

**School of Mechanical, Aerospace and Civil Engineering**

# Mathematical modelling of elastic waves in bearing raceways: the forward and inverse problem

Jessica Jordan Kent

*Supervisor 1:* Dr. Artur Gower

*Supervisor 2:* Prof. Rob Dwyer-Joyce

A thesis submitted to the School of Mechanical, Aerospace and Civil Engineering, University of Sheffield in partial fulfilment of the requirements for the degree of PhD

Dedicated to the memory of Cinder Lilith Foster-Smith (Kit)

*‘Love loudly!’* 

# Declaration

---

I, Jessica Jordan Kent, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means. This work has not been previously been presented for an award at this, or any other, university.

# Acknowledgements

The following thesis comprises a collection of research papers and manuscripts to which I have significantly contributed over the course of my PhD. I would like to acknowledge the significant contributions of my co-authors Matheus de Carvalho Loures and Dr Artur Lewis Gower. Without their aid and contributions, this thesis would not have been possible. I would also like to extend my thanks to the Engineering and Physical Sciences Research Council (EPSRC) for providing the vital funding needed for this project. My thanks also extend to my group the Dynamics Research Group (DRG); the friendliness and intellectual kinship my colleagues in the DRG provided over the course of my studies have been invaluable; I have yet to meet another group of friendlier and more intellectually stimulating researchers.

On a personal note, I would like to extend my warm thanks to my friends Lily Maxfield, Adam Pye and Evelyn Marsh. Their warmth, hospitality, and friendship have steered me through the most difficult periods of my studies and personal life. Without them, I am sure that this thesis would not have been possible. I would also like to thank my wonderful and supportive partner Lauren Byrne; the love, care, patience, and understanding she has shown me has opened up wide and beautiful new vistas that I never could've experienced without her.

I would further like to acknowledge my late friend Cinder Foster-Smith; to whose memory this thesis is dedicated. Cinder was my best friend from our days in sixth form up until her death in late 2023. She was funny, warm, and burningly intelligent. Her love for philosophy, mathematics, and coding was a source of endless inspiration to me.

I credit her with showing me the great potential of programming to address mathematical problems; when we were in college, I excitedly told her of a recursion formula I had derived for calculating $n$-dimensional triangular numbers. In no time at all, she set to work writing a short Python script on her old, beaten MacBook Air with Arch Linux installed. This simple recursive script could calculate these numbers more quickly than any human could hope to. Her love and encouragement helped to make this thesis possible.

# Abstract

Rolling element bearings are vital machine components for any application involving rotating machinery; hence, being able to quickly and efficiently detect and diagnose damage in its incipient stages is paramount from an industrial, economic and health and safety point of view. Traditionally, methods for condition monitoring of rolling element bearings have been based upon signal processing methods and very limited mathematical and physical modelling. In the works contained in this thesis, we propose a new method for condition monitoring of rolling element bearings based upon mathematical modelling of the elastic waves in the bearing raceways. From this, we hope to establish the limits of what can be measured and extracted from data and provide a more detailed account of the stresses inside a bearing raceway due to defects.

# Nomenclature

| | |
|---|---|
| $\bar{a}$ | The complex conjugate of $a$ |
| $\mathrm{J}_n(x)$ | First kind Bessel function of order $n$ |
| $\boldsymbol{\sigma}$ | Cauchy stress tensor, Pa = N m$^{-2}$ |
| $\boldsymbol{\varepsilon}$ | Strain tensor |
| $\delta(t)$ | Dirac delta function |
| $\delta_{ij}$ | The Kronecker delta |
| $\mathrm{e}^x$ | The exponential function |
| $\mathrm{H}_n^{(1)}(x)$ | First kind Hankel function of order $n$ |
| $\hat{\boldsymbol{\theta}}$ | Unit vector in the azimuthal direction |
| $\hat{\boldsymbol{r}}$ | Unit vector in the radial direction |
| $\mathrm{i}$ | The imaginary unit |
| $\lambda$ | Lame parameter, Pa |
| $\mathbb{C}^{n\times n}$ | The set of $n \times n$ matrices with complex entries |
| $\mathbb{R}$ | The real numbers |
| $\mathbb{R}^{m\times n}$ | The set of $m \times n$ matrices with real entries |
| $\mathbb{Z}$ | The integers |
| $\mathbf{A}^+$ | Moore-Penrose pseudo-inverse |
| $\mathbf{A}^T$ | Transpose of the matrix $\mathbf{A}$ |
| $\mathbf{A}^\dagger$ | Hermitian conjugate |
| $\mathbf{I}$ | Identity matrix |

$\mathbf{M}_n^{\text{for}}$ Governing matrix of the forward problem, Pa m$^{-2}$

$\mathbf{M}_n^{\text{inv}}$ Governing matrix of the inverse problem

$\mathcal{F}$ 1D Fourier transform

$\mathcal{F}_{\text{2D}}$ 2D Fourier transform

$\text{tr}()$ Trace

$\text{Y}_n(x)$ Second kind Bessel function of order $n$

$\mu$ Shear modulus, Pa

$\omega$ Angular frequency, rad s$^{-1}$

$\phi$ Pressure wave potential, m$^2$

$\psi$ Scalar shear wave potential, m$^2$

$\rho$ Mass density, kg m$^{-3}$

$\sigma_{r\theta}$ Stress in the azimuthal direction, Pa

$\sigma_{rr}$ Stress in the radial direction, Pa

$\star$ Convolution

$\tilde{f}$ Fourier transform of the function $f$

$\varphi$ Loading angle, rad

$\boldsymbol{\psi}$ Vector shear wave potential

$\boldsymbol{\tau}^1$ Traction on outer boundary, Pa

$\boldsymbol{\tau}^2$ Traction on inner boundary, Pa

$\boldsymbol{a}_n$ Vector of modal coefficients of Helmholtz potentials, m$^2$

$\boldsymbol{f}_n$ Vector of modal coefficients of boundary data, Pa

$\boldsymbol{u}$ Displacement vector, m

$BPFI$ Ballpass Frequency Inner, Hz

$BPFO$ Ballpass Frequency Outer, Hz

$BSF$ Ballspin Frequency, Hz

| | |
|---|---|
| $c_p$ | Pressure wave speed, m s$^{-1}$ |
| $c_s$ | Shear wave speed, m s$^{-1}$ |
| $C_{ijkl}$ | Stiffness tensor, Pa |
| $D$ | Cage diameter, m |
| $d$ | Roller diameter, m |
| $D(f)$ | Fourier transformed Dirac comb, Pa |
| $d(t)$ | Dirac comb |
| $d_0$ | Dirac comb amplitude, Pa |
| $f_d$ | Frequency domain period, Hz |
| $f_r$ | Shaft rotation speed, Hz |
| $FTF$ | Fundamental Train Frequency, Hz |
| $k_p$ | Pressure wave number, m$^{-1}$ |
| $k_s$ | Shear wave number, m$^{-1}$ |
| $L(\theta)$ | Loading distribution |
| $RSF$ | Rollerspin Frequency, Hz |
| $T_d$ | Time domain period, s |
| $Z$ | Number of rolling elements |

# Outline

The following work consists of four review chapters, which aim to familiarise the reader with roller bearing condition monitoring, linear elasticity and elastic waves, inverse problems, and Fourier analysis. These review chapters are then followed by a methodology chapter and two research papers/manuscripts produced throughout the author's PhD; the final chapter comprises concluding remarks. A summary of each chapter can be found below:

- Chapter 1: This chapter serves as a review of the field of roller bearing condition monitoring. Common techniques, such as envelope spectrum analysis, signal processing, and data acquisition, are reviewed in this chapter. This chapter ends with a review of mathematical modelling of roller bearing vibration signals; multi-body dynamic models and Finite Element models are reviewed.

- Chapter 2: This chapter gives an overview of the mathematics of linear elasticity and elastic wave propagation. Solutions to the homogeneous and forced Helmholtz equation are discussed here.

- Chapter 3: This chapter defines what is meant by an inverse problem and what it means for such a problem to be ill-posed. We further comment on what is known about inverse problems for the Helmholtz equation.

- Chapter 4: This chapter gives a brief overview of important tools and concepts in Fourier analysis, namely, the Fourier series and the Fourier transform. Important properties and results, pertinent to Chapters 6 and 7, are discussed.

- Chapter 5: This chapter serves as the methodology section. In this chapter, we discuss mathematical tools and concepts needed for Chapter 6 and Chapter 7.

- Chapter 6: This chapter is a draft of a paper later published in Proceedings A of the Royal Society. It gives a detailed overview of the model we propose for detecting and diagnosing bearing defects and shows detailed numerical simulations for some example cases. This paper also presents theoretical and numerical results on the well-posedness of the inverse problem.

- Chapter 7: This chapter is a research manuscript that extends the work found in Chapter 6 to the localisation of subsurface cracks. Numerical simulations are presented

to show that the method in Chapter 6 can reliably localise a single point source; sources of error, such as unknown amplitudes, are analysed and quantified. We also present a deconvolutional approach for localising multiple sources simultaneously; we propose further developments and directions for research that will make this approach viable.

- Chapter 8: This chapter comprises some concluding remarks. The preceding chapters' results are summarised, and directions for further study are presented.

# List of Figures


1.1 An illustration of a roller bearing with its components labelled. . . . . 1

1.2 An illustration of a roller bearing in its housing with sensors mounted to the outside. . . . . 6

2.1 (a) $\mathrm{J}_n(x)$ plotted for $n = 0, 1, 2, 3, 4$; (b) $\mathrm{Y}_n(x)$ plotted for $n = 0, 1, 2, 3, 4$. . 19

2.2 A plot showing $\mathrm{Re}\left(\frac{\mathrm{i}}{4}\mathrm{H}_0^{(1)}(k|\boldsymbol{r} - \boldsymbol{r}'|)\right)$ at frequency $\omega = 100000$. It shows waves radiating from the source location $\boldsymbol{r}'$ in all directions in concentric rings. 20

2.3 Illustration of Graff's addition theorem. The homogeneous solution, $\phi_0$, has its origin at $O_1$, whereas the fundamental solution has its origin at $O_2$, which is located at the point $\boldsymbol{r}'$ from $O_1$. $O_2$ in this context could be a scatterer in the medium or, as in Chapter 7, the location of a point source representing emission from a crack. If we wish to compute $\phi$ at a point $\boldsymbol{r}$ from the origin $O_1$ then we must use Equation (2.3.2). . . . . 21

3.1 An illustration showing an arbitrary domain $\Omega$ and its boundary $\partial\Omega$. . . . . 29

3.2 (a) shows a typical interior problem; the problem is that of uniquely continuing the solution, represented by the arrows, from the sub-domain $\Sigma$ into the sub-domain $\Omega - \Sigma$. (b) shows a typical exterior problem; the problem is that of uniquely continuing the solution, represented by the arrows, from the sub-domain $\Omega - \Sigma$ into the sub-domain $\Sigma$. . . . . 29

5.1 Image a) on the left shows the general boundary conditions we consider. That is, we describe elastic waves in the domain $\Omega$, and will use some combination of the boundary data $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$ on the boundaries $\partial\Omega_1$ and $\partial\Omega_2$ which are defined by $r = r_1$ and $r = r_2$ respectively. The image b) on the right illustrates the geometry of one type of roller bearing, where the radii of the outer raceway are shown as an example. The methods of this paper could be used to predict the stresses between either the rollers and the outer raceway or the rollers and the inner raceway. . . . . 39

5.2 Illustration of Graff's addition theorem. $r_0$ is the distance from $\boldsymbol{O}_2$ to each source, $r$ is the distance from $\boldsymbol{O}_2$ to some observation point on the outer boundary, and $\theta$ is the angle of the observation point from the source at $\boldsymbol{r}^{(1)}$. The angles $\alpha$ are needed to relate these sources to Fourier modes later. . . 43

5.3 Figure shows the condition number plot from Figure 6.5 with the diffraction limit in Equation (5.3.12) plotted on top in red. The diffraction limit gives a rough idea of when the inverse problem is well-posed; though it is only a loose idea as we have not taken account of multiple scattering events and have performed the calculation in free space. . . . . 45

6.1 True pressure on the left for one snapshot in time. The right shows the predicted stress for the same time when using 4 sensors shown as orange spots. The sensors measure displacement, and the outer boundary is stress free, which is why the stress tends to zero when it reaches the outer boundary. However, the displacements or accelerations caused by the stress are not zero on this boundary and can be measured. The recovery is not perfect as there is a 20% error added to the boundary data. More details given in Section 6.7.3. . . . . 51

6.2 A cross section illustration of a bearing in a mounting. Showing a wave for a fixed frequency (> 10kHzs) showing waves trapped in the raceway. Further shown are sensors mounted on the raceway as we propose. . . . . 52

6.3 The left shows a Schaeffler roller bearing for the main shaft of a wind turbine [74] while the right shows an example of a Miba tilting pad journal bearing used in turbines[75]. In both examples sensors could be placed on the outside of the casing before putting the bearing in its mounting. . . . . 53

6.4 The graph on the left shows the stress on the raceway boundary $y(\theta, t)$. Each blue spike is the result of one roller being in contact with the raceway. As the rollers move in time, the blue spikes in this graph also move, but they all trace the same curve $L(\theta)$, where we assume the load supported by the bearing does not change in time. The image on the right illustrates how load is transmitted through the rollers. . . . . 62

6.5 The above heatmaps show the condition number cond $\mathbb{M}_n^{\text{for}}$ on the left and cond $\mathbb{M}_n^{\text{inv}}$ on the right, after non-dimensionalisation. In the dark regions the errors in the boundary data can be amplified by 20 times. Note that for any fixed $k_p r_1$ if we keep increasing $n$ the system will at some point become ill-conditioned. The dark cross and lines on the forward problem are resonant modes. The parameters used are given in Table 6.1. . . . . 65

6.6 On the left is the error in the traction $\tau_1$, shown by (6.7.1), with $\tau_1$ predicted by the inverse problem after adding 2% error to all boundary data. Table 6.1 shows the parameters used. The right shows the condition number of $\mathbf{M}_n^{\text{for}}$ but for a thicker raceway: $r_1 = 1.0$ and $r_2 = 1.3$, which is easier to visualise the modes, where the modes are shown in Figure 6.7 for the parameters of the orange spots. . . . . 69

6.7 Above shows the real part of the pressure field from solving the forward system (6.3.3) for the modes $n = 2$ or $6$, and $k_p r_1 = 2.5, 10.5,$ or $18.5$. Red (blue) is positive (negative) pressure, while white is no pressure. The colour scaling is different for each mode, but the outer boundary is always traction-free so it is white. The outer radius $r_2 = 1.3$ which is easier to visualise then $r_1 = 1.1$. The chosen modes are shown as an orange scatter on Figure 6.6. Middle plots show resonance so large field inside for small boundary data. 70

6.8 The left shows the convergence of solving the inverse problem, with 1% error added to its boundary conditions, to recover the force shown on the right. As the number of sensors increases so does the number of Fourier modes $n$ that can be recovered. The right shows the forcing $f(\theta)$ given by (6.7.2) compared with the mean predicted force from the inverse problem for $k_p r_1 = 65$, using 25 sensors, and with a $10$ standard deviation ribbon to show the uncertainty. 71

6.9 True radial displacement on the left for one snapshot in time when using the Stribeck equation (6.7.4) and the properties in Table 6.2. The snapshot in time is a result of taking a Fourier transform over all frequencies. The right shows the predicted radial displacement when using only 4 sensors shown as orange spots. The sensors measure displacement, and the outer boundary is stress free. We can see that despite being stress free on the boundary $r = r_2$, the displacement is not zero there. The recovery is not perfect as 4% noise is added, and limit sensors imply limited Fourier modes are recovered. . . . . 73

6.10 The blue curve shows the Stribeck equation for the loading of rollers given in (6.7.4), the orange is prior method with 4 sensors and modes -2:2 and 4% added error to boundary data, other than 20% error caused from lacking Fourier mode. Also added 4% white noise error on top. Green is also modes -2:2 and 4% added error, but with 7 sensors able to resolve and ignore higher Fourier modes. . . . . 73

6.11 The right is an illustration of elastic waves being emitted when a roller hits a defect on the raceway. The left shows the loading profile (orange and dashed) with two sharp drops in pressure due to the presence of two defects on the inner boundary. The blue curves show the absolute value of the predicted loading profile when measuring the Fourier coefficients of the boundary data $\boldsymbol{u}_n$ for $n = -6, -5, \ldots, 6$ where 2% error was added. The rollers are rotating at a rate of $\Omega = 120$ rpm and the properties used for the raceway are shown in Table 6.1. . . . . 74

7.1 (a) shows a subsurface crack; (b) shows the crack propagating to the raceway surface, leading to spalling. Adapted from [11]. . . . . 78

7.2 Here we show plots of $f(\boldsymbol{r}')$ for (a) 1% Gaussian noise; (b) 5% Gaussian noise; (c) 10% Gaussian noise. We note that in all three cases we can discern an approximate source location, though in figures (b) and (c), we note that the region of interest is wider and more diffuse than in (a), where the level of noise is relatively low. . . . . 83

7.3 Here we show plots of $f(\boldsymbol{r}')$ for (a) 5% Gaussian noise and (b) 10% Gaussian noise. We note that the recovery of the source location in (a) and (b) for 11 sensors is far clearer than in Figure 7.2 (b) and (c), where the region of interest is far wider. . . . . 84

7.4 Here we show plots of $f(\boldsymbol{r}')$ for (a) $t_0 = 10^{-2}$; (b) $t_0 = 10^{-3}$; (c) $t_0 = 10^{-4}$; (d) $t_0 = 10^{-5}$. . . . . 85

7.5 Here we plot the error in the recovered source location $\hat{\boldsymbol{r}}'$ vs the exponent of arrival times of the form $t_0 = 10^{-m}$, we see that as $t_0 \to 0$, the error decreases dramatically. In particular, for $m \geq 4$, the source can be localised to a good degree of precision with just 3 sensors. . . . . 85

7.6 Here we show plots of $f(\boldsymbol{r}')$ for (a) $t_0 = 10^{-2}$; (b) $t_0 = 10^{-3}$. These results are achieved with 67 and 9 sensors, respectively. . . . . 86

7.7 Left: The field $\phi_{\text{src}}$, defined on the domain $[0, 1] \times [0, 1]$, due to a random distribution of 10 point sources each with amplitude 1. Right: the recovered density map using the deconvolution solution in Equation (7.4.14). . . . . 89

# List of Tables

6.1 the parameter values that approximate a steel raceway. These parameters are used for most numerical examples. . . . . 66
6.2 Parameter values used for numerical simulations in Section 6.7.3. . . . . 72

7.1 The parameter values that approximate a steel raceway. These parameters are used for all numerical simulations. . . . . 83

# Contents

**Acknowledgements**

**Abstract** i

**Nomenclature** v

**Outline** vii

**1 Roller bearing condition monitoring** 1

1.1 Rolling element bearings . . . 1

1.2 Vibration and acoustic data and damage characteristics . . . 3

1.3 Envelope spectrum analysis . . . 4

1.3.1 An outline of envelope spectrum analysis . . . 5

1.3.2 Concluding remarks on envelope spectrum analysis . . . 7

1.4 Mathematical models . . . 7

1.4.1 Hertzian Contact Theory . . . 8

1.4.2 Multi-body dynamic models . . . 9

1.4.3 Finite Element Models . . . 11

1.4.4 Concluding remarks on mathematical models . . . 12

**2 Linear elasticity and elastic waves** 15

2.1 Stress, strain and generalised Hooke's law . . . 15

2.2 Waves in homogeneous, isotropic solids . . . 16

2.3 Solving the wave equation in cylindrical coordinates . . . 18

2.3.1 Harmonic solutions in 2D . . . 18

2.3.2 Fundamental solutions in 2D . . . 19

**3 Inverse Problems** 23

3.1 What is an inverse problem? . . . 23

3.2 Well-posedness . . . 24

3.3 Methods for ill-posed inverse problems . . . 25

3.3.1 The Moore-Penrose pseudo-inverse . . . 25

3.3.2 Singular value decomposition and stability . . . 26
3.3.3 Tikhonov regularisation . . . 27
3.3.4 Conclusion . . . 28
3.4 The Cauchy problem for the Helmholtz equation . . . 28

**4 A Brief Overview of Fourier Analysis** **31**
4.1 The Fourier series . . . 31
4.1.1 Periodic Functions . . . 31
4.1.2 Definition of the Fourier series . . . 32
4.1.3 Aperiodic functions? . . . 33
4.2 The Fourier transform . . . 34
4.2.1 Definition of the Fourier transform . . . 34
4.2.2 Properties of the Fourier transform . . . 34

**5 Methodology** **37**
5.1 Problem statement . . . 37
5.1.1 Governing equations . . . 37
5.2 The Forward Problem . . . 38
5.2.1 Derivation of traction components . . . 39
5.2.2 Boundary conditions . . . 40
5.3 The Inverse Problem . . . 41
5.3.1 Boundary conditions . . . 41
5.3.2 The Diffraction Limit . . . 42
5.3.3 Non-dimensional parameters . . . 44
5.3.4 Priors . . . 46

**6 Elastic waves in bearing raceways: the forward and inverse problem** **49**
6.1 Introduction . . . 49
6.2 Elastic waves in raceways . . . 52
6.2.1 Modal solution . . . 53
6.3 Boundary conditions . . . 54
6.3.1 Traction boundary conditions: the forward problem . . . 55
6.3.2 Data on only one boundary: the inverse problem . . . 56
6.3.3 Measured points on the boundary . . . 57
6.4 Priors and recovering the load from the rollers . . . 58
6.4.1 Linear priors . . . 58
6.4.2 Prior due to boundary conditions . . . 59
6.4.3 One traction free boundary . . . 61
6.5 Roller bearings and the loading profile . . . 62
6.5.1 Static Loading profile . . . 63

6.5.2 Quasi-Static Loading profile . . . . . 64
6.6 What is measurable . . . . . 65
6.6.1 Roller bearings at constant speed . . . . . 66
6.6.2 Static vs dynamic regimes . . . . . 67
6.7 Examples . . . . . 68
6.7.1 The forward and inverse modal systems . . . . . 68
6.7.2 A localised force on the boundary . . . . . 70
6.7.3 Recovering the loading profile for rollers . . . . . 72
6.7.4 Localised defect in a roller bearing . . . . . 74
6.8 Conclusions . . . . . 75

**7 Source localisation in roller bearing raceways** **77**
7.1 Introduction . . . . . 77
7.2 Modelling raceway cracks . . . . . 78
7.2.1 A single crack . . . . . 79
7.2.2 Localising a single source . . . . . 81
7.3 Numerical Simulations . . . . . 82
7.3.1 Small number of sensors . . . . . 83
7.3.2 Amplitude and source localisation . . . . . 84
7.4 Multiple cracks; theoretical results and future prospects . . . . . 86
7.4.1 Crack density . . . . . 86
7.4.2 Recovering the density maps via deconvolution . . . . . 88
7.4.3 Future prospects . . . . . 89
7.5 Conclusions . . . . . 90

**8 Conclusions** **91**
8.1 Main results . . . . . 91
8.2 Future prospects . . . . . 92

# Chapter 1

# Roller bearing condition monitoring

Jessica J. Kent

## 1.1 Rolling element bearings

Rolling element bearings are crucial machine components for a variety of industrial, commercial, consumer, and defence applications. [1, 2] Their main purposes are to constrain motion; reduce friction and support loads. As such, their maintenance is often critical. For instance, in wind turbines, the failure rate of the main bearings is high; most do not reach the 20-year lifetime they are rated for, with some failing in less than 6 years [3, 4]. It is estimated in [5] that as many as 30% of wind turbine main bearings fail before their 20-year lifetime period, leading to reduced energy generation in wind farms and costly repairs [4, 6, 7].

Roller bearings come in a variety of types, shapes, and sizes. However, every roller bearing consists of four major components: an inner raceway, an outer raceway, rolling elements and the cage; the components are marked on a diagram of a roller bearing in Figure 1.1.

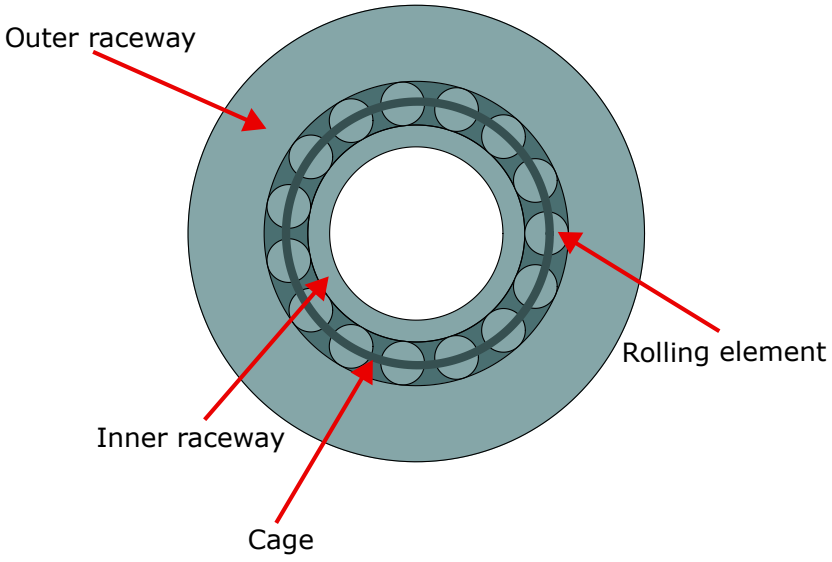



Figure 1.1: An illustration of a roller bearing with its components labelled.

In general, what distinguishes one type of bearing from another is the type of rolling element used. There are two main types: spherical and cylindrical rollers. Spherical roller bearings are used in a wide range of applications that involve both radial and axial loading. This is due to the symmetrical nature of the rollers themselves, which allows them to support more varied types of load, while cylindrical bearings are used in applications supporting large radial loads. [8]

Before surveying the literature on the topic, it is necessary to define what is meant when we talk about condition monitoring. Authors have given varying definitions; however, condition monitoring can be broadly defined as a process or tool for the detection of faults and damage at their incipient stages to prevent catastrophic damage. [9]

In [10], two main categories of roller bearing damage are identified

1. Localised faults: these types of faults are generally confined to a small area. For example, subsurface cracks [11]; spalls, where small pieces of the raceway surfaces flake off, usually due to surface breaking cracks [11].

2. Distributed defects: these defects are typically structural, e.g. raceway misalignment; raceway surface waviness; surface roughness; etc. [10]

The field of condition monitoring for rolling element bearings is rather broad and encompasses many differing techniques. In [10], the authors identify and describe five main classes of condition monitoring techniques for this problem;

1. Acoustic measurement: these methods monitor the bearing system with sensors that monitor acoustic emissions; these occur when fatigue cracks open up, generating a transient elastic wave. [12]

2. Temperature monitoring: this approach monitors the bearing housing and/or the bearing lubricant for significant temperature increases as distributed defects lead to the generation of extreme heat.

3. Motor current signature: these methods analyse the spectrum of the electrical background noise and look for changes in this noise that are characteristic of mechanical failures.

4. Wear debris analysis: these methods use sensitive sensors to identify metallic debris in the bearing lubricant. Spectroscopy can then be used to determine which metallic elements are present, which provide information about the location of defects in the bearing.

5. Vibration measurement: these methods use external sensors to monitor the passive vibrations of the bearing system. The vibrations can then be analysed using the methods discussed in Section 1.3.

For our purposes, we shall limit the scope of this review to methods based on vibration and acoustic monitoring and the signal processing techniques and algorithms used to extract damage characteristics from this data, as this is what is strictly relevant for the coming chapters.

Further, we examine mathematical models for rolling element bearing vibration signals. These encompass multi-body dynamic models, which attempt to model the entire bearing system from the point of view of Hertzian contact theory and finite element models. We find that there is a gap in the literature for a mathematical model which is both physically meaningful and useful in real-world condition monitoring applications; this model is set out in Chapters 5 to 7.

## 1.2 Vibration and acoustic data and damage characteristics

How are vibration and acoustic signals used in the detection and diagnosis of defects in rolling element bearings? The starting point of this kind of analysis is the recognition that different kinds of damage in differing locations will excite different frequencies in the signal's frequency spectrum. To put the matter more concretely, there are fundamental frequencies associated with the rates at which the rollers traverse the inner and outer raceways (the Ballpass Frequency Inner ($BPFI$) and Ballpass Frequency Outer ($BPFO$), respectively); the speed of the cage which holds the bearings (the Fundamental Train Frequency ($FTF$)), and the frequency at which the bearings rotate on their axes (the Ballspin Frequency ($BSF$) or Rollerspin Frequency ($RSF$)). [1, 13]

$$
\begin{aligned}
BPFO &= \frac{Zf_r}{2}\left\{1 - \frac{d}{D}\cos(\varphi)\right\} \\
BPFI &= \frac{Zf_r}{2}\left\{1 + \frac{d}{D}\cos(\varphi)\right\} \\
FTF &= \frac{f_r}{2}\left\{1 - \frac{d}{D}\cos(\varphi)\right\} \\
BSF(RSF) &= \frac{D}{2d}\left\{1 - \left(\frac{d}{D}\cos(\varphi)\right)^2\right\}
\end{aligned}
$$

where $Z$ is the number of rolling elements, $D$ is the diameter of the cage, $d$ is the diameter of the rolling elements, $f_r$ is the shaft's rotational speed and $\varphi$ is the angle of the load from the radial plane.

In principle, identifying these frequencies in the spectrum of vibration/acoustic emission data allows for faults to be localised to the rollers, cage or the inner/outer raceway. However, it is known that raw signals do not contain clearly identifiable diagnostic characteristics. [13, 14] Thus, a variety of signal processing techniques are employed to remove contaminants and enhance defect characteristics in the raw data. This includes the removal of periodic signals

which correspond to the motion of gears and the enhancement of any impulsive phenomena found in the data; the idea is that localised faults lead to a series of broadband impulses and that these impulses are *pseudo-cyclostationary*, i.e. that they are only approximately periodic, since the rollers randomly slip; this leads to a discrepancy in the signal on the order of 2-3%. [14, 15] This is predominantly based on the observations noted by McFadden and Smith in [15, 16]. In these papers, the authors build a model of the accelerometer signals measured from roller bearings with one or more localised defects. Their model is as follows: the signal measured from a stationary defect being periodically struck by rollers should correspond to a series of periodic impulses, i.e.,

$$d(t) = d_0 \sum_{k=-\infty}^{\infty} \delta(t - kT_d).$$

where $d_0$ is the amplitude of the impulses and $T_d$ is the time between successive impacts. It is important to note, however, that in reality, these impulses are only approximately periodic; usually, the way this is accounted for is by treating the spacing between periods as a random variable. [13, 14]

Performing a Fourier transform on the above equation, the authors deduce that the spectrum of the signal $d(t)$ is given by

$$D(f) = d_0 f_d \sum_{k=-\infty}^{\infty} \delta(f - kf_d).$$

where $f_d = 2\pi/T_d$.

In this description, we have skipped over further considerations that the authors make in their modelling, e.g. attenuation and modulation via a loading distribution. [15, 16] We do so, not because these details are unimportant, but rather to illustrate the main idea behind the methodology we highlight in the next section, i.e., that the spectrum of the signal $d(t)$ is a series of impulses which peak at multiples of the frequency $f_d$ and that by taking this together with the fundamental frequencies above suggests a powerful technique for detecting and diagnosing localised faults; namely, identifying peaks in the spectrum of a measured vibration/acoustic signal allows us to make deductions about where damage is located. This forms the basis of the high-frequency resonance technique, also known as envelope spectrum analysis. [13, 14, 15] In the next section, we will explain how this analysis is performed.

## 1.3 Envelope spectrum analysis

In the previous section, we set out the problem and laid out the theoretical work of McFadden and Smith in [15, 16]. In this section, we will discuss the benchmark methodology for detecting

and diagnosing defects in roller bearings, envelope spectrum analysis. [14]

The basic pipeline, through which damage features are extracted from raw acoustic/vibration signals, is described in [13, 14]. Essentially, we can separate the process into four main steps:

1. Raw data acquisition;
2. Isolation;
3. Enhancement;
4. Spectrum calculation.

In the following section, we shall discuss these stages of the process in some detail.

### 1.3.1 An outline of envelope spectrum analysis

**Raw data acquisition:** this stage is the measurement step of the process, through which acoustic or vibration data is acquired from a bearing in active operation. In the case of acoustic monitoring, this involves mounting acoustic sensors to the bearing housing with a low-attenuation, flexible glue, while in vibration monitoring, this involves rigidly mounted vibration sensors which directly record local vibrations. The sensor of choice in the latter method is usually an accelerometer of some sort. [9] An illustration showing sensors mounted to the bearing housing is shown in Figure 1.2. The type of data acquisition technique used depends on the types of damage that one hopes to detect [9, 13]; for example, acoustic sensors have proven to be very effective in the early detection of surface and subsurface cracks.[9] Whereas vibration monitoring has a broader range of applicability and can, given the use of appropriate pre- and post-processing techniques, detect most kinds of fault, damage and wear, although in certain applications, such as crack detection, acoustic sensors have proved superior. [9]

It is widely understood in the literature [9, 13, 14], that the spectrum of the raw signals must be processed before any relevant diagnostic information can be found; the reasons for this are numerous, for example, in the early stages of bearing damage, while high frequencies are excited, their amplitudes are close to the noise floor[9]; further, the rolling elements randomly slip as they traverse the raceways, this slippage means that the impulse response signals from a defect are not perfectly periodic, as a result, the higher harmonics in the spectrum of the raw spectrum smear over one another, making the identification of resonant frequencies corresponding to damage impossible. [13]

**Isolation:** this is the process of isolating pure bearing signals from environmental contaminants, e.g. gear signals. The methods used to isolate the bearing signals are numerous, however, the main idea present in all of these approaches is that bearing signals are pseudo-cyclostationary [13, 14], this is since the rollers randomly slip [13, 14], whereas signals from the

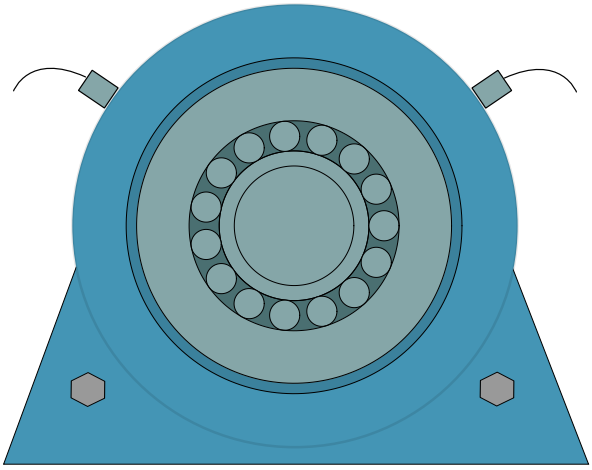


Figure 1.2: An illustration of a roller bearing in its housing with sensors mounted to the outside.

motion of gears are exactly periodic; thus, the approach often taken is that of isolating and removing periodic components in measured signals. [13] The removal of periodic noise often takes the form of employing a strategy to recover the deterministic part of the signal and then subtracting this from the raw signal; this is achieved through the use of many techniques, e.g. linear prediction; adaptive noise cancellation; self adaptive noise cancellation, etc. [13] Once this deterministic part of the signal is removed, then what remains is free of contamination from gear and other deterministic noise signals. The next step in the process is to identify impulsive phenomena in the raw signal and enhance it by choosing a demodulation band.

**Enhancement:** In this stage of the process, the isolated bearing data is enhanced to extract damage characteristics. In particular, two main enhancements are typically carried out; the first is the removal of the influence of the transfer path, i.e. to remove reflections, between the bearing and the receiver. Typically, methods based on Minimum Entropy Deconvolution are used to achieve this.[13] The idea is to find a filter function to convolve the measured signal with such that the convolved signal has a minimal entropy.[17] The assumption of minimal entropy arises from the assumption that the signal originates from a localised source, meaning the signal without the influence of multiple scattering phenomena should be highly ordered. [13, 17]

The second enhancement is the enhancement of impulsive phenomena in the data.[13] This involves the selection of an appropriate demodulation band, i.e. choosing a frequency range within which one believes defect frequencies may be found and then using this to filter out components of the signal that lie outside said demodulation band. A popular tool used in this connection is spectral kurtosis, where the kurtosis of the signal is calculated for all frequencies.[13] In [18], the authors explain that the spectral kurtosis can be used to identify transient and impulsive phenomena in vibration/acoustic data, as it takes zero values at frequencies where only Gaussian noise is present and high positive values at frequencies where impulsive and transient phenomena can be found. The authors elaborate that due

to this fact, the spectral kurtosis can be used to identify optimal demodulation bands for analysing vibration data. As a result, this tool and the methods derived from it have gained prominence in the literature. [13, 19, 20]

**Spectrum calculation:** having appropriately processed the raw data, we are now in a position to analyse the spectrum to perform diagnostics. The standard pipeline for this calculation is set out in [13, 14]; a Hilbert transform is performed on the raw signal; a band pass filter is then applied so that frequency components outside the demodulation band are excluded from consideration; this filtered signal is then transformed back to the time-domain; the envelope is then calculated in the time-domain by calculating the modulus of the signal; from there, this envelope is transformed back to the frequency domain where the defect frequencies can by identified in the frequency envelope.

### 1.3.2 Concluding remarks on envelope spectrum analysis

What we have described in this section is a basic outline of envelope spectrum analysis. This method and methods based upon it are widely used in industry. [10] Envelope spectrum analysis can pick up on most types of faults; however, it is noteworthy that it can struggle to detect advanced stages of damage. [10] This is because such signals are usually weak; to counter this issue, some have suggested that methods based on wavelet analysis might be better suited to analyse these types of signals.[10, 14, 20, 21] This points to a key issue in the condition monitoring methods that presently exist, namely, that there is no unified methodology which can detect and diagnose all possible types of damage; although envelope spectrum analysis is estimated to pick up on around 90% of potential bearing faults. [10] The method itself implicitly assumes that default signals are impulsive and resonate at high frequencies. This risks false positives and potentially misses more catastrophic damage, meaning that carefully trained experts are required to interpret the output, a major obstacle to automating detection and diagnosis. [22]

Further, envelope spectrum analysis is adapted to scenarios where only one (or very few) sensors are available. [23] This means that the diagnostic information obtained is relatively limited; the method can only give information about the component that is damaged and some limited information about the type of fault. [14] Therefore, there is a gap in the literature for a method that can give detailed information about the stresses and forces acting on the bearing.

## 1.4 Mathematical models

Mathematical modelling of rolling element bearing systems and the vibrations produced in their operation remains a popular topic in the literature.[5, 24] We begin this section with an overview of some key insights from Hertzian contact theory; we then move to discuss

multi-body dynamic models of rolling element bearing vibrations, and finally, we conclude with a section on Finite Element models.

Overall, we find that the multi-body dynamic models serve an important purpose in bearing design and simulation; however, they are ill-suited for diagnostic purposes. We also find that Finite Element models can produce detailed simulations of contact stresses and bearing vibration signals, but the considerable amount of time required to run these models is a major obstacle to their use as diagnostic tools.

### 1.4.1 Hertzian Contact Theory

Many mathematical models of bearing vibrations take Hertzian contact theory as their starting point. [5] As such, we find it pertinent to outline some key ideas which will be central to understanding the models presented in the literature. More detailed accounts can be found in [1, 25, 26].

Hertzian contact theory is primarily concerned with understanding stress and deformation when two elastic bodies are in contact. In [1, 25], we find that the normal contact stress is given by

$$Q = K\delta^n$$

where $K$ is the stiffness, $\delta$ is the radial deflection at the contact and $n = \frac{3}{2}$ for spherical rollers and $n = \frac{10}{9}$ for cylindrical rollers. The stiffness, $K$, depends on the material and geometric properties of the bodies in contact and is often evaluated numerically. [5] In reality, the contact patch between the roller and the raceway surface is not an ideal line or point, so typically, in analytic studies of rolling element bearing contact stress, the contact patch is divided into slices. Each slice is considered to have a perfect line contact with the above contact stress; these stresses are then integrated over the contact patch, resulting in a more accurate evaluation of the contact stress. [5]

Another key aspect considered is how the load is distributed among the rolling elements, as not every element is loaded equally. The exact loading distribution depends upon the type of load being supported. A key consideration for us is radial loading. In this case, rolling elements inside a loading zone of extent $-\theta_{\max} \leq \theta \leq \theta_{\max}$ are loaded along the radial direction, i.e. are pressed into the bearing raceways. In this case, the load on each rolling element is well modelled by the Stribeck equation [26], given by:

$$L(\theta) = L_0 \left(1 - \frac{1}{2\epsilon}(1 - \cos\theta)\right)^n \quad \text{for } -\theta_{\max} \leq \theta \leq \theta_{\max},$$

where $n = \frac{3}{2}$ for spherical rollers and $n = \frac{10}{9}$ for cylindrical rollers, $L_0$ is the peak load intensity and $\epsilon$ is the *load distribution factor* which determines the size of the loading zone. With this theory in place, we now move to examine the mathematical models of rolling element bearing vibration signals that can be found in the literature.

### 1.4.2 Multi-body dynamic models

A common approach to modelling vibration signals from operating rolling element bearings is a multi-body dynamic approach.[27] In these models, the contacts between the rolling elements and the bearing raceways are modelled as a series of masses connected by springs and dampers [27, 28] with the contact stiffnesses and stresses often coming directly from the Hertzian theory outlined in Section 1.4.1. Many bearing manufacturers use these models to aid in design and development, with many producing their own bespoke software packages or using commercially available packages.[5]

These models treat elements of the bearing system as masses connected by springs and dampers. In particular, it is common to model the bearing raceways as rigid masses and the contact surfaces between the rollers and the raceways as non-linear springs. [27, 28] Although in applications involving heavy loading, it is sufficient to use a linearised contact stiffness, as in this case, small changes in the load being supported lead to less marked changes in the stiffness.[29, 30]

An illustrative example of the models discussed in this section can be found in the work of Tandon and Choudhury in [31]. In this work, the authors develop an analytical model for rolling element bearing vibration signals due to a localised defect. Using Lagrange equations, the authors can describe the motion of the inner and outer raceways as a system of forced harmonic oscillators, i.e. as mass and spring systems; more specifically, they arrive at the following equations of motion for the outer raceway:

$$\ddot{q}_{i,\sin} + \omega_i^2 q_{i,\sin} = \frac{Q_{i,\sin}}{M_i},$$
$$\ddot{q}_{i,\cos} + \omega_i^2 q_{i,\cos} = \frac{Q_{i,\cos}}{M_i}$$

where, $q_{i,\sin}$ and $q_{i,\cos}$ are the generalised coordinates for the $i$th sine and cosine modes respectively and $\ddot{q}$ denotes the second derivative of $q$ with respect to time; $M_i$ are the generalised masses of the outer raceway, $\omega_i$ are the natural frequencies given by

$$\omega_i = \frac{i(i^2-1)}{\sqrt{1+i^2}}\sqrt{\frac{EI}{\rho a^4}},$$

where $E$ is the Young's modulus, $a$ is the radius of the neutral axis, $I$ is the moment of inertia of the cross section, and $\rho$ is the mass per unit length. $Q_{i,\sin}$ and $Q_{i,\cos}$ are the generalised forces for the $i$th sine and cosine modes respectively.

The authors arrive at a similar set of equations for the inner raceway

$$\ddot{q'}_{i,\sin} + \omega_i'^2 q'_{i,\sin} = \frac{Q'_{i,\sin}}{M'_i},$$
$$\ddot{q'}_{i,\cos} + \omega_i'^2 q'_{i,\cos} = \frac{Q'_{i,\cos}}{M'_i},$$

with all parameters being the inner raceway counterparts of the parameters described above.

The cornerstones of this model are the generalised forces $Q_i$ and $Q'_i$. By varying the forms of these quantities, the authors can simulate many different types of loading and localised defects common in bearing operation, e.g. an outer raceway defect or radial loading. The authors then conduct experimental tests to validate their model; they find that when predicting inner/outer raceway defect frequencies, the model compares favourably to experimental data; however, they found there were significant components of the experimental spectrum that the model could not predict, e.g. peaks at multiples of the shaft frequency.

The authors applied the model in [31] to a rotor bearing system in [30], taking into account the influence of the shaft and bearing housing on vibration signals. The major adjustment from [31] is the addition of a damping term, which arises due to the addition of the shaft and bearing housing into the model. Making this adjustment, the authors are now able to predict the signal amplitudes at multiples of the shaft frequency and overall achieve a fair agreement with experimental results.

The work in this field in recent years has mainly taken the form of incremental changes and modifications to models such as the ones in [30, 31]. This can be to account for different structures, attachments, environments, damage or other factors that may affect vibration signals. For example, Patil et al, in [32], building on the multi-body dynamic model of Sawahli and Randall in [33] and an acoustic emission model based on the work of Greenwood and Williamson on contact between rough surfaces in [34], attempt to model the vibrations from a rolling element bearing with random surface roughness.

The authors modify the multi-body dynamic model found in [33] by adding two high-frequency resonators in the $x$ and $y$ directions, respectively. These are modelled as two spring-damper systems, each having a mass $m_r$ attached. The result is six simultaneous, non-linear ODEs. This model is then fused with an acoustic emission model based on the work in [34]. This is a model for random surface roughness on the raceways of rolling element bearings. Here, surface roughness is modelled with a probability distribution for the heights, $z$, of asperities on the surface of the inner and outer races, respectively. Specifically, there are probability density functions for the heights of asperities on the inner raceway $f_i(z)$ and on the outer raceway $f_o(z)$. The probability that a selected asperity has a height between $z$

and $z + \mathrm{d}z$ is given by

$$P_i(z) = f_i(z)\mathrm{d}z,$$
$$P_o(z) = f_o(z)\mathrm{d}z.$$

The models are then fused by first solving the multi-body dynamic model and then feeding the results into the acoustic emissions model as inputs.

In [27], the authors point out that models of this type are not versatile enough to encompass the full range of factors that lead to damage and are limited in their diagnostic capabilities. We contend that these models aim to do much, to make them useful on a diagnostic level. Adding new factors, defects, etc., leads to systems of (potentially non-linear) equations with, in some cases, quite a large number of degrees of freedom [28]. While these models have utility in the realm of design [5], their complexity limits their usefulness as diagnostic tools.

### 1.4.3 Finite Element Models

In this section, we review Finite Element modelling of rolling element bearing vibrations. We note that due to the large amount of computational time and power required to run these models, they have mainly been employed to investigate rolling element contact stresses [5, 35, 36, 37]. However, there have been attempts to use Finite Element methods to model the entire bearing system [38, 39, 40]. In particular, the authors of [28] note that they could only find five papers in the literature at that time that pertained to explicit Finite Element modelling of the time domain response of rolling element bearings to a localised defect. This highlights the complexity of using Finite Element methods to address this type of problem, as modelling the response of the entire bearing system to a localised defect requires an adequate discretisation of the bearing geometry, combined with the complex multi-body dynamics associated with roller bearing systems.

The main advantage of using Finite Element methods, compared with those discussed in Section 1.4.2, is that fewer assumptions are needed to achieve good results. This is because Finite Element modelling is based upon discretisation of the domain of interest; from here, all that is required is a system of equations describing the elastic (or rigid body) behaviour in each element in the discretised domain. [41] As such, the major assumptions that need to be made pertain to the material properties of each element (rigid or elastic, material parameters, etc.). [28]

In [28], the authors sort the existing Finite Element Models found in the literature into two categories: *implicit static* and *explicit dynamic*. This categorisation is based on whether the models use implicit or explicit time integration and whether the problem considered is static or dynamic. By far, the most common type of Finite Element modelling of roller bearings

found in the literature is to model contact stresses, load distributions and load deflection relationships [5, 28, 35, 36, 37]. These models typically consist of a sphere, hemisphere, cylinder or half-cylinder in contact with an elastic half-space. For instance, in [36], the authors present an algorithm for Finite Element modelling of rolling cylindrical bodies in contact with an elastic half-space. The authors claim that the model is generalisable to linear and non-linear loading problems, i.e. it can account for light and heavy loading and deformation. In particular, they observe that the load-carrying capacity of a rolling body is reduced when travelling at higher (near critical) speeds and that in the case of moderate to high speeds, the load-deflection behaviour of the body is more non-linear than in the static case, due to inertia.

Some studies have sought to model the vibration response of the whole bearing system due to defects using Finite Element methods. [38, 39, 40] In [39], the authors produce a 2D Finite Element simulation of a roller bearing with 24 elements and a localised defect. The authors note that, in contrast to other Finite Element models found in the literature, they modelled the outer raceway surface as a flexible body; they found that this allowed for better prediction of bearing contact stiffnesses and vibration response, in particular, they found that it minimised the amplitude mismatch characteristic of other Finite Element and multi-body models found in the literature [28, 39]. In [40], Finite Element simulations of a cylindrical roller bearing with and without surface roughness are presented as a numerical validation of the multi-body dynamic model that the authors propose; in particular, they use this to analyse how contact stiffness and vibration responses are influenced by surface roughness.

### 1.4.4 Concluding remarks on mathematical models

Existing mathematical models of rolling element bearing vibrations have proven to be highly useful in the field of design. It is notable, however, that their role in condition monitoring is limited; mostly, they provide theoretically perfect signals that can be used to clean up data for envelope spectrum analysis. [13]

Finite Element models, as we have already noted, require a large amount of computational time and power [5] for their evaluation, limiting their use to theoretical analyses. These models can provide powerful insights, for example, in [39], the authors can replicate element destressing and restressing as it enters and leaves an etched line defect; something that other Finite Element models and multi-body dynamic models were unable to account for. [28, 39] This demonstrates the great potential found in mathematical modelling, which is found wanting in many techniques commonly employed in roller bearing condition monitoring.

We propose to remove much of the complexity of current models by focusing our efforts on modelling the vibrations of the roller bearing raceway(s), rather than modelling the entire bearing system, its housing, attachments, etc. It is hoped that we can preserve the great advantage of multibody models, i.e. that they can be evaluated for various kinds of loadings

and combinations of forces, while building systems of equations that are easily invertible and can provide good, detailed information about the stresses acting on this vital component. We also dispense with the underlying assumption of many of the multibody dynamic models, i.e. that the raceways are rigid, not elastic. We contend that dropping the assumption of rigidity provides deeper insights into the stresses acting on the bearing raceways, and as such, could greatly enhance the diagnostic capabilities of existing condition monitoring techniques; as we saw in [39], modelling the raceway as an elastic body leads to improved results when compared to rigid models. To properly build, understand and invert this model, there are two main topics that we must cover: elastic waves and the theory of inverse problems. These topics will be the subjects of the next two chapters.

# Chapter 2

# Linear elasticity and elastic waves

Jessica J. Kent

---

Elasticity is a branch of continuum mechanics that is concerned with how materials deform under the application of force. In this chapter, we give an overview of the theory of linear elasticity in homogeneous, isotropic solids. We confine our remarks only to what is relevant for the coming chapters.

## 2.1 Stress, strain and generalised Hooke's law

In the field of continuum mechanics, the most fundamental equation is the Cauchy momentum equation [42]

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} = \nabla \cdot \boldsymbol{\sigma} + \rho \boldsymbol{f}, \tag{2.1.1}$$

where $\boldsymbol{u}$ is displacement, $\boldsymbol{\sigma}$ is the Cauchy stress tensor, and $\boldsymbol{f}$ is a body force, e.g., gravity; for most applications and indeed for the rest of this work, we can neglect body forces. This equation furnishes us with a relationship between stress, $\boldsymbol{\sigma}$, and displacement, $\boldsymbol{u}$, describing how forces applied over the surface area of a volume lead to deformations in that volume.

The type of medium that we are dealing with is determined via a constitutive equation; this takes the form of a relationship between stress, $\boldsymbol{\sigma}$, and strain, $\varepsilon$ . Strain describes how distances in a medium change under deformation. Hence, the constitutive equation tells us how expansion and contraction lead to stresses. In elasticity, our constitutive equation is a linear relationship between stress and strain; this is generalised Hooke's law. [42]

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl} \tag{2.1.2}$$

Where $C_{ijkl}$ is the stiffness tensor and $\varepsilon_{kl}$ is the linear strain tensor, given by $\varepsilon_{kl} = \frac{1}{2}\left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k}\right)$. The form of $C_{ijkl}$ describes the stiffness and symmetries of a given material. For our purposes,

the form of the stiffness tensor we consider is that of a homogeneous, isotropic solid:

$$C_{ijkl} = \lambda\delta_{ij}\delta_{kl} + 2\mu\delta_{ik}\delta_{jl}, \tag{2.1.3}$$

where $\lambda$ is the material's Lame parameter, $\mu$ is the shear modulus and $\delta_{ij}$ is the Kronecker delta. Using Equation (2.1.3) in Equation (2.1.2) leads to the following expression for the stress tensor of a homogeneous, isotropic material

$$\sigma_{ij} = \lambda\varepsilon_{kk}\delta_{ij} + 2\mu\varepsilon_{ij}, \tag{2.1.4}$$

which we can write in tensor notation as,

$$\sigma_{ij} = \lambda\mathrm{tr}(\boldsymbol{\varepsilon})\mathbf{I} + 2\mu\boldsymbol{\varepsilon}, \tag{2.1.5}$$

where $\mathbf{I}$ is the $3 \times 3$ identity matrix. The stress tensor's main utility in this work is to compute the *traction*. We define the traction, $\boldsymbol{\tau}$, on a boundary with unit normal $\hat{\boldsymbol{n}}$ to be [42]

$$\boldsymbol{\tau} = \boldsymbol{\sigma} \cdot \hat{\boldsymbol{n}} = \sigma_n\hat{\boldsymbol{n}} + \sigma_{t_1}\hat{\boldsymbol{t}_1} + \sigma_{t_2}\hat{\boldsymbol{t}_2}, \tag{2.1.6}$$

where $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ are the basis vectors for the tangent plane to the boundary, $\sigma_n$ is the normal component of stress, and $\sigma_{t_1}$ and $\sigma_{t_2}$ are the components of stress in the $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ directions respectively. Notably, using Equation (2.1.5), we can see that normal stresses depend on tangential shear strains and that tangential stresses depend on normal strains. This fact plays an important role when we come to consider elastic waves in the next section.

If Equation (2.1.5) is used in Equation (2.1.1), an equation of motion for isotropic, homogeneous solids is reached. This is called the Navier equation[42];

$$\rho\frac{\partial^2\boldsymbol{u}}{\partial t^2} = (\lambda + \mu)\nabla(\nabla \cdot \boldsymbol{u}) + \mu\nabla^2\boldsymbol{u}. \tag{2.1.7}$$

With this equation, we are now in a position to discuss elastic waves in homogeneous, isotropic solids.

## 2.2 Waves in homogeneous, isotropic solids

In contrast to acoustic waves, elastic waves are vector waves; that is to say, they are polarised and, in contrast to electromagnetism, waves of different polarisation travel at different speeds.

In homogeneous, isotropic media, one obtains two clear, distinct types of bulk waves: P-waves (pressure waves) and S-waves (shear waves).[42] When working with materials in 3D, there are three different polarisations: P (Pressure), SV (Shear Vertical) and SH (Shear

Horizontal). In our case, we assume that the bearing raceway is of infinite extent in the $z$ direction; thus our problem is axisymmetric, meaning that only the in-plane polarisations (P and SV) are relevant to the work in the coming chapters. [43] Therefore, when describing S-waves in this work, we are referring to SV-waves.

P-waves and S-waves travel at different speeds given by

$$c_p = \sqrt{\frac{\lambda + \mu}{\rho}}, \tag{2.2.1}$$

$$c_s = \sqrt{\frac{\mu}{\rho}} \tag{2.2.2}$$

respectively.[42] From the above, it is easy to see that $c_p > c_s$, meaning that S-waves are always slower than P-waves. This fact has proved vital in the field of Seismology, where the arrival of a P-wave gives an advance warning of the approach of the far more destructive S-wave. [44] Further, P-waves propagate parallel to the direction of motion of the wavefront as a whole; S-waves, on the other hand, propagate perpendicular to the wavefront. [42] These facts about P and S-waves can be described mathematically using a Helmholtz decomposition [44]

$$\boldsymbol{u} = \nabla\phi + \nabla \times \boldsymbol{\psi}, \tag{2.2.3}$$

$$\nabla \cdot \boldsymbol{\psi} = 0. \tag{2.2.4}$$

As noted in the previous section, the normal components of the traction depend on tangential strains and vice versa. This fact has interesting consequences when considering the reflection and transmission of elastic waves from a boundary. Namely, Equation (2.1.6) and Equations (2.2.3) and (2.2.4) imply that normal and tangential stresses arise from P- and S-waves; this leads to what is termed *mode conversion* in the literature.[42] An incident wave of a single given polarisation (P or S) leads to reflected and transmitted waves of both polarisations. Mode conversions are also non-specular, meaning that the angle of incidence is not necessarily equal to the angle of reflection. [45]

Applying Equations (2.2.3) and (2.2.4) to the Navier equation leads to wave equations for $\phi$ and $\boldsymbol{\psi}$, the pressure and shear potentials, respectively:

$$\frac{\partial^2 \phi}{\partial t^2} = c_p^2 \nabla^2 \phi, \tag{2.2.5}$$

$$\frac{\partial^2 \psi}{\partial t^2} = c_s^2 \nabla^2 \psi. \tag{2.2.6}$$

this decomposition then allows us to solve for the potentials using standard methods for solving the wave equation. In the next section, we shall examine harmonic and fundamental solutions of the wave equation in cylindrical coordinates, as this is fundamental for the

coming chapters.

## 2.3 Solving the wave equation in cylindrical coordinates

In this section, we examine the solution of the wave equation in cylindrical coordinates. We limit our discussion to time-harmonic solutions, as these are the types of solutions that will be necessary for the coming chapters. This means that we take the time dependence of our solution to be given by a phase $\mathrm{e}^{-\mathrm{i}\omega t}$. Making this assumption, we may replace the wave equation by the Helmholtz equation

$$\nabla^2\phi + k^2\phi = 0$$

where $k = \frac{\omega}{c}$. The following sections will solve this equation in cylindrical coordinates. We shall only discuss solutions to the scalar wave equation, as these are the solutions that are used in all the works that follow.

### 2.3.1 Harmonic solutions in 2D

To solve the Helmholtz equation, we employ a separation of variables

$$\phi(r,\theta) = R(r)\Theta(\theta).$$

From this separation, we obtain the following equation for $\Theta$;

$$\frac{\mathrm{d}^2\Theta}{\mathrm{d}\theta^2} = -\alpha^2\Theta.$$

By inspection, we obtain $\Theta = \mathrm{e}^{\mathrm{i}\alpha\theta}$. Imposing cylindrical symmetry, $\Theta(0) = \Theta(2\pi)$, leads to the conclusion that $\alpha = n \in \mathbb{Z}$. Hence, $\Theta = \mathrm{e}^{\mathrm{i}n\theta}$. This leads us to the following equation for $R$

$$r^2\frac{\mathrm{d}^2R}{\mathrm{d}r^2} + r\frac{\mathrm{d}R}{\mathrm{d}r} + \left(r^2 - n^2\right)R = 0,$$

this is the Bessel equation; its solutions are two kinds of special functions; Bessel functions of the first kind, $\mathrm{J}_n(kr)$, and Bessel functions of the second kind, $\mathrm{Y}_n(kr)$. These special functions are plotted for different values of $n$ in Figure 2.1.

For our purposes, it is also convenient to introduce the first kind Hankel functions, $\mathrm{H}_n^{(1)}(kr)$, defined as

$$\mathrm{H}_n^{(1)}(kr) = \mathrm{J}_n(kr) + \mathrm{i}\mathrm{Y}_n(kr).$$

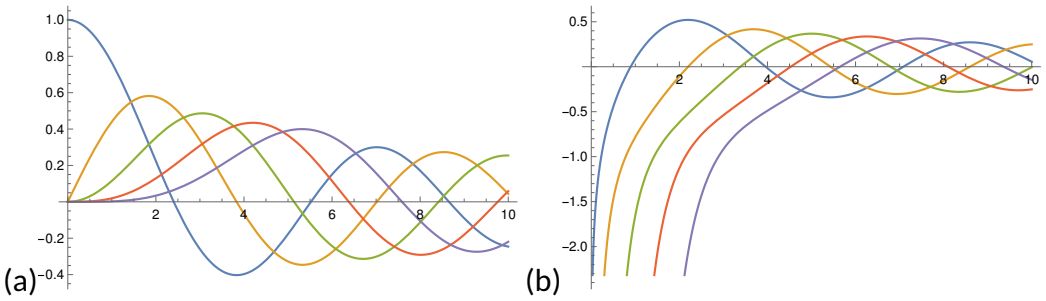



Figure 2.1: (a) $\mathrm{J}_n(x)$ plotted for $n = 0, 1, 2, 3, 4$; (b) $\mathrm{Y}_n(x)$ plotted for $n = 0, 1, 2, 3, 4$.

These functions represent outgoing waves.

The precise combination of harmonics that we use depends on the type of domain we are dealing with. If our domain is a solid cylinder with no holes, then our solutions will simply be a linear combination of first kind Bessel functions[43]:

$$\phi = \sum_{n=-\infty}^{\infty} a_n \mathrm{J}_n(kr)\mathrm{e}^{\mathrm{i}n\theta},$$

however, if our domain has a hole, then first kind Hankel functions are also necessary to obtain a complete representation of the waves inside the domain[43]:

$$\phi = \sum_{n=-\infty}^{\infty} \left(a_n \mathrm{J}_n(kr) + b_n \mathrm{H}_n^{(1)}(kr)\right) \mathrm{e}^{\mathrm{i}n\theta}.$$

### 2.3.2 Fundamental solutions in 2D

The previous section dealt with solutions to the homogeneous Helmholtz equation in cylindrical coordinates. Here, we look at the free-space Green's function for the Helmholtz equation in 2D. A Green's function is a solution to a differential equation which has a Dirac delta as its forcing term; hence, to find the Green's function for the Helmholtz equation, we must solve:

$$\nabla^2\phi + k^2\phi = \delta(\boldsymbol{r} - \boldsymbol{r}').$$

There are many such Green's function solutions for the Helmholtz equation. A form that is of particular interest is given by

$$\phi = \frac{\mathrm{i}}{4}\mathrm{H}_0^{(1)}(k|\boldsymbol{r} - \boldsymbol{r}'|).$$

This solution is of particular interest because it satisfies the Sommerfeld radiation condition [46]:

$$\lim_{r\to\infty} \sqrt{r}\left(\frac{\partial}{\partial r} - \mathrm{i}k\right)\phi(r,\theta) = 0, \tag{2.3.1}$$

thus it represents a point source centred at $\boldsymbol{r}'$ radiating in all directions, this is shown in Figure 2.2.

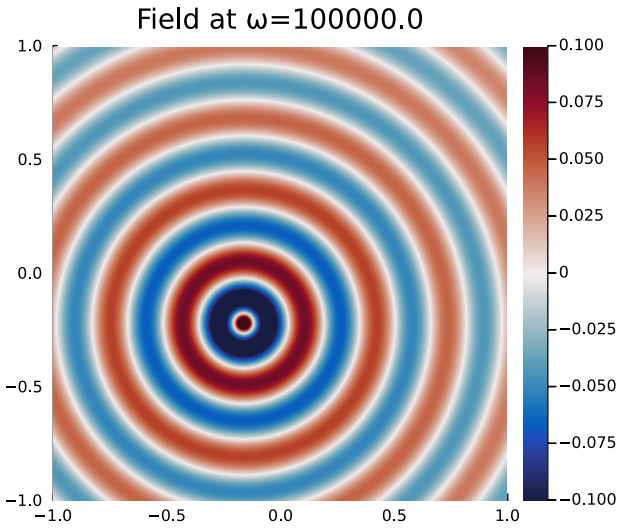



Figure 2.2: A plot showing $\mathrm{Re}\left(\frac{\mathrm{i}}{4}\mathrm{H}_0^{(1)}(k|\boldsymbol{r}-\boldsymbol{r}'|)\right)$ at frequency $\omega = 100000$. It shows waves radiating from the source location $\boldsymbol{r}'$ in all directions in concentric rings.

We can add on any homogeneous solution to the Helmholtz equation and obtain a valid solution, e.g.

$$\phi = \frac{\mathrm{i}}{4}\mathrm{H}_0^{(1)}(k|\boldsymbol{r}-\boldsymbol{r}'|) + \sum_{n=-\infty}^{\infty}\left(a_n\mathrm{J}_n(kr) + b_n\mathrm{H}_n^{(1)}(kr)\right)\mathrm{e}^{\mathrm{i}n\theta}.$$

The above representation will be useful in Chapter 7 as it will allow us to represent the emission from a subsurface crack inside a bearing raceway and its reflected waves. However, note that the homogeneous part of the solution and the Green's function are centred at different origins; the homogeneous solution at $[0,0]^T$ and the Green's function at $\boldsymbol{r}'$; to be able to apply boundary conditions, we need a representation that centres both parts of the solution at the same origin; this is the purpose of Graf's addition theorem [47, 48], which is given below

$$\phi(r,\theta) = \begin{cases} \frac{\mathrm{i}}{4}\sum_{n=-\infty}^{\infty}\mathrm{H}_{-n}^{(1)}(kr')\mathrm{J}_n(kr)\mathrm{e}^{\mathrm{i}n(\theta-\theta')} + \phi_0 & \text{for} \quad r < r', \\ \frac{\mathrm{i}}{4}\sum_{n=-\infty}^{\infty}\mathrm{J}_{-n}(kr')\mathrm{H}_n^{(1)}(kr)\mathrm{e}^{\mathrm{i}n(\theta-\theta')} + \phi_0 & \text{for} \quad r > r', \end{cases} \tag{2.3.2}$$

where $\phi_0 = \sum_{n=-\infty}^{\infty}\left(a_n\mathrm{J}_n(kr) + b_n\mathrm{H}_n^{(1)}(kr)\right)\mathrm{e}^{\mathrm{i}n\theta}$, $r = \|\boldsymbol{r}\|$ and $r' = \|\boldsymbol{r}'\|$. The relation between Equation (2.3.2), $\boldsymbol{r}$ and $\boldsymbol{r}'$ is illustrated in Figure 2.3. This representation is convenient, as it allows us to express the entire solution as a Fourier series, which, as we shall see in Chapter 7, makes applying boundary conditions far simpler.

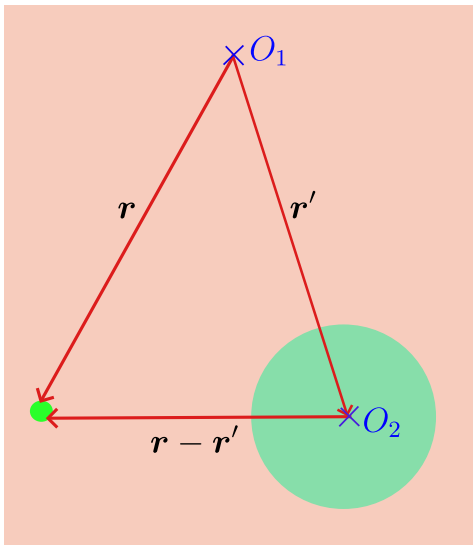



Figure 2.3: Illustration of Graff's addition theorem. The homogeneous solution, $\phi_0$, has its origin at $O_1$, whereas the fundamental solution has its origin at $O_2$, which is located at the point $\boldsymbol{r}'$ from $O_1$. $O_2$ in this context could be a scatterer in the medium or, as in Chapter 7, the location of a point source representing emission from a crack. If we wish to compute $\phi$ at a point $\boldsymbol{r}$ from the origin $O_1$ then we must use Equation (2.3.2).

# Chapter 3

# Inverse Problems

Jessica J. Kent

---

This chapter aims to introduce the concept of an inverse problem, give a basic overview of methods used in the analysis of inverse problems, and discuss what is said in the literature about a particular kind of inverse problem, which is pertinent to the rest of this work: Cauchy boundary value problems for the Helmholtz equation.

## 3.1 What is an inverse problem?

In scientific and engineering problems, two kinds of problems are of particular interest: forward problems and inverse problems. The definitions of these respective terms are interwoven; that is to say, we cannot define an inverse problem without defining a forward problem. As such, let us take as our point of departure the definition of forward problems.

A forward problem is the problem of mathematically modelling physical phenomena; i.e., it is the problem of selecting a model, of choosing governing equations and key parameters, which, if correctly chosen, predict and explain real-world data to a good degree of precision [49, 50]. Essentially, what we are searching for is some model, let us call this $\mathbf{A}$, which, when a set of parameters, say $\boldsymbol{x}$, is fed in, produces a result, $\boldsymbol{y}$, which corresponds to measured data to some precision $\boldsymbol{\epsilon}$:

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{\epsilon}.$$

An inverse problem is the problem of inverting the forward model to obtain the values of key parameters from real-world data [49, 50]. Mathematically, this corresponds to finding an inverse, $\mathbf{A}^{-1}$. As the reader might imagine, this is not always a straight-forward task; for instance, if $\mathbf{A}$ is a non-square matrix (or if $\mathbf{A}$ is a function, if $\mathbf{A}$ is not a bijection), then it is clear that the inverse problem cannot be uniquely solved; we shall return to this question in the next section.

To elucidate this distinction further, let us return to the problem of bearing diagnostics.

Recall from Chapter 1 the work of McFadden and Smith in [15, 16] attempts to model the signal measured from an accelerometer when one or multiple localised defects are present in a bearing; as such, this is a forward problem, and the authors are attempting to come up with a model which explains real-world measurements. Contrast this with the procedures found in [13, 14], where the aim is not a model that explains data but the extraction of characteristics from data which indicate the presence of a fault; this is an inverse problem, i.e., the problem of extracting useful information and key parameters from real-world data. This process is informed by the forward modelling, e.g., in [15], a series of impulses in the time domain leads to a series of impulses in frequency space, which should peak at multiples of the defect frequency, but it is an entirely separate problem.

With this distinction clearly established, a question might have occurred to the reader in the course of this section, namely, *how can we distinguish between inverse problems that are solvable and those that are not?* To answer this question, we must introduce the concept of well-posedness.

## 3.2 Well-posedness

An inverse problem is *well-posed in the sense of Hadamard* if the following three conditions are met [49]:

1. A solution to the problem exists.
2. That solution is unique.
3. That solution depends continuously upon the data.

If an inverse problem violates one or more of these three conditions, it is said to be ill-posed.

Condition 1 is self-explanatory; if there is no set of parameters such that our chosen model could reproduce our measured data, then our efforts to solve the problem will proceed in vain. Condition 2 was touched on briefly in the previous section, namely, the lack of a unique solution; this type of problem has two sources: either one has an overabundance of information (more equations than parameters) or one has an information deficit (more parameters than equations). In the first case, the problem can be remedied either by choosing a more detailed model or by a more discerning choice of information (i.e., not all measurements are strictly relevant to the problem at hand). In the latter case, the problem can be solved either by using a simpler model with fewer parameters or by adding assumptions and/or gaining additional information. In summary, conditions 1 and 2 are problems of model selection; if the forward model selected is inappropriate for the problem at hand, then conditions 1 and/or 2 fail, and the problem is ill-posed.

In practice, it is often condition 3 that fails.[49] That is to say that having selected the appropriate model, we now have to contend with the stability (or lack thereof) of our solution.

That is, we have to confront the question of our model's sensitivity to error. In the next section, we treat discuss the question of numerical instability and and how to treat it with regularisation.

## 3.3 Methods for ill-posed inverse problems

In the previous section, we discussed Hadamard's criteria for well-posed inverse problems. In this section, we expand on the treatment of ill-posed inverse problems. In particular, the treatment of inverse problems that are numerically unstable. We begin by defining the Moore-Penrose pseudo-inverse. We then move to describe singular value decomposition and its use in identifying unstable inverse problems; this leads on to a discussion of Tikhonov regularisation, a modification of the Moore-Penrose pseudo-inverse that deals with instability arising from small singular values.

### 3.3.1 The Moore-Penrose pseudo-inverse

When fitting a model to measured data, the simplest approach we can take is least squares regression. In this approach, given data $\boldsymbol{y}$, parameters $\boldsymbol{x}$ and a matrix $\mathbf{A}$ representing the model we seek to fit parameters $\hat{\boldsymbol{x}}$, such that, the square of the error between the model and the data is minimised

$$L = \|\boldsymbol{y} - \mathbf{A}\boldsymbol{x}\|^2. \tag{3.3.1}$$

To minimise $L$, we must first take the derivative with respect to $\boldsymbol{x}$ and then set this equal to $0$, i.e. we must solve

$$\frac{\partial L}{\partial \boldsymbol{x}} = 0.$$

Equation (3.3.1) can be expressed as

$$L = (\boldsymbol{y} - \mathbf{A}\boldsymbol{x})^\dagger(\boldsymbol{y} - \mathbf{A}\boldsymbol{x}), \tag{3.3.2}$$

where † denotes Hermitian conjugation, i.e. the complex conjugate of $(\boldsymbol{y} - \mathbf{A}\boldsymbol{x})^T$. Taking the derivative of Equation (3.3.2) with respect to $\boldsymbol{x}$ leads to

$$\frac{\partial L}{\partial \boldsymbol{x}} = 2\mathbf{A}^\dagger\mathbf{A}\boldsymbol{x} - 2\mathbf{A}^\dagger\boldsymbol{y}. \tag{3.3.3}$$

Setting Equation (3.3.3) equal to $0$ and rearranging leads to the following solution for $\hat{\boldsymbol{x}}$, the best fit parameters;

$$\hat{\boldsymbol{x}} = \mathbf{A}^+\boldsymbol{y}, \tag{3.3.4}$$

where,

$$\mathbf{A}^+ = (\mathbf{A}^\dagger\mathbf{A})^{-1}\mathbf{A}^\dagger. \tag{3.3.5}$$

Equation (3.3.5) is known as the *Moore-Penrose pseudo-inverse* or simply the pseudo-inverse. [50] It is a generalisation of the inverse of a square matrix to non-square matrices. Notably, the solution in Equation (3.3.4) is not unique if the matrix $\mathbf{A}$ is not square, in fact adding any $\boldsymbol{x}_0 \in \mathrm{Ker}(\mathbf{A})$ will produce identical results, since $\mathbf{A}\boldsymbol{x}_0 = 0$ for all $\boldsymbol{x}_0 \in \mathrm{Ker}(\mathbf{A})$.[50] This means that additional constraints are necessary to produce a unique solution.

### 3.3.2 Singular value decomposition and stability

A tool often used in analysing the stability of inverse problems is singular value decomposition (SVD). SVD allows us to analyse the numerical stability of the model represented by the matrix $\mathbf{A}$ by decomposing it as a product of three matrices

$$\mathbf{A} = \mathbf{U}^\dagger \mathbf{\Sigma} \mathbf{V}. \tag{3.3.6}$$

where $\mathbf{U} \in \mathbb{C}^{m\times m}$, $\mathbf{V} \in \mathbb{C}^{n\times n}$, are orthogonal matrices, and $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$ is a diagonal matrix. The above decomposition is guaranteed to exist and uniquely determines $\mathbf{A}$, if $\mathbf{A}$ is a *compact linear operator*, i.e. $\mathbf{A}$ maps finite subsets of vectors to finite subsets of vectors.[50, 51]

In the SVD, the diagonal elements of the matrix $\mathbf{\Sigma}$, $\sigma_j \geq 0$ for $j \in \{1, ..., \min(n, m)\}$, are one of the key determining factors for numerical stability since Equation (3.3.5) may be rewritten using Equation (3.3.6) as [50]

$$\mathbf{A}^+ = \mathbf{U}\mathbf{\Sigma}^+\mathbf{V}^\dagger. \tag{3.3.7}$$

Now, $\mathbf{\Sigma}^+ = (\mathbf{\Sigma}^T\mathbf{\Sigma})^{-1}\mathbf{\Sigma}^T$. Since, $\mathbf{\Sigma}$ is diagonal $(\mathbf{\Sigma}^T\mathbf{\Sigma})_{ij} = \delta_{ij}\sigma_i^2$, hence, we obtain the following for $\Sigma^+_{ij}$

$$\Sigma^+_{ij} = \frac{\Sigma_{ij}}{\sigma_i^2} \tag{3.3.8}$$

From Equation (3.3.8), it is clear that if the singular values drop to $0$ too rapidly, then the matrix $\mathbf{A}$ is numerically singular, i.e. $\mathbf{A}^+$ becomes infinite, leading to unstable solutions.

We can place an upper-bound on the stability of a given inverse problem by defining the condition number $\kappa(\mathbf{A})$,

$$\kappa(\mathbf{A}) = \frac{\sigma_{\max}}{\sigma_{\min}},$$

where $\sigma_{\max}$ and $\sigma_{\min}$ are the maximum and minimum singular values of $\mathbf{A}$ respectively. The condition number provides an upper bound for the relative error of our model since;

$$\frac{\|\boldsymbol{y} - \mathbf{A}\hat{\boldsymbol{x}}\|}{\|\boldsymbol{y}\|} \leq \frac{\sigma_{\max}\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|}{\sigma_{\min}\|\boldsymbol{x}\|} = \kappa(\mathbf{A})\frac{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|}$$

where $\hat{\boldsymbol{x}}$ are the fitted parameters obtained through inversion. In particular, this allows us to form an approximate relation between the condition number and the extent to which errors are exaggerated, i.e. if $\kappa(\mathbf{A}) = 10^k$ for some $k \in \mathbb{N}$, then we could lose up to $k$

decimal places of precision upon inversion. This is important when designing numerical simulations, as even if a matrix is invertible, it can have a large condition number, which may be on the order of $1/$machine precision. For example in the Julia programming language used for simulations in Chapters 6 and 7 machine precision is on the order of $10^{-16}$ for $64$ bit floating point numbers [52].

We make use of the condition number in Chapter 6 to estimate frequencies at which our inverse problem is ill-posed and also to identify resonances that make the forward problem more difficult to solve. In particular, we identify a relationship between frequency and the well-posedness of our inverse problem; this is expected and widely commented upon in the literature. [49, 53, 54]

### 3.3.3 Tikhonov regularisation

In Section 3.3.1, we derived the Moore-Penrose pseudo-inverse by minimising the cost function given in Equation (3.3.2). However, as noted in Section 3.3.2, solutions obtained using the pseudo-inverse are potentially numerically unstable. The instability arises when the singular values $\sigma_j$ fall to $0$ too rapidly.

One method to remedy this issue is *Tikhonov regularisation*, which works by adjusting Equation (3.3.2) to include the square of the magnitude of the parameters [50]. This leads to the following cost function

$$L_T = ||\boldsymbol{y} - \mathbf{A}\boldsymbol{x}||^2 + \delta||\boldsymbol{x}||^2, \tag{3.3.9}$$

where $\delta$ is a small positive real number. Minimising this cost function leads to the following [50]

$$\hat{\boldsymbol{x}}_T = (\mathbf{A}^\dagger\mathbf{A} + \delta\mathbf{I})^{-1}\mathbf{A}^\dagger\boldsymbol{y} \tag{3.3.10}$$

where $\mathbf{I}$ is the identity matrix.

We note here that, using Equation (3.3.10), we find that in index notation

$$\bar{A}_{ji}y_j = \left(\bar{A}_{jl}A_{il} + \delta\delta_{ij}\right)x_j, \tag{3.3.11}$$

using Equation (3.3.6) in Equation (3.3.11) yields

$$\bar{A}_{ji}y_j = \left(\bar{V}_{jl}\Sigma_{lk}\Sigma_{km}V_{im} + \delta\delta_{ij}\right)x_j, \tag{3.3.12}$$

since $\mathbf{\Sigma}$ is diagonal $\Sigma_{lk}\Sigma_{km} = \delta_{lm}\sigma_l^2$, hence

$$\bar{A}_{ji}y_j = \left(\bar{V}_{jl}V_{im}\delta_{il}\delta_{mj}\sigma_l^2 + \delta\right)\delta_{ij}x_j = \left(\bar{V}_{ji}V_{ij}\sigma_i^2 + \delta\right)\delta_{ij}x_j, \tag{3.3.13}$$

since $\mathbf{V}$ is unitary $\bar{V}_{ji}V_{ij} = 1$, leaving

$$\bar{A}_{ji}y_j = \left(\sigma_i^2 + \delta\right)\delta_{ij}x_j. \tag{3.3.14}$$

Equation (3.3.14) makes clear the role that $\delta$ plays in stabilising our solution. As the singular values tend to $0$, $\sigma_i^2 + \delta \to \delta$, meaning that for an appropriate choice of $\delta$, we know that $(\mathbf{A}^\dagger\mathbf{A} + \delta\mathbf{I})^{-1}\mathbf{A}^\dagger$ does not become infinite. The choice of $\delta$ depends on the level of expected noise in the measured data. In [50], the author describes one method for determining $\delta$ called the *Morozov discrepancy principle*, which states that given some estimate of the noise level in the measured data, $\epsilon$, $\delta$ should be such that

$$\|\mathbf{A}\hat{\boldsymbol{x}}_T(\delta) - \boldsymbol{y}\| = \epsilon.$$

### 3.3.4 Conclusion

In this section, we have dealt with the notion of numerical instability and ill-posedness in a broad sense; however, in most cases, the stability of an inverse problem depends on more complicated factors. [49, 50] The inverse problems we are interested in solving in Chapters 6 and 7 are Cauchy boundary value problems for the Helmholtz equation. This type of inverse problem displays a marked dependence on physical parameters, such as frequency. [49, 53, 54] In the next section, we explore these results in more detail.

## 3.4 The Cauchy problem for the Helmholtz equation

The type of inverse problem that we deal with in the chapters that follow is a Cauchy boundary value problem for the Helmholtz equation. This means that the boundary conditions we impose on our solutions are Cauchy boundary conditions, i.e;

$$\begin{aligned} \nabla^2 u + k^2 u &= 0, \\ u(\boldsymbol{x}) &= f(\boldsymbol{x}) \quad \boldsymbol{x} \in \partial\Omega, \\ \nabla u(\boldsymbol{x}) \cdot \hat{\boldsymbol{n}}(\boldsymbol{x}) &= g(\boldsymbol{x}) \quad \boldsymbol{x} \in \partial\Omega, \end{aligned}$$

where $\partial\Omega$ is the outer boundary of the domain $\Omega$ in which the solution is valid and $\hat{\boldsymbol{n}}(\boldsymbol{x})$ is the unit normal on the boundary $\partial\Omega$ at the point $\boldsymbol{x}$; we have illustrated an example of a domain, $\Omega$, and its boundary, $\partial\Omega$ in Figure 3.1. Essentially, this type of inverse problem aims to obtain a solution in the whole domain $\Omega$ from measurements of part or all of the boundary $\partial\Omega$.

In experiments and in numerical simulations, it is observed that well-posedness of this problem depends on the wave number $k$; in some cases, the stability of the problem decays

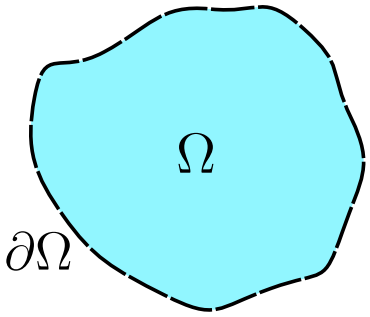



Figure 3.1: An illustration showing an arbitrary domain $\Omega$ and its boundary $\partial\Omega$.

as $k \to \infty$; in others, stability increases with $k$ [49, 53, 54]. In [54], the authors seek to explain why this is the case by obtaining stability estimates for the solution to the above problem. The authors conclude that the stability's relation to $k$ depends upon the *type* of problem one wishes to solve: interior or exterior. The distinction between interior and exterior problems is illustrated in Figure 3.2.

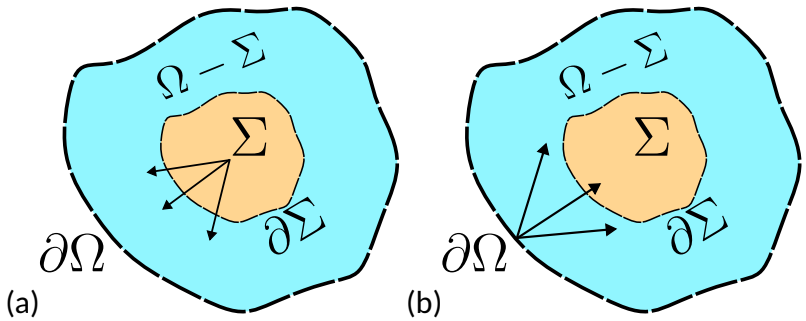



Figure 3.2: (a) shows a typical interior problem; the problem is that of uniquely continuing the solution, represented by the arrows, from the sub-domain $\Sigma$ into the sub-domain $\Omega - \Sigma$. (b) shows a typical exterior problem; the problem is that of uniquely continuing the solution, represented by the arrows, from the sub-domain $\Omega - \Sigma$ into the sub-domain $\Sigma$.

In essence, the distinction between these two problems is the type of information we have obtained; in an interior problem, we have knowledge of the solution in some sub-domain and we aim to extend this solution to predict the solution at points further afield in the domain; in exterior problems, we have far-field information (e.g. from the boundary $\partial\Omega$ in Figure 3.2 or some other subdomain), i.e. we know the solution in some more distant sub-domain and we hope to trace this back to a source. In [54], the authors present theoretical and numerical results on the numerical stability of these two problems. They begin by seeking a stability estimate for solutions to the Cauchy problem for the Helmholtz equation; they find that the stability of the solution depends on two factors the wave number $k$ and the distance $d$ of the subdomain $\Sigma$ to the boundary $\partial\Omega$; of particular interest are the stability estimates obtained for the low and high frequency parts of the solution, let these components be $u_{lf}$ and $u_{hf}$ respectively.

They derive the following stability estimates for $u_{lf}$ and $u_{hf}$

$$\|u_{lf}\| \leq CF,$$
$$\|u_{hf}\| \leq \frac{2C}{k}\left(F(k,d)^{\lambda}\left(\frac{M}{d^2}\right)^{1-\lambda} + F\right),$$

where $\|\cdot\|$ is the $\ell^2$ function norm, $M$, $F$ and $C$ are constants and $F(k,d)$ is a function of $d$ and $k$ specified by the authors.[54] In interior problems, we seek to propagate the solution from $\Sigma$ to $\Omega - \Sigma$, i.e., we want to obtain a solution for increasingly large $d$; whereas in exterior problems, we want to propagate the solution backwards towards $\Sigma$, i.e., we want solutions for increasingly small $d$.

For an interior problem, the authors find that the solution's stability decays as $k$ increases; whereas in the exterior problem, they find the opposite to be the case, i.e., stability increases with $k$. This can be observed in the numerical simulations in [54] and in [53], in relation to electromagnetic scattering. For our purposes, what is relevant is that for exterior problems, stability increases with wavenumber. Our inverse problem, the problem of taking in measured data from the outside of a bearing and attempting to predict the stress inside, is exterior; hence, we should expect to find that its stability increases with $k$; in Section 5.3.2, we find that this is the case.

# Chapter 4

# A Brief Overview of Fourier Analysis

Jessica J. Kent

---

In Chapters 6 and 7, knowledge of the Fourier series, the Fourier transform, and their properties is assumed, as such, we take the time here to give an overview. We begin by defining periodic functions; we then define and discuss the Fourier series and its properties; we then end by discussing the application of the Fourier series to aperiodic functions.

We then move to define the Fourier transform, we discuss the notion of the Fourier transform as a map between dual spaces (e.g. the time domain and the frequency domain) and detail some important properties of the Fourier transform that are assumed in Chapters 6 and 7; most important among these being the way in which the Fourier transform interacts with derivatives and the convolution theorem, both of which play an important role in Chapter 7.

## 4.1 The Fourier series

### 4.1.1 Periodic Functions

The Fourier series is a method for representing *periodic* functions as a linear combination of sinusoids. As such, we begin by defining the notion of a periodic function and discussing some examples.

**Definition:** A function $f(x)$, is said to be periodic, with period $P \in \mathbb{R}$ if the following holds for all $x \in A$, where $A$ is the domain of $f$:

$$f(x) = f(x + P). \tag{4.1.1}$$

The most commonly known examples of periodic functions are sinusoids, e.g., $\sin(x)$ and $\cos(x)$, which are $2\pi$ periodic. A related example, which will be relevant in the rest of this work, is

the complex exponential:

$$f(x) = \mathrm{e}^{\mathrm{i}x}.$$

The periodicity of this function is derived from Euler's formula [55]:

$$\mathrm{e}^{\mathrm{i}x} = \cos(x) + \mathrm{i}\sin(x),$$

since the complex exponential can be represented as a linear combination of $\sin(x)$ and $\cos(x)$, it is also $2\pi$ periodic. With the definition of a periodic function established, we now move on to define the Fourier series representation of a function.

### 4.1.2 Definition of the Fourier series

The Fourier series is a means of representing periodic functions as a linear combination of orthonormal sinusoids. [51] By orthonormal, we mean that, if we define the inner product of two functions, $f(x)$ and $g(x)$ with $x \in S \subset \mathbb{R}$, by the integral [51]

$$\langle f, g\rangle = \int_S \bar{f}(x)g(x)\mathrm{d}x, \tag{4.1.2}$$

where $\bar{f}(x)$ is the complex conjugate of $f$, then $f$ and $g$ are said to be orthonormal if the following holds

$$\frac{\langle f, g\rangle}{\langle f, f\rangle} = \delta_{fg} \tag{4.1.3}$$

where $\delta_{fg}$ is the Kronecker delta.[51]

Throughout this work, we use the *complex* Fourier series to solve boundary value problems, .i.e., we choose complex exponentials of the form $\mathrm{e}^{\frac{\mathrm{i}2\pi nx}{P}}$, where $n \in \mathbb{Z}$ and $P$ is the period, as the basis functions to represent the solutions of the boundary value problems we solve. It can be trivially shown [56] that when using a complex exponential basis as defined above, the orthonormality relation in Equation (4.1.3) reduces to

$$\frac{1}{P}\left\langle \mathrm{e}^{\frac{\mathrm{i}2\pi nx}{P}}, \mathrm{e}^{\frac{\mathrm{i}2\pi mx}{P}}\right\rangle = \delta_{nm}. \tag{4.1.4}$$

For a function $f(x)$, which is periodic over the interval $[-P/2, P/2]$, we define the Fourier series of $f$ over the interval $[-P/2, P/2]$ as [56]:

$$f(x) = \sum_n c_n \mathrm{e}^{\frac{\mathrm{i}2\pi nx}{P}}, \text{ where } c_n = \frac{1}{P}\int_{-P/2}^{P/2} f(x)\mathrm{e}^{-\frac{\mathrm{i}2\pi nx}{P}}\mathrm{d}x. \tag{4.1.5}$$

Note that the expression for $c_n$ found in Equation (4.1.5) follows from Equation (4.1.4), since

$$\frac{1}{P}\left\langle \mathrm{e}^{\frac{\mathrm{i}2\pi nx}{P}}, f\right\rangle = \frac{1}{P}\left\langle \mathrm{e}^{\frac{\mathrm{i}2\pi nx}{P}}, \sum_m c_m \mathrm{e}^{\frac{\mathrm{i}2\pi mx}{P}}\right\rangle,$$

and using the definition of $\langle f, g\rangle$ in Equation (4.1.2) we obtain

$$\frac{1}{P}\int_{-P/2}^{P/2} f(x)\mathrm{e}^{-\frac{\mathrm{i}2\pi nx}{P}}\mathrm{d}x = \frac{1}{P}\int_{-P/2}^{P/2}\sum_m c_m \mathrm{e}^{\frac{\mathrm{i}2\pi mx}{P}}\mathrm{e}^{-\frac{\mathrm{i}2\pi nx}{P}}\mathrm{d}x = \delta_{nm}c_m = c_n,$$

as required.

Of particular interest in this work is the case where the interval chosen is $[-\pi, \pi]$. In this case, Equation (4.1.5) simplifies to

$$f(x) = \sum_n c_n \mathrm{e}^{\mathrm{i}nx}, \text{ where } c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\mathrm{e}^{-\mathrm{i}nx}\mathrm{d}x. \tag{4.1.6}$$

This is because the work in Chapters 6 and 7 models a roller bearing raceway which has circular boundaries, and thus $2\pi$ periodicity.

### 4.1.3 Aperiodic functions?

The Fourier series works on the assumption that the function we are trying to represent is periodic; however, this representation can be extended to the representation of certain classes of functions which are not periodic. According to the *Riemann-Lebesgue lemma* [57], if a function, $f(x)$, is integrable on a circle, i.e. the interval $[-\pi, \pi]$, then the coefficients, $c_n$, of the function's Fourier series representation tend to $0$ as $|n| \to \infty$ and this, taken in conjunction with the assumption of local differentiability at a point $x_0 \in [-\pi, \pi]$, guarantees that the Fourier series representation of $f$ converges pointwise to $f$ in the interval $[-\pi, \pi]$. [57]

Effectively, this means that functions that are integrable and whose end points are equal, i.e. $f(-\pi) = f(\pi)$, can be represented with a Fourier series, without necessarily being periodic; an example used in Chapter 6 is that of a Gaussian centred at $0$, i.e.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-x^2/2\sigma^2}$$

where $\sigma$ is the standard deviation of the Gaussian.

It is important to note that the Fourier series representation we obtain for these aperiodic functions is only valid in the interval $[-\pi, \pi]$. Outside of this interval, we obtain infinite copies of the portion of $f$ in the interval $[-\pi, \pi]$; this can be straightforwardly deduced from the fact that the Fourier series representation of a function is always periodic. Thus, to consider more general types of aperiodic functions, we now consider the Fourier transform.

## 4.2 The Fourier transform

### 4.2.1 Definition of the Fourier transform

The Fourier transform, abstractly speaking, is a means of mapping a function to its equivalent representation in a dual vector space.[51] More concretely, the Fourier transform takes a signal, described either temporally or spatially, and maps it to a function that describes its frequency content.

In this work, two particular cases of the Fourier transform are relevant: the 1D Fourier transform and the 2D Fourier transform. Let $f(t)$ and $g(\boldsymbol{r})$ be complex-valued functions on $\mathbb{R}$ and $\mathbb{R}^2$ respectively; then we define the Fourier transform of $f$ and $g$ with the following integrals [57]:

$$\mathcal{F}\{f\}(\omega) := \tilde{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t, \tag{4.2.1}$$

$$\mathcal{F}_{\text{2D}}\{g\}(\boldsymbol{k}) := \tilde{g}(\boldsymbol{k}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\boldsymbol{r}) \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} \mathrm{d}x \mathrm{d}y. \tag{4.2.2}$$

Further, it can be shown [51, 57] that , given $\tilde{f}$ and $\tilde{g}$, we can recover the functions $f$ and $g$ by using the following integrals:

$$\mathcal{F}^{-1}\{\tilde{f}\}(t) = f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \tilde{f}(\omega) \mathrm{e}^{\mathrm{i}\omega t} \mathrm{d}\omega, \tag{4.2.3}$$

$$\mathcal{F}_{\text{2D}}^{-1}\{\tilde{g}\}(\boldsymbol{r}) = g(\boldsymbol{r}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{g}(\boldsymbol{k}) \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} \mathrm{d}k_x \mathrm{d}k_y. \tag{4.2.4}$$

This is the inverse Fourier transform. It will be used extensively in Chapters 6 and 7 to produce figures in the time domain; this is because the time domain represents how a function evolves in time/space rather than just showing its frequency content.

### 4.2.2 Properties of the Fourier transform

Over the course of this work, several properties of the Fourier transform are used in simplifying expressions and in solving equations. We begin by discussing the effect translation and scaling have on the Fourier transform:

$$\mathcal{F}\{f(t-a)\} = \mathrm{e}^{\mathrm{i}\omega a} \tilde{f}(\omega), \tag{4.2.5}$$

$$\mathcal{F}\{f(\mu t)\} = \frac{1}{\mu} \tilde{f}(\omega/\mu), \text{ for } \mu > 0. \tag{4.2.6}$$

Equation (4.2.5) is readily apparent from Equation (4.2.1) and Equation (4.2.6) can be shown to be true by making the substitution $\tilde{t} = \mu t$ in Equation (4.2.1).

A particularly important property is the way in which the Fourier transform interacts with

derivatives [57]:

$$\mathcal{F}\left\{\frac{\mathrm{d}f}{\mathrm{d}t}\right\} = \mathrm{i}\omega\mathcal{F}\{f\}. \tag{4.2.7}$$

Importantly, if $f$ is differentiable $n$ times and its derivatives are $L^2$ integrable, then

$$\mathcal{F}\left\{\frac{\mathrm{d}^n f}{\mathrm{d}t^n}\right\} = (\mathrm{i}\omega)^n\mathcal{F}\{f\}. \tag{4.2.8}$$

This property also generalises to the $\nabla$ operator. In particular, we obtain the following relations [51]

$$\mathcal{F}_{2\mathrm{D}}\{\nabla g\} = \mathrm{i}\boldsymbol{k}\mathcal{F}_{2\mathrm{D}}\{g\}, \tag{4.2.9}$$

$$\mathcal{F}_{2\mathrm{D}}\{\nabla \cdot \boldsymbol{g}\} = \mathrm{i}\boldsymbol{k} \cdot \mathcal{F}_{2\mathrm{D}}\{\boldsymbol{g}\}, \tag{4.2.10}$$

$$\mathcal{F}_{2\mathrm{D}}\{\nabla \times \boldsymbol{g}\} = \mathrm{i}\boldsymbol{k} \times \mathcal{F}_{2\mathrm{D}}\{\boldsymbol{g}\}, \tag{4.2.11}$$

$$\mathcal{F}_{2\mathrm{D}}\left\{\nabla^2 g\right\} = -|\boldsymbol{k}|^2\mathcal{F}_{2\mathrm{D}}\{g\}. \tag{4.2.12}$$

The relations given by Equations (4.2.8) to (4.2.12) are significant because they in combination with the linearity of the Fourier transform, they allow for differential equations to be converted into algebraic equations, which in many cases are easier to solve. This is especially relevant to Chapter 7, where Equation (4.2.12) is used to produce a theoretical solution to the problem of multiple source localisation in a roller bearing raceway.

Finally, we come to a property of the Fourier transform that is pertinent to Chapter 7; *the convolution theorem*. The convolution of two functions, $f$ and $g$, denoted by $(f \star g)(x)$ is given by the following integral

$$(f \star g)(x) = \int_{-\infty}^{\infty} f(y)g(x-y)\mathrm{d}y. \tag{4.2.13}$$

Convolution interacts with the Fourier transform in the following way [57];

$$\mathcal{F}\{(f \star g)\}(\omega) = \tilde{f}(\omega)\tilde{g}(\omega). \tag{4.2.14}$$

Equation (4.2.14) can be obtained by using Equations (4.2.1) and (4.2.13) in $\mathcal{F}\{(f \star g)\}$ and applying Equation (4.2.5) to deal with the translated argument in Equation (4.2.13).

# Chapter 5

# Methodology

Jessica J. Kent

---

## 5.1 Problem statement

In Chapter 6, we treat the problems of modelling and recovering the traction on the boundaries of a roller bearing raceway due to loading from the shaft and transient and periodic forcings from point defects on the boundaries, like those found in the work of McFadden and Smith in [15, 16]. In Chapter 7, this work is extended to encompass the modelling and localisation of subsurface point defects located at arbitrary locations in the bearing raceway.

### 5.1.1 Governing equations

As noted in Section 2.2, by applying a Helmholtz decomposition, Equations (2.2.3) and (2.2.4), to the Navier equation, Equation (2.1.7), we can obtain the general solution by solving two wave equations for the potentials $\phi(r, \theta)$, the pressure potential, and $\boldsymbol{\psi}(r, \theta)$, the vector shear potential. For our purposes, since we have an axisymmetric problem, we can reduce the vector shear wave potential to a scalar potential by setting $\boldsymbol{\psi}(r, \theta) = \psi(r, \theta)\hat{\boldsymbol{z}}$. From this, we obtain the two scalar wave equations seen in Equations (2.2.5) and (2.2.6).

Applying Equation (4.2.12) to the left-hand side of Equations (2.2.5) and (2.2.6) and rearranging gives two Helmholtz equations,

$$\nabla^2\phi + k_p^2\phi = 0, \tag{5.1.1}$$

$$\nabla^2\psi + k_s^2\psi = 0, \tag{5.1.2}$$

where $k_p = \omega/c_p$ and $k_s = \omega/c_s$ are the pressure and shear wave numbers respectively. Equations (5.1.1) and (5.1.2) can be easily solved using a separation of variables approach as in Section 2.3.

In Chapter 7, we model a point defect at an arbitrary location $\boldsymbol{r}'$ in the bearing raceway.

The primary motivation for this is the problem of detecting and locating subsurface cracks, as when these cracks open up, they release a burst of energy [12], which we model as an idealised point source at $\boldsymbol{r}'$ which initiates instantaneously at time $t = 0$. In the first instance, we consider Equation (5.1.1) forced by a Dirac delta centred at an arbitrary location, $\boldsymbol{r}'$,

$$\nabla^2\phi + k_p^2\phi = \delta(\boldsymbol{r} - \boldsymbol{r}'), \tag{5.1.3}$$

before moving to consider the more general case of Helmholtz equations forced by multiple delta functions at locations $\boldsymbol{r}_j$,

$$\nabla^2\phi + k_p^2\phi = \tilde{\Pi}_p(\boldsymbol{r}), \tag{5.1.4}$$

$$\nabla^2\psi + k_s^2\psi = \tilde{\Pi}_s(\boldsymbol{r}). \tag{5.1.5}$$

We call these more general forcings, $\tilde{\Pi}_p(\boldsymbol{r})$ and $\tilde{\Pi}_s(\boldsymbol{r})$, density maps and they are defined as

$$\tilde{\Pi}_p(\boldsymbol{r}, \omega) = \sum_{j=1}^{J_p} \alpha_j \delta(\boldsymbol{r} - \boldsymbol{r}_j) \mathrm{e}^{-\mathrm{i}\omega t_j} \tag{5.1.6}$$

$$\tilde{\Pi}_s(\boldsymbol{r}, \omega) = \sum_{k=1}^{J_s} \beta_k \delta(\boldsymbol{r} - \boldsymbol{r}_k) \mathrm{e}^{-\mathrm{i}\omega t_k} \tag{5.1.7}$$

where $\alpha_j$ and $\beta_k$ are the amplitudes for P and S sources, respectively, $t_j$ and $t_k$ are the initiation times for the P and S sources, respectively, and $J_p$ and $J_s$ are the number of P and S sources, respectively.

## 5.2 The Forward Problem

In Chapters 6 and 7, we define the *forward problem* as the problem of modelling the elastic waves inside a bearing raceway subject to a known traction or forcing. In both Chapters 6 and 7, we model the bearing raceway as a hollow, linearly elastic cylinder with inner radius $r_1$ and outer radius $r_2$ with infinite extent in the $z$-direction. See Figure 5.1 for a visualisation of the domain. Since the domain is axisymmetric, the problem can be treated as 2D.

In the forward problem, we consider the case where the traction is known on $r = r_1$ and $r = r_2$, $\boldsymbol{\tau}^1$ and $\boldsymbol{\tau}^2$ respectively, where

$$\boldsymbol{\tau}^1 = -\boldsymbol{\sigma}(r_1, \theta) \cdot \hat{\boldsymbol{r}} = -\left[\sigma_{rr}(r_1, \theta)\hat{\boldsymbol{r}} + \sigma_{r\theta}(r_1, \theta)\hat{\boldsymbol{\theta}}\right], \tag{5.2.1}$$

$$\boldsymbol{\tau}^2 = \boldsymbol{\sigma}(r_2, \theta) \cdot \hat{\boldsymbol{r}} = \sigma_{rr}(r_2, \theta)\hat{\boldsymbol{r}} + \sigma_{r\theta}(r_2, \theta)\hat{\boldsymbol{\theta}}, \tag{5.2.2}$$

where $\hat{\boldsymbol{r}}$ is the unit vector in the radial direction, $\hat{\boldsymbol{\theta}}$ is the unit vector in the azimuthal direction

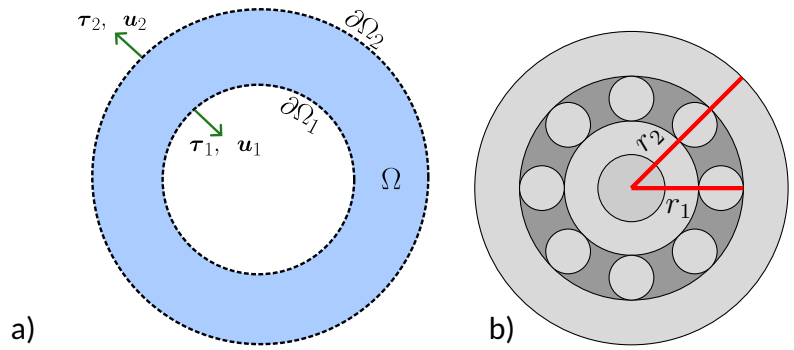



Figure 5.1: Image a) on the left shows the general boundary conditions we consider. That is, we describe elastic waves in the domain $\Omega$, and will use some combination of the boundary data $\boldsymbol{u}_1,\ \boldsymbol{u}_2,\ \boldsymbol{\tau}_1,\ \boldsymbol{\tau}_2$ on the boundaries $\partial\Omega_1$ and $\partial\Omega_2$ which are defined by $r = r_1$ and $r = r_2$ respectively. The image b) on the right illustrates the geometry of one type of roller bearing, where the radii of the outer raceway are shown as an example. The methods of this paper could be used to predict the stresses between either the rollers and the outer raceway or the rollers and the inner raceway.

and $\sigma_{rr}$ and $\sigma_{r\theta}$ are given by

$$\sigma_{rr} = \left(2c_s^2k_p^2 - \omega^2\right)\rho\phi + 2\rho c_s^2\left[\frac{\partial^2\phi}{\partial r^2} + \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial\psi}{\partial\theta}\right)\right], \tag{5.2.3}$$

$$\sigma_{r\theta} = -\rho\omega^2\psi - 2\rho c_s^2\left[\frac{\partial^2\psi}{\partial r^2} - \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial\phi}{\partial\theta}\right)\right]. \tag{5.2.4}$$

Figure 5.1 shows how Equations (5.2.1) and (5.2.2) relate to the geometry of the problem. In the next section, we derive (5.2.3) and (5.2.4).

### 5.2.1 Derivation of traction components

In this section, we show how to derive equations (5.2.3) and (5.2.4). We begin by specifying the form of the stress tensor. In our case, since the material we are considering is homogeneous and isotropic, the desired form is given by Equation (2.1.5). Using Equation (2.1.5), we obtain the following expressions for $\sigma_{rr}$ and $\sigma_{r\theta}$ it terms of the displacement $\boldsymbol{u}$ [42]

$$\sigma_{rr} = \lambda\left(\frac{\partial u_r}{\partial r} + \frac{1}{r}\left(\frac{\partial u_\theta}{\partial\theta} + u_r\right)\right) + 2\mu\frac{\partial u_r}{\partial r}, \quad \sigma_{r\theta} = \frac{\mu}{r}\left(\frac{\partial u_r}{\partial\theta} + r\frac{\partial u_\theta}{\partial r} - u_\theta\right). \tag{5.2.5}$$

We will deduce both equations in turn, beginning with $\sigma_{rr}$. Firstly, note that $(5.2.5)_1$ may be rewritten as

$$\sigma_{rr} = (\lambda + 2\mu)\nabla\cdot\boldsymbol{u} - \frac{2\mu}{r}\left(\frac{\partial u_\theta}{\partial\theta} + u_r\right). \tag{5.2.6}$$

From Equation (2.2.3) we see that $\nabla\cdot\boldsymbol{u} = \nabla^2\phi = -k_p^2\phi$, hence

$$\sigma_{rr} = -k_p^2(\lambda + 2\mu)\phi - \frac{2\mu}{r}\left(\frac{\partial u_\theta}{\partial\theta} + u_r\right). \tag{5.2.7}$$

Using Equation (2.2.3) and Equation (5.1.1) we find that

$$\frac{1}{r}\left(\frac{\partial u_\theta}{\partial \theta}+u_r\right)=-k_p^2\phi-\frac{\partial^2\phi}{\partial r^2}-\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial \psi}{\partial \theta}\right).$$

Substituting the above into Equation (5.2.7) leads to

$$\sigma_{rr}=-\lambda k_p^2\phi+2\mu\left(\frac{\partial^2\phi}{\partial r^2}+\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial \psi}{\partial \theta}\right)\right). \tag{5.2.8}$$

Finally, using Equations (2.2.1) and (2.2.2) we deduce $\lambda=\rho(c_p^2-2c_s^2)$ which together with $\omega=c_pk_p$ substituted into Equation (5.2.8) leads to (5.2.3).

To simplify $\sigma_{r\theta}$ in $(5.2.5)_2$, we rewrite it in the form

$$\sigma_{r\theta}=\mu\left(\nabla\times\boldsymbol{u}\right)\cdot\hat{\mathbf{z}}-\frac{2\mu}{r}\left(u_\theta-\frac{\partial u_r}{\partial \theta}\right).$$

Now, from (2.2.3) and (5.1.2), we have $(\nabla\times\boldsymbol{u})\cdot\hat{\mathbf{z}}=-\nabla^2\psi=k_s^2\psi$, which substituted into the above leads to

$$\sigma_{r\theta}=\mu k_s^2\psi-\frac{2\mu}{r}\left(u_\theta-\frac{\partial u_r}{\partial \theta}\right).$$

Using (2.2.3) and (5.1.2) we find that

$$\frac{1}{r}\left(u_\theta-\frac{\partial u_r}{\partial \theta}\right)=k_s^2\psi+\frac{\partial^2\psi}{\partial r^2}-\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial \phi}{\partial \theta}\right),$$

which substituted into $\sigma_{r\theta}$ above leads to

$$\sigma_{r\theta}=-\mu k_s^2\psi-2\mu\left(\frac{\partial^2\psi}{\partial r^2}-\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial \phi}{\partial \theta}\right)\right),$$

which is the same as (5.2.4) after using Equations (2.2.1) and (2.2.2).

### 5.2.2 Boundary conditions

As noted in Section 2.2, the general solutions of Equations (5.1.1) and (5.1.2) are given by the following Fourier series;

$$\phi=\sum_n\left(a_n\mathrm{J}_n(k_pr)+b_n\mathrm{H}_n^{(1)}(k_pr)\right)\mathrm{e}^{in\theta},$$

$$\psi=\sum_n\left(c_n\mathrm{J}_n(k_sr)+d_n\mathrm{H}_n^{(1)}(k_sr)\right)\mathrm{e}^{in\theta}.$$

Firstly, since the boundary conditions are applied on circular boundaries, the boundary data can be represented by a Fourier series

$$\sum_n \begin{bmatrix} \boldsymbol{\tau}_n^1 \\ \boldsymbol{\tau}_n^2 \end{bmatrix} \mathrm{e}^{\mathrm{i}n\theta} = \begin{bmatrix} \boldsymbol{\tau}^1 \\ \boldsymbol{\tau}^2 \end{bmatrix} \tag{5.2.9}$$

Inspecting the components of the traction given in Section 5.2.1, we note that they depend linearly on the partial derivatives of $\phi$ and $\psi$, hence, using the linearity of derivatives we can conclude that the Fourier coefficients of the boundary data must depend linearly on the coefficients of $\phi$ and $\psi$, $\boldsymbol{a}_n = [a_n, b_n, c_n, d_n]^T$. Thus, we can rewrite Equation (5.2.9) as:

$$\sum_n \mathbf{M}_n^{\text{for}} \boldsymbol{a}_n \mathrm{e}^{\mathrm{i}n\theta} = \begin{bmatrix} \boldsymbol{\tau}^1 \\ \boldsymbol{\tau}^2 \end{bmatrix} \tag{5.2.10}$$

where $\mathbf{M}_n^{\text{for}}$ is the $4 \times 4$ matrix relating $\boldsymbol{a}_n$ to $[\boldsymbol{\tau}_n^1, \boldsymbol{\tau}_n^2]^T$. Using the orthogonality relation, Equation (4.1.4), in Equation (5.2.10) yields

$$\mathbf{M}_n^{\text{for}} \boldsymbol{a}_n = \begin{bmatrix} \boldsymbol{\tau}_n^1 \\ \boldsymbol{\tau}_n^2 \end{bmatrix}.$$

## 5.3 The Inverse Problem

The inverse problem we aim to solve is simple; given a collection of boundary measurements of the displacement field, we wish to find unique solutions $\phi$ and $\psi$ of the elastic wave fields inside the bearing raceway.

### 5.3.1 Boundary conditions

In the inverse problem, we only have access to the data on the outer boundary, $r = r_2$. This leads to a set of mixed boundary conditions, one boundary condition for displacement and another for traction;

$$\boldsymbol{u}^2 = \boldsymbol{u}_{\text{data}}, \tag{5.3.1}$$

$$\boldsymbol{\tau}^2 = \boldsymbol{\sigma} \cdot \hat{\boldsymbol{r}}. \tag{5.3.2}$$

Figure 5.1 shows how Equations (5.3.1) and (5.3.2) relate to the geometry of the problem. This leads to a system of four equations. As before, we can express this system of equations as a Fourier series;

$$\sum_n \boldsymbol{u}_n \mathrm{e}^{\mathrm{i}n\theta} = \begin{bmatrix} \boldsymbol{u}^2 \\ \boldsymbol{\tau}^2. \end{bmatrix} \tag{5.3.3}$$

As before, we know that $\boldsymbol{u}_n$ depends linearly on the coefficients of the Helmholtz potentials, $\boldsymbol{a}_n$, so

$$\sum_n \mathbf{M}_n^{\text{inv}} \boldsymbol{a}_n \mathrm{e}^{\mathrm{i}n\theta} = \begin{bmatrix} \boldsymbol{u}^2 \\ \boldsymbol{\tau}^2 \end{bmatrix}, \tag{5.3.4}$$

where $\mathbf{M}_n^{\text{inv}}$ is the $4 \times 4$ matrix relating $\boldsymbol{u}_n$ to $\boldsymbol{a}_n$. Using the orthogonality relation, Equation (4.1.4), in Equation (5.3.4) yields

$$\mathbf{M}_n^{\text{inv}} \boldsymbol{a}_n = \begin{bmatrix} \boldsymbol{u}_n^2 \\ \boldsymbol{\tau}_n^2 \end{bmatrix}, \tag{5.3.5}$$

where $\boldsymbol{u}_n^2$ and $\boldsymbol{\tau}_n^2$ are the $n$th Fourier coefficients for $\boldsymbol{u}^2$ and $\boldsymbol{\tau}^2$ respectively.

Equation (5.3.5) holds for each $n \in \mathbb{Z}$, in reality, we can only choose a finite number of these to use in simulations and with real data, i.e., we approximate the solutions as

$$\phi = \sum_{n=-N}^{N} \left( a_n \mathrm{J}_n(k_p r) + b_n \mathrm{H}_n^{(1)}(k_p r) \right) \mathrm{e}^{\mathrm{i}n\theta},$$

$$\psi = \sum_{n=-N}^{N} \left( c_n \mathrm{J}_n(k_s r) + d_n \mathrm{H}_n^{(1)}(k_s r) \right) \mathrm{e}^{\mathrm{i}n\theta},$$

where $N$ is a non-negative integer.

This means that, for a given $N$, we seek $2N + 1$ vectors of coefficients, $\boldsymbol{a}_n$. Thus, to ensure that the problem is well-posed, we must have $2N+1$ measurements. In the examples presented in Sections 6.7 and 7.3 we take the simulated measurements $\boldsymbol{u}_{\text{data}}(\theta)$ from $2N + 1$ evenly spaced angular locations along the outer boundary, i.e. that the $j$th measurement, $\boldsymbol{u}_{\text{data}}(\theta_j)$ is taken from $\theta_j = (j-1)\pi/(2N+1)$. Further to this, we also assume that the 'sensors' making these measurements have size $0$, i.e., are ideal points. This raises a further question on how numerical stability is affected by the frequency $\omega$ and the mode $n$; this is explored in the next section.

### 5.3.2 The Diffraction Limit

One cause of the modal system becoming ill-posed, as shown in Section 6.6, is due to the diffraction limit [58]. Below, we provide an approximate formula based on a classical argument of the diffraction limit to easily determine when the system is ill-posed.

Instead of solving a boundary value problem, we consider a simpler case of determining the amplitudes $A$ and $B$ of two point sources on the boundary $r = r_1$. These, in a sense, imitate forcing on the boundary. The field emitted by these sources is given by

$$\phi_{\text{src}} = \frac{\mathrm{i}}{4} \left( A \mathrm{H}_0^{(1)}(k_p |\boldsymbol{r} - \boldsymbol{r}^{(1)}|) + B \mathrm{H}_0^{(1)}(k_p |\boldsymbol{r} - \boldsymbol{r}^{(2)}|) \right), \tag{5.3.6}$$

with $\boldsymbol{r}^{(1)}$ and $\boldsymbol{r}^{(2)}$ representing the position of the first and second point source respectively.

Now we ask, can we tell the difference between these two sources by measuring the field on the boundary $r = r_2$? That is, can we distinguish between sources at $\boldsymbol{r}^{(1)}$ and $\boldsymbol{r}^{(2)}$? As we want to determine the maximum amount of information available, we consider that we have access to the field everywhere on the boundary $r = r_2$. To achieve this, it is convenient to use the origin $\boldsymbol{O}_2$, the midpoint of the chord connecting the two sources. Then we can use Graff's addition theorem, Equation (2.3.2), to rewrite (5.3.6) as

$$\phi_{\text{src}} = \frac{\mathrm{i}}{4} \sum_n \left( A \mathrm{J}_{-n}(k_p r_0) \mathrm{H}_n^{(1)}(k_p r) + B \mathrm{J}_n(k_p r_0) \mathrm{H}_n^{(1)}(k_p r) \right) \mathrm{e}^{\mathrm{i}n\theta}, \tag{5.3.7}$$

where $r_0$ is the horizontal distance from $\boldsymbol{O}_2$ to each source, $r$ is the distance from $\boldsymbol{O}_2$ to some observation point on the outer boundary, and $\theta$ is the angle of the observation point from the source at $\boldsymbol{r}^{(1)}$, this is illustrated in Figure 5.2.

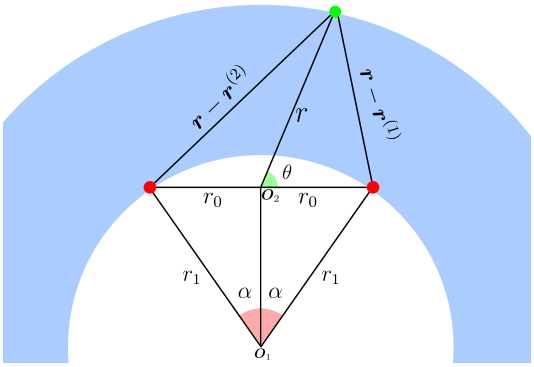



Figure 5.2: Illustration of Graff's addition theorem. $r_0$ is the distance from $\boldsymbol{O}_2$ to each source, $r$ is the distance from $\boldsymbol{O}_2$ to some observation point on the outer boundary, and $\theta$ is the angle of the observation point from the source at $\boldsymbol{r}^{(1)}$. The angles $\alpha$ are needed to relate these sources to Fourier modes later.

To reach a simple approximate formula, we consider the limit when the sources are close, $k_p r_0 \to 0$, and also evaluate (5.3.7) in the far field, $k_p r \to \infty$. The information count should not change significantly in the far-field, but it does simplify the field. Taking these asymptotic limits and retaining up too $\mathcal{O}(k_p^2 r_0^2)$, while keeping only the leading for $k_p r \to \infty$, leads to:

$$\phi_{\text{src}} = \frac{1+\mathrm{i}}{8\sqrt{\pi k_p r}} \mathrm{e}^{\mathrm{i}k_p r} \left[ 2(A+B) + 2\mathrm{i}(A-B) k_p r_0 \cos\theta - (A+B)(k_p r_0 \cos\theta)^2 \right]. \tag{5.3.8}$$

Now, to distinguish between the sources, we need to determine $r_0$ by measuring the above. We can further simplify this by specialising to the case where the boundary data is smooth, which will lead to a lower bound on what can be measured. In this case, we consider that $B$

is a smooth function of $r_0$ such that asymptotically:

$$B = A + \frac{\beta}{2} r_0^2,$$

which substituted into (5.3.8) leads to:

$$\phi_{\text{src}} = \frac{1+\mathrm{i}}{8\sqrt{\pi k_p r}} \mathrm{e}^{\mathrm{i}k_p r} \left[ 4A + r_0^2 \beta - 2A(r_0 k_p \cos\theta)^2 \right]. \tag{5.3.9}$$

The $\beta$ term can be anything, but will change from a negative to a positive value as the $\boldsymbol{r}^{(2)}$ changes. When $\beta$ has the same sign as $A$, it will make it easier to determine $r_0$, and when the signs are opposite, it will make it harder to determine $r_0$. To reach a simple approximation, we take $\beta = 0$.

Then, to resolve the difference between the sources at $\boldsymbol{r}^{(1)}$ and $\boldsymbol{r}^{(2)}$, we need to easily measure $r_0$, which implies that the quadratic term $r_0^2$ in (5.3.9) needs to be greater than or equal to the leading term, that is

$$|k_p r_0 \cos\theta|^2 \geq 2 \tag{5.3.10}$$

which is guaranteed to hold if $|k_p r_0| \geq \sqrt{2}$.

The main method of the paper uses a modal Fourier decomposition (6.2.4) to solve for elastic waves. For these, the level of detail of the boundary data increases with the Fourier order $n$. The minimal level of detail captured by order $n$ is the distance on the boundary between a trough and a crest, which is equal to $\theta = r_1|n|/\pi$. From Figure 5.2 we see that this leads to the choice $2\alpha = \pi/|n|$ and

$$r_0 = r_1 \sin\left(\frac{\pi}{2|n|}\right) \approx \frac{\pi r_1}{2|n|}, \tag{5.3.11}$$

where the approximation is accurate for $|n| > 1$. Substituting the above into $|k_p r_0| \geq \sqrt{2}$ yields the following useful result

$$\pi k_p r_1 \geq \sqrt{8}|n|. \tag{5.3.12}$$

Figure 5.3 shows when equality holds in the above, superimposed on the condition number plot in Figure 6.5. It is clear that the limit loosely indicates where we begin to lose precision due to an ill-conditioned system. In reality, the true bound depends on more factors than those found in Equation (5.3.12).

### 5.3.3 Non-dimensional parameters

Since the physical quantities used in the boundary conditions in Equation (5.3.4) have different units, it is important that we non-dimensionalise to scale out the units in the problem. Due to the fact that in steel $\lambda$ and $\mu$ are approximately $120$ GPa and $80$ GPa, respectively [59],

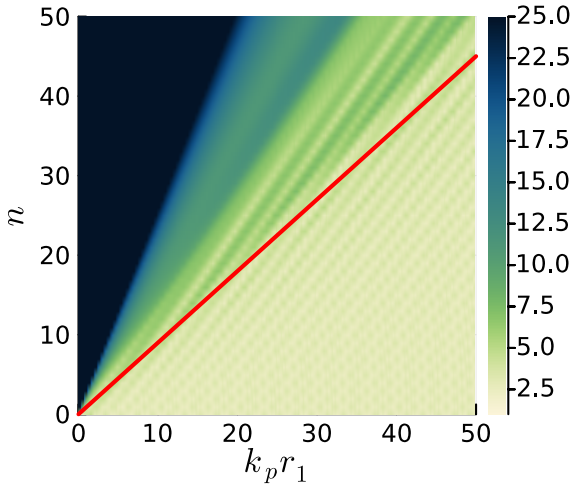



Figure 5.3: Figure shows the condition number plot from Figure 6.5 with the diffraction limit in Equation (5.3.12) plotted on top in red. The diffraction limit gives a rough idea of when the inverse problem is well-posed; though it is only a loose idea as we have not taken account of multiple scattering events and have performed the calculation in free space.

meaning that displacements are many orders of magnitude smaller than the traction [42]. This leads to the condition number of the dimensional $\mathbf{M}_n^{\text{inv}}$ becoming large, which in turn leads to numerical instability. Non-dimensionalising the problem, by removing the units of the problem and thus making all quantities comparable, remedies this issue [60].

To non-dimensionalise our system, we choose

$$\bar{r} = k_p r, \quad \bar{\phi} = k_p^2 \phi, \quad \bar{\psi} = k_p^2 \psi, \tag{5.3.13}$$

which leads to the non-dimensional Helmholtz equations:

$$\bar{\nabla}^2 \bar{\phi} + \bar{\phi} = 0, \quad \bar{\nabla}^2 \bar{\psi} + \bar{k}_s^2 \bar{\psi} = 0,$$

where $\bar{k}_s^2 = k_s^2/k_p^2$, and $\bar{\nabla}$ differentiates in terms of $\bar{\boldsymbol{x}} = k_p \boldsymbol{x}$. Substituting these non-dimensional potentials into the displacement and traction leads to

$$k_p \boldsymbol{u} = \bar{\nabla}\bar{\phi} + \bar{\nabla} \times (\bar{\psi}\hat{\boldsymbol{z}}) \quad \text{and} \quad \boldsymbol{\sigma} = \lambda \text{tr}(\boldsymbol{\varepsilon})\mathbf{I} + 2\mu\boldsymbol{\varepsilon},$$

which together with $\bar{k}_p = 1$ leads us to the natural choices

$$\bar{\boldsymbol{u}} = \boldsymbol{u} k_p \quad \text{and} \quad \bar{\boldsymbol{\sigma}} = \boldsymbol{\sigma}/(\lambda + 2\mu), \tag{5.3.14}$$

where $\bar{\lambda} = \lambda/(\lambda + 2\mu)$ and $\bar{\mu} = \mu/(\lambda + 2\mu)$.

Below, we summarise which material properties would lead to the above equations:

$$\bar{c}_p = \omega, \quad \bar{c}_s = \frac{\omega c_s}{c_p}, \quad \bar{\rho} = \frac{1}{\omega^2}, \quad \bar{r} = r k_p.$$

The dimensional versions can then be recovered from the solutions $\bar{\phi}$ and $\bar{\psi}$ by using (5.3.14) and (5.3.13). With this established, we now move to consider the definition and selection of priors.

### 5.3.4 Priors

In Section 6.4, we introduce the notion of a prior. In the context of the work, 'prior' refers to a parameterisation of the modal coefficients, $\boldsymbol{a}_n$, in terms of known information; this could be pre-established information about defects in the raceway or assumptions about the types of defects that are present in the data. The aim of such a parameterisation is the reduction of the dimension of the solution space, i.e., to recover more accurate solutions using as few measurements as possible.

In this work, we use a linear parameterisation to achieve this, i.e., we let

$$\boldsymbol{a}_n = \mathbf{B}_n \boldsymbol{x} + \boldsymbol{c}_n \tag{5.3.15}$$

where $\boldsymbol{x}$ are the parameters $\boldsymbol{a}_n$ depends on. Prior information is included via the matrix $\mathbf{B}_n$ and the bias vector $\boldsymbol{c}_n$. This parameterisation of $\boldsymbol{a}_n$ is found by solving the forward problem for our assumed prior information, i.e., suppose that we know that the traction of boundary $j = 1$ or $2$ depends linearly on parameters $\boldsymbol{x}^j$ and known traction $\boldsymbol{\tau}_0^j = \sum_n \boldsymbol{\tau}_n^j \mathrm{e}^{\mathrm{i}n\theta}$, then the boundary data, $\boldsymbol{\tau}^j$ in the forward problem is given by

$$\boldsymbol{\tau}^j = \sum_n \left( \mathbf{F}_n^j \boldsymbol{x}^j + \boldsymbol{\tau}_n^j \right) \mathrm{e}^{\mathrm{i}n\theta},$$

where the $\mathbf{F}_n^j$ are matrices relating the parameters $\boldsymbol{x}^j$ to the boundary data for mode $n$. From this, we form the following system of equations

$$\mathbf{M}_n^{\mathrm{for}} \boldsymbol{a}_n = \mathbf{F}_n \boldsymbol{x} + \boldsymbol{\tau}_n \tag{5.3.16}$$

where $\mathbf{F}_n = \begin{bmatrix} \mathbf{F}_n^1 & 0 \\ 0 & \mathbf{F}_n^2 \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}^1 \\ \boldsymbol{x}^2 \end{bmatrix}$ and $\boldsymbol{\tau}_n = \begin{bmatrix} \boldsymbol{\tau}_n^1 \\ \boldsymbol{\tau}_n^2 \end{bmatrix}$. Solving Equation (5.3.16) for each mode $n$ gives Equation (5.3.15) by setting $\mathbf{B}_n = \left(\mathbf{M}_n^{\mathrm{for}}\right)^{-1} \mathbf{F}_n$ and $\boldsymbol{c}_n = \left(\mathbf{M}_n^{\mathrm{for}}\right)^{-1} \boldsymbol{\tau}_n$.

In Chapter 6, this parameterisation scheme is mainly used to incorporate information about the bearing's specifications to recover the loading distribution; in particular, the work in Section 6.5.1 allows us to incorporate information about the number or rollers and their rotation speed, this in turn allows us to access lower frequency modes to fit a loading

distribution to the simulated data, see Section 6.7.3. In Chapter 7, the addition of a point source into the domain leads to the inverse problem becoming ill-posed; this is due to the addition of another set of Fourier coefficients, $\boldsymbol{a}_n^{\mathrm{src}}$, corresponding to the point source. To remedy this problem, we parametrise the $\boldsymbol{a}_n$ in terms of $\boldsymbol{a}_n^{\mathrm{src}}$, this reformulates the inverse problem in terms of a matrix that can be inverted to recover $\boldsymbol{a}_n^{\mathrm{src}}$, details can be found in Section 7.2.2.

# Chapter 6

# Elastic waves in bearing raceways: the forward and inverse problem

Jessica J. Kent, Matheus de C. Loures, Art L. Gower

## Abstract

Turbines are crucial to our energy infrastructure, and ensuring their bearings function with minimal friction while often supporting heavy loads is vital. Vibrations within a bearing can signal the presence of defects, friction, or misalignment. However, current detection methods are neither robust nor easy to automate. We propose a more quantitative approach by modelling the elastic waves within bearing raceways. By approximating the raceway as a hollow cylinder, we derive straightforward 4x4 systems for its vibrational modes, enabling both forward and inverse problem-solving. We also demonstrate how to significantly reduce the number of required sensors by using a simple prior: the known number of rollers and their angular speed. We present numerical examples showcasing the full recovery of contact traction between bearings and the raceway, as well as the detection of elastic emissions.

## 6.1 Introduction

Bearings are essential parts of modern industrial machinery; found everywhere from bicycles to wind turbines to jet engines [1]. Their main purpose is to reduce friction and constrain the motion of rotating components; as such, their maintenance and efficiency remains an important industrial problem and an active field of study [3].

**Current methods.** The most successful methods to monitor the condition of roller bearings are based on vibration analysis: analysing the frequency components of how the raceway, or mounting, vibrates in time [14, 61, 62]. Current methods perform best for identifying localised defects that create an impulsive response; given the right conditions these methods

can identify whether there is a localised defect on the inner or outer race, or on the rollers themselves. However, it has been challenging to make current methods robust and reliable enough to automate, and therefore experts are often required to both process and interpret the results.

**Too few sensors.** It is likely that robust diagnostics are not possible for a small number of sensors. Typically, sensors measure displacement or acceleration at one or two positions, and rotation speed on each large bearing [23]. That is, there are often too many unknowns for the number of sensors typically used. Further, acceleration sensors are often placed on the housing, in which case the *transfer path* of the signal can be unknown and complex. We call the *transfer path* the route an elastic wave takes from its source to a sensor.

**Be more quantitative.** In this work we suggest that to have more robust and automatic predictions we need to properly model the physics of elastic waves in the bearing raceway. After all, these elastic waves are what carry the information of the forces inside the bearing to the sensors. Modelling elastic waves enable us to: 1) determine how well we can truly expect a few sensors to perform, and 2) develop more quantitative methods to predict defects and contact forces in bearings.

A typical approach to account for the physics of elastic waves is to use finite element methods, however these are far too computationally intensive and opaque for inverse problems [39]. In this paper we show how elastic waves can be very efficiently described both for sensing and modelling, with details in the next section.

**Nonlinear dynamics at interfaces.** For roller and journal bearings the way that forces are transmitted through the bearing and the shaft at interfaces are non-linear [63, 64, 65]. For example, a roller rattling is a non-linear dynamic event. Non-linearities at the interfaces make it challenging to accurately solve for the dynamics of an entire system, which typically includes components like shafts, rotors, bearings, and varying foundations [66]. Over the past 40 years there has been significant work to model the entire system [15, 16, 23, 28, 67].

**Linear elastic waves.** While the interface conditions between machine components are often non-linear, the elastic waves within each component are primarily governed by linear elasticity. This allows us to break the problem into manageable parts: by measuring the vibrations at one boundary, we can predict the forces or vibrations at another boundary of the same component.

We demonstrate that it is possible to predict the stresses on a bearing raceway. See Figure 6.1 for a motivational example which predicts the stresses between the rollers and the raceway with just four sensors. There is currently no such method to predict these forces which are important for lifetime analysis and to understand the causes of defects [68].

**Acoustic emission and localisation.** One method to detect defects is to measure the sound they emit [69, 70, 71]. These methods rely on measuring only pressure waves in solid components and are almost exclusively based in the time domain, where the first signal that arrives is (likely) the bulk pressure mode. It is important to only use the first measured signal,

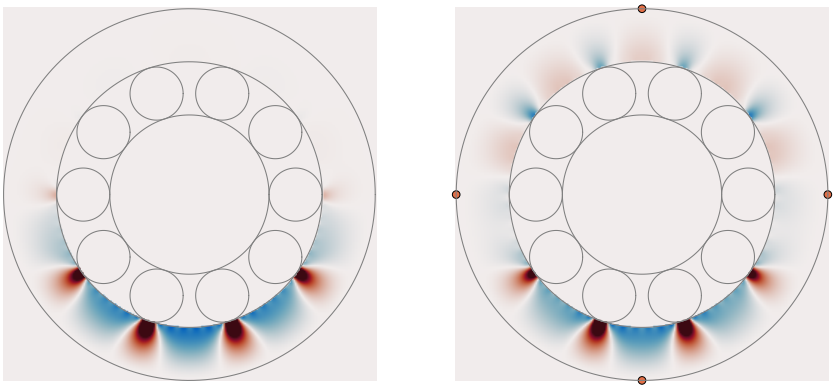


Figure 6.1: True pressure on the left for one snapshot in time. The right shows the predicted stress for the same time when using 4 sensors shown as orange spots. The sensors measure displacement, and the outer boundary is stress free, which is why the stress tends to zero when it reaches the outer boundary. However, the displacements or accelerations caused by the stress are not zero on this boundary and can be measured. The recovery is not perfect as there is a 20% error added to the boundary data. More details given in Section 6.7.3.

as the next signals will be a mix of pressure and shear waves that have likely mode converted at boundaries, and, especially for thin structures, potentially formed waveguide modes. To extract the first signal can be difficult when its amplitude is far less than the other wave modes, and, of course, its arrival time is unknown. [72]

Alternatively, instead of assuming that there are only acoustic waves, one could just measure the signal in time, potentially both pressure and shear displacement, and from that determine where the source is. To do so would require modelling elastic waves. This paper is the first step to an elastic emission method in bearing raceways that does not require extracting the first arrival time, and could continually measure and identify sources.

**Paper contents.** In this paper we show theoretical and numerical results on how to predict the stress on the raceway and between rollers and the raceway. The methods developed can be specialised to other bearings, though we focus on roller bearings here.

In Section 6.2 we discuss how, at high enough frequencies, elastic waves are mostly confined to the bearing raceway; this allows us to model the waves in a raceway as vibrations of a hollow cylinder, as opposed to modelling the full bearing system. In Section 6.3 we develop a modal method to quickly calculate elastic waves in the raceway from boundary conditions.

In Section 6.4 we show how assumptions about the boundary conditions, which we call priors, can greatly reduce the number of sensors required. As an example, in Section 6.5 we show how to deduce and use priors about rollers rotating at a constant speed.

From the elastic wave models we learn what is, and is not, measurable, which we summarise in Section 6.6. In Section 6.7 we show several numerical examples both for validation and to illustrate the main results.

## 6.2 Elastic waves in raceways

Bearings are mounted in many different ways, one example is shown in Figure 6.2. However, the raceway is usually a hollow cylinder and is fabricated as one solid piece, see Figure 6.3 for some examples. The raceway is then tightly fitted into the mounting, or if there is also an inner raceway then it is fitted over a shaft.

When the rollers press on the track, they emit elastic waves, which for high enough frequencies (> 10kHzs) are mostly trapped within the track [13] due to the air gap that remains between the raceway and the mounting (or shaft). However, for convenience, sensors are usually placed on the bearing mounting, rather than the raceway itself, which does not always get a good signal for bearing defects, or other features. In many cases, the waves originally emitted into the raceway can take a long (transmission) path until reaching a sensor on the mounting. During this journey, the wave is highly distorted; this can make it difficult to recognize defects signals [73]. There are methods which attempt to undo the effects of this path [14] for impulsive signals, e.g. minimum entropy deconvolution [73]. However, these can not be generalised to non impulsive signals; can enhance noise that is impulsive and have some difficulties in parameter choice such as window length [18].

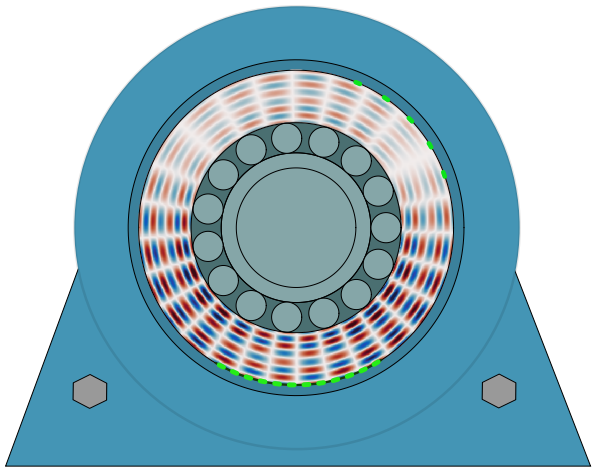


Figure 6.2: A cross section illustration of a bearing in a mounting. Showing a wave for a fixed frequency (> 10kHzs) showing waves trapped in the raceway. Further shown are sensors mounted on the raceway as we propose.

What if we could mount sensors on, or near, the raceway as shown in Figure 6.2? In this paper we show that these sensors would lead to many benefits such as: a direct prediction of the stresses in bearings, and clear signals on extended or localised defects. We can also answer exactly what is possible to predict, what frequencies to use, and where to place the sensors just by modelling the raceway as a hollow cylinder, as we do in the next section.



Figure 6.3: The left shows a Schaeffler roller bearing for the main shaft of a wind turbine [74] while the right shows an example of a Miba tilting pad journal bearing used in turbines[75]. In both examples sensors could be placed on the outside of the casing before putting the bearing in its mounting.

### 6.2.1 Modal solution

Below we show the simplest way to calculate elastic waves in the raceway, see Figure 5.1 for an illustration of the boundaries and domain where we calculate elastic waves.

Steel is well approximated as an isotropic material. Further, as even very high stresses of 1000 MPa only change elastic wave speeds by a few percent [76], the elastic waves within the raceway are well approximated by the linearised equations of elasticity in a homogeneous and isotropic solid [77]. We also assume that stresses applied to the raceway boundaries are approximately axially symmetric, at least after averaging over some time period, which implies that the elastic waves are axially symmetric.

The above allows us to write the small elastic displacement, for a harmonic angular frequency of $\omega$ in terms of the Helmholtz potentials in the form

$$\boldsymbol{u} = \nabla\phi + \nabla \times (\psi\hat{\boldsymbol{z}}) \tag{6.2.1}$$

where $\phi$ and $\psi$ are the pressure and shear potentials respectively, and noting that the vector shear potential automatically satisfies the divergence free condition $\nabla \cdot (\psi\hat{\boldsymbol{z}}) = 0$ when $\psi$ does not depend on $z$. The displacement in time can be calculated by taking an inverse Fourier transform, which is the same as integrating $\boldsymbol{u}\mathrm{e}^{-\mathrm{i}\omega t}$ over $\omega$ in the convention used here.

The advantage of using the Helmholtz decomposition (6.2.1) is that both potentials satisfy a Helmholtz equation:

$$\nabla^2\phi + k_p^2\phi = 0, \quad \nabla^2\psi + k_s^2\psi = 0, \tag{6.2.2}$$

where $k_p = \frac{\omega}{c_p}$ and $k_s = \frac{\omega}{c_s}$ are the wavenumbers of the P and S-waves, while $c_p$ and $c_s$ are the wave speeds which are related to the Lamé parameters $\lambda$ and $\mu$ through

$$\rho c_p^2 = \lambda + 2\mu \quad \text{and} \quad \rho c_s^2 = \mu, \tag{6.2.3}$$

where $\rho$ is the mass density.

As we consider the raceway to be a thick-walled circular cylinder, we can reach simple solutions by using cylindrical coordinates $(r, \theta)$, which leads to solutions of (6.2.2) in the form

$$\begin{aligned} \phi(r,\theta) &= \sum_{n=-\infty}^{\infty} \left(a_n \mathrm{J}_n(k_p r) + b_n \mathrm{H}_n^{(1)}(k_p r)\right) \mathrm{e}^{\mathrm{i}n\theta}, \\ \psi(r,\theta) &= \sum_{n=-\infty}^{\infty} \left(c_n \mathrm{J}_n(k_s r) + d_n \mathrm{H}_n^{(1)}(k_s r)\right) \mathrm{e}^{\mathrm{i}n\theta}, \end{aligned} \tag{6.2.4}$$

where $\mathrm{J}_n$ and $\mathrm{H}_n^{(1)}$ are Bessel and Hankel functions of the first kind respectively. To deduce the above you can use separation of variables [43, 46], and the coefficients $a_n, b_n, c_n, d_n$ can be determined from boundary conditions. Note that if the cylinder had no hole, then $b_n = d_n = 0$ and there would also be one less boundary to prescribe boundary conditions.

To prescribe boundary conditions we need the traction on the boundary in polar coordinates. In general the Cauchy stress tensor is given by

$$\boldsymbol{\sigma} = \lambda \mathrm{tr}(\boldsymbol{\varepsilon})\mathbf{I} + 2\mu\boldsymbol{\varepsilon} \tag{6.2.5}$$

where $\boldsymbol{\varepsilon} = \frac{1}{2}\left(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^{\mathrm{T}}\right)$. The traction $\boldsymbol{\tau}$ on the outer boundary of a cylinder is given by

$$\boldsymbol{\tau} = \boldsymbol{\sigma} \cdot \hat{\boldsymbol{r}} = \sigma_{rr}\hat{\boldsymbol{r}} + \sigma_{r\theta}\hat{\boldsymbol{\theta}}, \tag{6.2.6}$$

where $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{\theta}}$ are unit vectors along the directions that the radius and polar angle increase. The traction on the inner boundary is given by $\boldsymbol{\tau} = -\boldsymbol{\sigma} \cdot \hat{\boldsymbol{r}}$ as the outward normal vector in this case is $-\hat{\boldsymbol{r}}$. See [78] for more details on stress tensors in polar coordinates.

Substituting (6.2.1) into (6.2.5) leads to

$$\sigma_{rr} = \left(2c_s^2 k_p^2 - \omega^2\right)\rho\phi + 2\rho c_s^2 \left[\frac{\partial^2 \phi}{\partial r^2} + \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial \psi}{\partial \theta}\right)\right], \tag{6.2.7}$$

$$\sigma_{r\theta} = -\rho\omega^2\psi - 2\rho c_s^2 \left[\frac{\partial^2 \psi}{\partial r^2} - \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial \phi}{\partial \theta}\right)\right]. \tag{6.2.8}$$

See Section 5.2.1 for derivation.

## 6.3 Boundary conditions

How much boundary data is needed to determine the potentials in Equation (6.2.4)? To answer this question, consider the simpler case of how much data is needed to determine just the coefficients $a_n$ in a series $f(\theta) = \sum_n a_n \mathrm{e}^{\mathrm{i}n\theta}$; representing any square integrable periodic function $f(\theta)$ (almost everywhere). To determine $a_n$ we need to supply a function $f(\theta)$. Therefore, to determine $a_n$, $b_n$, $c_n$, and $d_n$ we must supply four functions.

For example, let us turn to Figure 5.1 and consider the outer raceway. To determine all the coefficients, it would be sufficient to have the boundary data of the displacement $\boldsymbol{u}$ and traction $\boldsymbol{\tau}$ on just one boundary, say at $r = r_1$, because

$$\boldsymbol{u}(r_1, \theta) = u_r(r_1, \theta)\hat{\boldsymbol{r}} + u_\theta(r_1, \theta)\hat{\boldsymbol{\theta}},$$

which is composed of two scalar functions in $\theta$, and likewise for $\boldsymbol{\tau}$. If the source of the elastic waves was also on the boundary $r = r_1$, then this would be an initial value problem, and we expect it then to be well-posed, and therefore this boundary data would completely determine the coefficients.

Below we consider different combinations of boundary data and show how to determine the coefficients $a_n$, $b_n$, $c_n$, and $d_n$. In this work we do not discuss uniqueness of the solution, and simply verify that the solution is indeed well-posed where we expect it to be. That is, we expect the solution to become ill posed for low enough frequencies, and to be ill posed when approaching the diffraction limit.

We call the *forward problem* the case where the traction $\boldsymbol{\tau}$ on the inside and outside of the raceway are given. The name is just for convenience, and because knowing the traction on both boundaries often implies we know the source of the waves. However, this is still a boundary value, and could equally be considered to be an inverse problem. We consider this case first and then turn to more general boundary conditions, such as the case where the displacement $\boldsymbol{u}$ and traction $\boldsymbol{\tau}$ on the outer boundary are known, which we call the *inverse problem*. Finally, we look at some example with stresses inspired by an operating bearing.

The inverse problem is of more practical importance, as it is possible to place sensors on the outside of the raceway for the case shown in Figure 5.1. In Section 6.3.3 we discuss the details on taking boundary data from a, possibly small, finite number of measurements at specific points on the boundary.

### 6.3.1 Traction boundary conditions: the forward problem

Here we consider prescribing only traction boundary conditions on both the boundaries of the raceway, see Figure 5.1 for an illustration.

Let the traction on the boundary $r = r_1$ be $\boldsymbol{\tau}^1(\theta)$, and the traction on the boundary $r = r_2$ be $\boldsymbol{\tau}^2(\theta)$, be given by;

$$\boldsymbol{\tau}^1(\theta) = -p^1(\theta)\hat{\boldsymbol{r}} - s^1(\theta)\hat{\boldsymbol{\theta}}, \quad \boldsymbol{\tau}^2(\theta) = p^2(\theta)\hat{\boldsymbol{r}} + s^2(\theta)\hat{\boldsymbol{\theta}}, \tag{6.3.1}$$

where the negative sign in the top equation is due to the unit normal being $-\hat{\boldsymbol{r}}$.

To solve the problem now we need the Fourier series representation:

$$p^j(\theta) = \sum_{n=-\infty}^{\infty} p_n^j \mathrm{e}^{\mathrm{i}n\theta}, \quad s^j(\theta) = \sum_{n=-\infty}^{\infty} s_n^j \mathrm{e}^{\mathrm{i}n\theta}, \quad \text{for } \; j = 1, 2, \tag{6.3.2}$$

then substitute the potentials (6.2.4) into (6.2.7) and (6.2.8) and then substitute the result into (6.3.1). Using that the $\mathrm{e}^{\mathrm{i}n\theta}$ in the Fourier series are orthonormal leads to the matrix equation

$$\mathbf{M}_n^{\text{for}} \boldsymbol{a}_n = \boldsymbol{f}_n, \tag{6.3.3}$$

for the mode $n$; where,

$$\boldsymbol{a}_n = [a_n, b_n, c_n, d_n]^{\mathrm{T}} \quad \text{and} \quad \boldsymbol{f}_n = [p_n^1, s_n^1, p_n^2, s_n^2]^{\mathrm{T}}.$$

The expressions for the components of the $4\times4$ matrix $\mathbf{M}_n^{\text{for}}$ involve known special functions and can be found in the supplementary materail. If the matrix is well conditioned, then we can solve (6.3.3) for the coefficients $\boldsymbol{a}_n$. In practice, we numerically check whether $\mathbf{M}_n^{\text{for}}$ is well conditioned for $n = 0$, and then increase $|n| = 1, 2, \ldots$ until $\mathbf{M}_n^{\text{for}}$ is not well conditioned. Further details on when this problem is ill-posed can be found in Section 6.6.

### 6.3.2 Data on only one boundary: the inverse problem

In practice it is not possible to know the traction on both boundaries. For example in the image on the right of Figure 5.1 it is clearly not feasible to have sensors on the boundary $r = r_1$, however the boundary $r = r_2$ is often approximately traction free, due to the small air gap between the raceway and mounting. If we place ultrasonic sensors on the boundary $r = r_2$ then we would also know the displacement $\boldsymbol{u}_2$. See Figure 6.11 for an illustration. In the next section we discuss using a finite number of sensors, but for this section we consider that both the traction and displacement are known on the boundary $r = r_2$.

Analogous to the previous section, we write the displacement on the outer-boundary as a Fourier series

$$\boldsymbol{u}(r_2, \theta) = \boldsymbol{u}^2(\theta) = \sum_n u_n^{(r)} \hat{\boldsymbol{r}} \mathrm{e}^{\mathrm{i}n\theta} + \sum_n u_n^{(\theta)} \hat{\boldsymbol{\theta}} \mathrm{e}^{\mathrm{i}n\theta},$$

then by substituting the potentials (6.2.8) (6.2.7) into the expression for $\boldsymbol{u}$ in (6.2.1), and then substituting the result into the above leads to two separate equations. Combining these two equations with the two equations for the traction boundary data $\boldsymbol{\tau}^2$ from the previous section, and again using that the modes of the Fourier series are orthogonal leads to another $4\times4$ matrix equation:

$$\mathbf{M}_n^{\text{inv}} \boldsymbol{a}_n = \boldsymbol{u}_n. \tag{6.3.4}$$

If $\mathbf{M}_n^{\text{inv}}$ is well conditioned, then we can solve for $\boldsymbol{a}_n$, and this solution will also solve the forward problem (6.3.3). An example of solving a transient point force is given in Section 6.7.1

### 6.3.3 Measured points on the boundary

Let us start by summarising our results so far. In Section 6.3 we showed how to form a system

$$\mathbf{M}_n \boldsymbol{a}_n = \boldsymbol{f}_n, \tag{6.3.5}$$

for some given choice of boundary conditions, where the vector $\boldsymbol{f}_n$ contains the Fourier modes from measurements on the boundaries, i.e. the measured displacement and/or traction, and $\mathbf{M}_n$ is a known 4-by-4 matrix that depends on the type of boundary conditions. In practice, we do not have direct access to the Fourier modes of the boundary data $\boldsymbol{f}_n$, but instead measure the elastic wave displacement at specific points on the boundary. That is, by summing up the Fourier modes on both sides of the modal system (6.3.5) we obtain:

$$\sum_n \mathrm{e}^{\mathrm{i}n\theta} \mathbf{M}_n \boldsymbol{a}_n = \sum_n \mathrm{e}^{\mathrm{i}n\theta} \boldsymbol{f}_n, \tag{6.3.6}$$

where $\boldsymbol{y}(\theta) := \sum_n \mathrm{e}^{\mathrm{i}n\theta} \boldsymbol{f}_n$ represents all the boundary data as a function of $\theta$. In practice we may measure $\boldsymbol{y}(\theta)$ at specific angles $\theta$ and from this want to obtain the $\boldsymbol{a}_n$.

A sophisticated approach would consider that $\boldsymbol{y}(\theta)$ is some statistical distribution that is estimated from measured data. However, in this work we want to keep the presentation as simple as possible. So instead, we show how to rewrite the system (6.3.6) in terms of a finite number of deterministic measurements on the boundaries.

First note that $\boldsymbol{y}(\theta)$ covers two different boundaries, each of which could be sampled at different angles $\theta$. To accommodate this we rewrite (6.3.6) sampled at discrete angles

$$\sum_n \begin{bmatrix} \mathrm{e}^{\mathrm{i}n\theta^1_{m_1}} \mathbf{M}^1_n \\ \mathrm{e}^{\mathrm{i}n\theta^2_{m_2}} \mathbf{M}^2_n \end{bmatrix} \boldsymbol{a}_n = \begin{bmatrix} \boldsymbol{y}^1_{\text{inv}}(\theta^1_{m_1}) \\ \boldsymbol{y}^2_{\text{inv}}(\theta^2_{m_2}) \end{bmatrix}, \tag{6.3.7}$$

where we evaluate $m_1 = 1, 2, \ldots, M_1$ to iterate over the $M_1$ measured points $\theta^1_1, \theta^2_1, \ldots, \theta^{M_1}_1$. Likewise, we evaluate $m_2 = 1, 2, \ldots, M_2$. However, to facilitate implementation, we want to iterate over just one index $m$, rather than $m_1$ and $m_2$, which leads us to rewrite the left of (6.3.7) in the form

$$\sum_n \mathbf{E}_{mn} \boldsymbol{a}_n = \begin{bmatrix} \chi^1_m \boldsymbol{y}^1_{\text{inv}}(\theta^1_m) \\ \chi^2_m \boldsymbol{y}^2_{\text{inv}}(\theta^2_m) \end{bmatrix} \quad \text{with} \quad \mathbf{E}_{mn} := \begin{bmatrix} \chi^1_m \mathrm{e}^{\mathrm{i}n\theta^1_m} \mathbf{M}^1_n \\ \chi^2_m \mathrm{e}^{\mathrm{i}n\theta^2_m} \mathbf{M}^2_n \end{bmatrix}, \tag{6.3.8}$$

where $\chi^1_m = 1$ if $1 \leq m \leq M_1$, and otherwise $\chi^1_m = 0$, and $\chi^2_m$ has the analogous definition. We note the technicality that $\theta^j_m$ is not defined if $\chi^j_m = 0$, which we can remedy by setting $\theta^j_m = 0$ when $\chi^j_m = 0$.

To solve (6.3.8) it is best to rewrite it in the block matrix form:

$$\mathbf{E}\boldsymbol{a} = \boldsymbol{y}, \tag{6.3.9}$$

where $\mathbf{E}$ is a block matrix with the matrix block components $\mathbf{E}_{mn}$, $\boldsymbol{a}$ is a block vector from vertically stacking the vectors $\boldsymbol{a}_n$, and likewise $\boldsymbol{y}$ is a block vector which results from vertically stacking the vectors on the right side of (6.3.8).

Finally, for (6.3.9) to have a unique solution for $\boldsymbol{a}$ then the number of modes $N$ considered for $\boldsymbol{a}_n$ has to satisfy $N \leq M_1$ and $N \leq M_2$.

## 6.4 Priors and recovering the load from the rollers

The methods shown in Section 6.3 make no assumptions about the boundary conditions. If we make no assumptions about the internal geometry, or sources of the elastic waves, then we may need a lot of sensors to obtained detailed prediction of the boundary data, as shown in Section 6.7.2. To use a small number of sensors we need to provide some information, which we call priors.

For example, using a tachometer (a revolution counter), together with the design specifications of the roller bearing, we would know approximately the speed of the rollers and their contact points. We show later that this in itself is a powerful prior.

### 6.4.1 Linear priors

Any prior information about the source of waves, such as a known number of roller bearings, will allow us to parameterise $\boldsymbol{a}$ in some way. For instance, a linear parameterisation:

$$\boldsymbol{a} = \mathbf{B}\boldsymbol{x} + \boldsymbol{c}, \tag{6.4.1}$$

where the matrix $\mathbf{B}$, and bias vector $\boldsymbol{c}$, are known from the prior information, while $\boldsymbol{x}$ is now the unknown. We assume that $\mathbf{B}$ is full-column rank. For the above to be a restriction on $\boldsymbol{a}$ the matrix $\mathbf{B}$ has to have more rows than columns. We will show later how a linear relationship between $\boldsymbol{a}$ and $\boldsymbol{x}$ covers many important cases.

Substituting (6.4.1) into the block modal equation (6.3.9) then leads to

$$\mathbf{E}\mathbf{B}\boldsymbol{x} + \mathbf{E}\boldsymbol{c} = \boldsymbol{y}, \tag{6.4.2}$$

where we use a pseudo inverse to obtain a solution

$$\boldsymbol{x}_\star = (\mathbf{E}\mathbf{B})^+(\boldsymbol{y} - \mathbf{E}\mathbf{c}). \tag{6.4.3}$$

For the above to give a unique solution, we have also assumed that $\mathbf{E}$ is full rank. Substituting the above into (6.4.1) leads to

$$\boldsymbol{a}_\star = \mathbf{B}(\mathbf{EB})^+(\boldsymbol{y} - \mathbf{Ec}) + \boldsymbol{c}. \tag{6.4.4}$$

The solution $\boldsymbol{a}_\star$ would be equal to $\boldsymbol{a}$ when $\mathbf{EB}$ is a square matrix, otherwise $\boldsymbol{a}_\star$ is a least squares approximation to $\boldsymbol{a}$. We give an example for roller bearings in Section 6.7.3.

Note that while the inverse of $\mathbf{E}$ could be ill defined, the pseudo inverse of $\mathbf{EB}$ could be well defined, as we expect the dimension of $\boldsymbol{x}$ to be much smaller than the dimension of $\boldsymbol{a}$.

### 6.4.2 Prior due to boundary conditions

One of the most general ways to have a linear prior, as shown in Section 6.4.1, is to have a linear basis for the boundary conditions. For example, the boundary where the rollers make contact with the raceway leads to a basis as shown in Section 6.5.

In this section we use $\mathbf{M}_n^{\text{for}}$ and $\mathbf{E}^{\text{for}}$ to represent the modal matrix and block matrix in (6.3.5) and (6.3.9) for the boundary conditions for which we have prior knowledge. The "$\text{for}$" in $\mathbf{M}^{\text{for}}$ stands for forward problem[1]. We use $\mathbf{M}_n^{\text{inv}}$ and $\mathbf{E}^{\text{inv}}$ for problems which are ill posed, where "$\text{inv}$" represents inverse problem.

To completely determine the elastic waves within the bearing would require the boundary data $\boldsymbol{y}^1_{\text{for}}(\theta)$ and $\boldsymbol{y}^2_{\text{for}}(\theta)$, each of which can be written in terms of scalar functions in the form

$$\boldsymbol{y}^1_{\text{for}}(\theta) = \begin{bmatrix} p^1(\theta) \\ s^1(\theta) \end{bmatrix} \quad \text{and} \quad \boldsymbol{y}^2_{\text{for}}(\theta) = \begin{bmatrix} p^2(\theta) \\ s^2(\theta) \end{bmatrix}. \tag{6.4.5}$$

For example, if the boundary data $\boldsymbol{y}^1_{\text{for}}(\theta)$ represents the traction (see Figure 5.1), then $p^1(\theta)$ and $s^1(\theta)$ would represent the pressure and shear force as a function of the angle $\theta$.

To reach a linear prior (6.4.1), we assume there is a known basis for the boundary data:

$$\boldsymbol{y}^1_{\text{for}}(\theta) = \sum_{\ell=0}^{L_1} x^1_\ell \boldsymbol{y}^1_\ell(\theta) + \boldsymbol{b}^1(\theta), \quad \boldsymbol{y}^2_{\text{for}}(\theta) = \sum_{\ell=0}^{L_2} x^2_\ell \boldsymbol{y}^2_\ell(\theta) + \boldsymbol{b}^2(\theta), \tag{6.4.6}$$

where the $\boldsymbol{y}^j_\ell(\theta)$ and $\boldsymbol{b}^j(\theta)$ are known, and the $x^j_\ell$ are, for now, unknown. Each of these functions can be decomposed in Fourier modes:

$$\boldsymbol{y}^j_\ell(\theta) = \sum_n \boldsymbol{f}^j_{\ell n} \mathrm{e}^{\mathrm{i}n\theta} \quad \text{and} \quad \boldsymbol{b}^j(\theta) = \sum_n \boldsymbol{b}^j_n \mathrm{e}^{\mathrm{i}n\theta}, \quad \text{for } \ j = 1, 2. \tag{6.4.7}$$

[1] Although it is debatable what is a forward or inverse problem here.

Using the above, we can write the boundary conditions for one mode in the form:

$$\mathbf{M}_n^{\text{for}} \boldsymbol{a}_n = \begin{bmatrix} \mathbf{F}_n^1 \boldsymbol{x}^1 \\ \mathbf{F}_n^2 \boldsymbol{x}^2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{b}_n^1 \\ \boldsymbol{b}_n^2 \end{bmatrix}, \tag{6.4.8}$$

where we define

$$\mathbf{F}_n^j = \begin{bmatrix} \boldsymbol{f}_{1n}^j & \boldsymbol{f}_{2n}^j & \cdots & \boldsymbol{f}_{L_j n}^j \end{bmatrix}, \tag{6.4.9}$$

so that $\mathbf{F}_n^j \boldsymbol{x}^j = \sum_\ell \boldsymbol{f}_{\ell n}^j x_\ell^j$. To write the above in a block matrix form we define

$$\boldsymbol{b}_n = \begin{bmatrix} \boldsymbol{b}_n^1 \\ \boldsymbol{b}_n^2 \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}^1 \\ \boldsymbol{x}^2 \end{bmatrix}, \quad \text{and} \quad \mathbf{F}_n = \begin{bmatrix} \mathbf{F}_n^1 & 0 \\ 0 & \mathbf{F}_n^2 \end{bmatrix}, \quad \text{so that} \quad \mathbf{F}_n \boldsymbol{x} = \begin{bmatrix} \mathbf{F}_n^1 \boldsymbol{x}^1 \\ \mathbf{F}_n^2 \boldsymbol{x}^2 \end{bmatrix}, \tag{6.4.10}$$

and then rewrite (6.4.8) in a block form to obtain

$$\mathbf{M}^{\text{for}} \boldsymbol{a} = \mathbf{F}^{\text{for}} \boldsymbol{x} + \boldsymbol{b}^{\text{for}}, \quad \text{with} \quad \mathbf{F}^{\text{for}} = \begin{bmatrix} \vdots \\ \mathbf{F}_{-1} \\ \mathbf{F}_0 \\ \mathbf{F}_1 \\ \vdots \end{bmatrix} \quad \text{and} \quad \boldsymbol{b}^{\text{for}} = \begin{bmatrix} \vdots \\ \boldsymbol{b}_{-1} \\ \boldsymbol{b}_0 \\ \boldsymbol{b}_1 \\ \vdots \end{bmatrix}, \tag{6.4.11}$$

where $\mathbf{M}^{\text{for}}$ is a block diagonal matrix with $\mathbf{M}_n^{\text{for}}$ on the diagonals, and we have added the superscript "$\text{for}$" to emphasize that these above quantities are related to the forward problem. Note that if we knew the boundary data of the forward problem we would have $\boldsymbol{f}^{\text{for}} = \mathbf{F}^{\text{for}} \boldsymbol{x} + \boldsymbol{b}^{\text{for}}$. Finally, we take the inverse of $\mathbf{M}^{\text{for}}$ on both sides of (6.4.11) to obtain

$$\boldsymbol{a} = (\mathbf{M}^{\text{for}})^{-1} \mathbf{F}^{\text{for}} \boldsymbol{x} + (\mathbf{M}^{\text{for}})^{-1} \boldsymbol{b}^{\text{for}}, \tag{6.4.12}$$

where we have assumed that the type of boundary conditions that lead to $\mathbf{M}^{\text{for}}$ lead to a well conditioned problem so that calculating the inverse $(\mathbf{M}^{\text{for}})^{-1}$ is stable and well defined.

The restriction (6.4.12) on $\boldsymbol{a}$ now matches the abstract form given by (6.4.15), where by comparison we obtain:

$$\mathbf{B}^{\text{for}} = (\mathbf{M}^{\text{for}})^{-1} \mathbf{F}^{\text{for}} \quad \text{and} \quad \boldsymbol{c}^{\text{for}} = (\mathbf{M}^{\text{for}})^{-1} \boldsymbol{b}^{\text{for}}. \tag{6.4.13}$$

We can use the above restriction to solve for $\boldsymbol{a}$ even when given incomplete boundary data. Let us write this out in full for clarity.

Let $\mathbf{E} = \mathbf{E}^{\text{inv}}$ and $\boldsymbol{y} = \boldsymbol{y}^{\text{inv}}$ in (6.3.9) to indicate that calculating $(\mathbf{E}^{\text{inv}})^{-1}$ is either ill-posed

or that the measured $\boldsymbol{y}^{\text{inv}}$ is incomplete boundary data. Our aim is now to solve

$$\mathbf{E}^{\text{inv}}\boldsymbol{a} = \boldsymbol{y}^{\text{inv}}. \tag{6.4.14}$$

Using the result (6.4.4) together with the substitutions (6.4.13) leads to the solution

$$\boldsymbol{a}_\star = \mathbf{B}^{\text{for}}(\mathbf{E}^{\text{inv}}\mathbf{B}^{\text{for}})^{+}(\boldsymbol{y}^{\text{inv}} - \mathbf{E}^{\text{inv}}\mathbf{c}^{\text{for}}) + \boldsymbol{c}^{\text{for}}. \tag{6.4.15}$$

It is likely easier to understand this result, and its consequences, with a concrete example which we provide for roller bearings in Section 6.5. Nonetheless, let us consider some here important features of this solution.

To simplify the discussion here, let us assume that the number of boundary measurements $\boldsymbol{y}^{\text{inv}}$ is equal to the number of unknowns in $\boldsymbol{x}$ so that $\boldsymbol{a}_\star = \boldsymbol{a}$ and $\mathbf{E}^{\text{inv}}\mathbf{B}^{\text{for}}$ is a square matrix. So if we have a representation for $\boldsymbol{a}$ that uses a small number of basis elements $L_1$ and $L_2$ in (6.4.6), then we need only a small number of measurements in $\boldsymbol{y}^{\text{inv}}$ to obtain the unique solution $\boldsymbol{a}$. In Section 6.5 we show how assuming a smooth loading of a roller bearing leads to small values for $L_1$ and $L_2$. To further emphasize this point, note that the resolution of the solution is governed by the number of modes $N$ in $\boldsymbol{a}$. That is, the block vector $\boldsymbol{a}$ is formed of the vectors $\boldsymbol{a}_n$ with $n$ having $N$ possible values. For a fixed number of basis' $L_1$ and $L_2$ in (6.4.6) we can increase $N$ and still obtain the unique solution $\boldsymbol{a}$ as long as $\mathbf{M}^{\text{for}}$ in (6.4.12) continues to be well conditioned. This is why the images in Figure 6.1 have such high resolution, despite having only 3 sensors.

### 6.4.3 One traction free boundary

Here we give an example which is typical for bearings: the traction on one boundary is known, and the boundary in contact with the roller bearings has a basis function. For example the outside boundary in Figure 5.1 could be traction free.

For this case, we need to make a small adjustment to the prior method shown in the previous section. Here the boundary data is of the form

$$\boldsymbol{y}^1_{\text{for}}(\theta) = \sum_{\ell=0}^{L_1} x^1_\ell \boldsymbol{y}^1_\ell(\theta), \quad \boldsymbol{y}^2_{\text{for}}(\theta) = \boldsymbol{b}^2, \tag{6.4.16}$$

so that $\boldsymbol{b}^1(\theta) = \mathbf{0}$ and $\boldsymbol{x}^2 = \mathbf{0}$.

Following the same steps shown in the previous section, we would need to make a small adjustment by redefining

$$\boldsymbol{x} = \boldsymbol{x}^1, \quad \text{and} \quad \mathbf{F}_n = \begin{bmatrix} \mathbf{F}^1_n \\ 0 \end{bmatrix}. \tag{6.4.17}$$

## 6.5 Roller bearings and the loading profile

Here we develop an application for roller bearings that shows the great potential of describing the elastic waves in the raceway in more detail.

Consider a roller bearing as shown on the left of Figure 6.3 and on the right of Figure 5.1. Any load applied to the shaft in the middle of the bearing, or applied to the outer raceway, will be transmitted through the rollers themselves with each roller in contact with only a small region of the raceway, as illustrated in Figure 6.4. We do not need to know the exact shape of this small contact region if the goal is just to measure the overall load passed through the bearing; as long as the contact region is small compared to the bearing geometry. Below we show how knowing the rotation speed of the bearing, number of bearings, and their contact points, can lead us to predict the load transmitted through each bearing with very few sensors.

We make a number of simplifying assumptions, which can improved on in future work. First, in practice, rollers slip as they go around [14], making their contact points better described as a random variable. Further, the contact points of the bearings with the raceway are more accurately modelled as Hertzian contacts [79]. However, here we show only how to use deterministic priors both for simplicity but because it is necessary to develop the deterministic version first before developing more precise models with random variables. Second, we assume the bearing is rotating at a constant speed $\Omega$. The framework we present can accommodate any change in rotation speed, but the conclusions shown below would need to be adjusted.

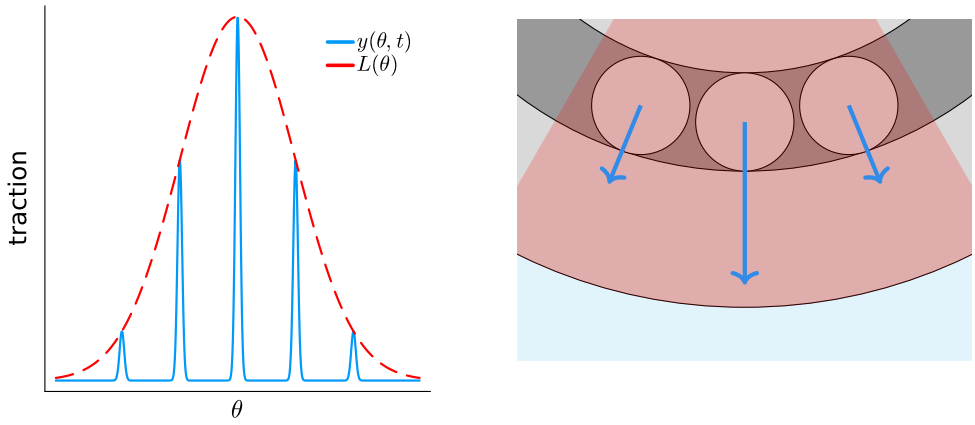



Figure 6.4: The graph on the left shows the stress on the raceway boundary $y(\theta, t)$. Each blue spike is the result of one roller being in contact with the raceway. As the rollers move in time, the blue spikes in this graph also move, but they all trace the same curve $L(\theta)$, where we assume the load supported by the bearing does not change in time. The image on the right illustrates how load is transmitted through the rollers.

### 6.5.1 Static Loading profile

Imagine the bearings are loaded just due to gravity, or some other static forces on the shaft, mounting, or casing. We use the function $L(\theta)$ to denote the stress transmitted through a roller when it is in contact with the raceway at an angle $\theta$. We call $L(\theta)$ the loading profile and in this section assume it is independent of time. So this excludes enviroment effects, for example.

Assume that $L(\theta)$ represents the radial stress for simplicity, and that the contact region of the roller is small, then we can write that the radial stress on the boundary $y(\theta, t)$ of the raceway is given by

$$y(\theta, t) = L(\theta)d(\theta, t), \quad \text{where} \quad d(\theta, t) = \sum_{s=-\infty}^{\infty} \delta\left((\theta - \Omega t)Z + 2\pi s\right), \tag{6.5.1}$$

where $Z$ is the number of roller bearings, $\Omega$ their angular speed, $t$ is time, and the function $\delta(\theta)$ represents the stress distribution due to one roller. See Figure 6.4 for an illustration. A form similar to (6.5.1) for the stress on the boundary was introduced in [15] and [16].

The function $d(\theta, t)$ moves the contact points of the bearing as time passes, and assumes that the $Z$ bearings have the same distance between each other. We assume that each $\delta\left((\theta - \Omega t)Z + 2\pi s\right)$ when integrated over $\theta$ is equal to 1, so that the magnitude of the load transmitted is always $L(\theta)$, no matter the shape of the function $\delta(\theta)$. We could, for example have the function $\delta$ be as scaled Dirac delta function. However to avoid Gibbs phenomena it is best to use a Gaussian function:

$$\delta\left(x\right) = \frac{Z}{\sigma}\mathrm{e}^{-\pi x^2/\sigma^2}, \tag{6.5.2}$$

where $\sigma$ is the standard deviation of the contact spread. In the limit of $\sigma \to 0$ the above $\delta$ would become a scaled Dirac delta. Naturally, we could use other type of contact points, but when the aim is to measure $L(\theta)$ we do not need to model precisely the contact region.

The function $d(\theta, t)$ is periodic in time with period $T = 2\pi/(Z\Omega)$, which means we can write $d(\theta, t)$ in terms of its Fourier series in time, which (after some calculations) is given by

$$d(\theta, t) = \frac{Z}{2\pi} \sum_{m=-\infty}^{\infty} \mathrm{e}^{-\pi\sigma^2 m^2} \cos(mZ[\theta - \Omega t]), \quad \text{where} \quad \omega_m = mZ\Omega. \tag{6.5.3}$$

The loading profile $L(\theta)$ is also $2\pi$ periodic in $\theta$, so we use a Fourier series representation:

$$L(\theta) = \sum_n c_n \mathrm{e}^{\mathrm{i}n\theta}, \tag{6.5.4}$$

which substituted into (6.5.1) together with (6.5.3), and after some calculations, leads to

$$y(\theta,t)=\sum_{n,m} f_n(\omega_m)\mathrm{e}^{\mathrm{i}n\theta}\mathrm{e}^{-\mathrm{i}\omega_m t},\quad \text{with}\quad f_n(\omega_m)=\frac{Z}{2\pi}c_{n-mZ}\mathrm{e}^{-\pi\sigma^2m^2}, \tag{6.5.5}$$

which matches the notation from the previous sections.

Using the above, together with the prior method shown in Section 6.4.2 and Section 6.4.3, we can recover the coefficients of the loading profile $c_n$ by measuring the displacement on the boundary of the raceway that is traction free. To do so, we first identify the unknowns $x_\ell = c_\ell$, then the matrix $\mathbf{F}_n^1$ would be full of zero except for the column number $\ell = n - mZ$ which would be

$$\boldsymbol{f}_{\ell,n}^1=\frac{Z}{2\pi}\begin{bmatrix}\mathrm{e}^{-\pi\sigma_P^2m^2}\\ \mu_S\mathrm{e}^{-\pi\sigma_S^2m^2}\end{bmatrix},$$

where we assume the contact force distribution for the pressure $\sigma_P$ is potentially different than the contact force for the shear $\sigma_S$, and we also assume that if the pressure is known, then the shear is known as consequence.

In the examples section, we show that four different sensors are needed to recover the load accurately, in steel, if the loading profile is smooth, as shown in Section 6.7.3. This is because a smooth loading profile implies that the series (6.5.4) needs few terms to converge.

### 6.5.2 Quasi-Static Loading profile

The most general force due to the rollers bearings on the raceway is given by

$$y(\theta,t)=L(\theta,t)d(\theta,t), \tag{6.5.6}$$

instead of (6.5.6). By taking the Fourier transform of both sides and using (6.5.3) with convolution theorem we obtain

$$\hat{y}(\theta,\omega)=\frac{Z}{2\pi}\sum_m \mathrm{e}^{-\pi\sigma^2m^2}\left(\hat{L}(\theta,\omega-\omega_m)\mathrm{e}^{\mathrm{i}mZ\theta}+\hat{L}(\theta,\omega+\omega_m)\mathrm{e}^{-\mathrm{i}mZ\theta}\right), \tag{6.5.7}$$

where $\hat{y}(\theta,\omega)$ and $\hat{L}(\theta,\omega)$ are the Fourier transforms of $y(\theta,t)$ and $L(\theta,t)$ respectively.

The form (6.5.7) would not be a useful prior if we knew nothing about $\hat{L}$. However, there is a useful and practical assumption that the loading is quasi-static, i.e. does not change rapidly. The simplest scenario being that $\hat{L}(\theta,\omega)\approx 0$ for $|\omega| > Z\Omega$, in which case the sum in (6.5.7) reduces to just one value for $m$, leading to

$$\hat{y}(\theta,\omega)=\frac{Z}{2\pi}\mathrm{e}^{-\pi\sigma^2m^2}\left(\hat{L}(\theta,\omega-\omega_m)\mathrm{e}^{\mathrm{i}mZ\theta}+\hat{L}(\theta,\omega+\omega_m)\mathrm{e}^{-\mathrm{i}mZ\theta}\right), \tag{6.5.8}$$

where $m=\left\lfloor\frac{\omega}{Z\Omega}\right\rceil$, with $\lfloor x\rceil$ being equal to $x$ rounded to the nearest integer.

Analogous to the previous section, we decompose $\hat{y}$ and $\hat{L}$ in their Fourier modes:

$$f_n(\omega) = \frac{Z}{2\pi} \mathrm{e}^{-\pi\sigma^2 m^2} c_{n-mZ}(\omega - \omega_m). \tag{6.5.9}$$

## 6.6 What is measurable

It is not always possible to robustly estimate the stresses between the rollers, or other elements with elastic waves. There are two main phenomena that cause this: 1) resonance and 2) the diffraction limit. Numerically, we observe a relationship between the stability of the inverse problem and the frequency $\omega$. In particular, we find that the numerical stability of the inverse problem increases as we increase the frequency, see Figure 6.5. This relationship between numerical stability and frequency is a well-established phenomenon in inverse problems for Helmholtz equations with Cauchy boundary conditions [49, 53, 54]. In particular [49, 54] show that for problems such as ours, the numerical stability of the inverse problem increases with the wavenumber $k$.

When hitting a resonance, the field inside (6.2.4) the raceway varies significantly with small changes of the boundary data. This occurs for the forward problem and interferes in using the prior method. To determine this precisely, we can turn to the modal system (6.3.3) or (6.3.4) and check if the matrix $\mathbf{M}_n$ is well conditioned, with one example shown in Figure 6.5.

Due to the diffraction limit [58], for any fixed frequency there is a limited amount of information, or resolution that can be extracted. The maximum spatial resolution that can be recovered from the boundary is given by the largest mode number $n$ used in the expansions (6.2.4). When fixing $\omega$, the problem becomes more ill-posed as $n$ grows larger.

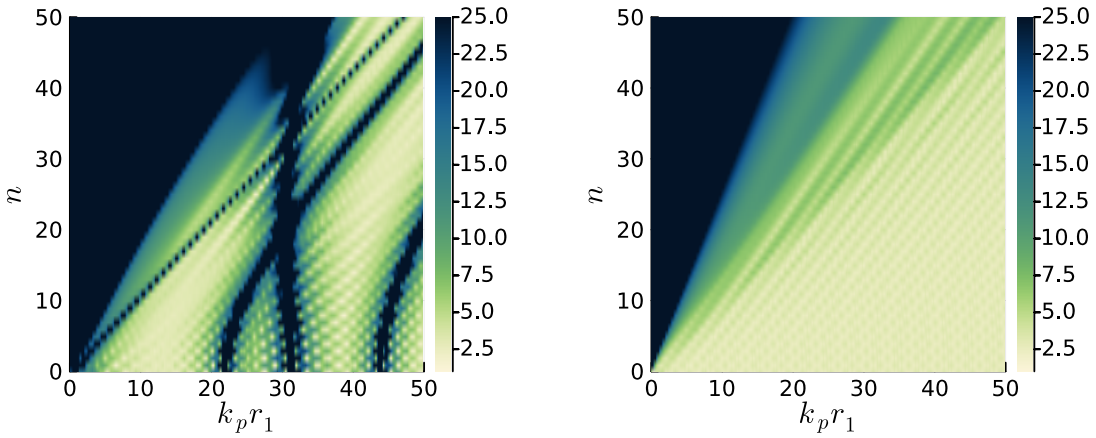



Figure 6.5: The above heatmaps show the condition number cond $\mathbf{M}_n^{\text{for}}$ on the left and cond $\mathbf{M}_n^{\text{inv}}$ on the right, after non-dimensionalisation. In the dark regions the errors in the boundary data can be amplified by 20 times. Note that for any fixed $k_p r_1$ if we keep increasing $n$ the system will at some point become ill-conditioned. The dark cross and lines on the forward problem are resonant modes. The parameters used are given in Table 6.1.

The condition number of $\mathbf{M}_n$ depends on the material parameters and geometry. For instance, for the raceway in Table 6.1 we can see from Figure 6.5 that both $\mathbf{M}_n^{\text{for}}$ and $\mathbf{M}_n^{\text{inv}}$ can only be well conditioned if, approximately:

$$|k_p| r_1 > |n|, \tag{6.6.1}$$

although there are many frequencies and modes $n$ that are ill posed for $\mathbf{M}_n^{\text{for}}$ inside this region. Specifically, the condition number cond $\mathbf{M}_n^{\text{for}}$ shows dark lines where the condition number is high. These indicate that the system is close to resonance, as small values of the boundary data lead to large values of the field. These dark lines depend on both $r_2$ and $r_1$, however we note that: when the ratio $r_2/r_1$ gets closer to 1 the lines move to higher frequencies but get thicker, and when $r_2/r_1$ gets larger, more and more lines move from high frequencies to lower frequencies, but get thinner and thinner.

When the modal system is well conditioned, then the mode number $n$ can be measured when using $2n$ sensors, as illustrated in the example in Section 6.7.1. This is a rather high demand on sensors to reach a reasonable resolution. The number of sensors needed for roller bearings, developed in Section 6.5, is very different as we discuss below.

| Parameter | value | description |
|---|---|---|
| $r_1$ | $1.0$ m | inner radius |
| $r_2$ | $1.1$ m | outer radius |
| $c_p$ | $5000$ m/s | pressure speed |
| $c_s$ | $3500$ m/s | shear speed |
| $\rho$ | $7000$ kg/m$^3$ | mass density |

Table 6.1: the parameter values that approximate a steel raceway. These parameters are used for most numerical examples.

### 6.6.1 Roller bearings at constant speed

Here our focus is to learn the Fourier coefficients of the loading profile $c_n$, shown in (6.5.4). In this section we explain an important lesson from the mathematics: the higher the rotation speed of the rollers $\Omega$, the better conditioned the modal system becomes, and the more coefficients $c_n$ it is possible to measure. Specifically, the lowest order coefficients $c_0$, $c_{-1}$, $c_1$, $c_{-2}$, ..., become inaccessibly if the speed $\Omega$ is too slow.

Let us consider an example of a steel raceway with the properties shown in Table 6.1. To use the prior method, we need to invert the matrices $\mathbf{M}_n^{\text{for}}$ as shown in (6.4.12). This inversion is only stable when (6.6.1) holds. If we substitute the angular frequency $\omega_m = mZ\Omega$ from (6.5.3) into (6.6.1) we obtain

$$Z|m|C > |n|, \quad \text{where } C = \Omega r_1/\alpha,$$

note that the $\omega_m$ are the only frequencies available for constant rotation speed. Now the goal is to obtain the coefficients $c_{n-mZ}$ from (6.5.9). For clarity we define $\ell = n - mZ$ and substitute $n = \ell + mZ$ in the above, and manipulating, to reach the restriction:

$$-Z(|m|C + m) < \ell < Z(|m|C - m). \tag{6.6.2}$$

We can measure different frequencies $\omega_m$, which in turn implies we can choose different values for $m$. For each value $m$, the restriction (6.6.2) determines which values for $\ell$ are possible to measure. Despite this liberty, if $C$ is small, then (6.6.2) will still significantly restrict all possible values for $\ell$. For an example, assume that $C < 1$. For the raceway defined in Table 6.1 we have that $C < 1$ when $\Omega < 10$ rev/min.

Let us consider the cases:

$$-Z(C + 1) < \ell < -Z(1 - C), \qquad \text{for } \; m = 1. \tag{6.6.3}$$

$$Z(1 - C) < \ell < Z(1 + C), \qquad \text{for } \; m = -1, \tag{6.6.4}$$

If $C = 0.5$ then the first and second inequality above would read $-1.5Z < \ell < -0.5Z$ and $0.5Z < \ell < 1.5Z$ respectively, which together imply that $|\ell| > 0.5Z$. Larger values for $|m|$ would lead to restrictions where $|\ell|$ has to be larger. The number of rollers $Z$ can be anything larger than 10, so that $|\ell| > 0.5Z$ would become $|\ell| > 5$. In other words, the loading coefficients $c_\ell$ for $\ell = -4, -3, \ldots, 4$ could not be reliably measured.

The parameter $C$ can only increase, for one fixed raceway, when the speed of rotation $\Omega$ increases. With this increased speed, more modes of the loading become available to measure by measuring elastic waves. One way to interpret this is in terms of the static limit.

### 6.6.2 Static vs dynamic regimes

In the previous section we learned that if the rollers spin too slowly then the lowest order modes of the loading $|\ell|$ cannot be robustly measured. This is because as $\Omega$ slows down, we approach the static limit. A simple way to check if we are approaching the static limit is to compare the elastic wave speed with the speed of rotation of the rollers. That is, the ratio:

$$\frac{\Omega r_1}{c_p} \quad = \quad \text{the roller to wave speed ratio.} \tag{6.6.5}$$

If the above is very small, then the rollers are almost standing still relative to the wave speed, and therefore the solution could be calculated by using static stress balance, which is known to be ill-posed [80, 81, 82], and the equations for the potentials (6.2.4) tend to Laplace equations which are also ill-posed[49, 50]. As discussed in the previous section, for the steel bearing in Table 6.1 having $c = 1$ implies that $\Omega \approx 1.05$ rad/s, which substituted into the ratio above leads to $2 \times 10^{-4}$. In conclusion, to predict the complete load due to the rollers

(rotating at a constant speed) becomes well-posed if the rotation speed $\Omega$ is large enough. In practice, there are several ways around this limitation, as we explain next.

**Localised defects and forces.** Some important goals do not require a complete measurement of the loading through the rollers. An example of this is to detect a localised defect on, or near, the boundary in contact with the rollers. In this case, the Fourier coefficients $c_\ell$, of the loading profile, for larger $\ell$ will be significant. These can be measured as shown in Section 6.7.4. Here we explain why this is possible in terms of algebra.

Suppose we are using a low frequencies $\omega_m$ which, due to the diffraction limit, implies we can only measure small values of $n$ of the boundary conditions $f_n$. Turning to (6.5.9), and setting as an example $n = 0$, we could measure the coefficients $c_{-mZ}$ of the loading profile, where $mZ$ are high modes (as $Z > 10$) which are related to locating defects, as shown in Section 6.7.4.

## 6.7 Examples

A large number of scenarios are rigorously tested in the folder `test` of the package `ElasticWaves.jl` [83] where `MultipleScattering.jl` [84] was also used. Below we show a few examples of methods developed in this paper to both validate and illustrate our method.

In Section 6.7.1 we show the modes for both the forward and inverse problem and explain where they are ill-posed. Section 6.7.2 shows an example of generating, and predicting, a localised for on the inner raceway. There we learn that many sensors are needed to accurately predict a localised mode if we make no assumptions about the forces. If we assume there are rollers travelling at a constant speed, then as shown in Section 6.7.3 we can greatly reduce the number of sensors. Finally, Section 6.7.4 shows an example where vibrations are due to rollers hitting a localised defect, and what can be recovered using the inverse system..

### 6.7.1 The forward and inverse modal systems

In the first sections of the paper we introduced the forward and inverse modal systems which are shown in (6.3.3) and (6.3.4). Figure 6.5 shows where these systems likely lead to stable solutions. Here we provide examples that the inverse problem truly recovers the boundary conditions of the forward problem. We start with a sweep over all modes and frequencies.

**Boundary conditions.** For every mode $n$ and frequency $\omega$ we choose:

$$\boldsymbol{f}_n = [1, 1, 0, 0]^{\mathrm{T}},$$

for the forward problem to immitate some forcing on the inner face of the raceway. We then add a uniform random 2% error and solve $\mathbf{M}_n^{\text{for}} \boldsymbol{a}_n = \boldsymbol{f}_n$ for $\boldsymbol{a}_n$. Then, to setup the inverse problem, we substitute $\boldsymbol{a}_n$ in (6.2.4) and from these calculate the boundary data $\boldsymbol{u}_n$, the

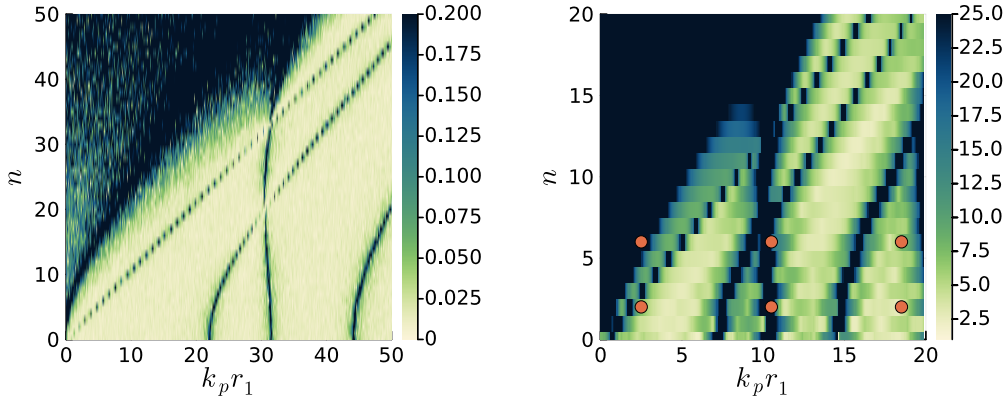



Figure 6.6: On the left is the error in the traction $\boldsymbol{\tau}_1$, shown by (6.7.1), with $\boldsymbol{\tau}_1$ predicted by the inverse problem after adding 2% error to all boundary data. Table 6.1 shows the parameters used. The right shows the condition number of $\mathbf{M}_n^{\text{for}}$ but for a thicker raceway: $r_1 = 1.0$ and $r_2 = 1.3$, which is easier to visualise the modes, where the modes are shown in Figure 6.7 for the parameters of the orange spots.

traction and displacement on the outer boundary. We then add 2% error to $\boldsymbol{u}_n$ and solve $\mathbf{M}_n^{\text{inv}}\boldsymbol{a}_n = \boldsymbol{u}_n$ for $\boldsymbol{a}_n$, and finally use this $\boldsymbol{a}_n$ to predict $\boldsymbol{\tau}_1$, the traction for $r = r_1$.

**The heatmap of error.** The error given by

$$\text{error} = |\boldsymbol{\tau}_1 - [1,1]^{\text{T}}|/\sqrt{2}, \tag{6.7.1}$$

and is shown as a heatmap over all modes and frequencies in Figure 6.6 (the image on the left). We can see that most of the heatmap has an error of around $2\%$, meaning that most modes and frequencies lead to a well posed problem.

**Visualising the modes.** As explained in Section 6.6, the main causes that increase the error are 1) the diffraction limit and 2) resonant modes. To help visualise, we plot some modes in Figure 6.7, with the mode number $n$ and wavenumber $k_p r_1$ of these modes shown by the orange spots in the heatmap on the right of Figure 6.6. We have chosen to use a thicker raceway with $r_1 = 1.0$ and $r_2 = 1.3$, but with the same material properties in Table 6.1, to better visualise the modes. The condition numbers of the forward problem for the thicker raceway is similar to the thinner raceway and is shown on the right of Figure 6.6.

**The diffraction limit modes.** The top and bottom left modes in Figure 1 are close to the diffraction limit, meaning the potentials approximately obey Laplace's equation. Notably, solutions to Laplace's equation are also solutions to a diffusion equation, where any source smoothly dissipates as it moves away from its origin. Solutions to the Laplacian are known to be ill posed and are essentially the same as the diffraction limit [49, 50].

**Near resonant modes.** The two images in the middle column of Figure 6.7 illustrate near resonant modes. That is the pressure on both boundaries is near zero, while the pressure away from the boundaries grows. This is why small errors in the boundaries lead to large errors in the fields for these modes.

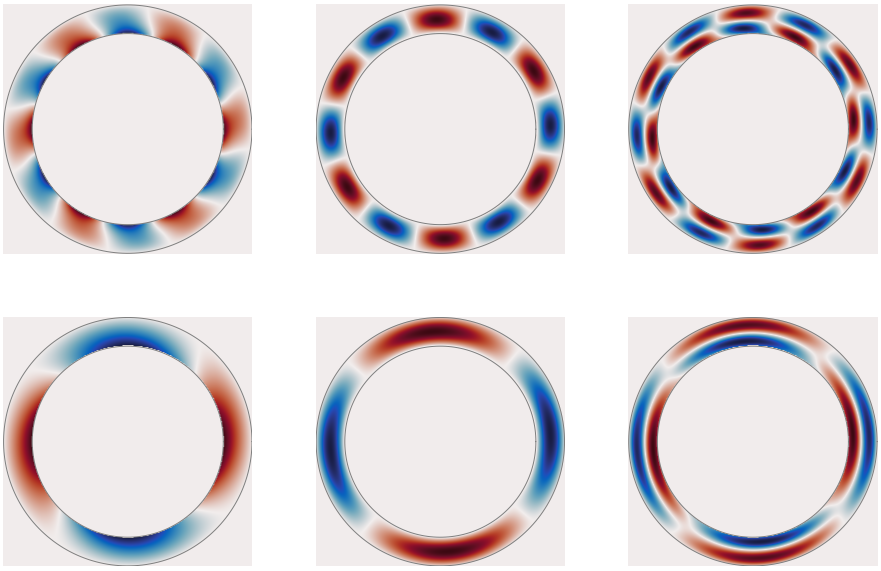


Figure 6.7: Above shows the real part of the pressure field from solving the forward system (6.3.3) for the modes $n = 2$ or $6$, and $k_p r_1 = 2.5, 10.5,$ or $18.5$. Red (blue) is positive (negative) pressure, while white is no pressure. The colour scaling is different for each mode, but the outer boundary is always traction-free so it is white. The outer radius $r_2 = 1.3$ which is easier to visualise then $r_1 = 1.1$. The chosen modes are shown as an orange scatter on Figure 6.6. Middle plots show resonance so large field inside for small boundary data.

**Well-posed modes.** The top and bottom right modes shown in Figure 6.7 are well posed, as both phase information is still present, and the pressure is not small on all boundaries.

### 6.7.2 A localised force on the boundary

In the previous section we saw that the inverse problem in general works, when outside of resonance or the low frequency limit. Here we show an example where although the inverse problem is well posed you need many Fourier modes to converge, and therefore many sensors. The material parameters and dimensions of the bearing used here are shown in Table 6.1. This section acts as motivation for using the prior method developed in Section 6.4.

**Boundary conditions and method.** Consider a sharp Gaussian force applied to the inner boundary given by

$$f(\theta) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(\theta-\pi)^2}{2\sigma^2}}, \tag{6.7.2}$$

where $\sigma = 0.1$. For the forward model we use the boundary conditions

$$\boldsymbol{\tau}^1 = f(\theta)\hat{\boldsymbol{r}} \quad \text{and} \quad \boldsymbol{\tau}^2 = \mathbf{0},$$

which for one fixed frequency $\omega$ leads to the modal system

$$\mathbf{M}_n^{\text{for}} \boldsymbol{a}_n = [f_n, 0, 0, 0]^T, \tag{6.7.3}$$

where $f_n$ is the $n$th coefficient of the Fourier series expansion of the forcing (6.7.2). A large number of coefficients $f_n$ are needed to accurately represent $f(\theta)$, which is why this example will need many sensors to obtain a good resolution. Solving Equation (6.7.3) for each mode $n$ gives a solution to the forward problem.

Just as before, to setup the inverse problem, we use the forward problem to predict the displacement $\boldsymbol{u}_n$ on the outer boundary $r = r_2$, and then add 1% error to $\boldsymbol{u}_n$ and then solve the inverse modal problem (6.3.4) for $\boldsymbol{a}_n$. With $\boldsymbol{a}_n$ we then predict the traction $\boldsymbol{\tau}_1$ and compare it with the true traction. The results for the frequency $k_p r_1 = 65$ are shown in Figure 6.8 where the ribbon is $10$ the standard deviation from solving this problem many times each with a different error added.

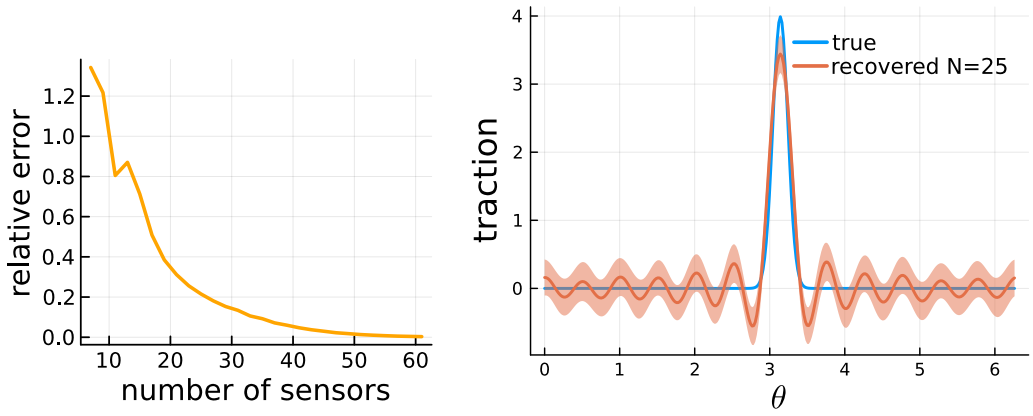



Figure 6.8: The left shows the convergence of solving the inverse problem, with 1% error added to its boundary conditions, to recover the force shown on the right. As the number of sensors increases so does the number of Fourier modes $n$ that can be recovered. The right shows the forcing $f(\theta)$ given by (6.7.2) compared with the mean predicted force from the inverse problem for $k_p r_1 = 65$, using 25 sensors, and with a $10$ standard deviation ribbon to show the uncertainty.

**Results.** From the thickness of the ribbon in the plot on the right of Figure 6.8 we can see that the problem is well posed for $k_p r_1 = 65$. This relatively high frequency avoids most resonances, and allows us to recover very high Fourier modes by avoiding the diffraction limit. However, the number of sensors needed to reach a relative error less than 20% is around 25, as can be seen from the graph on the left of Figure 6.8.

Clearly a Fourier series representation $f(\theta) = \sum_n f_n \mathrm{e}^{\mathrm{i}n\theta}$ is not the best choice for a localised force. However, by making assumptions about what led to the force, or more generally the boundary conditions, we can use many other representations. For example, we can assume the traction is due to contact with rollers bearings rotating at a constant speed, as we do in the next section.

### 6.7.3 Recovering the loading profile for rollers

Here we show an example of predicting the loading on the bearings by using the methods developed in Section 6.5.1 that assume the rollers are rotating at a constant speed.

**Stribeck boundary condition.** To create a realistic boundary conditions for the inverse problem we use the Stribeck equation for the loading profile of roller bearings [1], it is given by

$$L(\theta) = L_0 \left(1 - \frac{1}{2\epsilon}(1 - \cos\theta)\right)^{10/9}, \tag{6.7.4}$$

where $\epsilon$ is called the load distribution factor. This parameter determines the loading zone, that is, the region where the load is being applied. For a radial loading, it is related to the loading region through:

$$\epsilon = \frac{1}{2}\left(1 - \cos(\varphi/2)\right) \tag{6.7.5}$$

where $\varphi$ is the angular extent of the loading zone.

For the numerical simulations, we used $L_0 = 1$ and $\epsilon = 0.5$, which implies a loading zone of angular length of $\varphi = \pi$. Figure 6.1 shows this Stribeck loading profile. The other parameters used for this section are shown in Table 6.2, where we use a thicker raceway with thickness 1m to make the plots below easier to see.

| Parameter | value | description |
|---|---|---|
| $r_1$ | $2.5$ m | inner radius |
| $r_2$ | $3.5$ m | outer radius |
| $c_p$ | $5000$ m/s | pressure speed |
| $c_s$ | $3500$ m/s | shear speed |
| $\rho$ | $7800$ kg/m$^3$ | mass density |
| $\Omega$ | $2000$ rpm | rotation speed |

Table 6.2: Parameter values used for numerical simulations in Section 6.7.3.

**Data from the forward problem.** Like the previous sections, we create the boundary data for the inverse problem, represented by $\boldsymbol{y}^{\text{inv}}$ in (6.4.14), by solving the forward problem. In this section, for the forward problem we used the loading (6.7.4), from which the $c_n$ coefficients in (6.5.4) can be calculated, which then lead to the Fourier modes $f_n$ shown in (6.5.5) which we use for the boundary conditions of the forward problem. Again we assume the outer boundary $r = r_2$ is traction free with $\boldsymbol{\tau}_1 = \mathbf{0}$.

After solving the forward problem, we can then calculate $\boldsymbol{y}^{\text{inv}}$ from (6.3.9) for a chosen number of sensors, where $\boldsymbol{a}$ is given by solving the forward problem, and $\mathbf{E}$ is composed of the modal matrices for the inverse problem.

**Results.** To solve the inverse problem using the prior method we can follow the steps shown at the end of Section 6.5.1. The result is that at least 4 sensors are needed to recover the loading profile, when adding 4% noise, as shown by the Figure 6.1 in the introduction, which shows the predicted pressure distribution, and Figure 6.9 which shows the predicted displacement.

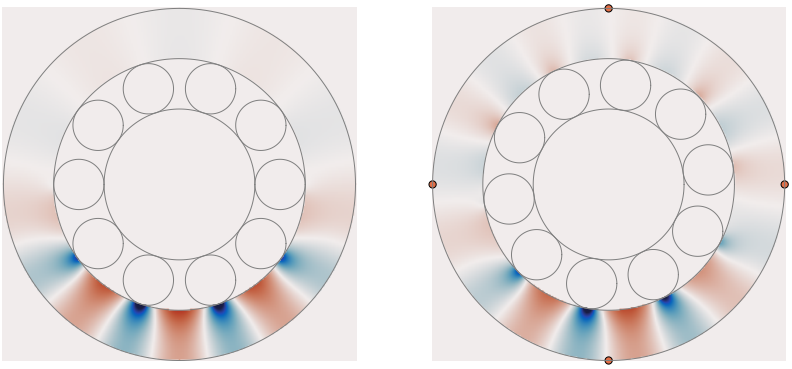


Figure 6.9: True radial displacement on the left for one snapshot in time when using the Stribeck equation (6.7.4) and the properties in Table 6.2. The snapshot in time is a result of taking a Fourier transform over all frequencies. The right shows the predicted radial displacement when using only 4 sensors shown as orange spots. The sensors measure displacement, and the outer boundary is stress free. We can see that despite being stress free on the boundary $r = r_2$, the displacement is not zero there. The recovery is not perfect as 4% noise is added, and limit sensors imply limited Fourier modes are recovered.

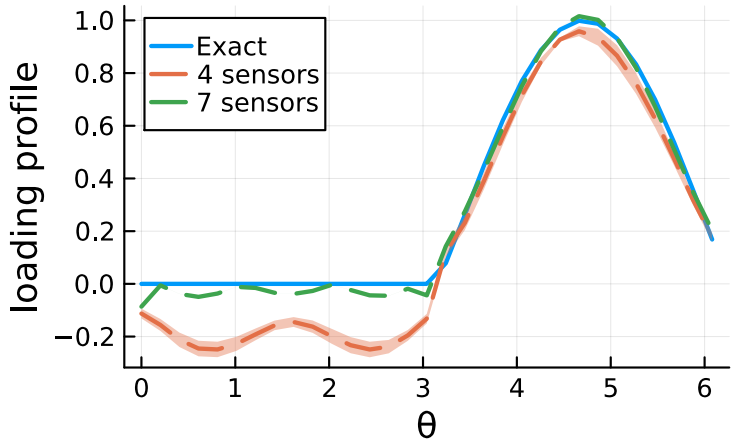



Figure 6.10: The blue curve shows the Stribeck equation for the loading of rollers given in (6.7.4), the orange is prior method with 4 sensors and modes -2:2 and 4% added error to boundary data, other than 20% error caused from lacking Fourier mode. Also added 4% white noise error on top. Green is also modes -2:2 and 4% added error, but with 7 sensors able to resolve and ignore higher Fourier modes.

Figure 6.10 shows a more quantitative view with just the predicted loading profile $L(\theta)$ when using 4 and 7 sensors, shown against the exact loading profile used. When using 4 or 7 sensor we are only trying to recover the Fourier coefficients $c_{-2}, c_{-1}, c_0, c_1$, and $c_2$. When using 4 sensors there is a reasonably large error because the higher Fourier coefficients, which are ignored, make a substantial contribution to the Stribeck equation shown in Figure 6.10. With only 4 sensors, the ignored higher Fourier coefficients are treated liked an add error (20%), which explains the error in recovering the coefficients $c_{-2}, c_{-1}, c_0, c_1$, and $c_2$. When using 7 sensors, we are able to differentiate the modes associate with $c_{-3}, c_3$ from the lower modes.

### 6.7.4 Localised defect in a roller bearing

As our final example, we consider a a localised defect on, or near, the boundary in contact with the rollers. A schematic is shown on the right of Figure 6.11.

**Defects and slow rotation.** In the previous example we showed how a smooth loading profile can be predicted with only a few sensors by using the prior method, together with a Fourier series expansion of the loading profile, as shown in Section 6.5.1. To predict the loading profile, the rollers need to rotate fast enough, as discussed in Section 6.6.2. If the rollers are rotating slower, then we can only predict the higher Fourier coefficients $c_\ell$ for larger $\ell$, which are associated to localised defects, as we illustrate in this example.

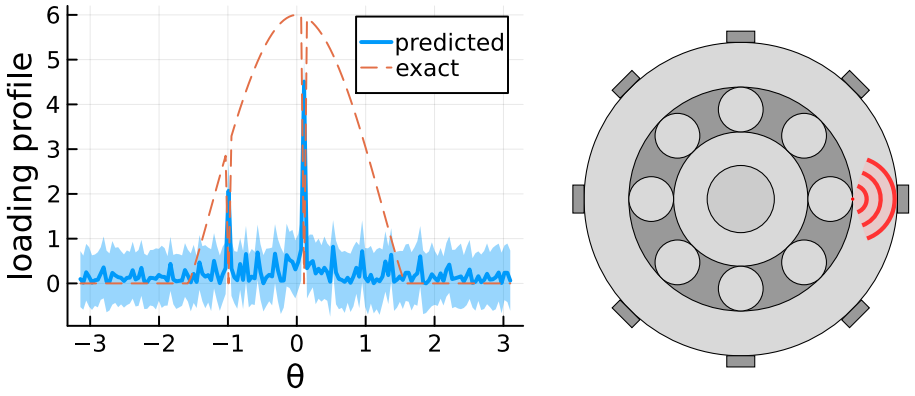



Figure 6.11: The right is an illustration of elastic waves being emitted when a roller hits a defect on the raceway. The left shows the loading profile (orange and dashed) with two sharp drops in pressure due to the presence of two defects on the inner boundary. The blue curves show the absolute value of the predicted loading profile when measuring the Fourier coefficients of the boundary data $\boldsymbol{u}_n$ for $n = -6, -5, \ldots, 6$ where 2% error was added. The rollers are rotating at a rate of $\Omega = 120$ rpm and the properties used for the raceway are shown in Table 6.1.

**Use the inverse modal system.** Consider the inside raceway with two localised defects which we assume leads to a loading profile shown by the orange dashed curve on the left of Figure 6.11. This time, rewriting the loading profile in terms of a Fourier series, as done in Section 6.5.1, does not help because the Fourier series will converge very slowly. So instead, just to illustrate, we just directly solve the system (6.3.9) using the inverse modal system (6.3.4) and boundary data $\boldsymbol{y}^{\text{inv}}$ solely from the outer boundary.

We return to using the parameters in Table 6.1, as these more closely match real applications, but use now a slower rotation speed of the rollers $\Omega = 120$ rpm. As discussed in Section 6.6.1, for slow rotation speeds there are restrictions on which of the Fourier coefficients $c_\ell$, of the loading profile, can be measured.

**What can be measured.** Each frequency $\omega_m$ gives access to a range of values for $\ell$. For this example we use the frequencies $m = 1, 2, \ldots 5$ and for each solve (6.3.9) for $\boldsymbol{a}$, and then predict the inner traction $\boldsymbol{\tau}_1$. From $\boldsymbol{\tau}_1$, and depending on the choice of $m$, we then estimate some of the coefficients $c_\ell$ of the loading profile by using (6.5.5). By combining all

the predicted coefficients $c_\ell$ from all five frequencies $\omega_m$ we then predict the absolute value loading profile shown by the blue curve in Figure 6.11. We have shown the absolute value for visual clarity, as we can more clearly see that the blue and orange spikes match up perfectly.

Clearly, Figure 6.11 shows that we can locate the defects, and determine their magnitude, at least in terms of the pressure difference, by directly solving the inverse problem.

**Envelope analysis.** There is a method commonly used to detect localised defects called *envelope analysis* [85]. Here we only make a brief comment on how this method is connected to the work in this paper.

For the example in this section, we are not able to recover the first modes $c_0$, $c_{-1}$, or $c_1$ of the loading profile. However, these modes are not small, and therefore do make a contribution to our boundary data. They in fact act like an error term. For example, for the frequency $\omega_m$, the coefficient $c_0$ contributes to $f_{mZ}(\omega_m)$ of the boundary data as shown in (6.5.5). If we have less than $m \times Z$ sensors, and attempt to calculate a Fourier series of the loading data $y(\theta, \omega_m)$, then the mode of $f_{mZ}(\omega_m)$ will be mixed in with the other modes $f_n(\omega_m)$, for $|n| < m \times Z$, and lead to errors for these modes. This error could be avoided if the function $y(\theta, \omega_m)$ was first smoothed in $\theta$ before calculating the Fourier modes, as the smoothening would remove the higher modes such as $f_{mZ}(\omega_m)$. We believe this can be linked with the smoothening in time used in Envelope Analysis, though this deserves a more lengthy analysis elsewhere.

## 6.8 Conclusions

In this paper we have shown how to model elastic waves confined within a hollow thick walled cylinder with symmetry along the axis. As discussed in Section 6.2, the dynamics of these waves captures the dominant vibrations within a raceway. By deriving simple systems for the modes, we provided tools to better understand and quickly solve for these elastic waves. A detailed outline of the paper's content is given at the end of Section 6.1.

**Results.** The main results are how to: 1) model waves, 2) use prior assumptions about the boundary conditions, and 3) determine what it is, and is not, possible to predict the traction or displacement within the raceway. Notably, in Section 6.6, we show that solving for the elastic waves becomes ill-posed when hitting a resonant frequency or near the diffraction limit. These results hold for any transfer path of the signal. For roller bearings we demonstrate that if the rotation speed is slow, then it is only possible to predict localised contact forces. Extended, or smooth, contact forces lead to ill-posed problems for elastic waves.

**Modelling - raceways.** Our models lay the foundation for many future avenues. For instance, instead of considering waves which are just confined in the raceway, as shown by Figure 6.7, the boundary conditions can be adjusted to let waves leak out towards the rollers or the oil. Additionally, the raceway's bolted supports can be incorporated in the boundary conditions by assuming that waves dissipate through these bolts and do not return.

**Modelling - bearings.** It is also possible to extend the models to consider bearings that do not have axial symmetry, such as ball bearings or steeply inclined tapered roller bearings. Further extensions could account for the slip and slide of roller bearings [13] which lead to transient waves with a high-frequency content. Incorporating these phenomena into the models as more elaborate priors would lead to more accurate predictions.

**Inverse problem - uncertainty.** To develop robust predictions, and methods to determine defects, uncertainty needs to be accounted for [71]. A first step in this direction is to consider the boundary data to be samples of a distribution, and also to consider that priors, as discussed in Section 6.4.1, are also distributions. This would help properly account for roller bearings slipping, or fluid interaction in journal bearings.

**Detect localised defects.** In industrial applications, there are only a few sensors per bearing, so defect detection needs to relies on monitoring the amplitude of specific frequencies, such as the ball pass frequencies. Can we rely on these methods? The models in this paper provide a clear path to address this question: imagine continually measuring the vibration of a bearing. When a change occurs, we can assume it is due to a localised defect with an unknown position and size. By adopting a Bayesian approach, we can then estimate the defect size by marginalising over its possible positions. This approach would clarify how robust current methods are and lead to more reliable physics-based diagnostics.

# Chapter 7

# Source localisation in roller bearing raceways

Jessica J. Kent, Art L. Gower

## Abstract

Early detection of subsurface cracks is essential for the maintenance of roller bearings. Many of the methods presently used are time-domain-based methods; this leads to methods that are highly dependent upon correctly identifying first arrival times, something which is not always possible. Further, current methodologies, by and large, are based on acoustics; this ignores the fact that we are dealing with an elastic medium, and therefore, information is lost. In this paper, we propose a frequency-domain method for the localisation of bearing faults from first-principles mathematical modelling.

## 7.1 Introduction

In the condition monitoring of rolling element bearings, the early detection of subsurface cracks induced by roller contact fatigue remains an important problem. [86] This is because subsurface cracks, once they propagate to the surface, can turn into spalls, leading to debris flaking from the rolling surfaces [11, 86, 87]; this is illustrated in Figure 7.1.

The initiation of subsurface cracks produces an impulse response in the bearing raceway, which can be detected via acoustic emission monitoring methods.[12] In particular, the methods that are applied in acoustic emission monitoring of rolling element bearings are directed towards detection rather than localisation.

Current methods for the localisation of subsurface faults are predominantly time-domain based, this leads to many difficulties, in particular, the detection of shear wave emissions pose a particular problem, as in thin structures the arrival of a shear wave can be masked in

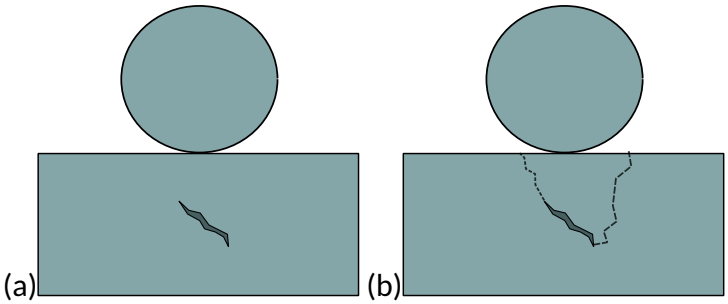



Figure 7.1: (a) shows a subsurface crack; (b) shows the crack propagating to the raceway surface, leading to spalling. Adapted from [11].

a pressure wave signal; this problem can be solved by using a higher sample rate however, this comes at the cost of data storage and efficiency. [72]

The method we propose is a frequency domain approach based on a Green's function representation of the emission signals. This work is an extension of the method proposed in [88]; we begin by extending the theory in [88] to incorporate a forcing term which enables us to model subsurface cracks, we then move to the problem of localising a single crack, we show that our method can localise a single crack given that first arrival times are known; finally, we set out theory which could be further developed to localise multiple sources.

## 7.2 Modelling raceway cracks

In [88], we introduced a method for detecting and diagnosing defects in a bearing raceway. This method was based on modelling the bearing raceway as a hollow steel cylinder. Before we proceed, we find it pertinent to summarise the key aspects of the model.

We begin by modelling the bearing raceway as a hollow cylinder made of a homogeneous, isotropic material, with inner radius $r_1$ and outer radius $r_2$. We define the domain, $\Omega$, of our solution as

$$\Omega := \{(r, \theta)^{\mathrm{T}} : r \in [r_1, r_2], \theta \in [0, 2\pi)\}. \tag{7.2.1}$$

This domain is illustrated in Figure 5.1 (a).

With the domain $\Omega$ defined, we apply the following Helmholtz decomposition to the Navier equation [42]:

$$\boldsymbol{u} = \nabla\phi + \nabla \times (\psi\hat{\boldsymbol{z}}) \quad \text{and} \quad \nabla \cdot (\psi\hat{\boldsymbol{z}}) = 0 \tag{7.2.2}$$

this leads to the following equations for $\phi$ and $\psi$:

$$\nabla^2\phi + k_p^2\phi = 0 \quad \text{and} \quad \nabla^2\psi + k_s^2\psi = 0. \tag{7.2.3}$$

Solving the system in Equation (7.2.3), we found that we could represent $\phi$ and $\psi$ as a

superposition of regular and outgoing waves, i.e.

$$\phi = \sum_{n=-\infty}^{\infty} \left(a_n \mathrm{J}_n(k_p r) + b_n \mathrm{H}_n^{(1)}(k_p r)\right) \mathrm{e}^{\mathrm{i}n\theta} \tag{7.2.4}$$

$$\psi = \sum_{n=-\infty}^{\infty} \left(c_n \mathrm{J}_n(k_s r) + d_n \mathrm{H}_n^{(1)}(k_s r)\right) \mathrm{e}^{\mathrm{i}n\theta}. \tag{7.2.5}$$

We then proceeded to define two main boundary value problems that were of interest: the problem of prescribing traction on both boundaries and the problem of prescribing traction and displacement on a single boundary; we termed these the forward and inverse problems, respectively. In particular, we found that these boundary value problems could be represented as a system of matrix equations for each mode $n$:

$$\mathbf{M}_n^{\text{for}} \boldsymbol{a}_n = \boldsymbol{\tau}_n \tag{7.2.6}$$

$$\mathbf{M}_n^{\text{inv}} \boldsymbol{a}_n = \boldsymbol{u}_n \tag{7.2.7}$$

Where $\mathbf{M}_n^{\text{for}}$ is the matrix corresponding to the forward problem and $\mathbf{M}_n^{\text{inv}}$ is the matrix corresponding to the inverse problem. Having recounted these details, we now move to the main content of this section: modelling subsurface cracks.

### 7.2.1 A single crack

Suppose we have a single crack located at position $\boldsymbol{r}'$. For simplicity, in this section, we only consider the case of a single P-wave source; in Section 7.4 we shall consider the more general problem of multiple sources emitting P and S-waves. In this case, the governing equations in Equation (7.2.3) become;

$$\nabla^2 \phi + k_p^2 \phi = \alpha \delta(\boldsymbol{r} - \boldsymbol{r}') \quad \text{and} \quad \nabla^2 \psi + k_s^2 \psi = 0. \tag{7.2.8}$$

We assume that the crack radiates a P-wave in all directions; as such, we require a particular solution which satisfies the Sommerfeld radiation condition [46]:

$$\lim_{r \to \infty} \sqrt{r} \left( \frac{\partial}{\partial r} - \mathrm{i}k_p \right) \phi(r, \theta) = 0 \tag{7.2.9}$$

We can find the appropriate particular solution to Equation (7.2.8) satisfying Equation (7.2.9) in [46], and the form of the homogeneous solution can be found in [88]. Putting these

together, we obtain the following potentials:

$$\phi = \sum_{n=-\infty}^{\infty} \left(a_n \mathrm{J}_n(k_p r) + b_n \mathrm{H}_n^{(1)}(k_p r)\right) \mathrm{e}^{\mathrm{i}n\theta} + \frac{\mathrm{i}\alpha}{4} \mathrm{H}_0^{(1)}(k_p|\boldsymbol{r} - \boldsymbol{r}'|), \tag{7.2.10}$$

$$\psi = \sum_{n=-\infty}^{\infty} \left(c_n \mathrm{J}_n(k_s r) + d_n \mathrm{H}_n^{(1)}(k_s r)\right) \mathrm{e}^{\mathrm{i}n\theta}. \tag{7.2.11}$$

The homogeneous part of the solution in Equation (7.2.10) represents the reflected P-waves that are scattered from the boundaries, and Equation (7.2.11) represents the reflected S-waves from mode conversion. [42]

In Equation (7.2.10), the source term, represented by $\mathrm{H}_0^{(1)}(k_p|\boldsymbol{r} - \boldsymbol{r}'|)$, is written in a coordinate system centred at the source location, $\boldsymbol{r}'$, whereas the homogeneous solution is written in a coordinate system centred at the origin of the raceway. To make use of the boundary conditions on the raceway, we need to rewrite Equation (7.2.10) so that it is centred at the origin of the raceway. To do this, we must utilise Graf's addition theorem [89]:

$$\mathrm{H}_0^{(1)}(k_p|\boldsymbol{r} - \boldsymbol{r}'|) = \sum_{n=-\infty}^{\infty} (-1)^n \mathrm{H}_n^{(1)}(k_p r') \mathrm{J}_n(k_p r) \mathrm{e}^{\mathrm{i}n(\theta - \theta')} \quad \text{for} \quad r < r', \tag{7.2.12}$$

$$\mathrm{H}_0^{(1)}(k_p|\boldsymbol{r} - \boldsymbol{r}'|) = \sum_{n=-\infty}^{\infty} (-1)^n \mathrm{J}_n(k_p r') \mathrm{H}_n^{(1)}(k_p r) \mathrm{e}^{\mathrm{i}n(\theta - \theta')} \quad \text{for} \quad r > r', \tag{7.2.13}$$

where $r'$ is the distance of the source from the origin and $\theta'$ is the angular position of the source from the $x$ axis.

Using Equations (7.2.12) and (7.2.13) in Equation (7.2.10) leads to the following representations for $\phi$:

$$\phi(r, \theta) = \frac{\mathrm{i}\alpha}{4} \sum_{n=-\infty}^{\infty} (-1)^n \mathrm{H}_n^{(1)}(k_p r') \mathrm{J}_n(k_p r) \mathrm{e}^{\mathrm{i}n(\theta - \theta')} + \phi_0 \quad \text{for} \quad r < r', \tag{7.2.14}$$

$$\phi(r, \theta) = \frac{\mathrm{i}\alpha}{4} \sum_{n=-\infty}^{\infty} (-1)^n \mathrm{J}_n(k_p r') \mathrm{H}_n^{(1)}(k_p r) \mathrm{e}^{\mathrm{i}n(\theta - \theta')} + \phi_0 \quad \text{for} \quad r > r', \tag{7.2.15}$$

where $\phi_0 = \sum_{n=-\infty}^{\infty} \left(a_n \mathrm{J}_n(k_p r) + b_n \mathrm{H}_n^{(1)}(k_p r)\right) \mathrm{e}^{\mathrm{i}n\theta}$.

Since we're only interested in sources inside the bearing raceway, we find that, when measuring from the boundary at $r_1$, the solution takes the form of Equation (7.2.14) and at $r_2$ it takes the form of Equation (7.2.15). Thus, using Equation (7.2.14), Equation (7.2.15),

and Equation (7.2.11) in the equations for stress and displacement in [88] leads to a matrix equation for each mode as before, but now with an additional source term;

$$\mathbf{M}_n^{\text{for}}\boldsymbol{a}_n + \mathbf{T}_n\boldsymbol{a}_n^{\text{src}}(\boldsymbol{r}') = \boldsymbol{\tau}_n, \tag{7.2.16}$$

$$\mathbf{M}_n^{\text{inv}}\boldsymbol{a}_n + \mathbf{U}_n\boldsymbol{a}_n^{\text{src}}(\boldsymbol{r}') = \boldsymbol{u}_n. \tag{7.2.17}$$

where

$$\boldsymbol{a}_n^{\text{src}}(\boldsymbol{r}') = \frac{(-1)^n \mathrm{i}\alpha}{4}\left[\mathrm{H}_n^{(1)}(k_p r')\mathrm{e}^{-\mathrm{i}n\theta'}, \mathrm{J}_n(k_p r')\mathrm{e}^{-\mathrm{i}n\theta'}, 0, 0\right]^{\mathrm{T}}.$$

### 7.2.2 Localising a single source

The problem of locating a single source is made somewhat difficult by the fact that the source fields have different representations depending upon our distance from the source. This difficulty can be readily demonstrated by inspecting Equation (7.2.17) more closely.

In Equation (7.2.17), we encounter the term $\mathbf{U}_n\boldsymbol{a}_n^{\text{src}}$; let us arrange the vector $\boldsymbol{a}_n^{\text{src}}$ so that information on $r = r_1$ is found in the first two entries and data on $r = r_2$ is in the last two entries. This leads to the following block matrix representation;

$$\begin{bmatrix} \mathbf{0} & \mathbf{U}_1 \\ \mathbf{0} & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{n,r_1}^{\text{src}} \\ \boldsymbol{a}_{n,r_2}^{\text{src}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_{n,r_2} \\ \boldsymbol{\tau}_{n,r_2} \end{bmatrix}.$$

Clearly, this system is not invertible; further, without knowing the source locations, calculating the traction condition is not possible. This poses a problem, as information pertaining to the source location and amplitude is contained in $\boldsymbol{a}_n^{\text{src}}$. To solve this problem, we must restate the problem to incorporate data from both boundaries. A simple strategy is to parametrise $\boldsymbol{a}_n$ in terms of $\boldsymbol{a}_n^{\text{src}}$.

From Equation (7.2.16) we obtain

$$\boldsymbol{a}_n = (\mathbf{M}_n^{\text{for}})^{-1}\left(\boldsymbol{\tau}_n - \mathbf{T}_n\boldsymbol{a}_n^{\text{src}}(\boldsymbol{r}')\right), \tag{7.2.18}$$

which substituted in Equation (7.2.17) leads to

$$\left(\mathbf{U}_n - \mathbf{M}_n^{\text{inv}}(\mathbf{M}_n^{\text{for}})^{-1}\mathbf{T}_n\right)\boldsymbol{a}_n^{\text{src}}(\boldsymbol{r}') = \boldsymbol{u}_n - \mathbf{M}_n^{\text{inv}}(\mathbf{M}_n^{\text{for}})^{-1}\boldsymbol{\tau}_n. \tag{7.2.19}$$

For simplicity, we assume that the boundaries are traction-free, i.e., $\boldsymbol{\tau}_n = 0$, for all $n \in \mathbb{Z}$. Thus,

$$\left(\mathbf{U}_n - \mathbf{M}_n^{\text{inv}}(\mathbf{M}_n^{\text{for}})^{-1}\mathbf{T}_n\right)\boldsymbol{a}_n^{\text{src}}(\boldsymbol{r}') = \boldsymbol{u}_n. \tag{7.2.20}$$

The simplest way to find $\boldsymbol{r}'$ that solves the above is to simply try them all. That is, for each $\boldsymbol{r}'$

we can check the residual of Equation (7.2.20) given by

$$e_n(\boldsymbol{r}') = \left\| \left( \mathbf{U}_n - \mathbf{M}_n^{\text{inv}} (\mathbf{M}_n^{\text{for}})^{-1} \mathbf{T}_n \right) \boldsymbol{a}_n^{\text{src}}(\boldsymbol{r}') - \boldsymbol{u}_n \right\|^2, \tag{7.2.21}$$

then, in abstract terms, we can obtain the true location of the source $\hat{\boldsymbol{r}}'$ from

$$\hat{\boldsymbol{r}}' = \arg\min_{\boldsymbol{r}' \in \Omega} \sum_n \int e_n(\boldsymbol{r}') / \|\boldsymbol{u}_n\|^2 \mathrm{d}\omega. \tag{7.2.22}$$

With this solution to the inverse problem in place, we are now in a position to present numerical simulations which examine the sensitivity of the solution in Equation (7.2.22) to a range of errors.

## 7.3 Numerical Simulations

In this section, we present the results of numerical simulations. The parameters used in the simulations can be found in Table 7.1. For these simulations, we use the following procedure to identify the source location:

1. Discretise the domain, $\Omega$, as $\Omega_N = \{\boldsymbol{r}' \in \Omega : \boldsymbol{r}' \in \Omega \cap \mathcal{G}_N \times \mathcal{G}_N\}$, where $\mathcal{G}_N := \{-r_2, -r_2 + \frac{2r_2}{N}, ..., -r_2 + \frac{2(N-1)r_2}{N}, r_2\}$.

2. Calculate $e_n(\boldsymbol{r}')$ for each mode $n$, frequency $\omega$ and $\boldsymbol{r}' \in \Omega_N$.

3. Compute $\sum_n \sum_\omega e_n(\boldsymbol{r}') / \|\boldsymbol{u}_n\|^2 \Delta\omega$.

4. Identify the $\boldsymbol{r}' \in \Omega_N$ that gives the maximal value of

$$f(\boldsymbol{r}') := \frac{1}{\sum_n \sum_\omega e_n(\boldsymbol{r}') / \|\boldsymbol{u}_n\|^2 \Delta\omega}.$$

   This is identified as the approximate source location.

We begin this section by analysing the identification of the source location for a small number of sensors; in particular, we examine the sensitivity of the method outlined in Section 7.2.2 to measurement errors. We find that, if the amplitude of the source, $\alpha$, is known, then the source location can be recovered to a reasonable degree of accuracy with very few sensors. However, when $\alpha$ is not known, recovering the source location becomes quite difficult. In particular, in Section 7.3.2 we find that incorrectly identifying the *phase* of $\alpha$ is a significant source of error.

| Parameter | value | description |
|---|---|---|
| $r_1$ | $1.0$ m | inner radius |
| $r_2$ | $1.5$ m | outer radius |
| $c_p$ | $5000$ m/s | pressure speed |
| $c_s$ | $3500$ m/s | shear speed |
| $\rho$ | $7800$ kg/m$^3$ | mass density |
| $N$ | $100$ | resolution |

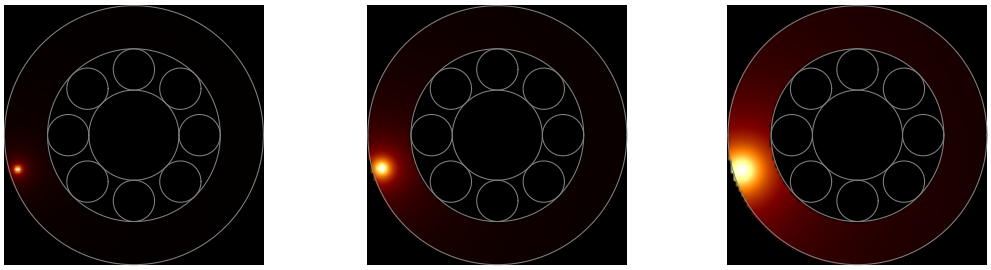
Table 7.1: The parameter values that approximate a steel raceway. These parameters are used for all numerical simulations.

### 7.3.1 Small number of sensors

Let us assume that we only know the displacement boundary conditions at the points

$$\mathcal{D} = \{(r_2, \theta_j) \quad \text{for} \quad j = 1, 2, \ldots, J\}.$$

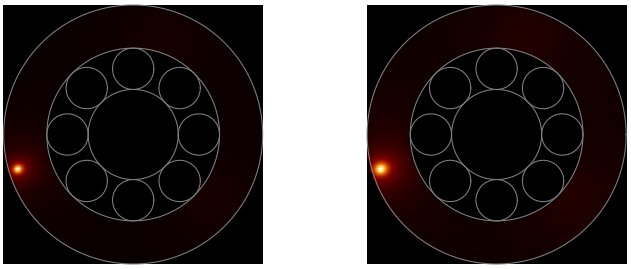
If there are only a small number of sensors, say $J = 1$ to $3$, then we would only know the displacement $\boldsymbol{u}$ at the points $\mathcal{D}$. In which case, we could only calculate the Fourier modes $\boldsymbol{u}_n$ for $n = -1, 0, 1$. As can be seen in Figure 7.2, we can obtain a good degree of accuracy in identifying the source location with this small number of sensors. In this section, we make the assumption that the amplitude $\alpha$ is known.



(a) (b) (c)

Figure 7.2: Here we show plots of $f(\boldsymbol{r}')$ for (a) 1% Gaussian noise; (b) 5% Gaussian noise; (c) 10% Gaussian noise. We note that in all three cases we can discern an approximate source location, though in figures (b) and (c), we note that the region of interest is wider and more diffuse than in (a), where the level of noise is relatively low.

In the cases of Figure 7.2 (b) and (c), we further note that the influence of higher levels of noise can be combatted with the addition of more sensors; this can be seen in Figure 7.3 where we have plotted $f(\boldsymbol{r}')$ for 11 sensors with 5 and 10 % Gaussian noise. In summary, if the amplitude of the source is known, then the method set out in Section 7.2.2 can localise the source to a good degree of accuracy with as few as 3 sensors. When dealing with higher levels of noise, the addition of more sensors helps to counter the smearing effect seen in Figure 7.2 (b) and (c). This is in good agreement with the most basic methods used to localise sources, e.g., triangulation. [72] An important assumption we have made in this section is



(a) (b)

Figure 7.3: Here we show plots of $f(\boldsymbol{r}')$ for (a) 5% Gaussian noise and (b) 10% Gaussian noise. We note that the recovery of the source location in (a) and (b) for 11 sensors is far clearer than in Figure 7.2 (b) and (c), where the region of interest is far wider.

that the amplitude of the source is $1$; this corresponds to the case where the amplitude is known and can be scaled out of the problem. However, as we shall see in the next section, the impact of amplitude on source localisation with this method is far greater than the impact of measurement error; in particular, the amplitude of the source, $\alpha$, contains information on the time at which the crack initiated. In the next section, we will show how a lack of knowledge of the amplitude phase can be devastating for locating sources.

### 7.3.2 Amplitude and source localisation

In the simulations we have presented thus far, we have assumed a unit amplitude. In this section, we examine the effect of the source amplitude, $\alpha$, on localisation. In particular, we investigate the impact of the phase of $\alpha$ on the error.

Incorrectly identifying the phase of $\alpha$ corresponds to incorrectly identifying the initiation time of the crack, or, what amounts to the same, the first arrival time at each sensor. To this end, we define a set of initiation times

$$T_0 := \{10^{-m} : m \in \{1, ..., 10\}\}. \tag{7.3.1}$$

From this, we define a set of amplitudes for each frequency:

$$A(\omega) := \{\mathrm{e}^{-\mathrm{i}\omega t_0} : t_0 \in T_0\} \tag{7.3.2}$$

we then apply the method described in Section 7.2.2 as before, with the assumption that $\alpha = 1$. From this, we are able to discern a clear relationship between the order of magnitude, $m$, of the initiation time and the error in the recovery of the source location. In Figure 7.4 we show results for $m = 2, 3, 4, 5$, when using only 3 sensors:

From Figure 7.4, we can see that errors in identifying the initiation time of a crack lead to errors in the identification of $\boldsymbol{r}'$. In Figure 7.5 we plot $\|\boldsymbol{r}' - \hat{\boldsymbol{r}}'\|$ for each $m$. We can identify a clear relationship between the order of magnitude of the error in initiation time $t_0$ and the

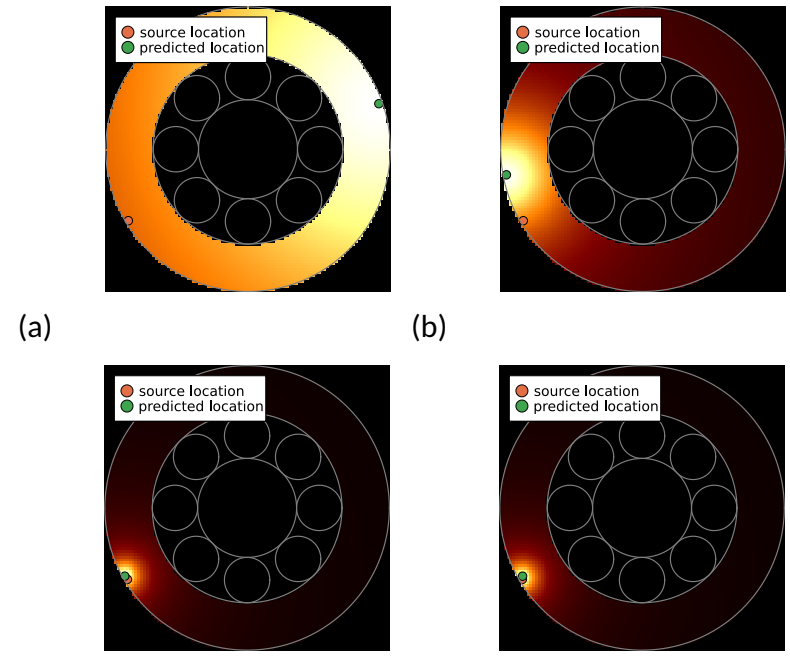



Figure 7.4: Here we show plots of $f(\boldsymbol{r}')$ for (a) $t_0 = 10^{-2}$; (b) $t_0 = 10^{-3}$; (c) $t_0 = 10^{-4}$; (d) $t_0 = 10^{-5}$.

error in predicting the location of the source $\boldsymbol{r}'$:

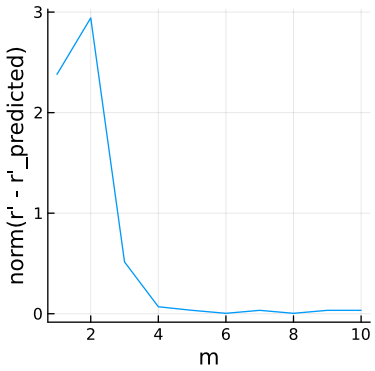
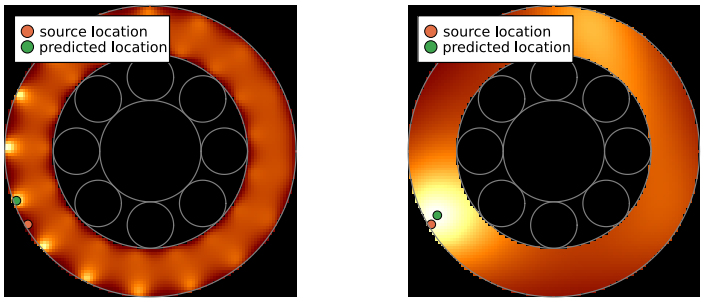



Figure 7.5: Here we plot the error in the recovered source location $\hat{\boldsymbol{r}}'$ vs the exponent of arrival times of the form $t_0 = 10^{-m}$, we see that as $t_0 \to 0$, the error decreases dramatically. In particular, for $m \geq 4$, the source can be localised to a good degree of precision with just 3 sensors.

The addition of more sensors can help to alleviate the error in locating the source; but importantly, it does not remove this error entirely as can be seen in Figure 7.6; this suggests that correctly identifying the phase of $\alpha$ for each frequency is vital to the process of localising a source; this aligns well with the results of methods based on the measurement of arrival time [72].




(a) (b)

Figure 7.6: Here we show plots of $f(\boldsymbol{r}')$ for (a) $t_0 = 10^{-2}$; (b) $t_0 = 10^{-3}$. These results are achieved with 67 and 9 sensors, respectively.

## 7.4 Multiple cracks; theoretical results and future prospects

In Sections 7.2 and 7.3, we tackled the problem of localising a single P-wave source. In this section, we consider the more general problem of recovering a map of P and S-wave sources inside a bearing raceway. To do this, we introduce the density maps $\Pi_p$ and $\Pi_s$; these maps tell us the location and amplitude of P- and S-wave sources in the raceway, respectively.

We begin this section by setting out the equations which will allow us to model the field in the raceway due to an arbitrary number of sources; we then move to presenting the theoretical solution of the problem of recovering the density maps, we will find that this solution only works when the fields $\phi$ and $\psi$ are fully specified; we then conclude this section by indicating directions for future research which will enable us to solve this problem when limited information in is available.

### 7.4.1 Crack density

Rather than attempting to locate one or two cracks, here we turn to a different approach: we assume that elastic waves are emitted from all locations and instead look to solve for the amplitude of these emissions, which is given by a density map.

Let $\Pi_p(\boldsymbol{r}, t)$ and $\Pi_s(\boldsymbol{r}, t)$ be the emission maps for the pressure and shear potentials. The field inside the bearing can thus be described by two Helmholtz potentials $\phi$ and $\psi$ satisfying

$$\nabla^2\phi + k_p^2\phi = \tilde{\Pi}_p(\boldsymbol{r}, \omega) \quad \text{and} \quad \nabla^2\psi + k_s^2\psi = \tilde{\Pi}_s(\boldsymbol{r}, \omega), \tag{7.4.1}$$

where $\tilde{\Pi}_p(\boldsymbol{r}, \omega) = \mathcal{F}\{\Pi_p(\boldsymbol{r}, t)\}$ and $\tilde{\Pi}_s(\boldsymbol{r}, \omega) = \mathcal{F}\{\Pi_s(\boldsymbol{r}, t)\}$. The solutions to Equation (7.4.1) are given by

$$\phi = \phi_0 + \phi_{\text{src}}, \quad \psi = \psi_0 + \psi_{\text{src}}, \tag{7.4.2}$$

where

$$\phi_{\mathrm{src}} = \frac{\mathrm{i}}{4}\int_{\Omega} \mathrm{H}_0^{(1)}(k_p|\boldsymbol{r}-\boldsymbol{r}'|)\tilde{\Pi}_p(\boldsymbol{r}',\omega)\mathrm{d}\boldsymbol{r}', \quad \psi_{\mathrm{src}} = \frac{\mathrm{i}}{4}\int_{\Omega} \mathrm{H}_0^{(1)}(k_s|\boldsymbol{r}-\boldsymbol{r}'|)\tilde{\Pi}_s(\boldsymbol{r}',\omega)\mathrm{d}\boldsymbol{r}'. \tag{7.4.3}$$

Suppose we have the following boundary value problem

$$\boldsymbol{\tau}(r_1,\theta) = \boldsymbol{\tau}^1 \quad \text{and} \quad \boldsymbol{\tau}(r_2,\theta) = \boldsymbol{\tau}^2 \tag{7.4.4}$$

where $\boldsymbol{\tau}^1$ and $\boldsymbol{\tau}^2$ are given. The traction written in terms of the potentials [88] is given by:

$$\sigma_{rr} = \lambda\nabla^2\phi + 2\mu\left(\frac{\partial^2\phi}{\partial r^2} + \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial\psi}{\partial\theta}\right)\right), \quad \sigma_{r\theta} = \mu\nabla^2\psi - 2\mu\left(\frac{\partial^2\psi}{\partial r^2} - \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial\phi}{\partial\theta}\right)\right). \tag{7.4.5}$$

Using (7.4.2) we find that Equation (7.4.4) may be rewritten as

$$\boldsymbol{\tau}_0(r_1,\theta) + \boldsymbol{\tau}_{\mathrm{src}}(r_1,\theta) = \boldsymbol{\tau}^1 \quad \text{and} \quad \boldsymbol{\tau}_0(r_2,\theta) + \boldsymbol{\tau}_{\mathrm{src}}(r_2,\theta) = \boldsymbol{\tau}^2, \tag{7.4.6}$$

where $\boldsymbol{\tau}_0$ is the contribution to the traction from $\phi_0$ and $\psi_0$ and $\boldsymbol{\tau}_{\mathrm{src}}$ is the contribution to the traction from $\phi_{\mathrm{src}}$ and $\psi_{\mathrm{src}}$. Hence, we find that

$$\sigma_{rr}^{\mathrm{src}} = \lambda\tilde{\Pi}_p(\boldsymbol{r},\omega) - \lambda k_p^2\phi_{\mathrm{src}} + 2\mu\left(\frac{\partial^2\phi_{\mathrm{src}}}{\partial r^2} + \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial\psi_{\mathrm{src}}}{\partial\theta}\right)\right), \tag{7.4.7}$$

$$\sigma_{r\theta}^{\mathrm{src}} = \mu\tilde{\Pi}_s(\boldsymbol{r},\omega) - \mu k_s^2\psi_{\mathrm{src}} - 2\mu\left(\frac{\partial^2\psi_{\mathrm{src}}}{\partial r^2} - \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial\phi_{\mathrm{src}}}{\partial\theta}\right)\right). \tag{7.4.8}$$

Here we note the presence of the density maps $\tilde{\Pi}_p$ and $\tilde{\Pi}_s$ in the equations for stress. Since we assume that $\tilde{\Pi}_p$ and $\tilde{\Pi}_s$ vanish at the boundaries when we come to consider the boundary conditions, these terms can be dropped.

For the remaining sections we shall assume that the maps $\tilde{\Pi}_p$ and $\tilde{\Pi}_s$ take the form:

$$\tilde{\Pi}_p(\boldsymbol{r},\omega) = \sum_{j=1}^{J_p} \alpha_j\delta(\boldsymbol{r}-\boldsymbol{r}_j)\mathrm{e}^{-\mathrm{i}\omega t_j} \tag{7.4.9}$$

$$\tilde{\Pi}_s(\boldsymbol{r},\omega) = \sum_{k=1}^{J_s} \beta_k\delta(\boldsymbol{r}-\boldsymbol{r}_k)\mathrm{e}^{-\mathrm{i}\omega t_k} \tag{7.4.10}$$

where $\alpha_j$ and $\beta_k$ are the amplitudes for P and S sources, respectively, $t_j$ and $t_k$ are the initiation times for the P and S sources, respectively, and $J_p$ and $J_s$ are the number of P and S sources, respectively.

Using Equations (7.4.9) and (7.4.10) and applying Graf's addition theorem to $\phi_{\mathrm{src}}$ and $\psi_{\mathrm{src}}$

leads us to the following system of equations for the forward problem:

$$\mathbf{M}_n^{\text{for}}\boldsymbol{a}_n + \sum_{j=1}^{J}\mathbf{T}_n\boldsymbol{a}_n^{\text{src}}(\boldsymbol{r}_j) = \boldsymbol{\tau}_n \tag{7.4.11}$$

and the following for the inverse problem:

$$\mathbf{M}_n^{\text{inv}}\boldsymbol{a}_n + \sum_{j=1}^{J}\mathbf{U}_n\boldsymbol{a}_n^{\text{src}}(\boldsymbol{r}_j) = \boldsymbol{u}_n \tag{7.4.12}$$

where $J = J_p + J_s$ and

$$\boldsymbol{a}_n^{\text{src}}(\boldsymbol{r}_i) = \frac{(-1)^n\mathrm{i}}{4}\left[\alpha_i\mathrm{H}_n^{(1)}(k_p r_i)\mathrm{e}^{-\mathrm{i}n\theta_i}, \alpha_i\mathrm{J}_n(k_p r_i)\mathrm{e}^{-\mathrm{i}n\theta_i}, \beta_i\mathrm{H}_n^{(1)}(k_s r_i)\mathrm{e}^{-\mathrm{i}n\theta_i}, \beta_i\mathrm{J}_n(k_s r_i)\mathrm{e}^{-\mathrm{i}n\theta_i}\right]^{\mathrm{T}}.$$

### 7.4.2 Recovering the density maps via deconvolution

We wish to solve for the densities $\Pi_p$ and $\Pi_s$, so we have two independent integral equations to solve. Theoretically, this task is quite straightforward; the potentials in Equation (7.4.3) are convolution integrals, with the convolution kernel;

$$K(\boldsymbol{r}) = \frac{\mathrm{i}}{4}\mathrm{H}_0^{(1)}(k|\boldsymbol{r}|),$$

thus we can exploit the convolution theorem to recover the density maps; applying the 2D Fourier transform to the kernel gives;

$$\tilde{K}(\boldsymbol{k}, \boldsymbol{k}') = \frac{1}{k^2 - k'^2 + \mathrm{i}\epsilon},$$

where the $\mathrm{i}\epsilon$ term is added to deal with the pole at $k = k'$ when applying the inverse transform. Using the convolution theorem and the above in Equation (7.4.3) leads to the following;

$$\tilde{\phi} - \tilde{\phi}_0 = \frac{\mathcal{F}_{\text{2D}}\{\tilde{\Pi}_p(\boldsymbol{r}', \omega)\}}{k_p^2 - k_p'^2 + \mathrm{i}\epsilon}, \quad \tilde{\psi} - \tilde{\psi}_0 = \frac{\mathcal{F}_{\text{2D}}\{\tilde{\Pi}_s(\boldsymbol{r}', \omega)\}}{k_s^2 - k_s'^2 + \mathrm{i}\epsilon}. \tag{7.4.13}$$

where $\mathcal{F}_{\text{2D}}$ is the 2D spatial Fourier transform.

Hence, the solution to the inverse problem will be

$$\Pi_p = \mathcal{F}^{-1}\left\{\lim_{\epsilon\to 0}\mathcal{F}_{\text{2D}}^{-1}\left\{(k_p^2 - k_p'^2 + \mathrm{i}\epsilon)(\tilde{\phi} - \tilde{\phi}_0)\right\}\right\} \tag{7.4.14}$$

$$\Pi_s = \mathcal{F}^{-1}\left\{\lim_{\epsilon\to 0}\mathcal{F}_{\text{2D}}^{-1}\left\{(k_s^2 - k_s'^2 + \mathrm{i}\epsilon)(\tilde{\psi} - \tilde{\psi}_0)\right\}\right\} \tag{7.4.15}$$

where $\mathcal{F}$ is the 1D time-to-frequency Fourier transform. An example using the solution Equation (7.4.14) to localise P-wave sources in a box domain can be found in Figure 7.7.

This solution, however, poses significant difficulties in practical implementation; primarily,

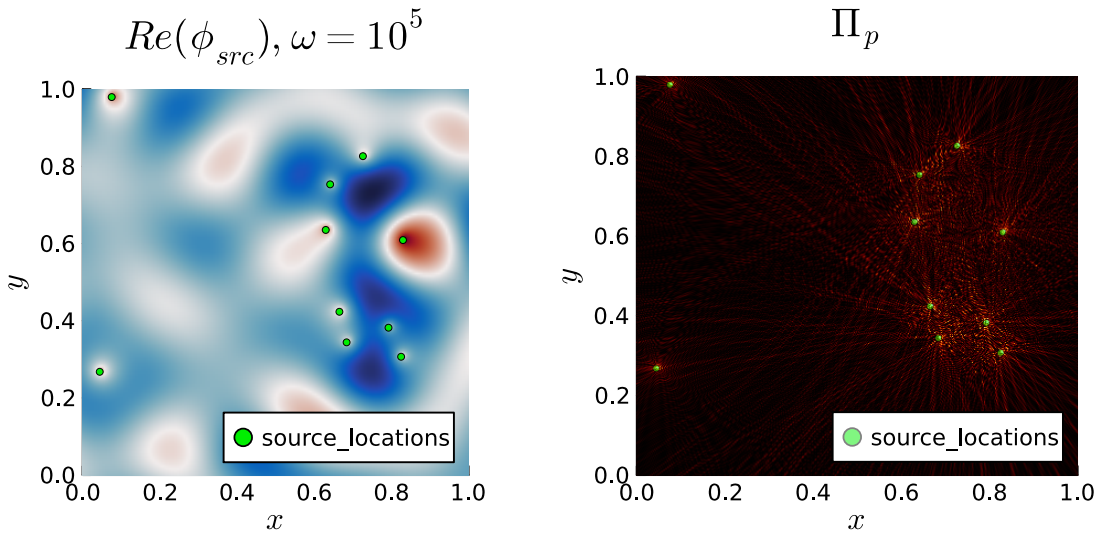



Figure 7.7: Left: The field $\phi_{\text{src}}$, defined on the domain $[0,1] \times [0,1]$, due to a random distribution of 10 point sources each with amplitude 1. Right: the recovered density map using the deconvolution solution in Equation (7.4.14).

it assumes that $\phi_{\text{src}}$ is known over the whole domain $\Omega$; this is not the case, as we only have access to the reflected waves inside the domain and the full field on the boundary $\partial\Omega_2$. This means that we must either find a means to reconstruct the full field inside the bearing or attack the problem from a different angle.

### 7.4.3 Future prospects

At this juncture, we take the opportunity to point out future directions that may be taken based on the results of this section. It is clear that the theoretical solution in Section 7.4.2 requires us to know the fields $\phi_{\text{src}}$ and $\psi_{\text{src}}$ everywhere in $\Omega$. As such, to make use of this solution, we would need to first reconstruct the fields $\phi_{\text{src}}$ and $\psi_{\text{src}}$.

This could be achieved by using Equation (7.4.2) with Equation (7.4.11) to rewrite the reflected waves $\phi_0$ and $\psi_0$ in terms of the density maps $\tilde{\Pi}_p$ and $\tilde{\Pi}_s$:

$$\boldsymbol{a}_n = \frac{(-1)^{n+1}\mathrm{i}}{4} \left(\mathbf{M}_n^{\text{for}}\right)^{-1} \int_{\Omega} \mathbf{P}(\boldsymbol{r}')\boldsymbol{a}_n^{\text{src}}(\boldsymbol{r}')\mathrm{d}\boldsymbol{r}' \tag{7.4.16}$$

where $\mathbf{P}(\boldsymbol{r}') = \text{diag}(\tilde{\Pi}_p, \tilde{\Pi}_p, \tilde{\Pi}_s, \tilde{\Pi}_s)$. Using Equation (7.4.16), we would obtain direct integral equations for $\phi$ and $\psi$ in terms of the maps $\tilde{\Pi}_p$ and $\tilde{\Pi}_s$. From this, we can then use Equation (7.4.12) to obtain an integral equation for the boundary conditions. This equation could then be solved as in Section 7.4.2. Such an inverse problem would no doubt be ill-posed. Hence, we would need to employ classical regularisation techniques, e.g., Tikhonov regularisation, to obtain a stable solution. [50]

## 7.5 Conclusions

Here, we have presented the beginnings of a robust method for localising subsurface cracks in rolling element bearings. This method is based on full modelling of the elastic waves inside the bearing raceway. In particular, in Section 7.3 we have shown that the method is able to provide good localisation from only 3 sensors if the amplitude of the source is known; we noted that the largest source of error is incorrect identification of the source amplitude $\alpha$. We believe that this error can be controlled by adding a step to the inversion process where the amplitude $\alpha$ is fitted from the data.

In Section 7.4, we presented the theoretical groundwork for an elaboration on the method found in Section 7.3. Here, our goal is the recovery of density maps $\Pi_p$ and $\Pi_s$, rather than a source location $\boldsymbol{r}'$. We found that this problem can be solved, theoretically, by deconvolution; however, we noted that this is only possible if the source fields $\phi_{\mathrm{src}}$ and $\psi_{\mathrm{src}}$ are known in the whole domain $\Omega$.

There remain significant challenges to be overcome with this method. In particular, in the case of localising a single emission event, we have identified the phase of the amplitude as a major source of error. This, in essence, is the same issue basic triangulation algorithms run into: the problem of correctly identifying first arrival times. [72] Thus, more work is required to determine the optimal method for determining the amplitude of an emission event.

# Chapter 8

# Conclusions

Jessica J. Kent

---

The condition monitoring of rolling element bearings is vital for modern industry. In this thesis, we have presented what we hope is the beginning of a promising tool to aid in this vital work. In this chapter, the main results of Chapters 5 to 7 are summarised, and potential directions for further study are identified.

## 8.1 Main results

In Chapters 5 and 6, we set out and tested a new physics-based approach to condition monitoring in rolling element bearings. Both papers are centred around the development of an analytic model of elastic waves trapped in bearing raceways. We show how prior assumptions and information on loading can be used to enhance the recovery of stress information. Further, in Chapter 6, we provided numerical and theoretical results on the well-posedness of the inverse problem; in particular, we showed that the numerical stability of the inverse problem increases with the wave numbers, $k_p$ and $k_s$; a result that is well established in the literature for exterior Cauchy boundary value problems [49, 53, 54]. We further linked the well-posedness of the inverse problem to the diffraction limit; in particular, we found that as we approach the diffraction limit, the condition number of $\mathbf{M}_n^{\text{inv}}$ increases, leading to a loss in precision. (More details can be found in Section 5.3.2.)

In Chapter 7, we presented the groundwork for an extension of the methods found in Chapters 5 and 6 to the problem of source localisation. We outlined two approaches to the problem of source localisation; one which relies upon a brute force solution of the boundary conditions to uncover a single defect location and another approach which relies on the fact the the total field can be modelled as a density $\Pi$ convolved with the Green's solution to the Helmholtz equation; we showed that theoretically this approach could recover multiple point defects simultaneously, but that in practice further study is needed to make this approach viable.

## 8.2 Future prospects

As with any scientific endeavour, there remain avenues and possibilities to explore. In this section, possible future directions for research are discussed, some of which are already being explored as this thesis is being written. Of primary importance is the need for experimental validation. Thus far, we have undertaken a rigorous theoretical exposition and numerical validation of the methodology we propose; however, without experimental testing, it is hard to determine the practical utility of the model we have proposed for real-world condition monitoring applications.

There also remains the problem of static and quasi-static loading, like that found in slowly rotating machinery. Since the stability of our dynamic method begins to decay as we approach the static limit, it is clear that an analogous approach based on elastostatic theory holds promise. In particular, such a method would hold promise for detecting and monitoring unusual loading in applications where the rotational speed of the bearings is low enough to approach the static limit, e.g., wind turbine main bearings. [5].

Further, in Chapter 6, when discussing the monitoring of a healthy bearing, we make some key simplifying assumptions which, while acceptable for a proof of concept, would need to be addressed in a practical implementation. Firstly, when modelling the contact stress between the bearing raceway and the rolling elements, we assume a perfect line contact; in reality, the contact would not be a line contact but would be more elliptical. [5] Secondly, we do not account for the random slip of the rolling elements as they traverse the raceway; this random slippage affects the measured response signal quite significantly. [13]

It must also be noted that, in Chapter 6, we set out what we call the prior method, where we parametrise solutions to the inverse problem in terms of information already known. We found that using these priors allowed us to recover the stress inside the raceway with much greater accuracy. However, these are not priors in the Bayesian sense but just parameterisations. Thus, a promising direction for further development would be to develop this method into a more rigorous statistical method; this would involve treating the parameterisations as statistical distributions with some degree of inherent uncertainty. [50]

Further, the work presented in Chapter 7 requires refinement and testing. Firstly, a robust methodology must be developed to ensure that the phase of the amplitude can be recovered adequately. Moreover, the deconvolution method we propose needs yet more refinement to ensure that we are able to recover the field inside the entire bearing raceway, as without this, the method is not readily applicable. Lastly, we have yet to put our method to the test against more firmly established methods and techniques in source localisation. To carry out this work, acoustic emission data are needed; the method we outlined should be tested against common source localisation techniques such as triangulation. This would serve a dual purpose: (a) it would act as a validation step to ensure our method recovers what we claim it does, and (b) it would act as a benchmark for the efficiency of our method.

# Bibliography


[1] TA Harris. *Rolling Bearing Analysis*. A Wiley-interscience publication. Wiley, 2001. url: https://books.google.co.uk/books?id=Pt9SAAAAMAAJ.

[2] Idriss El-Thalji et al. "A summary of fault modelling and predictive health monitoring of rolling element bearings". In: *Mechanical systems and signal processing* 60 (2015), pp. 252–272.

[3] Edward Hart et al. "Wind Turbine Main-Bearing Loading and Wind Field Characteristics". In: *Wind Energy* 22.11 (2019), pp. 1534–1547.

[4] Latha Sethuraman et al. *Main Bearing Dynamics in Three-Point Suspension Drivetrains for Wind Turbines; National Wind Technology Center (NWTC), NREL (National Renewable Energy Laboratory)*. Tech. rep. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2015.

[5] E. Hart et al. "A review of wind turbine main-bearings: design, operation, modelling, damage mechanisms and fault detection". In: *Wind Energy Science Discussions* (2019). url: https://doi.org/10.5194/wes-2019-25.

[6] Michael Hornemann et al. "Establishing Failure Modes for Bearings in Wind Turbines". In: *Windpower Engineering & Development* (2013). url: https://www.windpowerengineering.com/establishing-failure-modes-for-bearings-in-wind-turbines/.

[7] Bin Lu et al. "A review of recent advances in wind turbine condition monitoring and fault diagnosis". In: *2009 IEEE power electronics and machines in wind applications* (2009), pp. 1–7.

[8] William Alex Gray. "In-Situ Measurement of the Roller Bearing Inlet Meniscus Using Ultrasonic Spectroscopy". PhD thesis. University of Sheffield, 2023.

[9] Henrique Dias Machado de Azevedo et al. "A review of wind turbine bearing condition monitoring: State of the art and challenges". In: *Renewable and Sustainable Energy Reviews* 56 (2016), pp. 368–379.

[10] Chandrabhanu Malla et al. "Review of condition monitoring of rolling element bearing using vibration analysis and other techniques". In: *Journal of Vibration Engineering & Technologies* 7 (2019), pp. 407–414.

[11] Thomas Peter Howard. “Development of a novel bearing concept for improved wind turbine gearbox reliability”. PhD thesis. University of Sheffield, 2016.

[12] Jun Liu et al. “Overview of dynamic modelling and analysis of rolling element bearings with localized and distributed faults”. In: *Nonlinear Dynamics* 93 (2018), pp. 1765–1798. url: https://doi.org/10.1007/s11071-018-4314-y.

[13] Robert B Randall et al. “Rolling element bearing diagnostics—A tutorial”. In: *Mechanical systems and signal processing* 25.2 (2011), pp. 485–520.

[14] R.B. Randall. *Vibration-based Condition Monitoring: Industrial, Automotive and Aerospace Applications*. Wiley, 2021. url: https://books.google.co.uk/books?id=InoqEAAAQBAJ.

[15] P.D. McFadden et al. “Model for the vibration produced by a single point defect in a rolling element bearing”. In: *Journal of Sound and Vibration* 96.1 (1984), pp. 69–82.

[16] P.D. McFadden et al. “The vibration produced by multiple point defects in a rolling element bearing”. In: *Journal of Sound and Vibration* 98.2 (1985), pp. 263–273.

[17] Ralph A Wiggins. “Minimum entropy deconvolution”. In: *Geoexploration* 16.1-2 (1978), pp. 21–35.

[18] Jérôme Antoni. “The spectral kurtosis: a useful tool for characterising non-stationary signals”. In: *Mechanical systems and signal processing* 20.2 (2006), pp. 282–307.

[19] Jesper Berntsen et al. “Enhanced demodulation band selection based on Operational Modal Analysis (OMA) for bearing diagnostics”. In: *Mechanical Systems and Signal Processing* 181 (2022), p. 109300.

[20] Wade A Smith et al. “Optimal demodulation-band selection for envelope-based diagnostics: A comparative study of traditional and novel tools”. In: *Mechanical Systems and Signal Processing* 134 (2019), p. 106303.

[21] Zhi Ke Peng et al. “Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography”. In: *Mechanical systems and signal processing* 18.2 (2004), pp. 199–221.

[22] Peter W Tse et al. “Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—their effectiveness and flexibilities”. In: *J. Vib. Acoust.* 123.3 (2001), pp. 303–310.

[23] Arthur W Lees. *Vibration problems in machines: diagnosis and resolution*. CRC Press, 2020.

[24] Urko Leturiondo et al. “Methodology for the physics-based modelling of multiple rolling element bearing configurations”. In: *Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics* 231.1 (2017), pp. 194–212.

[25] TA Harris et al. *Advanced concepts of bearing technology: rolling bearing analysis*. CRC press, 2006.

[26] TA Harris et al. "Rolling element bearing dynamics". In: *Wear* 23.3 (1973), pp. 311–337.

[27] M Wrzochal et al. "The problems of mathematical modelling of rolling bearing vibrations". In: *Bulletin of the Polish Academy of Sciences. Technical Sciences* 68.6 (2020).

[28] Sarabjeet Singh et al. "An extensive review of vibration modelling of rolling element bearings with localised and extended defects". In: *Journal of Sound and Vibration* 357 (2015), pp. 300–330.

[29] MF While. "Rolling element bearing vibration transfer characteristics: effect of stiffness". In: (1979).

[30] A Choudhury et al. "Vibration response of rolling element bearings in a rotor bearing system to a local defect under radial load". In: (2006).

[31] N Tandon et al. "An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect". In: *Journal of sound and vibration* 205.3 (1997), pp. 275–292.

[32] AP Patil et al. "Vibration based modelling of acoustic emission of rolling element bearings". In: *Journal of Sound and Vibration* 468 (2020), p. 115117.

[33] Nader Sawalhi et al. "Simulating gear and bearing interactions in the presence of faults: Part II: Simulation of the vibrations produced by extended bearing faults". In: *Mechanical Systems and Signal Processing* 22.8 (2008), pp. 1952–1966.

[34] James A Greenwood et al. "Contact of nominally flat surfaces". In: *Proceedings of the royal society of London. Series A. Mathematical and physical sciences* 295.1442 (1966), pp. 300–319.

[35] Robert L Jackson et al. "A finite element study of elasto-plastic hemispherical contact against a rigid flat". In: *J. Trib.* 127.2 (2005), pp. 343–354.

[36] Ibrahim Zeid et al. "Finite element modeling of rolling contact". In: *Computers & Structures* 14.1-2 (1981), pp. 163–170.

[37] VGTCA Bhargava et al. "An elastic-plastic finite element model of rolling contact, Part 1: Analysis of single contacts". In: (1985).

[38] M Grujicic et al. "Multiphysics computational analysis of white-etch cracking failure mode in wind turbine gearbox bearings". In: *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications* 230.1 (2016), pp. 43–63.

[39] Sarabjeet Singh et al. "Analyses of contact forces and vibration response for a defective rolling element bearing using an explicit dynamics finite element model". In: *Journal of Sound and Vibration* 333.21 (2014), pp. 5356–5377.

[40] Hui Ma et al. “Nonlinear vibration of cylindrical roller bearing considering the surface roughness: simulation and experiment: Hui Ma et al.” In: *Nonlinear Dynamics* (2025), pp. 1–17.

[41] Nam-Ho Kim et al. *Introduction to finite element analysis and design*. John Wiley & Sons, 2018.

[42] Peter Howell et al. *Applied Solid Mechanics*. Cambridge, UK; New York: Cambridge University Press, 2009.

[43] Jan D. Achenbach. *Wave Propagation in Elastic Solids*. Amsterdam: Elsevier Science, 2005.

[44] Peter M. Shearer. *Introduction to Seismology*. Cambridge; New York: Cambridge University Press, 2009.

[45] Vlastislav Červenỳ. “Seismic ray theory”. In: *Geophysics*. Springer, 1989, pp. 1098–1108.

[46] Paul A Martin. *Multiple scattering: interaction of time-harmonic waves with N obstacles*. 107. Cambridge University Press, 2006.

[47] Milton Abramowitz et al. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. ninth Dover printing, tenth GPO printing. New York: Dover, 1964.

[48] K. K. Napal et al. “Effective T-matrix of a cylinder filled with a random two-dimensional particulate”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 480.2292 (2024).

[49] Victor Isakov. *Inverse Problems for Partial Differential Equations*. Springer, 2017, pp. 80–84.

[50] Jari Kaipio et al. *Statistical and computational inverse problems*. Vol. 160. Springer Science & Business Media, 2006.

[51] Lokenath Debnath et al. *Introduction to Hilbert spaces with applications*. 3rd Edition. Boston ; London: Elsevier Academic Press, 2005.

[52] *Integers and Floating-Point Numbers; The Julia Language — gensoft.pasteur.fr*. `https://gensoft.pasteur.fr/docs/julia/1.4.1/manual/integers-and-floating-point-numbers.html#`. [Accessed 01-06-2026].

[53] David Colton et al. “The linear sampling method in inverse electromagnetic scattering theory”. In: *Inverse Problems* 19.6 (2003), S105–S137.

[54] Tomasz Hrycak et al. “Increased stability in the continuation of solutions to the Helmholtz equation”. In: *Inverse Problems* 20.3 (2004), pp. 697–712.

[55] Jörg Kortemeyer. *Complex Numbers: An Introduction for First Year Students*. Springer International Publishing, 2022. url: `https://books.google.com/books/about/Complex_Numbers.html?id=42RXEAAAQBAJ`.

[56] Anders Vretblad. *Fourier Analysis and Its Applications*. eng. 1st ed. 2003. Graduate Texts in Mathematics, 223. New York, NY: Springer New York, 2003.

[57] Elias M. Stein et al. *Fourier Analysis: An Introduction*. Vol. 1. Princeton Lectures in Analysis. Princeton, NJ: Princeton University Press, 2003.

[58] AA Maznev et al. “Upholding the diffraction limit in the focusing of light and sound”. In: *Wave Motion* 68 (2017), pp. 182–189.

[59] Benny Lautrup. *Physics of Continuous Matter: Exotic and Everyday Phenomena in the Macroscopic World*. 1st. Bristol, UK: IOP Publishing, 2005.

[60] Mark H. Holmes. *Introduction to the Foundations of Applied Mathematics*. 2nd. Vol. 56. Texts in Applied Mathematics. Cham, Switzerland: Springer International Publishing, 2019. url: `https://springer.com`.

[61] Jyoti Kumar Sinha. *Industrial approaches in vibration-based condition monitoring*. CRC Press, 2020.

[62] A. Rezaei. “Fault Detection and Diagnosis on the Rolling Element Bearing”. Thesis. Carleton University, 2007.

[63] Alan Champneys et al. “Characterising Spectral Anomalies in Rotating Machine Sensors”. In: *Mathematics in Industry Reports* (2021).

[64] Alexander D Shaw et al. “Normal form analysis of bouncing cycles in isotropic rotor stator contact problems”. In: *International Journal of Mechanical Sciences* 155 (2019), pp. 83–97.

[65] S Janjarasjitt et al. “Bearing condition diagnosis and prognosis using applied nonlinear dynamical analysis of machine vibration signal”. In: *Journal of Sound and Vibration* 317 (2008), pp. 112–126.

[66] A W Lees et al. “Model-based identification of rotating machines”. In: *Mechanical Systems and Signal Processing* 23.6 (2009), pp. 1884–1893.

[67] Tobias Souza Morais et al. “Rotating machinery health evaluation by modal force identification”. In: *Inverse Problems in Science and Engineering* 28.5 (2020), pp. 695–715.

[68] Ian Howard. “A Review of rolling element bearing vibration "Detection, Diagnosis and Prognosis“”. In: (Oct. 1994).

[69] Mark J Eaton et al. “Towards improved damage location using acoustic emission”. In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 226.9 (2012), pp. 2141–2153.

[70] Matthew R Jones et al. “A Bayesian methodology for localising acoustic emission sources in complex structures”. In: *Mechanical Systems and Signal Processing* 163 (2022), p. 108143.

[71] Matthew R Jones et al. “Bayesian localisation of acoustic emission sources for wind turbine bearings”. In: *Health monitoring of structural and biological systems XV*. Vol. 11593. SPIE. 2021, pp. 443–453.

[72] C.U. Grosse et al. *Acoustic Emission Testing: Basics for Research – Applications in Engineering*. Springer Tracts in Civil Engineering. Springer International Publishing, 2021. url: https://books.google.co.uk/books?id=EYk4EAAAQBAJ.

[73] Nader Sawalhi et al. “The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis”. In: *Mechanical Systems and Signal Processing* 21.6 (2007), pp. 2616–2633.

[74] Mike Santora. *Schaeffler’s asymmetric spherical roller bearing for wind turbines*. 2017. url: https://www.bearingtips.com/schaefflers-asymmetric-spherical-roller-bearing-wind-turbines/.

[75] Miba AG. *A Miba Industrial Tilting Pad Journal Bearings*. 2024. url: https://www.miba.com/en/product-areas/industrial-bearings/tilting-pad-journal-bearing.

[76] Guo-Yang Li et al. “An ultrasonic method to measure stress without calibration: The angled shear wave method”. In: *The Journal of the Acoustical Society of America* 148.6 (2020), pp. 3963–3970.

[77] Jerrold E Marsden et al. *Mathematical foundations of elasticity*. Courier Corporation, 2012.

[78] Phillip L. Gould et al. *Introduction to Linear Elasticity*. Cham: Springer, 2018.

[79] J.R. Barber. *Contact Mechanics*. Cham, Switzerland: Springer, 2018. url: https://books.google.co.uk/books?id=4GNLDwAAQBAJ.

[80] T.J. Martin et al. “An inverse method for finding unknown surface tractions and deformations in elastostatics”. In: *Computers & Structures* 56.5 (1995). Boundary Element Method, pp. 825–835. url: https://www.sciencedirect.com/science/article/pii/0045794995000115.

[81] L Marin et al. “Boundary element method for the Cauchy problem in linear elasticity”. In: *Engineering Analysis with Boundary Elements* 25.9 (2001), pp. 783–793. url: https://www.sciencedirect.com/science/article/pii/S0955799701000625.

[82] D.B. Ingham L. Marin L. Elliott et al. “Boundary Element Regularisation Methods for Solving the Cauchy Problem in Linear Elasticity”. In: *Inverse Problems in Engineering* 10.4 (2002), pp. 335–357. eprint: https://doi.org/10.1080/1068276021000004698. url: https://doi.org/10.1080/1068276021000004698.

[83] AL Gower et al. *ElasticWaves.jl: A Julia library to calculate propagation and scattering of elastic waves*. 2024. url: https://github.com/JuliaWaveScattering/ElasticWaves.jl/tree/v0.1.0.

[84] AL Gower et al. *MultipleScattering.jl: A Julia library for simulating, processing, and plotting multiple scattering of waves*. 2020. url: https://github.com/JuliaWaveScattering/MultipleScattering.jl.

[85] Robert B Randall et al. "The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals". In: *Mechanical systems and signal processing* 15.5 (2001), pp. 945–962.

[86] Farshid Sadeghi et al. "A review of rolling contact fatigue". In: *ASME. J. Tribol.* (2009).

[87] Jing Liu et al. "An investigation of contact characteristics of a roller bearing with a subsurface crack". In: *Engineering Failure Analysis* 116 (2020), p. 104744.

[88] Jessica J. Kent et al. "Elastic waves in bearing raceways: the forward and inverse problem". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 481.2322 (Sept. 2025), p. 20240972. eprint: https://royalsocietypublishing.org/rspa/article-pdf/doi/10.1098/rspa.2024.0972/2813564/rspa.2024.0972.pdf. url: https://doi.org/10.1098/rspa.2024.0972.

[89] DLMF. *DLMF: §10.23 Sums > Bessel and Hankel Functions > Chapter 10 Bessel Functions*. https://dlmf.nist.gov/10.23. Accessed: 2024-09-10. n.d.