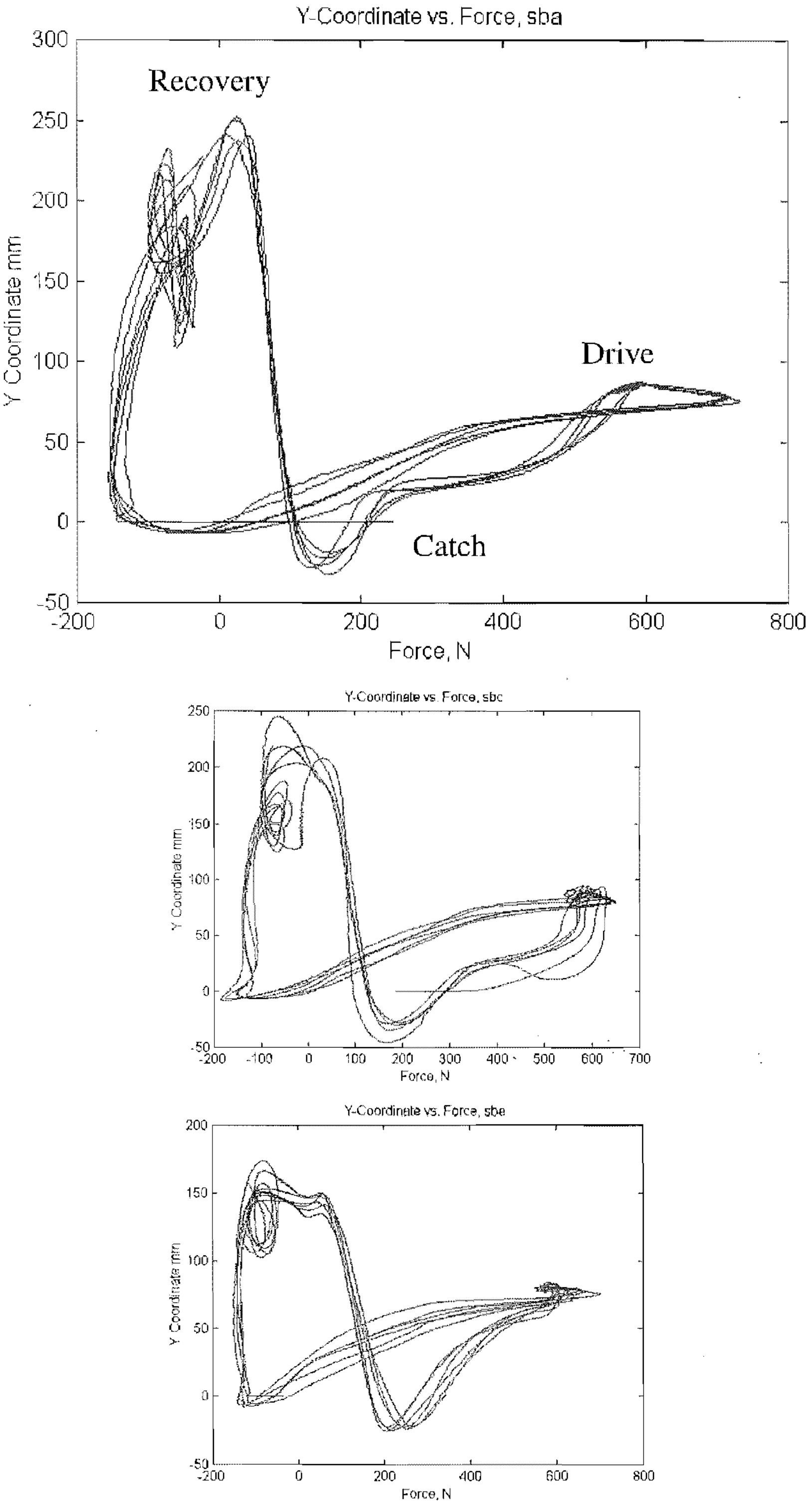


Figure 3.18 Y-Coordinate of centre of force vs. applied normal force for the subject 'sb' at three different values of foot stretcher angle.

Shown in Fig. 3.19 is a Y – coordinate vs. normal force plot in which the direction of time has been drawn in.

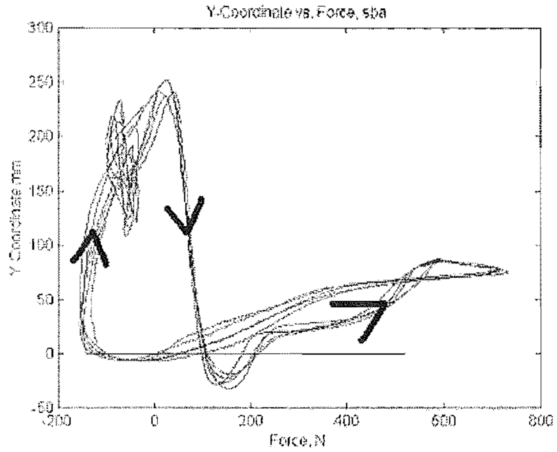


Figure 3.19 The general direction of time for Y-coordinate vs. Force plots.

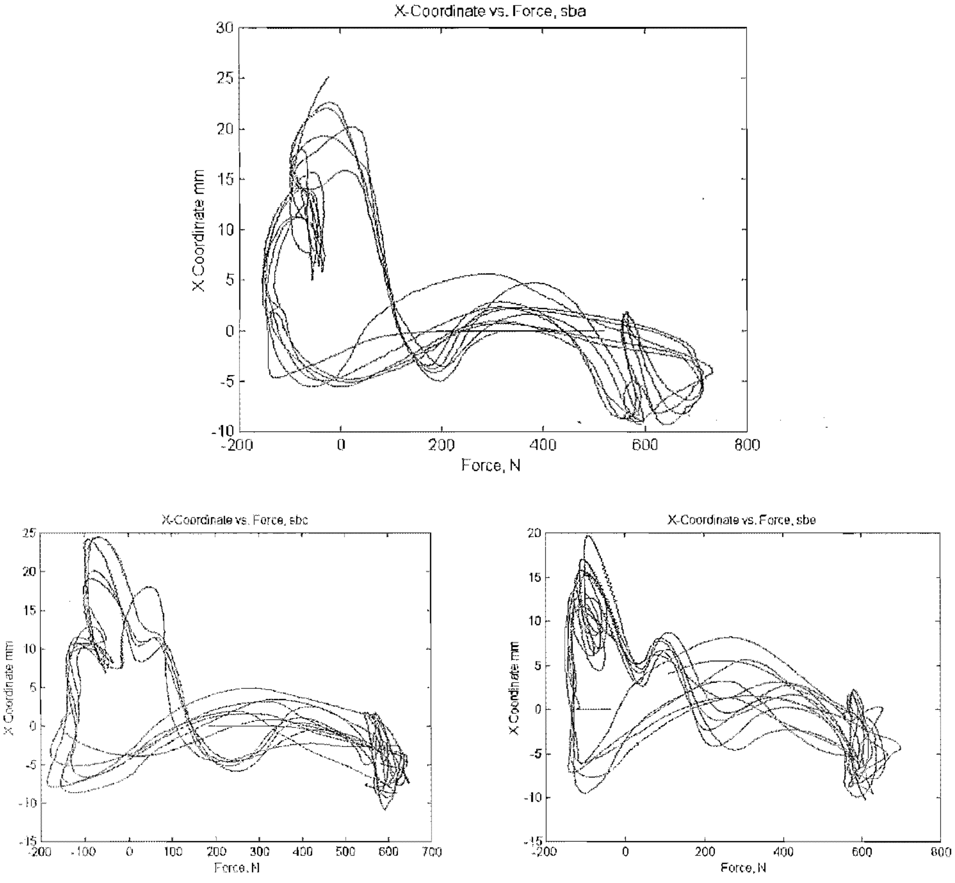


Figure 3.20 X-Coordinate of centre of force vs. applied normal force for the subject 'sb' at three different foot stretcher angle.

As previously mentioned, the X – coordinate of normal force was found to be almost constant during ergometer rowing. This is not to say that 'interesting shapes' were not generated, as shown in Fig. 3.18. Each rower tended to have signature profiles that varied with the changing foot stretcher angles.

While the normal force and coordinate data obtained using the sensor was found to be good, it was disappointing that the shear characteristics of the sensor were so poor. The next section explores the reasons for this poor performance.

3.9 Sensor Analysis & Improvement

Thus far the sensor design, calibration, method of operation and sample results have been discussed. It has also been stated that the sensor did not function exactly as had been hoped in that shear was not reliably measured. In this section the reasons for this poor performance are suggested and possible methods of improving functionality are discussed. Prior to this, a geometric visualisation of least squares estimation is developed that helps in the understanding of both how the sensor was intended to work, and how it failed. Unsurprisingly, the characteristics of the sensor are contained within the coefficients of the **C** matrix generated during calibration. The structure of these matrices is explored and it is shown how the low accuracy model of shear response can degrade the other estimates, regardless of the accuracy of the other column vectors.

3.9.1 Algebraic/Geometric Least Squares Derivation

Least squares estimation is very easy to understand if it is simply stated that it is the linear estimate that minimizes some quadratic function of the error, but this does not fully explain the geometric concepts involved. A simple 3D case will now be explained so that geometric concepts can be exploited. In this case, terminology is skewed towards our application: we have a measurement vector **V** that we wish to approximate using a linear combination of the column vectors of a coefficient matrix **C**. The coefficients of the vectors of **C**, f_1 and f_2, form the vector **F**

$$f_1\mathbf{C}_1 + f_2\mathbf{C}_2 = \mathbf{CF} \tag{3.21}$$

If (and only if) the vector **V** lies in the subspace spanned by the column vectors of **C** (the *column-space* of **C**), then it is possible to find values of f_1 and f_2 such that

$$\mathbf{CF} = \mathbf{V} \tag{3.22}$$

This will not normally be the case, i.e. **V** will usually lie outside the column space of **C**, this is the situation shown in Fig. 3.21.

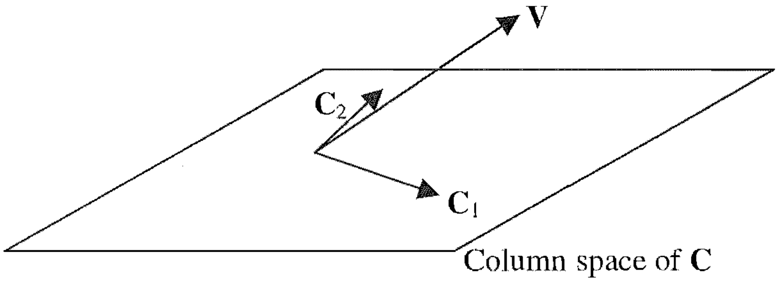


Figure 3.21 The geometry of the estimation problem. The column space of C consists of all linear combinations of C_1 and C_2, and the measurement vector, V, lies outside this space.

In this case the best that can be achieved is the estimation of **V** by a linear combination of $\mathbf{C}_1$ and $\mathbf{C}_2$ such that the length of the error (**V**-**CF**) is minimised. The length of the error vector is minimised when **CF** is that vector resulting from the orthogonal projection of **V** onto the column space of **C**.

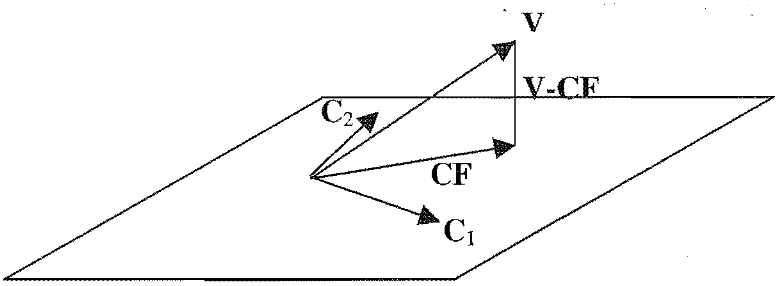


Figure 3.22 The least squares estimate, formed by the orthogonal projection of V onto the column space of C.

When **F** is chosen in this way, it is clear that **V**-**CF** is orthogonal to both $\mathbf{C}_1$ and $\mathbf{C}_2$, and in general is orthogonal to the entire column space of **C**. This may be written using the scalar product notation:

$$\mathbf{C}_1.(\mathbf{V}-\mathbf{CF}) = 0 \qquad \mathbf{C}_2.(\mathbf{V}-\mathbf{CF}) = 0 \qquad (3.23)$$

Since the scalar product **a.b** can also be expressed $\mathbf{a}^T\mathbf{b}$, it follows that $\mathbf{C}_i^T(\mathbf{V}-\mathbf{CF}) = 0$ and

$$\begin{bmatrix} \mathbf{C}_1^T \\ \mathbf{C}_2^T \end{bmatrix}(\mathbf{V}-\mathbf{CF})=\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\mathbf{C}^T(\mathbf{V}-\mathbf{CF}) = \mathbf{0} \qquad (3.24)$$

Manipulation of (3.24) leads to the familiar least squares solution

$$\mathbf{F} = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{V} \qquad (3.25)$$

In the actual problem at hand, the voltage vector is in R^6, and we are trying to estimate it by an optimal linear estimation of the four (6x1) coefficient vectors $\mathbf{C}_f$, $\mathbf{C}_{fx}$, $\mathbf{C}_{fy}$ & $\mathbf{C}_s$. The sub-space generated by these vectors is a hyperplane in R^6. The geometric interpretation offered by this derivation of the least squares technique helps in the understanding of what 'went wrong' with the sensor. The characteristics of the least squares estimation are obviously contained in the coefficient matrix, or equivalently the column space, that in this work, because of poor results, was constructed in a number of different ways.

3.9.2 Coefficient Matrix Generation

Driven by the quest for better results, three different methods were used to create **C** matrices. Each of these matrices, and the results that they yielded when used with purely normal and purely shear loadings are now explained. Following this is an attempt to explain why shear output was so poor.

The first matrix, $\mathbf{C}_{\text{normal}}$, was created via a least squares estimation using only normal data, i.e. tensile and compressive loadings at various locations on the sensor surface. When this matrix was used to estimate the loading condition (i.e. running the calibration data back through the derived coefficient matrix), both the magnitude of force and the coordinates of the centre of force were of good quality. (See for example Fig. 3.14). When the coefficient matrix was used with shear only data, the estimated force was close to zero, as it should have been, while the coordinates of loading, which were undefined, were large and arbitrary. The nature of the coordinate estimates in response to purely shear loading is easy to explain since the pair (x,y) is estimated by dividing the second and third elements (Fx and Fy) of the estimated load condition vector by the first element (F). When the first element is estimated to be small, as it is when no normal loading is applied, dividing by this element results in a magnification of the error in the estimate of Fx and Fy. While this explains that magnification takes place, it does not explain what is being magnified, i.e. why the estimate of Fy is greater than F (as must be the case for magnification to occur). A possible reason for this is a moment caused by uneven application of shear force during calibration. Another reason is suggested once required concepts have been introduced.

Before the rig was modified so that shear and normal forces could be applied simultaneously the method of creating a $\mathbf{C}$ matrix with both normal and shear 'capabilities' was to concatenate the results of a shear calibration to the normal force calibration. Thus the (6x1) vector $\mathbf{C}_{\text{shear}}$ resulting from a shear calibration was joined to the previously described matrix $\mathbf{C}_{\text{normal}}$ to form $\mathbf{C}_{\text{super}} = [\mathbf{C}_{\text{normal}}\ \mathbf{C}_{\text{shear}}]$, where the subscript 'super' is used to reinforce the fact that this matrix is built on the assumption of superposition. When $\mathbf{C}_{\text{super}}$ was applied to normal force data, it was found that the force was estimated with medium accuracy, the force coordinates were estimated very poorly and shear of considerable magnitude was often indicated where none was applied. In the second instance, where $\mathbf{C}_{\text{super}}$ was used on shear data, normal force was estimated well, i.e. it estimated a force very close to zero, and shear was close to being correct, but the estimated coordinates of force fluctuated violently, presumably for the same reasons as stated above. Recall that the shear response of the sensor was found to be directional, i.e. the response in one direction bore no resemblance to shear in the opposite sense. This could potentially be overcome using the assumption that the direction of shear force is dependent upon the sense of normal force. More precisely, the shear force during compressive loadings, such as the drive, will always be towards the toe, while shear during tensile forces, if it exists, will be in the other direction. This assumption can be used to dynamically update the **C** matrix by first calculating the sense of the normal force using $\mathbf{C}_{normal}$ then applying the stated assumption and including the relevant $\mathbf{C}_s$ vector.

The final matrix, $\mathbf{C}_{combined}$, was created using a least squares fit on data obtained during combined (shear and normal) loadings. This approach gave results that were poorer than the previous two methods in response to both shear and normal loading data, although, again, the normal force estimate was usually accurate.

These were the main methods used to generate coefficient matrices. Some alternatives were considered for increasing the accuracy in specific loadings. For example, if the centre of force was estimated to be in the front left corner of the sensor plate, a new **C** matrix generated by data only from that area of the plate could be used to refine the estimate. This potentially increases the accuracy because the performance surfaces of the plate are not exactly linear. When a small quadrant of the plate is used, a linear fit should be more accurate. The other way in which a more accurate solution could be found is to first estimate the force using a general **C** matrix and then use a matrix based only on data similar to the estimated force to improve the estimate. This method would allow for variation of the performance surfaces with respect to load. During calibration, no great fluctuation in the normalised performance surfaces was observed, but at higher levels of load, some deviations may occur.

3.9.3 Coefficient Matrix Properties

Now that the results of the methods used to generate the **C** matrices have been discussed, it remains to try and relate the results obtained to the matrices themselves. To do this we return to the geometric principles at the introduction of this section, and also explore the structure of the matrices employed.

Regardless of the method of generation of coefficient matrix, **C**, it was found that the estimate of normal force was generally good. The results from using $\mathbf{C}_{\text{normal}}$ on normal force data were of high quality, suggesting that the vectors of these matrices were accurate. The estimates of these vectors were also very stable with time, i.e. over a period of a few months of testing and various calibrations the numerical values of the elements of the vectors changed very little. This also adds weight to assumption that these three vectors are 'correct'.

In comparison, the response of the sensor to shear was found to be non-linear (and therefore had low accuracy with a linear fit), directional and time varying. This last component of uncertainty must surely be due to unintentional variation in loading condition. Disappointingly the channels that were included principally to monitor shear did not have good response to shear. This fact, and the directionality are illustrated by the results of shear calibration in two directions shown in Table 3.3. Also given is the $\mathbf{C}_s$ vector of $\mathbf{C}_{\text{combined}}$. Note the large discrepancy between the numerical values.

	1	2	3	4	5	6
Positive	-0.0038	-0.0120	**-0.0003**	0.0137	0.0020	**-0.0098**
Negative	-0.0006	0.0053	**0.0038**	-0.0083	0.0000	**-0.0001**
Combined	-0.0203	0.0250	**-0.0047**	-0.0129	0.0009	**0.0010**

Table 3.3 Generated shear coefficients for positive, negative and combined calibrations. The bold columns relate to the 'shear' channels.

Regardless of what caused the sensor's poor shear response (some suggestions are given below), the translation of the effects of shear to a linear representation was an inaccurate process. The low accuracy of the $\mathbf{C}_s$ vector translates to a low accuracy of shear estimation, but does not directly explain why coordinate estimation and normal force estimation were adversely affected to varying degrees. The phenomenon of estimate degradation due to the inaccuracy of the shear coefficient vector is now explored.

Recall equation (3.24) expressing the orthogonality of the column space of **C** to the orthogonal projection of **V** onto it encountered during the least squares derivation:

$$\mathbf{C}^T(\mathbf{V}-\mathbf{CF}) = \mathbf{0}$$

or

$$\mathbf{C}^T\mathbf{CF} = \mathbf{C}^T\mathbf{V} \tag{3.26}$$

Using the scalar product, this (3.26) can be rewritten

$$\begin{bmatrix} \mathbf{C}_f.\mathbf{C}_f & \mathbf{C}_f.\mathbf{C}_{fx} & \mathbf{C}_f.\mathbf{C}_{fy} & \mathbf{C}_f.\mathbf{C}_s \\ \mathbf{C}_f.\mathbf{C}_{fx} & \mathbf{C}_{fx}.\mathbf{C}_{fx} & \mathbf{C}_{fx}.\mathbf{C}_{fy} & \mathbf{C}_{fx}.\mathbf{C}_s \\ \mathbf{C}_f.\mathbf{C}_{fy} & \mathbf{C}_{fx}.\mathbf{C}_{fy} & \mathbf{C}_{fy}.\mathbf{C}_{fy} & \mathbf{C}_{fy}.\mathbf{C}_s \\ \mathbf{C}_f.\mathbf{C}_s & \mathbf{C}_{fx}.\mathbf{C}_s & \mathbf{C}_{fy}.\mathbf{C}_s & \mathbf{C}_s.\mathbf{C}_s \end{bmatrix} \begin{bmatrix} F \\ Fx \\ Fy \\ S \end{bmatrix} = \begin{bmatrix} \mathbf{C}_f.\mathbf{V} \\ \mathbf{C}_{fx}.\mathbf{V} \\ \mathbf{C}_{fy}.\mathbf{V} \\ \mathbf{C}_s.\mathbf{V} \end{bmatrix} \tag{3.27}$$

which explicitly shows the way in which the vector $\mathbf{C}_s$ is manifested in the least squares solution. It is clear that the degree to which the inaccuracy of $\mathbf{C}_s$ affects the estimates of F, Fx and Fy is contained in the scalar products $\mathbf{C}_f.\mathbf{C}_s$, $\mathbf{C}_{fx}.\mathbf{C}_s$ and $\mathbf{C}_{fy}.\mathbf{C}_s$ respectively. If the vectors are orthogonal then the error in $\mathbf{C}_s$ does not contaminate the estimation of the other parameters. In general, also, the smaller the included angle between $\mathbf{C}_s$ and another coefficient vector, the greater effect $\mathbf{C}_s$ has on the associated parameter estimate. An investigation of the orthogonality of the columns of **C** for the formulations described above is shown in tabular form below, where all values appear twice for clarity. The values are the included angle between the two indicated vectors divided by $\pi/2$, i.e. orthogonal vectors will have a value equal to unity. Significant deviations from unity indicate that the involved column vectors are not orthogonal, and hence the accuracy of the parameters with which these vectors are associated is affected by the accuracy of the second vector.

	$\mathbf{C}_f$	$\mathbf{C}_{fx}$	$\mathbf{C}_{fy}$	$\mathbf{C}_s$
$\mathbf{C}_f$	-	1.0196	0.9942	1.0738
$\mathbf{C}_{fx}$	1.0196	-	1.0160	0.7208
$\mathbf{C}_{fy}$	0.9942	1.0160	-	1.2335
$\mathbf{C}_s$	1.0738	0.7208	1.2335	-

Table 3.4 Included angle between indicated vectors divided by $\pi/2$ for $\mathbf{C}_{\text{combined}}$

	$\mathbf{C}_f$	$\mathbf{C}_{fx}$	$\mathbf{C}_{fy}$	$\mathbf{C}_s$
$\mathbf{C}_f$	-	1.0334	0.9944	0.9789
$\mathbf{C}_{fx}$	1.0334	-	1.0460	0.9320
$\mathbf{C}_{fy}$	0.9944	1.0460	-	0.6511
$\mathbf{C}_s$	0.9789	0.9320	0.6511	-

Table 3.5 Included angle between indicated vectors divided by $\pi/2$ for $\mathbf{C}_{\text{super}}$ (shear positive)

Note that:

- the vectors $\mathbf{C}_f$, $\mathbf{C}_{fx}$ and $\mathbf{C}_{fy}$ are nearly mutually orthogonal in both cases
- $\mathbf{C}_f$ and $\mathbf{C}_s$ are almost orthogonal in both cases
- $\mathbf{C}_s$ and $\mathbf{C}_{fy}$ are significantly removed from orthogonality in both cases
- $\mathbf{C}_s$ and $\mathbf{C}_{fx}$ are significantly removed from orthogonality in the combined loading case

These results show numerically why the estimates of coordinate are greatly degraded by including a shear vector shear in the **C** matrix. Geometrically, these comments are easily translated. Each of the column vectors in **C** represents a direction of increase due to a particular loading. For example the vector $\mathbf{C}_f$ is oriented such that it points in the direction of increased normal load, F, i.e. if an increasing normal load was applied at (0,0) the direction of increase in R^6 would be $\mathbf{C}_f$, and the magnitude of the load would be given by the ratio of the magnitudes $\|\mathbf{V}\|/\|\mathbf{C}_f\|$. The triple $\mathbf{C}_f$, $\mathbf{C}_{fx}$ and $\mathbf{C}_{fy}$ are three near orthogonal vectors in R^6. This orthogonality was predicted early in the chapter, where the complementary nature of the signs of the responses to a normal loading at arbitrary coordinates was observed. The vector $\mathbf{C}_s$, while almost orthogonal to $\mathbf{C}_f$ is oriented so that it has non-negligible components in the direction of $\mathbf{C}_{fx}$ and $\mathbf{C}_{fy}$. Recall the least squares estimate is obtained through the orthogonal projection of the measurement vector on the column space of the coefficient matrix, i.e. the subspace defined by linear combinations of $\mathbf{C}_f$, $\mathbf{C}_{fx}$, $\mathbf{C}_{fy}$ and $\mathbf{C}_s$. Since $\mathbf{C}_s$ has components in the same direction as $\mathbf{C}_{fx}$ and $\mathbf{C}_{fy}$, when $\mathbf{V}$ is projected onto the subspace there is an ambiguity that is heightened by the poor accuracy of $\mathbf{C}_s$ (i.e. the presence of un-modelled but deterministic components in $\mathbf{V}$) and the presence of noise.

In summary, the response of the sensor to shear, as it has been loaded proved to be such that it could not be well modelled linearly. Further, the linear model that was fitted to the response was not in the anticipated 'direction' in that the elements of the vector that were expected to be large were not. If the vector relating the response to shear had been in error and also, by some chance, orthogonal to all other vectors, its inaccuracy would not affect the other estimates. As it is, the erroneous $\mathbf{C}_s$ vector 'soaks' up some of the projection that would otherwise be distributed between $\mathbf{C}_{fx}$, $\mathbf{C}_{fy}$ and estimation error, i.e. the component orthogonal to the column space of $\mathbf{C}$. It is possible that the true response of the sensor to shear is not orthogonal to the other vectors. If this is the case and the vectors are all well modelled then there will be little error.

As the sensor is, it yields good results in response to purely normal loadings in that the load and coordinates of load are estimated to a reasonable level of accuracy. Since the shear response was not modelled well it is impossible to accurately state to what extent the presence of shear effects normal load condition estimation during combined loads. The results obtained from the ergometer trials, however, yielded results that were consistent with intuition. If it is desired that the sensor's combined loading characteristics be accurately quantified, some modifications will need to be made.

3.9.4 Sensor Improvement

Clearly, to improve the sensor's performance, its shear response must be more accurately determined. Before discussing methods by which the shear response could be improved some general comments are made on the **C** matrix and the effects that good shear modelling would have.

3.9.4.1 The 'Ideal' C Matrix

If the sensor had behaved as planned, the third and sixth elements of the calculated shear vector would have been large, with all other elements (corresponding to the response of the normal channels to shear loading) small. This would increase the orthogonality of the shear vector with respect to three other coefficient vectors, since none of the other channels have significant components in these directions. In addition, if the sensor had functioned as was desired, there would not be such a large error in $\mathbf{C}_s$ and hence, the small amount of error propagation would not be problematic.

In an ideal situation a square (6x6) **C** matrix would be used, since this would allow for 'perfect' estimation of the loading conditions, i.e. the measurement vector, **V**, would always lie in the column space of **C**. One cannot, however, simply add two arbitrary columns to the coefficient matrix. The columns, for the inverse of $\mathbf{C}^T\mathbf{C}$ to exist must be linearly independent. In physical terms this means that additional parameters associated with the new columns must be selected so that the vector response of the sensor to parameter variation is distinct from existing column vectors. In particular the vector must be non-zero. Even if a full square matrix could be generated, the estimation error would still only be 'theoretically' zero. The word theoretically is used because of potential errors in both the measurements vector and the coefficients matrix. If the number of unknowns (corresponding to the number of columns) is anything less than the dimension of the measurements vector then there is, additional to errors in **C** or **V**, an estimation error. This estimation error is not purely a function of the discrepancy between the dimension of the column space and the dimension of the vectors to be minimized, but also of the form of the coefficient matrix and more particularly the accuracy of each of the column vectors and their relative orthogonality.

3.9.4.2 Improving Shear Response

Inherent in the discussion of improving shear response is identification of possible causes of the originally poor characteristic.

It is possible that a portion of the error in $\mathbf{C}_s$ was due to the way in which shear was applied. Both methods had the potential for a simultaneous application of bending moment that would alter the outputs. The bad repeatability and directionality could also be a function of the sensor constraints, i.e. the way in which the sensor is fastened to the calibration rig. To truly find the sensor's response to shear loading the transducer could be mounted on a reliable load cell set-up, or a force plate. The latter suggestion would allow for very accurate dynamic calibration of the sensor.

The small magnitude of the shear response could be improved, thereby increasing signal to noise ratio, by making the sensing beams thinner. This would not drastically alter the response of the other four channels due to the definition of the second moment of area for bending in each direction. There is enough room, considering strain gauge placement, for easy removal of almost 2mm from the width of the beams. Higher gauge factor strain gauges, such as polymer varieties, could also be used.

Another possible minor modification to the sensor would be an increase of the width of the slots that define the sensing beams so that it would be possible to fix strain gauges to both faces of the elements. It is thought that a possible contribution to the poor shear response was the placement of strain gauges on adjacent beams. Reasons for this are: the possibly asymmetrical shear distribution and location of the gauges and different distances along the beams.

If the shear response of the sensor cannot be improved by the discussed methods, then it is possible that a compound sensor could be designed. Such a unit could be designed by mounting the existing sensor, or a subtle variation, on the magnetoresistive shear sensors discussed in the foot force measurement review section. The method of function would remain unchanged; the output of the shear sensors would simply be substituted for the outputs of the shear sensing channels. There would be the option of increasing the dimension of the space, for example using a shear sensor in each corner would result in eight channels in total, or a single sensor could be placed at each end, maintaining the original dimension. If the shear sensors can be designed so that they have either negligible, or well modelled, response to normal loadings, as well as good shear response, then the compound sensor would be likely to yield very good results.

While the preceding section has briefly explained ways in which the shear response of the sensor can be improved, there is also room for improvement in the normal force estimation. Investigation of the deformation caused by even the simplest case of central normal loads shows that the ideal situation considered at the beginning of the chapter was a great oversimplification. This is discussed briefly in Chapter 6.

3.10 Conclusions

The design, calibration and function of a foot force sensor that is theoretically capable of both normal and shear force measurement has been described. The structure of the sensor was enforced through the sensing requirements, spatial constraints and desired generality.

While the sensor had good characteristics in response to normal loadings the shear facility was very poor.

Least squares estimation is used in both the calibration and function of the sensor, and an investigation of error propagation due to this method has been made.

Suggestions for improvement of the sensor have been made. Modifications are generally concerned with the increase in accuracy of the sensor's shear response. It is proposed that a sensor hybrid is perhaps the most attractive option.

Chapter 4

While the rower pulls his oar through an arc during the rowing stroke, only the force in the direction of motion of the boat has impact upon the motion of the boat. There are therefore portions of the stroke where it is more efficient to create a greater force. Thus, to investigate the efficiency of the stroke, it is necessary to measure the rotation of the oar. Other kinematic characteristics of the stroke that should be measured are the size of the arc subtended during the stroke and the sequencing of the oar movement with the motion of the seat and the force at the foot-stretcher.

All the aforementioned aspects of the stroke require only the measurement of the angle between the loom of the oar and a normal projected from the boat 'parallel with the plane of the water'. Rather than being a simple hinge, allowing only one degree of freedom, the oarlock permits full rotation; only translation of the oar is prevented. During the rowing stroke, in addition to the previously described angle, the oar is rotated so that it rises out of the water during the recovery and is submerged into the water for the catch. It is also rotated about its own axis during feathering. Just like any general three-dimensional motion, to fully describe the rowing stroke requires three angles. These three angles can be quantified as the yaw (sweep angle), pitch (tilt of loom with respect to the horizontal) and roll (rotation of the oar about its own longitudinal axis). Many insights into rowing technique could be gathered from investigation of the rotation of the oar.

This chapter gives a brief history of oar angle measurement to measure the sweep angle (yaw). Following this, a new sensor combination capable measuring the three oar angles is developed theoretically. These sensors were manufactured, but results were poor, for reasons that are elaborated upon in later sections of this chapter. Methods by which these problems could be overcome are fully detailed.

4.1 Oar Angle Measurement Review

Atkinson's 'Rowing Indicator' [5], [6] described in Chapter 3, was capable of measuring both oar force and sweep angle. It in fact produced, as output, a chart plotting force against rowing angle. In Atkinson's original version [5], a pencil rotating with the gate drew onto specially configured charts. His later method was much more mechanically sophisticated.

With the exception of Atkinson, all researches who have measured oar angle have used a rotary potentiometer in one form or another [27], [39], [64]. The advantages of the potentiometer in this application are that they are compact, relatively cheap, require only very basic signal processing and can be obtained in splash-proof configurations and non-contact varieties. The shortcomings of the potentiometer are that it measures only one angle, and that it must of course be fixed somehow so that it measures the rotation of the oarlock with respect to the rigger.

Candidate sites for oar angle measuring potentiometers are limited and include: the oarlock cavity (using a shaft type potentiometer), with the potentiometer fixed inside the cavity, with the hole pressed onto the pin; above the oarlock (probably necessitating an elongated pin. An alternative to these sites where the potentiometer is directly driven by the rotation of the oarlock is to situate the sensor on the rigger, distinct from the oarlock and drive it using some sort of belt. Such a method, the 'rubber band goniometer' was mentioned by Gerber [27].

4.2 Sensors Applied to Measure 3D Oar Rotation

Since it was already decided that an attempt would be made to measure all three angles of the oar, none of the previously used methods of measuring oar rotation were suitable, and a new method was devised. Later in this section, it is described how the output of a sensor such as a rotary potentiometer could be used in conjunction with a general 3D 'orientation sensor' to resolve certain ambiguities.

The new method, as developed in the following sections measures the rotation of the oar by finding the relative rotation matrix that transforms measurements of the earth's magnetic field and a general acceleration field, from an orthogonal set of sensors mounted on the oar to a corresponding set fixed to the boat. The sensors employed are magnetoresistive (Honeywell HMC1021,1022) and accelerometers (Analog Devices AD202).

Magnetoresistive (MR) sensors are a relatively new technology, so a brief description of their construction and method of operation is now given. Each axis of an MR sensor consists of a Wheatstone bridge made up of NiFe thin film deposits on a silicon substrate. These sensing elements are oriented opposite to each other in pairs. When the magnetoresistive elements are subjected to a perpendicular magnetic field, the magnetisation vector within the elements is rotated. The resistivity of the elements depends on the angle between the electric current in the element and the magnetisation, thus each axis outputs a voltage proportional to the strength of the field perpendicular to the axis. General characteristics for the HMC1021/1022 microcircuits are shown below.

Field Range	+/-6 Gauss
Field Resolution	85u Gauss
Bandwidth	Over 5 MHZ
Sensitivity	1.0 mV/V/Gauss
Linearity	+/-0.5-1% full scale

Table 4.1 HMC 1021/1022 MR Sensor Characteristics

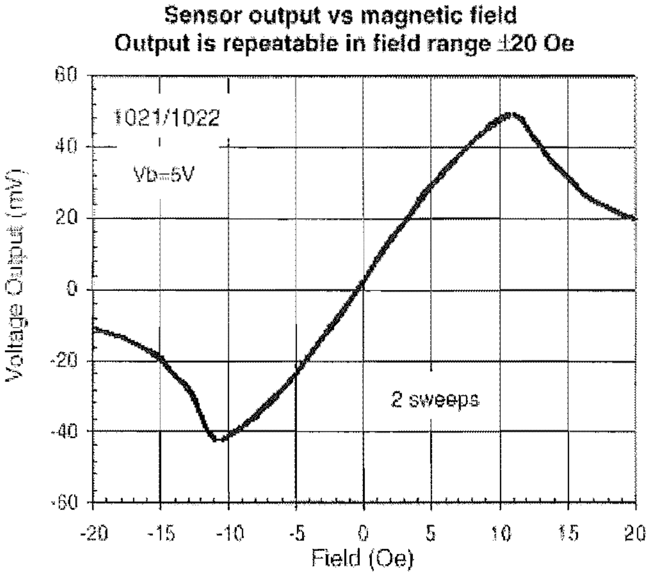


Figure 4.1 HMC 1021/1022 typical sensor output

The earth's magnetic field is well within the linear range of the sensor, at approximately 1 Oe or 1 Gauss, as shown in Figure 4.1. From this point onwards, each axis of an MR sensor is regarded purely mathematically as an axis that maps the incident magnetic field to a scalar output via a linear transformation. This is more fully described in the following sections.

4.3 Sensor Design and Construction

Sensor clusters consisting of triaxial MR sensors and accelerometers mounted inside a common enclosure were used in this work. A brief description of these sensors is given below.

4.3.1 MR Sensors

3-Axis sensors consisting of a HMC1022 two-axis MR microcircuit and a HMC1021Z single-axis MR microcircuit were constructed so that a 3D representation of the earth's magnetic field could be measured. The orientations of the axes within the microcircuits and the 'pin outs' are shown in Figure 4.2.

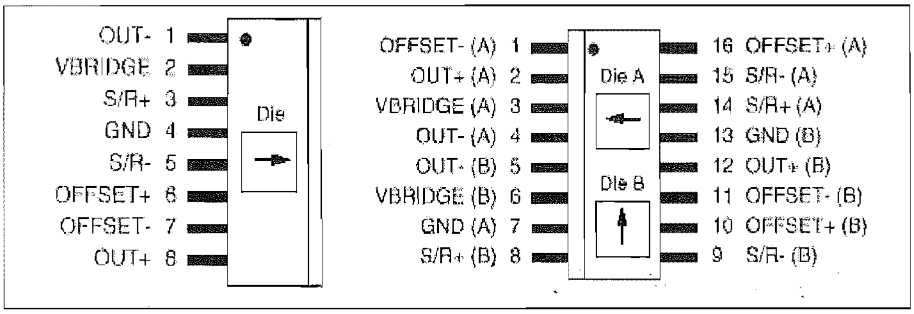


Figure 4.2 Pin-Out Diagrams for HMC1021Z (left) and HMC1022 [34]

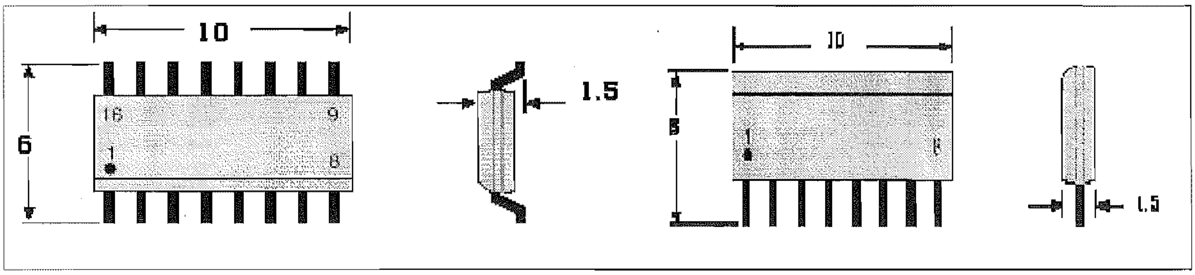


Figure 4.3 Approximate Dimensions of HMC1022 (left) and HMC1021Z [34]

The two-axis sensor was mounted flat to the board, while the HMC1021Z was mounted with its axis orthogonal to the plane of the board. This gave a set of 3 'orthogonal' axes. Orthogonal appears in quotes because all the microcircuits were mounted 'by eye'. This had consequences that are fully discussed in later sections.

The circuitry on the board (designed by Electronics Technician Julian Philips) was very simple, basically consisting of bridge supply and output amplification. The pins in the above diagram marked S/R are 'Set/Reset Straps' and can be used to reset the circuit if saturation occurs. (Saturation due to large incident fields can reverse the polarity of the sensing elements.)

The triaxial sensor is shown below in Fig 4.4. The external dimensions of the board was chosen to give a tight fit in a photographic 35mm film canister, as this was at the time considered to be a cheap method of housing the sensors. Problems with sensor orientation, due to the curved surface of the canisters, lead to a new housing design.

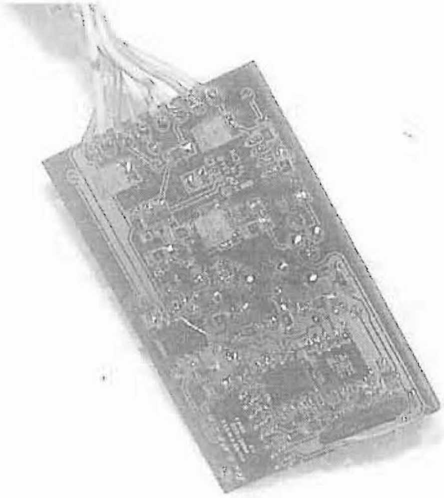


Figure 4.4 Triaxial MR sensor. The single axis sensor is seen projecting from the lower right corner. The dual axis sensor is the micro-circuit in the lower left corner.

4.3.2 Accelerometers

Dual-axis solid-state accelerometers manufactured by Analog Devices were used to create a three-axis accelerometer. Two boards joined along a common edge had one dual axis accelerometer each (the two axes are in the plane of the microcircuit). Again, the microcircuits and the boards were mounted without a special jig.

The boards on which the sensors were mounted were then glued into a plastic enclosure, without any jig to ensure that they were square with respect to edges of the enclosure. As mentioned previously, these approximate methods of mounting lead to problems. While it would surely have been sensible to create jigs to mount the sensors properly, there were a number of factors that prevented this. Among these are the cost of producing a jig of useful accuracy, and the time this would take to produce, offset against the desire to quickly create a new sensor methodology. Additional to these excuses, there will always be some degree of error in the mounting of the sensor; even if the microcircuits are mounted perfectly, there is no guarantee that the axes within the sensor are 'true', thus it is actually more useful to combat these errors in an ad hoc sense, i.e. finding the error and then compensating rather trying to eliminate the error during manufacturing. Thus, by being initially sloppy, and paying the consequences, a much more useful technique was developed!

4.4 Introduction to Measurement of 3D Rotations

The following sections develop the theory required to use the output of the previously described sensors to measure relative orientation during general 3D motions. The results from the field of theoretical kinematics are taken from various texts [11], [29], [47] & [48] and have been combined so that only the useful material is present. The subject matter in its pure sense is far removed from rowing, but in practicality, the described technique will be useful in an almost endless array of areas where motion, human generated in particular, needs to be measured. The aim of this work is to use vector observations of the earth's magnetic field and a 'general acceleration field' to discern the relative orientations of two objects from which the observations were made.

Before the problem of attitude estimation is discussed, an amount of theory needs to be laid down. In particular, theoretical spherical kinematics, discussed in Section 4.5, has many essential, and beautiful results. Section 4.6 provides a geometric bridge between spherical kinematics and attitude estimation, showing the minimum information requirements to uniquely discern attitude.

Attitude determination, that is, estimating the relative orientation of two objects using vector observations, in this case the orientation of the oar with respect to the boat, has applications in many fields, and thus many solutions to the problem have been proposed. A review of existing methods, including the derivations of the algorithms, of attitude determination is included in Appendix A2. In Section 4.7, a simple new method of attitude determination that simulations have shown to out-perform all-comers in our application is derived.

Having sufficient theory to solve the problem, the next section of the chapter, Section 4.8, is concerned with the actual application of accelerometers and magnetoresistive sensors in the role of generating the required vector observations, still at a theoretical level. Section 4.9 gives some more practical details concerning the use of the sensors with the new attitude estimation algorithm.

Next, in Section 4.10, the problems that have been alluded to in the previous sections, concerned with the non-orthogonality of the sensor axes, are discussed. While these problems cannot be entirely eliminated physically, a new method of calibration is proposed, which 'orthogonalises' the axes of the sensors.

The output of the attitude estimation is not in a form that is immediately useful to rowers or coaches, thus some simple processing is needed to transform the output to physically meaningful angles. This is the subject of Section 4.11.

4.5 Spherical Kinematics

Spherical kinematics is concerned with rotations of bodies in three-dimensional space. The displacements of spherical kinematics have the property that one point (normally taken as the origin) remains stationary, and thus the paths of all points on the body as it rotates are constrained to lie on concentric spheres, with the stationary point as their centre.

In the case of the oar, the problem for which this system was designed, the oar is constrained from translation at the oarlock, while all other motions are possible. It is possible to place one non-rotating frame at the oarlock and one on the oar such that the origins of the two frames are always coincident, thus the motion of the oar with respect to the boat is one consistent with the requirements of spherical kinematics.

The more general multi-body case can also be considered as one of spherical kinematics as long as the body does not involve sliding joints, e.g. the upper arm rotates relative to the shoulder, while the lower arm rotates relative to the elbow (and hence the upper arm).

4.5.1 Rotation Matrices

Say the vector, $\boldsymbol{r}$, represents a field that is being 'measured' in two bases, M and F. The vector $\boldsymbol{r}$ is constant, while the column matrices, $\mathbf{r}_M$ and $\mathbf{r}_F$ that represent the coordinates of the vector in M and F vary as the orientation of the two bases change. Consider F to be fixed and aligned with the global frame so that the unit vectors of F are $[\mathbf{i}, \mathbf{j}, \mathbf{k}]$. Then

$$r = x_F\,\mathbf{i} + y_F\,\mathbf{j} + z_F\,\mathbf{k}. \tag{4.1}$$

where $\mathbf{r}_F = [x_F \quad y_F \quad z_F]^{\mathrm{T}}$. M, with defining unit vectors $[\mathbf{m}_{\mathrm{x}}, \mathbf{m}_{\mathrm{y}}, \mathbf{m}_{\mathrm{z}}]$ is arbitrarily oriented with respect to F such that

$$r = x_M\,\mathbf{m}_{\mathrm{x}} + y_M\,\mathbf{m}_{\mathrm{y}} + z_M\,\mathbf{m}_{\mathrm{z}}. \tag{4.2}$$

Equating (4.1) and (4.2) gives

$$x_F\,\mathbf{i} + y_F\,\mathbf{j} + z_F\,\mathbf{k} = x_M\,\mathbf{m}_x + y_M\,\mathbf{m}_y + z_M\,\mathbf{m}_z \tag{4.3}$$

If the scalar products of this equation with **i**, **j**, **k** are taken, the resulting equations are, respectively:

$$\begin{aligned} x_F &= x_M\,\mathbf{m}_x.\mathbf{i} + y_M\,\mathbf{m}_y.\mathbf{i} + z_M\,\mathbf{m}_z.\mathbf{i} \\ y_F &= x_M\,\mathbf{m}_x.\mathbf{j} + y_M\,\mathbf{m}_y.\mathbf{j} + z_M\,\mathbf{m}_z.\mathbf{j} \\ z_F &= x_M\,\mathbf{m}_x.\mathbf{k} + y_M\,\mathbf{m}_y.\mathbf{k} + z_M\,\mathbf{m}_z.\mathbf{k} \end{aligned} \tag{4.4}$$

This can be written in matrix form as

$$\begin{bmatrix} x_F \\ y_F \\ z_F \end{bmatrix} = \begin{bmatrix} \mathbf{m}_x.\mathbf{i} & \mathbf{m}_y.\mathbf{i} & \mathbf{m}_z.\mathbf{i} \\ \mathbf{m}_x.\mathbf{j} & \mathbf{m}_y.\mathbf{j} & \mathbf{m}_z.\mathbf{j} \\ \mathbf{m}_x.\mathbf{k} & \mathbf{m}_y.\mathbf{k} & \mathbf{m}_z.\mathbf{k} \end{bmatrix} \begin{bmatrix} x_M \\ y_M \\ z_M \end{bmatrix} \tag{4.5}$$

or

$$\mathbf{r}_F = \mathbf{A}\mathbf{r}_M \tag{4.6}$$

Note that the matrix, **A**, is formed by three columns that are the scalar products of each of the unit vectors of M with the unit vectors of F. The matrix therefore has as its columns, the coordinates of the unit vectors of M with respect to F. It is clear, then, that knowing the matrix that relates a vector that is measured in two bases is tantamount to knowing the orientation of one basis with respect to the other (and therefore the relative orientation of two bodies in which the bases are fixed). This is not to say that the problem of orientation estimation is simply to find a matrix that relates the vector $\mathbf{r}_M$ to $\mathbf{r}_F$, as the matrix is in general non-unique. This problem is described in the following sections.

If rather than forming the matrix to relate the vector measured in M to the vector in F, the converse is undertaken, the result is

$$\begin{aligned} x_M &= x_F\,\mathbf{i.m_x} + y_F\,\mathbf{j.m_x} + z_F\,\mathbf{k.m_x} \\ y_M &= x_F\,\mathbf{i.m_y} + y_F\,\mathbf{j.m_y} + z_F\,\mathbf{k.m_y} \\ z_M &= x_F\,\mathbf{i.m_z} + y_F\,\mathbf{j.m_z} + z_F\,\mathbf{k.m_z} \end{aligned} \tag{4.7}$$

leading to:

$$\begin{bmatrix} x_M \\ y_M \\ z_M \end{bmatrix} = \begin{bmatrix} \mathbf{m}_x.\mathbf{i} & \mathbf{m}_x.\mathbf{j} & \mathbf{m}_x.\mathbf{k} \\ \mathbf{m}_y.\mathbf{i} & \mathbf{m}_y.\mathbf{j} & \mathbf{m}_y.\mathbf{k} \\ \mathbf{m}_z.\mathbf{i} & \mathbf{m}_z.\mathbf{j} & \mathbf{m}_z.\mathbf{k} \end{bmatrix} \begin{bmatrix} x_F \\ y_F \\ z_F \end{bmatrix} \tag{4.8}$$

or $\mathbf{r}_M = \mathbf{B}\mathbf{r}_F$ (4.9)

where the commutative property of the scalar product has been used in the writing of $\mathbf{B}$. Note that since $\mathbf{r}_F = \mathbf{A}\mathbf{r}_M$ and $\mathbf{r}_M = \mathbf{B}\mathbf{r}_F$ it follows that $\mathbf{r}_F = \mathbf{A}\mathbf{B}\mathbf{r}_F$, i.e. $\mathbf{AB} = \mathbf{1}$, or $\mathbf{B} = \mathbf{A}^{-1}$. Investigation of the matrices $\mathbf{A}$ and $\mathbf{B}$ shows that $\mathbf{B} = \mathbf{A}^{\mathrm{T}}$. Thus

$$\mathbf{A}^{-1} = \mathbf{A}^{\mathrm{T}} \tag{4.10}$$

This is the property of orthogonal matrices. It is easily shown that orthogonality of the rotation matrices is required for displacements to be rigid, i.e. the distance between two points is invariant under a rotation.

Orthogonal matrices with a determinant of +1 are known as rotation matrices. (Every orthogonal matrix has a determinant of either +1 or −1. The matrices with a determinant of −1 are reflection matrices).

The rotations associated with the rotation matrices may be considered in two ways, either:

- the rotation of a vector in a fixed basis or

- the measurement of a vector in two bases that are oriented such that their origins are coincident.

The second visualisation is the more natural in this work for reasons that will become obvious. The maths is the same for both visualisations; only the sign convention associated with rotations is different. (The above development was based on [11], [29] and [47]).

4.5.2 Cayley's Formula, the Rodrigues' Vector and the Axis Angle Formulation of the Rotation Matrix

Regardless of how the rotation is visualised, Euler's Rotation Theorem: *the displacement of a body with one point fixed is a rotation about an axis through that point,* is valid and useful. Since the rotation matrices have been shown to represent rotations, it follows that each rotation matrix is associated with an axis and angle of rotation.

4.5.2.1 The Rodrigues' Vector and Equation

Any vector that is collinear with the direction of rotation remains unchanged by the rotation matrix

$$\mathbf{x} = \mathbf{A}\mathbf{x} \tag{4.11}$$

where $\mathbf{x}$ is a vector of arbitrary magnitude along the axis of rotation. Cayley's formula, which is now derived, shows that every orthogonal matrix can be defined by three parameters. Knowing these three parameters is equivalent to knowing the associated rotation matrix and therefore the relative orientation of the frames in which the vector is measured.

Consider the rotation $\mathbf{R} = \mathbf{A}\mathbf{r}$. Since the magnitude of vectors are invariant under operation by rotation matrices, $\|\mathbf{R}\| = \|\mathbf{r}\|$, ($\|\mathbf{R}\|^2 = \|\mathbf{r}\|^2$) which can also be written:

$$\mathbf{R}.\mathbf{R} - \mathbf{r}.\mathbf{r} = 0. \tag{4.12}$$

Note that

$$(\mathbf{R} - \mathbf{r}).(\mathbf{R} + \mathbf{r}) = \mathbf{R}.\mathbf{R} + \mathbf{R}.\mathbf{r} - \mathbf{r}.\mathbf{R} - \mathbf{r}.\mathbf{r} = \mathbf{R}.\mathbf{R} - \mathbf{r}.\mathbf{r} = 0. \qquad (4.13)$$

The first and last elements of this chain of equalities show that the vectors ($\mathbf{R} - \mathbf{r}$) and ($\mathbf{R} + \mathbf{r}$) are orthogonal. This is shown geometrically in Figure 4.5.

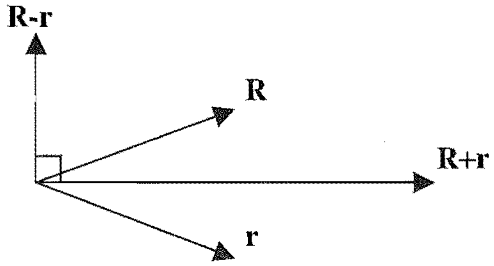


Figure 4.5 The orthogonality of R-r and R+r when R=Ar

Now ($\mathbf{R} - \mathbf{r}$) = ($\mathbf{A} - \mathbf{I}$)$\mathbf{r}$ and ($\mathbf{R} + \mathbf{r}$) = ($\mathbf{A} + \mathbf{I}$)$\mathbf{r}$, so that $\mathbf{r} = (\mathbf{A} + \mathbf{I})^{-1}(\mathbf{R} + \mathbf{r})$. Combining the first and last equalities gives

$$(\mathbf{R} - \mathbf{r}) = (\mathbf{A} - \mathbf{I})(\mathbf{A} + \mathbf{I})^{-1}(\mathbf{R} + \mathbf{r})$$
$$(\mathbf{R} - \mathbf{r}) = \mathbf{B}(\mathbf{R} + \mathbf{r}), \qquad (4.14)$$

where

$$\mathbf{B} = (\mathbf{A} - \mathbf{I})(\mathbf{A} + \mathbf{I})^{-1}. \qquad (4.15)$$

Defining $\mathbf{s} = (\mathbf{R} - \mathbf{r})$ and $\mathbf{t} = (\mathbf{R} + \mathbf{r})$ (which have been shown above to be orthogonal) (4.14) may be rewritten,

$$\mathbf{s} = \mathbf{Bt}, \qquad (4.16)$$

thus the matrix $\mathbf{B}$ is seen to have the property that the vector $\mathbf{Bx}$ is orthogonal to $\mathbf{x}$ that is,

$$\mathbf{x}.\mathbf{Bx} = 0, \qquad (4.17)$$

where **x** is an arbitrary vector. Expressing the elements of the matrix product (4.16) as a sum gives

$$s_i = \sum_j b_{ij} t_j \quad \text{for } i = 1,2,3 \tag{4.18}$$

where b_{ij} is the element in the i^{th} row and j^{th} column of **B**. Thus the scalar product (4.18) may be rewritten

$$\sum_i t_i \sum_j b_{ij} t_j = 0. \tag{4.19}$$

Expanding (4.19) leads to

$$t_1(b_{11}t_1 + b_{12}t_2 + b_{13}t_3) + t_2(b_{21}t_1 + b_{22}t_2 + b_{23}t_3) + t_3(b_{31}t_1 + b_{32}t_2 + b_{33}t_3) = 0$$

$$t_1^2\, b_{11} + t_2^2 b_{22} + t_3^2 b_{33} + t_1t_2(b_{12} + b_{21}) + t_1t_3(b_{13} + b_{31}) + t_2t_3(b_{23} + b_{32}) = 0.$$

or

$$\frac{1}{2}\sum_{i,j}(b_{ij} + b_{ji})t_i t_j = 0 \tag{4.20}$$

Equation (4.20) reveals, that for an arbitrary vector, **t**, the orthogonality of **t** and **Bt** can only be assured if **B** is skew-symmetric, i.e. $b_{ij} = -b_{ij}$ and $b_{ii} = 0$. Skew symmetric matrices have the form:

$$\mathbf{B} = \begin{bmatrix} 0 & -b_z & b_y \\ b_z & 0 & -b_x \\ -b_y & b_x & 0 \end{bmatrix} \tag{4.21}$$

and the property that $\mathbf{B}^{\mathrm{T}} = -\mathbf{B}$. The three elements of **B** are named so that the matrix product **Bt** is the same as the vector product **b** x **t**, where $\mathbf{b} = [b_x\ b_y\ b_z]^{\mathrm{T}}$. The vector, **b**, is known as the Rodrigues' vector. Using the original definition of **B** (4.15), we can find an expression for the rotation matrix **A** (Cayley's Formula):

$$\mathbf{B} = (\mathbf{A} - \mathbf{I})(\mathbf{A} + \mathbf{I})^{-1}$$

$$\mathbf{B}(\mathbf{A}+\mathbf{I}) = (\mathbf{A}-\mathbf{I})$$

$$\mathbf{BA}+\mathbf{B} = \mathbf{A} - \mathbf{I}$$

$$\mathbf{I}+\mathbf{B} = \mathbf{A}-\mathbf{AB} = (\mathbf{I}-\mathbf{B})\mathbf{A}$$

$$\mathbf{A} = (\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}+\mathbf{B}). \qquad (4.22)$$

Since **B** is defined by three parameters, the elements of the vector **b**, it is seen that the rotation matrix **A** is indeed defined by just three parameters. Further, note that due to the properties of the matrix **B**, the equation (4.14) may be rewritten as the Rodrigues' Equation

$$(\mathbf{R}-\mathbf{r}) = \mathbf{b} \times (\mathbf{R}+\mathbf{r}) \qquad (4.23)$$

This is the equation required for the method of orientation calculation employed in this work. The remaining part of this subsection explores the equivalence to the axis/angle formulation, which is useful in the next section. (This development was based on [11] and [47]).

4.5.2.2 The Relationship Between the Rodrigues' Vector and Axis of Rotation

Returning to the definition (4.11) of the axis of rotation:

$$(\mathbf{A}-\mathbf{I})\mathbf{x} = \mathbf{0}, \qquad (4.24)$$

where **x** is a vector collinear with the axis of rotation. Using Cayley's formula (4.22) to substitute for **A**

$$[(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}+\mathbf{B})-\mathbf{I}]\mathbf{x} = \mathbf{0}, \qquad (4.25)$$

and premultiplying (4.25) by $(\mathbf{I}-\mathbf{B})$ gives:

$$[(\mathbf{I}+\mathbf{B})-(\mathbf{I}-\mathbf{B})]\mathbf{x} = 2\mathbf{Bx} = \mathbf{0}. \qquad (4.26)$$

Recall that $\mathbf{Bx} = \mathbf{b} \times \mathbf{x}$, and $\|\mathbf{b} \times \mathbf{x}\| = \|\mathbf{b}\|\|\mathbf{x}\|\sin\theta$, where θ is the included angle between the vectors **b** and **x**. Hence (4.26) states that (for non-zero **b**) the vectors **b** and **x** are parallel and therefore the Rodrigues' vector, **b**, is along the axis of rotation. Define the unit vector along **b** as s. It has now been shown that

- Only three parameters are required to fully specify a rotation in 3-space
- The components of the Rodrigues' vector can be used to specify a rotation

These facts imply that within the definition of **b**, the Rodrigues' vector must be information concerning the angle of rotation. Since this vector has been shown to be collinear with the axis of rotation, the only 'degrees of freedom' left in the vector are its magnitude and sense. It seems logical that the magnitude of the vector must contain the information. This is now shown to be the case.

Consider the rotation of a point with position vector **r**, through an angle ϕ about the unit vector **s** to yield the vector **R**. Construct a plane through the points r and R (the terminal points of **r** and **R**) that is normal to the axis of rotation. This plane intersects the axis at the point s_0, with associated vector $\mathbf{s}_0$ as shown in Figure 4.6

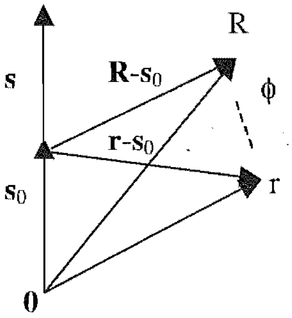


Figure 4.6 Rotation of r to R about s and the construction of a plane normal to s, through the tips of r and R.

A view normal to the plane shows two vectors $\mathbf{R}\text{-}\mathbf{s}_0$ and $\mathbf{r}\text{-}\mathbf{s}_0$, of equal magnitude separated by an angle, ϕ. This situation is shown in Figure 4.7.

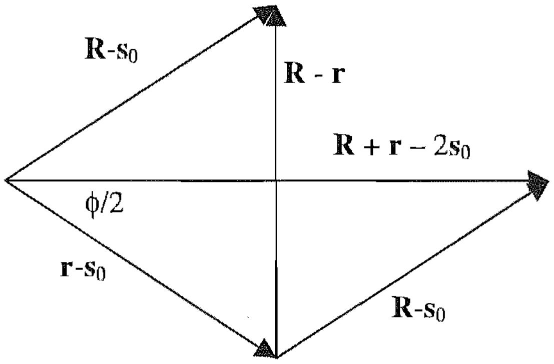


Figure 4.7 A view of the rotation from r to R from a plane normal to the axis of rotation.

Using simple trigonometry it is seen that $\tan(\phi/2) = \|\mathbf{R} - \mathbf{r}\|/\|\mathbf{R} + \mathbf{r} - 2s_0\|$, thus

$$\|\mathbf{R} - \mathbf{r}\| = \tan(\phi/2)\|\mathbf{R} + \mathbf{r} - 2s_0\|, \tag{4.27}$$

which can be used to form a vector product. To do this requires that we find a vector of the same orientation and magnitude as **R-r**. A vector parallel to **R-r**, would result from the cross product of s (out of the page in Fig. 4.7) and $(\mathbf{R} + \mathbf{r} - 2s_0)$. Since the vectors s and $(\mathbf{R} + \mathbf{r} - 2s_0)$ are orthogonal, the definition of the magnitude of the vector product gives

$$\|s \times (\mathbf{R} + \mathbf{r} - 2s_0)\| = \|s\| \, \|(\mathbf{R} + \mathbf{r} - 2s_0)\| = \|(\mathbf{R} + \mathbf{r} - 2s_0)\| \tag{4.28}$$

To obtain the vector **R-r**, this vector product must be multiplied by the scalar $\tan(\phi/2)$,

$$(\mathbf{R} - \mathbf{r}) = \tan(\phi/2)s \times (\mathbf{R} + \mathbf{r} - 2s_0). \tag{4.29}$$

Note that the axis of rotation, s, and the position vector of the intersection of the axis with the constructed plane, s_0, are collinear, hence $s \times s_0 = \mathbf{0}$ and this equation may be rewritten,

$$(\mathbf{R} - \mathbf{r}) = \tan(\phi/2)s \times (\mathbf{R} + \mathbf{r}). \tag{4.30}$$

Recall Rodrigues' equation (4.23) $(\mathbf{R} - \mathbf{r}) = \mathbf{b} \times (\mathbf{R} + \mathbf{r})$, comparing this with the previous equation shows that

$$\mathbf{b} = \tan(\phi/2)\mathrm{s}, \qquad (4.31)$$

where ϕ is the angle of rotation about s. This is the result that relates the Rodrigues' vector to the axis and angle of rotation. (A similar development can be found in [11]).

4.5.2.3 Axis/Angle Formulation of the Rotation Matrix

It is possible to use Cayley's formula (4.22) and the relationship between the Rodrigues' vector and axis of rotation (4.31) to derive an axis/angle representation of the rotation matrix. This, however, requires a large amount of manipulation, and it is considered to be much more useful to derive the result from a graphical approach, which is based upon [29].

Consider a vector **r** rotated about s (unit vector) through an angle ϕ, to yield **R**.

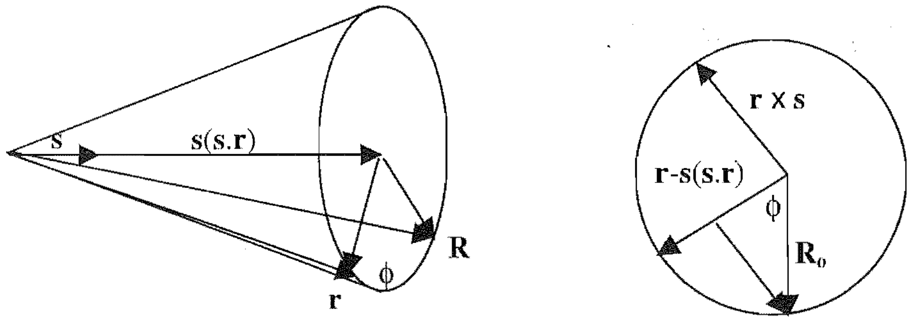


Figure 4.8 Two views of a general rotation. The vector R is obtained by rotating r about s, through the angle ϕ.

The two vectors, $\mathbf{r} \times \mathrm{s}$ and $\mathbf{r}\text{-}\mathrm{s}(\mathrm{s}.\mathbf{r})$ are of the same magnitude (the radius of the circle) and are orthogonal. The component of the vector **R** orthogonal to s, defined as $\mathbf{R}_0$:

$$\mathbf{R} = \mathrm{s}(\mathrm{s}.\mathbf{r}) + \mathbf{R}_0 \qquad (4.32)$$

can be seen to be given by

$$\mathbf{R}_0 = (\mathbf{r}\text{-s(s.}\mathbf{r}))\cos\phi - (\mathbf{r} \times \text{s})\sin\phi \qquad (4.33)$$

as shown in Fig 4.8. Combining (4.32) and (4.33) an expression for **R** is therefore

$$\mathbf{R} = \text{s(s.}\mathbf{r}) + (\mathbf{r}\text{-s(s.}\mathbf{r}))\cos\phi - (\mathbf{r} \times \text{s})\sin\phi. \qquad (4.34)$$

To use this expression as the definition for a rotation matrix it must be possible to extract the vector **r** from all terms to the right hand side of the equation. Terms of the form s(s.r) can be rewritten as $\text{ss}^{\text{T}}\mathbf{r}$. The only other non-trivial term is **r** X s. Using the previously defined skew symmetric matrix definition, a matrix **S** is defined such that **Sr** = s X **r** = -**r** X s. Using these notations, (4.34) may now be rewritten

$$\begin{aligned}\mathbf{R} &= [\text{ss}^{\text{T}} + (\mathbf{I}\text{- ss}^{\text{T}})\cos\phi + \mathbf{S}\sin\phi]\mathbf{r} \\ &= [\mathbf{I}\cos\phi + \text{ss}^{\text{T}}(1 - \cos\phi) + \mathbf{S}\sin\phi]\mathbf{r} \\ &= \mathbf{Ar} \qquad (4.35)\end{aligned}$$

Note that

$$\text{ss}^{\text{T}} = \begin{bmatrix} s_x^2 & s_x s_y & s_x s_z \\ s_x s_y & s_y^2 & s_y s_z \\ s_x s_z & s_y s_z & s_z^2 \end{bmatrix}$$

while

$$\mathbf{S}^2 = \begin{bmatrix} -s_y^2 - s_z^2 & s_x s_y & s_x s_z \\ s_x s_y & -s_x^2 - s_z^2 & s_y s_z \\ s_x s_z & s_y s_z & -s_x^2 - s_y^2 \end{bmatrix}$$

Since s is a unit vector the diagonal terms of $\mathbf{S}^2$ can be rewritten $s_x^2 - 1$, $s_y^2 - 1$ and $s_z^2 - 1$, meaning that

$$\mathbf{S}^2 = \text{ss}^{\text{T}} - \mathbf{I}. \qquad (4.36)$$

Using this (4.36), (4.35) may be rewritten,

$$\begin{aligned}\mathbf{A} &= \mathbf{I}\cos\phi + (\mathbf{S}^2 + \mathbf{I})(1 - \cos\phi) + \mathbf{S}\sin\phi \\ &= \mathbf{I} + \mathbf{S}^2(1 - \cos\phi) + \mathbf{S}\sin\phi. \qquad (4.37)\end{aligned}$$

Thus, given the axis of rotation and the angle of rotation about that axis, it is possible to generate the associated rotation matrix. At this point, sufficient tools have been developed to proceed to the actual aim of this section: orientation/attitude estimation.

4.6 Calculation of Relative Orientation

As previously discussed, the aim of this work is to use vector observations of a field from two bodies to find their relative orientation. The previous section showed this to be equivalent to finding a rotation matrix that relates the two vector measurements. This section shows, using geometric concepts, that it is impossible to uniquely define the relative orientation of two bodies using the observation of a single vector quantity from the two bodies. In other words, it is shown that given two vectors of the same magnitude there are an infinite set of rotation matrices that relate the two vectors.

4.6.1 The Disc Argument

When a vector is rotated about an axis, the component of the vector along the axis remains constant. This being the case, given a pair of vectors, one a rotation of the other, the axis of rotation must have the same included angle with both vectors. This condition yields a unit disc of possible rotation axes that bisects the angle between the two vectors. Each of the infinitude of axes has an associated angle of rotation, and when the axis and angle are combined, using the expression (4.37) of the previous section, different rotation matrices result. The result of the non-uniqueness of the rotation axis is that there are an infinite group of rotation matrices, **A**, that fulfil the relation:

$$\mathbf{x}_F = \mathbf{A}\mathbf{x}_M. \tag{4.38}$$

4.6.2 The Cone Argument

Another explanation of what will be referred to as the non-uniqueness problem can be given by the 'cone argument' as follows. Consider $\|\mathbf{x}_M\| = \|\mathbf{x}_F\| = 1$. It must be possible to find a matrix **A** such that $\mathbf{x}_F = \mathbf{A}\mathbf{x}_M$, where the columns of **A** are the unit vectors of M with respect to F (i.e. $\mathbf{A} = [\mathbf{X}_M\ \mathbf{Y}_M\ \mathbf{Z}_M]$ where $\mathbf{X}_M$ is the unit vector representing the orientation of the X-axis of M w.r.t F etc.). The equation $\mathbf{x}_F = \mathbf{A}\mathbf{x}_M$ is a compact expression of three scalar products, which using the properties of orthogonal matrices ($\mathbf{A}^T = \mathbf{A}^{-1}$) can be written

$$x_{M1} = \mathbf{X}_M . \mathbf{x}_F$$
$$x_{M2} = \mathbf{Y}_M . \mathbf{x}_F \tag{4.39}$$
$$x_{M3} = \mathbf{Z}_M . \mathbf{x}_F$$

where $\mathbf{x}_M = [x_{M1}\, x_{M2}\, x_{M3}]^{\mathrm{T}}$. Consider the first of these relations. Since both vectors are of unit magnitude, this equations states that the included angle of the X-axis of *M* and $\mathbf{x}_F$ is arcos(x_{M1}). Vectors, $\mathbf{X}_M$, that satisfy this requirement form a cone with $\mathbf{x}_F$ as longitudinal axis. The same argument can be used to show that each of the axes of *M* lie on cones that share $\mathbf{x}_{\mathrm{F}}$ as their axis. Aside from lying on the cones, the axes must form an orthogonal right-handed system. Since it is possible to find one rotation matrix **A** that relates $\mathbf{x}_F$ and $\mathbf{x}_M$ it can be seen that there are an infinite set of matrices that fulfil the requirements. The set is formed by rotating the original frame (arbitrarily defined), *M*, about $\mathbf{x}_F$. The rotation of the initial solution about $\mathbf{x}_F$ results in three cones, corresponding to the possible solution spaces of $\mathbf{X}_{\mathrm{M}}$,$\mathbf{Y}_M$ and $\mathbf{Z}_M$.

Using either the infinite set of rotation axes or the cone argument shows that it is impossible to uniquely and consistently relate two vectors based only upon the vectors themselves.

If the non-uniqueness problem is to be resolved, additional information is required. The minimum information that can be used to make the solution unique is any pair of angles between the axes of *F* and *M*. In the rowing problem, one of these additional angles, say the sweep angle, could be provided through use of a potentiometer. Alternatively, multiple vector observations can be made from each frame. Considering two vector observations from each of the bodies the disc argument is modified. In possession of $\mathbf{x}_M$, $\mathbf{y}_M$, $\mathbf{x}_F$ & $\mathbf{y}_F$ we seek a unique rotation matrix that satisfies $\mathbf{x}_F = \mathbf{A}\mathbf{x}_M$ and $\mathbf{y}_F = \mathbf{A}\mathbf{y}_M$

4.6.3 Two Vector Observations

Potential axes of rotation lie on the intersection of the discs that bisect the angles between the pairs ($\mathbf{x}_F$, $\mathbf{x}_M$) and ($\mathbf{y}_F$, $\mathbf{y}_M$). If $\boldsymbol{x}$ and $\boldsymbol{y}$ (the vectors that are measured in the two frames) are non-collinear then under most conditions the discs are distinct and intersect along a line through the origin, giving two possible values for the axis of rotation: s and –s, see Figure 4.9. The two axes are associated with angles of ϕ and $-\phi$ and therefore generate the same, unique, rotation matrix.

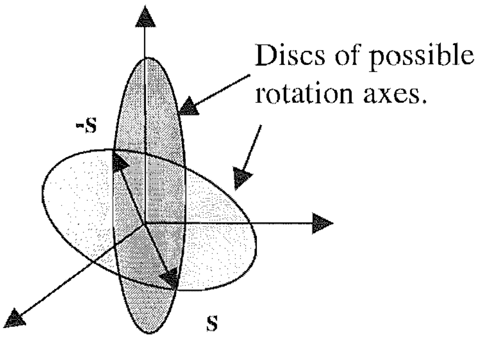


Figure 4.9 The intersection of two unit discs of possible axes of rotation defines two possible axes of rotation s and -s.

What may seem a counterintuitive result is that even if the discs do coincide it is still possible, as long as $\boldsymbol{x}$ and $\boldsymbol{y}$ are non-collinear, to identify the relative orientation of two bodies. To understand this requires a more thorough investigation of the formation of the discs.

Recall that the disc is formed by the requirement that the axis of rotation has the same included angle with any given pair of vectors, one a rotation of the other.

To find a method by which the axis of rotation can be identified, when the discs are coincident requires that we investigate the situations under which this occurs. This is undertaken geometrically. Recall the discs bisect the angles between the vector pairs ($\mathbf{x}_F$, $\mathbf{x}_M$) and ($\mathbf{y}_F$, $\mathbf{y}_M$). Clearly, for the discs to be coincident requires that there exists a single plane that bisects the angles between these pairs. This plane must also pass through the origin, since it also bisects any scaled versions of the vectors.

The problem is approached, somewhat 'backwards', i.e. we assume we know the orientation of the axis and angle of rotation and then show the conditions under which the discs bisecting the vector pairs are coincident. This is preferable to the 'other direction' since the vectors and their rotations must be physically realizable, whereas if arbitrary vectors are chosen, it may not be possible that they are related by a single rotation.

Assume the axis of rotation lies along the X-axis, this makes the drawing easier but does not cause any loss of generality. We now make use of the idea of a sub-space, actually a plane, that passes through two arbitrary vectors, $\mathbf{x}_M$ and $\mathbf{y}_M$ and the axis of rotation, $s = \mathbf{X}$. The plane including $\mathbf{x}_M$ contains all vectors of the form $\alpha\mathbf{x}_M + \beta s$. Upon rotation, the plane maps to $\mathbf{A}(\alpha\mathbf{x}_M + \beta s) = \alpha\mathbf{A}\mathbf{x}_M + \beta\mathbf{A}s = \alpha\mathbf{x}_F + \beta s$, that is, the plane that passes through the two vectors $\mathbf{x}_M$ and s becomes the plane including $\mathbf{x}_F$ and s. This situation is shown below in Figure 4.10. The case for two arbitrary vectors, $\mathbf{x}_M$ and $\mathbf{y}_M$ is also shown in Figure 4.11.

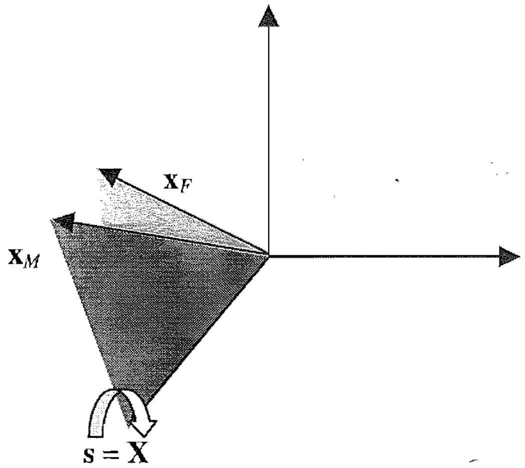


Figure 4.10 The plane including the axis of rotation, s, and the vector $\mathbf{x}_M$ rotates to form the plane including s and $\mathbf{x}_F$.

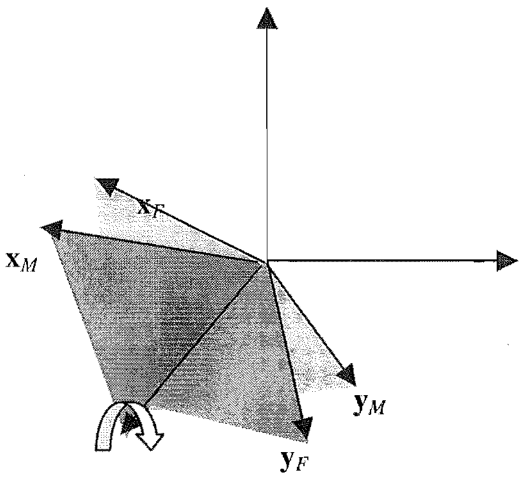


Figure 4.11 The planes generated when two arbitrary vectors, x_M and y_M are rotated about the X-axis.

Two discs of candidate axes of rotation, D_x and D_y, bisect the planes created by the **x**'s and **y**'s respectively. The only condition under which D_x and D_y are coincident is seen to be when $\mathbf{x}_M$, $\mathbf{y}_M$ and **s** all lie on the same plane. In this situation the axis of rotation lies at the intersection of the planes containing ($\mathbf{x}_M$, $\mathbf{y}_M$, **s**) and ($\mathbf{x}_F$, $\mathbf{y}_F$, **s**). This is since the intersection is the invariant direction, and using the definition of the planes, i.e. if $\alpha\mathbf{x}_M + \beta\mathbf{s} = \beta\mathbf{s}$, then $\mathbf{A}(\alpha\mathbf{x}_M + \beta\mathbf{s}) = \mathbf{A}(\beta\mathbf{s}) = \beta\mathbf{s}$. This is shown in Figure 4.12 for an arbitrary axis of rotation.

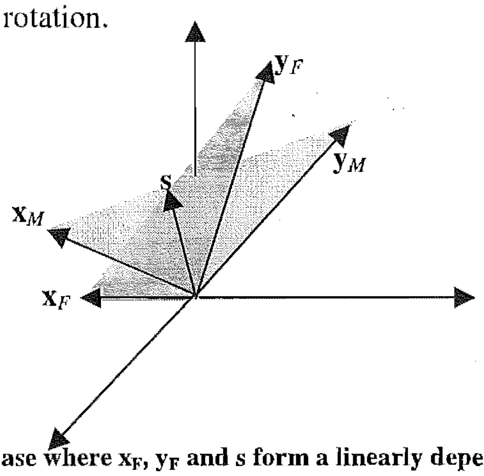


Figure 4.12 The case where x_F, y_F and s form a linearly dependent set.

This geometric method of identification of the axis of rotation does not work for cases in which ($\mathbf{x}_M$, $\mathbf{y}_M$, **s**) are linearly independent (do not lie in a single plane).

The only case in which taking two vector measurements will not lead to a unique solution is when the vectors are collinear. This is exactly equivalent to the case in which only one measurement is taken. Conceptually this leaves a single disc, or an infinite number of planes that can go through each of the 'vector pairs' ($\mathbf{x}_F = \mathbf{y}_F$) and ($\mathbf{x}_M = \mathbf{y}_M$).

It may seem that two completely different methods are being employed to find the axis of rotation for the two cases ($\mathbf{x}_M$, $\mathbf{y}_M$, $\mathbf{s}$) linearly dependent and independent. However, all that is different is the method by which the initial defining relations are applied. It is not proposed that the above-mentioned methods, or their mathematical translations are used to calculate the axis of rotation. The diversion was simply made to prove that even if the discs of potential axes of rotation are coincident, it is still possible to find the axis of rotation.

For the interested reader, the mathematical 'translation' of the disc method in the linearly independent case is easily reduced to finding a unit vector that is orthogonal to both ($\mathbf{x}_M - \mathbf{x}_F$) and ($\mathbf{y}_M - \mathbf{y}_F$), which can be accomplished through normalising the vector product of these two 'observation differences'. This is impossible in the linearly dependent case, since the observation differences are parallel, and the vector product therefore results in the zero vector.

These methods only allow for the estimation of the axis of rotation. In possession of the axis it is, however, possible to find the required angle of rotation, which is unique within added integer multiples of 2π due to an agreed sign convention ('right hand rule').

The method that is actually used in this work to estimate orientation, simultaneously calculates the axis and angle of rotation, using the previously introduced Rodrigues' vector, introduced in Section 4.5. It is more robust to sensor noise than the mathematical translations of the above methods.

4.7 Relative Orientation Estimation

While it is theoretically possible to find the rotation matrix using the geometric approaches implicitly outlined in the previous section, it must be remembered that we are dealing with sensors that will have noisy outputs. Error in the vector observations will cause the discs to rotate, meaning that the intersection, which is a function of the orientation of the two discs, will be removed from the ideal. Rather than calculating the orientation exactly, we are therefore reduced to estimating the orientation so that some optimality criterion is fulfilled.

Orientation estimation has applications in wide ranging areas from aerospace to computer vision. A number of approaches that have been used are reviewed in full detail in Appendix A2. A new method, that of estimating the Rodrigues' vector, is presented below. This method is computationally inexpensive, allows relative weightings of measurements and was found to either have accuracy greater than or equal to the reviewed methods in the case at hand.

4.7.1 A New Method of Orientation Estimation

A feature of all the reviewed algorithms in (see Appendix A2) is that for $n \leq 3$ it is impossible to weight one observation above another. A method of estimating rotation from $n \geq 2$ observations that uses only a 3x3 matrix inversion and allows differential weighting of a sort is now presented. This algorithm has not been found anywhere in the literature and could be quite a useful new addition to the arsenal of orientation estimation techniques.

In the following, to simplify notation, **R** and **r** are observations of the same vector from two bases. Recall the Rodrigues' equation (4.23):

$$(\mathbf{R} - \mathbf{r}) = \mathbf{b} \times (\mathbf{R} + \mathbf{r})$$

where $\mathbf{R} = \mathbf{A}\mathbf{r}$, and **b** is the Rodrigues' vector. Now $\mathbf{c} \times \mathbf{d} = -\mathbf{d} \times \mathbf{c}$, so

$$\mathbf{R} - \mathbf{r} = -(\mathbf{R} + \mathbf{r}) \times \mathbf{b} \tag{4.40}$$

Define a matrix Σ, so that $\Sigma\mathbf{c} = -(\mathbf{R} + \mathbf{r}) \times \mathbf{c}$, i.e.

$$\Sigma = \begin{bmatrix} 0 & R_3 + r_3 & -(R_2 + r_2) \\ -(R_3 + r_3) & 0 & R_1 + r_1 \\ R_2 + r_2 & -(R_1 + r_1) & 0 \end{bmatrix} \tag{4.41}$$

and the vector $\Delta = \mathbf{R} - \mathbf{r}$, so that the (4.40) may be rewritten

$$\Delta = \Sigma\mathbf{b}. \tag{4.42}$$

Σ is singular (as are all skew-symmetric matrices), so the system cannot be solved simply by inverting Σ. This is proof in itself that a single vector observation is insufficient to specify relative orientation. Consider two separate vector quantities that are measured in the two bases to be related. Let $\mathbf{Y} = \mathbf{A}\mathbf{y}$, $\Delta_r = \mathbf{R} - \mathbf{r}$, $\Delta_y = \mathbf{Y} - \mathbf{y}$, $\Sigma_r = \Sigma$ (defined above), and

$$\Sigma_y = \begin{bmatrix} 0 & Y_3 + y_3 & -(Y_2 + y_2) \\ -(Y_3 + y_3) & 0 & Y_1 + y_1 \\ Y_2 + y_2 & -(Y_1 + y_1) & 0 \end{bmatrix} \tag{4.43}$$

For simultaneous measurements we then have two matrix equations involving the Rodrigues' vector

$$\Delta_r = \Sigma_r\mathbf{b} \tag{4.44(a)}$$

$$\Delta_y = \Sigma_y\mathbf{b} \tag{4.44(b)}$$

or combining them (4.44(a)&(b)):

$$\begin{bmatrix} \Delta_r \\ \Delta_y \end{bmatrix} = \begin{bmatrix} \Sigma_r \\ \Sigma_y \end{bmatrix} \mathbf{b} \tag{4.45}$$

$$\Delta_C = \mathbf{\Sigma}_C\mathbf{b} \tag{4.46}$$

Define $\varepsilon = \Delta_C - \mathbf{\Sigma}_C\mathbf{b}$, then the least squares solution minimises

$$J(\mathbf{b}) = \varepsilon^T\varepsilon = (\Delta_C - \mathbf{\Sigma}_C\mathbf{b})^T(\Delta_C - \mathbf{\Sigma}_C\mathbf{b}). \tag{4.47}$$

The solution is

$$\mathbf{b} = (\mathbf{\Sigma}_C^T\mathbf{\Sigma}_C)^{-1}\mathbf{\Sigma}_C^T\Delta_C. \tag{4.48}$$

If the vector observations are error free, this is the exact solution for **b**, i.e. there exists a **b**, such that $J(\mathbf{b}) = 0$.

Corresponding to the discussion in Section 4.6.3, the only condition under which the matrix $\mathbf{\Sigma}_C^T\mathbf{\Sigma}_C$ is singular is if the vectors being observed are collinear. If one of the vector observations is a linear combination of the other observation and the axis of rotation, the case for which the discs of potential axes of rotation are coincident, the matrix remains non-singular.

Typically, measurements of the same quantity will be obtained using similar instruments. If one instrument is known to be more accurate than another it may be desired that the observations from this sensor be weighted more heavily than the other. This is simply achieved using 'weighted least squares', where the aim is to minimise

$$\varepsilon^T\mathbf{W}\varepsilon = (\Delta_C - \Sigma_C\mathbf{b})^T\mathbf{W}(\Delta_C - \Sigma_C\mathbf{b}) \tag{4.49}$$

The weighted least squares solution is

$$\mathbf{b} = (\Sigma_C^T\mathbf{W}\Sigma_C)^{-1}\Sigma_C^T\mathbf{W}\Delta_C. \tag{4.50}$$

In possession of an estimate of the Rodrigues' vector associated with the rotation, there are a number of ways in which the associated rotation matrix can be formed. These include: forming the skew symmetric matrix **B** associated with **b** and then using the direct consequence of Cayley's formula $\mathbf{A} = (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} + \mathbf{B})$; finding the angle of rotation by $2\tan^{-1}\|\mathbf{b}\|$, and the axis of rotation by $\mathbf{b}/\|\mathbf{b}\|$ and using the axis-angle formulation (4.37); or the following closed form expression [37]

$$\mathbf{A} = \frac{2}{d}\begin{bmatrix} b_1^2+1 & b_1b_2-b_3 & b_1b_3+b_2 \\ b_1b_2+b_3 & b_2^2+1 & b_2b_3-b_1 \\ b_1b_3-b_2 & b_2b_3+b_1 & b_3^2+1 \end{bmatrix} - \mathbf{1} \qquad (4.51)$$

where $\mathbf{b} = [b_1\ b_2\ b_3]$ and $d = \|\mathbf{b}\|$. The theoretical advantages of the new proposed method: 'Least Squares Estimation of the Rodrigues' Vector' are:

1. it always produces a rotation matrix (i.e. orthogonal matrix with det = +1)
2. it functions with $n \geq 2$ measurements
3. the major mathematical function is a 3x3 matrix inversion (as compared to SVD or eigendecomposition)
4. it is possible to weight observations even for $n = 2$

These characteristics can be compared with those of the reviewed methods in Table 4.2.

While the advantages appear positive, it was thought that testing should be done to confirm the method's utility. The case for which testing was conducted is similar to that to be used in this work – that of two noisy vector observations. So that the new method could be compared to a large number of alternatives, Black's method was used to generate a third observation from the original two. (If this were not done, the only method that could be used as a comparison would be that of Arun.)

Vector observations $\mathbf{r}_{1\text{true}}$ and $\mathbf{r}_{2\text{true}}$ were generated randomly, with $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{R}_1$ and $\mathbf{R}_2$ formed in the following way:

$$\begin{aligned} \mathbf{r}_1 &= \mathbf{r}_{1\text{true}} + \mathbf{n}_1 & \mathbf{R}_1 &= \mathbf{A}\mathbf{r}_{1\text{true}} + \mathbf{n}_1 \\ \mathbf{r}_2 &= \mathbf{r}_{2\text{true}} + \mathbf{n}_2 & \mathbf{R}_2 &= \mathbf{A}\mathbf{r}_{2\text{true}} + \mathbf{n}_2 \end{aligned} \qquad (4.52)$$

where: **A** is a rotation matrix generated by a random axis and angle of rotation

$\mathbf{n}_1$ and $\mathbf{n}_2$ are zero mean Gaussian white noise vectors of variance σ_1 and σ_2

Using Black's method the third observations are

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \qquad \mathbf{R}_3 = \mathbf{R}_1 \times \mathbf{R}_2 \tag{4.53}$$

Name	Orthogonal/Non-Orthogonal	Required Independent Measurements	Main Computation	Comments
Brock Constrained [14]	Orthogonal	≥ 3	Matrix Square Root	A standard solution to the question posed by Wahba.
Arun et al [4]	Orthogonal	≥ 2	SVD	Can yield a reflection matrix or a rotation matrix.
Markley & Bar-Itzhack [45]/ Brock Unconstrained [14]	Non-Orthogonal	≥ 3	(3x3) Matrix Inversion	Can be made orthogonal by conditioning the measurement matrices.
Carta & Lackowski [16], Markley & Bar-Itzhack [45]	Orthogonal	≥ 3	(3x3) Matrix Inversion & Matrix Square Root	Not guaranteed to be the least squares orthogonal estimate.

Table 4.2 Characteristics of reviewed attitude estimation algorithms.

Note: Any algorithm that requires $n \geq 3$ linearly independent measurements can be used with Black's method of using the vectors defined by the cross product of the vectors in each frame.

A master function was written to recursively call all the previously described functions with the same observations before changing both **A** and the observations. At the conclusion of each cycle the Frobenius norms of the difference between the true rotation matrix and that calculated by each of the methods was calculated and stored. For reference, the methods tested were:

1. Arun's method with two observations (A2)
2. Arun's method with three observations (A3)
3. Brock's constrained method (B)
4. Markley's unconstrained method (MU)
5. Markley's unconstrained followed by 'Carta Orthogonalisation' (MC)
6. Least squares estimation of Rodrigues' vector (RLS)
7. Weighted least squares estimation of Rodrigues' vector (WRLS2)
8. Weighted least squares estimation of Rodrigues' vector using three observations (WRLS3)

A number of different values for σ_1 and σ_2 were used. Increasing both indicated how robust the solutions were, while having one larger and tuning the weighted least squares method accordingly suggested this approach's efficacy. The **W** matrices were chosen to be of the form

$$\mathbf{W} = \begin{bmatrix} w_1\mathbf{1} & \mathbf{0} \\ \mathbf{0} & w_2\mathbf{1} \end{bmatrix}, \tag{4.54}$$

where **1** and **0** are 3x3 identity and zero matrices respectively, and w_1 and w_2 are the weights attributed to the measurement of the pairs $(\mathbf{r}_1,\mathbf{R}_1)$ and $(\mathbf{r}_2,\mathbf{R}_2)$. Obvious extensions were made for the three-measurement case.

At the conclusion of each run, which consisted of 1000 calls to each routine, the means and variances of the Frobenius norms of the matrix differences were calculated. Making comparison easier is the fact that the three solutions to Wahba's problem (A2, A3, B) yielded exactly the same results. (It is likely, however, that their computation times would be different.) This being the case, the comparison was reduced to that of the general Wahba solution, Markley's unconstrained method, its orthogonalised form, and the three listed variants on the proposed Rodrigues' vector estimation scheme.

The Wahba solutions (A2, A3, B) were always better than the unconstrained (MU) and orthogonalised versions (MC), but only as good as the Rodrigues trio in the case of zero or very low-level noise. Orthogonalising the Markley unconstrained estimate reduced error in all circumstances.

In the case of different values for σ_1 and σ_2 it was easy to assign weightings in the two-measurement Rodrigues' estimation scheme to achieve consistently lower error norms. It was more difficult, although possible, to select a third value to get better performance still. The only conditions under which RLS performed better than WRLS2 and WRLS3, were if different weightings were assigned to the two actual measurements when the noise strengths were the same. Even in this condition the error norms from WRLS2 and WRLS3 were less than that for the Wahba solutions (for reasonable weightings).

This qualitative testing suggests that the best option for our purpose is either RLS or WRLS2. The extra computation time and small performance benefit of WRLS3 as compared to WRLS2 mean that its choice was not justified. WRLS2 would be the algorithm of choice if 'sensible' choices for the weighting matrices could be made. A method for choosing such matrices is discussed in Section 4.9.

Having chosen a method by which vector observations will be processed, the next step is to show that the outputs yielded by the previously described sensors (accelerometers and magnetoresistive sensors) are suitable in the application.

4.8 Theoretical Relative Orientation Estimation Using Accelerometers and Magnetoresistive Sensors

Up to this point, while sensors have been mentioned, it has simply been assumed that two independent vector measurements can be gathered. No mention of the relationship between these vector quantities and the actual motion has been made. This section rectifies this situation by considering a general two-body configuration. The mapping between the physical motion of the bodies, i.e. linear and rotational quantities, and the outputs of theoretical sensors mounted on the bodies is then derived.

4.8.1 Problem Definition

Two bodies, U and L, with associated coordinate frames *U* and *L*, are connected at a spherical joint. A frame, *J*, that is at all times parallel to the inertial frame, *F*, is coincident with the centre of the joint. The position vector $O_F \rightarrow O_J$ that measures the translation of the joint is denoted by $\boldsymbol{p}$. The frames *J*, *U* and *L* have coincident origins at the centre of the joint. This situation is shown in Fig. 4.13.

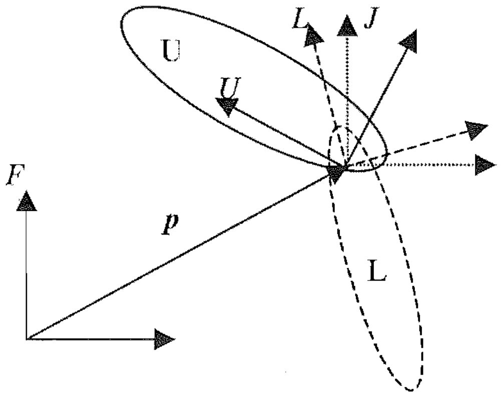


Figure 4.13 A 2D representation of the problem geometry.

A point x is fixed in the upper body, U. The coordinate matrices of this point with respect to the bases U and J are related by

$$\mathbf{x}_J = \mathbf{A}\mathbf{x}_U \tag{4.55}$$

where **A** is the rotation matrix that has as its columns the coordinate matrices of the unit vectors of U w.r.t J (and therefore F). Similarly, for a point, y, fixed in L:

$$\mathbf{y}_J = \mathbf{B}\mathbf{y}_L. \tag{4.56}$$

where the columns of **B** give the orientation of L w.r.t. J (F).

Located at both x and y are sensor clusters consisting of triaxial magnetoresistive (MR) sensors and triaxial accelerometers. The ultimate aim is to find the orientation of U with respect to L using the outputs of these sensors. Before this may be achieved, the outputs of these sensors in response to a general motion must be derived. This analysis is performed first for the MR sensors, and then the triaxial accelerometers. Once the relationships have been derived, schemes for selecting parameters in the weighting matrix in the orientation estimation algorithm are discussed in Section 4.9.

4.8.2 Theoretical Output of the Magnetoresistive Sensor Under General Motion

In the following, the triaxial magnetoresistive sensor, M, is assumed to be a perfect sensor of the incident magnetic field, **E**. This means that each axis of M outputs a voltage exactly proportional to the component of **E** perpendicular to itself. Denoting the axes of M as the right-handed orthogonal set of unit vectors $M = \{\mathbf{x}_M\ \mathbf{y}_M\ \mathbf{z}_M\}$ (where the unit vectors are defined in the reference basis F), and a field that is orthogonal to **E** as **F**, the output of the sensor, M, is given by the voltage vector $\mathbf{V}_M = [\mathrm{v}_{\mathrm{x}M}\ \mathrm{v}_{\mathrm{y}M}\ \mathrm{v}_{\mathrm{z}M}]^{\mathrm{T}}$. The components of $\mathbf{V}_M$ fulfil the following relations:

$$\begin{aligned} \mathrm{v}_{\mathrm{x}M} &= \mathbf{F}.\mathbf{x}_M \\ \mathrm{v}_{\mathrm{y}M} &= \mathbf{F}.\mathbf{y}_M \\ \mathrm{v}_{\mathrm{z}M} &= \mathbf{F}.\mathbf{z}_M. \end{aligned} \tag{4.57}$$

If the field **F** is assumed to be a constant vector field, an assumption that is soon relaxed, translating the sensor without changing its orientation relative to the field will not change its output. Further, since the field is assumed to be constant, rotating the sensor will always yield a vector of the same magnitude. It is seen then that regardless of the displacement experienced by the sensor, only a rotation is sensed, and thus even if the sensors are submitted to non-spherical displacements, the output vectors will be able to be related by rotation matrices. Note that these comments relating to the rotation of a single triaxial MR sensor also apply to their use in a pair, i.e. it is always possible to relate the output of two ideal MR sensors via a rotation matrix. If the pair is positioned such that the field incident at each of their locations is identical, then the field may be time varying without disturbing the function of the sensors since the magnitude of the two observation vectors will be the same.

Define the field that would be measured by an MR sensor aligned with J as $\mathbf{f}_J$. Since the system's motion does not alter the field, the sensors located in U and L simply yield an output that is related to $\mathbf{f}_J$ by the relevant rotation matrix:

$$\mathbf{f}_J = \mathbf{A}\mathbf{f}_U \tag{4.58(a)}$$

$$\mathbf{f}_J = \mathbf{B}\mathbf{f}_L \tag{4.58(b)}$$

or

$$\mathbf{f}_U = \mathbf{A}^{\mathrm{T}}\mathbf{f}_J \tag{4.59(a)}$$

$$\mathbf{f}_L = \mathbf{B}^{\mathrm{T}}\mathbf{f}_J. \tag{4.59(b)}$$

For the estimation of the relative orientation of U and L, we require vector observations from each of these bodies related by a relative rotation matrix. The relative rotation matrix is easily found by combining (4.58) and (4.59):

$$\mathbf{f}_U = \mathbf{A}^{\mathrm{T}}\mathbf{B}\mathbf{f}_L = \mathbf{C}\mathbf{f}_L \tag{4.60(a)}$$

$$\mathbf{f}_L = \mathbf{B}^{\mathrm{T}}\mathbf{A}\mathbf{f}_U = \mathbf{C}^{\mathrm{T}}\mathbf{f}_U \tag{4.60(b)}$$

This section has shown that MR sensors are theoretically very useful for measuring the relative rotation of two bodies. This is since they effectively 'filter out' any translational aspects of the motion and derivatives of the angular motion are inconsequential; only the angular orientation of the body affects the output of the MR sensors. Unfortunately, the situation is not so simple for accelerometers, as the next section shows.

4.8.3 Theoretical Output of the Accelerometer Under General Motion

The triaxial accelerometers are considered as ideal sensors of acceleration, where measured acceleration is due to both gravity and motion of the sensor itself. Because the 'field' measured by the accelerometers is affected by motion, the analysis is considerably more complex than that for the MR sensors. Actually, this section shows that the general outputs of two triaxial sensors mounted on two bodies undergoing general motion *cannot* be related via a rotation matrix. It is also shown, however, that under certain conditions they can be approximately related.

The first step is to find the velocities and accelerations of points x and y as measured in F. Since x (y) is fixed within U (L), the velocities and accelerations of these points with respect to their associated frames is zero:

$$\dot{\mathbf{x}}_U = \ddot{\mathbf{x}}_U = \dot{\mathbf{y}}_L = \ddot{\mathbf{y}}_L = 0 \tag{4.61}$$

therefore

$$\dot{\mathbf{x}}_J = \dot{\mathbf{A}}\,\mathbf{x}_U \tag{4.62}$$

$$\ddot{\mathbf{x}}_J = \ddot{\mathbf{A}}\,\mathbf{x}_U \tag{4.63}$$

with similar equations for the lower body, L. Using $\mathbf{x}_U = \mathbf{A}^{\mathrm{T}}\mathbf{x}_J$, (4.62) gives

$$\dot{\mathbf{x}}_J = \dot{\mathbf{A}}\,\mathbf{A}^{\mathrm{T}}\mathbf{x}_J \tag{4.64}$$

Consider the derivative of the expression of orthogonality: $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{1}$,

$$\dot{\mathbf{A}}\,\mathbf{A}^T + \mathbf{A}\,\dot{\mathbf{A}}^{\mathrm{T}} = \mathbf{0}$$

or
$$\dot{\mathbf{A}}\,\mathbf{A}^{\mathrm{T}} = -\mathbf{A}\,\dot{\mathbf{A}}^{\mathrm{T}} = -\left(\dot{\mathbf{A}}\,\mathbf{A}^{\mathrm{T}}\right)^{\mathrm{T}} \tag{4.65}$$

which shows that the matrix $\dot{\mathbf{A}}\mathbf{A}^T$ is skew-symmetric. Denote this matrix, the *angular velocity matrix*, by Ω, and its time derivate by Ξ (which is skew-symmetric by definition).

Now

$$\dot{\mathbf{x}}_J = \Omega\,\mathbf{x}_J \tag{4.66}$$

and

$$\ddot{\mathbf{x}}_J = \dot{\mathbf{\Omega}}\mathbf{x}_J + \mathbf{\Omega}\dot{\mathbf{x}}_J = \dot{\mathbf{\Omega}}\mathbf{x}_J + \mathbf{\Omega}^2\mathbf{x}_J = \left[\mathbf{\Xi} + \mathbf{\Omega}^2\right]\mathbf{x}_J \tag{4.67}$$

or

$$\ddot{\mathbf{x}}_J = \left[\mathbf{\Xi} + \mathbf{\Omega}^2\right]\mathbf{A}\mathbf{x}_U \tag{4.68}$$

(Note: strictly the matrices Ω and Ξ should be Ω_U and Ξ_U but since only the case of the upper body is considered here, the subscripts are omitted for simplicity.) Equating the expressions (4.63) and (4.68) for the acceleration of the point *x*, in the fixed frame it is seen that

$$\ddot{\mathbf{A}} = (\mathbf{\Xi} + \mathbf{\Omega}^2)\mathbf{A} \tag{4.69}$$

This matrix relates the coordinate matrix of *x*, in *U*, to the acceleration due to the rotation of *U* in *J*. Since

$$\mathbf{x}_F = \mathbf{x}_J + \mathbf{p} \tag{4.70}$$

where **p** is defined above, we have

$$\ddot{\mathbf{x}}_F = \ddot{\mathbf{x}}_J + \ddot{\mathbf{p}} = \ddot{\mathbf{A}}\mathbf{x}_U + \ddot{\mathbf{p}} \tag{4.71}$$

This is the acceleration of *x* due to the rotation of *U* and the translation of *J*. Two changes need to be made before this quantity will represent the output of an accelerometer mounted in *U* at *x*. Firstly, note that accelerometers register the acceleration not just due to motion, but that due to gravity as well. Making this correction, the vector representing the sum of the accelerations is

$$\mathbf{a}_F = \ddot{\mathbf{x}}_F + \mathbf{g} = \ddot{\mathbf{A}}\mathbf{x}_U + \ddot{\mathbf{p}} + \mathbf{g} \tag{4.72}$$

Secondly, the calculated expression is 'measured' in *F*, whereas the accelerometer is mounted in *U*, with axes that are in general non-parallel with *F*. To account for this, the vector of acceleration measured in *F* is rotated to *U*. Since **A** is the rotation matrix that relates $U \rightarrow J$ and the axes of J and F are parallel, the output of the accelerometer at x is given by

$$\mathbf{acc}_U = \mathbf{A}^T\,(\ddot{\mathbf{A}}\mathbf{x}_U + \ddot{\mathbf{p}} + \mathbf{g}) \qquad (4.73)$$

and following exactly the same derivation for the accelerometer in L gives

$$\mathbf{acc}_L = \mathbf{B}^T\,(\ddot{\mathbf{B}}\mathbf{x}_L + \ddot{\mathbf{p}} + \mathbf{g}) \qquad (4.74)$$

Following the approach of the MR section, it is desired that the outputs of the accelerometers be related via a relative rotation matrix. It can be seen that this is actually impossible unless the first quantities in the brackets of the two expressions are equal. The following development, however, shows that in cases where reasonable limits are put on the magnitude of angular acceleration and velocity and the accelerometer is placed close to the joint, the first term in the brackets (henceforth known as the rotational term) is negligible compared to the second two. Alternatively, mounting arrangements, requiring additional sensors, are discussed that allow for general motions.

In the following $\|\,.\,\|$ denotes the Euclidean (2-norm) or the associated induced matrix norm. The following properties of norms are used:

1. $\|\mathbf{AB}\| \le \|\mathbf{A}\|\,\|\mathbf{B}\|$
2. $\|\mathbf{A} + \mathbf{B}\| \le \|\mathbf{A}\| + \|\mathbf{B}\|$ & $\|\mathbf{x} + \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$
3. $\|\mathbf{Ax}\| \le \|\mathbf{A}\|\,\|\mathbf{x}\|$

where **A** and **B** are matrices and **x** and **y** are vectors. Also used is the property of rotation matrices (and orthogonal matrices in general) that their induced 2-norm is equal to one. This has a physical interpretation in that rotating a vector does not change its magnitude.

Since **A** is a rotation matrix,

$$\|\mathbf{acc}_U\| = \|\mathbf{A}^{\mathrm{T}}(\ddot{\mathbf{A}}\mathbf{x}_U + \ddot{\mathbf{p}} + \mathbf{g})\| = \|\ddot{\mathbf{A}}\mathbf{x}_U + \ddot{\mathbf{p}} + \mathbf{g}\| \le \|\ddot{\mathbf{A}}\mathbf{x}_U\| + \|\ddot{\mathbf{p}} + \mathbf{g}\| \qquad (4.75)$$

denote the 'signal' $\mathbf{s} = \ddot{\mathbf{p}} + \mathbf{g}$. Consider first the rotational term of (4.75):

$$\| \ddot{\mathbf{A}}\mathbf{x}_U \| = \|(\Xi + \boldsymbol{\Omega}^2)\mathbf{A}\mathbf{x}_U\| \leq \|(\Xi + \boldsymbol{\Omega}^2)\| \, \|\mathbf{A}\mathbf{x}_U\| = \|(\Xi + \boldsymbol{\Omega}^2)\| \, \|\mathbf{x}_U\| \quad (4.76)$$

and

$$\|(\Xi + \boldsymbol{\Omega}^2)\| \leq \| \Xi \| + \| \Omega^2 \| \quad (4.77)$$

It was stated earlier that the matrices Ω and Ξ are associated with vectors ω and α. The structure of the matrices is as follows

$$\boldsymbol{\Omega} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}$$
$$\Xi = \begin{bmatrix} 0 & -\alpha_3 & \alpha_2 \\ \alpha_3 & 0 & -\alpha_1 \\ -\alpha_2 & \alpha_1 & 0 \end{bmatrix} \quad (4.78)$$

where $\omega = [\omega_1 \; \omega_2 \; \omega_3]^T$ is the angular velocity vector and $\alpha = [\alpha_1 \; \alpha_2 \; \alpha_3]^T$ is the angular acceleration (rate of change of angular velocity) vector. These matrices have the property that their product with a vector is equal to the cross product of the associated vector, e.g.

$$\Omega \, \mathbf{y} = \omega \times \mathbf{y} \quad (4.79)$$

Since, using the properties of the cross-product, the maximum norm that this vector can assume is

$$\|\Omega \, \mathbf{y}\|_{max} = \|\omega \times \mathbf{y}\|_{max} = \|\omega\| \, \|\mathbf{y}\| \quad (4.80)$$

when $\|\mathbf{y}\| = 1$, this maximum takes the value of the induced norm, and it is seen that

$$\| \Omega \| = \|\omega\| \quad (4.81)$$

and similarly $\|\Xi\| = \|\alpha\|$. A similar argument can be used to show that $\|\Omega^2\| = \|\omega\|^2$. This gives

$$\|(\Xi+\mathbf{\Omega}^2)\| \leq \|\alpha\| + \|\omega\|^2 \tag{4.82}$$

and the (4.76) can be rewritten

$$\|\ddot{\mathbf{A}}\mathbf{x}_U\| \leq \|(\Xi+\mathbf{\Omega}^2)\|\,\|\mathbf{x}_U\| \leq (\|\alpha\| + \|\omega\|^2)\,\|\mathbf{x}_U\| \tag{4.83}$$

This gives an upper bound to the magnitude of the accelerometer signal that is due to the rotational velocity and acceleration of the body. Expected bounds on each of the quantities can be reasonably estimated through consideration of the motions that are to be monitored. The relative sizes of the terms due to **s** and the rotational term may then be discerned by assuming equality in the initial relationship:

$$\|\mathbf{acc}_U\| = \|\ddot{\mathbf{A}}\mathbf{x}_U\| + \|\mathbf{s}\|. \tag{4.84}$$

Recall **s** is the vector sum of the acceleration due to gravity and the linear acceleration of the joint. Numerical values are now associated with these quantities. The following assumptions are made for the expected smooth motion:

$\|\alpha\| < 1$ rad/s^2

$\|\omega\| < 2$ rad/s

$\|\mathbf{s}\| > 7$ m/s^2

The other quantity to be estimated is $\|\mathbf{x}_U\|$, the Euclidean norm of the vector from the accelerometer to the joint about which rotation is occurring. The magnitude of the rotational term therefore becomes a linear function of the distance of the accelerometer from the joint:

$$\|\ddot{\mathbf{A}}\mathbf{x}_U\| \leq 5\|\mathbf{x}_U\| \tag{4.85}$$

It is of course of interest that this quantity be made as small as possible, i.e. the accelerometer should be very close to the joint, but this conflicts with the case in which more than two bodies are being monitored. In this truly multi-body application, it would be preferable to have one sensor on each body, rather than one at each ‘end’ of the body, close to the joints. The obvious choice in this situation is to place the sensors midway between the two joints. Approximating the maximum length of a human limb segment as 400mm, the distance to the joint is half this value, and the magnitude of the rotational term is unity. This is unacceptably large when it is considered that the signal term could reasonably be expected to be as small as 7 (when the joint is accelerating directly upwards at 2m/s^2). Central placement of a single accelerometer/MR is therefore not recommended for multi-body instrumentation. For multi-body instrumentation, an obvious solution is to place a cluster at either end of the limb segment, but this introduces a large number of channels. Some alternative methods of reducing, or potentially eliminating the effects of the rotational term are now discussed.

While the previous discussion details the minimisation of rotational effects on the accelerometer through physical placement, it may also be possible to apply a sophisticated signal processing technique to do the same job and therefore reduce the required number of sensors. Consider, for example, a 'magnitude filter'. The magnitude of the output of each of the accelerometers is a function of the linear acceleration of the joint and the rotational term associated with the body on which the accelerometer is mounted. This being the case, in the frequency domain the spectra of the signals from the two accelerometers may share some similar characteristics (due to the signal term) and some different (due to the rotational term). It may be possible to use the similarity in the frequency domain to eliminate the effects of the rotational term. The exception to this would be when the spectra from each effect occupy the same region of the frequency domain.

What might be the most reliable method of elimination of the rotational term is a combination of very basic signal processing and the use of two rigidly linked accelerometers. Consider a second triaxial accelerometer, placed on a rigid extension from the accelerometer/MR sensor cluster. The output of the additional accelerometer ($\mathbf{acc}_2$) will be the same as that of the first ($\mathbf{acc}_1$) except for the rotational term. If the vectors from the joint to the accelerometers are $\mathbf{l}_1$ and $\mathbf{l}_2$, then ideally

$$\mathbf{acc}_1 - \mathbf{acc}_2 = \mathbf{A}^T \ddot{\mathbf{A}} (\mathbf{l}_1 - \mathbf{l}_2) = \mathbf{A}^T \ddot{\mathbf{A}}\, \Delta \qquad (4.86)$$

where Δ is the vector between the two accelerometers. If the vectors $\mathbf{l}_1$ and $\mathbf{l}_2$ are collinear (or near to it), the ratio, $\|\mathbf{l}_1\|/\|\Delta\|$ can be used to modify a combination of the original accelerometer output and the vector difference of the outputs so that the new vector is free of the rotational term:

$$\mathbf{acc}^* = \mathbf{acc}_1 - (\mathbf{acc}_1 - \mathbf{acc}_2)\|\mathbf{l}_1\|/\|\Delta\| \qquad (4.87)$$

This is since $(\mathbf{acc}_1 - \mathbf{acc}_2)/\|\Delta\|$, the normalised vector difference, gives the rotational term that would result by placing an accelerometer 1 unit from the joint in the direction of $\mathbf{l}_1$ and $\mathbf{l}_2$. Modifications need to be made if $\mathbf{l}_1$ and $\mathbf{l}_2$ (and hence Δ) are not collinear, although it is probably accurate enough to assume that this is the case.

Regardless of the method in which the rotational term is made small, the result is that:

$$\mathbf{acc}_U \approx \mathbf{A}^T(\ddot{\mathbf{p}} + \mathbf{g}) \qquad \mathbf{acc}_L \approx \mathbf{B}^T(\ddot{\mathbf{p}} + \mathbf{g}) \qquad (4.88)$$

so that $\mathbf{acc}_U \approx \mathbf{A}^T\mathbf{B}\,\mathbf{acc}_L = \mathbf{C}\mathbf{acc}_L$, giving the accelerometer equivalent of (4.60(a)).

This section has shown that under certain conditions, an accelerometer may provide the additional observations needed to estimate the orientation of U w.r.t. L. The next section details the exact way in which these observations are used. The basic ideas of this section are used in the design of 'weighting matrices' for more accurate orientation estimation.

4.9 Orientation Estimation Revisited

Recall that in Section 4.7, a method for the weighted least squares estimation of the Rodrigues' vector was derived (4.50):

$$\mathbf{b} = (\Sigma_C{}^T\mathbf{W}\Sigma_C)^{-1}\Sigma_C{}^T\mathbf{W}\Delta_C. \tag{4.89}$$

The matrix Σ_C (6x3) contains the sum of the vector outputs of the MR sensors and accelerometers (3x3 matrices stacked on top of each other), and the vector Δ_C (6x1) the differences. Before the data is summed and differenced it is normalised so that it is actually possible that a rotation matrix can relate the measurements. Clearly if vectors measured by the same type of instrument are of different magnitude, no rotation will bring the two into agreement. This normalisation removes any gain discrepancy between sensors and also the dynamic effects discussed for the accelerometer in the previous section.

As documented above, it is likely that the accelerometer measurements will have what can be considered an 'error signal' due to the angular velocity and acceleration of the body on which they are mounted. The data from the two sorts of sensor should be weighted so that the lower error level information of the MR sensors is 'believed' more than that of the accelerometers in the determination of orientation.

The accuracy of the accelerometer outputs is a function of the motion that the accelerometers are undergoing, thus the optimal (in a very loose sense of the word) weighting matrix is non-constant. A method to dynamically produce a sensible diagonal weighting matrix **W** is now described.

Recall that the exact expression for the two vector outputs was given by (4.73), (4.74):

$$\mathbf{acc}_U = \mathbf{A}^T(\ddot{\mathbf{A}}\mathbf{x}_U + \ddot{\mathbf{p}} + \mathbf{g}) \tag{4.90(a)}$$

$$\mathbf{acc}_L = \mathbf{B}^T(\ddot{\mathbf{B}}\mathbf{x}_L + \ddot{\mathbf{p}} + \mathbf{g}) \tag{4.90(b)}$$

Since the magnitudes of the second (signal) terms are the same, the only reason for the accelerometers to give outputs of different magnitudes is due to the differences of the rotational terms. If the accelerometers give signals of vastly different magnitude then it is safe to assume that the first term in one of the expressions is non-negligible. In the case where the magnitudes are practically the same there are two possibilities

- the effect of the rotational terms is negligible
- the direction and magnitude of the first terms are such that the signals are of similar magnitude

The first possibility is more likely to occur, requiring only low-level angular accelerations and velocities. Based on this reasoning, the greater the discrepancy between the magnitudes of the two accelerometer signals, the more significant the rotational term and the lower the weighting to be attributed to the accelerometer data. The MR data is constantly weighted, as the only error assumed to present in the output of these sensors is noise. Thus, the chosen weighting (6x6) matrix is block-diagonal

$$\mathbf{W} = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{W}_{acc} \end{bmatrix} \qquad (4.91)$$

with the identity matrix corresponding to the MR data. $\mathbf{W}_{acc}$ is defined

$$\mathbf{W}_{acc} = (c| \, \|\mathbf{acc}_U\| - \|\mathbf{acc}_L\| \, |)^{-1}\mathbf{I}. \qquad (4.92)$$

where c is a positive constant, discussed below. Note that the greater the difference between the magnitudes of the two acceleration measurements the lower the weighting assigned to the accelerometer data. In the case where $c| \, \|\mathbf{acc}_U\| - \|\mathbf{acc}_L\| \, | < 1$, this will result in 'more notice' being taken of the accelerometer data than the magnetoresistive sensor data. Unless the MR data has been adversely affected by localized magnetic fields, there is no reason for a higher weighting to be placed on the accelerometer data, since the accelerometer data is more 'approximate' due to rotational effects. The high weighting of accelerometer data is prevented by checking the magnitude of the difference of the vector norms, i.e. if $c| \|\mathbf{acc}_U\| - \|\mathbf{acc}_L\| | < 1$, then $\mathbf{W} = 1$, otherwise it is as previously defined. The constant, c, is chosen by considering the magnitude of the vector norm difference after which, the data can be considered to be practically unaffected by the rotational terms. For example, if it were considered that $| \|\mathbf{acc}_U\| - \|\mathbf{acc}_L\| | < .05$ were a suitable level, then $c = 20$. For any vector norm difference with magnitude greater than .05, the weighting is given by $c| \|\mathbf{acc}_U\| - \|\mathbf{acc}_L\| |$, for a difference less than 0.05, the weighting is unity.

At this point, it is probably beneficial to summarise the method by which orientation is to be estimated. The steps are:

1. Obtain $\mathbf{f}_L$, $\mathbf{f}_U$, $\mathbf{acc}_U$, $\mathbf{acc}_L$.
2. Calculate norms of each of the quantities
3. Calculate $w_{\text{acc}} = 1/(c| \|\mathbf{acc}_U\| - \|\mathbf{acc}_L\| |)$
4. Create weighting matrix, based on $w_{\text{acc}}<1$ or $w_{\text{acc}}\geq 1$
5. Normalise all vector quantities by dividing by the associated norm
6. Form vector Δ_C and matrix Σ_C using normalised data
7. Solve for $\mathbf{b}$ using weighted least squares

The exact form of the vector Δ_C is

$$\Delta_C = \begin{bmatrix} \mathbf{f}_U - \mathbf{f}_L \\ \mathbf{acc}_U - \mathbf{acc}_L \end{bmatrix} \tag{4.93}$$

i.e. the MR data is stacked on top of the accelerometer data, and the rotation matrix associated with the obtained Rodrigues' vector is the relative rotation matrix $\mathbf{C}$, where:

$$\mathbf{f}_U = \mathbf{C}\mathbf{f}_L \text{ and } \mathbf{acc}_U = \mathbf{C}\,\mathbf{acc}_L \tag{4.94}$$

Having described now, the theoretical aspects of the use of outputs of accelerometers and MR sensors in the least squares estimation of orientation, the next section discusses the calibration of the sensors.

4.10 Sensor Calibration

This section contains two methods of calibration for the sensors. The first is simple, and assumes that the axes of the sensors are orthogonal, and that the sensors are mounted orthogonally within their enclosure. This is quite a large assumption to make when all assembly, as previously mentioned, was done 'by eye' with no jig to ensure accurate mounting. This method of calibration was applied while the author was able to access the sensors and had very little time to gather results. After the collected results were found to be poor, the second, more comprehensive method was derived, while at a distance of many thousands of kilometres from the sensors. Thus, the second method remains unproved. Both are included so that it may be seen where the original method went wrong, and also so that future workers may apply the second method.

Before describing the methods of calibration, it is perhaps useful that the purpose of calibration is properly defined. In this case, the purpose is to make the physical sensors behave as much as possible like their mathematical idealizations. The axes of both sensors are modelled as operators that will yield the scalar product of the associated field vector with a unit vector collinear with the physical axes. Deviations from this ideality include

1. Constant offsets: output from the axes when either no field is present, or the field is exactly orthogonal to the axis.
2. Non-linear effects: hysteresis and general deviation of the output of the sensor from the scalar product model.

The first possibility is relatively easy to account for, while the correction of the non-linearities of individual axes would be a very time consuming task to undertake. While there will always be some degree of non-linearity in the sensor's characteristics, it is hoped/trusted that this will be a negligible component.

In addition to making the individual axes behave as scalar product operators, certain requirements are placed upon the collection of axes as a whole. These are

1. Uniform gain among axes of the same sensor type: Once the offsets are accounted for, the response of two axes of a similar sensor that are collinear and of the same polarity should be equal.
2. Orthogonality: The axes should form a right-handed orthogonal set.

These two requirements combined ensure that the field that is being measured by the sensor is uniquely and accurately portrayed. The orthogonality of the axes also has important consequences that are fully discussed below.

It should be noted that when talking of calibration purely through use of the ambient fields (magnetic or gravitational), there is no need to make explicit reference to the type of sensor that is being calibrated. While most people know more about the gravitational field than they do about the magnetic field, and may sometimes use this knowledge to simplify calibration procedures, e.g. the '2g test' when an accelerometer's axis is oriented so that it is positive upward and then downward, this is potentially dangerous. If the method of finding vertical is not reliable, or the axis is not accurately aligned with the enclosure, errors will result. It is better to make no assumptions about the fields that are being measured and make the calibration routine general. Further, since both the accelerometers and MR sensors are mounted within the same enclosure, it saves time if both sensor types are calibrated using the same technique.

Finally, in this calibration it is not necessary that the voltage output of the sensors be converted to a physically meaningful number with associated units. It is required only that the equivalent sensors in the same orientation yield the same results. Having discussed the purpose of calibration, the first, crude method of calibration is now presented.

4.10.1 Calibration Assuming Orthogonality of Sensor Axes

This first method of calibration makes three assumptions about the sensor axes. The first is the orthogonality of the axes within a single sensor; the second that the axes of the two sensors within an enclosure have parallel axes; and finally that the axes of the sensors are parallel with a coordinate frame defined by the edges of the enclosure.

Based on these assumptions, when the enclosure is rotated 180° about any of the axes of its associated frame, the outputs of all the sensor axes, except the two (one from each type of sensor) parallel to the axis of rotation, should have their outputs remain constant in magnitude, while they change polarity. This is the basis of the method of calibration, which is now described below for a general sensor.

Consider the incident field, **f**, to be arbitrarily oriented with respect to the X-axis (in orientation $\mathbf{x}(1)$) of a sensor. The field may be decomposed into two components, that orthogonal to $\mathbf{x}(1)$ ($\mathbf{f}_{\mathrm{orth}}$) and the projection of **f** onto $\mathbf{x}(1)$ ($\mathbf{f}_{\mathrm{proj}}$). The projection term, which is calculated using the scalar product $\mathbf{f}.\mathbf{x}(1)$, is the output of the sensor. Rotating $\mathbf{x}(1)$ 180° about any axis orthogonal to itself gives $\mathbf{x}(2) = -\mathbf{x}(1)$. The projection term is now given by $\mathbf{f}.\mathbf{x}(2) = -\mathbf{f}.\mathbf{x}(1)$, i.e. rotating the axis through 180° gives two values equal in magnitude and opposite in sign. This is shown diagrammatically in Figure 4.14.

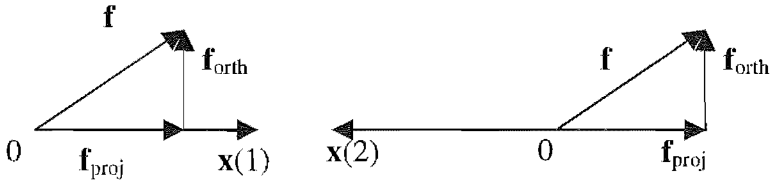


Figure 4.14 Rotating an axis through 180° gives values of equal magnitude and opposite sign.

This intuitive fact treats the sensors as ideal mathematical objects. In reality the axes of each sensor has an associated offset so that even when **f** is orthogonal to an axis, the voltage output is non-zero. This offset must be accounted for if the outputs of the sensors are to be related by a rotation matrix. This correction may be achieved by either subtracting the offset from the output during processing or altering the offset using hardware. The former option was undertaken as it was considered to be both faster and more accurate. (The offsets of the MR sensors can be altered using three trim potentiometers attached to the sensor boards.) If the offsets are to be accounted for, they must first be determined. This is the first part of the calibration process.

As shown above, taking two readings from a single axis before and after a 180° rotation should, in the absence of an offset, give two values symmetrical dispersed about zero, thus the offset can be calculated as the average of the two measured outputs.

The characteristic magnitude of an axis may be defined as the voltage output (measured from the offset) due to a unit field collinear with the axis. For the sensors' outputs to be relatable via a rotation matrix, each axis must have the same characteristic magnitude, or the outputs must be modified so that the effects of the individual characteristic magnitudes are accounted for. While it would be difficult to calculate the characteristic magnitude as it is defined, it is a simpler task to ensure that each axis has an identical characteristic magnitude. All the latter requires is that the processed output of all axes are similar when the axes are subject to a similar field. Consider, for example, taking the output of the X-axis of a sensor, and then rotating the sensor so that the Y-axis has the orientation that X used to occupy. Ensuring similar characteristic magnitudes simply means that the deviations of the measured outputs of the X and Y-axes from the known offsets are identical. In this work the similarity of the axes' response was achieved by normalising all the axes, i.e. finding the magnitudes of the responses to a standard field and then dividing the axes' outputs by these magnitudes. Since the magnitude can be calibrated using the same data as was required for the offset determination, calibration can theoretically be completed using six tests (two for each axis).

There are potential problems with using only two tests that may be reduced by using more tests. Firstly the direction of the axes might be such that reversing the orientation causes only a small change in output. This would occur when the field is close to being orthogonal to the chosen direction. The smaller the actual change in the magnetic field, the greater the effect of measurement error in the determination of offset and magnitude normalisation. Also, when using only two tests, the exact orientation of the sensor is of high importance. If more tests are used, errors in orientation (assuming a symmetrical spread about the mean) tend to 'average out'. A preferable method of calibration uses six tests for each of the axes of the sensor. The six tests correspond to three orthogonal axes in both directions (as shown in Figure 4.15).

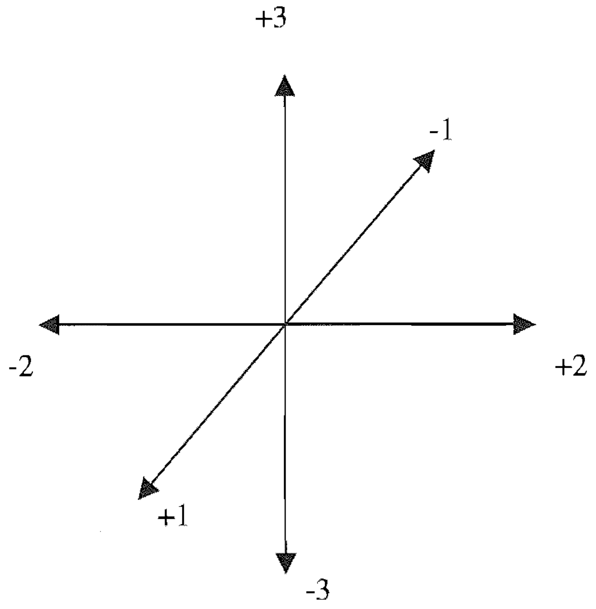


Figure 4.15 The six orientations of each axis used in calibration of accelerometers and MR sensors.

When using six tests, the offset is given simply as the average of the values of each output. This is because each pair of tests has outputs that should be symmetrical about the offset. Using this three-dimensional approach allows for a more thorough definition of the magnitude of the sensors response to the incident field. The magnitude of the incident field, as measured by an axis is given by the Euclidean norm of the output (measured from the offset) in three orthogonal directions. Rather than using the deviations from the offset in three directions, e.g. +1, +2 and +3, the following expression is used

$$\mathrm{Mag} = \tfrac{1}{2}\{([+1]-[-1])^2 + ([+2]-[-2])^2 + ([+3]-[-3])^2\}^{1/2} \qquad (4.95)$$

where [+1] indicates the value measured when the axis is collinear with direction +1 etc.

As well as an offset and a characteristic magnitude, each axis has a polarity that must be correctly determined if the axes of the sensor are to form a right-hand coordinate system. This polarity can be due either to the orientation of the chip on the board, or the wiring of the sensor. While the definition of a positive axis is arbitrary, all axes must be treated similarly. A simple method using only two of the six directions is to arbitrarily say that [+1] > [-1], i.e. the output when the sensor is in the +1 direction should be greater than that when it is in the opposite direction. If the recorded data for a particular channel gives [+1] < [-1], then the previously calculated normalising factor is made negative. If this method is used, it should be ensured that the chosen direction entails a large change in voltage when reversed. The minimum number of tests that can be used so that each axis of a three-axis set occupies all six positions was found to be nine. This minimum was used in the interests of saving time.

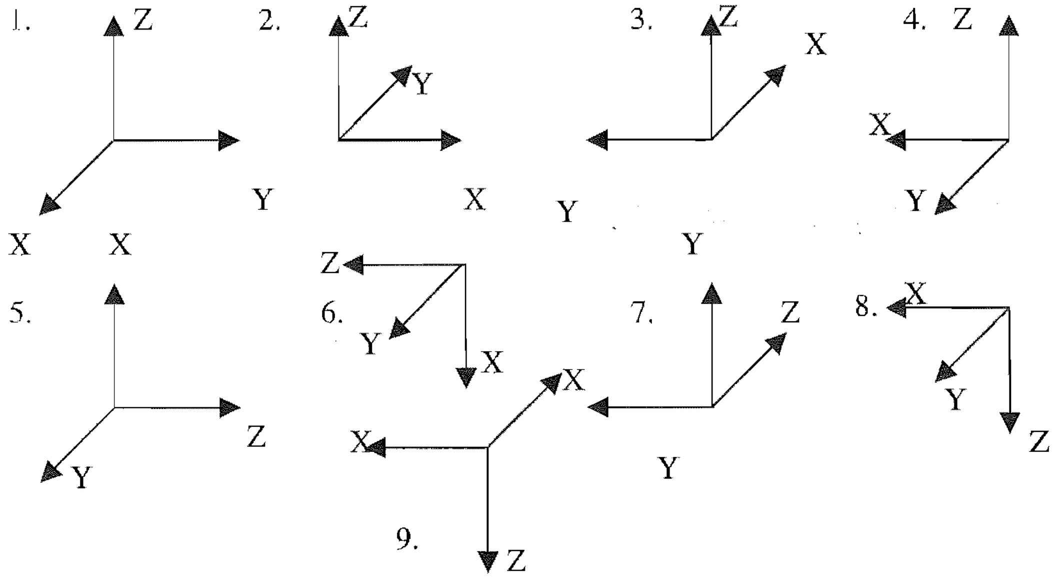


Figure 4.16 The nine orientations of the enclosure frame during calibration

As previously mentioned, this method of determining axis offset and magnitude demands orthogonality of the sensor axes. The actual calibration implicitly requires this as it uses the fact that rotating the enclosure about one of its axes is equivalent to rotating the sensors about one of their own axes. Since this is assumed for all three axes, each of the axes must be parallel with the associated axis of the enclosure frame. This applies for both sensors within the same enclosure. If the axes are not orthogonal in this way, then, assuming the X-axis of a sensor has components $[a, b, c]$ in the enclosure frame, rotating about the Z-axis of the enclosure frame gives the axis new components of $[-a, -b, c]$ in a fixed frame with which the enclosure frame was previously aligned. In the six orientations [1]-[6] for which the output of the X-axis is recorded, the components of the axis in the fixed frame are:

[1]: $[a, b, c]$
[2]: $[-b, a, c]$
[3]: $[-a, -b, c]$
[4]: $[b, -a, c]$
[5]: $[b, c, a]$
[6]: $[b, -c, -a]$

In each of these configurations, the output is of the form

$$m_x f_x + m_y f_y + m_z f_z + \Delta \tag{4.96}$$

where $[m_x, m_y, m_z]$ are the components of the vector representation of the axis in the fixed frame, and $\mathbf{f} = [f_x, f_y, f_z]$ is the field in the fixed frame. Summing the six outputs gives:

$$(a - b - a + b + b + b)f_x + (b + a - b - a + c - c)f_y + (c + c + c + c + a - a)f_z + 6\Delta$$
$$= 2bf_x + 4cf_z + 6\Delta \tag{4.97}$$

Whereas under the assumption of orthogonality the sum gives 6Δ. Similarly erroneous results are generated by applying the previously described method for finding the characteristic magnitude of the axis. Clearly, the problems caused by non-orthogonal axes are many. Firstly it leads to errors in the calculation of axis offsets and characteristic magnitudes. During the actual operation of the sensor, non-orthogonality also causes problems, even if the offsets and axis gains are correctly determined. Two of the problems are to do with consistency within and between sensor clusters. Even if the axes of the accelerometer and MR sensor are orthogonal themselves, if the two types of sensor are arbitrarily oriented with respect to each other within the enclosure, the vector observations are being made from different bases and hence cannot be used in the proposed method of estimating orientation. If the accelerometer and MR sensor axes of each cluster are parallel, but arbitrarily oriented within the enclosure, then a simpler type of error results. The discrepancy caused by this error in orientation is theoretically a constant rotation matrix that relates the orientation of one cluster to the other (in the ideal case, this is the identity matrix).

The following method of calibration makes no assumptions about the orientations of the axes within the enclosure, yet theoretically yields data that would be obtained if the sensor axes were orthogonal and parallel to the enclosure frame.

4.10.2 Calibration Assuming Axes Non-Orthogonal

Consider three non-orthogonal axes, $\mathbf{m}_i$ i = 1,2,3 of a field sensor fixed in a cubic enclosure, which itself has an orthogonal coordinate frame [X_C Y_C Z_C] (the axes of which are parallel with the edges of the cube and forming a right-handed system). The enclosure frame is initially aligned with a fixed global coordinate system [X_F Y_F Z_F]. The field, **f**, has unknown components [f_x, f_y, f_z] in the fixed frame. Consider first the axis $\mathbf{m}_1$ which has unknown components [*a*, *b*, *c*] in the enclosure frame. The situation is shown in Figure 4.17

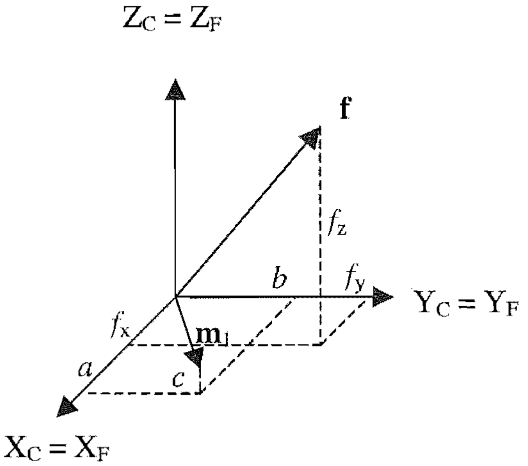


Figure 4.17 The orientation of $\mathbf{m}_1$ and f in the enclosure and fixed frames, which are shown as coincident.

Ideally, the axis would give an output proportional to the scalar product:

$$af_x + bf_y + cf_z, \tag{4.98}$$

However, there is an unknown offset on the output of the axis that must be determined, and the output is therefore given by

$$c_1 = af_x + bf_y + cf_z + \Delta, \tag{4.98}$$

where Δ is the unknown (assumed constant) offset. If the axis is rotated about Z_F by 180°, the axis now has components, as described in the previous section, [$-a$ $-b$ c] in the fixed frame, and the output in this configuration is therefore given by

$$c_2 = -af_x - bf_y + cf_z + \Delta \tag{4.99}$$

Similarly rotating the original configuration about the X_F and Y_F axes gives, respectively

$$c_3 = af_x - bf_y - cf_z + \Delta \tag{4.100}$$

$$c_4 = -af_x + bf_y - cf_z + \Delta \tag{4.101}$$

It is seen that by adding together the outputs collected with the axis in these four configurations, the result is simply 4Δ, i.e. averaging the recorded data gives Δ (or at least an approximation of Δ). There, is more information in these tests, however, than the value of Δ. Writing the scalar relations in the following matrix form:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} f_x a \\ f_y b \\ f_z c \\ \Delta \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} \tag{4.102}$$

$$\mathbf{Ad} = \mathbf{c} \tag{4.103}$$

Since $\mathbf{A}$ is non-singular it is possible to calculate $\mathbf{d}$, which contains information about the orientation of $\mathbf{m}_1$ in the enclosure frame. If the enclosure is reoriented in the fixed frame so that the components of the axis in that frame are $[c, a, b]$, which is achieved by making $X_C = Y_F$, $Y_C = Z_F$ and $Z_C = X_F$ and the same rotations are performed, then the matrix equation becomes

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} f_x c \\ f_y a \\ f_z b \\ \Delta \end{bmatrix} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} \tag{4.104}$$

Similarly orienting the enclosure so that $[b, c, a]$ are the coordinates of the axis in the fixed frame, i.e. $X_C = Z_F$, $Y_C = X_F$ and $Z_C = Y_F$, then the equations are

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} f_x b \\ f_y c \\ f_z a \\ \Delta \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix} \tag{4.105}$$

From these twelve tests, the results are three estimates of Δ, (of which the average is used), and the following products, written in matrix form:

$$\begin{bmatrix} af_x & af_y & af_z \\ bf_x & bf_y & bf_z \\ cf_x & cf_y & cf_z \end{bmatrix}. \tag{4.106}$$

Thus, three scalar multiples of the desired coordinate matrix are furnished, i.e. $f_x\mathbf{m}_1$, $f_y\mathbf{m}_1$ and $f_z\mathbf{m}_1$. Performing similar operations for the other two axes of the triaxial sensor, the results are estimates of the axis offsets and the scaled representations of the axes: $(f_x\mathbf{m}_2, f_y\mathbf{m}_2, f_z\mathbf{m}_2)$ and $(f_x\mathbf{m}_3, f_y\mathbf{m}_3, f_z\mathbf{m}_3)$.

Ideally the vectors represented by the coordinate matrices $f_x\mathbf{m}_i$, $f_y\mathbf{m}_i$, $f_z\mathbf{m}_i$, should be collinear, with lengths and sense dictated by f_x, f_y, f_z. Due to a large number of factors, including orientation errors (i.e. errors made during rotation of the enclosure) and sensor noise, the vectors will not be collinear. There are two pieces of information in the representation $\mathbf{m}_i = [a_i, b_i, c_i]$, the orientation of the axis with respect to the enclosure, and the gain of the sensor axis. Ideally, all axes would have the same gain, but this is not a safe assumption due to manufacturing variability. If the outputs of the sensor axes are to be related using a rotation matrix, the characteristic gain of each axis must be calculated and accounted for. While it is not possible to find the exact gain of each axis, without precisely knowing the field in which the sensor is located, it is easy to define the gain within a multiplicative constant, and if this is done consistently, this is sufficient. A method for finding the gain that uses all the collected information is now described. The sum of the norms of the vector outputs is seen to be:

$$||f_x\mathbf{m}_i|| + ||f_y\mathbf{m}_i|| + ||f_z\mathbf{m}_i|| = (|f_x| + |f_y| + |f_z|)||\mathbf{m}_i|| = F||\mathbf{m}_i|| \qquad (4.107)$$

The positive scalar, F, is common for all axes, therefore, if we assume that $||\mathbf{m}_1|| = g_1 = 1$, then

$$F||\mathbf{m}_1|| = F. \qquad (4.108)$$

Using this value for F, the characteristic gains of the other axes may be found by

$$\frac{\|f_x\mathbf{m}_i\| + \|f_y\mathbf{m}_i\| + \|f_z\mathbf{m}_i\|}{F} = \|\mathbf{m}_i\| = g_i \qquad (4.109)$$

The numerator is simply the sum of the norms of the three calculated scaled versions of the coordinate matrices of $\mathbf{m}_i$. It should be remembered that if two triaxial sensors are to be used in an application they must both be calibrated in the same magnetic field. If this is not the case, the value F may be different for the two sensors and hence the calculated gains will not be consistent between sensors.

Now that both the offsets and gains of each axis have been determined, the issue turns to the orthogonalisation of the sensor axes. The first step in this process is the normalization of the three coordinate matrices representing the orientation of the axes in the enclosure frame. The resulting unit vectors are denoted $\mathbf{u}_i$, i.e.

$$\mathbf{u}_i = (f_x\mathbf{m}_i + f_y\mathbf{m}_i + f_z\mathbf{m}_i)/||f_x\mathbf{m}_i + f_y\mathbf{m}_i + f_z\mathbf{m}_i|| \qquad (4.110)$$

The vectors $\mathbf{u}_i$ are in general not orthogonal. Orthogonality is required if the fields are to be able to be measured uniquely. Therefore, the next part of the process is the orthogonalisation of the vectors so that a linear combination of axis outputs can be used to give the same output as would be the case if the sensor axes were truly orthogonal, but still arbitrarily oriented within the enclosure. This orthogonalisation is achieved using the Gram-Schmidt process and transforms $\mathbf{u}_i$ to $\mathbf{n}_i$. First, one of the vectors is chosen as the reference, in this case $\mathbf{u}_1$. Next a unit vector, $\mathbf{n}_2$, is created by removing the component of $\mathbf{u}_2$ in the direction of $\mathbf{n}_1 = \mathbf{u}_1$ and then normalizing the resulting vector. Similarly, $\mathbf{n}_3$ is created by removing the components of $\mathbf{u}_3$ in the directions of $\mathbf{n}_1$ and $\mathbf{n}_2$ and then normalizing, i.e.

$$\mathbf{n}_1 = \mathbf{u}_1 \tag{4.111(a)}$$

$$\mathbf{n}_2 = (\mathbf{u}_2 - (\mathbf{u}_2.\mathbf{n}_1)\mathbf{n}_1)/\|\mathbf{u}_2 - (\mathbf{u}_2.\mathbf{u}_1)\mathbf{n}_1\| \tag{4.111(b)}$$

$$\mathbf{n}_3 = (\mathbf{u}_3 - (\mathbf{u}_3.\mathbf{n}_2)\mathbf{n}_2 - (\mathbf{u}_3.\mathbf{n}_1)\mathbf{n}_1)/\|\mathbf{u}_3 - (\mathbf{u}_3.\mathbf{n}_2)\mathbf{n}_2 - (\mathbf{u}_3.\mathbf{n}_1)\mathbf{n}_1\| \tag{4.111(c)}$$

At the conclusion of this process, $[\mathbf{n}_1\ \mathbf{n}_2\ \mathbf{n}_3]$ is an orthogonal set of unit vectors. While there is no guarantee that the set will be 'right-handed', this is likely since the physical axes have been arranged in this way and the Gram-Schmidt process simply removes the components of the vector representations that are non-orthogonal to the other vectors in the frame. Should it be found that the set is not right-handed, i.e. $\mathbf{n}_1 \times \mathbf{n}_2 = -\mathbf{n}_3$, then the components of $\mathbf{n}_3$ can be negated, without affecting the orthogonality of the set, and the frame will then be right-handed.

The matrix composed of the unit vectors, $\mathbf{n}_i$:

$$\mathbf{N} = [\mathbf{n}_1\ \mathbf{n}_2\ \mathbf{n}_3] \tag{4.112}$$

is the rotation matrix that expresses the orientation of the *orthogonalised sensor axes* in the enclosure frame. The final step in the process is to rotate this orthogonalised frame so that it coincides with the enclosure frame. This is required for a number of reasons. Firstly, two triaxial sensors are located within each enclosure, and it is required that the vector observations are referred to a common frame. It is also physically useful to have the enclosure frame as a visible reference to the true orientation of the sensors (or at least the transformed versions of them).

The derived rotation matrix, N, relates the coordinate matrices of vectors measured in the arbitrarily oriented orthogonal frame ($\mathbf{r}_a$) to the enclosure frame ($\mathbf{r}_e$) , i.e.

$$\mathbf{r}_e = \mathbf{N}\mathbf{r}_a \tag{4.113}$$

Hence, knowing the orientation of the orthogonalised sensor within the enclosure frame and the output that this sensor would have is equivalent to having an orthogonal sensor mounted collinear with the enclosure frame.

While the method of determination of **N** has been described, no mention has been made of the method of calculation of $\mathbf{r}_a$, the output that the arbitrarily oriented orthogonal sensor. The processing of the outputs of the actual sensor, which is now described, runs in parallel to that applied to the vector representations of the sensor axes.

After subtracting the calculated offsets from the axis outputs, each of the corrected values, denoted m_i i = 1,2,3, are divided by the calculated gains:

$$u_i = m_i/g_i \tag{4.114}$$

where u_i is the output associated with the arbitrarily oriented unit vector $\mathbf{u}_i$. This would give the outputs of all the sensor axes if they had the same gain, but were still arbitrarily oriented and non-orthogonal. The next step is the processing of u_i to produce the theoretical outputs of the orthogonal, but arbitrarily oriented sensor, the part of the processing associated with the Gram-Schmidt process and the unit vectors $\mathbf{n}_i$. Just as with the physical axes, the outputs of the theoretical axes is given by the scalar product of the vector representation with the field vector, i.e. $n_i = \mathbf{n}_i.\mathbf{f}$. Recalling the definitions of $\mathbf{n}_i$, this gives the following relations:

$$n_1 = \mathbf{n}_1.\mathbf{f} = \mathbf{u}_1.\mathbf{f} = u_1 \tag{4.115(a)}$$

$$\begin{aligned} n_2 = \mathbf{n}_2.\mathbf{f} &= (\mathbf{u}_2.\mathbf{f} - (\mathbf{u}_2.\mathbf{n}_1)\mathbf{n}_1.\mathbf{f})/\|\mathbf{u}_2 - (\mathbf{u}_2.\mathbf{n}_1)\mathbf{n}_1\| \\ &= (u_2 - (\mathbf{u}_2.\mathbf{n}_1)n_1)/\|\mathbf{u}_2 - (\mathbf{u}_2.\mathbf{n}_1)\mathbf{n}_1\| \end{aligned} \tag{4.115(b)}$$

$$\begin{aligned} n_3 = \mathbf{n}_3.\mathbf{f} &= (\mathbf{u}_3.\mathbf{f} - (\mathbf{u}_3.\mathbf{n}_2)\mathbf{n}_2.\mathbf{f} - (\mathbf{u}_3.\mathbf{n}_1)\mathbf{n}_1.\mathbf{f})/\|\mathbf{u}_3 - (\mathbf{u}_3.\mathbf{n}_2)\mathbf{n}_2 - (\mathbf{u}_3.\mathbf{n}_1)\mathbf{n}_1\| \\ &= (u_3 - (\mathbf{u}_3.\mathbf{n}_2)n_2 - (\mathbf{u}_3.\mathbf{n}_1)n_1)/\|\mathbf{u}_3 - (\mathbf{u}_3.\mathbf{n}_2)\mathbf{n}_2 - (\mathbf{u}_3.\mathbf{n}_1)\mathbf{n}_1\|. \end{aligned} \tag{4.115(c)}$$

While these calculations may look a little complex, they can be rewritten in the form

$$n_1 = u_1 \tag{4.116(a)}$$

$$n_2 = (u_2 - c_{21}n_1)/N_2 \tag{4.116(b)}$$

$$n_3 = (u_3 - c_{32}n_2 - c_{31}n_1)/N_3 \qquad (4.116(c))$$

where $c_{ij} = \mathbf{u}_i.\mathbf{n}_j$ and N_2 and N_3 (the norm terms associated with (4.115)) are constants. At this stage, the n_i are the outputs that would be obtained if an orthogonal sensor were arbitrarily oriented within the enclosure. The final step is to multiply the coordinate matrix $[n_1\ n_2\ n_3]^T$ by the rotation matrix **N**. This gives the output that would be obtained if three orthogonal axes of the same gain were oriented so that they were collinear with the orthogonal enclosure frame.

Use of this method of calibration should improve the results gathered by the sensor clusters. Once good results have been collected, the issue turns to methods by which they may be usefully presented (the rowing coach who has discussed Rodrigues' vectors with his athletes is rare!).

4.11 Data Presentation

While the actual output of the new orientation estimation procedure is the least-squares estimate of the Rodrigues' vector, the relative rotation matrix can be easily calculated (as previously shown). Still the rotation matrix is not an ideal method of presenting results to the rowing community, more processing is clearly required. Recall that the relative rotation matrix has its columns the orientation of the axes of one basis with respect to the other. In the situation where it is the motion of one object relative to another that is of interest, this means that by judicious placement of sensor clusters the presentation of results can be simplified. In the case where one cluster is fixed to the oar so that one axis, say the Y-axis, is collinear with the longitudinal axis of the oar and the second cluster is oriented so that when the oar is at a 'zero position' all axes of both clusters are parallel, then the second column of the relative rotation matrix gives the orientation of the axis of the oar-shaft with respect to the boat. The orientation of the longitudinal axis of the oar gives two of the three angles required to exactly specify the orientation of the oar. The third angle is the rotation of the oar about its own longitudinal axis, the feathering of the oar, which must be determined using either of the other columns of the rotation matrix.

Based on the structure of the calculated relative rotation matrix, the results could clearly be presented graphically. In addition to a line representing the Y-axis (collinear with the shaft of the oar), either the Z or X-axes would have to be added so that feathering could be observed. While this method of presentation would be aesthetically pleasing, it is unlikely that sufficient information could be gleaned by coaches/athletes. It is thought that numerical representation of the oar angles is a more suitable method of presentation, or at the least that three separate plots of the oar angles are generated.

There are a number of conventions used to express the orientation of an object using three angles, including many Euler angle variants, but to be useful to the rower/coach the angles must be physically meaningful. Such an angle system, based on that presented by Zatsiorsky and Yakunin [70], is used in this work. The angles measured are:

α: the swing angle created by projecting the loom of the oar onto the horizontal plane

γ: the angle of the oar with respect to the horizontal

β: the rotation of the oar about its axis (feathering)

The situation is shown in Figure 4.18, which is followed by descriptions of methods by which the three angles may be determined.

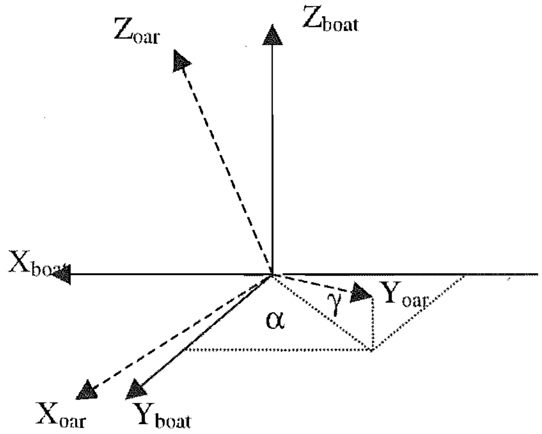


Figure 4.18 The orientation of the oar-mounted cluster with respect to the boat-mounted cluster, and the positive definitions of α and γ.

α

Since the Y-axis represents the shaft of the oar in the boat coordinate system, projecting it onto the horizontal plane is simply using the first two components of the second column of the relative rotation matrix. The angle between this projection and Y_{boat} is found using the dot product. Defining the unit vector in the direction of Y_{boat} as $\mathbf{y}_{boat}$ and the projection of the unit vector representing the oar onto the X_{boat}, Y_{boat} plane as $\mathbf{y}_{oar}^{x,y} = [y_{oar}(1)\ y_{oar}(2)\ 0]^T$ (i.e. the vector created by the X and Y components of $\mathbf{y}_{oar}$ and zero for the Z component)

$$\mathbf{y}_{boat} \cdot \mathbf{y}_{oar}^{x,y} = \|\mathbf{y}_{boat}\|\ \|\mathbf{y}_{oar}^{x,y}\| \cos\alpha$$

$$y_{oar}(2) = \|\mathbf{y}_{oar}^{x,y}\| \cos\alpha \qquad (4.117)$$

$$\cos\alpha = y_{oar}(2)/(y_{oar}(1)^2 + y_{oar}(2)^2)^{1/2} \qquad (4.118)$$

Since cosine is an even function, the sign of alpha must be determined by another method. Arbitrarily assigning α positive when Y_{oar} is forward of Y_{boat} as shown in Fig. 4.18, α may be fully defined by

$$|\alpha| = \mathrm{acos}((y_{oar}(2)/(y_{oar}(1)^2 + y_{oar}(2)^2)^{1/2}) \qquad (4.119)$$

$$\alpha = |\alpha| \text{ if } y_{oar}(1) < 0, \text{ else } \alpha = -|\alpha|$$

γ

The tilt angle is also calculated using the projection $\mathbf{y}_{oar}{}^{x,y}$. This time, however, the angle is found using the dot product of $\mathbf{y}_{oar}$ with $\mathbf{y}_{oar}{}^{x,y}$:

$$\mathbf{y}_{oar}\bullet\mathbf{y}_{oar}{}^{x,y} = \|\mathbf{y}_{oar}\|\ \|\mathbf{y}_{oar}{}^{x,y}\| \cos\gamma$$

$$\|\mathbf{y}_{oar}{}^{x,y}\|^2 = \|\mathbf{y}_{oar}{}^{x,y}\| \cos\gamma$$

$$\cos\gamma = \|\mathbf{y}_{oar}{}^{x,y}\| \qquad (4.120)$$

Again, the sign of the angle must be determined through consideration of the elements of $\mathbf{y}_{oar}$. In this case, if the Z component of $\mathbf{y}_{oar}$ is positive ($y_{oar}(3) > 0$), then the tilt is defined positive, i.e.

$$|\gamma| = \mathrm{acos}\|\mathbf{y}_{oar}{}^{x,y}\| \qquad (4.121)$$

$$\gamma = |\gamma| \text{ if } y_{oar}(3) > 0, \text{ else } \gamma = -|\gamma|$$

β

It may have been noticed that β is not included in the previous diagram. This is because its definition and determination is a little more involved than those for α and γ. Following Zatsiorsky and Yakunin, β is the angle between Z_{oar} and an 'auxiliary axis', Z^{o}, that forms the angle γ with Z_{boat}. This clumsy sounding definition is hopefully clarified in Figure 4.19.

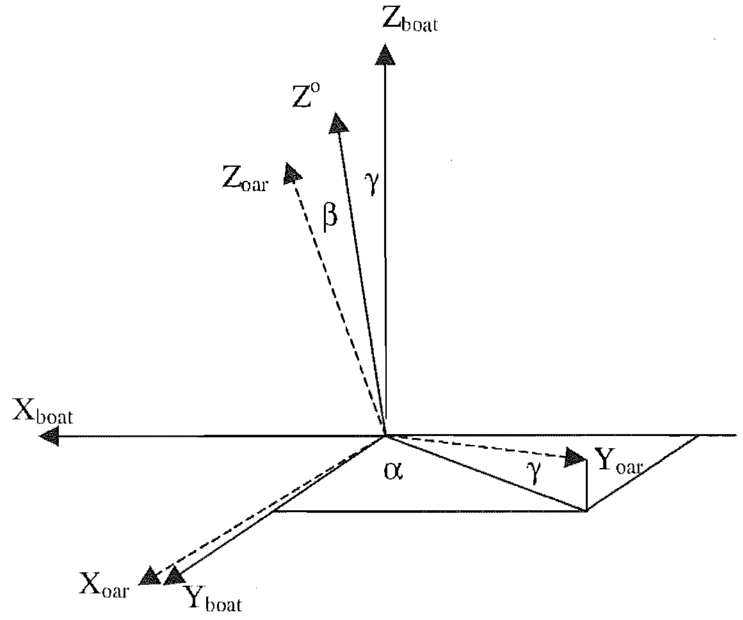


Figure 4.19 The auxiliary axis, Z^o, and the feathering angle, β.

The purpose of the auxiliary axis is to negate the part of the rotation due to pitching of the oar, which would be included if the scalar product of $\mathbf{z}_{boat}$ and $\mathbf{z}_{oar}$ were simply used. The auxiliary axis lies in the plane defined by $\mathbf{y}_{oar}$ and $\mathbf{z}_{boat}$ and has the same included angle and sense with $\mathbf{z}_{boat}$ that $\mathbf{y}_{oar}$ has with its projection onto the $\mathbf{x}_{boat},\mathbf{y}_{boat}$ plane, i.e. γ. This fact can be used to create $\mathbf{z}^o$, by treating $\mathbf{z}_{boat}$ and the unit vector in the direction of $\mathbf{y}_{oar}^{x,y}$ as the basis for the plane in which $\mathbf{z}^o$ lies. The new unit vector in the direction of $\mathbf{y}_{oar}^{x,y}$ is denoted $\mathbf{y}^o$ and is found by normalising $\mathbf{y}_{oar}^{x,y}$, i.e. $\mathbf{y}^o = \mathbf{y}_{oar}^{x,y}/\| \mathbf{y}_{oar}^{x,y} \|$. Reference to Fig. 4.19 shows that $\mathbf{z}^o$ has the same relationship with $\mathbf{z}_{boat}$ that $\mathbf{y}_{oar}$ has with $\mathbf{y}^o$, and that the relationship of $\mathbf{y}_{oar}$ with $\mathbf{z}_{boat}$ is the same as $\mathbf{z}^o$ and $-\mathbf{y}^o$. Thus, $\mathbf{z}^o$ can be written as

$$\mathbf{z}_o = -(\mathbf{y}_{oar}.\mathbf{z}_{boat})\mathbf{y}^o + (\mathbf{y}_{oar}.\mathbf{y}^o)\mathbf{z}_{boat} \qquad (4.122)$$

Upon the calculation of this unit vector, the feathering angle can be determined by

$$\mathbf{z}_{oar}.\mathbf{z}^o = \|\mathbf{z}_{oar}\| \, \|\mathbf{z}^o\|\cos\beta = \cos\beta \qquad (4.123)$$

Again due to the even nature of the cosine function, extra information is required to assign a sense to β. This can be achieved by finding the sign of the cosine of the angle between $\mathbf{z}^0$ and the cross product of $\mathbf{y}^0$ and $\mathbf{z}_{boat}$, i.e. if the cross product of $\mathbf{y}^0$ and $\mathbf{z}_{boat}$ is defined as $\mathbf{x}^0$, i.e.

$$\mathbf{x}^0 = \mathbf{y}^0 \times \mathbf{z}_{boat} \tag{4.124}$$

where $\mathbf{x}^0$ clearly lies in the $\mathbf{x}_{boat}$, $\mathbf{y}_{boat}$ plane, and forms a right handed orthogonal axis system with $\mathbf{y}^0$ and $\mathbf{z}_{boat}$, then β can be said to be positive if $\mathbf{z}_{oar}.\mathbf{x}^0$ is positive (the included angle is less than 90 degrees) and negative otherwise.

4.12 Discussion

While it is believed that with the new calibration method in place, the data yielded by the sensors, in combination with the relative orientation algorithm, should yield sound results, this will have to be verified by a future researcher, as the length and breadth of the Pacific Ocean now separate the author from the apparatus. It is hoped that research is continued in this area, as there are many applications for this technology, in sports and biomedical engineering in particular.

Chapter 5

The ultimate rowing performance indicator is not the peak force generated during the stroke, the stroke 'shape', or the acceleration profile of the rower within the boat, but the culmination of all these effects, the actual motion of the boat. From a systems point of view, up to this point only the inputs have been measured. The boat motion: displacement, velocity and acceleration are the outputs that the rower is trying to control through his technique.

Obviously, the time taken to cross a 'piece' of race length, say 2km, conveys overall performance, but there are very many interesting effects within the general motion that warrant investigation. Such effects include those caused by the motion of the rower, such as periodic speed variation.

This chapter describes a method of measuring the motion of the system's centre of mass. This is achieved through individual measurement of the kinematic parameters of the two substantial system components: the rower and the boat.

Since the rower's bulk is centred more or less over the seat, measurement of the motion of the seat is a good indicator of the motion of the rower's centre of mass. An optical rotary encoder was chosen to measure the seat displacement. This necessitated the design of a mounting bracket to force the encoder wheel to run on the chosen surface. To determine velocity and acceleration from the displacement signal requires differentiation. This can be problematic due to the quantisation noise on the output signal. A simple method of real time differentiation using a Kalman filtering approach is discussed and results are presented.

Two sensors are used to measure the motion of the boat: an accelerometer, and a submerged impeller. The outputs of these two sensors are fused using a Kalman filter of very similar design to that used as a differentiator for the seat motion.

The first part of this chapter reviews methods that have previously been used to measure rowing seat displacement, both on and off the water (in boats and on ergometers). After justifying the selection of the rotary encoder for the task, the design of the mounting bracket is described. Following this is a section on the design of Kalman filters for use as a differentiator of random 'periodic' signals. Numerous other techniques of numerical differentiation are also explored. After the presentation of sample results, attention is turned to the problem of measuring the motion of the boat. While the accelerometer used is exactly the same as that described in Chapter 4, some details regarding the additional sensor, a commercially available submerged impeller are required. The simple modifications required to transform the Kalman 'differentiator' to a basic sensor fusion technique are then described.

5.1 Seat Motion Measurement

5.1.1 Sensing Requirements

The displacement, velocity and acceleration of the seat are all parameters of interest, thus all three must be able to be determined from the output of the transducer used to sense the motion. It was considered, however, that as a base requirement the displacement should be available in real-time. This is because it is more easily interpreted than the velocity and acceleration, i.e. a direct correspondence can be drawn between the motion of the rower and the data on the screen, whereas in the cases of acceleration, more 'processing' has to be performed in the mind of the observer to negate the effects of boat motion. It was also considered that such a 'transparent' signal could act as a check to prove that the system was working.

5.1.2 Previous and Considered Methods

Seat motion does not seem to have been a priority for previous researchers, as until very recently it has not been measured. This is strange when it is considered that the motion of the rower(s) within the boat has a large impact on the vessel's progress.

Three very different methods have been used. Martin et al [50] analysed film of rowing motion and used the observed beginning and end of seat motion as indicators in the study of the effect of stroke rate on the velocity of the rowing shell. It is conceivable that such a system could be used to actually measure the position of the seat, although this would be problematic since a camera would have to operate at all times, creating a large computational burden if the camera were computer driven. Also the film has to be analysed after it is collected to yield any results (unless some real-time computer vision system is used, which would probably be expensive both computationally and monetarily.)

More recently, Rosow [58], [59] used a rotary potentiometer to study the motion of the seat on a rowing ergometer. It is not stated, but it is presumed that a geared wheel attached to the potentiometer ran on the surface of the ergometer beam. Farquhar [24] used a similar approach, with the potentiometer replaced by a rotary encoder. The use of encoders in the measurement of seat movement is considered below.

Another method of using a potentiometer, implemented by Loschner and Smith [43], to measure seat displacement is to drive a potentiometer via a cable attached to the seat. A spring-loaded drum unit that houses the potentiometer prevents slack in the cable that would cause measurement error. This method is also discussed below.

McBride [46] used a 'Hall-effect sensor' and 'magnetic track' to measure seat displacement. It is not stated whether the sensor is incremental, i.e. pulses are counted, or whether the magnetic track allows for actual position measurement. Certainly the non-contact approach of the sensor makes it attractive, but the lack of information meant the approach was not considered further. The methods of instrumentation that were considered are now described.

At first it was thought that a non-contact sensor would be the ideal solution, since this would not impede the motion of the seat. To this end an ultrasonic displacement sensor was considered and a Banner Q45-UL was obtained for trial (Fig. 5.1). This is an analogue sensor, produced for industrial applications operated on a time of arrival basis, with stated operating range of 100 to 1400mm, resolution of 0.25mm and repeatability of 0.1% of the sensing distance. This range is acceptable since the movement of the seat is limited by the tracks, which are approximately 800mm in length. The characteristics of the sensor were dependent upon both the material and physical size of the implemented reflector, with a 30mmx30mm aluminium plate recommended. The Q45-UL has attractive features, such as a programmable sensing window and splash proof casing in addition to the non-contact modality. In operation the sensor would be mounted a distance greater than 100mm behind the extreme rear seat position and a reflector would be fixed to the rear of the seat.

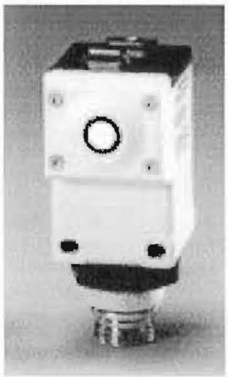


Figure 5.1 The Banner Q45-UL Ultrasonic Sensor

Qualitative testing of the Q45-UL was undertaken, with the sensor output level viewed on an oscilloscope. The amount of fluctuation and noise on the signal when the sensor and reflector were stationary appeared to be high. The process of differentiation that would be required to estimate velocity and acceleration magnifies any noise on the displacement signal. Another potential problem is the sensor mistakenly identifying another surface as the reflector and returning an erroneous signal. All of these problems could probably be overcome through the careful placement of the reflector and good signal processing, but at a cost of ~$800, it was decided that the price was too high for the required amount of work. Also, in some boats, particularly singles, there is very little room behind the extreme travel of the seat, meaning that mounting the sensor would be difficult.

An alternative non-contact method of measuring the seat motion is to use accelerometers. Measuring acceleration rather than position means that rather than requiring the noisy differentiation process, the signal is integrated to give velocity and position. Whereas integration can be said to average the effects of signal noise (over the short term at least), it certainly does not eliminate it. The first and second integral of a white noise signal are called random walks and ramps respectively. Examples of these phenomena are shown in Figure 5.2. A method of reducing the effect of the integrated signal noise is to periodically reset the integrator at a particular point. In the case of monitoring seat motion this can be achieved through the use of a limit switch. In reality this could be a reed switch activated by a magnet on the underside of the seat. The position would be reset each time the switch is activated.

Figure 5.2 A Gaussian white noise random signal and the first two integrals (cumulative sums), a random walk and a random ramp. Note the vastly different scales for unit divisions in the plots: for the original signal, the division is unity, for the random walk, the division is 50 and for the ramp, it is 1x10^5

To trial several methods of measuring seat motion using an accelerometer, a test rig was set up. An accelerometer and optical encoder were attached to a sliding rowing seat and sampled during representative motion. The output of the encoder was used for two purposes, firstly to act as a reference to appraise the integrated signal of the accelerometer and secondly to act as a 'virtual limit switch'. After sampling, the data was imported into a function written in MATLAB that:

1. converted the encoder and accelerometer outputs into compatible units (m and m/s^2)
2. identified the zero crossings* as defined by the encoder and made a vector with sampling instants corresponding to the locations of the zero crossings
3. doubly integrated the accelerometer signal using the Simpson rule and reset the displacement at each instant contained in the 'reset vector'.

*When the instrumentation was started, the value of the encoder was set to zero, thus zero crossings are associated with instants where the seat is in the same position at the culmination of sampling.

Trials showed that between resets, the integrated accelerometer output diverged from the encoder signal considerably. An obvious method of minimising this effect is to increase the number of limit switches. A set-up that consisted of three magnets on the seat and five limit switches was conceived, as shown in Figure 5.3. The spacing between magnets A&C is such that two reed switches may be simultaneously activated. If only one switch is activated it is known that magnet B must be directly over that switch (the extreme switches (-2&2) are placed so that it is impossible for the opposing extreme magnets (C&A) to activate them). Using this method it is possible to identify the position B as one of nine discrete possibilities.

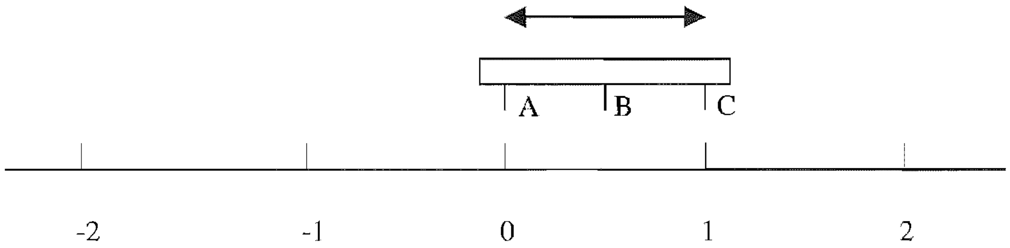


Figure 5.3 The seat, with three reed switches attached, traverses the track, which has five evenly spaced magnets on it.

Results obtained through integration and resets using an obvious extension of the aforementioned computational method are shown in Figures 5.4 & 5.5. Each time the integrators are reset, a spike results in the output. Filtering can reduce this effect, but as shown in Figure 5.6, this introduces a delay to the estimated signal.

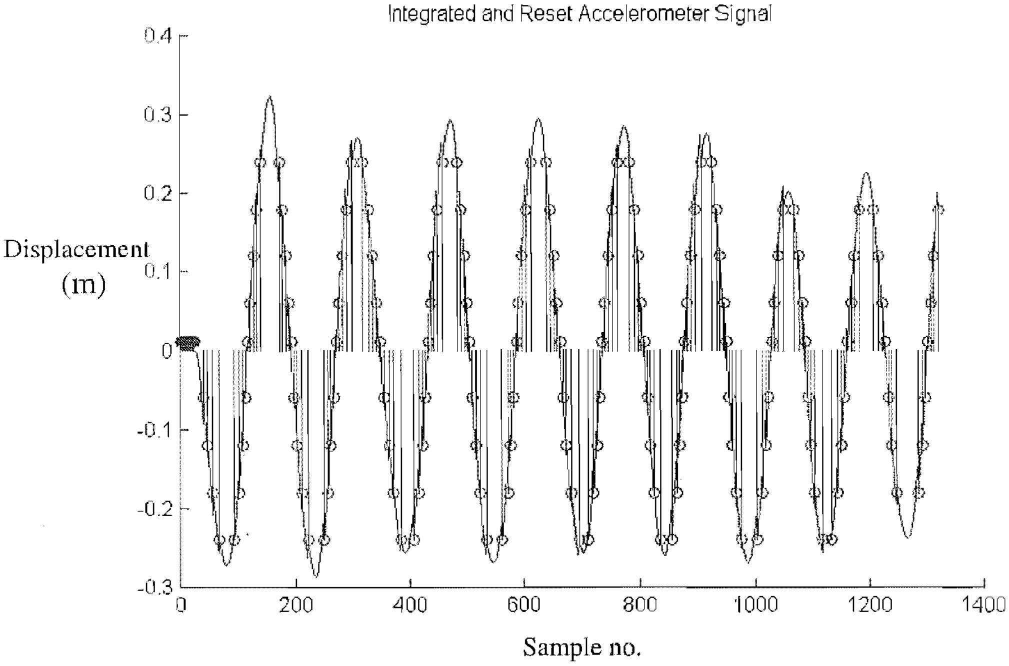
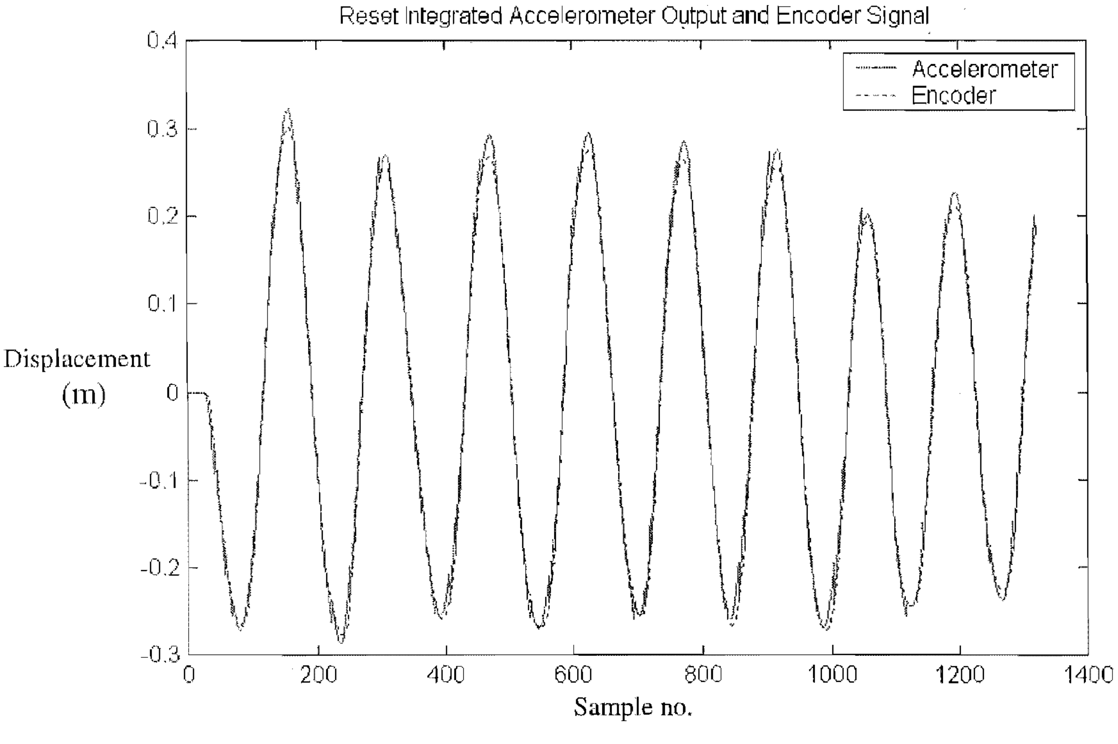


Figure 5.4 Typical results obtained by integrating and resetting accelerometers, using the three switch, 5 magnet approach. The circles indicate reset points.



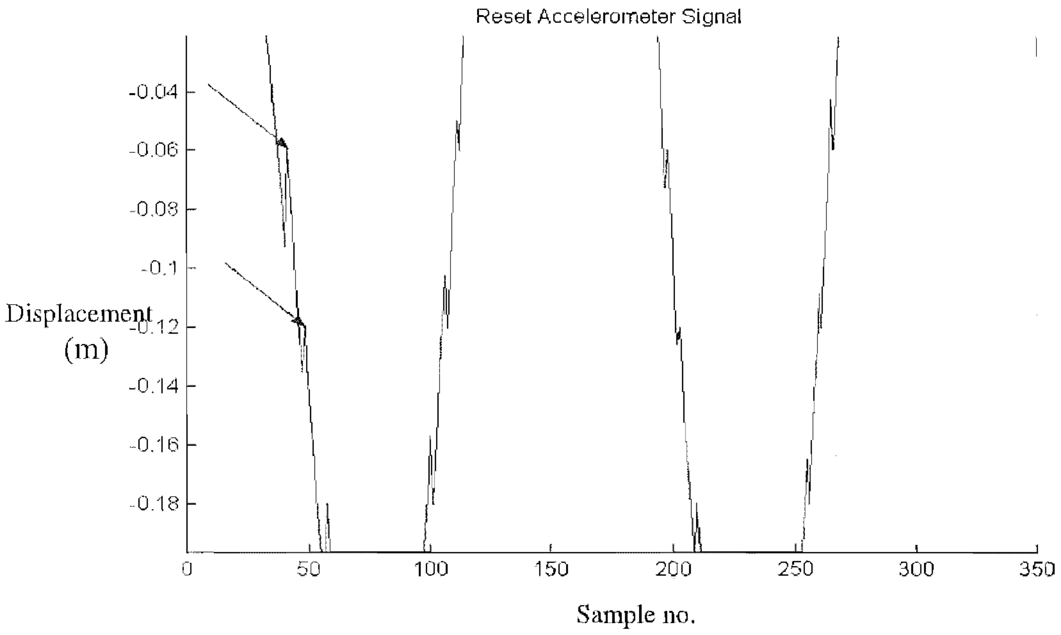


Figure 5.5 Two more examples of the results obtained using the reset accelerometer approach. The upper plot is the same as Figure 5.4, with the reset indicators removed, while the lower plot shows a close up of the spikes caused by resetting.

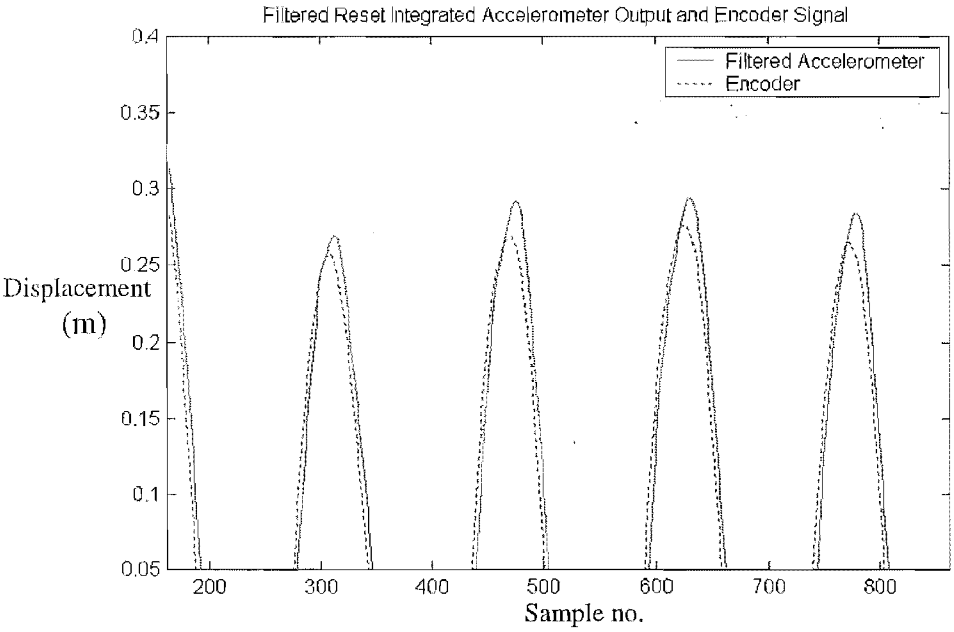


Figure 5.6 The delay introduced by filtering the integrated and reset output.

While the results of integration, reset and filtering of the accelerometer output were encouraging there were three problems, the first of which is the lag introduced by the filtering. Since timing of events was of interest it was desirable that lag be avoided.

This is actually a small problem when the magnitude of the delay caused by filtering is considered. The second problem has not yet been mentioned and is concerned with the motion of the boat. Since the boat is undergoing motion the acceleration measured by the accelerometer on the seat is a combination of the boat motion, seat motion and gravity. It was considered that compensating for the motion of the boat would incur too much computational burden to produce a seat displacement signal in real time. Lastly, the method of resetting the integrators at known positions does not aid in the estimation of velocity. Thus, while position may be found by this method, velocity would require another procedure. Such a procedure could be setting the velocity to zero every time the estimated position experiences a maximum. This would by necessity be a post-processing measure. While the accelerometer and switch method was abandoned, it was found, as a matter of interest, that the reed switches by themselves could emulate the motion of the seat to a high degree of accuracy using cubic splines. Examples of this are shown in Figures 5.7 & 5.8.

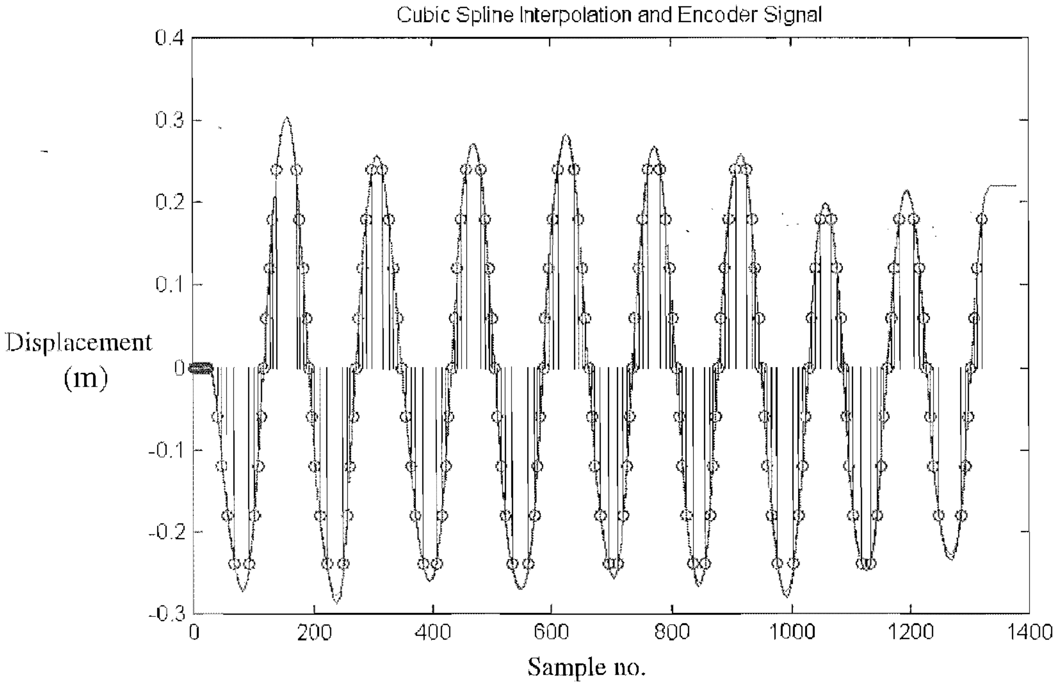


Figure 5.7 Cubic spline interpolation of the reset points (indicated by circles) and the encoder output.

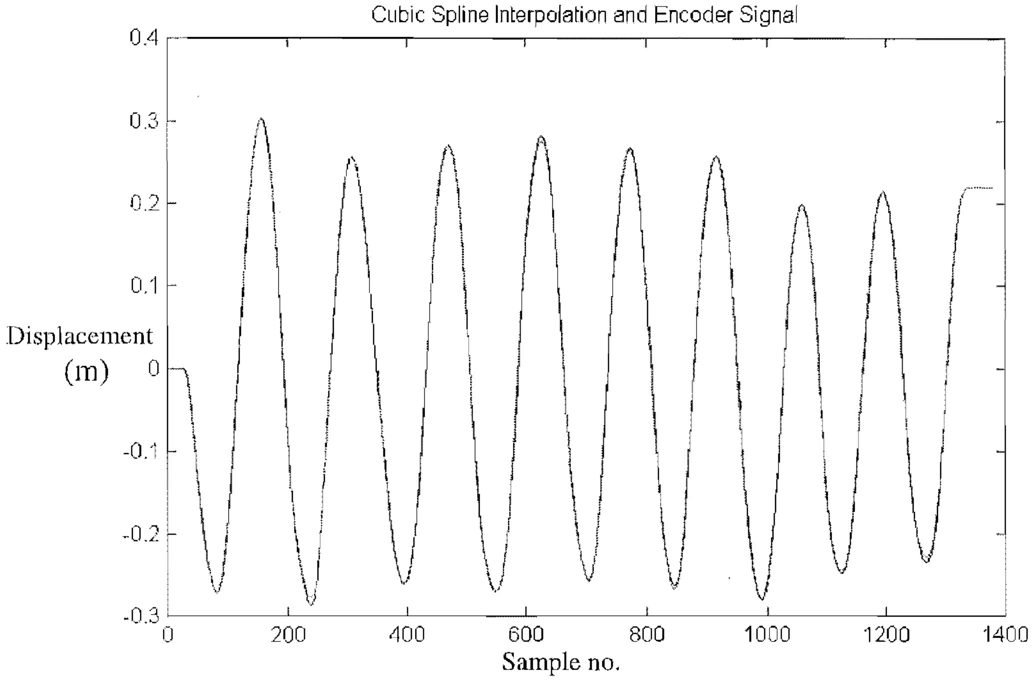


Figure 5.8 Position estimates achieved by cubic spline interpolation of limit switch data. The differences between the interpolant and the encoder data are only evident at some of the turning points.

On the basis of these results, various attempts were made to devise a real time cubic spline interpolation scheme. Two algorithms were designed, both of which were initialised in the same way, by creating an initialising cubic and then enforcing continuity. This is probably most easily understood by considering Figure 5.9.

Position fixes, corresponding to instants at which reed switches are closed give temporal-spatial coordinates (t_i,d_i). At every instant, noisy acceleration measurements are available. Four conditions are required to specify the initial cubic interpolant, $S_1(t)$. At t_2 these conditions are available, i.e. the positions and accelerations at t_1 and t_2. Using these conditions is equivalent to giving the endpoints of the curve, as well as the second derivative of the curve with respect to t.

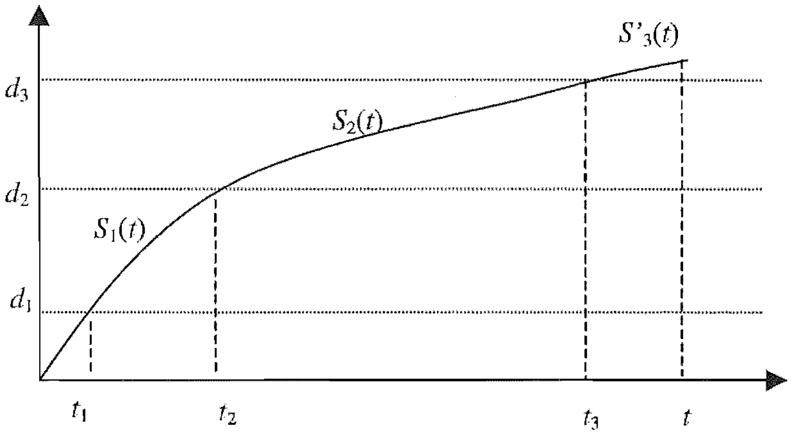


Figure 5.9 An illustration of the implemented cubic spline method. Position fixes d_i occur at times t_i and interpolants S_i are created between the coordinates (t_i,d_i) and (t_{i+1},d_{i+1}).

The initial interpolant is of the form:

$$S_1(t) = A_i + B_i(t - t_i) + C_i(t - t_i)^2 + D_i(t - t_i)^3, \qquad t_1 \le t \le t_2 \qquad (5.1)$$

Consider the seat to be in between position fixes at an arbitrary time t ($t_2 \le t \le t_3$). The only information available at t is the acceleration, $a(t)$. It is possible to create a cubic, $S'_2(t)$, to approximate the motion, using this value of acceleration by imposing continuity of $S'_2(t)$ with $S_1(t_2)$. This continuity involves the actual value of the new cubic $S'_2(t)$ at t_2, as well as its slope and 'curvature' (velocity and acceleration). At each consecutive sampling instant only one coefficient of the cubic has to be recalculated, that corresponding to the $(t - t_2)^3$ term, since all others are fixed by the continuity requirements. As soon as the seat reaches position d_3, the initialising cubic becomes $S_2(t)$, which is calculated in a way entirely analogous to $S_1(t)$, using (t_2,d_2), (t_3,d_3), $a(t_2)$ and $a(t_3)$ (where $a(t_i)$ is the acceleration at t_i). The continuity of $S'_3(t)$ at t_3 is then used to specify all but one of the coefficients, which is calculated using $a(t)$. This method suffered from the fact that the value of the interpolant at t depended only on the value of three values of acceleration, the two that specified the initialising interpolant, and that at t. If, in particular, a value of acceleration used to generate the initialising cubic had a large amount of error, the interpolating cubic was of poor quality. This is in contrast to numerical integration schemes where the (assumed) zero mean nature of noise signals is somewhat 'averaged out' through summation over short periods.

A second algorithm was designed that differed only in the way in which the cubic term of the function is calculated. The method was to impose a particular slope (velocity) at the point of interest rather than a 'curvature' (acceleration). To determine the required slope necessitates numerical integration of the acceleration measurements. The numerical integration begins at the initial point of the interpolating cubic (e.g. for $S'_3(t)$, the integration begins at t_3), at which point the initial value is found by evaluating the differential of the initialising cubic ($S_2(t_2)$). While this approach did mean that more values of acceleration were taken into account in the determination of the value of the interpolant between position fixes, it did not change the fact that only two are used to determine three of the coefficients of the cubic, thus errors in these two measurements are propagated and magnified into the estimated position.

The problem with both the above schemes are similar, there are not sufficient acceleration measurements used in the determination of the initialising cubic to 'average out' the effects of signal noise. A method by which this may be overcome is to specify the initialising cubic using the position fixes and the velocity (obtained by numerical integration) at the end points. If this approach was combined with the use of the second of the above-mentioned methods, the estimate may be of much better quality. Essentially a position estimate is being obtained through single integration of noisy acceleration data, thus the error can be expected to behave as a periodically reset random walk. This is in comparison to regular numerical integration schemes in which the required double integration results in a random ramp type error. The downside of the spline approach is a slightly higher computational burden (due to the calculation of the spline coefficients).

While this approach is appealing, and may warrant further investigation for a simpler case, it suffers from the same problem as the regular use of accelerometers and limit switches – the requirement of accounting for the motion of the boat and the possible influence of gravity. The reason that simple spline interpolation using only the position fixes worked so well is the continuity imposed during the calculation, i.e. the velocity and acceleration are continuous at given data points. The smooth nature of the seat displacement signal makes it ideal for this type of interpolation, while other applications may produce signals where the acceleration is not so 'well-behaved'. Since it was thought that accelerometers would be too difficult to work with in this application, transducers that would yield displacement directly were explored.

The use of a cable potentiometer and optical encoder were considered concurrently. An identified suitable cable-type potentiometer was the LX-PA from Unimeasure, at a cost of US$160 (see Figure 5.10). The main advantage of the potentiometer over the encoder is that it has a simple analog output as compared to the encoder, which requires a digital input card with encoder capabilities. The disadvantage of the cable type sensor is the possibility of slack in the cable during the recovery, i.e. the seat moving at a rate higher than the drum can retract. Since there was already an encoder available for use (that used by Farquhar [24]) the cost of the potentiometer was compared to that of the necessary digital I/O card. Further, it was found that a card designed in-house, the Universal Pulse Processor (UPP), had the required capabilities and could be used at little cost. It should be noted that a digital I/O card was also necessitated by the use of other sensory devices (impeller speed sensor and heart rate unit).

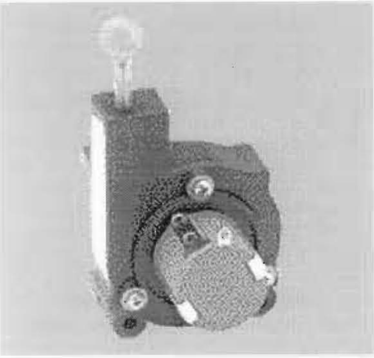


Figure 5.10 The LX-PA Cable Displacement Sensor

5.1.3 Chosen Concept

In the interests of minimising required computer time while maintaining a real time seat position signal it was considered that a position sensor, e.g. encoder or potentiometer, that did not significantly alter the characteristics of the motion of the seat would be the most sensible option. (The magnetic system used by McBride may be optimal in that it offers a real time position signal using a non-contact sensor, but no precise information was available on this at the time.) Based on the economic advantages discussed in the previous section, a rotary encoder was chosen.

The main issue with using a rotary encoder to measure position is ensuring that there is no slip between the wheel of the encoder and the surface on which the wheel is running. The method by which this is normally ensured, and the method employed here, is to create a significant force between the wheel and the surface and use a high grip surface on the wheel. Two potential running surfaces for the encoder wheel were identified, the deck of the boat and a small lip on the seat track, present to prevent the seat from lifting from the track. The latter site was chosen as it seemed that this was almost standard from boat to boat, and also, due to the tilt of the track on which the seat runs, the distance between the seat and deck varies considerably during the stroke. This difference in distance would have made design of a mounting bracket difficult.

A bracket was designed to fit under the seat and push the encoder wheel onto the track (see Figures 5.11 & 5.12). This was an interesting design problem due to the tight spatial constraints and the requirement that the bracket be adaptive enough to fit a variety of seats. The main part of the bracket was designed as a single piece to be manufactured by EDM (Electro Discharge Machining). Each end of this part rests on the top of the seat axles. Socket head cap screws tightened onto the axles prevent translation of the bracket with respect to the seat in the direction of motion. The plate on which the encoder is mounted runs on two brass screws and is pushed downwards by two compression springs, thus providing the force required to prevent slipping of the encoder wheel. Blocks can be inserted between the main part of the bracket and the upper piece to accommodate for larger distances between the top of the axles and track lip.

While the main problem with using a rotary encoder to measure linear displacement is ensuring that the encoder wheel does not slip, there is another problem associated with using the encoder to generate velocity and acceleration estimates. This is due to the fact that the encoder does not yield a continuous waveform, but a quantized signal, which can be thought of as a noise contaminated signal, and this 'noise' is magnified by the required differentiation.

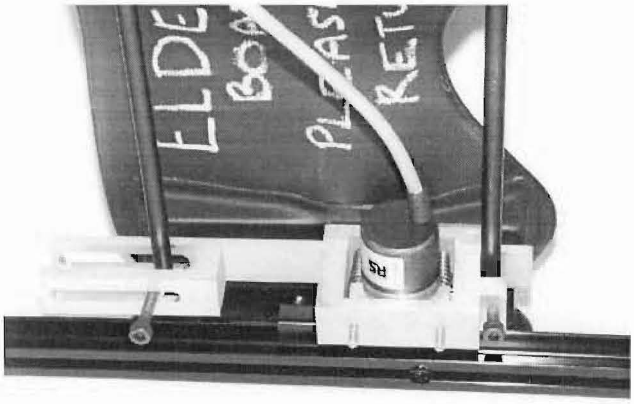


Figure 5.11 Photograph of encoder and bracket in situ.

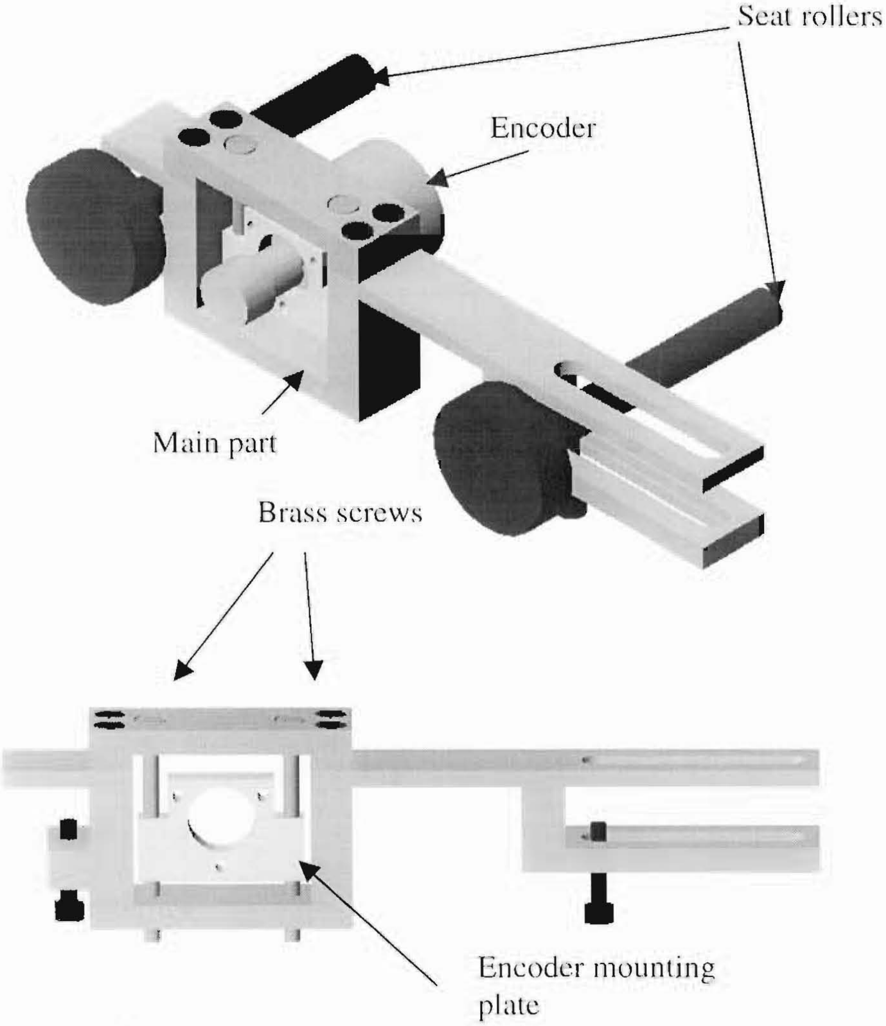


Figure 5.12 SolidWorks® generated views of the encoder mounting bracket.

5.2 Numerical Differentiation of Encoder Signals

If a rotary encoder could be continuously sampled while the shaft underwent continuous angular velocity motion, the conditioned signal of the encoder would appear as a staircase approximation to the true displacement. The depth of the steps is dependent upon the resolution of the encoder, i.e. the number of pulses per revolution. In actual application the encoder is attached to a counter card. Simplistically, the card counts pulses accounting for direction changes and outputs the number representing the net number of rotations (and parts thereof) to a particular register. The value contained in the register is then sampled at regular intervals. The quantisation of the encoder can cause problems at both extremes of speed. When rotation is very slow, sampling may be such that no pulses are recorded for certain intervals. Conversely when rotation is occurring at high speed a large step occurs in the output data.

Assuming that no slip is involved, an encoder does provide the exact position at the instant at which a pulse is emitted. It is rare however that a pulse coincides with a sampling instant, and thus a position error is introduced, which is dependent on the average velocity over the period and the time between the pulse and the sampling instant. Carpenter et al [15] reviewed a number of different algorithms used to estimate velocity given encoder measurements mainly from a frequency domain perspective. Two interesting methods are the 'least squares filters' and the 'trained least squares filters'. In the former method, a polynomial of order m is passed through n points using a least squares fit. In possession of the polynomial, the velocity is simply the derivative of the polynomial at the sampling instant of interest. In the second method, filter coefficients are generated by running the least squares filter on representative data. In this way a time-invariant filter is generated. Independent experimentation showed that if the filter is being used on real-time data there is a difficult trade-off between filter lag and noise attenuation, i.e. increasing the order of the filter yields smoother results, but also introduces considerable lag. Post-processing the data using approximating polynomials centred on the data point of interest yielded good results.

Differentiating an encoder signal to get velocity and acceleration estimates is a special case of numerical differentiation, that of differentiation of noisy signals. The amount of literature in this field is large, most of it in mathematical numerical method journals. The results in these articles are prohibitive in their mathematical sophistication.

An interesting alternative to normal or modified numerical differentiation is the use of a 'Real Time Fourier Series' as proposed by Tang et al [63]. In this method a finite length of data terminating at the current sample is represented as a truncated Fourier series and differentiation is performed analytically on the individual terms. The effects of high frequency noise are reduced by the truncation of the series, i.e. if n points are included in the data series, then the first n-1 spectral components are calculated. A recursive formula for the computation of the Fourier series coefficients is presented. While this is an attractive idea, the authors state that the calculation of the Fourier series coefficients is a time consuming task that makes it unsuitable for real time use on current computers.

The method of numerical differentiation used in this work employs a Kalman filter (KF). A few previous instances of the use of optimal estimation theory in numerical differentiation were identified, which is not to say that it is not much more widespread.

A few brief comments are probably required to indicate the way in which a Kalman filter aids numerical differentiation. (A probably excessively full derivation and description of the Kalman filter algorithm is included in Appendix A2). A discrete time state space model is derived so that the states 'resemble' the position, velocity and acceleration of the seat when the model is driven by white noise. Such a model is shown in Fig. 5.13, where Δ represents a unit delay, Φ is the state transition matrix, Γ is the noise coupling matrix, $\mathbf{w}(k)$ is the white noise input, and $\mathbf{x}(k)$ is the state (position, velocity and acceleration).

Rather than being able to measure all of the states, there is only a noisy measurement of the position, $z(k)$, which is modelled as a scaled version of position with additive white noise, *v*(k). The aim of the Kalman filter is to optimally estimate the state of the system, based on the measurement and the assumed signal model. An alternative form of the filter would be to model the dynamics of the rower's sliding motion and use the force produced at the feet and that at the oarlock as inputs to the system. Even in this 'deterministic' case a white noise input is still required.

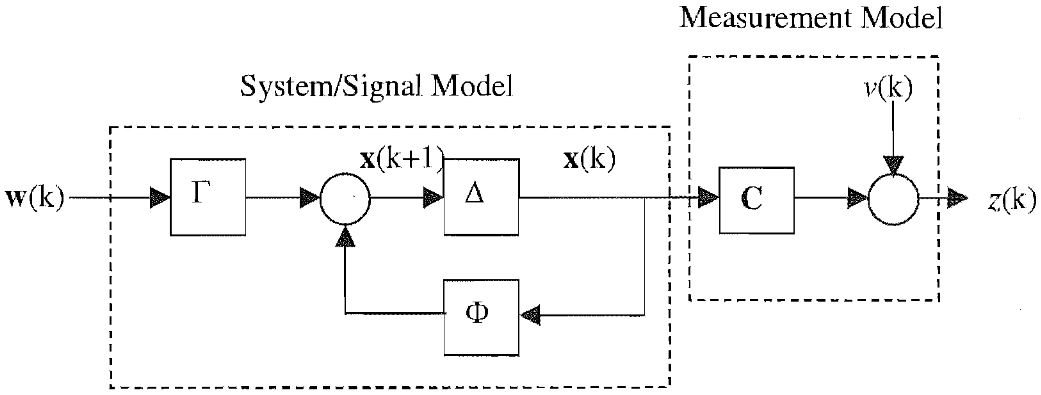


Figure 5.13 The discrete time state space model used in the Kalman filter

The discrete-time Kalman filter state estimate at the k+1th instant, is given by the equation:

$$\hat{\mathbf{x}}(k+1|k+1) = \Phi\hat{\mathbf{x}}(k|k) + \mathbf{K}(k+1)[z(k+1) - \mathbf{C}\Phi\hat{\mathbf{x}}(k|k)] \qquad (5.2)$$

where $\hat{\mathbf{x}}(k|k)$ denotes the estimate of the state at instant k based on all measurements up to and including k and **K**(k+1) is the recursively calculated Kalman gain matrix at k+1. The calculation of **K**(k) involves three equations, that are not shown here (see Appendix A2). The estimate equation has exactly the same form as any other estimator, and is very easily interpreted. Before the measurement at k+1 is made, the best estimate that can be made is the propagation of the estimate at the kth instant using the state transition matrix. Once the measurement becomes available, this a priori estimate can be corrected using a weighted residual, i.e. the difference between the measurement and the propagated estimate.

What differentiates the Kalman filter from other estimators is the definition of the **K** matrix. This gain matrix is calculated to minimise the mean square error of the state estimate. The statistical information that is required to yield this optimal form is included in the covariance matrices of the system and measurement noise processes,

Q and **R**. The noise covariance matrices are usually set roughly using knowledge about the noise processes, e.g. the variance of noise expected from a certain sensor, and then tuned to give good performance. Once the model of the system is fixed, i.e. the state transition and noise coupling matrices have been designed; the **Q** and **R** matrices are the only degrees of freedom available to the designer. Without complicating matters by introducing the equations by which **K** is calculated, it can be stated that when **R**, the measurement noise covariance matrix, is small compared to **Q**, the system noise covariance matrix, this 'tells' the filter to 'believe' the measurements more than the propagated estimates, and **K** is altered accordingly. For those who are interested in the derivation of the Kalman filter, a full development is included in the appendix.

Before progressing to discuss previous methods through which Kalman filters have been applied to the differentiation of signals, and the development of a differentiator for periodic signals, a couple of notes are required. In the case where **Q** and **R** are constant matrices, it is possible to find **K**, the static Kalman gain, through solution of an algebraic Riccati equation, and the filter algorithm is reduced to a single invariant equation [26].

While the optimality of the Kalman filter algorithm is obviously attractive, the abuse of the **Q** and **R** matrices as tuning factors and the difference between the ideal and implemented models means that the optimality is unlikely to be realised in the application at hand. Even in this disabused state the Kalman filter is of great utility, as is shown in the following sections.

5.2.1 Previous Approaches to Numerical Differentiation Using Kalman Filtering

Three papers concerned with numerical differentiation using Kalman filters were found. As the Kalman filtering algorithm is fairly standard, the main differences in the approaches are in the models used in the filter, i.e. the choice of the Φ and Γ matrices, or the corresponding continuous time **A** and **G** matrices. Two of the papers [12], [25] used very similar models even though one was designed for general use and the other was created specifically for the differentiation of data obtained in tracking points of the human body as recorded by video cameras. The third paper [8], by Belanger, is concerned specifically with the estimation of angular velocity and acceleration given shaft encoder measurements. Surprisingly the approach of this paper is not relevant to the problem at hand and hence is not reviewed in detail here. The approach taken by Belanger to justify his choice of model is asymptotic analysis of a general state space model in companion form as the sampling period tends to zero. Through this he arrives at a model that is similar, although a little more simplistic, to those described below.

The first step in the application of the Kalman filter to numerical differentiation is to design a random sequence that models the signal or one of its derivatives. This is where Bortolami [12] and Fioretti [25] have slightly different approaches. Bortolami suggests that a discrete time model of the derivative of acceleration can be adequately described by the first order Gauss-Markov model

$$\varepsilon_{k+1} = \phi\varepsilon_k + w_k \tag{5.3}$$

Fioretti is perhaps a little more conservative in suggesting that the N+1th derivative of the signal of interest has this model, and that if poor performance is experienced with the original design, the order of the filter should be increased. In continuous time formulation, Fioretti's approach basically means that the N+1th derivative of the signal is a white noise process.

Both approaches use a single noisy position measurement

$$z_k = \mathbf{C}\mathbf{x}_k + v_k \tag{5.4}$$

where $\mathbf{C}_{Bortolami} = [1\ 0\ 0]$ and $\mathbf{C}_{Fioretti} = [1\ 0\ 0\ \ldots.]$.

Bortolami used his model in a Kalman filter, while Fioretti used a Kalman smoother. Using a smoother rather than a filter, i.e. using past, present and future values of output, rather than just past and present to approximate $\mathbf{x}_k$ can reduce the error variance of the estimate. (Incidentally, it seems that Bortolami made an error in the discretisation of his model in that he did not modify his 'noise input' matrix.)

5.2.2 A Kalman Filter Differentiator for Smooth Oscillatory Signals Measured in Additive Noise

There were two proposed modifications to the previously reviewed methods, although only one was ultimately adopted.

The major difference in the filter is due to the fact that quite a lot of information is known about the seat displacement. Later it will become clear that it is local knowledge of the waveforms, e.g. maximum frequencies of oscillation, rather than global characteristics such as the gross frequency of the seat motion that aid in the design of the filter. More comments will be made on this below, but for now it suffices to state that a 'periodic random variable' model is more suitable than the Gauss-Markov variants implemented by previous researches, as described above.

Apart from the model 'within the filter', a second proposed modification was a method by which measurements 'enter' the filter. It was found however that this method yielded little performance benefit for the increased model size. This proposed modification is fully described below.

The continuous time kinematic equations relating distance, d, velocity, v, and acceleration, a, can be written in state space form as:

$$\begin{bmatrix} \dot{d} \\ \dot{v} \\ \dot{a} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} d \\ v \\ a \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \dot{a} \tag{5.5}$$

For use in a Kalman filter (before discretisation), the model should be of the form

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{G}w \tag{5.6}$$

where w is a white noise process. Since the motion is known to be low frequency, it is unlikely that modelling the derivative of acceleration as a white noise process will give good results. To yield a model in the required form, the state of (5.5) is augmented, e.g.

$$\begin{bmatrix} \dot{d} \\ \dot{v} \\ \dot{a} \\ \dot{\alpha} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & a_{11} & a_{12} \\ 0 & 0 & a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} d \\ v \\ a \\ \alpha \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ g_1 \\ g_2 \end{bmatrix} w \tag{5.7}$$

where one state, α, has been added. The new state does not have any physical meaning, but has been added simply so that the model has a white noise input while allowing a better model of the derivative of acceleration. This method of augmentation is sometimes known as a shaping filter, where the input noise, w, is shaped to statistically resemble a known process. The main part of the design of the Kalman filter was in this case the design of this shaping filter, the a's and g's of the above state space equation. This is now described for the particular case in which the output of the shaping filter is to be a random oscillatory waveform. Upon the completion of the design of the shaping filter, a discretised form of the continuous state space model is applied in a discrete time Kalman filter.

Stochastic modelling, which consists of designing a system that gives a desired response to a random process input, can be considered in either the time or frequency domain. These closely related approaches are briefly explained for the special case in which the input to the system is a white noise process. Additionally, what appears to be a new method of checking the nature of a signal generated using a discrete time state-space model in response to a white noise input is presented.

5.2.2.1 Modelling a Periodic Random Variable

The aim of this section is to design a linear time invariant (LTI) system that will generate a signal with periodic 'tendencies' in response to a stationary white noise input. This signal will be known as a periodic random variable, although the output is not strictly periodic, but oscillatory with a prescribed base frequency. The designed model is to be incorporated into the state matrix of a discrete Kalman filter. Since the signal is random, it is impossible to actually specify the output of the system; the best that can be done is to indicate the output's statistics. In particular, specifying the mean and autocorrelation of the output gives a large amount of information about the expected signal. Both discrete and continuous time concepts are used at different times to make the mathematical relationships simpler and to appeal to the mechanical engineer's intuition respectively.

When the input to a discrete LTI system is the stationary white noise random sequence $\{w(n)\}$, the output sequence is a wide sense stationary sequence (the mean is constant and the autocorrelation sequence is a function of a single variable $R_{yy}(n,n+m) = E[y(n)y(n+m)] = R_{yy}(m)$). Further, the autocorrelation of the system output is given by a scaled version of the autocorrelation of the impulse response of the system. The developments of these facts are in standard signal processing texts [7], [54]. Mathematically stated, the output autocorrelation in response to a general stationary random input sequence $\{w(n)\}$ is given by the convolution sum:

$$R_{yy}(m) = \sum_{l=-\infty}^{\infty} R_{ww}(m-l)c(l) \tag{5.8}$$

where $c(l)$ is the autocorrelation of the system impulse response sequence

$$c(l) = \sum_{k=-\infty}^{\infty} h(k)h(l+k) \tag{5.9}$$

and $R_{ww}(m-l) = E[w(k)w(k+m-l)]$. In the case where $\{w(n)\}$ is zero mean stationary white noise sequence, the autocorrelation function $R_{ww}(m-l) = \sigma\delta(m-l)$, i.e. a unit impulse at $m = l$. This simplifies the above convolution sum to give

$$R_{yy}(m) = \sum_{l=-\infty}^{\infty} \delta(m-l)c(l) = c(m)\sigma\,. \tag{5.10}$$

This result enables time domain design of systems to give outputs of desired autocorrelation in response to white noise excitation. In the case at hand, an output signal with periodic tendencies is desired. Using the fact that the autocorrelation of a periodic sequence has the same period as the original signal [51], it is seen that the impulse response of the designed system should be oscillatory with a base frequency similar to that of the desired stochastic signal. The simplest system with an oscillatory impulse response is of second order. The design of an oscillatory second order system is obviously possible in discrete time, using the Z-transform, but it is believed that most readers will be more familiar with continuous time and the associated Laplace transform. While the above relations for the output autocorrelation are for the discrete time case, there are entirely analogous results that relate the autocorrelation of the input and output sequence in continuous time. Alternatively, since the model is going to be implemented on a digital computer, the continuous time model can be discretised and then the aforementioned autocorrelation relationships can be used. Regardless, the impulse response of a general second order system is given by an exponentially decaying sinusoid of frequency $\omega_n(1-\zeta^2)$ and decay of $\omega_n\zeta$, where ω_n and ζ are the natural frequency and damping ratio respectively. Examples of this response for $\omega=1$ rad/s and varying ζ are shown in Figure 5.14. The autocorrelation functions of the impulse responses are shown in Figure 5.15. The autocorrelation sequences are only shown for positive shift values (since they are symmetric about zero) and have been normalised to have unit mean squared value (value of autocorrelation at zero shift).

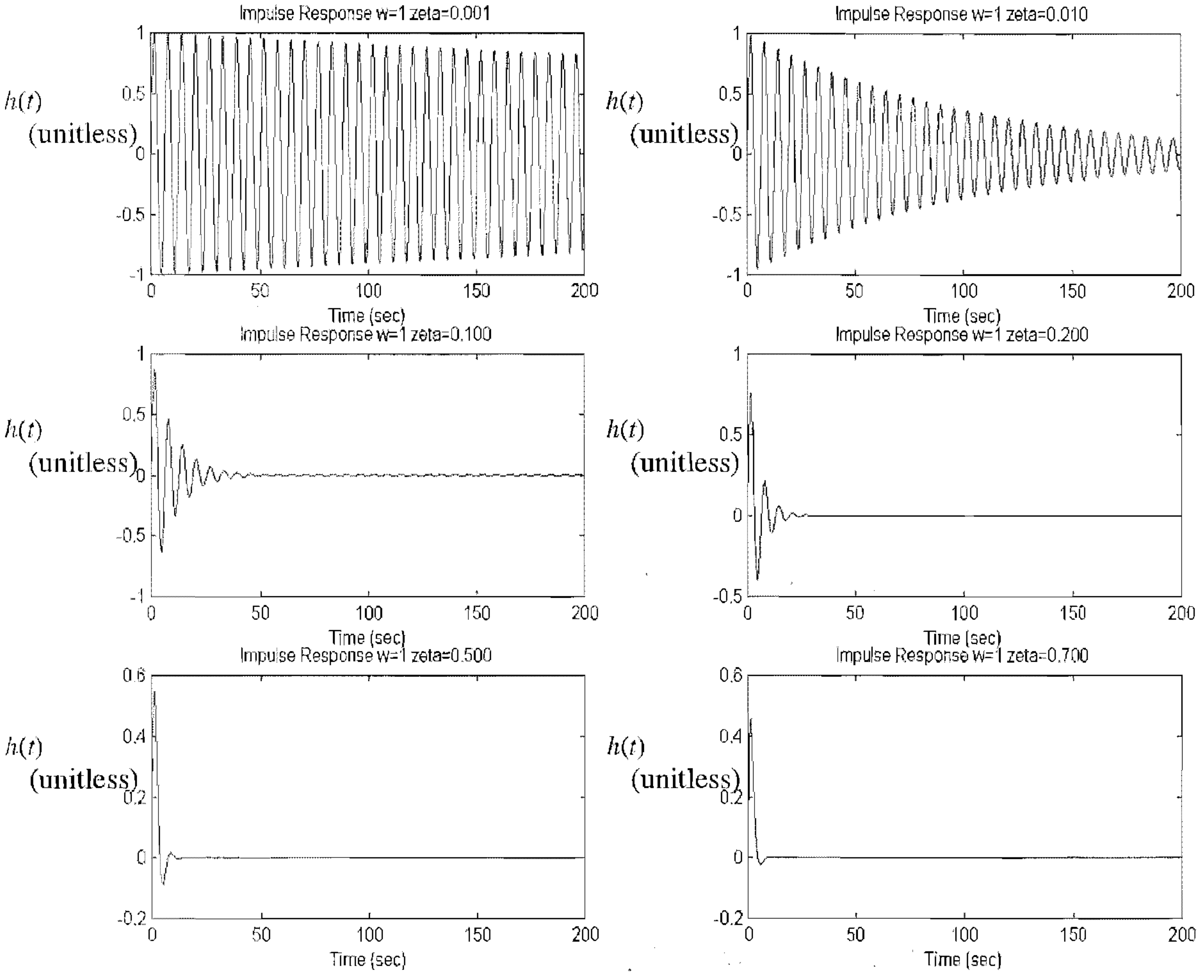


Figure 5.14 Impulse responses of second order systems with ω=1 and ζ=0.001, 0.01, 0.1, 0.2 0.5 and 0.7 (from top left to bottom right)

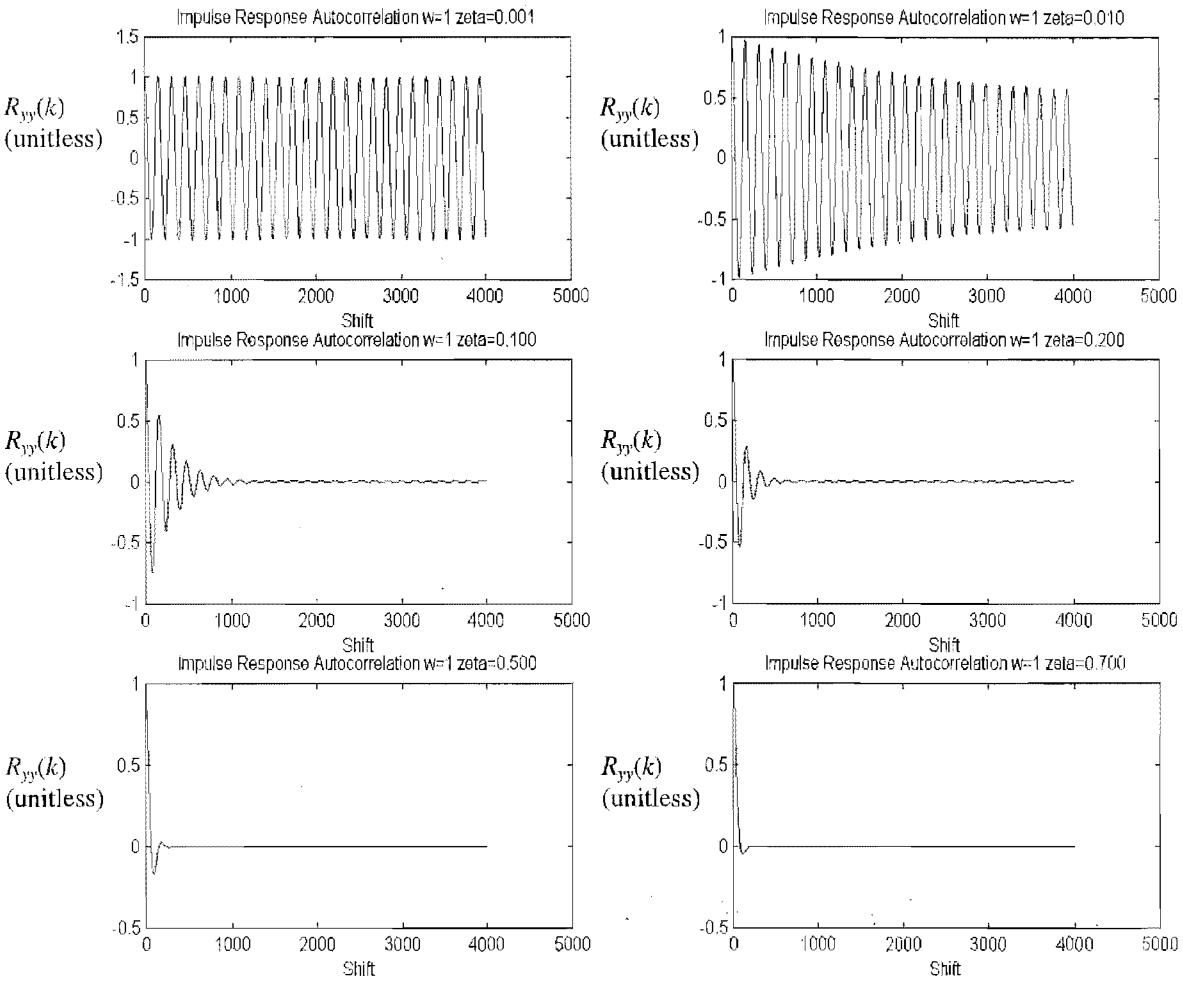


Figure 5.15 Autocorrelation of impulse responses of second order systems with ω=1 and ζ=0.001, 0.01, 0.1, 0.2 0.5 and 0.7 (from top left to bottom right). These represent the theoretical autocorrelations of the output of the second order systems when the input is white noise.

Because the decay of the autocorrelation shows how dissimilar the signal is to itself as shift increases, the decay of the impulse response, which is controlled by choice of ζ, impacts on the variability of the generated signal. The faster the decay, the less similar shifted versions of the output signal are to each other. In the case of zero damping, i.e. a pure oscillator, the autocorrelation function indicates that the random signal should be entirely periodic. This is of course impossible, but it is true that the larger the value of ζ chosen, the higher degree of variability the random signal will show 'period' to 'period'. Designing a model in the time domain therefore consists of choosing the approximate desired frequency of the generated signal, ω, and then choosing a value for ζ that reflects the desired randomness of the signal. (Note that the choice of ζ does have an effect on the frequency of the impulse response oscillation and also therefore on the output random signal.) There are many ways in which the designed system may be represented, but since it is desired that the model be implemented as part of a discrete time Kalman filter it should eventually be transformed to discrete time state space model. Knowledge of ω and ζ allows the construction of a transfer function

$$\frac{Y(s)}{U(s)}=\frac{\omega_n^2}{s^2+2\zeta\omega_n+\omega_n^2} \tag{5.11}$$

This transfer function is then simply converted to continuous time state space, with x_1 as 'position', and x_2 as 'velocity':

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}=\begin{bmatrix}0 & 1\\ -\omega_n^2 & -2\zeta\omega_n\end{bmatrix}\begin{bmatrix}x_1\\ x2\end{bmatrix}+\begin{bmatrix}0\\ \omega_n^2\end{bmatrix}u \tag{5.12}$$

$$y=\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} \tag{5.13}$$

This model is finally converted to discrete time state space using a zero-order hold transformation. The discrete formulation does not have a convenient closed form. A range of generated outputs for the values of ω and ζ previously considered are shown below in Figure 5.16. Note that since these plots are generated by the discrete state space formulation, the outputs are sequences and should therefore not be plotted as being continuous. They have been plotted as such for clarity. Note that the deviation of the signal from pure oscillation increases with ζ.

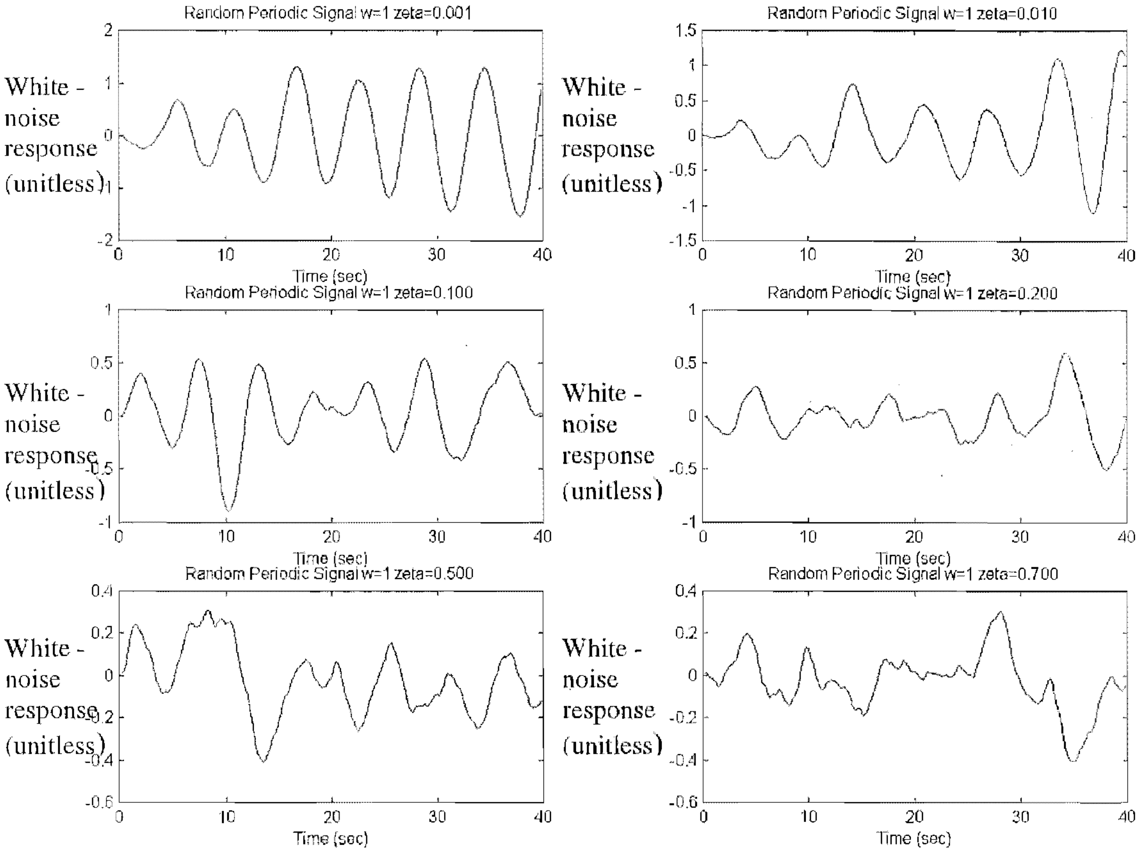


Figure 5.16 Signals generated by second order systems with ω=1 and ζ=0.001, 0.01, 0.1, 0.2 0.5 and 0.7 (from top left to bottom right) when input is unity variance white noise.

In the frequency domain, the development of the equations is entirely equivalent to those in the time domain to the point where the autocorrelation sequence of the system output is shown to be equal to the autocorrelation of the system's impulse response (5.10):

$$R_{yy}(m) = \sigma c(m) \tag{5.14}$$

Taking the Fourier transforms of both sides leads to [54]

$$P_{yy}(e^{j\omega}) = \sigma|H(e^{j\omega})|^2 \tag{5.15}$$

where $P_{yy}(e^{j\omega})$ is the power spectral density of the output sequence and $H(e^{j\omega})$ is the system's frequency response function. This result shows that the spectral distribution of power in the output sequence is exactly the same as that of the square of the system's amplitude frequency response function (excluding the multiplicative constant of the noise power). Since it is desired that the output signal be periodic, the bulk of the power of the output signal should be centred at the desired frequency. This indicates that there should be a peaking behaviour in the system's frequency response. Again this leads to the choice of a second order function, since it is a low order system that exhibits a peak in the frequency domain.

In the time domain ζ was seen to have the role of roughly indicating the randomness of the signal. In the frequency domain the parallel is that ζ controls the peak of the frequency response function. The smaller the value of ζ, the higher and narrower the peak is, and thus the smaller the range of the 'pass band'. Note also that the choice of ζ impacts upon the power of the generated signal since it controls the height of the peak at ω. The basic trend is that the lower the value of ζ, the higher the power of the generated signal. This is evident in Figure 5.16, where the signals with low ζ are seen to have amplitude higher than that of those with high ζ. In the time domain, the impulse response of a system with low ζ lasts a lot longer than a system with high ζ. This means that the mean-squared value of the impulse response is higher for a system with low ζ. This effect was masked in the plots of the impulse response autocorrelation functions because all the mean-squared values were normalised to more clearly show the effects of ζ on the decay of the autocorrelation. The amplifying role of ζ is secondary to that controlling the randomness (or equivalently the spectral composition) of the generated signal, and can easily be cancelled out by altering the power of the input sequence.

It should be noted that below the natural frequency of the system, the spectrum, while not amplified to the same extent as the 'pass-band', is not attenuated and thus these low frequency components will always be present in the output. Plots of $|H(e^{j\omega})|^2$ for the previously considered values of ω and ζ are shown in Figure 5.17.

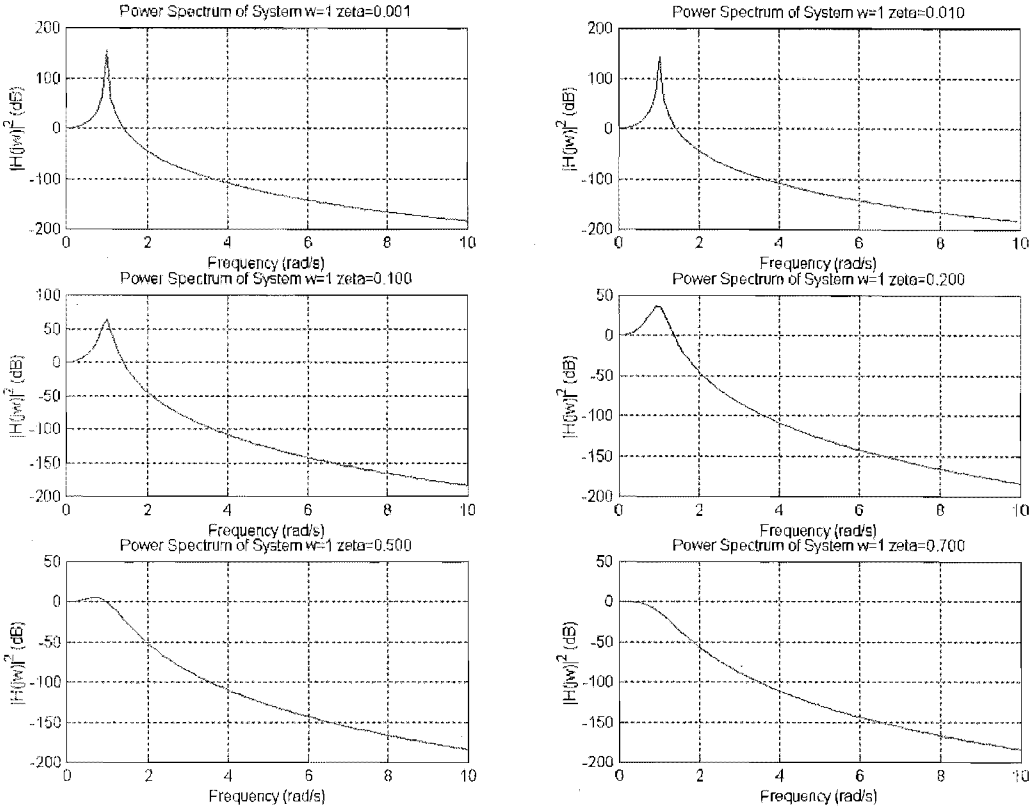


Figure 5.17 Second order system power spectrums for ω=1 and ζ=0.001, 0.01, 0.1, 0.2 0.5 and 0.7 (from top left to bottom right).

Since the model is to eventually be transformed into a discrete time representation, it seems sensible that the full design procedure take place in this format. Toward this aim, the following result was attained. (The methods described above, i.e. the autocorrelation and frequency domain methods, are well known in the literature, whereas the author derived the following method.) This method is not completely useful as a design tool since it involves the solution of a discrete Lyapunov equation. While this makes the technique less useful for design, it can be used as a simulation tool, as it does show the evolution of the state autocorrelation matrix. This is advantageous since the cross-correlation between the states of the system is shown. Also, the possibility of multiple inputs is covered easily. This is the situation considered in the development below. Consider the discrete time state space model

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{w}_{k-1}, \qquad (5.16)$$

where $\{\mathbf{w}_k\}$ is a zero mean stationary white noise vector sequence with covariance matrix $\mathbf{Q}$. (The time index is indicated by the subscript $_k$ to make the presentation more clear.) Since the stationary sequence is being processed by a linear time invariant system, $\mathbf{x}_k$, is also a stationary process. By finding the expression for the state at $m>k$-1 in terms of the state at k-1, e.g.

$$\begin{aligned}\mathbf{x}_k &= \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{w}_{k-1}\\ \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{w}_k = \mathbf{A}^2\mathbf{x}_{k-1} + \mathbf{A}\mathbf{B}\mathbf{w}_{k-1} + \mathbf{B}\mathbf{w}_k\\ \mathbf{x}_{k+2} &= \mathbf{A}\mathbf{x}_{k+1} + \mathbf{B}\mathbf{w}_{k+1} = \mathbf{A}^3\mathbf{x}_{k-1} + \mathbf{A}^2\mathbf{B}\mathbf{w}_{k-1} + \mathbf{A}\mathbf{B}\mathbf{w}_k + \mathbf{B}\mathbf{w}_{k+1} \qquad (5.17)\end{aligned}$$

The following general expression is found:

$$\mathbf{x}_{k+m} = \mathbf{A}^{m+1}\mathbf{x}_{k-1} + \mathbf{A}^m\mathrm{B}\mathbf{w}_{k-1} + \mathbf{A}^{m-1}\mathbf{B}\mathbf{w}_k + \ldots + \mathbf{A}\mathbf{B}\mathbf{w}_{k+m-2} + \mathbf{B}\mathbf{w}_{k+m-1} \quad (5.18)$$

The correlation between the state at k and the state at k+m is given by

$$E[\mathbf{x}_k\mathbf{x}_{k+m}^{\mathrm{T}}] = E[(\mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{w}_{k-1})(\mathbf{A}^{m+1}\mathbf{x}_{k-1} + \mathbf{A}^m\mathbf{B}\mathbf{w}_{k-1} + \ldots + \mathbf{A}\mathbf{B}\mathbf{w}_{k+m-2} + \mathbf{B}\mathbf{w}_{k+m-1})^{\mathrm{T}}] \quad (5.19)$$

This can be simplified using the fact that $\mathbf{x}_{k-1}$ is uncorrelated with $\mathbf{w}_j\, j \geq k$-1, and since $\{\mathbf{w}_k\}$ is a zero mean sequence, i.e. $E[\mathbf{x}_{k-1}\mathbf{w}_j^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}] = E[\mathbf{x}_{k-1}]E[\mathbf{w}_j^{\mathrm{T}}]\mathbf{B}^{\mathrm{T}} = 0\, j \geq k\text{-}1$. Thus

$$\begin{aligned}E[\mathbf{x}_k\mathbf{x}_{k+m}^{\mathrm{T}}] &= E[(\mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{w}_{k-1})(\mathbf{A}^{m+1}\mathbf{x}_{k-1} + \mathbf{A}^m\mathbf{B}\mathbf{w}_{k-1})^{\mathrm{T}}]\\ &= E[\mathbf{A}\mathbf{x}_{k-1}\mathbf{x}_{k-1}^{\mathrm{T}}(\mathbf{A}^{m+1})^{\mathrm{T}} + \mathbf{B}\mathbf{w}_{k-1}\mathbf{w}_{k-1}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}(\mathbf{A}^m)^{\mathrm{T}}] \qquad (5.20)\end{aligned}$$

Denoting the correlation matrix of the state vector, $\mathbf{x}$, by $\mathbf{R}$, e.g. $\mathbf{R}(k,k+m) = E[\mathbf{x}_k\mathbf{x}_{k+m}^{\mathrm{T}}]$, and the correlation matrix of the noise process by $\mathbf{Q}$ (which is equivalent to the covariance matrix since the noise sequence is zero mean):

$$\mathbf{R}(k,k+m) = \mathbf{A}\mathbf{R}(k\text{-}1,k\text{-}1)(\mathbf{A}^{m+1})^{\mathrm{T}} + \mathbf{B}\mathbf{Q}\mathbf{B}^{\mathrm{T}}(\mathbf{A}^m)^{\mathrm{T}} \quad (5.21)$$

Since the state vector sequence is wide sense stationary, the autocorrelation function should be a function only of the difference between the two arguments leading to the following statement

$$\mathbf{R}(m) = \mathbf{AR}(0)(\mathbf{A}^{m+1})^{\mathrm{T}} + \mathbf{BQB}^{\mathrm{T}}(\mathbf{A}^{m})^{\mathrm{T}} \tag{5.22}$$

In particular, for m=0,

$$\mathbf{R}(0) = \mathbf{AR}(0)\mathbf{A}^{\mathrm{T}} + \mathbf{BQB}^{\mathrm{T}}, \tag{5.23}$$

which is a Lyapunov equation, the solution of which is the 'mean squared matrix' **R**(0). The solution **R**(m) can be expanded into the form

$$\mathbf{R}(m) = [\mathbf{AR}(0)\mathbf{A}^{\mathrm{T}} + \mathbf{BQB}^{\mathrm{T}}](\mathbf{A}^{\mathrm{m}})^{\mathrm{T}}, \tag{5.24}$$

which is seen to be equivalent to

$$\mathbf{R}(m) = \mathbf{R}(0)(\mathbf{A}^{m})^{\mathrm{T}}. \tag{5.25}$$

The matrix **R**(m) has the following structure for a two state system:

$$\mathbf{R}(m) = \begin{bmatrix} R_{11}(m) & R_{12}(m) \\ R_{21}(m) & R_{22}(m) \end{bmatrix} \tag{5.26}$$

where $R_{12}(m)$ is the cross-correlation of the states at a difference of m. There are situations in which it may be of benefit to design states that are uncorrelated, and using this method, this may be possible, although knowledge of the behaviour of the Lyapunov equation would be required. The evolution of the autocorrelation matrix of the second order system for the previously covered values of ω and ζ are shown in Figure 5.18.

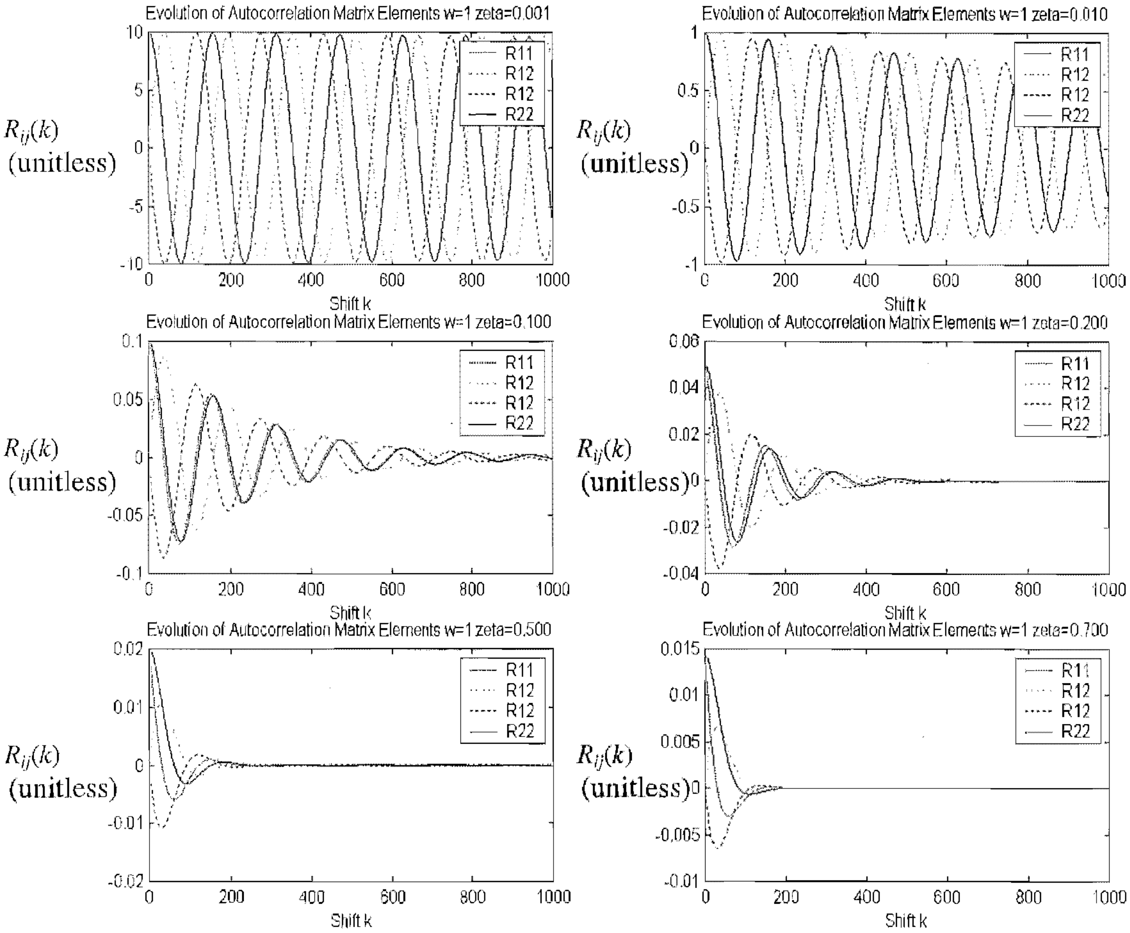


Figure 5.18 The evolution of the elements of the state autocorrelation matrices of second order systems in response to white noise excitation.

In Fig 5.18, the autocorrelations have not been normalised, and thus it is possible to see the effect of ζ on the power of the output sequence. Note the periodic nature of the cross correlations in the highly oscillatory cases. The low value of cross correlation at zero shift, combined with the similar autocorrelation functions show that the two states are close to being shifted versions of one another.

Having considered the task of designing a shaping filter using both time and frequency domain concepts, and identifying a likely model, the generic second order system, this shaping filter is now added into the continuous time system equation, i.e. the a's and g's of the augmented state space model are determined, at least parametrically:

$$\begin{bmatrix} \dot{d} \\ \dot{v} \\ \dot{a} \\ \dot{\alpha} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\omega_n^2 & -2\zeta\omega_n \end{bmatrix} \begin{bmatrix} d \\ v \\ a \\ \alpha \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} w \tag{5.27}$$

As previously mentioned, the value of ζ is chosen to represent the degree of variability in the signal, whereas ω_n is the base frequency of oscillation, in rad/s.

The continuous time model is discretised to give

$$\mathbf{x}_{k+1} = \Phi\mathbf{x}_k + \Gamma w_k \tag{5.28}$$

where Φ and Γ do not have elements with closed form. The only measurement actually yielded by the encoder is the displacement, i.e.

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \tag{5.29}$$

and this measurement is contaminated by the effects of quantisation noise. This is represented by the white sequence $\{v_k\}$. Two alternative filter structures were tested. The first artificially generated velocity and acceleration signals by the method described below. The second used only the position measurement. This artificial generation of measurements is the second departure from previous works.

At each instant the filter generates an estimate of the state vector that includes as its elements position, velocity and acceleration. It was thought that these estimates could be used in conjunction with the measured data to create reasonable estimates of velocity and acceleration that could then be treated as measurements. After collecting each position measurement, therefore, the following calculations were made

$$\begin{aligned} \mathbf{z}_k(2) &= \frac{1}{T}\left[\mathbf{z}_k(1) - \hat{\mathbf{x}}_{k-1}(1)\right] \\ \mathbf{z}_k(3) &= \frac{1}{T}\left[\mathbf{z}_k(2) - \hat{\mathbf{x}}_{k-1}(2)\right] \end{aligned} \tag{5.30}$$

where: $\hat{\mathbf{x}}_{k-1}(1)$ is the first element of the estimated state vector from the (k-1)th instant $\mathbf{z}_k(1)$ is the measurement of position at the k^{th} instant, $\mathbf{z}_k(2)$ and $\mathbf{z}_k(3)$ are the 'measurements' of velocity and acceleration at the k^{th} instant and T is the sampling period. Using three measurements, of course, necessitated a new **C** matrix

$$\mathbf{C}_{\mathrm{new}} = [\mathbf{I}_3 \ \mathbf{0}_{3\mathrm{x}1}] \qquad (5.31)$$

as well as a 3x3 sensor noise error covariance matrix **R**, where in the single measurement case, a scalar, R, was used. In assigning values to the matrix, **R**, it was assumed that the 'noises' on the true and synthesised measurements were uncorrelated, resulting in a diagonal matrix. In reality the noises are related due to the way in which the synthesised measurements are calculated, but the aforementioned assumption means that **R** is always non-singular, and hence no problems are encountered during matrix inversion. A diagonal form also obviously makes the specification of **R** much easier.

To compare the two filter designs, they were run concurrently, i.e. on the same noisy simulated position measurements, for a range of different data sets. In all cases the 'true' position was a sum of trigonometric functions so that the derivatives could be known exactly. Estimation errors for the position, velocity and acceleration were collected in vectors during the operation of the filters. At the conclusion of the filter operation, the sum of the norms of the three estimation-error vectors for both of the filter designs were calculated. These values were used to tune the filter, i.e. pick values for **R** (or R) and Q to minimise the norms, and also to compare the relative efficacy of the two designs. It was found that while the synthesised data KF could perform better than the single measurement KF, the difference in performance was made small when the assumed covariance of the position measurement noise was small. It was also found that the filters both exhibited good characteristics with this value chosen to be very small (R~0.0001).

5.2.3 Results from Simulated and Experimental Data

This section first presents results showing the effectiveness of the designed Kalman filter on discretised versions of signals that have analytic derivatives in additive noise. This was done so that the approximated derivatives could be compared with the ideal. These results are followed by a representative sample of the collected encoder signal and the estimated derivatives.

The first simulated position signal is $d(t) = 10\sin(t) + \sin((1.5)t)$. Added to the position signal is a Gaussian white noise sequence with variance 0.01. The results below used a Kalman filter with $\omega = 1$, $\zeta = .01$, $R = 1\text{e-}5$ and $Q = 0.05$. The value of ω was chosen since the main part of the signal is 'unit' frequency, while ζ was chosen to be small since the variation in the signal was expected to be small (actually zero in this case). The numerical values of R and Q were found by trial and error. The 'correct value' for R, based on the variance of the measurement noise is 0.01, but in actual operation R, like Q becomes a tuning factor. When 'tuning the filter' heuristically, increasing Q relative to R instructs the filter to weight the measurements more heavily than the propagated estimate and vice versa. Placing too small a weighting on Q (i.e. a relatively large number for R since R represents the strength of the noise on the measurement) makes the filter 'lose track' and a large lag is introduced. Conversely if too small a value is used for R, the measurements are 'trusted' almost entirely and the resulting estimates are not much better than those obtained by using finite-differencing on the noisy measurement.

Figure 5.19 shows the position measurement used in the first trial along with the results of finite-differencing. Note that the positions measurement signal appears to be 'clean' but the noise is significantly magnified when numerical differentiation is used.

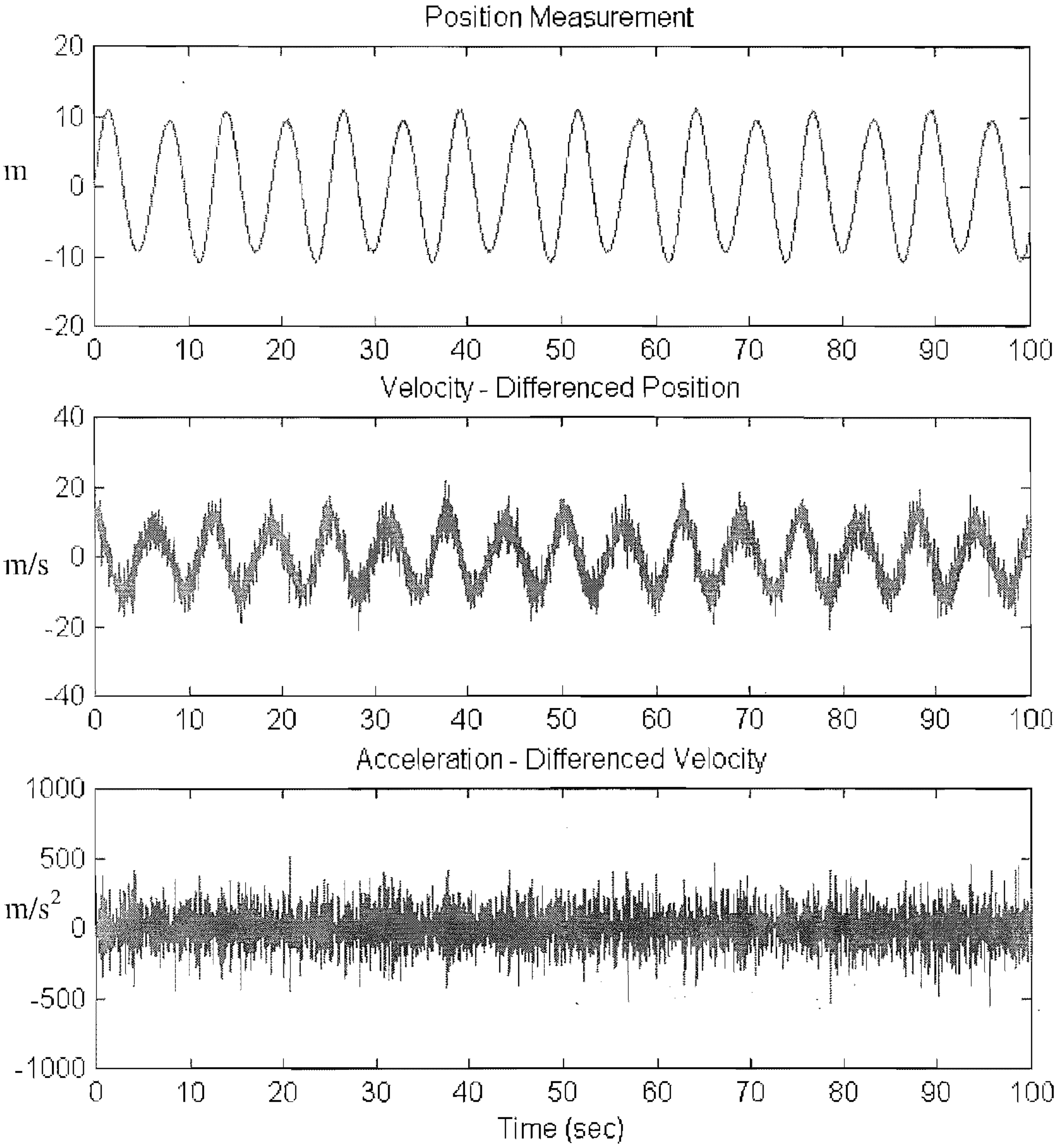


Figure 5.19 The position 'measurement' $d(t)$ = 10sin(t) + sin((1.5)t) in additive white noise of variance = 0.01, and the results of the finite-differencing procedure.

The results corresponding to those of Figure 5.19, for the Kalman filter are shown in Fig. 5.20. The derivatives are clearly of much higher accuracy than those obtained by finite-differencing. A point of interest is the initial fluctuation in the filter outputs observable in the velocity and acceleration plots. This behaviour is due to erroneous initial estimates (position, velocity and acceleration were assumed to initially be zero), and a high value chosen for the state error covariance matrix.

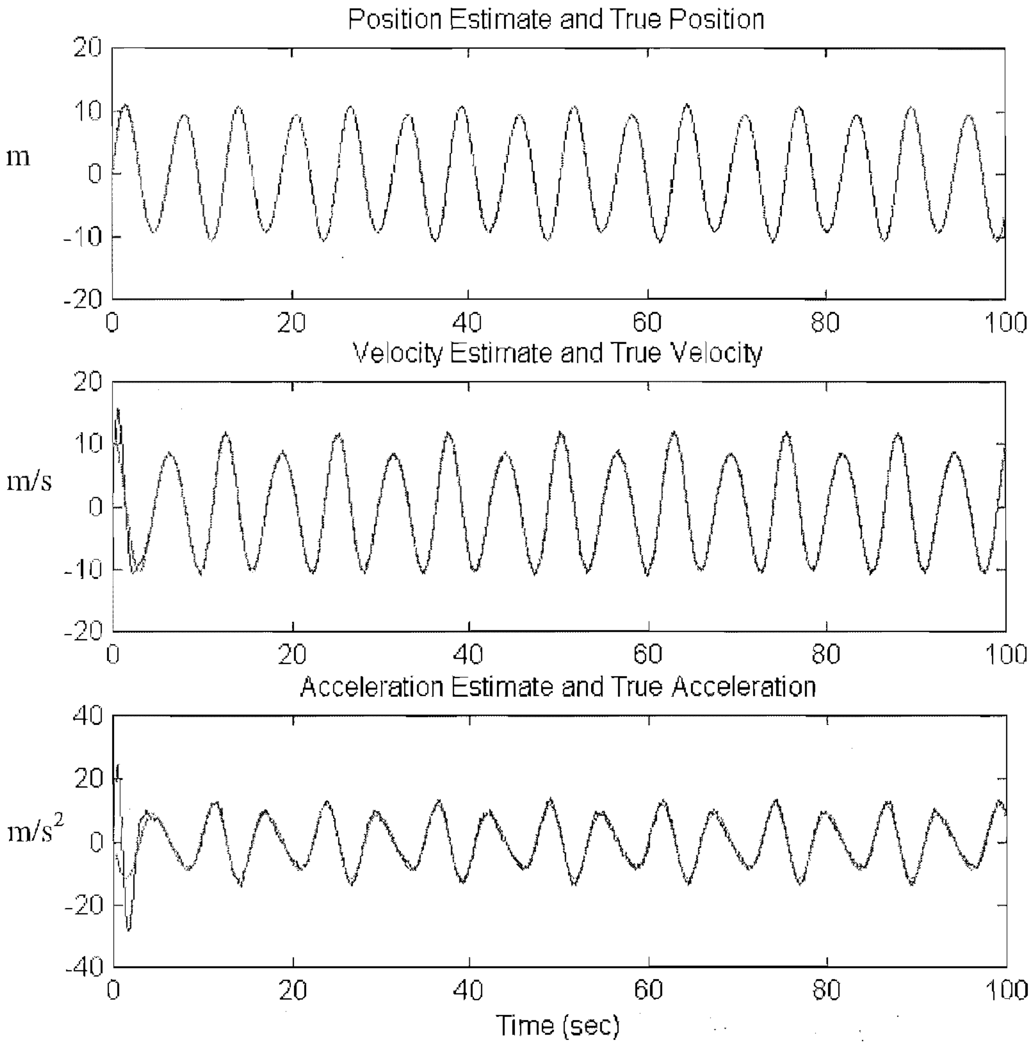


Figure 5.20 The position, velocity and acceleration estimates from the Kalman filter for the measurement $d(t) = 10\sin(t) + \sin((1.5)t)$ in additive white noise of variance = 0.01. The true values are shown on the same plots.

These results show that the filter works for simulated signals. The design method is slightly different when the filter is used on real displacement data. While the seat *displacement* appears to be almost sinusoidal with a base frequency of the stroke rating, it is the acceleration that is to be modelled by the output of the shaping filter, and since the displacement actually consists of a wide range of frequencies (consider a Fourier series of the seat displacement), the higher components of which are magnified by the process of differentiation, the shaping filter must be designed so that these high frequency components are not lost. This magnification of higher frequency components by differentiation can be mathematically displayed very simply. If a Fourier series expansion of the seat displacement is considered, there will be an infinite sum of terms of the form $A_i\cos(\omega_i t + \phi_i)$, the derivative of which is $\omega_i A_i \sin(\omega_i t + \phi_i)$, and for $\omega_i > 1$ this leads to a frequency component of greater magnitude than that in the original signal. The parameter that sets the frequency band of the filter, ω, must be chosen so that it includes all the 'frequency information' of the seat acceleration, without allowing undue measurement noise through. As an aside, it is because white noise theoretically has a flat spectrum, i.e. it contains all frequencies in equal quantities, and therefore includes very high frequencies, that signals contaminated with white noise have such poor signal to noise ratios when they are differentiated.

A method of designing the filter taking into account the above comments would be to find, via an FFT, the frequency spectrum of a representative seat displacement, paying particular attention to the highest frequency component of any great magnitude, and then setting ω to be this value. The method employed here, however, is much more qualitative; the derivatives obtained using the filter are compared to those using finite differencing, and ω is varied so that the filter output matches the gross variations in the finite difference data, while rejecting the visible noise. The value chosen for ω, is better to be chosen slightly too high rather than too low, since if ω is set too low, valid oscillations within the derivatives are smoothed out, low frequency oscillations are introduced where there should be none, and significant delay is introduced. As long as ω is not set at an unrealistically high level, the filter yields a frequency-limited output that has very little delay. Obviously setting ω far too high allows a large amount noise through the filter, and the results are degraded. The output of the filter is shown in Figures 5.21 & 5.22 for one rower, with $\omega = 1$ and 4 ($\zeta = .0001$)

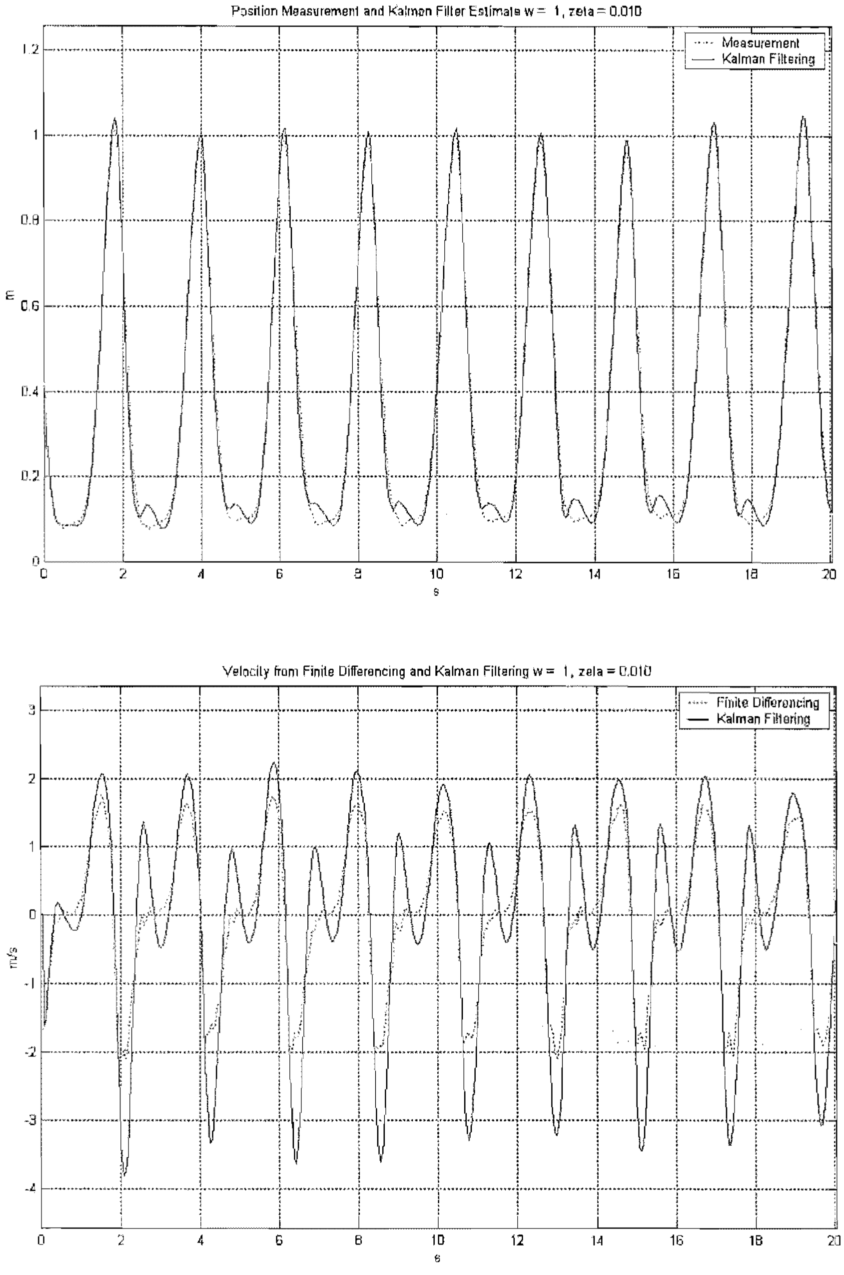


Figure 5.21(a)&(b) Kalman filter displacement (a) and velocity estimates (b) when ω = 1 (too low). Note the large erroneous oscillations in the estimates.

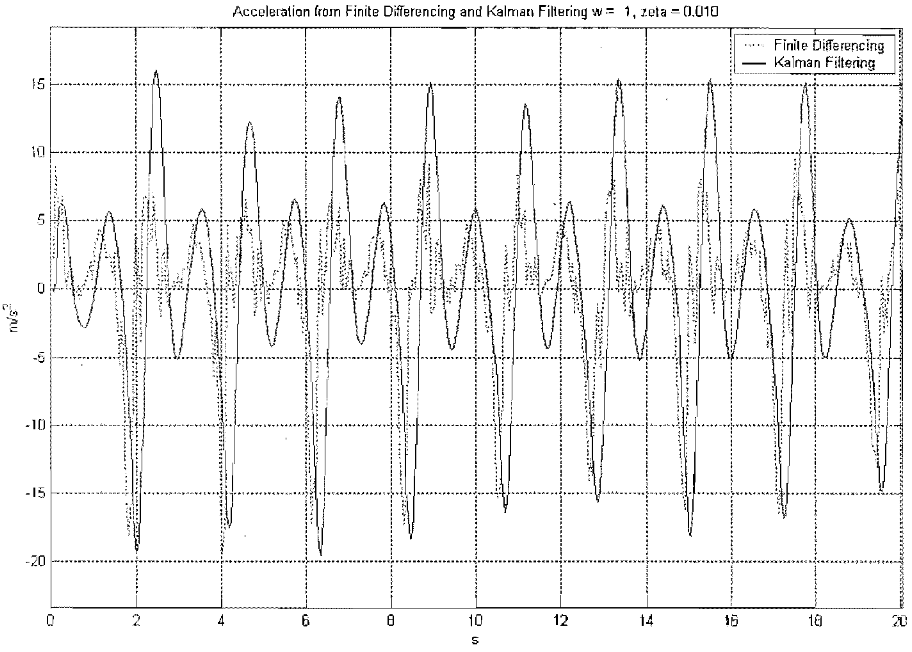


Figure 5.21(c) Acceleration estimates by finite differencing and Kalman filtering with ω = 1. The finite differencing estimate is very noisy, and the Kalman filtering estimate has a large delay and is overly smoothed.

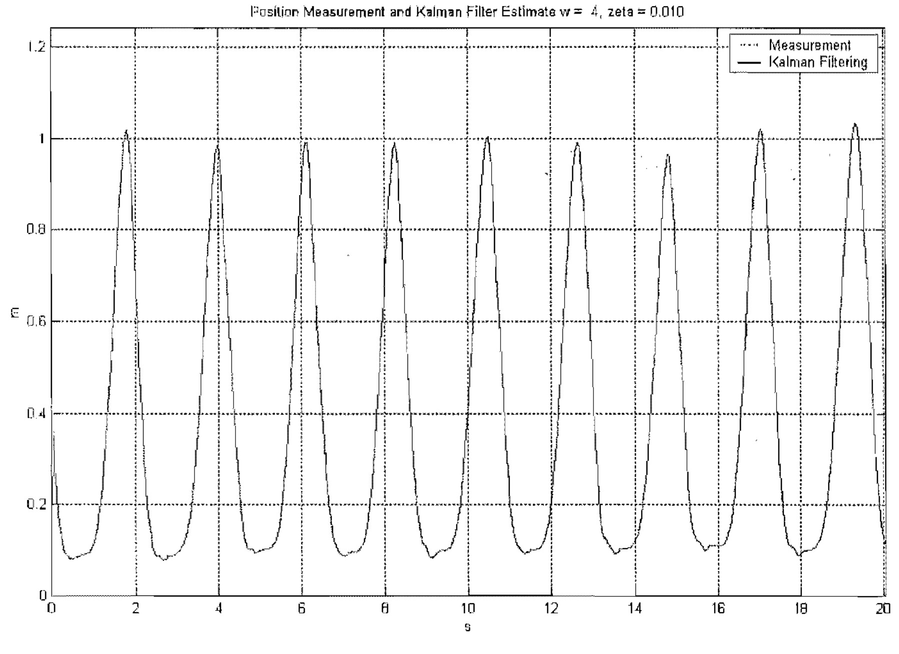


Figure 5.22(a) The Kalman filter displacement estimate is almost indiscernible from the measurement for ω = 4.

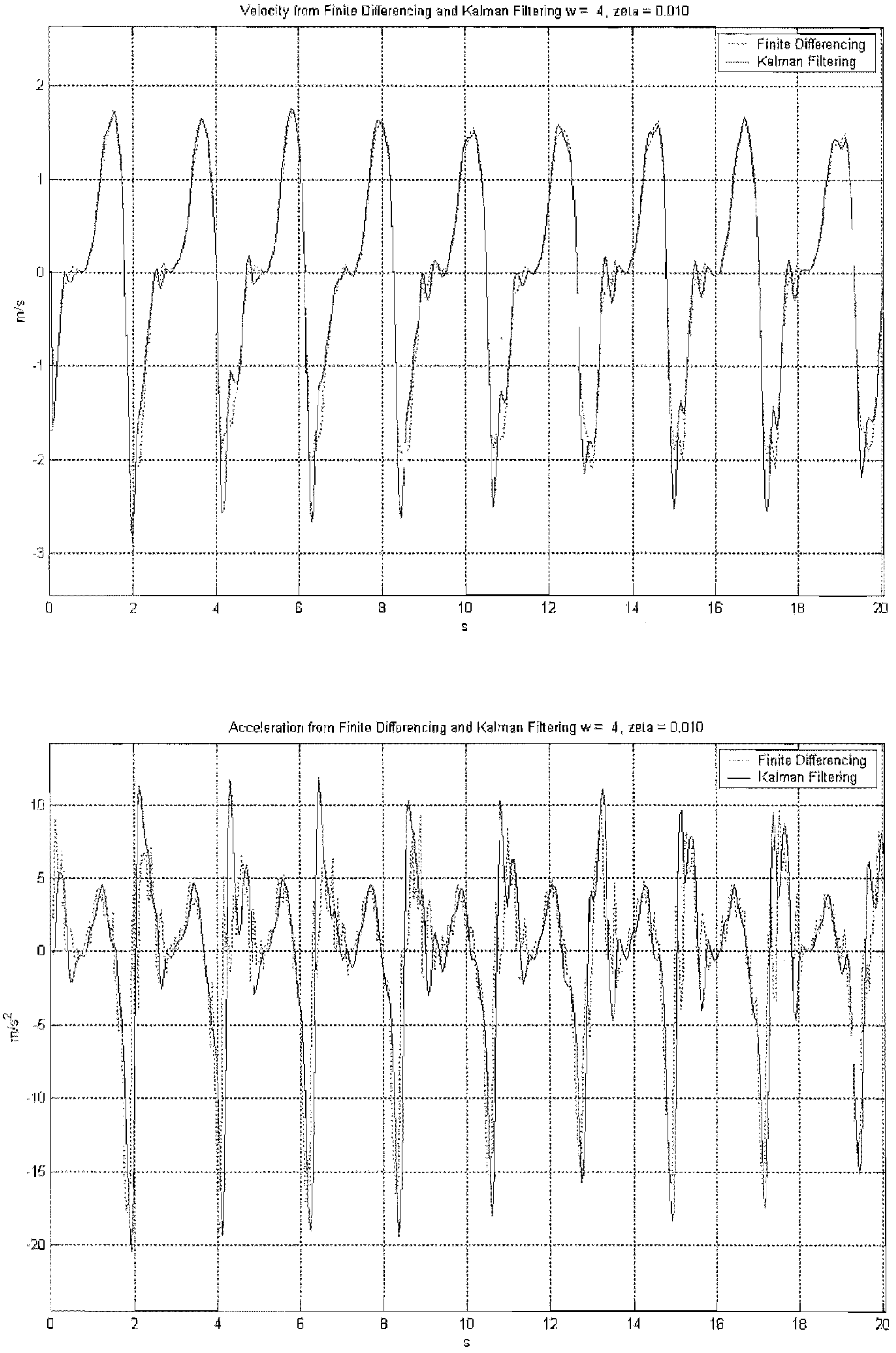


Figure 5.22(b)&(c) The velocity (b) and acceleration (c) estimates generated by finite differencing and Kalman filtering for $\omega = 4$.

Figures 5.22(b)&(c) show pleasing results. There is very little delay between the finite difference and Kalman filter generated estimates while the Kalman filter estimates clearly have a higher signal to noise ratio. It should be noted that the only reason that this method of filter tuning was possible is that the noise level in the displacement measurement was very low. As previously mentioned, the noise level is a function of the displacement, the number of pulses per revolution of the encoder, and the sampling rate that is used. The acceleration estimate generated by finite differencing shows the high frequency magnification characteristic of the process of differentiation, but gross shapes within the signal are still easily discernable, meaning that it is possible to qualitatively tune the Kalman filter to mimic the true acceleration. If this were not the case, i.e. the noise level on the displacement measurement was higher, then it would be necessary to resort to finding the frequency spectrum of a typical seat displacement signal, and tuning the filter based on these results.

As discussed at the very beginning of this section, it is not believed that the Kalman filter is generating an optimal estimate of the kinematical variables of the seat displacement for a number of reasons. Firstly, the simple model used within the shaping filter to generate a signal 'somewhat like' acceleration and its derivative, while certainly a better fit than the reviewed Gauss-Markov variants, will not be exact; it was chosen as a trade-off between simplicity and goodness of fit. Secondly, the fact that the quantisation noise of the encoder is not Gaussian white noise will cause it to function in a sub-optimal manner. The Kalman filter can be derived under the assumption of Gaussian white noise processes, in which case the filter is optimal with respect to a large range of cost functions, or a general white noise, in which case it is optimal with respect to a quadratic cost function. The smaller the quantisation level of the encoder the less effect the non-white noise is likely to have. Even with these two caveats, the above results show that the Kalman filter performed well. An insight into a method by which the Kalman filter can be used as a frequency limited differentiator was also gained.

Up to this point, while this section is concerned with the measurement of the motion of the system centre of mass, only the movement of the seat, which approximates that of the rower's centre of mass has been measured. The other significant component of the system is the boat, the motion of which is now considered.

5.3 Measurement of Boat Motion

Before rowing data was captured it was believed that the motion of the boat would have to be considered as two dimensional, with considerable pitching and dipping components. This made the analysis more challenging, and if it were the case, the measurement of boat motion would surely have received a chapter all of its own!

However, collected data showed that tilt was only on the order of a few degrees, meaning that to estimate the boat motion as being one-dimensional was acceptable, and hence the analysis was considerably simplified.

This section reviews the few methods that have previously been applied to measure the motion of a rowing shell before explaining the choice of sensors that were used in this work. A pair of sensors was used to estimate instantaneous boat displacement, velocity and acceleration. A Kalman filter sensor fusion technique, which is a very basic extension of the differentiator developed in 5.2.2 is used to combine the outputs of the two sensors.

5.3.1 Previous Methods

Aside from Martin, who considered the effect of stroke rate on boat velocity through film analysis [50], two sensors have dominated the area of boat motion measurement. These sensors are the accelerometer and the submerged magnetic impeller. These two sensors have very different characteristics and measure the boats motion relative to two different frames, one moving with the water, and the other fixed on the land.

Young and Muirhead [69] used a 1g single axis accelerometer to measure longitudinal acceleration during rowing. Velocity was obtained through integration, but no details are given as to special signal processing measures employed to eliminate drift. In a somewhat more bizarre application of accelerometry, Lin et al [41] fixed accelerometers to various parts of the rowing system, including the seat, oars and the rower's shoulders during ergometer and on-the-water rowing. The intention of Lin's work was to find characteristics of good rowing, i.e. accelerometer outputs yielded by expert rowers, and then compare these results with novices. There is no discussion of any transformation from accelerometer output to physical motion.

The other sensor commonly applied in the measurement of boat motion is a submerged impeller [39], [43], [46], [64]. Each of the references citing the use of these impeller, simply state that they were used, without discussing signal processing or sensor characteristics, thus little was known about their performance. The SpeedCoach is a small commercially available magnetic impeller system that is commonly used in rowing training. A magnet is mounted in the impeller, and there is a coil pickup mounted inside the boat directly above the impeller. The rotation of the magnet creates a current in the coil that is then converted to pulse waveform by a high-gain amplifier acting as a comparator. The faster the magnet spins, the greater the changing flux and the higher the frequency of the output pulse waveform. The overall system has a small signal-processing/display unit that displays information in a variety of forms, including current speed, distance travelled and projected 500m times. The distance is presumably calculated by multiplying the number of pulses counted by a factor, and velocity is estimated by multiplying the number of pulses in a certain period by another constant.

The company who manufacture the SpeedCoach system, Nielsen-Kellerman, insist that the current of the water in which the boat is moving does not affect the distance or speed indicated by the sensor. This strange claim is justified by the statement that it is the motion through the water that is measured, i.e. if a boat is allowed to drift with the current, the sensor will indicate a speed of 0m/s and a distance of 0m. Rowing a distance measured on the land upstream therefore gives a different result from rowing the same distance downstream.

Almost all the previous researches who have used impeller sensors such as the SpeedCoach have also used accelerometers although none state how, or indeed if, the outputs of the two sensors are combined to yield estimates of the boat kinematics. Before the method by which the data is combined in this work is described, the sensors used are briefly described.

5.3.2 Sensors and Sensor Calibration

The two sensors that are used to measure boat motion are the ADXL210 accelerometer (Analog Devices) and the SpeedCoach impeller (Nielsen-Kellerman). The accelerometer, which was used for orientation estimation in Chapter 4, is triaxial, but due to the assumed one-dimensional motion of the boat, only the axis in the longitudinal direction of the boat is used for the measurement of boat motion. (Obviously it is one of the axes of the fixed accelerometer that is used for this purpose.) The method of accelerometer calibration is exactly that described in Chapter 4 (the first method). Now, however, the accelerometer output has a physical meaning, and hence must be associated with units. Therefore, during calibration the gain must be calculated as Volts/[m/s^2].

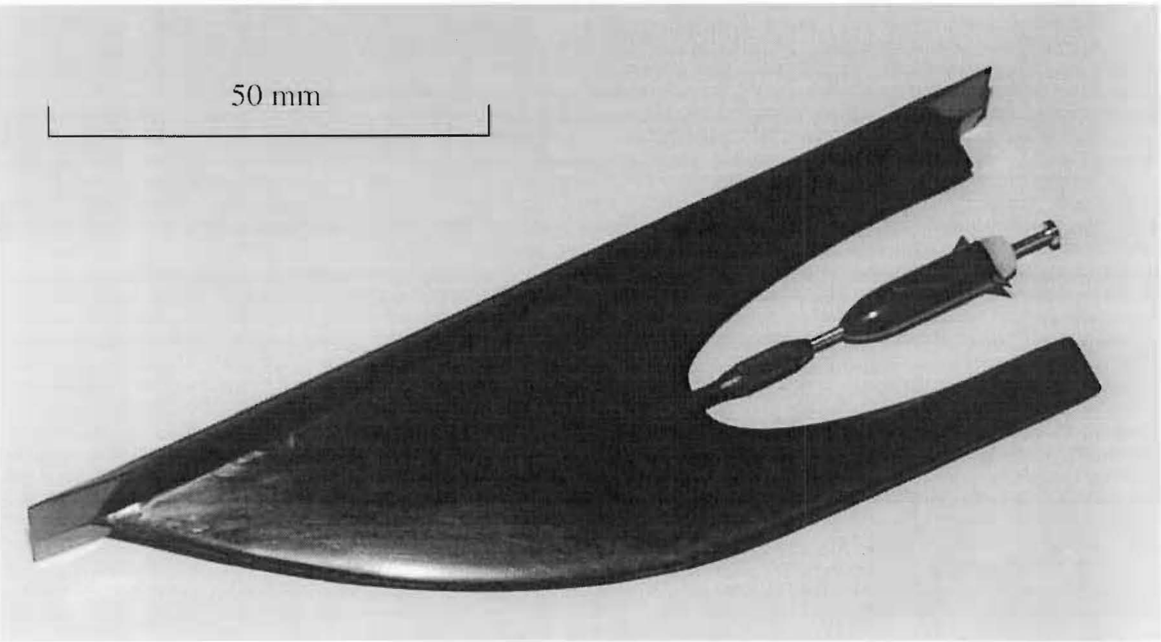


Figure 5.23 The SpeedCoach impeller unit.

A description of the operating principle of the SpeedCoach has already been given. The system as a whole comes with built in calibration factors to convert the incoming pulses to measures of distance travelled and speed. These parameters are assumedly calculated by experimental means and set to whole batches of the sensors as a preshipping operation. It is possible to update the values by rowing with the impeller over a known distance and, at the conclusion of the piece, 'telling' the display/processing unit, via button presses, what the true distance was. The values of the constants are not given with the sensor. Since the impeller was to be used without the display unit, it was required that the values of these constants, or some more accurate set, be determined. This is the process of calibration. The location of the impeller within the boundary layer makes it very difficult to calibrate without actually mounting on a shell. This difficulty is compounded by the fact that the 'true' constants depend to some extent on the location of the impeller on the hull of the boat, i.e. placing the same sensor at different locations on the hull will yield different results. Ideally, the sensor should be calibrated using a reliable external source, such as a radar speed gun. It was intended that such a calibration take place, using the Department's Stalker Radar Gun, but questions to the manufacturers of the device regarding the interface of their product with a general data acquisition system were unanswered. Calibration, therefore was a crude affair, consisting of connecting the SpeedCoach display unit to a signal generator producing a square wave and recording the input frequency and speed indicated on the display. This approach gave a good linear fit (see Fig 5.24) and exposed the factory calibration, but revealed nothing of the accuracy of this calibration. Rowing a known distance in still water and comparing the indicated and true distances could approximately appraise the accuracy.

In operation the SpeedCoach is used as a distance rather than a speed sensor. Since the calculated constant, c, that relates pulse frequency to speed

$$\text{Speed} = c(\text{Pulse Frequency}) \qquad (5.32)$$

Is relating the distance travelled per unit time to the number of pulses counted per unit time, the corresponding relationship for distance is

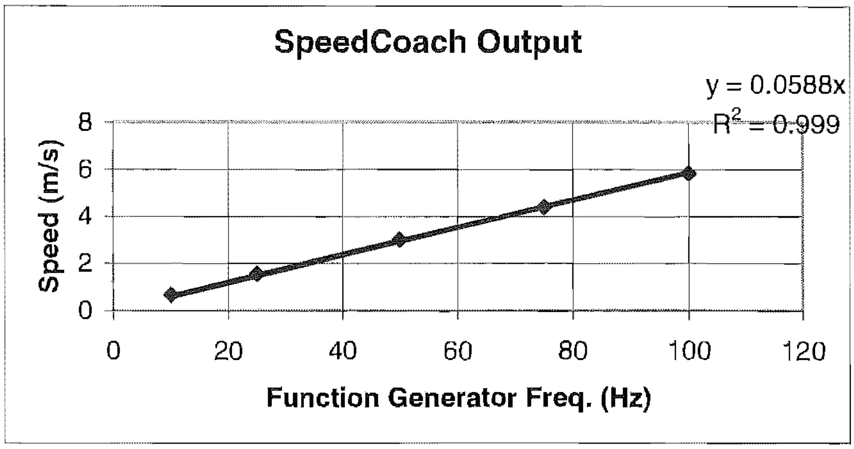


Figure 5.24 The data used to determine the constant relating speed to pulse frequency for the SpeedCoach.

$$\text{Distance} = c(\text{Number of Pulses Counted}) \qquad (5.33)$$

In the following section the means by which the outputs of an accelerometer and a SpeedCoach sensor are combined using a very simple extension of the Kalman filter developed in the previous section is described. This basic method yields apparently sound results. (Apparently, because there is nothing to check the results against, but the waveforms are consistent with those of previous researches.)

5.3.3 Sensor Fusion via Kalman Filtering

The traditional approach to sensor fusion using Kalman filtering techniques has the sensors' error characteristics included within the state as parameters to estimate. One particular technique is to take the difference of two measurements of the same parameter from two sources, say a velocity readings obtained through: integration of an acceleration measurement, and a speed reading from a Doppler radar. The difference between the two measurements will be due to the errors of the Doppler reading, and the integral of the error characteristics of the accelerometer. Thus, the state of the Kalman filter includes the error characteristics of the two measurement sources and is 'fed' by the difference of the two sources. At each instant the estimates of the measurement errors are subtracted from the sensor data to yield better estimates of the body's velocity and acceleration.

This is not the approach taken here, for a number of reasons. Firstly, as in the case of seat motion, the general nature of the boat motion is reasonably well known. The displacement will be a non-decreasing function, the velocity will fluctuate about a positive mean, and for most of the time (once a 'steady-state' has been achieved) the acceleration of the boat will be oscillatory with zero mean. Again, it is not the gross variations that are modelled, but the local oscillations of the waveform. This oscillation lends itself to the same model as was used in the previous section, i.e. the motion of the boat can be modelled along with the characteristics of the sensors. This is not the case in applications such as inertial navigation systems for aircraft where during straight flight there may be no predictable 'dynamics' in the kinematical parameters, i.e. all motion may be due to random factors such as turbulence. For short periods of time, the output of the accelerometer can be modelled to a reasonable degree of accuracy as a quantity directly proportional to acceleration in additive noise, thus if the accelerometer output 'enters' the filter as an acceleration measurement, rather than as integrated velocity or position, there are no error dynamics to model. The goodness of this hypothesis can be tested by observing the output of the accelerometer in response to a known acceleration. The simplest possible case is constant acceleration, i.e. subjecting the stationary accelerometer to some component of gravity. If the error on the signal is indeed 'white', the autocorrelation of the deviation of the signal from the mean will be a 'spike at zero shift' (As can be seen from Fig 5.25 the approximation of the accelerometer error as white noise is justified, assuming of course that the error is independent of the incident acceleration.) Equivalently, the Fourier transform of the output should give a spectrum that is constant at all frequencies. In reality, this will never happen exactly, but if significant deviations occur, this is when modelling of the error characteristic must be used. Typical approaches to use are the applications of Gauss-Markov models, or use of the Yule-Walker equations [32].

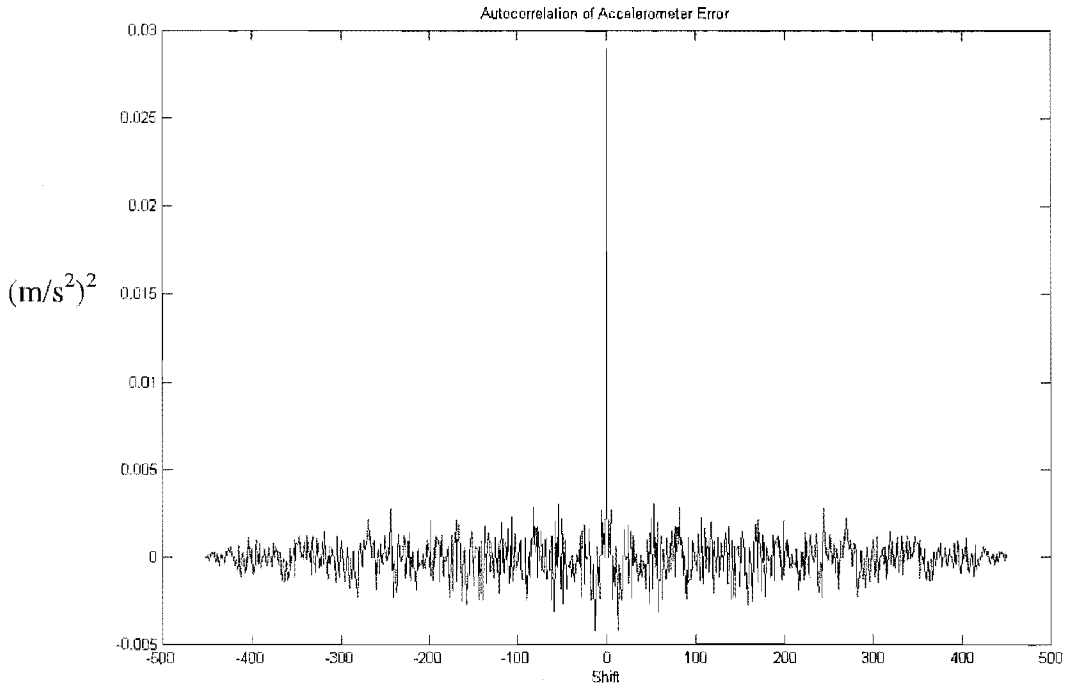


Figure 5.25 Typical autocorrelation of accelerometer error. If the error were totally 'white' the autocorrelation would be a 'spike' at zero shift and zero everywhere else.

The SpeedCoach sensor, as has already been mentioned measures the speed of the boat relative to the water, thus even if the sensor functions perfectly, it will still, in a constant current, have a constant error in speed, and a linearly increasing error in distance. Additional to this error due to the method of sensor operation, rather than any defect, will be the sensor characteristics. If the current is very slow, as it is in most cases where rowing training is undertaken, then the effect may possibly be ignored. Certainly, this should be tried before unnecessarily complicating the filter design. Since pulses are counted every sampling period, as they were in the case of the encoder, there is the possibility of quantisation error, which was modelled as white noise for the encoder. Operating, as it does, on a pulse counting basis, it is difficult to see how any drift in the output could occur, provided the impeller is operating as intended. Thus, under the assumption of no (or low) current, and in possession of no extra information to suggest otherwise, the SpeedCoach sensor can also be modelled as a measurement of distance travelled in combination with white noise. It would not be a simple task to model the error characteristics of the SpeedCoach sensor, this is because of its place within the turbulent boundary layer of the boat, a location that also prohibited any real calibration, as previously discussed.

In addition to the easily modelled variation of boat motion and the lack of real dynamics for the sensor error characteristics, the situation is simplified by the short timescales over which the system will be used. Whereas over many hours an accelerometer may heat and therefore have its offset change, it is expected that this system will only be used for short periods of time. The conditions in which the accelerometer will be operating are also relatively standard. The temperatures will not be extreme (rowing is not possible when the water is solid!) and the accelerometer is located within an enclosure that is itself sheltered from radiant and convectional effects. The combination of these factors mean that the Kalman filter for sensor fusion may be of very simple design.

Just as the motion of the seat was modelled as a periodic process, the motion of the boat, which is of course affected by the periodic fluctuation of oar force and seat motion, is also modelled as an oscillating random process. Thus, the system matrices, the system transition and noise input matrices, remain parametrically similar, i.e. in continuous form the state space equations can again be written

$$\begin{bmatrix} \dot{d} \\ \dot{v} \\ \dot{a} \\ \dot{\alpha} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\omega_n^2 & -2\zeta\omega_n \end{bmatrix} \begin{bmatrix} d \\ v \\ a \\ \alpha \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} w \qquad (5.34)$$

where ω_n is the base frequency of oscillation (in radians/sec), and ζ is chosen to reflect the variability of the signal. The difference is in the measurements, where previously only one measurement was available, there are now two, distance and acceleration, each with associated additive white noise. Thus the measurement equations are

$$\mathbf{z} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} d \\ v \\ a \\ \alpha \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \qquad (5.35)$$

where ν_1 is the noise associated with the SpeedCoach furnished position measurement, and ν_2 is that of the accelerometer. It is assumed that the two noise processes are independent, and thus the covariance matrix, $\mathbf{R}$, is diagonal, with elements R_1 and R_2. As before, the variance of the system noise is Q. The design parameters and their criteria for selection for the filter are therefore:

ω_n – chosen to reflect the highest frequency component of the boat motion

ζ -controls the width of the frequency band of the modelled process, or equivalently the variability of the motion from oscillation to oscillation

Q – has dual purpose. Firstly it ensures that the model has enough power to simulate a motion of the correct amplitude, and secondly 'tells' the filter how much to believe the model in comparison to the measured data.

R_1,R_2 – tell the filter how reliable each of the sensors is at each of their tasks, e.g. a large value of R_2 (relative to R_1 and Q) suggests that the accelerometer measurements are very noisy, so that the model, and the derivative of the position measurements are weighted more heavily.

5.3.4 Results

This section presents results from the Kalman filter sensor fusion technique. To show the utility of this method, the signals obtained from direct numerical integration and differentiation of the appropriate signals are also given.

Figure 5.26 shows the data obtained from the SpeedCoach impeller, which, as mentioned above, was used to measure distance travelled. Also shown are the estimates of boat velocity and acceleration obtained using finite differencing. A large amount of noise is present here due to the quantisation involved in the measuring system, i.e. a finite number of pulses occur during a sampling period. For comparison, the output of the accelerometer is shown in the same plot as the acceleration estimate.

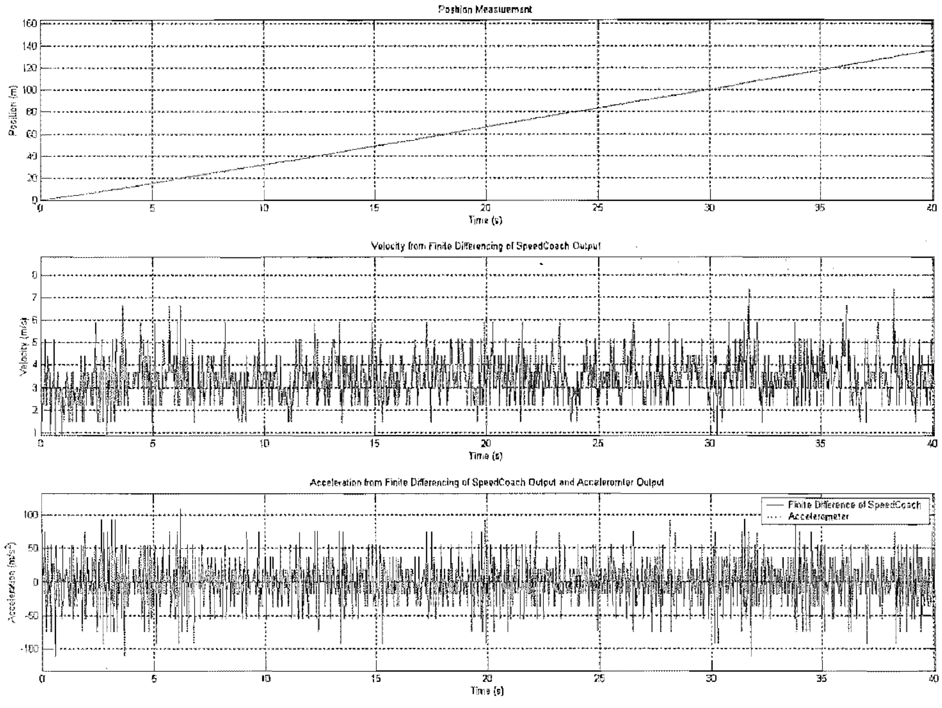


Figure 5.26 The distance output of the SpeedCoach impeller and the velocity and acceleration estimates obtained through finite differencing.

Shown in Figure 5.27 are the accelerometer output and the velocity and position estimates obtained through numerical integration. Shown for comparative purposes is the position measurement obtained from the SpeedCoach sensor. The large amount of drift in the velocity and position estimates belie the amount of noise on the signal, and perhaps also a DC offset in the accelerometer, which could very easily be caused by the accelerometer axis being subjected to gravity.

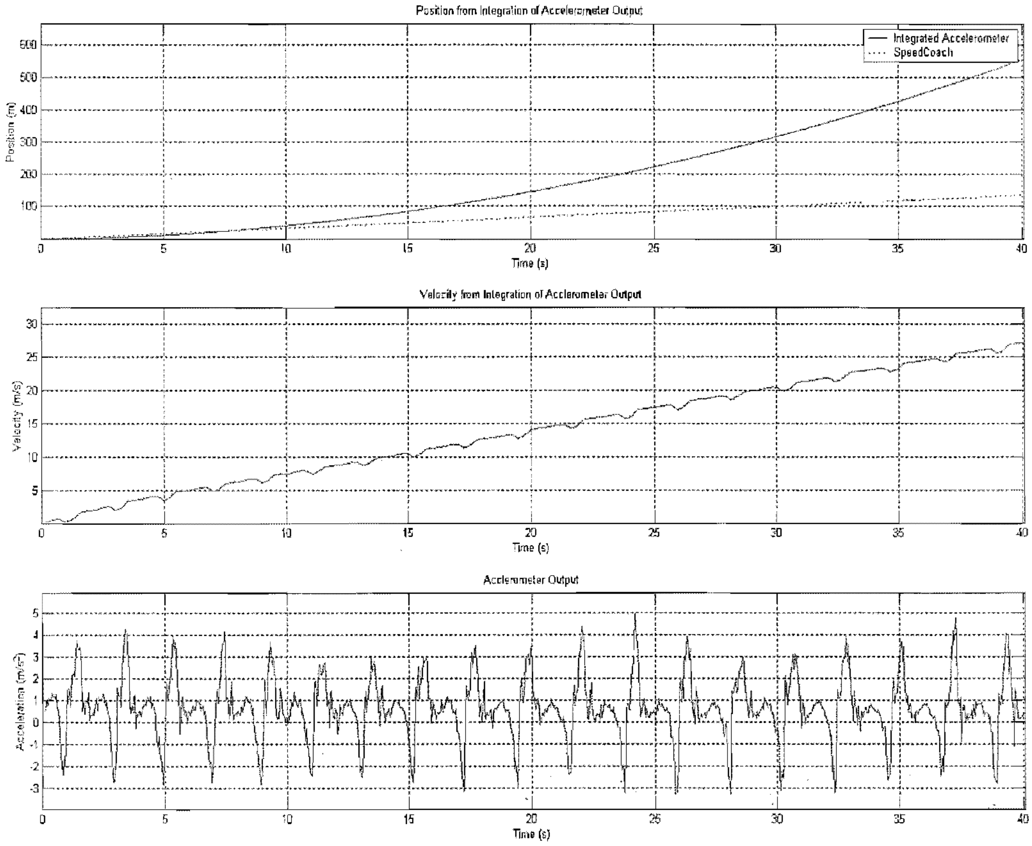


Figure 5.27 The accelerometer output and the velocity and position estimates obtained through numerical integration. The SpeedCoach position estimate is shown to show the degree of drift in the integrated accelerometer output.

Figures 5.26 & 5.27 have shown that, as implemented, neither of the sensors can by themselves reliably record the kinematic parameters of the boat. As already discussed in detail, the Kalman filter combines the sensors' outputs to produce 'optimal' estimates of the parameters of interest.

The figures below show what are very believable results. First, in Figure 5.28, the acceleration estimate of the Kalman filter is plotted with the accelerometer output. They clearly match each other well. Any lag in the KF estimate at this stage would suggest that the frequency of the model needs to be increased, and in fact, this delay was used as a measure of adequacy of the filter frequency (as discussed in 5.2.3).

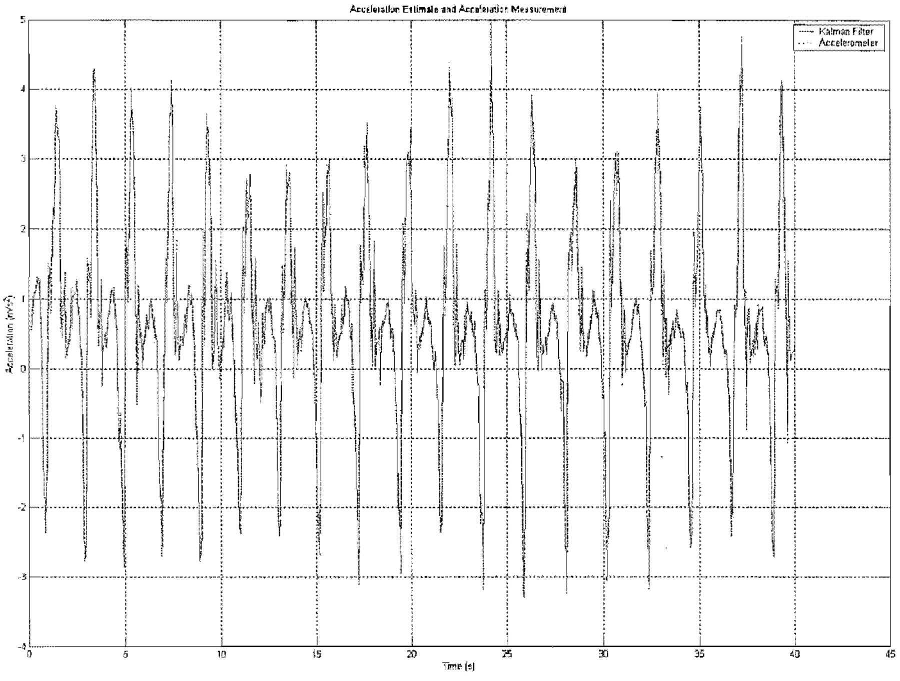


Figure 5.28 The accelerometer output, and the Kalman filter generated acceleration estimate.

While the accelerometer output and KF estimate appear to match very closely, there is a difference, as is made evident through the velocity estimate, Figure 5.29, which now, after an initial increase, oscillates about a reasonably steady mean (as compared to the integral of the accelerometer signal, which exhibited ramping)..

Finally, the position estimate of the KF is shown in Figure 5.30 . The difference between the estimate and the SpeedCoach data could suggest that the factory calibration of the sensor was a little low.

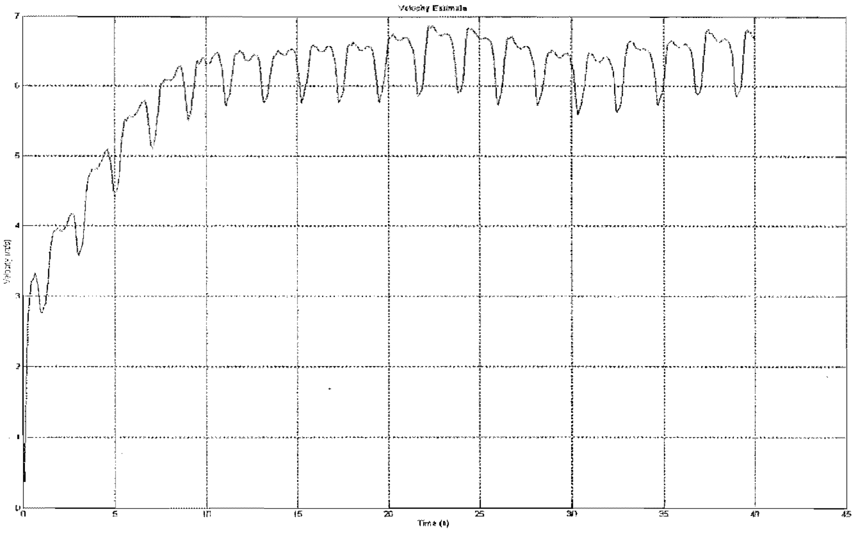


Figure 5.29 The Kalman filter generated velocity estimate.

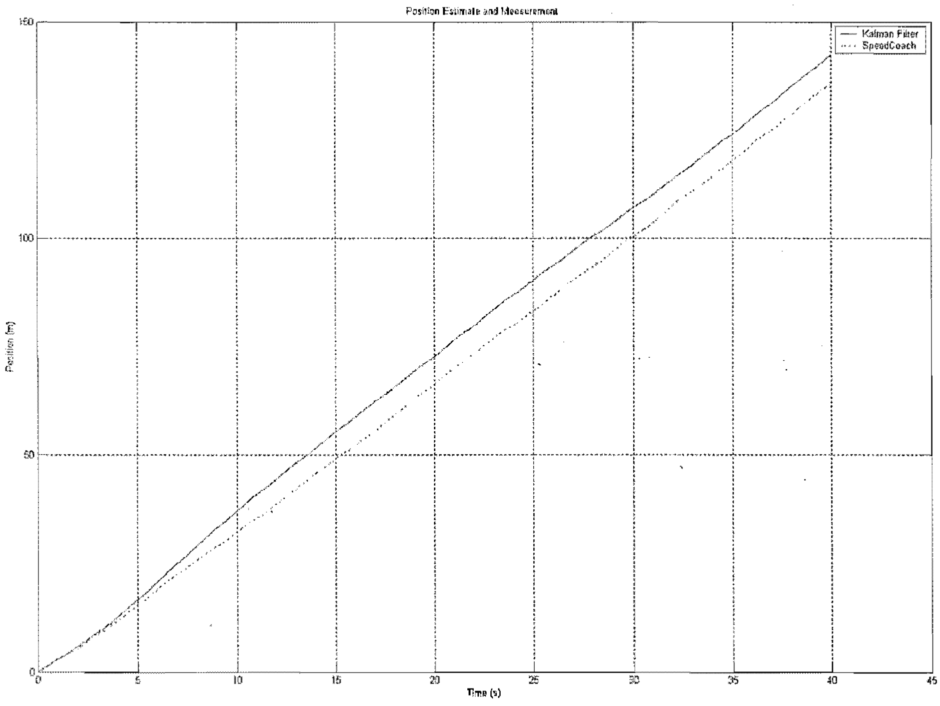


Figure 5.30 The Kalman filter position estimate and the SpeedCoach position measurement.

The results of Kalman filtering as differentiator, for seat motion measurement, and sensor fusion algorithm, for boat motion estimation, have now been presented. The next brief section combines the estimates produced by these two filters to investigate the motion of the system centre of mass.

5.4 Centre of Mass

The following equation of motion for the boat/rower system was developed in Chapter 1, (1.9):

$$F_{blade}\cos\theta_{oar} - m_{rower}a_{rower} - D = (m_{boat} + m_{rower})\,a_{boat}. \qquad (5.36)$$

Using the substitution:

$$(m_{rower} + m_{boat})a_{sys} = m_{rower}(a_{rower} + a_{boat}) + m_{boat}a_{boat} \qquad (5.37)$$

where a_{rower} is the acceleration of the rower relative to the boat and a_{sys} is the acceleration of the centre of mass of the two-body system, (5.36) may be rewritten

$$\begin{aligned} F_{blade}\cos\theta_{oar} - D &= (m_{boat} + m_{rower})\,a_{boat} + m_{rower}a_{rower} \\ &= m_{rower}(a_{rower} + a_{boat}) + m_{boat}a_{boat} \\ &= (m_{rower} + m_{boat})a_{sys} \qquad (5.38) \end{aligned}$$

Thus, the acceleration of the system, or the centre of mass of the system (the two major components of which are the rower and the boat) is seen to be a variable of interest. It is also clear, as was indicated at the beginning of this chapter, that the acceleration of the system centre of mass can be estimated through knowledge of the boat acceleration and the acceleration of the seat relative to the boat (assuming one has knowledge of the masses of the components of the system). The velocity of the centre of mass is also of interest, and is perhaps more easily comprehended by the viewer. For example, neglecting the effects of drag and oar forces (quite a significant neglect!) and considering only the rower and boat in the system, system momentum should be conserved. This means that when the rower slides towards the stern during the recovery, the boat should move faster in the direction of the bow. When the water is reinstated, this effect will still be present, but will be somewhat damped. The periodic application of oar force also, of course, causes fluctuation in the boat motion, upon which the other effects are superimposed.

Along with plots showing the motion of the system centre of mass, this section includes plots showing the timing between oar force, seat and boat motion.

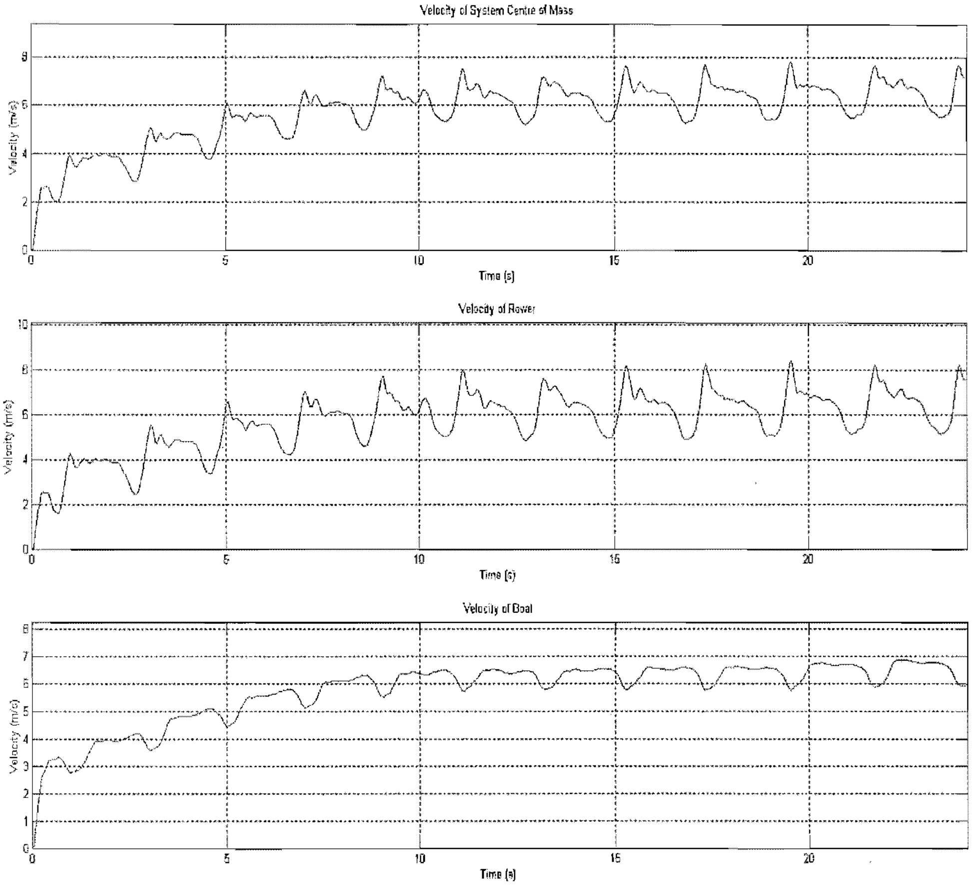


Figure 5.31 The velocity of the system centre of mass, the absolute velocity of the rower, and the velocity of the boat.

Figure 5.31 shows the velocity of the centre of mass (COM) along with the absolute velocities of the rower and boat. Note how closely the velocity of the COM matches the velocity of the rower. This is of course due to the fact that the rower is the major component of the system. In this case, the rower's mass was 100kg, while that of the boat and all components moving with it (notably ORAC and sensory devices) was less than 30kg.

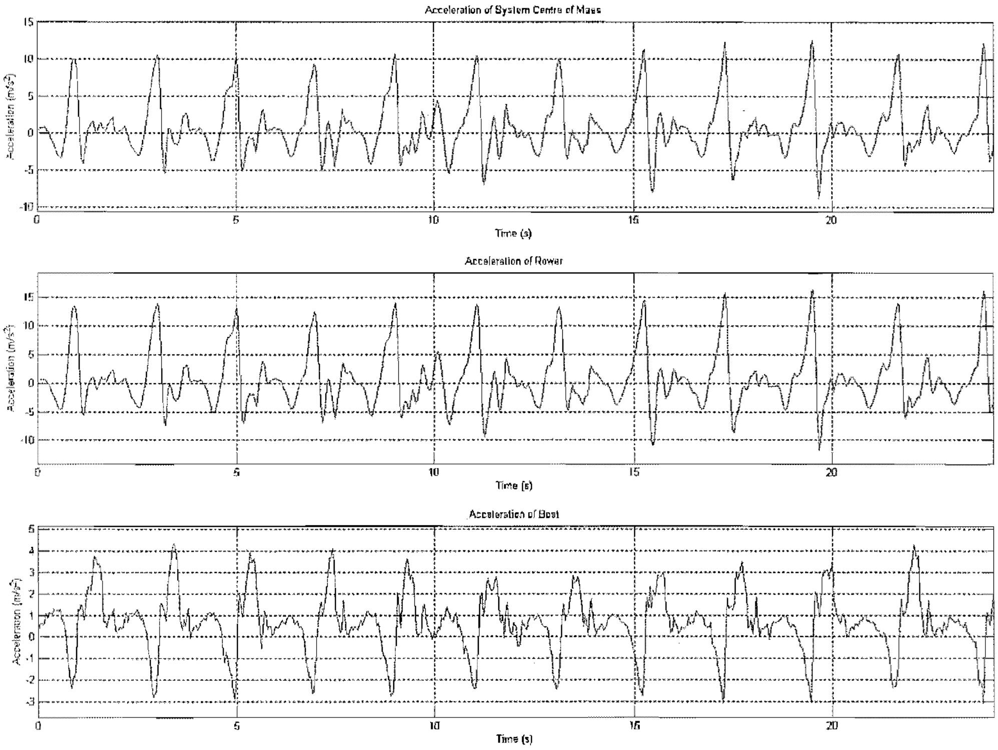


Figure 5.32 The acceleration of the system centre of mass, the absolute acceleration of the rower, and the acceleration of the boat.

The acceleration of the system components and COM are shown in Figure 5.32. While this plot is not particularly aesthetically pleasing, there are a number of interesting aspects to note. Chiefly, note the when the rower's acceleration is most positive, during the drive phase, the boat actually shows negative acceleration, due to the previously discussed momentum effects. This is perhaps more evident in plots shown below.

The timing of events during the rowing cycle is displayed in Figure 5.33. As expected, the force is applied as the rower slides in the direction of motion of the boat. The rower from whom this data was collected was a little out of practice, as can be seen from the lack of

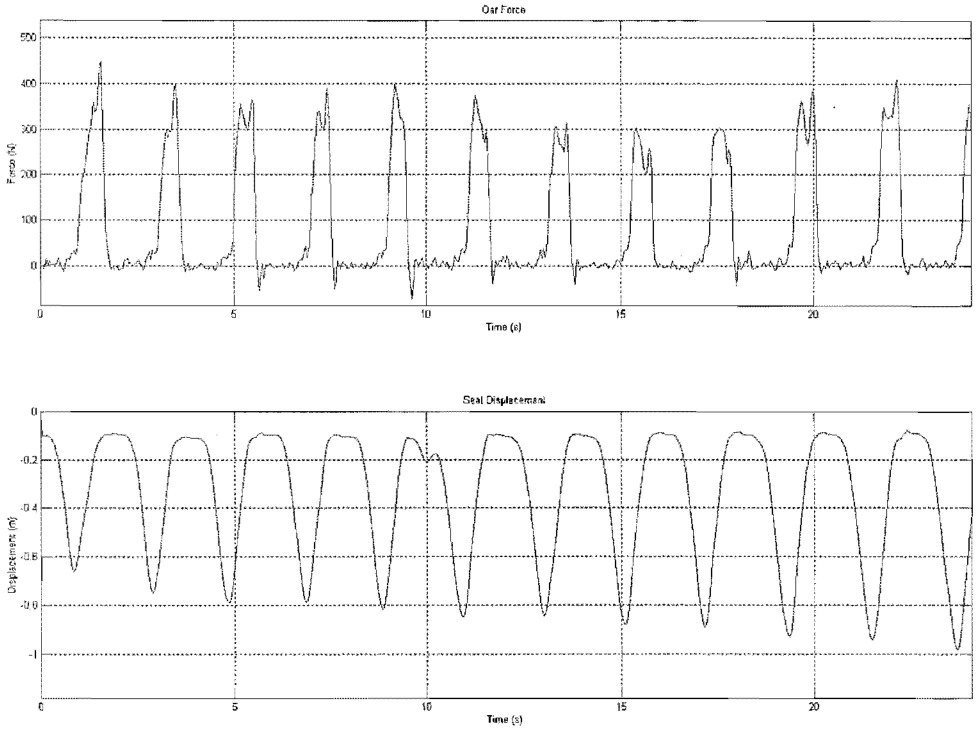


Figure 5.33 Oar force and seat displacement

consistency in both the seat movement and force profiles. An area of interest for athletes and coaches is the relative timing of both the initiation of seat movement and force generation, and the conclusion of force production and seat motion. If force is generated before the seat moves, it is mainly the back or arms that are doing work, while if the rower continues to move in the direction of motion of the boat after he stops pulling on the oars, he has wasted a portion of the stroke. Both of these effects are visible in Figure 5.33.

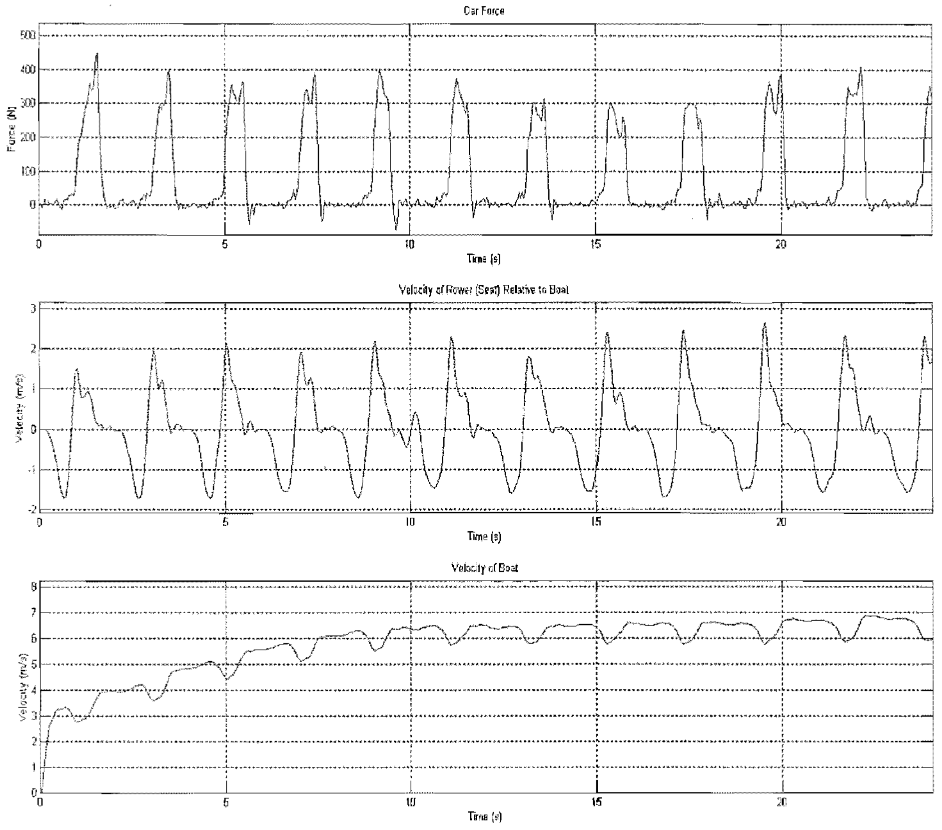


Figure 5.34 Oar force, relative seat velocity and absolute boat velocity.

Very interesting results are shown in Figure 5.34, the relative velocity of the rower is shown along with the oar force and the velocity of the boat. Considering the first variables, it is seen that the force is initially generated with a fast increase in velocity. This may be identified as the portion of the drive immediately following the catch. During the drive development, the velocity decreases, but remains positive during force generation. Inconsistencies in the seat movement are easily seen in this plot as peaks during the rower's deceleration. Considering now the oar force and seat movement with the boat velocity, it is seen, that when the oar force is peaking, the boat velocity is actually in a trough. This is due to the previously mentioned momentum effects.

5.5 Discussion & Conclusions

The results of the preceding section indicate that the methods used to measure seat and boat motion are useful.

The encoder-mounting bracket was sometimes a little difficult to position but held the encoder in position well, the only problems with the encoder's use being the tendency of the encoder wheel to slide along the spindle and become disengaged with the track. This can easily be fixed.

Using a Kalman filter as a differentiator for encoder measurements was an interesting pursuit. While the high resolution of the encoder lead to relatively accurate velocity estimation by finite differencing, the Kalman filter acceleration estimates were qualitatively seen to be of better quality than those obtained by standard numerical differentiation. The design of system transfer matrices so soon after working with rotation matrices allowed some interesting comparisons. The matrices chosen for the shaping filters are effectively 2D rotation matrices with a small amount of decay added.

While previous researchers have mentioned the use of impellers in combination with accelerometers in the measurement of boat motion, none have discussed how, or indeed if, the data from these two instruments are combined. The simple Kalman filter presented in this chapter appears to be a useful sensor fusion technique.

While only a small attempt was made to analyse the collected results, it is clear that there is significant scope for research in the area of the timing of the events during the rowing stroke and the biomechanical implications of this timing.

Chapter 6

Many encouraging developments were made during this work. Perhaps a more robust system could have been created, but the work was a development exercise and it is thought that as such, it has been very successful. The author enjoyed unexpected excursions into the worlds of estimation theory and theoretical kinematics along the way, and it is hoped, and believed, that these will prove fruitful.

This brief chapter makes recommendations for further research. Aside from the section covering the performance of the data acquisition system, which is presented in Appendix A4, the items are discussed in the same order as they were presented in the thesis.

6.1 System

Several components of the system hardware limited the utility of the system. Among these was the very poor performance of the wireless LAN card. This card was specified to have a range of around 100m, but during testing it seemed that a more reasonable estimate would be around 10m. This meant that testing was performed in standalone mode, i.e. the data acquisition program was started manually and the boat was sent off with no feedback to ensure that the sensors were indeed working. This would not have been so problematic if all the other system hardware was functioning correctly, but unfortunately this was not the case. Output from the SpeedCoach sensor was very temperamental. Sending it out on two runs with seemingly identical operating conditions, it would function very differently; sometimes it would work well, at other times there would be no output at all. In addition to this, some strange effects were sometimes noted on the analog channels, where one channel would greatly alter the output of its neighbours. This problem was also sporadic, and despite the best efforts of the electronics technicians could not be identified or eliminated. It was initially thought that this problem was due to the short length of time between consecutive samples of the multiplexer, but increasing this time did not fix the problem. Luckily some data runs were free of this problem.

Apart from problems with system performance and reliability, it is desirable that the physical size and mass of the system be reduced. Rowers sometimes looked at the computer balanced by the foot-stretcher with suspicion, and both the mass and the consequences of capsizing the boat affected their rowing styles.

An interesting idea, conceived of by electronics technician, Julian Phillips, is constructing the sensors to have 'on-board' power supplies, signal processing and short-term data capture facilities. This would reduce the mass of the system and the set-up time. The obvious problems are the increased cost, and the synchronization of data from the numerous sensors. The former problem will reduce with time, and the latter may be overcome by some simple wireless form of communication between the sensors, making these 'smart sensors' a very attractive option for future development.

6.2 Oar Force

The developed oar force sensor is appealing because of its simplicity, wide applicability, and the quality of the results it offers. It is recommended that future work be directed into improving the design of the sensor, and eliminating the mysterious long-term drift that was observed. With this drift removed, it is believed that the oar force sensor could be a useful tool for rowers, coaches and biomechanists.

6.3 Foot Force

As mentioned at great length in Chapter 3, the foot force sensors shear force characteristics were very poor. Although this is the case, it is believed that the sensor justifies further work, due to the encouraging normal force and coordinate estimation capabilities that were displayed. Improving the shear response could be achieved in a number of ways that are now briefly detailed.

Increasing the dimensions of the slots at either end of the sensor would allow for the placement of strain gauges on both sides of the shear sensing beams. This should increase the sensitivity and linearity of the shear response. The response, however, will never be fully linear, due to the stress state of the beams, caused by the end constraints. This can be seen, since when a central downward force is applied to the plate, the ends of the sensing beams will deflect downward, but rather than being a case of simple bending, additional compressive stress is superimposed due to the manner in which the beams join the plate.

6.4 Oar Orientation

This part of the work was the most inspiring and also the most annoying. Once the sensors are calibrated as described at the end of Chapter 4, the sensors should yield sound estimates of relative orientation in 3D space. Among the fields of application for this technology are: virtual reality, prosthesis control, haptic interfaces, personal navigation and, of course, sports performance measurement. It is hoped that future researchers will pursue this technology to its conclusion. The author is currently undertaking work to expand the generality of the sensors from 'spherical' to truly spatial applications.

6.5 Seat and Boat Motion

As mentioned in Chapter 5, the use of an accelerometer to measure the motion of the seat was ruled out because of the necessity of accounting for the motion of the boat, which prior to actually collecting data was expected to be a complex affair. Based on the 1D motion of the boat, and a little hindsight, using an accelerometer in combination with the rotary encoder, as described below, would be an interesting option. The 1D boat motion may even make a variant of the described real-time-spline methods viable.

If an accelerometer and rotary encoder, were both connected to the seat, the output data of these sensors and the SpeedCoach could be combined in one Kalman filter. This filter would be a simple extension of that derived in Chapter 5, with two shaping filters, one for the derivative of the seat acceleration relative to the boat, the other for the derivative of the boat acceleration. Clearly the seat accelerometer measures the absolute acceleration of the seat; that is the sum of the acceleration of the seat relative to the boat and the acceleration of the boat itself. With the filter designed accordingly, the outputs would include all the kinematic parameters of interest for the system.

While the above approach would be more academically interesting, the results presented in Chapter 5 were pleasing.

6.6 Comments on Possible Studies

The above comments, and indeed the entire thesis, have been concerned with the development of the instrumentation system. The collected data has only been analysed so far as to show that the macroscopic features agree with intuition. Once the elements of the instrumentation system are developed, a wealth of information will be available for wide ranging investigations of rowing. The studies rendered possible include both biomechanical studies, of the rowers and their interactions with the boat, and the pure mechanical performance of the boat itself. It will, for example, using a combination of boat motion, oar force, foot force and/or rower motion data, be possible to estimate, using system identification techniques, the drag characteristics of a boat during actual rowing, rather than during simulated tests.

There is also scope for extension or alteration of the instrumentation system. In particular, a simpler oar angle sensor, such as a potentiometer, may be applied to yield the sweep angle of the oar. Extensions to the system include the use of feedback for the athlete, which may be presented using dedicated goggles with head up display, a small touch screen within the boat, or in the simplest case, some audible indications of the rowing parameters.

It is of interest to rowing coaches and athletes that the boat be configured in the most optimal way, i.e. the energy of the athlete is used efficiently. There are a number of parameters involved in the rigging of a boat, including foot-stretcher position and angle, distance between oarlocks, oarlock height and pitch and oar length, and the optimality is therefore a function of each of these parameters. To find the optimal combination of rigging parameters would be a large task, but varying single parameters at a time and collecting data should lead to meaningful results and improved athlete specific rigging. For example, measuring the oar sweep angle and the force at the oarlock shows how much of the generated force is in the direction of motion, and how much is 'wasted' by compression of the hull of the boat. Altering the stroke characteristics by shifting the foot-stretches or oarlock spacing will change the oarlock-force/time and oar-angle/time curves, resulting in different efficiencies.

Measuring the multi-axial force components in the oar shaft at the same time as the oarlock force and rotation will give some indication of the force generation mechanisms. In this way, in combination with video of the blade-water interaction, the fluid dynamics of the rowing at various stroke ratings could be well investigated.

When the relative orientation sensor is functional, it will be possible to add extra dimensions to the study of the blade-water interaction. In particular, the effects of the pitch and roll of the oar during the drive will be able to be investigated. The pitch will indicate how deep the blade is in the water, while the roll measures how orthogonal the blade is with respect to the water. The symmetry of the rower's technique will also be able to be studied down to the degree. This tool is not limited to rowing studies, and should find applications in diverse fields such as feedback for prosthesis control and virtual reality. The further development and miniaturisation of these sensors is considered to be very worthwhile.

6.7 Achievements and Contributions of Research

The part of this research that had the most potential; the estimation of relative orientation using accelerometers and magnetoresistive sensors, was unfortunately not realised beyond the theoretical development. The developed theory, including the revised method of calibration, which takes account of assembly (non-orthogonality) errors, should enable the construction of sensors that will accurately estimate the relative orientation of consecutive rigid bodies in a kinematic chain connected by spherical joints undergoing general spatial motion. With the exception that the bodies be non-ferrous (so as not to saturate the magnetoresistive sensors) the applications of these sensors are boundless. Within the conceptual developments of the theory, a new method of orientation estimation was developed, which was found to outperform all reviewed methods.

The combination of accelerometer and impeller has been used by previous researchers to estimate the position, velocity and acceleration of the boat. In these instances, however, no mention was made of the method in which the outputs of the sensors were combined. The discrete Kalman filter was shown to be very suitable for this purpose. A very similar filter was also used to differentiate a quantized random signal. As this filter was being designed, a new (as far as the author is aware) method of checking the evolution of the autocorrelation matrix of a discrete-time state-space model driven by white noise. This method could conceivably be used to design, given required autocorrelation functions, state-space models for random sequences.

Aside from theoretical developments, two strain gauge sensors were designed. The oar force sensor is of a new design, which does not alter the external geometry of the oarlock and does not require the rower to use a specially instrumented oar. The minimal disturbance to the feel of rowing, and the maintenance of the ability to easily alter the pitch angle of the oarlock are pleasing aspects to the sensor's design. In addition to fulfilling these design constraints, the sensor is easy to calibrate in a way consistent with loading during rowing and yields sound data. A sensor of similar design may be useful in other situations in which the compressive force on an axle (or shaft) inside a cylindrical enclosure be measured. Examples of such cases are:

measuring the loads on the axles of a truck, to ensure that cargo is well distributed; measuring the forces on an axle during mountain biking to aid in the design of optimal wheels.

The design of the foot force sensor was based upon commercially available force plates. Unlike a force plate, the sensor incorporates the sensing elements into the actual structure and has the advantages of relatively low cost and weight. The use of least squares estimation techniques in the processing of the outputs of the foot force sensor proved to be effective, and was also useful in the analysis of the sensor's response, especially its poor shear characteristics.

A1 Alternative Method for Determination of Normal Force and Coordinates for Foot Force Sensor

The method described below assumes (incorrectly) that the four 'normal force channels' labelled 1,2,3,4 are not sensitive to shear, and hence the normal force and coordinates are estimated without consideration of shear. The method was found to give almost exactly the same results as the least squares method that was used in its place, its main disadvantage was the large number of constants that had to be determined for its implementation. Even using Maple® to perform the calculus and algebra, the method was very time-consuming.

Recall, from Chapter 3, that the approximate equation for the output of the i^{th} channel of the foot force sensor is given by:

$$V_i = fC_{fi} + fxC_{Xi} + fyC_{Yi} \tag{A1.1}$$

Treating x and y (the coordinates of the centre of force) as independent variables, this equation represents a continuum of 'voltage planes' in R^3, corresponding to a range of f. At each sampling instant, each channel has an associated measurement plane , which passes through this force dependent continuum of planes. The intersection of the i^{th} measurement plane (parallel to the xy plane) and the i^{th} voltage plane continuum is the line

$$y_i = \frac{-C_{Xi}}{C_{Yi}}x + \frac{V_i - fC_{fi}}{fC_{Yi}} \tag{A1.2}$$

These lines, for given f give the required relationship between x and y to yield the measured i^{th} voltage. The equations of the lines can be rewritten in a simpler form:

$$y_i = m_i x + c_i(f) \tag{A1.3}$$

which makes it explicit, that the force, f, only changes the intercept of the line, i.e. the gradient is invariant to force.

The intersection of the lines at an arbitrary instant is shown below. The intersection of lines y_i and y_j occurs at the point (x_{ij} , y_{ij}). The diagram approximately reproduces the 'sense' of the solution lines in that they are nearly in orthogonal parallel pairs (as is discussed in Chapter 3). The intersection of two lines, of course, represents an agreement between two channels on the coordinate of the centre of force for a particular f. As a consequence of the approximate nature of the planes, and also the fact that the lines are in near parallel pairs, an agreement between all four channels, represented by an intersection of all four lines, is very unlikely. Intuitively, the best agreement between the channels, which are all assumed to be providing valid information, is achieved when the area bounded by the four lines is minimised. When the area is small, the lines representing the coordinates deemed 'allowable' by each of the channels will be in closest agreement.

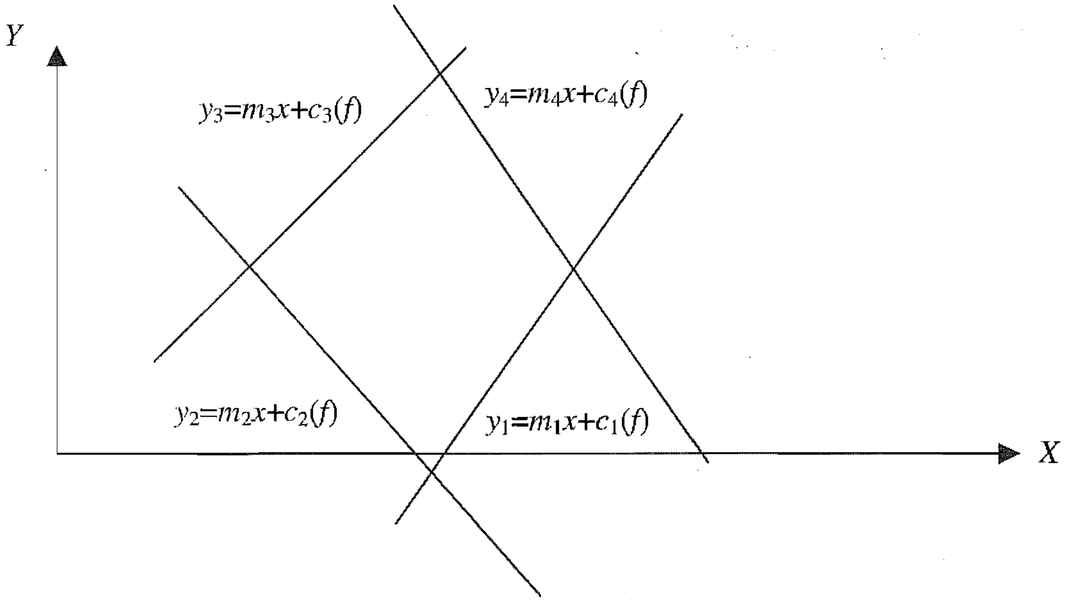


Figure A1.1 Estimation problem geometry

A procedure for solution, therefore, is to find f for which the enclosed area is smallest and average the coordinates of the four intersections at this value. This yields, what is in some sense, the most likely combination of f, x and y.

Consider the problem of minimising the quadrilateral area bounded by four lines of fixed gradient and variable intercept. It is evident that the minimisation of the area is equivalent to the minimisation of either of the regions diagonals, as the opposing intercepts are dependent upon all four lines. The mathematics required is simplified if the minimised function is the square of the length of the diagonal, and as this does not affect the result, this is the approach taken.

Referring to the Figure A1.1, the square of the length of one of the diagonals is:

$$D=(x_{34}-x_{12})^2+(y_{34}-y_{12})^2 \qquad \text{(A1.4)}$$

Substituting:

$$\begin{aligned} x_{34} &= \frac{c_4(f)-c_3(f)}{m_3-m_4} \\ x_{12} &= \frac{c_4(f)-c_1(f)}{m_1-m_2} \\ y_{34} &= m_3x_{34}+c_3(f) \\ y_{12} &= m_1x_{12}+c_1(f) \end{aligned} \qquad \text{(A1.5)}$$

and,

$$\begin{aligned} c_1(f) &= \frac{V_1}{fC_{Y1}}-\frac{C_{f1}}{C_{Y1}} \\ c_3(f) &= \frac{V_3}{fC_{Y3}}-\frac{C_{f3}}{C_{Y3}} \end{aligned} \qquad \text{(A1.6)}$$

gives

$$D = \left(\frac{\frac{V_4}{fC_{y4}} - \frac{C_{f4}}{C_{y4}} - \frac{V_3}{fC_{y3}} + \frac{C_{f3}}{C_{y3}}}{m_3 - m_4} - \frac{\frac{V_2}{fC_{y2}} - \frac{C_{f2}}{C_{y2}} - \frac{V_1}{fC_{y1}} + \frac{C_{f1}}{C_{y1}}}{m_3 - m_4} \right)^2 +$$

$$\left(\frac{m_3\left(\frac{V_4}{fC_{y4}} - \frac{C_{f4}}{C_{y4}} - \frac{V_3}{fC_{y3}} + \frac{C_{f3}}{C_{y3}} \right)}{m_3 - m_4} + \frac{V_3}{fC_{y3}} - \frac{C_{f3}}{C_{y3}} - \frac{m_1\left(\frac{V_2}{fC_{y2}} - \frac{C_{f2}}{C_{y2}} - \frac{V_1}{fC_{y1}} + \frac{C_{f1}}{C_{y1}} \right)}{m_3 - m_4} - \frac{V_1}{fC_{y1}} + \frac{C_{f}}{C_{y1}} \right.$$

(A1.7)

This function is then minimised with respect to f (i.e. the first differential w.r.t. f is set to zero and the resulting equation solved for f). The resulting expression can be simplified to give the following:

$$f = \frac{a_1V_1^2 + a_2V_2^2 + a_3V_3^2 + a_4V_4^2 + a_5V_1V_2 + a_6V_1V_3 + a_7V_1V_4 + a_8V_2V_3 + a_9V_2V_4 + a_{10}V_3V_4}{b_1V_1 + b_2V_2 + b_3V_3 + b_4V_4}$$

(A1.8)

where the constants are obviously algebraic combinations of the constant terms of the expressions of the equations of the channel performance surfaces.

Having estimated f, the coordinates of the intersections of the lines are found by substituting f into equations of the form as those given above for x_{34}, y_{34}. The final estimate of the load coordinates is given as the average of the coordinates of the line intersections.

A2 Relative Orientation Estimation

The approaches reviewed in this Appendix are all least squares techniques of estimating the rotation matrix. The rotation matrix is not of course the only method by which orientation can be specified. A popular alternative, the use of unit quaternions is mentioned in section A2.4. While it may seem a little extraneous to include limited derivations of previously used methods, it acts to highlight possible problems. It is also interesting to see the way in which new mathematical methods have altered the algorithms over the last few decades.

A2.1 Wahba's Problem

In 1965, Wahba [66], posed the following problem for fellow mathematicians to solve (where the notation has been modified for simplicity):

Given two sets of n points $\{\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_n\}$ and $\{\mathbf{R}_1, \mathbf{R}_2, \ldots \mathbf{R}_n\}$, where $n \geq 2$, find the rotation matrix $\mathbf{A}$ which brings the first set into the best least squares coincidence with the second. That is, find $\mathbf{A}$ which minimises

$$\sum_{j=1}^{n} \left\| \mathbf{R}_j - \mathbf{A}\mathbf{r}_j \right\|^2 . \tag{A2.1}$$

There were several replies to this problem. We review here two methods of solution, one that was first stated in reply to Wahba, and the other that strangely does not even refer to the original challenge.

Of the replies to Wahba's problem, the most often quoted is that devised by Brock, which is now explained in detail. The initial steps of the derivation are common to both reviewed methods of solution and consist of transforming the minimum norm problem into the equivalent 'minimum trace' problem. This makes the solution easier, since the following properties of traces can be exploited [14]

tr($\mathbf{QR}$) = tr($\mathbf{RQ}$) for conformable matrices $\mathbf{Q}$ and $\mathbf{R}$

tr($\mathbf{Q} + \mathbf{R}$) = tr($\mathbf{Q}$) + tr($\mathbf{R}$) for matrices, $\mathbf{Q}$ and $\mathbf{R}$ of the same dimension

$\mathrm{tr}(\mathbf{Q}^{\mathrm{T}}) = \mathrm{tr}(\mathbf{Q})$ for any square matrix, $\mathbf{Q}$

The equivalence of the norm problem to a trace is now shown. For simplicity, suppose that we are concerned with the quantity:

$$\sum_{j=1}^{n}\left\|\mathbf{b}_j\right\|^2 = \sum_{j=1}^{n}\mathbf{b}_j^T\,\mathbf{b}_j = \mathbf{b}_1{}^{\mathrm{T}}\mathbf{b}_1 + \mathbf{b}_2{}^{\mathrm{T}}\mathbf{b}_2 + \ldots + \mathbf{b}_n{}^{\mathrm{T}}\mathbf{b}_n \tag{A2.2}$$

If we form a matrix, $\mathbf{B} = [\mathbf{b}_1\ \mathbf{b}_2 \ldots \mathbf{b}_n]$, then clearly

$$\sum_{j=1}^{n}\left\|\mathbf{b}_j\right\|^2 = \mathrm{tr}\mathbf{B}^{\mathrm{T}}\mathbf{B} \tag{A2.3}$$

Following this approach, we define measurement matrices $\mathbf{R} = [\mathbf{R}_1\ \mathbf{R}_2 \ldots \mathbf{R}_n]$ and $\mathbf{r} = [\mathbf{r}_1\ \mathbf{r}_2 \ldots \mathbf{r}_n]$ so that we have the following equality, from (A2.1):

$$\sum_{j=1}^{n}\left\|\mathbf{R}_j - A\mathbf{r}_j\right\|^2 = \mathrm{tr}(\mathbf{R} - \mathbf{Ar})^{\mathrm{T}}(\mathbf{R}\text{-}\mathbf{Ar}) \tag{A2.4}$$

Defining this cost function as $J(\mathbf{A})$, and expanding gives

$$\begin{aligned} J(\mathbf{A}) &= \mathrm{tr}(\mathbf{R}\text{-}\mathbf{Ar})^{\mathrm{T}}(\mathbf{R}\text{-}\mathbf{Ar}) = \mathrm{tr}(\mathbf{R}^{\mathrm{T}}\mathbf{R} - \mathbf{R}^{\mathrm{T}}\mathbf{Ar} - \mathbf{r}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{R} + \mathbf{r}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{Ar}) \\ &= \mathrm{tr}(\mathbf{R}^{\mathrm{T}}\mathbf{R} - \mathbf{R}^{\mathrm{T}}\mathbf{Ar} - \mathbf{r}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{R} + \mathbf{r}^{\mathrm{T}}\mathbf{r}), \end{aligned} \tag{A2.5}$$

where the orthogonality of **A** has been used to simplify the last term. Clearly the cost, J, is minimised by maximising the trace of the two middle (negative) terms, i.e. we aim to maximise

$$K(\mathbf{A}) = \mathrm{tr}(\mathbf{R}^{\mathrm{T}}\mathbf{Ar} + \mathbf{r}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{R}). \tag{A2.6}$$

The method by which this number is minimised is where the differences between solutions normally begin.

Brock's approach is to find the orthogonal matrix **A** such that the cost, K, is stationary with respect to variation in any of the elements of **A**. Denoting partial differentiation with respect to an arbitrary element of **A**, a, by (-)*, and using the readily observed fact that [tr(-)]* = tr[(-)*]

$$K^*(\mathbf{A}) = \mathrm{tr}([\mathbf{R}^T\mathbf{A}\mathbf{r}]^* + [\mathbf{r}^T\mathbf{A}^T\mathbf{R}]^*) = \mathrm{tr}[\mathbf{R}^T\mathbf{A}^*\mathbf{r} + \mathbf{r}^T(\mathbf{A}^T)^*\mathbf{R}] \qquad \text{(A2.7)}$$

Differentiating the equation, $\mathbf{A}\mathbf{A}^T = \mathbf{1}$, with respect to a gives

$$\mathbf{A}^*\mathbf{A}^T + \mathbf{A}(\mathbf{A}^T)^* = \mathbf{0},$$
$$\mathbf{A}^* = -\mathbf{A}(\mathbf{A}^T)^*\mathbf{A}, \qquad \text{(A2.8)}$$

and substituting (A2.8) into (A2.7) gives

$$K^*(\mathbf{A}) = \mathrm{tr}[-\mathbf{R}^T\mathbf{A}(\mathbf{A}^T)^*\mathbf{A}\mathbf{r} + \mathbf{r}^T(\mathbf{A}^T)^*\mathbf{R}]. \qquad \text{(A2.9)}$$

Now, reordering the matrix products within the trace

$$K^*(\mathbf{A}) = \mathrm{tr}[-(\mathbf{A}^T)^*\mathbf{A}\mathbf{r}\mathbf{R}^T\mathbf{A} + (\mathbf{A}^T)^*\mathbf{R}\mathbf{r}^T] = \mathrm{tr}[(\mathbf{A}^T)^*\{\mathbf{A}\mathbf{r}\mathbf{R}^T\mathbf{A} - \mathbf{R}\mathbf{r}^T\}] \quad \text{(A2.10)}$$

we see that $K^*(\mathbf{A})$ is stationary with respect to variation in the elements of $\mathbf{A}$, if

$$\mathbf{A}\mathbf{r}\mathbf{R}^T\mathbf{A} = \mathbf{R}\mathbf{r}^T. \qquad \text{(A2.11)}$$

Multiplying on the right by $\mathbf{r}\mathbf{R}^T$ we get

$$(\mathbf{A}\mathbf{r}\mathbf{R}^T)^2 = \mathbf{R}\mathbf{r}^T\mathbf{r}\mathbf{R}^T, \qquad \text{(A2.12)}$$

or $$\mathbf{A} = (\mathbf{R}\mathbf{r}^T\mathbf{r}\mathbf{R}^T)^{1/2}(\mathbf{r}\mathbf{R}^T)^{-1} = (\mathbf{B}\mathbf{B}^T)^{1/2}(\mathbf{B}^T)^{-1} \qquad \text{(A2.13)}$$

where $\mathbf{B} = \mathbf{R}\mathbf{r}^T$. There are a number of things to note with regard to this solution. Firstly for the solution to exist, $\mathbf{B}$ must be non-singular, i.e. of rank three. The minimum number of measurements that will ensure the full rank of $\mathbf{B}$ is three, with the requirement that the measurements be linearly independent. Consider first the case of two non-collinear vector observations, i.e. $\mathbf{R} = [\mathbf{R}_1\ \mathbf{R}_2]$, $\mathbf{r} = [\mathbf{r}_1\ \mathbf{r}_2]$, then

$$\mathbf{B} = \mathbf{R}_1\mathbf{r}_1^T + \mathbf{R}_2\mathbf{r}_2^T = [(r_{11}\mathbf{R}_1 + r_{21}\mathbf{R}_2)\ (r_{12}\mathbf{R}_1 + r_{22}\mathbf{R}_2)\ (r_{13}\mathbf{R}_1 + r_{23}\mathbf{R}_2)] \qquad \text{(A2.14)}$$

and the columns of $\mathbf{B}$ are clearly linearly dependent since they all lie in the plane of $\mathbf{R}_1$ and $\mathbf{R}_2$. Similarly for three measurements:

$$\mathbf{B} = [(r_{11}\mathbf{R}_1 + r_{21}\mathbf{R}_2 + r_{31}\mathbf{R}_3)\ (r_{12}\mathbf{R}_1 + r_{22}\mathbf{R}_2 + r_{32}\mathbf{R}_3)\ (r_{13}\mathbf{R}_1 + r_{23}\mathbf{R}_2 + r_{33}\mathbf{R}_3)]. \tag{A2.15}$$

This matrix is non-singular unless $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ all lie in a plane, since then the matrix is effectively

$$\mathbf{B} = [(r_{11}+r_{31}\alpha)\mathbf{R}_1+(r_{21}+r_{31}\beta)\mathbf{R}_2\ \ (r_{12}+r_{32}\alpha)\mathbf{R}_1+(r_{22}+r_{32}\beta)\mathbf{R}_2\ \ (r_{13}+r_{33}\alpha)\mathbf{R}_1+(r_{23}+r_{33}\beta)\mathbf{R}_2], \tag{A2.16}$$

where $\mathbf{R}_3 = \alpha\mathbf{R}_1 + \beta\mathbf{R}_2$. Thus the requirement for Brock's solution to be useful is that we are in possession of three linearly independent measurements in each frame.

The second point to note with Brock's solution is that a matrix square root is required. The method by which this square root is obtained through an eigendecomposition is now described. The matrix, $\mathbf{C} = \mathbf{B}^T\mathbf{B}$, of which the square root is required, is symmetric and thus can be expressed,

$$\mathbf{C} = \mathbf{E}\Lambda\mathbf{E}^T \tag{A2.17}$$

where $\mathbf{E}$ is an orthogonal matrix of eigenvectors and Λ is a diagonal matrix containing the eigenvalues of $\mathbf{C}$. The square-root of $\mathbf{C}$ is that matrix $\mathbf{D}$ such that

$$\mathbf{D}^2 = \mathbf{C} = \mathbf{E}\Lambda\mathbf{E}^T. \tag{A2.18}$$

Note that $(\mathbf{E}\Lambda\mathbf{E}^T)(\mathbf{E}\Lambda\mathbf{E}^T) = \mathbf{E}\Lambda^2\mathbf{E}^T$, therefore $\mathbf{D} = \mathbf{E}\Lambda^{1/2}\mathbf{E}^T$. Brock states that there is freedom in calculating the square root of the diagonal matrix Λ, since each element may be either $\pm(\lambda_i)^{1/2}$, i.e.

$$\Lambda^{1/2} = \mathrm{diag}\{\pm(\lambda_1)^{1/2}\ \pm(\lambda_2)^{1/2}\ \pm(\lambda_3)^{1/2}\} \tag{A2.19}$$

It can be shown however, as was stated in Carta and Lackowski [16], that the maximum value of K is always assumed when the positive roots are taken. This is proved by manipulating the definition of K from (A2.6):

$$\begin{aligned} K(\mathbf{A}) &= \mathrm{tr}(\mathbf{R}^{\mathrm{T}}\mathbf{A}\mathbf{r} + \mathbf{r}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{R}) = 2\mathrm{tr}(\mathbf{R}^{\mathrm{T}}\mathbf{A}\mathbf{r}) \\ &= 2\mathrm{tr}[\mathbf{R}^{\mathrm{T}}(\mathbf{R}\mathbf{r}^{\mathrm{T}}\mathbf{r}\mathbf{R}^{\mathrm{T}})^{1/2}(\mathbf{r}\mathbf{R}^{\mathrm{T}})^{-1}\mathbf{r}] \\ &= 2\mathrm{tr}[\mathbf{r}\mathbf{R}^{\mathrm{T}}(\mathbf{R}\mathbf{r}^{\mathrm{T}}\mathbf{r}\mathbf{R}^{\mathrm{T}})^{1/2}(\mathbf{r}\mathbf{R}^{\mathrm{T}})^{-1}] = 2\mathrm{tr}[(\mathbf{R}\mathbf{r}^{\mathrm{T}}\mathbf{r}\mathbf{R}^{\mathrm{T}})^{1/2}] \\ &= 2\mathrm{tr}(\mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{E}^{\mathrm{T}}) = 2\mathrm{tr}(\mathbf{\Lambda}^{1/2}) \end{aligned} \qquad \text{(A2.20)}$$

Thus K is maximised by choosing all positive roots of the eigenvalues of $\mathbf{C}$. This method is Brock's 'constrained' method. It is constrained in that the generated matrix is forced to be orthogonal. An unconstrained method Brock detailed in the same paper seems to be more popular with following researchers, principally because of its simplicity. Before this, and more modern unconstrained methods are detailed, a more recent solution of Wahba's problem is discussed. It is considered worthwhile to discuss this second method because

- it offers solutions when only two vector observations are available
- certain geometric insights are offered
- It is also interesting to see how a 'new' mathematical method, singular value decomposition (SVD) has both simplified the derivation and allowed for a more general solution.

The first steps of 'Arun's Solution' [4] are the same as those of Brock's in that the problem is reduced to that of maximizing (see (A2.6)):

$$K(\mathbf{A}) = \mathrm{tr}(\mathbf{R}^{\mathrm{T}}\mathbf{A}\mathbf{r} + \mathbf{r}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{R}) = 2\mathrm{tr}(\mathbf{R}^{\mathrm{T}}\mathbf{A}\mathbf{r}) = 2\mathrm{tr}(\mathbf{A}\mathbf{r}\mathbf{R}^{\mathrm{T}}) = 2\mathrm{tr}(\mathbf{A}\mathbf{H}) \qquad \text{(A2.21)}$$

where $\mathbf{H} = \mathbf{r}\mathbf{R}^{\mathrm{T}}$. The problem is therefore solved by the rotation matrix, $\mathbf{A}$, that maximises the trace of ($\mathbf{AH}$). The singular value decomposition of $\mathbf{H}$ is given by

$$\mathbf{H} = \mathbf{Q}\mathbf{\Lambda}\mathbf{S}^{\mathrm{T}} \qquad \text{(A2.22)}$$

where **Q** and **S** are orthogonal matrices and **Λ** is a diagonal matrix of non-negative elements (the singular values of **H**). If we select $\mathbf{A}^{\circ} = \mathbf{S}\mathbf{Q}^{T}$, which is orthogonal, since it is the product of two orthogonal matrices, then

$$\mathbf{A}^{\circ}\mathbf{H} = \mathbf{S}\mathbf{Q}^{T}\mathbf{H} = \mathbf{S}\mathbf{Q}^{T}\mathbf{Q}\boldsymbol{\Lambda}\mathbf{S}^{T} = \mathbf{S}\boldsymbol{\Lambda}\mathbf{S}^{T} \qquad \text{(A2.23)}$$

which is both symmetric and positive definite. The lemma below shows that this choice of **A** is optimal since $\mathrm{tr}(\mathbf{A}^{\circ}\mathbf{H}) \geq \mathrm{tr}(\mathbf{C}\mathbf{A}^{\circ}\mathbf{H})$, for any orthogonal matrix, **C**. This fully satisfies the requirements of the optimal orthogonal matrix since to form another orthogonal matrix requires that $\mathbf{A}^{\circ}$ be multiplied by a second orthogonal matrix. While $\mathbf{A}^{\circ}$ is assured of being orthogonal it is not necessarily a rotation matrix since the determinant has not been restricted to being +1. The other possibility is that $\det(\mathbf{A}^{\circ}) = -1$, in which case $\mathbf{A}^{\circ}$ describes a reflection. Since the columns of **r** and **R** are related by a rotation, in the situation in which $\mathbf{A}^{\circ}$ is a reflection matrix it must be possible to relate the vectors by both types of transformation. Geometrically it can be seen that the requirement for this case is that the observation vectors from each frame are linearly dependent. This case has been described in Section 4.6.3. Studying Fig. 4.12, it is clear that the vectors may be related by a reflection or a rotation.

If the vectors are linearly dependent, then the matrix $\mathbf{H} = \mathbf{r}\mathbf{R}^{T}$ will not be of full rank (it will be of rank 2, unless the vectors are collinear). (Note that $\mathbf{H} = \mathbf{B}^{T}$ from Brock's method.) Correspondingly the last singular value of **H** will be zero, and the SVD can be written

$$\mathbf{H} = \sigma_1\mathbf{q}_1\mathbf{s}_1^{T} + \sigma_2\mathbf{q}_2\mathbf{s}_2^{T} + 0\mathbf{q}_3\mathbf{s}_3^{T} \qquad \text{(A2.24)}$$

where σ_i is a singular value of **H** and $\mathbf{q}_i$ and $\mathbf{s}_i$ are the columns of **Q** and **S**. The zero singular value means that the sign of the last column vector of **S** can be changed without affecting the decomposition. (Neither does it affect the orthogonality of **S**). Making this change alters the sign of the determinant of $\mathbf{A}^{\circ}$, i.e. if the matrix is originally a reflection matrix, changing the sign of the last vector of **S** leads to a rotation matrix and vice versa.

If the measurements are coplanar (which they are by necessity if only two measurements are taken) there is no way of telling in advance whether the matrix $\mathbf{A}^{\circ}$ will be a reflection or rotation matrix, it is purely a matter of chance. Once $\mathbf{A}^{\circ}$ has been calculated one has to check its determinant, and if necessary change a column of **S**.

As a matter of interest, the reader may compare Arun's method with that used by Farrell and Stuelpnagel in the original reply to Wahba's Problem [66]. In this method, the **B** matrix is first decomposed into the product of orthogonal and symmetric matrices so that when the eigendecomposition of the symmetric matrix is taken, orthogonal eigenvector matrices result. SVD is much more concise in its creation of orthogonal matrices.

Lemma: For any positive definite matrix $\mathbf{DD}^{\mathrm{T}}$ and any orthogonal matrix $\mathbf{C}$

$$\mathrm{tr}(\mathbf{DD}^{\mathrm{T}}) \geq \mathrm{tr}(\mathbf{CDD}^{\mathrm{T}}) \tag{A2.25}$$

Proof: $$\mathrm{tr}(\mathbf{CDD}^{\mathrm{T}}) = \mathrm{tr}(\mathbf{D}^{\mathrm{T}}\mathbf{CD}) = \sum_{i=1}^{n} \boldsymbol{d}_i^{\mathrm{T}}\,\mathbf{C}\boldsymbol{d}_i = \sum_{i=1}^{n} \boldsymbol{d}_i.\mathbf{C}\boldsymbol{d}_i \tag{A2.26}$$

By the Cauchy-Schwarz inequality, $\boldsymbol{d}_i.\mathbf{C}\boldsymbol{d}_i \leq \| \boldsymbol{d}_i \| \, \| \mathbf{C}\boldsymbol{d}_i \| = (\boldsymbol{d}_i.\boldsymbol{d}_i)^{1/2}(\boldsymbol{d}_i^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{C}\boldsymbol{d}_i)^{1/2} = \boldsymbol{d}_i.\boldsymbol{d}_i$. Since each element of the sum is bounded by $\boldsymbol{d}_i.\boldsymbol{d}_i$ it follows that

$$\mathrm{tr}(\mathbf{CDD}^{\mathrm{T}}) = \sum_{i=1}^{n} \boldsymbol{d}_i.\mathbf{C}\boldsymbol{d}_i \leq \sum_{i=1}^{n} \boldsymbol{d}_i.\boldsymbol{d}_i = \mathrm{tr}(\mathbf{DD}^{\mathrm{T}}) \tag{A2.27}$$

which proves (A2.25).

Before considering unconstrained methods, what may be a useful trick is presented. This technique was first presented by Black [9], and is potentially useful in situations where only two vector observations are available but three are required by the orientation estimation procedure.

In possession of two linearly independent (i.e non-collinear) vector observations there are two equations,

$$\mathbf{R} = \mathbf{Ar} \tag{A2.28(a)}$$

$$\mathbf{B} = \mathbf{Ab}. \tag{A2.28(a)}$$

Black's method was to state that as well as these two equations there is also the equation that relates the vector that is normal to the two observations in either orientation:

$$\mathbf{R} \mathrm{x} \mathbf{B} = \mathbf{A}(\mathbf{r} \mathrm{x} \mathbf{b}). \qquad \text{(A2.29)}$$

Further, forming the following matrices,

$$[\mathbf{R}\ \mathbf{B}\ \mathbf{R}\mathrm{x}\mathbf{B}] = \mathbf{A}[\mathbf{r}\ \mathbf{b}\ \mathbf{r}\mathrm{x}\mathbf{b}] \qquad \text{(A2.30)}$$

it is possible to solve for **A** simply by inverting the matrix on the right. The non-singularity of the matrix [**r b rxb**] is guaranteed since the vectors **r** and **b** are linearly independent by definition, and their cross product is orthogonal to them both. Black also notes that the matrices may be made orthogonal by choosing them as follows

$$[\ \mathbf{B}\mathrm{x}(\mathbf{R}\mathrm{x}\mathbf{B}/\|\mathbf{R}\mathrm{x}\mathbf{B}\|)\ \mathbf{B}\ \mathbf{R}\mathrm{x}\mathbf{B}/\|\mathbf{R}\mathrm{x}\mathbf{B}\|\ ] = \mathbf{A}[\ \mathbf{b}\mathrm{x}(\mathbf{r}\mathrm{x}\mathbf{b}/\|\mathbf{r}\mathrm{x}\mathbf{b}\|)\ \mathbf{b}\ \mathbf{r}\mathrm{x}\mathbf{b}/\|\mathbf{r}\mathrm{x}\mathbf{b}\|\ ], \qquad \text{(A2.31)}$$

under the conditions that all vectors are normalised. Since the matrices are now orthogonal, the matrix inversion is reduced to transposition. Additionally, normalising and orthogonalising the matrices in this way forces the calculated matrix, **A**, to be orthogonal (since the product of orthogonal matrices is itself orthogonal). It will also always be a rotation matrix since it is impossible for a reflection to be associated with full rank measurement matrices. It should be noted that this method takes no account of any noise on the vector observations. Also, the cross product of two noisy vectors is intuitively 'noisier' than the original two vectors. Thus, while Black's method is useful in that it permits the calculation of a rotation matrix in the possession of only two sets of independent vector observations, the lack of optimization in the computation limits its use in the case of real sensor outputs.

A2.2 Unconstrained Orientation Estimation

Brock's unconstrained method [14] follows the same approach as the constrained method described above, with the exception that the orthogonality of **A** is not enforced. Rather than present this method, a technique that allows for a more general result is introduced. This is a simple variant on the technique of Markley and Bar-Itzhack [45]. The added generality allowed by this method is the inclusion of an *n*x*n* symmetric matrix, **W**, that weights the observations to varying degrees:

$$J(\mathbf{A}) = \mathrm{tr}[\mathbf{W}(\mathbf{R}\text{-}\mathbf{A}\mathbf{r})^{\mathrm{T}}(\mathbf{R}\text{-}\mathbf{A}\mathbf{r})]. \qquad \text{(A2.32)}$$

The method of derivation used is called a 'directional derivative', the cost function is formulated for a general matrix $\mathbf{A}_0 + \varepsilon\mathbf{H}$, where **H** is a general non-zero matrix and $\mathbf{A}_0$ is the optimal (non-orthogonal) matrix. The derivative of the cost function with respect to ε is then taken at ε = 0 so that the cost is stationary with respect to any general variation in $\mathbf{A}_0$. When the cost function

$$J(\mathbf{A}_0 + \varepsilon\mathbf{H}) = \mathrm{tr}[\mathbf{W}(\mathbf{R}\text{-}(\mathbf{A}_0 + \varepsilon\mathbf{H})\mathbf{r})^{\mathrm{T}}(\mathbf{R}\text{-}(\mathbf{A}_0 + \varepsilon\mathbf{H})\mathbf{r})] \qquad \text{(A2.33)}$$

is expanded, it results in terms that are constant, linear and quadratic in ε. Differentiating and evaluating the result at ε = 0, leaves only the coefficient of the linear term, which leaving out some trace manipulations is

$$\mathbf{H}^{\mathrm{T}}[\mathbf{A}_0\mathbf{r}\mathbf{W}\mathbf{r}^{\mathrm{T}} - \mathbf{R}\mathbf{W}\mathbf{r}^{\mathrm{T}}]. \qquad \text{(A2.34)}$$

Therefore for the cost to be invariant to any change in $\mathbf{A}_0$:

$$\mathbf{A}_0 = \mathbf{R}\mathbf{W}\mathbf{r}^{\mathrm{T}}(\mathbf{r}\mathbf{W}\mathbf{r}^{\mathrm{T}})^{-1} \qquad \text{(A2.35)}$$

In the case of an identity weighting matrix $\mathbf{A}_0 = \mathbf{R}\mathbf{r}^{\mathrm{T}}(\mathbf{r}\mathbf{r}^{\mathrm{T}})^{-1}$, which is Brock's solution for the unconstrained problem. Clearly $\mathbf{A}_0$ can only be calculated if $(\mathbf{r}\mathbf{r}^{\mathrm{T}})$ is of full rank, the requirement for this being that at least three linearly independent measurements are made (or two non-collinear measurements and Black's approach). In the case of three linearly independent measurements, **r** is a square matrix, thus $\mathbf{r}^{-1}$ exists and

$$\mathbf{A}_0 = \mathbf{RWr}^{\mathrm{T}}(\mathbf{rWr}^{\mathrm{T}})^{-1} = \mathbf{RWr}^{\mathrm{T}}(\mathbf{r})^{-\mathrm{T}}\mathbf{W}^{-1}(\mathbf{r})^{-1} = \mathbf{R}(\mathbf{r})^{-1}, \tag{A2.36}$$

which is exactly Black's solution. Note that this solution is independent of the weighting matrix, **W**, and it is therefore impossible to weight observations differently for the case $n = 3$. For vector observations to be weighted differently requires that more than three pairs of measurements be made. Markley and Bar-Itzhack do a simple error analysis for the case in which the **r** measurements are error free and the **R** measurements are subject to zero mean white noise and show that the deviation of $\mathbf{A}_0$ from orthogonality is directly related to the noise on the measurements.

Now that both constrained and unconstrained methods of attitude estimation have been presented, one may ask what the benefits of each branch are. The unconstrained methods, at least those described here, have the advantage of computational simplicity over the constrained methods (matrix inverse vs. eigendecomposition or singular value decomposition). Also, in some cases where the vector observations are very noisy, it is possible that a non-orthogonal matrix will have less error than the corresponding matrix formed via a constrained method. The advantage of creating an orthogonal matrix is the inherent structure. Each of the columns may be used directly to estimate angles of one body with respect to the other. Also, if the matrix is a rotation matrix, it is possible to find the associated axis and angle of rotation, which may be useful in some situations. When the matrix is non-orthogonal, the columns have no particular structure, although they will presumably be 'close' to having the orthogonal structure, if the vector observations are indeed able to be related by a rotation matrix. With this 'almost orthogonal' structure it is unclear which of a continuum of rotation matrices in the neighbourhood of the unconstrained solution is indeed true. If information such as included angles between the axes of the two coordinate frames, or the orientation of the axis of rotation are to be easily gathered, it is necessary that the matrix be orthogonal. Driven by the simpler computation offered by unconstrained methods, some researches have devised methods by which the yielded non-orthogonal matrices may be ‘optimally orthogonalised’. One such method is now presented.

A2.3 Orthogonalising Unconstrained Estimates

Carta and Lackowski [16] have presented a method by which unconstrained estimates can be 'orthogonalised' using Lagrange multipliers. Denoting the deviation of $\mathbf{A}_o$ (the optimal non-orthogonal matrix) from orthogonality by the matrix $\mathbf{E}$, i.e. $\mathbf{A}_o = \mathbf{A} + \mathbf{E}$, where $\mathbf{A}$ is the desired orthogonal matrix, the aim is to minimise

$$L(\mathbf{A}) = \mathrm{tr}(\mathbf{E}^T\mathbf{E}) = \mathrm{tr}[(\mathbf{A}_o - \mathbf{A})^T(\mathbf{A}_o - \mathbf{A})]. \qquad \text{(A2.37)}$$

That is, to minimise the sum of the squares of the elements of $\mathbf{E}$ (this is the square of the Frobenius norm of $\mathbf{E}$). The constraint, which is adjoined to the cost function via a symmetric Lagrange multiplier matrix $\mathbf{\Lambda}$, ensures the orthogonality of the solution matrix $\mathbf{A}$, i.e

$$L(\mathbf{A},\mathbf{\Lambda}) = \mathrm{tr}[(\mathbf{A}_o{}^T\mathbf{A}_o - 2\mathbf{A}_o{}^T\mathbf{A} + \mathbf{1} + \mathbf{\Lambda}(\mathbf{A}^T\mathbf{A}-\mathbf{1})] \qquad \text{(A2.38)}$$

This function is then differentiated with respect to $\mathbf{A}$, using the rules

$$\frac{\partial \mathrm{tr}\left(\mathbf{B}^T\mathbf{C}\right)}{\partial \mathbf{C}} = \mathbf{B} \text{ and } \frac{\partial tr\, \mathbf{D}\mathbf{C}^T\, \mathbf{C}}{\partial \mathbf{C}} = 2\,\mathbf{C}\mathbf{D} \text{ if } \mathbf{D}=\mathbf{D}^T.$$

The resulting equation is then set equal to zero to give

$$-2\mathbf{A}_o + 2\mathbf{A}\mathbf{\Lambda} = \mathbf{0}. \qquad \text{(A2.39)}$$

Thus $\mathbf{A} = \mathbf{A}_o\mathbf{\Lambda}^{-1}$. The inverse of the Lagrange multiplier matrix is found by enforcing the orthogonality of $\mathbf{A}$:

$$\begin{aligned}
&\mathbf{A}^T\mathbf{A} - \mathbf{1} = 0 \\
&(\mathbf{A}_o\mathbf{\Lambda}^{-1})^T(\mathbf{A}_o\mathbf{\Lambda}^{-1}) - \mathbf{1} = 0 \\
&\mathbf{\Lambda}^{-1}\mathbf{A}_o{}^T\mathbf{A}_o\mathbf{\Lambda}^{-1} = \mathbf{1} \\
&(\mathbf{\Lambda}^{-1}\mathbf{A}_o{}^T\mathbf{A}_o)^2 = \mathbf{A}_o{}^T\mathbf{A}_o,
\end{aligned} \qquad \text{(A2.40)}$$

where the last equality is found by multiplying on the right by $\mathbf{A}_0^T\mathbf{A}_0$. Now $\mathbf{A}^{-1}\mathbf{A}_0^T\mathbf{A}_0 = (\mathbf{A}_0^T\mathbf{A}_0)^{1/2}$ so

$$\mathbf{A}^{-1} = (\mathbf{A}_0^T\mathbf{A}_0)^{1/2}(\mathbf{A}_0^T\mathbf{A}_0)^{-1} = (\mathbf{A}_0^T\mathbf{A}_0)^{-1/2} \qquad \text{(A2.41)}$$

Giving (from (A2.39)):

$$\mathbf{A} = \mathbf{A}_0(\mathbf{A}_0^T\mathbf{A}_0)^{-1/2}. \qquad \text{(A2.41)}$$

The matrix square root is found by the same eigendecomposition method as previously described (Brock's solution), and again, all positive roots of the eigenvalues are chosen. While this method does indeed yield an orthogonal matrix from a non-orthogonal matrix, the actual benefit using an unconstrained estimate has been lost, since the computation required for the orthogonalisation process is the same as that used in Brock's method, which yields an orthogonal matrix directly.

A2.4 Other Methods of Orientation Estimation

The above algorithms are concerned with the estimation of the rotation matrix, but just as there are other methods of specifying an orientation, there are other approaches to its estimation through vector observations. Probably most notable among these alternatives is the use of quaternions [11], such an approach was taken by Horn [35]. This method, which was developed primarily for computer vision applications, is applicable to the problem at hand but is not elaborated upon, as it would require the introduction of quaternions.

Another alternative is the computation of the Rodrigues' vector, which was shown in Section 4.5 to fully describe a rotation. Only one instance of Rodrigues' vector estimation was found in a brief literature search, and the method, which was based upon an Extended Kalman Filter algorithm, required both vector observations and measurements of the angular velocity through the use of gyroscopes [37]. It was not desired that more instrumentation be added, so this method also, was not considered further.

A3 The Kalman Filter

The discrete Kalman filter is a recursive algorithm that generates the minimum mean square estimate of a vector **x**, given linearly related measurements observed with additive white noise and a discrete time state space model for the evolution of the state vector in response to a white noise input. This derivation first covers general minimum mean square estimation and properties that are required for the development of the Kalman filter algorithm. Next the problem is defined and the development of the algorithm is completed.

A3.1 Minimum Mean Square Estimation

Given a random vector **y**, we seek a linear estimate of a related random vector β, i.e. **Ky** + **b**, such that the sum of the variances of the elements of the estimation error vector are minimised. Mathematically stated, let ε be the estimation error vector:

$$\varepsilon = \mathbf{Ky} + \mathbf{b} - \beta. \tag{A3.1}$$

The covariance matrix of ε is given by $E[(\varepsilon - E[\varepsilon])(\varepsilon - E[\varepsilon])^{\mathrm{T}}]$, which has as its diagonal entries, the variances of the components of estimation error, we therefore aim to minimise the sum of the diagonal elements, i.e. the trace of the matrix

$$J = \mathrm{tr}\{E[(\varepsilon - E[\varepsilon])(\varepsilon - E[\varepsilon])^{\mathrm{T}}]\} \tag{A3.2}$$

Additionally it is desired that the estimate be unbiased, i.e. $E[\varepsilon] = \mathbf{0}$. The combination of the minimum variance and unbiased properties give the minimum mean square error estimate, which is equivalent to minimising the length of the estimation error vector in n-space. Expanding and simplifying J,

$$J = \mathrm{tr}\{E[\varepsilon\varepsilon^{\mathrm{T}}] - E[\varepsilon]E[\varepsilon]^{\mathrm{T}}\} = \mathrm{tr}E[\varepsilon\varepsilon^{\mathrm{T}}] - \mathrm{tr}E[\varepsilon]E[\varepsilon]^{\mathrm{T}} \tag{A3.3}$$

Substituting for ε and expanding the first term of (A3.3):

$$\begin{aligned}\mathrm{tr}E[\varepsilon\varepsilon^T] &= \mathrm{tr}\{E[\mathbf{K}\mathbf{y}\mathbf{y}^T\mathbf{K}^T + \mathbf{K}\mathbf{y}\mathbf{b}^T - \mathbf{K}\mathbf{y}\beta^T + \mathbf{b}\mathbf{y}^T\mathbf{K}^T + \mathbf{b}\mathbf{b}^T - \mathbf{b}\beta^T - \beta\mathbf{y}^T\mathbf{K}^T - \beta\mathbf{b}^T + \beta\beta^T]\} \\ &= \mathrm{tr}\{E[(\mathbf{K}\mathbf{y}\mathbf{y}^T\mathbf{K}^T - 2\mathbf{K}\mathbf{y}\beta^T + 2\mathbf{K}\mathbf{y}\mathbf{b}^T + \mathbf{b}\mathbf{b}^T - 2\beta\mathbf{b}^T + \beta\beta^T]\} \\ &= \mathrm{tr}\{(\mathbf{K}E[\mathbf{y}\mathbf{y}^T]\mathbf{K}^T - 2\mathbf{K}E[\mathbf{y}\beta^T] + 2\mathbf{K}E[\mathbf{y}]\mathbf{b}^T + \mathbf{b}\mathbf{b}^T - 2E[\beta]\mathbf{b}^T + E[\beta\beta^T]\}\end{aligned} \tag{A3.4}$$

where the linearity of the expectation operator has been used (also property that tr(**A**) = tr($\mathbf{A}^T$)). Similarly, expanding the second term:

$$\mathrm{tr}E[\varepsilon]E[\varepsilon]^T = \mathrm{tr}\{(\mathbf{K}E[\mathbf{y}]E[\mathbf{y}]^T\mathbf{K}^T - 2\mathbf{K}E[\mathbf{y}]E[\beta]^T + 2\mathbf{K}E[\mathbf{y}]\mathbf{b}^T + \mathbf{b}\mathbf{b}^T - 2E[\beta]\mathbf{b}^T + [\beta]E[\beta^T]\} \tag{A3.5}$$

Summing (A3.4) and (A3.5) to form J gives:

$$J = \mathrm{tr}\{\mathbf{K}E[\mathbf{y}\mathbf{y}^T]\mathbf{K}^T - 2\mathbf{K}E[\mathbf{y}\beta^T] + E[\beta\beta^T] - \mathbf{K}E[\mathbf{y}]E[\mathbf{y}]^T\mathbf{K}^T + 2\mathbf{K}E[\mathbf{y}]E[\beta^T] - E[\beta]E[\beta^T]\}. \tag{A3.6}$$

Defining the covariance matrices $\mathbf{P}_{yy} = E[\mathbf{y}\mathbf{y}^T] - E[\mathbf{y}]E[\mathbf{y}]^T$ and $\mathbf{P}_{y\beta} = E[\mathbf{y}\beta^T] - E[\mathbf{y}]E[\beta]^T$, $\mathbf{P}_{\beta\beta} = E[\beta\beta^T] - E[\beta]E[\beta]^T$, (A3.6) simplifies to

$$J = \mathrm{tr}\{\mathbf{K}\mathbf{P}_{yy}\mathbf{K}^T - 2\mathbf{K}\mathbf{P}_{y\beta} + \mathbf{P}_{\beta\beta}\}. \tag{A3.7}$$

The matrix **K** for which J is minimum is found by differentiating J with respect to **K** and setting the resulting equation to zero.

$$2\mathbf{K}\mathbf{P}_{yy} - 2\mathbf{P}_{\beta y} = \mathbf{0}, \tag{A3.8}$$

where $\mathbf{P}_{\beta y} = \mathbf{P}_{y\beta}^T$. Solving for **K** gives

$$\mathbf{K} = \mathbf{P}_{\beta y}\mathbf{P}_{yy}^{-1} \tag{A3.9}$$

In the case $E[\mathbf{y}] = E[\beta] = \mathbf{0}$, $\mathbf{P}_{\beta y} = E[\beta\mathbf{y}^T]$, $\mathbf{P}_{yy} = E[\mathbf{y}\mathbf{y}^T]$ and the result is

$$\mathbf{K} = E[\beta \mathbf{y}^T](E[\mathbf{y}\mathbf{y}^T])^{-1} \tag{A3.10}$$

While it is seen that the choice of the vector **b** does not effect the variance of the estimation error, it is required, in the case of non-zero mean vectors, to ensure that the estimate is unbiased. Having determined **K**, we are now in the position to calculate **b** (see A3.1):

$$E[\varepsilon] = E[\mathbf{K}\mathbf{y} + \mathbf{b} - \beta] = \mathbf{K}E[\mathbf{y}] - E[\beta] + \mathbf{b}. \tag{A3.11}$$

For $E[\varepsilon] = \mathbf{0}$, i.e. unbiased error:

$$\mathbf{b} = E[\beta] - \mathbf{K}E[\mathbf{y}]. \tag{A3.12}$$

Thus the minimum mean squared error estimate of β given $\mathbf{y}$ is

$$\mathbf{K}\mathbf{y} + \mathbf{b} = \mathbf{P}_{\beta y}\mathbf{P}_{yy}^{-1}(\mathbf{y} - E[\mathbf{y}]) + E[\beta] \tag{A3.13}$$

which in the case of zero mean random vectors $\mathbf{y}$ and β ($E[\mathbf{y}] = E[\beta] = \mathbf{0}$) results in

$$E[\beta \mathbf{y}^T](E[\mathbf{y}\mathbf{y}^T])^{-1}\mathbf{y} \tag{A3.14}$$

This estimate (both the non-zero and zero means cases), which will be denoted by $E^*[\beta|\mathbf{y}]$ has many important properties.

1. If $\mathbf{y}$ and β are jointly distributed Gaussian vectors $E^*[\beta|\mathbf{y}] = E[\beta|\mathbf{y}]$, i.e. the conditional expectation, and the estimate is optimal with respect to a large range of criteria [38], [2].

2. A property of the estimation error, $\varepsilon = E^*[\beta|\mathbf{y}] - \beta$, that is very useful is that $E[\varepsilon \mathbf{y}^T] = \mathbf{0}$. Random vectors that have this property are termed *orthogonal*. This is useful in simplifying terms during the derivation of the Kalman filter. The orthogonality can be shown by direct calculation:

$$\begin{aligned} E[(E^*[\beta|\mathbf{y}] - \beta)\mathbf{y}^{\mathrm{T}}] &= E[\{\mathbf{P}_{\beta y}\mathbf{P}_{yy}^{-1}(\mathbf{y} - E[\mathbf{y}]) + E[\beta] - \beta\}\mathbf{y}^{\mathrm{T}}] \\ &= \mathbf{P}_{\beta y}\mathbf{P}_{yy}^{-1}E(\mathbf{y}\mathbf{y}^{\mathrm{T}} - E[\mathbf{y}]E[\mathbf{y}]^{\mathrm{T}}) + E[\beta]E[\mathbf{y}]^{\mathrm{T}} - E[\beta\mathbf{y}^{\mathrm{T}}] \\ &= \mathbf{P}_{\beta y} - \mathbf{P}_{\beta y} = \mathbf{0} \end{aligned} \tag{A3.15}$$

It can also be shown that orthogonality is a sufficient condition for an optimal estimate [60]. It is due to the orthogonality that $E^*[\beta|\mathbf{y}]$ is known as the 'orthogonal projection' of β onto $\mathbf{y}$. Furthermore, the property of orthogonality can be used to derive the filter in a Hilbert space where the inner product is defined using the expectation operator [44].

3. The notation $E^*[\beta|\mathbf{y}_1,\mathbf{y}_2]$ implies that the estimate of β is conditioned upon two random vectors, $\mathbf{y}_1$ and $\mathbf{y}_2$. This condition is also denoted $E^*[\beta|\mathbf{Y}]$, with corresponding covariance and cross-covariance matrices $\mathbf{P}_{YY}$ and $\mathbf{P}_{\beta Y}$ are as follows

$$\begin{aligned} \mathbf{P}_{\beta Y} &= \begin{bmatrix} \mathbf{P}_{\beta y1} & \mathbf{P}_{\beta y2} \end{bmatrix} \\ \mathbf{P}_{YY} &= \begin{bmatrix} \mathbf{P}_{y1y1} & \mathbf{P}_{y1y2} \\ \mathbf{P}_{y1y2} & \mathbf{P}_{y2y2} \end{bmatrix} \end{aligned} \tag{A3.16}$$

A particularly important, and useful, case is that in which $\mathbf{y}_1$ and $\mathbf{y}_2$ are uncorrelated. In this case the covariance matrix $\mathbf{P}_{YY}$ is diagonal:

$$\mathbf{P}_{yy} = \begin{bmatrix} \mathbf{P}_{y1y1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{y2y2} \end{bmatrix}, \tag{A3.17}$$

since the off diagonal terms are of the form $\mathbf{P}_{y1y2} = E[\mathbf{y}_1\mathbf{y}_2^{\mathrm{T}}] - E[\mathbf{y}_1]E[\mathbf{y}_2]^{\mathrm{T}}$ and for uncorrelated random vectors $E[\mathbf{y}_1\mathbf{y}_2^{\mathrm{T}}] = E[\mathbf{y}_1]E[\mathbf{y}_2]^{\mathrm{T}}$. The inverse of this matrix is simply the inverse of the matrices on the diagonal. The cross-covariance matrix is of the same form as that given above:

$$\begin{aligned} E^*[\beta|\mathbf{Y}] &= \mathbf{P}_{\beta Y}\mathbf{P}_{YY}^{-1}[(\mathbf{y}_1 - E[\mathbf{y}_1])(\mathbf{y}_2 - E[\mathbf{y}_2])^{\mathrm{T}}]^{\mathrm{T}} + E[\beta] \\ &= \mathbf{P}_{\beta 1y1}\mathbf{P}_{y1y1}^{-1}(\mathbf{y}_1 - E[\mathbf{y}_1]) + \mathbf{P}_{\beta 2y2}\mathbf{P}_{y2y2}^{-1}(\mathbf{y}_2 - E[\mathbf{y}_2]) + E[\beta]. \end{aligned}$$

(A3.18)

Thus it is seen that

$$E^*[\beta|\mathbf{Y}] = E^*[\beta|\mathbf{y}_1,\mathbf{y}_2] = E^*[\beta|\mathbf{y}_1] + E^*[\beta|\mathbf{y}_2] - \mathrm{E}[\beta] \qquad \text{(A3.19)}$$

and in general

$$E^*[\beta|\mathbf{y}_1,\mathbf{y}_2,\ldots\mathbf{y}_n] = E^*[\beta|\mathbf{y}_1] + E^*[\beta|\mathbf{y}_2] + \ldots + E^*[\beta|\mathbf{y}_n] - (n-1)E[\beta] \qquad \text{(A3.20)}$$

or in the case of $E[\beta] = \mathbf{0}$:

$$E^*[\beta|\mathbf{y}_1,\mathbf{y}_2,\ldots\mathbf{y}_n] = E^*[\beta|\mathbf{y}_1] + E^*[\beta|\mathbf{y}_2] + \ldots + E^*[\beta|\mathbf{y}_n] \qquad \text{(A3.21)}$$

This has a good geometrical interpretation, especially when the **y**'s are uncorrelated *and* zero mean (which means the vectors are orthogonal, with respect to the inner product defined by the expectation operator). Projecting β onto $\{\mathbf{y}_1,\mathbf{y}_2,\ldots\mathbf{y}_n\}$ in the case in which all the **y**'s are orthogonal is the same as the sum of the projections of β onto each of the individual **y**'s. This is equivalent to finding the components of a vector in Euclidean space by taking the scalar product with each of the coordinate axes.

4. If **A** is an arbitrary non-singular square matrix, and **c** is an arbitrary vector, $E^*[\beta|\mathbf{Ay} + \mathbf{c}] = E^*[\beta|\mathbf{y}]$, i.e. conditioning on a random variable **y** is equivalent to conditioning on a linear transformation of **y**. This can be seen by direct calculation, i.e. let $\mathbf{z} = \mathbf{Ay} + \mathbf{c}$, then

$$\mathbf{z} - E[\mathbf{z}] = \mathbf{Ay} + \mathbf{c} - \mathbf{A}E[\mathbf{y}] - \mathbf{c} = \mathbf{A}(\mathbf{y} - E[\mathbf{y}]) \qquad \text{(A3.22)}$$

$$\begin{aligned}
\mathbf{P}_{\beta z} &= E[(\beta - E[\beta])(\mathbf{z} - E[\mathbf{z}])^T] = E[(\beta - E[\beta])\{\mathbf{A}(\mathbf{y} - E[\mathbf{y}])\}^T] \\
&= E[\beta\mathbf{y}\mathbf{A}^T + \beta E[\mathbf{y}]^T\mathbf{A}^T - E[\beta]\mathbf{y}\mathbf{A}^T + E[\beta]E[\mathbf{y}]^T\mathbf{A}^T] \\
&= E[\beta\mathbf{y} + \beta E[\mathbf{y}]^T - E[\beta]\mathbf{y} + E[\beta]E[\mathbf{y}]^T]\mathbf{A}^T \\
&= \mathbf{P}_{\beta y}\mathbf{A}^T
\end{aligned} \qquad \text{(A3.23)}$$

$$\begin{aligned}\mathbf{P}_{zz} &= E[(\mathbf{z} - E[\mathbf{z}])(\mathbf{z} - E[\mathbf{z}])^{\mathrm{T}}] = E[\{\mathbf{A}(\mathbf{y} - E[\mathbf{y}])\}\{\mathbf{A}(\mathbf{y} - E[\mathbf{y}])\}^{\mathrm{T}}] \\ &= E[\mathbf{A}\mathbf{y}\mathbf{y}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} - \mathbf{A}\mathbf{y}E[\mathbf{y}]^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} - \mathbf{A}E[\mathbf{y}]\mathbf{y}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} + \mathbf{A}E[\mathbf{y}]E[\mathbf{y}]^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}] \\ &= E[\mathbf{y}\mathbf{y}^{\mathrm{T}} - \mathbf{y}E[\mathbf{y}]^{\mathrm{T}} - E[\mathbf{y}]\mathbf{y}^{\mathrm{T}} + E[\mathbf{y}]E[\mathbf{y}]^{\mathrm{T}}]\mathbf{A}^{\mathrm{T}} \\ &= \mathbf{A}\mathbf{P}_{yy}\mathbf{A}^{\mathrm{T}} \qquad \text{(A3.24)}\end{aligned}$$

$$\mathbf{P}_{zz}^{-1} = \mathbf{A}^{-\mathrm{T}}\mathbf{P}_{yy}^{-1}\mathbf{A}^{-1} \qquad \text{(A3.25)}$$

$$E^*[\beta|\mathbf{z}] = \mathbf{P}_{\beta z}\mathbf{P}_{zz}^{-1}[\mathbf{z} - E[\mathbf{z}]] + E[\beta] = \mathbf{P}_{\beta y}\mathbf{A}^{\mathrm{T}}\mathbf{A}^{-\mathrm{T}}\mathbf{P}_{yy}^{-1}\mathbf{A}^{-1}\mathbf{A}(\mathbf{y} - E[\mathbf{y}]) + E[\beta] = E^*[\beta|\mathbf{y}]. \qquad \text{(A3.26)}$$

5. If random vectors are related by the general equation $\beta = \mathbf{Ac} + \mathbf{d}$, where $\mathbf{A}$ is a known matrix and $\mathbf{d}$ is a vector (a number of properties of which are discussed below). The minimum mean square estimate of β conditioned on correlated random vector $\mathbf{y}$ is given by

$$E^*[\beta|\mathbf{y}] = \mathbf{P}_{\beta y}\mathbf{P}_{yy}^{-1}(\mathbf{y} - E[\mathbf{y}]) + E[\beta], \qquad \text{(A3.27)}$$

where obviously $\mathbf{P}_{yy}$ is as before, and $\mathbf{P}_{\beta y}$ and $E[\beta]$ are given by

$$E[\beta] = \mathbf{A}E[\mathbf{c}] + E[\mathbf{d}] \qquad \text{(A3.28)}$$

$$\begin{aligned}\mathbf{P}_{\beta y} &= E[(\beta - E[\beta])(\mathbf{y} - E[\mathbf{y}])^{\mathrm{T}}] = E[\{\mathbf{A}(\mathbf{c} - E[\mathbf{c}]) + (\mathbf{d} - E[\mathbf{d}])\}(\mathbf{y} - E[\mathbf{y}])^{\mathrm{T}}] \\ &= \mathbf{A}E[(\mathbf{c} - E[\mathbf{c}])(\mathbf{y} - E[\mathbf{y}])^{\mathrm{T}}] + E[(\mathbf{d} - E[\mathbf{d}])(\mathbf{y} - E[\mathbf{y}])^{\mathrm{T}}] \\ &= \mathbf{A}\mathbf{P}_{cy} + \mathbf{P}_{dy} \qquad \text{(A3.29)}\end{aligned}$$

Substituting this into (A3.27) gives

$$\begin{aligned}E^*[\beta|\mathbf{y}] &= [\mathbf{A}\mathbf{P}_{cy} + \mathbf{P}_{dy}]\mathbf{P}_{yy}^{-1}(\mathbf{y} - E[\mathbf{y}]) + E[\beta] \\ &= \mathbf{A}\mathbf{P}_{cy}\mathbf{P}_{yy}^{-1}(\mathbf{y} - E[\mathbf{y}]) + \mathbf{A}E[\mathbf{c}] + \mathbf{P}_{dy}\mathbf{P}_{yy}^{-1}(\mathbf{y} - E[\mathbf{y}]) + E[\mathbf{d}] \\ &= \mathbf{A}E^*[\mathbf{c}|\mathbf{y}] + E^*[\mathbf{d}|\mathbf{y}] \qquad \text{(A3.30)}\end{aligned}$$

Based on this result the estimator is seen to be linear, i.e. the best estimate of a linear combination is the corresponding linear combination of estimates of the components.

An important case arises when $\mathbf{d}$ is uncorrelated with $\mathbf{y}$, meaning that $\mathbf{P}_{dy} = \mathbf{0}$. If in addition $E[\mathbf{d}] = 0$, then

$$E^*[\beta|\mathbf{y}] = \mathbf{A}E^*[\mathbf{c}|\mathbf{y}] \qquad \text{(A3.31)}$$

The definition of the linear minimum mean square estimate and the five properties listed above are the main theoretical basis for the Kalman filter.

A3.2 Kalman Filter Problem Statement

The Kalman Filter is a recursive algorithm that calculates the linear minimum mean square estimate of the state of a dynamic system driven by white noise, based on the model of the system and noisy measurements. Using the notation of the previous section, the state of the system, $\mathbf{x}$, is equivalent to the vector β, and the measurement $\mathbf{z}$ is comparable to $\mathbf{y}$. The state, $\mathbf{x}$, however is not a constant vector, but is related by the discrete time state space model:

$$\mathbf{x}(k+1) = \Phi\mathbf{x}(k) + \Gamma\mathbf{w}(k) \qquad \text{(A3.32(a))}$$

$$\mathbf{z}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{v}(k) \qquad \text{(A3.32(b))}$$

where [-](k) denotes the value of a random vector sequence at a discrete instant.

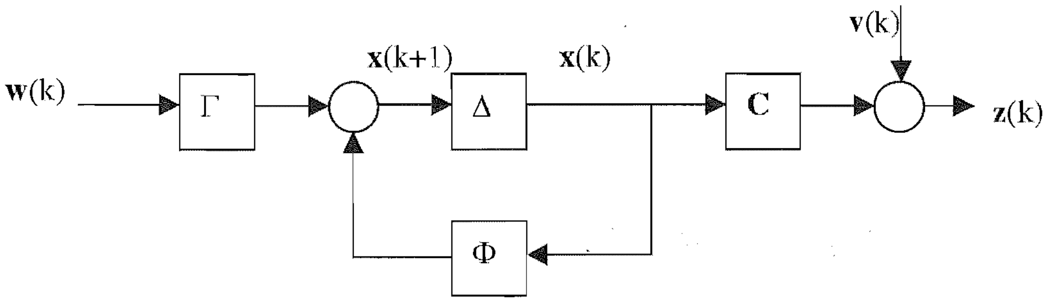


Figure A3.1 Discrete Kalman filter block diagram

The process $\{\mathbf{w}(k)\}$ is known as the *system noise* and has the following properties

$$E[\mathbf{w}(k)] = \mathbf{0}\ \forall\ k \qquad \text{(A3.33(a))}$$

$$E[\mathbf{w}(k)\mathbf{w}(j)^T] = \mathbf{Q}\delta_{jk} \qquad \text{(A3.33(b))}$$

i.e. w(k) is a zero mean white noise vector sequence (it is uncorrelated with itself from one instant to the next). Similarly, $\{\mathbf{v}(k)\}$ is the *measurement noise* with properties:

$$E[\mathbf{v}(k)] = \mathbf{0}\ \forall\ k \qquad \text{(A3.34(a))}$$

$$E[\mathbf{v}(k)\mathbf{v}(j)^T] = \mathbf{R}\delta_{jk}. \qquad \text{(A3.34(b))}$$

It is assumed that the measurement and system noise processes are uncorrelated

$$E[\mathbf{w}(k)\mathbf{v}(j)^T] = \mathbf{0} \;\forall\; k,j \qquad (A3.35)$$

and that the initial value of the state is uncorrelated with both $\{\mathbf{w}(k)\}$ and $\{\mathbf{v}(k)\}$

$$E[\mathbf{x}(0)\mathbf{w}(k)^T] = E[\mathbf{x}(0)\mathbf{v}(k)^T] = \mathbf{0} \;\forall\; k \qquad (A3.36)$$

The model for the system can either be composed in discrete time or a discretised version of a continuous model. The above model is not completely general for a number of reasons. Most importantly, the model matrices are assumed to be constant, and there is no deterministic input. These situations are easily dealt with and are commented on below. Extra generality can also be added by allowing the system and measurement noises to be correlated, or for the system noise to have a 'feed-through' term to the measurements. The above model is all that is required in this work, and simplifies the derivation. The general derivation is covered in Anderson and Moore [2].

In the Kalman filter, at each instant, k, an estimate of the state of the system is conditioned on all available data, i.e. $\mathbf{z}(j)$, j = 1, 2, ..., k. For simplicity, and to be consistent with almost all Kalman filtering literature, the following notation is adopted

$$E^*[\mathbf{x}(k)|\mathbf{z}(1),\mathbf{z}(2), \ldots,\mathbf{z}(k)] = E^*[\mathbf{x}(k)|\mathbf{Z}(k)] = \hat{\mathbf{x}}(k|k) \qquad (A3.37)$$

i.e. the estimate of the state at the kth instant based on all measurements available up to the kth instant. It is stated that the estimate $E^*[\mathbf{x}(k)|\mathbf{Z}(k)]$ is *conditioned* on $\mathbf{Z}(k)$. This is suggestive of the fact that in the Gaussian case $E^*[\mathbf{x}(k)|\mathbf{Z}(k)] = E[\mathbf{x}(k)|\mathbf{Z}(k)]$ (as mentioned in the previous section), the conditional expectation.

A3.3 Development of the Algorithm

Consider the estimate based on all available data at the $(k+1)^{th}$ instant

$$\hat{\mathbf{x}}(k+1|k+1) = E^*[\mathbf{x}(k+1)|\mathbf{z}(1),\mathbf{z}(2), \ldots,\mathbf{z}(k), \mathbf{z}(k+1)] = E^*[\mathbf{x}(k+1)|\mathbf{Z}(k+1)] \tag{A3.38}$$

The following previously derived properties are now exploited:

1. the linearity of the estimator
2. the linear minimum mean squared error estimation error is orthogonal to the data on which it is conditioned.
3. conditioning on a linear transformation of a random vector is equivalent to conditioning on the original random vector
4. an estimate based on orthogonal random variables is equal to the sum of the estimates based on the individual random variables

Using the first property note that

$$\hat{\mathbf{z}}(k+1|k) = \mathbf{C}\hat{\mathbf{x}}(k+1|k) + \hat{\mathbf{v}}(k+1|k) \tag{A3.39}$$

Consideration of the measurement equation (A3.32(b)) and the white characteristics of $\{\mathbf{v}(k)\}$ (A3.34) shows that $\mathbf{v}(k+1)$ is uncorrelated with $\mathbf{z}(j)$ $j = 1,2,..,k$ and since it is zero mean the linear minimum mean square estimate is zero, meaning (A3.39) simplifies to:

$$\hat{\mathbf{z}}(k+1|k) = \mathbf{C}\hat{\mathbf{x}}(k+1|k) \tag{A3.40}$$

The predictive error in estimating $\mathbf{z}(k+1)$ known as the *innovation* is denoted

$$\mathbf{e}_z(k+1|k) = \mathbf{z}(k+1) - \hat{\mathbf{z}}(k+1|k) \tag{A3.41}$$

Using the second property, this estimation error is orthogonal to all the data on which it is conditioned, i.e. $\mathbf{Z}(k)$. Since $\mathbf{e}_z(k+1|k)$ it is a linear transformation of $\mathbf{z}(k+1)$, the third property shows that the following is true

$$\hat{\mathbf{x}}(k+1|k+1) = E^*[\mathbf{x}(k+1)|\mathbf{Z}(k+1)] = E^*[\mathbf{x}(k+1)|\mathbf{z}(1),\mathbf{z}(2), \ldots,\mathbf{z}(k), \mathbf{z}(k+1)]$$
$$= E^*[\mathbf{x}(k+1)|\mathbf{Z}(k),\mathbf{e}_z(k+1|k)] \qquad \text{(A3.42)}$$

Finally, the fourth property shows that since $\mathbf{Z}(k)$ and $\mathbf{e}_z(k+1|k)$ are orthogonal the estimate can be expanded to the sum of the estimates based on each of these quantities:

$$\hat{\mathbf{x}}(k+1|k+1) = E^*[\mathbf{x}(k+1)|\mathbf{Z}(k),\mathbf{e}_z(k+1|k)]$$
$$= E^*[\mathbf{x}(k+1)|\mathbf{Z}(k)] + E^*[\mathbf{x}(k)|\mathbf{e}_z(k+1|k)] - E[\mathbf{x}(k)] \qquad \text{(A3.43)}$$

It should be noted that it is customary to show all conditioning as occurring on the innovations sequence, rather than the actual measurements, i.e.

$$\hat{\mathbf{x}}(k+1|k+1) = E^*[\mathbf{x}(k+1)|\mathbf{e}_z(1),\mathbf{e}_z(2),\ldots,\mathbf{e}_z(k),\mathbf{e}_z(k+1)] = E^*[\mathbf{x}(k+1)|\mathbf{E}_z(k+1)] \qquad \text{(A3.44)}$$

this has the conceptual benefit that the estimate is the sum of the projections of each of $\mathbf{x}(k+1)$ onto each of the k+1 orthogonal vectors; conditioning occurs on a white zero mean sequence. While this is theoretically and intuitively interesting, it is not required, since the estimates are generated recursively, as is now shown.

To evaluate $E^*[\mathbf{x}(k+1)|\mathbf{Z}(k+1)]$, is equivalent to calculating the sum of $E^*[\mathbf{x}(k+1)|\mathbf{Z}(k)]$ and $E^*[\mathbf{x}(k+1)|\mathbf{e}_z(k+1|k)]$. The former is known as the *predictive* estimate since it predicts the value of the state at $\mathbf{x}(k+1)$ using all *previous* measurements. The latter is the *corrective* estimate and uses only the new information at the $(k+1)^{th}$ instant. The term 'new information' is ambiguous since it means both the new measurement and, more correctly, also that part of the new measurement that could not be predicted from all previous measurements. We proceed now by evaluating the predictive and corrective estimates.

Using the linearity of the estimator and the model of the state (A3.32(a)) it is seen that the predictive estimate of the state (the estimate of the state at (k+1)th instant based on information up to the kth instant) is given by

$$\hat{\mathbf{x}}(k+1|k) = \Phi\hat{\mathbf{x}}(k|k) + \Gamma\hat{\mathbf{w}}(k|k) \qquad \text{(A3.45)}$$

Note from (A3.32(b)) and (A3.33) that $\mathbf{w}(k)$ is uncorrelated with $\mathbf{z}(k)$, therefore the linear minimum mean squared error estimate of $\mathbf{w}(k)$ is $E[\mathbf{w}(k)] = \mathbf{0}$, and

$$\hat{\mathbf{x}}(k+1|k) = \Phi\hat{\mathbf{x}}(k|k) \qquad \text{(A3.46)}$$

Before the second term can be calculated, definitions of estimation errors and the associated covariance matrices need to be made. There are estimation errors associated with both the predictive and corrected estimate, both of which are zero mean due to the unbiased characteristic of the linear minimum mean square estimator.

The prediction error is denoted

$$\mathbf{e}(k+1|k) = \mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1|k) = \Phi[\mathbf{x}(k) - \hat{\mathbf{x}}(k|k)] + \Gamma\mathbf{w}(k). \qquad \text{(A3.47)}$$

The corrected estimation error in the estimate when all possible information is included is defined

$$\mathbf{e}(k|k) = \mathbf{x}(k) - \hat{\mathbf{x}}(k|k) \qquad \text{(A3.48)}$$

and thus it is seen (using (A3.32(a)), (A3.45) and (A3.48)) that the predictive and corrected estimation errors actually fulfil the difference equation:

$$\mathbf{e}(k+1|k) = \Phi\mathbf{e}(k|k) + \Gamma\mathbf{w}(k). \qquad \text{(A3.49)}$$

The covariance matrix of the predictive estimation error (which is zero mean) is calculated as follows

$$\begin{aligned}\mathbf{P}(k+1|k) &= E[\mathbf{e}(k+1|k)\mathbf{e}(k+1|k)^T]\\ &= E[\{\Phi\mathbf{e}(k|k) + \Gamma\mathbf{w}(k)\}\{\Phi\mathbf{e}(k|k) + \Gamma\mathbf{w}(k)\}^T]\\ &= \Phi E[\mathbf{e}(k|k)\mathbf{e}(k|k)^T]\Phi^T + \Phi E[\mathbf{e}(k|k)\mathbf{w}(k)^T]\Gamma^T + \Gamma E[\mathbf{w}(k)\mathbf{e}(k|k)^T]\Phi^T\\ &\quad + \Gamma E[\mathbf{w}(k)\mathbf{w}(k)^T]\Gamma^T \qquad \text{(A3.50)}\end{aligned}$$

Examination of the definitions of the quantities $\mathbf{e}(k|k)$ and $\mathbf{w}(k)$ show that they are uncorrelated, hence (A3.50) simplifies to

$$\begin{aligned}\mathbf{P}(k+1|k) &= \Phi E[\mathbf{e}(k|k)\mathbf{e}(k|k)^T]\Phi^T + \Gamma E[\mathbf{w}(k)\mathbf{w}(k)^T]\Gamma^T\\ &= \Phi\mathbf{P}(k|k)\Phi^T + \Gamma\mathbf{Q}\Gamma^T, \qquad \text{(A3.51)}\end{aligned}$$

where $\mathbf{P}(k|k)$ is the covariance matrix associated with $\mathbf{e}(k|k)$ and $\mathbf{Q}$ is the previously defined system noise covariance matrix. Some more work needs to be done before $\mathbf{P}(k|k)$ can be evaluated, which is actually the last step of the derivation.

Recall that the filtered estimate is given by (A3.43):

$$\begin{aligned}\hat{\mathbf{x}}(k+1|k+1) &= E^*[\mathbf{x}(k+1)|\mathbf{Z}(k)] + E^*[\mathbf{x}(k)|\mathbf{e}_z(k+1|k)] - E[\mathbf{x}(k)]\\ &= \hat{\mathbf{x}}(k+1|k) + E^*[\mathbf{x}(k)|\mathbf{e}_z(k+1|k)] - E[\mathbf{x}(k)] \qquad \text{(A3.52)}\end{aligned}$$

the second part of the estimate, the component conditioned upon the innovation at the $(k+1)^{th}$ instant, is now calculated. Following from the definition of the linear minimum mean square estimator:

$$E^*[\mathbf{x}(k+1)|\mathbf{e}_z(k+1|k)] =$$
$$\mathrm{cov}[\mathbf{x}(k+1),\mathbf{e}_z(k+1)]\mathrm{cov}[\mathbf{e}_z(k+1),\mathbf{e}_z(k+1)]^{-1}(\mathbf{e}_z(k+1) - E[\mathbf{e}_z(k+1)]) + E[\mathbf{x}(k+1)]$$
$$\text{(A3.53)}$$

where cov[**a**,**b**] denotes the covariance matrix of the random vectors **a** and **b**. Note firstly that since the estimate of $\mathbf{z}(k+1)$ is unbiased, $E[\mathbf{e}_z(k+1)] = \mathbf{0}$, therefore

$$E^*[\mathbf{x}(k+1)|\mathbf{e}_z(k+1|k)] = \mathrm{cov}[\mathbf{x}(k+1),\mathbf{e}_z(k+1)]E[\mathbf{e}_z(k+1)\mathbf{e}_z(k+1)^T]^{-1}\mathbf{e}_z(k+1) + E[\mathbf{x}(k+1)] \quad (A3.54)$$

Consider the cross-covariance matrix $\mathrm{cov}[\mathbf{x}(k),\mathbf{e}_z(k+1)]$:

$$\mathrm{cov}[\mathbf{x}(k+1),\mathbf{e}_z(k+1)] = E[(\mathbf{x}(k+1) - E[\mathbf{x}(k+1)])\mathbf{e}_z(k+1)^T] \quad (A3.55)$$

using $\mathbf{x}(k+1) = \hat{\mathbf{x}}(k+1|k) - \mathbf{e}(k+1|k)$ and $\mathbf{e}_z(k+1) = \mathbf{Ce}(k+1|k)+\mathbf{v}(k+1)$:

$$\mathrm{cov}[\mathbf{x}(k+1),\mathbf{e}_z(k+1)] = E[(\hat{\mathbf{x}}(k+1|k) - \mathbf{e}(k+1|k) - E[\mathbf{x}(k+1)])(\mathbf{Ce}(k+1|k) + \mathbf{v}(k+1))^T] \quad (A3.56)$$

This expression can thankfully be significantly simplified, since the predictive estimate is orthogonal to the associated error, and $\mathbf{v}(k+1)$ is uncorrelated with both quantities. Combining these facts with $E[\mathbf{e}(k+1|k)] = \mathbf{0}$ and $E[\mathbf{v}(k+1)] = \mathbf{0}$ means (A3.55) simplifies to:

$$\mathrm{cov}[\mathbf{x}(k+1),\mathbf{e}_z(k+1)] = E[\mathbf{e}(k+1|k)\mathbf{e}(k+1|k)^T]\mathbf{C}^T = \mathbf{P}(k+1|k)\mathbf{C}^T \quad (A3.57)$$

The next matrix of (A3.54) to evaluate is $E[\mathbf{e}_z(k+1)\mathbf{e}_z(k+1)^T]$. This is achieved by again using the relation $\mathbf{e}_z(k+1) = \mathbf{Ce}(k+1|k)+\mathbf{v}(k+1)$:

$$E[\mathbf{e}_z(k+1)\mathbf{e}_z(k+1)^T] = E[\{\mathbf{Ce}(k+1|k)+\mathbf{v}(k+1)\}\{\mathbf{Ce}(k+1|k)+\mathbf{v}(k+1)\}^T] \quad (A3.58)$$

Since $\mathbf{e}(k+1|k)$ and $\mathbf{v}(k+1)$ are uncorrelated and zero mean this simplifies to give

$$\begin{aligned} E[\mathbf{e}_z(k+1)\mathbf{e}_z(k+1)^T] &= \mathbf{C}E[\mathbf{e}(k+1|k)\mathbf{e}(k+1|k)^T]\mathbf{C}^T + E[\mathbf{v}(k+1)\mathbf{v}(k+1)^T] \\ &= \mathbf{CP}(k+1|k)\mathbf{C}^T + \mathbf{R} \end{aligned} \quad (A3.59)$$

Combining these matrix definitions shows that

$$E^*[\mathbf{x}(k+1)|\mathbf{e}_z(k+1|k)] = \mathbf{P}(k+1|k)\mathbf{C}^T[\mathbf{C}\mathbf{P}(k+1|k)\mathbf{C}^T + \mathbf{R}]^{-1}\mathbf{e}_z(k+1) + E[\mathbf{x}(k+1)]. \tag{A3.60}$$

Substituting this result into the (A3.52) gives

$$\begin{aligned}\hat{\mathbf{x}}(k+1|k+1) &= E^*[\mathbf{x}(k+1)|\mathbf{Z}(k)] + E^*[\mathbf{x}(k)|\mathbf{e}_z(k+1|k)] - E[\mathbf{x}(k)] \\ &= \hat{\mathbf{x}}(k+1|k) + \mathbf{P}(k+1|k)\mathbf{C}^T[\mathbf{C}\mathbf{P}(k+1|k)\mathbf{C}^T + \mathbf{R}]^{-1}\mathbf{e}_z(k+1) \\ &= \hat{\mathbf{x}}(k+1|k) + \mathbf{K}(k+1)[\mathbf{z}(k+1) - \mathbf{C}\hat{\mathbf{x}}(k+1|k)], \end{aligned} \tag{A3.61}$$

where

$$\mathbf{K}(k+1) = \mathbf{P}(k+1|k)\mathbf{C}^T[\mathbf{C}\mathbf{P}(k+1|k)\mathbf{C}^T + \mathbf{R}]^{-1} \tag{A3.62}$$

is the Kalman gain matrix.

The only matrix of the KF algorithm yet to be evaluated is $\mathbf{P}(k|k)$, which is required for the calculation of $\mathbf{P}(k+1|k)$ and consequently $\mathbf{K}(k+1)$. Recall that $\mathbf{P}(k|k)$ is the covariance matrix associated with the corrected error measurement:

$$\mathbf{P}(k+1|k+1) = E[\mathbf{e}(k+1|k+1)\mathbf{e}(k+1|k+1)^T], \tag{A3.63}$$

where

$$\begin{aligned}\mathbf{e}(k+1|k+1) &= \mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1|k+1) \\ &= \mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1|k) + \mathbf{K}(k+1)[\mathbf{z}(k+1) - \mathbf{C}\hat{\mathbf{x}}(k+1|k)] \\ &= \mathbf{e}(k+1|k) + \mathbf{K}(k+1)[\mathbf{C}\mathbf{e}(k+1|k) + \mathbf{v}(k+1)] \\ &= [1 - \mathbf{K}(k+1)\mathbf{C}]\mathbf{e}(k+1|k) + \mathbf{K}(k+1)\mathbf{v}(k+1) \end{aligned} \tag{A3.64}$$

Letting $\mathbf{K} = \mathbf{K}(k+1)$ for brevity

$$\mathbf{P}(k+1|k+1) = E[\{[1 - \mathbf{K}\mathbf{C}]\mathbf{e}(k+1|k) + \mathbf{K}\mathbf{v}(k+1)\}\{[1 - \mathbf{K}\mathbf{C}]\mathbf{e}(k+1|k) + \mathbf{K}\mathbf{v}(k+1)\}^T] \tag{A3.65}$$

Since $\mathbf{e}(k+1|k)$ and $\mathbf{v}(k)$ are uncorrelated and zero mean the expectations of their products are equal to zero and

$$\begin{aligned}\mathbf{P}(k+1|k+1) &= [\mathbf{1}-\mathbf{KC}]E[\mathbf{e}(k+1|k)\mathbf{e}(k+1|k)^T][\mathbf{1}-\mathbf{KC}]^T+\mathbf{K}E[\mathbf{v}(k+1)\mathbf{v}(k+1)^T]\mathbf{K}^T \\ &= [\mathbf{1}-\mathbf{KC}]\mathbf{P}(k+1|k)[\mathbf{1}-\mathbf{KC}]^T+\mathbf{KRK}^T \\ &= \mathbf{P}(k+1|k)[\mathbf{1}-\mathbf{KC}]^T-\mathbf{KCP}(k+1|k)+\mathbf{KCP}(k+1|k)\mathbf{C}^T\mathbf{K}^T+\mathbf{KRK}^T \\ &= \mathbf{P}(k+1|k)[\mathbf{1}-\mathbf{KC}]^T-\mathbf{KCP}(k+1|k)+\mathbf{K}[\mathbf{CP}(k+1|k)\mathbf{C}^T+\mathbf{R}]\mathbf{K}^T\end{aligned} \tag{A3.66}$$

Since $\mathbf{K} = \mathbf{P}(k+1|k)\mathbf{C}^T[\mathbf{CP}(k+1|k)\mathbf{C}^T+\mathbf{R}]^{-1}$,

$$\mathbf{K}[\mathbf{CP}(k+1|k)\mathbf{C}+\mathbf{R}] = \mathbf{P}(k+1|k)\mathbf{C}^T, \tag{A3.67}$$

and therefore (A3.66) simplifies to:

$$\begin{aligned}\mathbf{P}(k+1|k+1) &= \mathbf{P}(k+1|k)[\mathbf{1}-\mathbf{KC}]^T-\mathbf{KCP}(k+1|k)+\mathbf{P}(k+1|k)\mathbf{C}^T\mathbf{K}^T \\ &= \mathbf{P}(k+1|k)-\mathbf{P}(k+1|k)\mathbf{C}^T\mathbf{K}^T-\mathbf{KCP}(k+1|k)+\mathbf{P}(k+1|k)\mathbf{C}^T\mathbf{K}^T \\ &= \mathbf{P}(k+1|k)-\mathbf{KCP}(k+1|k) \\ &= [\mathbf{1}-\mathbf{KC}]\mathbf{P}(k+1|k)\end{aligned} \tag{A3.68}$$

In summary the Kalman filter equations are

$$\hat{\mathbf{x}}(k+1|k) = \Phi\hat{\mathbf{x}}(k|k) \tag{A3.46}$$

$$\mathbf{P}(k+1|k) = \Phi\mathbf{P}(k|k)\Phi^T+\Gamma\mathbf{Q}\Gamma^T, \tag{A3.51}$$

$$\mathbf{K}(k+1) = \mathbf{P}(k+1|k)\mathbf{C}^T[\mathbf{CP}(k+1|k)\mathbf{C}^T+\mathbf{R}]^{-1} \tag{A3.62}$$

$$\hat{\mathbf{x}}(k+1|k+1) = \hat{\mathbf{x}}(k+1|k)+\mathbf{K}(k+1)[\mathbf{z}(k+1)-\mathbf{C}\hat{\mathbf{x}}(k+1|k)] \tag{A3.61}$$

$$\mathbf{P}(k+1|k+1) = [\mathbf{1}-\mathbf{KC}]\mathbf{P}(k+1|k) \tag{A3.68}$$

Furnished with measurements, $\mathbf{z}(k)$, a system model consisting of the state transition matrix Φ, the noise input matrix Γ, the input and measurement noise covariance matrices, $\mathbf{Q}$ and $\mathbf{R}$, an initial state estimate, and an initial estimate of the state estimation error covariance matrix $\mathbf{P}(0|0)$, this algorithm recursively generates the minimum mean squared estimate of the state for all k.

A4 Data Acquisition Hardware, Software and Protocols

The subsequent chapters describe sensors and signal processing methods that were developed for the purpose of rowing instrumentation. These sensors, of course, need to be connected to a data acquisition system, for the collection and storage of data. This brief chapter describes the hardware and software that were developed for this purpose. The method by which data was collected is also described.

Both the hardware and software were designed in consultation with the author and Dr. David Aitchison, but Julian Phillips, Julian Murphy and Dejan Metrovic performed the bulk of the work described in this chapter.

A4.1 Computer Hardware

The aim of this work was to create an instrumentation system for rowing capable of producing real time results. As such it was necessary to somehow transmit the data from the sensors to a remote viewing station. The system was therefore comprised of two computers, one that acts as a data acquisition system on board the boat (ORAC), and one that received and displayed data on the shore (Rocky). The two computers were to be linked using a wireless LAN (Local Area Network) connection, but trials showed the range to be insufficient. An alternative product for data transmission is being searched for.

A4.1.1 ORAC

The data acquisition computer was dubbed ORAC, On-the-water Rowing data Acquisition Computer. This acronym will be familiar to those aufait with archaic British science fiction.

ORAC is a fully functional PC, with 400 MHz Celeron processor, mounted on a small footprint KA-6110 motherboard, chosen for its size and multiple ISA (Industry

Standard Architecture) slots, 64 MB ram and a 7.5 GB hard drive. The case for the computer is a large polypropylene clear tub, with an aluminium lid for hear dissipation. The motherboard is fastened to the bottom of the tub, while other components, described below, are fastened to the lid. A rubber gasket is located between the lid and tub to give a reasonably watertight seal.

Two data acquisition cards were used, one for the analog sensors, and the other for the digital sensors. The NUDAQ 9114 has 32 analog channels (single ended) and 16 digital channels available. It was used only for the analog channels, of which there were a total of 19. Space is available for system expansion, the most pressing of these items is a second oar force sensor, so that sculling can be fully monitored. The 9114 occupied an ISA slot. The Universal Pulse Processor was designed and constructed by Julian Murphy of the Electronics Workshop. In this work it was used to process the quadrature output of a rotary encoder as well as the pulse type outputs of miscellaneous digital devices requiring counter/timer facilities.

In addition to the sensors described in the following chapters, the system was designed to capture linked video images. ‘Linked’ is used in the sense that the sensory data is synchronised with the video data, providing very useful information for biomechanists/coaches. A Matrox video card was used for video capture and also allowed for, in combination with a small transmitter, the use of ORAC with a portable television instead of a regular computer monitor. While this was initially seen as a good way to keep costs down, it was found, as is described below, to be too troublesome for the saving.

ORAC was controlled using a wireless keyboard (an infrared receiver is located on top of the case) or an FM wireless mouse. The keyboard normally performed well, but required a clear line of sight between the transmitter and receiver.

A number of options for display were trailed, the least successful of which was using a small wireless TV, with the transmitted within the case. A lot of time was spent manoeuvring a small aerial, with one foot in the river, sheltering the small screen from any incident light! Towards the end of the work, a flat screen monitor was trailed. In operation, one ‘docked’ with the boat and passed the monitor connection cable to the rower to connect to ORAC. This technology made data collection a lot faster. The final choice, however, was a flat touch screen, so that the user-interface and display could be combined into one.

Power is provided to the computer by an external 6.5 Ahr Lead Acid battery, which is connected through a watertight socket on the lid. Mounted on the lid are two switch-mode power supplies, one providing 3.3V, +12V and -12V lines, and the other dedicated to provide 5V. One battery allowed approximately one hour of operation. A similar battery was also used to power the flat screen display.

A4.1.2 Rocky

Rocky is the 'base unit' of the data acquisition system. While one could perhaps think of a clever acronym, the fact is that Rocky is the brand name of the ruggedised, splash-proof laptop. Apart from robustness, Rocky did not have any special features.

A4.1.3 Methods of Operation; Planned and Reality

As previously mentioned, it was desired that the system be able to produce real time results, that is results viewable at a distance from the boat, without any appreciable delay. To achieve the transmission of data, a Diamond Homefree® Wireless LAN system was purchased. This system consisted of an ISA card, that was installed within ORAC and a PCMCIA card inserted into Rocky. Within the laboratory, communication between the two machines was often difficult to initiate, a process that seemingly had to be repeated every time a new piece of hardware was installed. In its favour, some simple land based testing showed that the transmission distance was around 150 m. For some reason, this value dropped to closer to 5m as soon as one of the computers was over water. This meant that communication between the computers was not a viable option. Regardless of this fact, the planned methods of operation, one of which was luckily a 'standalone' mode are described below.

Radio-linked Operation

In this method, Rocky is used to remotely start data collection; ORAC is set adrift, powered on, with the Dataview software (described below) running, but no user intervention is required. Once data collection is initiated by the operator, ORAC processes the incoming data, writes it to hard disk and also transfers to Rocky for viewing purposes. When Rocky's user issues the command to stop data collection duplicate data files are created on Rocky and ORAC.

Standalone

ORAC is operated using the previously mentioned infrared keyboard and flat screen display. Dataview is started, data collection initiated and the monitor disconnected and then the boat is set adrift. Data collection is stopped in a similar way. Once the data is stored within ORAC, one has the option of either connecting a drive and writing the data to a floppy disk, or using the wireless LAN to transmit it to Rocky for further analysis.

A4.2 DataView

DataView is the name of the data collection and display program written by Dejan Metrovic for this work. It was written in Visual C++ to give a simple and attractive user interface while still allowing for reasonably low level commands required for fast data acquisition.

The program is operable in three modes. One collects, records and displays data, one collects, saves and sends data, and the last receives data, via the LAN, displays and saves the data. Due to the aforementioned problems with the LAN, DataView was most often run in the second mode, with no display.

During operation the traces of the input signals scroll across the screen. Most of the signals appear in their raw state, i.e. voltages, although it would be a small task to include the transformations from voltage to physical parameter. The program allows the user to toggle the display of the channels, as well as control the colour of the trace.

At the conclusion of data capture, a '.dat' file is created. Each sample is 'timestamped', and the columns of the file correspond to the recorded channel. The timestamping was performed to facilitate synchronisation between video and sensory data. At the time of writing, there were still some problems with the video capture facility.

A4.3 System Performance Summary

The construction of computer hardware and software required for this project was a large undertaking. The data acquisition system, ORAC, performs well, apart from a strange problem that was sometimes found in the collection of data from the analog channels. One channel would sporadically influence the output of the next. It was initially thought that this was caused by an insufficient delay during multiplexing, but extending the period did not remove the effect. This problem baffled all who witnessed it! ORAC has the benefit of being a total PC, which gives flexibility; an important quality for future research. The downsides of this generality are the power requirement and physical size of the computer.

The software, DataView, is also general, and if ORAC, or a successor, are reduced in size, perhaps to a 'single card' computer system, the program will still be of utility.

To offer real time results, as was initially intended, it is required that a more effective telemetry package be secured. Such a system is currently being sought.

References

1. Abu-Faraj, Z.O., Harris, G.F., Abler, J.H., Wertsch, J.J., Smith, P.A., A holter-type microprocessor-based rehabilitation instrument for acquisition and storage of plantar pressure data in children with cerebral palsy, IEEE Transactions on Rehabilitation Engineering, March 1996, 4(1), 33-37

2. Anderson, B.D.O, & Moore, J.B., Optimal Filtering, Prentice-Hall, 1979

3. Arvikar, R. & Seireg, A., Pressure distribution under the foot during static activities, Engineering in Medicine, 1980, 9(2), 99-103

4. Arun, K.S., Huang, T.S. & Blostein, S.D., Least-Squares Fitting of Two 3-D Point Sets, IEEE Transactions on Pattern Analysis and Machine Intelligence, September 1987, 9(5), 698-700

5. Atkinson, E.C., A Rowing Indicator, Natural Science, March 1896, 179-185

6. Atkinson, E.C., Some More Rowing Experiements, Natural Science, August 1898,89-102

7. Baher, H., Analog and Digital Signal Processing, 1990, John Wiley and Sons

8. Belanger, P.R., Estimation of Angular Velocity and Acceleration form Shaft Encoder Measurements', Proceedings of the 1992 IEEE International Conference on Robotics and Automation, Nice, France, May 1992, 585-592

9. Black, H.D., A Passive System for Determining the Attitude of a Satellite, AIAA Journal, July 1964, 2(7), 1350-1351

10. Bompa, T.O., Hebbelinck, M. & Van Gheluwe, B., Force Analysis of the Rowing Stroke Employing Two Differing Oar Grips, Canadian Journal of Applied Sports Sciences, (10), 2, 64-67

11. Bottema, O, & Roth, B., Theoretical Kinematics, Dover Edition, 1990

12. Bortolami, S.B., Riley, P.O. & Krebs, D.E., Numerical Differentiation of Tracking Data of Human Motion: The Virtual Accelerometer, Journal of Dynamic Systems Measurement and Control, June 1997, 119, 355-358

13. Brearley, M.N., de Mestre, N.J., Modelling the Rowing Stroke and Increasing its Efficiency

14. Brock, J.E., Optimal Matrices Describing Linear Systems, AIAA Journal, July 1968, 6(7), 1292-1296

15. Carpenter, P.S., Brown, R.H., Heinen, J.A., Schneider, S.C., On Algorithms for Velocity Estimation Using Discrete Position Encoders, Proceedings of the 1995 IEEE 21st International Conference on Industrial Electronics, Control and Instrumentation, Orlando, Florida, pp. 844-849, 1995.

16. Carta, D.G., Lackowski, D.H., Estimation of Orthogonal Transformations in Strapdown Inertial Systems, IEEE Transactions on Automatic Control, Februrary 1972, 97-100

17. Cobb, J. & Claremont, D. J., Transducers for foot pressure measurement: survey of recent developments, Medical & Biological Engineering & Computing, July 1995, 33, 525-532

18. Ctercteko, G.C., Dhanendran, M., Hutton, W.C., Le Quesne, L.P., Vertical forces acting on the feet of diabetic patients with neuropathic ulceration, British Journal of Surgery, 1981, 68, 608-614

19. Dal Monte, A. & Komor, A., Rowing and Sculling Mechanics, in Biomechanics of Sport, CRC Press, Florida, 1989, 53-119

20. Davis, B.L., Perry, J.E., Neth, D.C., Waters, K.C., A device for simultaneous measurement of pressure and shear forces distribution on the plantar surfaces of the foot, Journal of Applied Biomechanics, 1998, 14(1), 93-104

21. Dhanendran, M., Hutton, W.C., Paker, Y., The distribution of force under the human foot – an on-line measuring system, Measurement and Control, July 1978, 11, 261-264

22. Domenici, C., De Rossi, D., Bacci, A., Bennati, S., Shear stress detection in an elastic layer by a piezoelectric polymer tactile sensor, IEEE Transactions on Electrical Insulation, December 1989, 24(6), 1077-1081

23. Dudhia, A., FAQ: Basic physics of rowing, http://www-atm.ox.ac.uk/rowing/basics/htm

24. Farquhar, G., 'Ergometer Instrumentation', Third Professional Year Project, University of Canterbury

25. Fioretti, S. & Jetto, L., Accurate Derivative Estimation from Noisy Data: A State Space Approach, International Journal of Systems Science, 1989, 20 (1), 33-53

26. Franklin, G.F., Powell, J.D. & Workman M., Digital Control of Dynamic Systems, Addison Wesley Longman, 1998

27. Gerber, H., Jenny, H., Sudan, J. & Stuessi, E., Biomechanical Performance Analysis in Rowing with a New Measuring System, Biomechanics X – 10th International Congress on Biomechanics, Sweden, 1987, 721-724

28. Giacomozzi, C. & Macellari, V., Piezo-dynamometric platform for a more complete anaylsis of foot-to-floor interaction, IEEE Transactions on Rehabilitation Engineering, December 1997, 5(4), 322-330

29. Goldstein, H., Classical Mechanics, 2nd Edition, 1980, Addison-Wesley

30. Gross, T.S. & Bunch, R.P., Measurement of discrete vertical in-shoe stress with piezoelectric transducers, Journal of Biomedical Engineering, May 1988, 10, 261-265

31. Grundy, M., Tosh, P.A., McLeish, R.D., Smidt, L., An investigation of the centres of pressure under the foot while walking, The Journal of Bone and Joint Surgery, February 1975, 57B (1), 98-103

32. Hayes, M.H, Statistical digital signal processing and modeling, John Wiley and Sons, 1996

33. Hennig, M., Cavanagh, P.R., Albert, H.T., Macmillan, N.H., Piezoelectric method of measuring the vertical contact stress beneath the human foot, Journal of Biomedical Engineering, July 1982, 4, 213-222

34. Honeywell, 1&2 Axis Magnetoresistive Microcircuits, Application Notes, 5-99

35. Horn, B.K.P., Closed-Form Solution of Absolute Orientation using Unit Quaternions, Journal of the Optical Society of America, A, April 1987, 4(4), 629-642

36. Hume, P., Soper, C., Joe, G., Williams, T., Aitchison, D., Effects of foot-stretcher angle on ergometer rowing kinematics and kinetics, New Zealand Sports Science Report, 2000

37. Idan, M., Estimation of Rodrigues parameters from vector observations, IEEE Transactions on Aerospace and Electronic Systems, April 1996, 32(2), 578-586

38. Kalman, R.E., A new approach to linear filtering and prediction problems, Journal of Basic Engineering, Transactions of the ASME, D, 82(1), 1960, 35-45

39. Kleshnev, V., Propulsive Efficiency in Rowing, ISBS 1999 Proceedings

40. Lebar, A.M., Harris, G.F., Wertsch, J.J., Zhu, H., An optoelectric plantar "shear" sensing transducer: design, validation, and preliminary subject tests, IEEE Transactions on Rehabilitation Engineering, December 1996, 4(4), 310-317

41. Lin, A., Mullins, R., Pung, M. & Theofilactidis, L, Application of Accelerometers in Sports Training, WWW: http://www.analog.com.iMEMS/markets/consumer/sports_training/sports_tr.html

42. Lord, M., Foot pressure measurement: a review of methodology, Journal of Biomedical Engineering, April 1981, 3, 91-99

43. Loschner C., Smith R., The Relationship Between Seat Movement and Boat Acceleration During Sculling, ISBS 1999

44. Luenberger, David G., Optimization by Vector Space Methods, 1969, John Wiley & Sons

45. Markley, F.L. & Bar-Itzhack, I.Y., Unconstrained Optimal Transformation Matrix, IEEE Transactions on Aerospace and Electronic Systems, January 1998, 34(1), 338-340

46. McBride M.E., Elliott B. C., Use of Real-Time Telemetry to Monitor Instantaneous Seat and Boat Velocity in Paired Oared Rowing, ISBS 1999

47. McCarthy, J.M., An Introduction to Theoretical Kinematics, Massachusetts Institute of Technology, 1990

48. McCarthy, J.M., Geometric Design of Linkages, 2000, Springer-Verlag, New York

49. Macellari, V. & Giacomozzi, C., Multistep pressure platform as a stand-alone system for gait assessment, Medical & Biological Engineering & Computing, July 1996, 34,299-304

50. Martin, T.P & Bernfield, J.S., Effect of stroke rate on velocity of a rowing shell, Medicine and Science in Sports and Exercise, 12 (4), 250-256, 1980

51. Mix, Dwight F., Random Signal Processing, 1995, Prentice-Hall

52. Miyazaki, S. & Ishida, A., Capacitive transducer for continuous measurement of vertical foot force, Medical & Biological Engineering & Computing, July 1984, 309-316

53. Miyazaki, S., Ishida, A., Iwakura, H., Takino, K., Ohkawa, H., Tsubakimoto, H., Hayashi, N., Portable limb-load monitor utilizing a thin capacitive transducer, Journal of Biomedical Engineering, January 1986, (8), 67-72

54. Oppenheim, A.V. & Schafer, R.W., Discrete Time Signal Processing, International Edition, 1989, Prentice Hall

55. Pollard, J.P., Le Quesne, L.P., Tappin, J.W., Forces under the foot, Journal of Biomedical Engineering, January 1985, 5, 37-40

56. Pope, D.L. On the dynamics of men and boats and oars, Mechanics and Sport (Bleustein, J.L. Ed), 113-130, ASME 1973

57. Rose, A., Rowing rigging optimisation, Third Professional Year Project, University of Canterbury, 1999

58. Rosow, E & Davis, R.B., An instrumented ergometer for the examination of rowing biomechanics, IEEE 1989

59. Rosow, E., A Comparative Study of Rowing Biomechanics, IEEE 1991

60. Santina, Mohammed S.; Stubberud, Allen R.; Hostetter, Gene H., Digital Control System Design, 2nd Edition, 1988, Saunders College Publishing

61. Soames, R. W., Stott, J. R. R., Goodbody, A., Blake, C. D., Brewerton, D. A., Measurement of pressure under the foot during function, Medical & Biological Engineering & Computing, July 1982, 489-495

62. Stott, J. R. R., Hutton, W. C., Stokes, I. A. F., Forces under the foot, The Journal of Bone and Joint Surgery, May 1973, 55B (2), 335-344

63. Tang, X., Cai, L. & Huang, W., Acceleration Feedback of Tracking Control Based on Real Time Fourier Series, Proceedings of the American Control Conference, Philadelphia, 1998, 3322-3326

64. Teague, P.E., Gold Rush! Software, sensors, mototrs, and more help athletes go for the gold, Design News, 6-24-96, 130-134

65. Virginia Tech Mechanical Engineering, ME4016 Rowing rigger design, http://fbox.vt.edu:10021/eng/mech/tidwell/me4016/index

66. Wahba, G., Farrell, J.L., Stuelpnagel, J.C., Wessner, R.H. & Velman J.R., A Least Squares Estimate of Satellite Attitude, SIAM Review, July 1966, 8(3), 384-386

67. Williams, J.P.G. & A.C. Scott (Eds.), Rowing a Scientific Approach, 1967, Kaye and Ward

68. Williams, R.B., Porter, D., Roberts, V.C., Regan, J.F., Triaxial force transducer for investigating stresses at the stump/socket interface, Medical & Biological Engineering & Computing, January 1982, 89-96

69. Young, K. & Muirhead, R., On-Board-Shell Measurements of Acceleration, WWW: http://courses.washington.edu.phys208/shell.acceleration.htm

70. Zatsiorsky, V.M. & Yakunin, N., Mechanics and Biomechanics of Rowing: A Review, Intenational Journal of Sport Biomechanics, 7, 1991, 229-281

71. Zhu, H., Wertsch, J.J., Harris, G.F., Loftsgaarden, J.D., Price, M.B., Foot pressure distribution during walking and shuffling, Arch Phys Med Rehabil, 72, May 1991, 390-397