A Guided Tour of Mathematical Physics

Roel Snieder

Dept. of Geophysics, Utrecht University, P.O. Box 80.021,
3508 TA Utrecht, The Netherlands

Published by Samizdat Press
Golden · White River Junction

Samizdat Press publications are available
via Internet from http://samizdat.mines.edu

November 16, 1998

Contents

Chapter 1

Introduction

The topic of this course is the application of mathematics to physical problems. In practice, mathematics and physics are taught separately. Despite the fact that education in physics relies on mathematics, it turns out that students consider mathematics to be disjoint from physics. Although this point of view may strictly be correct, it reflects an erroneous opinion when it concerns an education in physics or geophysics. The reason for this is that mathematics is the *only* language at our disposal for quantifying physical processes. One cannot learn a language by just studying a textbook. In order to truly learn how to use a language one has to go abroad and start using a language. By the same token one cannot learn how to use mathematics in physics by just studying textbooks or attending lectures, the only way to achieve this is to venture into the unknown and apply mathematics to physical problems.

It is the goal of this course to do exactly that; a number of problems is presented in order to apply mathematical techniques and knowledge to physical concepts. These examples are not presented as well-developed theory. Instead, these examples are presented as a number of problems that elucidate the issues that are at stake. In this sense this book offers a guided tour; material for learning is presented but true learning will only take place by active exploration.

Since this book is written as a set of problems you may frequently want to consult other material to refresh or deepen your understanding of material. In many places we will refer to the book of *Boas*[11]. In addition, the books of *Butkov*[14] and *Arfken*[3] are excellent. When you are a physics of geophysics student you should seriously consider buying a comprehensive textbook on mathematical physics, it will be of great benefit to you.

In addition to books, colleagues either in the same field or other fields can be a great source of knowledge and understanding. Therefore, don't hesitate to work together with others on these problems if you are in the fortunate positions to do so. This may not only make the work more enjoyable, it may also help you in getting "unstuck" at difficult moments and the different viewpoints of others may help to deepen yours.

This book is set up with the goal of obtaining a good working knowledge of mathematical geophysics that is needed for students in physics or geophysics. A certain basic knowledge of calculus and linear algebra is needed for digesting the material presented here. For this reason, this book is meant for upper-level undergraduate students or lower-level graduate students, depending on the background and skill of the student.

At this point the book is still under construction. New sections are regularly added, and both corrections and improvements will be made. If you are interested in this material therefore regularly check the latest version at Samizdat Press. The feedback of both teachers and students who use this material is vital in improving this manuscript, please send you remarks to:

Roel Snieder

Dept. of Geophysics
Utrecht University
P.O. Box 80.021
3508 TA Utrecht
The Netherlands

telephone: +31-30-253.50.87
fax: +31-30-253.34.86
email: snieder@geo.uu.nl

Acknowledgements: This manuscript has been prepared with the help of a large number of people. The feedback of John Scales and his proposal to make this manuscript available via internet is very much appreciated. Barbara McLenon skillfully drafted the figures. The patience of Joop Hoofd, John Stockwell and Everhard Muyzert in dealing with my computer illiteracy is very much appreciated. Numerous students have made valuable comments for improvements of the text. The input of Huub Douma in correcting errors and improving the style of presentation is very much appreciated.

Chapter 2

Summation of series

2.1 The Taylor series

In many applications in mathematical physics it is extremely useful to write the quantity of interest as a sum of a large number of terms. To fix our mind, let us consider the motion of a particle that moves along a line as time progresses. The motion is completely described by giving the position $x(t)$ of the particle as a function of time. Consider the four different types of motion that are shown in figure 2.1.

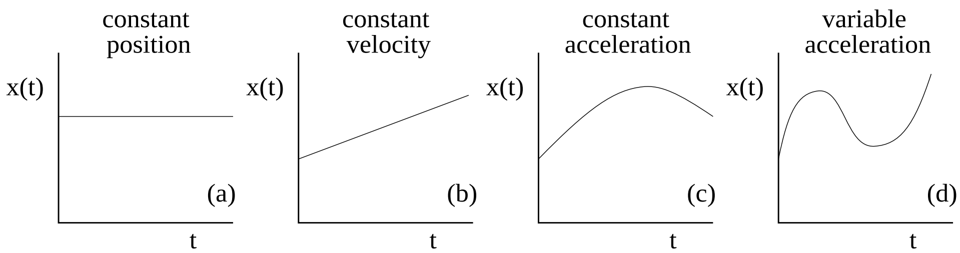


Figure 2.1: Four different kinds of motion of a particle along a line as a function of time.

The simplest motion is a particle that does not move, this is shown in panel (a). In this case the position of the particle is constant:

$$x(t) = x_0 \ . \tag{2.1}$$

The value of the parameter x_0 follows by setting $t = 0$, this immediately gives that

$$x_0 = x\left(0\right) \ . \tag{2.2}$$

In panel (b) the situation is shown of a particle that moves with a constant velocity, in that case the position is a linear function of time:

$$x(t) = x_0 + v_0 t \ . \tag{2.3}$$

Again, setting $t = 0$ gives the parameter x_0, which is given again by (2.2). The value of the parameter v_0 follows by differentiating (2.3) with respect to time and by setting $t = 0$.

Problem a: Do this and show that

$$v_0 = \frac{dx}{dt}(t=0) \ . \tag{2.4}$$

This expression reflects that the velocity v_0 is given by the time-derivative of the position. Next, consider a particle moving with a constant acceleration a_0 as shown in panel (c). As you probably know from classical mechanics the motion is in that case a quadratic function of time:

$$x(t) = x_0 + v_0 t + \frac{1}{2} a_0 t^2 \ . \tag{2.5}$$

Problem b: Evaluate this expression at $t = 0$ to show that x_0 is given by (2.2). Differentiate (2.5) once with respect to time and evaluate the result at $t = 0$ to show that v_0 is again given by (2.4). Differentiate (2.5) twice with respect to time, set $t = 0$ to show that a_0 is given by:

$$a_0 = \frac{d^2x}{dt^2}(t=0) \ . \tag{2.6}$$

This result reflects the fact that the acelleration is the second derivative of the position with respect to time.

Let us now consider the motion shown in panel (d) where the acceleration changes with time. In that case the displacement as a function of time is not a linear function of time (as in (2.3) for the case of a constant velocity) nor is it a quadratic function of time (as in (2.5) for the case of a constant acceleration). Instead, the displacement is in general a function of all possible powers in t:

$$x(t) = c_0 + c_1 t + c_2 t^2 + \cdots = \sum_{n=0}^{\infty} c_n t^n \ . \tag{2.7}$$

This series, where a function is expressed as a sum of terms with increasing powers of the independent variable, is called a *Taylor series.* At this point we do not know what the constants c_n are. These coefficients can be found in exactly the same way as in **problem b** where you determined the coefficients a_0 and v_0 in the expansion (2.5).

Problem c: Determine the coefficient c_m by differentiating expression (2.7) m-times with respect to t and by evaluating the result at $t = 0$ to show that

$$c_m = \frac{1}{m!}\frac{d^m x}{dt^m}(x=0) \ . \tag{2.8}$$

Of course there is no reason why the Taylor series can only be used to describe the displacement $x(t)$ as a function of time t. In the literature, one frequently uses the Taylor series do describe a function $f(x)$ that depends on x. Of course it is immaterial how we call a function. By making the replacements $x \to f$ and $t \to x$ the expressions (2.7) and (2.8) can also be written as:

$$f(x) = \sum_{n=0}^{\infty} c_n x^n \ , \tag{2.9}$$

with

$$c_n = \frac{1}{n!}\frac{d^n f}{dx^n}(x=0) \ . \tag{2.10}$$

You may find this result in the literature also be written as

$$f(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}\frac{d^n f}{dx^n}(x=0) = f(0) + x\frac{df}{dx}(x=0) + \frac{1}{2}\frac{d^2 f}{dx^2}(x=0) + \cdots \tag{2.11}$$

Problem d: Show by evaluating the derivatives of $f(x)$ at $x = 0$ that the Taylor series of the following functions are given by:

$$\sin(x) = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots \tag{2.12}$$

$$\cos(x) = 1 - \frac{1}{2}x^2 + \frac{1}{4!}x^4 - \cdots \tag{2.13}$$

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots = \sum_{n=0}^{\infty}\frac{1}{n!}x^n \tag{2.14}$$

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots = \sum_{n=0}^{\infty} x^n \tag{2.15}$$

$$(1-x)^\alpha = 1 - \alpha x + \frac{1}{2!}\alpha(\alpha-1)x^2 - \frac{1}{3!}\alpha(\alpha-1)(\alpha-2)x^3 + \cdots \tag{2.16}$$

Up to this point the Taylor expansion was made around the point $x = 0$. However, one can make a Taylor expansion around any arbitrary point x. The associated Taylor series can be obtained by replacing the distance x that we move from the expansion point by a distance h and by replacing the expansion point 0 by x. Making the replacements $x \to h$ and $0 \to x$ the expansion (2.11) is given by:

$$f(x+h) = \sum_{n=0}^{\infty}\frac{h^n}{n!}\frac{d^n f}{dx^n}(x) \tag{2.17}$$

The Taylor series can not only be used for functions of a single variable. As an example consider a function $f(x, y)$ that depends on the variables x and y. The generalization of the Taylor series (2.9) to functions of two variables is given by

$$f(x,y) = \sum_{n,m=0}^{\infty} c_{nm} x^n y^m \ . \tag{2.18}$$

At this point the coefficients c_{nm} are not yet known. They follow in the same way as the coefficients of the Taylor series of a function that depends on a single variable by taking the derivatives of the Taylor series and by evaluating the result in the point where the expansion is made.

Problem e: Take suitable derivatives of (2.18) with respect to x and y and evaluate the result in the expansion point $x = y = 0$ to show that up to second order the Taylor expansion (2.18) is given by

$$\begin{aligned} f(x,y) &= f(0,0) + \frac{\partial f}{\partial x}(0,0)\ x + \frac{\partial f}{\partial y}(0,0)\ y \\ &+ \tfrac{1}{2}\frac{\partial^2 f}{\partial x^2}(0,0)\ x^2 + \frac{\partial^2 f}{\partial x \partial y}(0,0)\ xy + \tfrac{1}{2}\frac{\partial^2 f}{\partial y^2}(0,0)\ y^2 + \cdots \end{aligned} \tag{2.19}$$

Problem f: This is the Taylor expansion of $f(x,y)$ around the point $x = y = 0$. Make suitable substitutions in this result to show that the Taylor expansion around an arbitrary point (x,y) is given by

$$\begin{aligned} f(x+h_x, y+h_y) &= f(x,y) + \frac{\partial f}{\partial x}(x,y)\ h_x + \frac{\partial f}{\partial y}(x,y)\ h_y \\ &+ \tfrac{1}{2}\frac{\partial^2 f}{\partial x^2}(x,y)\ h_x^2 + \frac{\partial^2 f}{\partial x \partial y}(x,y)\ h_x h_y + \tfrac{1}{2}\frac{\partial^2 f}{\partial y^2}(x,y)\ h_y^2 + \cdots \end{aligned} \tag{2.20}$$

Let us now return to the Taylor series (2.9) with the coefficients c_m given by (2.10). This series hides a very intriguing result. It follows from (2.9) and (2.10) that a function $f(x)$ is specified for all values of its argument x when all the derivatives are known at a single point $x = 0$. This means that the global behavior of a function is completely contained in the properties of the function at a single point. In fact, this is not always true.

First, the series (2.9) is an infinite series, and the sum of infinitely many terms does not necessarily lead to a finite answer. As an example look at the series (2.15). A series can only converge when the terms go to zero as $n \to \infty$, because otherwise every additional term changes the sum. The terms in the series (2.15) are given by x^n, these terms only go to zero as $n \to \infty$ when $|x| < 1$. In general, the Taylor series (2.9) only *converges* when x is smaller than a certain critical value called the *radius of convergence*. Details on the criteria for the convergence of series can be found for example in *Boas*?? or *Butkov*??.

The second reason why the derivatives at one point do not necessarily constrain the function everywhere is that a function may change its character over the range of parameter values that is of interest. As an example let us return to a moving particle and consider a particle with position $x(t)$ that is at rest until a certain time t_0 and that then starts moving with a uniform velocity $v \neq 0$:

$$x(t) = \begin{cases} x_0 & \text{for } t \le t_0 \\ x_0 + v(t - t_0) & \text{for } t > t_0 \end{cases} \tag{2.21}$$

The motion of the particle is sketched in figure 2.2. A straightforward application of (2.8) shows that all the coefficients c_n of this function vanish except c_0 which is given by x_0. The Taylor series (2.7) is therefore given by $x(t) = x_0$ which clearly differs from (2.21). The reason for this is that the function (2.21) changes its character at $t = t_0$ in such a way that nothing in the behavior for times $t < t_0$ predicts the sudden change in the motion at time $t = t_0$. Mathematically things go wrong because the higher derivatives of the function do not exist at time $t = t_0$.

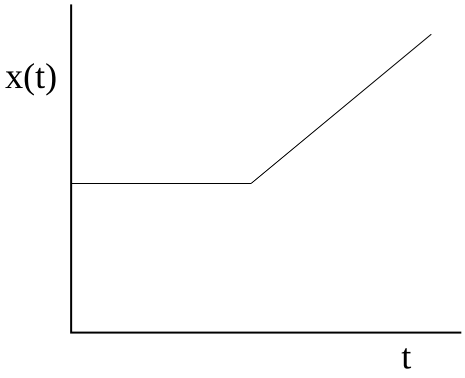


Figure 2.2: The motion of a particle that suddenly changes character at time t_0.

Problem g: Compute the second derivative of $x(t)$ at $t = t_0$.

The function (2.21) is said to be not analytic at the point $t = t_0$. The issue of analytic functions is treated in more detail in the sections 12.1 and 13.1.

Problem h: Try to compute the Taylor series of the function $x(t) = 1/t$ using (2.7) and (2.8). Draw this function and explain why the Taylor series cannot be used for this function.

Problem i: Do the same for the function $x(t) = \sqrt{t}$.

Frequently the result of a calculation can be obtained by summing a series. In section 2.2 this is used to study the behavior of a bouncing ball. The bounces are "natural" units for analyzing the problem at hand. In section 2.3 the reverse is done when studying the total reflection of a stack of reflective layers. In this case a series expansion actually gives physical insight in a complex expression.

2.2 The bouncing ball

In this exercise we study a problem of a rubber ball that bounces on a flat surface and slowly comes to rest as sketched in figure (2.3). You will know from experience that the ball bounces more and more rapidly with time. The question we address here is whether the ball can actually bounce infinitely many times in a finite amount of time. This problem is not an easy one. In general with large difficult problems it is a useful strategy to divide the large and difficult problem that you cannot solve in smaller and simpler problems that you can solve. By assembling these smaller sub-problems one can then often solve the large problem. This is exactly what we will do here. First we will solve how much time it takes for the ball to bounce once given its velocity. Given a prescription of the energy-loss in one bounce we will determine a relation between the velocity of subsequent bounces. From these ingredients we can determine the relation between the times needed for subsequent bounces. By summing this series over an infinite number of bounces we can determine the total time that the ball has bounced. *Keep this general strategy in mind when solving complex problems. Almost all of us are better at solving a number of small problems rather than a single large problem!*

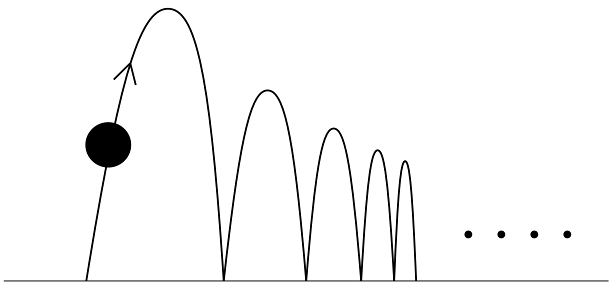


Figure 2.3: The motion of a bouncing ball that looses energy with every bounce.

Problem a: A ball moves upward from the level $z = 0$ with velocity v. Determine the height the ball reaches and the time it takes for the ball to return to its starting point.

At this point we have determined the relevant properties for a single bounce. During each bounce the ball looses energy due to the fact that the ball is deformed anelastically during the bounce. We will assume that during each bounce the ball looses a fraction γ of its energy.

Problem b: Let the velocity at the beginning of the $n-$th bounce be v_n. Show that with assumed rule for energy loss this velocity is related to the velocity v_{n-1} of the previous bounce by

$$v_n = \sqrt{1-\gamma}\, v_{n-1}. \tag{2.22}$$

Hint: when the ball bounces upward from $z = 0$ all its energy is kinetic energy.

In **problem a** you determined the time it took the ball to bounce once, given the initial velocity, while expression (2.22) gives a recursive relation for the velocity between subsequent bounces. By assembling these results we can find a relation for the time t_n for the $n-$th bounce and the time t_{n-1} for the previous bounce.

Problem c: Determine this relation. In addition, let us assume that the ball is thrown up the first time from $z = 0$ to reach a height $z = H$. Compute the time t_0 needed for the ball to make the first bounce and combine these results to show that

$$t_n = \sqrt{\frac{8H}{g}}(1-\gamma)^{n/2}, \tag{2.23}$$

where g is the acceleration of gravity.

We can now use this expression to determine the total time T_N it takes to carry out N bounces. This time is given by $T_N = \sum_{n=0}^{N} t_n$. By setting N equal to infinity we can compute the time T_∞ it takes to bounce infinitely often.

Problem d: Determine this time by carrying out the summation and show that this time is given by:

$$T_\infty = \sqrt{\frac{8H}{g}} \frac{1}{1-\sqrt{1-\gamma}}. \tag{2.24}$$

Hint: write $(1-\gamma)^{n/2}$ as $\left(\sqrt{1-\gamma}\right)^n$ and treat $\sqrt{1-\gamma}$ as the parameter x in the appropriate Taylor series of section (2.1).

This result seems to suggest that the time it takes to bounce infinitely often is indeed finite.

Problem e: Show that this is indeed the case, except when the ball looses no energy between subsequent bounces. Hint: translate the condition that the ball looses no energy in one of the quantities in the equation (2.24).

Expression (2.24) looks messy. It happens often in mathematical physics that a final expression is complex; very often final results look so messy it is difficult to understand them. However, often we know that certain terms in an expression can assumed to be very small (or very large). This may allow us to obtain an approximate expression that is of a simpler form. In this way we trade accuracy for simplicity and understanding. In practice, this often turns out to be a good deal! In our example of the bouncing ball we assume that the energy-loss at each bounce is small, i.e. that γ is small.

Problem f: Show that in this case $T_\infty \approx \sqrt{\frac{8H}{g}}\frac{2}{\gamma}$ by using the leading terms of the appropriate Taylor series of section (2.1).

This result is actually quite useful. It tells us *how* the total bounce time approaches infinity when the energy loss γ goes to zero.

In this problem we have solved the problem in little steps. In general we will take larger steps during this course, you will have to discover how to divide a large step in smaller steps. The next problem is a "large" problem, solve it by dividing it in smaller problems. First formulate the smaller problems as ingredients for the large problem before you actually start working on the smaller problems. *Make it a habit whenever you solve problems to first formulate a strategy how you are going to attack a problem before you actually start working on the sub-problems. Make a list if this helps you and don't be deterred if you cannot solve a particular sub-problem. Perhaps you can solve the other sub-problems and somebody else can help you with the one you cannot solve.* Keeping this in mind solve the following "large" problem:

Problem g: Let the total distance travelled by the ball during infinitely many bounces be denoted by S. Show that $S = 2H/\gamma$.

2.3 Reflection and transmission by a stack of layers

Lord Rayleigh[48] addressed in 1917 the question why some birds or insects have beautiful iridescent colors. He explained this by studying the reflective properties of a stack of thin reflective layers. This problem is also of interest in geophysics; in exploration seismology one is also interested in the reflection and transmission properties of stacks of reflective layers in the earth. Lord Rayleigh solved this problem in the following way. Suppose we have one stack of layers on the left with reflection coefficient R_L and transmission coefficient T_L and another stack of layers on the right with reflection coefficient R_R and

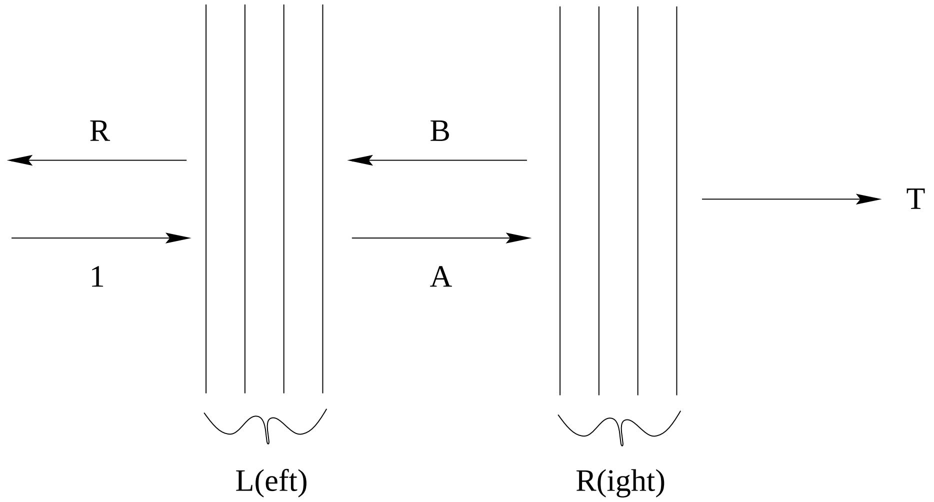


Figure 2.4: Geometry of the problem where stacks of n and m reflective layers are combined. The notation of the strength of left- and rightgoing waves is indicated.

transmission coefficient T_R. If we add these two stacks together to obtain a larger stack of layers, what are the reflection coefficient R and transmission coefficient T of the total stack of layers? See figure (2.4) for the scheme of this problem. Note that the reflection coefficient is defined as the ratio of the strength of the reflected wave and the incident wave, similarly the transmission coefficient is defined as the ratio of the strength of the transmitted wave and the incident wave. For simplicity we will simplify the analysis and ignore that the reflection coefficient for waves incident from the left and the right are in general not the same. However, this simplification does not change the essence of the coming arguments.

Before we start solving the problem, let us speculate what the transmission coefficient of the combined stack is. Since the transmission coefficient T_L of the left stack determines the ratio of the transmitted wave to the incident wave, and since T_R is the same quantity of the right stack, it seems natural to assume that the transmission coefficient of the combined stack is the product of the transmission coefficient of the individual stacks: $T = T_L T_R$. However, this result is wrong and we will try to discover why this is so.

Consider figure (2.4) again. The unknown quantities are R, T and the coefficients A and B for the right-going and left-going waves between the stacks. An incident wave with strength 1 impinges on the stack from the left. Let us first determine the coefficient A of the right-going waves between the stacks. The right-going wave between the stacks contains two contributions; the wave transmitted from the left (this contribution has a strength $1 \times T_L$) and the wave reflected towards the right due the incident left-going wave with strength B (this contribution has a strength $B \times R_L$). This implies that:

$$A = T_L + BR_L \ . \tag{2.25}$$

Problem a: Using similar arguments show that:

$$B = AR_R \ , \tag{2.26}$$

$$T = AT_R \ , \tag{2.27}$$

$$R = R_L + BT_L \; . \tag{2.28}$$

This is all we need to solve our problem. The system of equations (2.25)-(2.28) consists of four linear equations with four unknowns A, B, R and T. We could solve this system of equations by brute force, but some thinking will make life easier for us. Note that the last two equations immediately give T and R once A and B are known. The first two equations give A and B.

Problem b: Show that

$$A = \frac{T_L}{(1 - R_L R_R)} \; , \tag{2.29}$$

$$B = \frac{T_L R_R}{(1 - R_L R_R)} . \tag{2.30}$$

This is a puzzling result, the right-going wave A between the layers does not only contain the transmission coefficient of the left layer T_L but also and additional term $1/(1-R_L R_R)$.

Problem c: Make a series expansion of $1/(1 - R_L R_R)$ in the quantity $R_L R_R$ and show that this term accounts for the waves that bounce back and forth between the two stacks. Hint: use that R_L gives the reflection coefficient for a wave that reflects from the left stack, R_R gives the reflection coefficient for one that reflects from the right stack so that $R_L R_R$ is the total reflection coefficient for a wave that bounces once between the left and the right stack.

This implies that the term $1/(1-R_L R_R)$ accounts for the waves that bounce back and forth between the two stacks of layers. It is for this reason that we call this term a *reverberation* term. It plays an important role in computing the response of layered media.

Problem d: Show that the reflection and transmission coefficient of the combined stack of layers is given by:

$$R = R_L + \frac{T_L^2 R_R}{(1 - R_L R_R)} , \tag{2.31}$$

$$T = \frac{T_L T_R}{(1 - R_L R_R)} . \tag{2.32}$$

In the beginning of this section we conjectured that the transmission coefficient of the combined stacks is the product of the transmission coefficient of the separate stacks.

Problem e: Is this correct? Under which conditions is it approximately correct?

Equations (2.31) and (2.32) are very useful for computing the reflection and transmission coefficient of a large stack of layers. The reason for this is that it is extremely simple to determine the reflection and transmission coefficient of a very thin layer using the Born approximation. Let the reflection and transmission coefficient of a *single* thin layer n be denoted by r_n respectively t_n and let the reflection and transmission coefficient of a *stack* of n layers be denoted by R_n and T_n respectively. Suppose the left stack consists on n layers and that we want to add an $(n + 1)$-th layer to the stack. In that case the right stack consists of a single $(n+1)$-th layer so that $R_R = r_{n+1}$ and $T_R = t_{n+1}$ and the reflection and transmission coefficient of the left stack are given by $R_L = R_n$, $T_L = T_n$. Using this in expressions (2.31) and (2.32) yields

$$R_{n+1} = R_n + \frac{T_n^2 r_{n+1}}{(1 - R_n r_{n+1})}, \tag{2.33}$$

$$T_{n+1} = \frac{T_n t_{n+1}}{(1 - R_n r_{n+1})}. \tag{2.34}$$

This means that given the known response of a stack of n layers, one can easily compute the effect of adding the $(n+1) - th$ layer to this stack. In this way one can recursively build up the response of the complex reflector out of the known response of very thin reflectors. Computers are pretty stupid, but they are ideally suited for applying the rules (2.33) and (2.34) a large number of times. Of course this process has to be started when we start with a medium in which no layers are present.

Problem f: What are the reflection coefficient R_0 and the transmission coefficient T_0 when there are no reflective layers present yet? Describe how one can compute the response of a thick stack of layers once we know the response of a very thin layer.

In developing this theory, Lord Rayleigh prepared the foundations for a theory that later became known as *invariant embedding* which turns out to be extremely useful for a number of scattering and diffusion problems[6][61].

The main conclusion of the treatment of this section is that the transmission of a combination of two stacks of layers is not the product of the transmission coefficients of the two separate stacks. Paradoxically, *Berry and Klein*[8] showed in their analysis of "transparent mirrors" that for a large stacks of layers with random transmission coefficients the total transmission coefficients *is* the product of the transmission coefficients of the individual layers, despite the fact that multiple reflections play a crucial role in this process.

Chapter 3

Spherical and cylindrical coordinates

Many problems in mathematical physics exhibit a spherical or cylindrical symmetry. For example, the gravity field of the Earth is to first order spherically symmetric. Waves excited by a stone thrown in water usually are cylindrically symmetric. Although there is no reason why problems with such a symmetry cannot be analyzed using Cartesian coordinates (i.e. (x, y, z)-coordinates), it is usually not very convenient to use such a coordinate system. The reason for this is that the theory is usually much simpler when one selects a coordinate system with symmetry properties that are the same as the symmetry properties of the physical system that one wants to study. It is for this reason that spherical coordinates and cylinder coordinates are introduced in this section. It takes a certain effort to become acquainted with these coordinate system, but this effort is well spend because it makes solving a large class of problems much easier.

3.1 Introducing spherical coordinates

In figure (3.1) a Cartesian coordinate system with its x, y and z-axes is shown as well as the location of a point $\mathbf{r}$. This point can either be described by its x, y and z-components or by the radius r and the angles θ and φ shown in figure (3.1). In the latter case one uses spherical coordinates. Comparing the angles θ and φ with the geographical coordinates that define a point on the globe one sees that φ can be compared with *longitude* and θ can be compared with *co-latitude*, which is defined as (*latitude - 90 degrees)*. The angle φ runs from 0 to 2π, while θ has values between 0 and π. In terms of Cartesian coordinates the position vector can be written as:

$$\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \ , \tag{3.1}$$

where the caret $(\hat{\cdot})$ is used to denote a vector that is of unit length. An arbitrary vector can of course also be expressed in these vectors:

$$\mathbf{u} = u_x\hat{\mathbf{x}} + u_y\hat{\mathbf{y}} + u_z\hat{\mathbf{z}} \ . \tag{3.2}$$

We want to express the same vector also in basis vectors that are related to the spherical coordinate system. Before we can do so we must first establish the connection between the Cartesian coordinates (x, y, z) and the spherical coordinates (r, θ, φ).

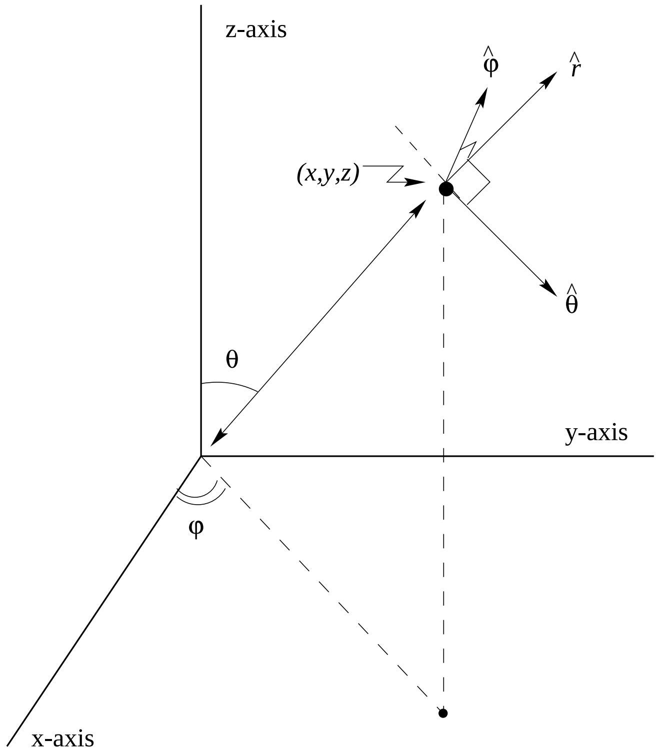


Figure 3.1: Definition of the angles used in the spherical coordinates.

Problem a: Use figure (3.1) to show that the Cartesian coordinates are given by:

$$\begin{aligned} x &= r \sin\theta \cos\varphi \\ y &= r \sin\theta \sin\varphi \\ z &= r \cos\theta \end{aligned} \tag{3.3}$$

Problem b: Use these expressions to derive the following expression for the spherical coordinates in terms of the Cartesian coordinates:

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2} \\ \theta &= \arccos\left(z/\sqrt{x^2 + y^2 + z^2}\right) \\ \varphi &= \arctan\left(y/x\right) \end{aligned} \tag{3.4}$$

We now have obtained the relation between the Cartesian coordinates (x, y, z) and the spherical coordinates (r, θ, φ). We want to express the vector $\mathbf{u}$ of equation (3.2) also in spherical coordinates:

$$\mathbf{u} = u_r \hat{\mathbf{r}} + u_\theta \hat{\boldsymbol{\theta}} + u_\varphi \hat{\boldsymbol{\varphi}} \ , \tag{3.5}$$

and we want to know the relation between the components (u_x, u_y, u_z) in Cartesian coordinates and the components $(u_r, u_\theta, u_\varphi)$ of the same vector expressed in spherical coordinates. In order to do this we first need to determine the unit vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$. In Cartesian coordinates, the unit vector $\hat{\mathbf{x}}$ points along the x-axis. This is a different way of saying that it is a unit vector pointing in the direction of increasing values of x for constant values of y and z; in other words, $\hat{\mathbf{x}}$ can be written as: $\hat{\mathbf{x}} = \partial\mathbf{r}/\partial x$.

Problem c: Verify this by carrying out the differentiation that the definition $\hat{\mathbf{x}} = \partial\mathbf{r}/\partial x$ leads to the correct unit vector in the x-direction: $\hat{\mathbf{x}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

Now consider the unit vector $\hat{\boldsymbol{\theta}}$. Using the same argument as for the unit vector $\hat{\mathbf{x}}$ we know that $\hat{\boldsymbol{\theta}}$ is directed towards increasing values of θ for constant values of r and φ. This means that $\hat{\theta}$ can be written as $\hat{\boldsymbol{\theta}} = C\partial\mathbf{r}/\partial\theta$. The constant C follows from the requirement that $\hat{\boldsymbol{\theta}}$ is of unit length.

Problem d: Use this reasoning for all the unit vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ and expression (3.3) to show that:

$$\hat{\mathbf{r}} = \frac{\partial\mathbf{r}}{\partial r} \quad , \quad \hat{\boldsymbol{\theta}} = \frac{1}{r}\frac{\partial\mathbf{r}}{\partial\theta} \quad , \quad \hat{\boldsymbol{\varphi}} = \frac{1}{r\sin\theta}\frac{\partial\mathbf{r}}{\partial\varphi} \, , \tag{3.6}$$

and that this result can also be written as

$$\hat{\mathbf{r}} = \begin{pmatrix} \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\theta \end{pmatrix} \quad , \quad \hat{\boldsymbol{\theta}} = \begin{pmatrix} \cos\theta\cos\varphi \\ \cos\theta\sin\varphi \\ -\sin\theta \end{pmatrix} \quad , \quad \hat{\boldsymbol{\varphi}} = \begin{pmatrix} -\sin\varphi \\ \cos\varphi \\ 0 \end{pmatrix} . \tag{3.7}$$

These equations give the unit vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ in Cartesian coordinates.

In the right hand side of (3.6) the derivatives of the position vector are divided by 1, r and $r\sin\theta$ respectively. These factors are usually shown in the following notation:

$$h_r = 1 \quad , \quad h_\theta = r \quad , \quad h_\varphi = r\sin\theta \, . \tag{3.8}$$

These scale factors play a very important role in the general theory of curvilinear coordinate systems, see *Butkov*[14] for details. The material presented in the remainder of this chapter as well as the derivation of vector calculus in spherical coordinates can be based on the scale factors given in (3.8). However, this approach will not be taken here.

Problem e: Verify explicitly that the vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ defined in this way form an orthonormal basis, i.e. they are of unit length and perpendicular to each other:

$$(\hat{\mathbf{r}}\cdot\hat{\mathbf{r}}) = \left(\hat{\boldsymbol{\theta}}\cdot\hat{\boldsymbol{\theta}}\right) = (\hat{\boldsymbol{\varphi}}\cdot\hat{\boldsymbol{\varphi}}) = 1 \, , \tag{3.9}$$

$$\left(\hat{\mathbf{r}}\cdot\hat{\boldsymbol{\theta}}\right) = (\hat{\mathbf{r}}\cdot\hat{\boldsymbol{\varphi}}) = \left(\hat{\boldsymbol{\theta}}\cdot\hat{\boldsymbol{\varphi}}\right) = 0 \, . \tag{3.10}$$

Problem f: Using the expressions (3.7) for the unit vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ show by calculating the cross product explicitly that

$$\hat{\mathbf{r}}\times\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\varphi}} \quad , \quad \hat{\boldsymbol{\theta}}\times\hat{\boldsymbol{\varphi}} = -\hat{\mathbf{r}} \quad , \quad \hat{\boldsymbol{\varphi}}\times\hat{\mathbf{r}} = \hat{\boldsymbol{\theta}} \, . \tag{3.11}$$

The Cartesian basis vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ point in the same direction at every point in space. This is not true for the spherical basis vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$; for different values of the angles θ and φ these vectors point in different directions. This implies that these unit vectors are functions of both θ and φ. For several applications it is necessary to know how the basis vectors change with θ and φ. This change is described by the derivative of the unit vectors with respect to the angles θ and φ.

Problem g: Show by direct differentiation of the expressions (3.7) that the derivatives of the unit vectors with respect to the angles θ and φ are given by:

$$\begin{array}{ll} \partial\hat{\mathbf{r}}/\partial\theta = \hat{\theta} & \partial\hat{\mathbf{r}}/\partial\varphi = \sin\theta\,\hat{\boldsymbol{\varphi}} \\ \partial\hat{\boldsymbol{\theta}}/\partial\theta = -\hat{\mathbf{r}} & \partial\hat{\boldsymbol{\theta}}/\partial\varphi = \cos\theta\,\hat{\boldsymbol{\varphi}} \\ \partial\hat{\boldsymbol{\varphi}}/\partial\theta = 0 & \partial\hat{\boldsymbol{\varphi}}/\partial\varphi = -\sin\theta\,\hat{\mathbf{r}} - \cos\theta\,\hat{\boldsymbol{\theta}} \end{array} \tag{3.12}$$

3.2 Changing coordinate systems

Now that we have derived the properties of the unit vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ we are in the position to derive how the components $(u_r, u_\theta, u_\varphi)$ of the vector $\mathbf{u}$ defined in equation (3.5) are related to the usual Cartesian coordinates (u_x, u_y, u_z). This can most easily be achieved by writing the expressions (3.7) in the following form:

$$\begin{array}{l} \hat{\mathbf{r}} = \sin\theta\cos\varphi\,\hat{\mathbf{x}} + \sin\theta\sin\varphi\,\hat{\mathbf{y}} + \cos\theta\,\hat{\mathbf{z}} \\ \hat{\boldsymbol{\theta}} = \cos\theta\cos\varphi\,\hat{\mathbf{x}} + \cos\theta\sin\varphi\,\hat{\mathbf{y}} - \sin\theta\,\hat{\mathbf{z}} \\ \hat{\boldsymbol{\varphi}} = -\sin\varphi\,\hat{\mathbf{x}} + \cos\varphi\,\hat{\mathbf{y}} \end{array} \tag{3.13}$$

Problem a: Convince yourself that this expression can also be written in a symbolic form as

$$\begin{pmatrix} \hat{\mathbf{r}} \\ \hat{\boldsymbol{\theta}} \\ \hat{\boldsymbol{\varphi}} \end{pmatrix} = \mathbf{M} \begin{pmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{y}} \\ \hat{\mathbf{z}} \end{pmatrix} , \tag{3.14}$$

with the matrix $\mathbf{M}$ given by

$$\mathbf{M} = \begin{pmatrix} \sin\theta\cos\varphi & \sin\theta\sin\varphi & \cos\theta \\ \cos\theta\cos\varphi & \cos\theta\sin\varphi & -\sin\theta \\ -\sin\varphi & \cos\varphi & 0 \end{pmatrix} . \tag{3.15}$$

Of course expression (3.14) can only be considered to be a shorthand notation for the equations (3.13) since the entries in (3.14) are vectors rather than single components. However, expression (3.14) is a convenient shorthand notation.

The relation between the spherical components $(u_r, u_\theta, u_\varphi)$ and the Cartesian components (u_x, u_y, u_z) of the vector $\mathbf{u}$ can be obtained by inserting the expressions (3.13) for the spherical coordinate unit vectors in the relation $\mathbf{u} = u_r\hat{\mathbf{r}} + u_\theta\hat{\boldsymbol{\theta}} + u_\varphi\hat{\boldsymbol{\varphi}}$.

Problem b: Do this and collect all terms multiplying the unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ to show that expression (3.5) for the vector $\mathbf{u}$ is equivalent with:

$$\begin{array}{rcl} \mathbf{u} & = & (u_r\sin\theta\cos\varphi + u_\theta\cos\theta\cos\varphi - u_\varphi\sin\varphi)\,\hat{\mathbf{x}} \\ & + & (u_r\sin\theta\sin\varphi + u_\theta\cos\theta\sin\varphi + u_\varphi\cos\varphi)\,\hat{\mathbf{y}} \\ & + & (u_r\cos\theta - u_\theta\sin\theta)\,\hat{\mathbf{z}} \end{array} \tag{3.16}$$

Problem c: Show that this relation can also be written as:

$$\begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} = \mathbf{M}^T \begin{pmatrix} u_r \\ u_\theta \\ u_\varphi \end{pmatrix} . \tag{3.17}$$

In this expression, $\mathbf{M}^T$ is the transpose of the matrix $\mathbf{M}$: $M_{ij}^T = M_{ji}$, i.e. it is the matrix obtained by interchanging rows and columns of the matrix $\mathbf{M}$ given in (3.15).

We have not reached with equation (3.17) our goal yet of expressing the spherical coordinate components $(u_r, u_\theta, u_\varphi)$ of the vector $\mathbf{u}$ in the Cartesian components (u_x, u_y, u_z). This is most easily achieved by multiplying (3.17) with the inverse matrix $\left(\mathbf{M}^T\right)^{-1}$, which gives:

$$\left(\begin{array}{c} u_r \\ u_\theta \\ u_\varphi \end{array}\right) = \left(\mathbf{M}^T\right)^{-1} \left(\begin{array}{c} u_x \\ u_y \\ u_z \end{array}\right) . \tag{3.18}$$

However, now we have only shifted the problem because we don't know the inverse $\left(\mathbf{M}^T\right)^{-1}$. One could of course painstakingly compute this inverse, but this would be a laborious process that we can avoid. It follows by inspection of (3.15) that all the columns of $\mathbf{M}$ are of unit length and that the columns are orthogonal. This implies that $\mathbf{M}$ is an orthogonal matrix. Orthogonal matrices have the useful property that the transpose of the matrix is identical to the inverse of the matrix: $\mathbf{M}^{-1} = \mathbf{M}^T$.

Problem d: The property $\mathbf{M}^{-1} = \mathbf{M}^T$ can be verified explicitly by showing that $\mathbf{M}\mathbf{M}^T$ and $\mathbf{M}^T\mathbf{M}$ are equal to the identity matrix, do this!

Note that we have obtained the inverse of the matrix by making a guess and by verifying that this guess indeed solves our problem. This approach is often very useful in solving mathematical problems, there is nothing wrong with making a guess (as long as you check afterwards that your guess is indeed a solution to your problem). Since we know that $\mathbf{M}^{-1} = \mathbf{M}^T$, it follows that $\left(\mathbf{M}^T\right)^{-1} = \left(\mathbf{M}^{-1}\right)^{-1} = \mathbf{M}$.

Problem e: Use these results to show that the spherical coordinate components of $\mathbf{u}$ are related to the Cartesian coordinates by the following transformation rule:

$$\left(\begin{array}{c} u_r \\ u_\theta \\ u_\varphi \end{array}\right) = \left(\begin{array}{ccc} \sin\theta\cos\varphi & \sin\theta\sin\varphi & \cos\theta \\ \cos\theta\cos\varphi & \cos\theta\sin\varphi & -\sin\theta \\ -\sin\varphi & \cos\varphi & 0 \end{array}\right) \left(\begin{array}{c} u_x \\ u_y \\ u_z \end{array}\right) \tag{3.19}$$

3.3 The acceleration in spherical coordinates

You may wonder whether we really need all these transformation rules between a Cartesian coordinate system and a system of spherical coordinates. The answer is yes! An important example can be found in meteorology where air moves along a sphere. The velocity $\mathbf{v}$ of the air can be expressed in spherical coordinates:

$$\mathbf{v} = v_r\hat{\mathbf{r}} + v_\theta\hat{\boldsymbol{\theta}} + v_\varphi\hat{\boldsymbol{\varphi}} . \tag{3.20}$$

The motion of the air is governed by Newton's law, but when the velocity $\mathbf{v}$ and the force $\mathbf{F}$ are both expressed in spherical coordinates it would be wrong to express the θ-component of Newton's law as: $\rho dv_\theta/dt = F_\theta$. The reason is that the basis vectors of the spherical coordinate system depend on the position. When a particle moves, the direction of the basis vector change as well. This is a different way of saying that the spherical coordinate system is not an inertial system. When computing the acceleration in such a system additional terms appear that account for the fact that the coordinate system is not an inertial system. The results of the section (3.1) contains all the ingredients we need.

Let us follow a particle or air particle moving over a sphere, the position vector $\mathbf{r}$ has an obvious expansion in spherical coordinates:

$$\mathbf{r} = r\hat{\mathbf{r}} \ . \tag{3.21}$$

The velocity is obtained by taking the time-derivative of this expression. However, the unit vector $\hat{\mathbf{r}}$ is a function of the angles θ and φ, see equation (3.7). This means that when we take the time-derivative of (3.21) to obtain the velocity we need to differentiate $\hat{\mathbf{r}}$ as well with time. Note that this is not the case with the Cartesian expression $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$ because the unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ are constant, hence they do not change when the particle moves and they thus have a vanishing time-derivative.

An as example, let us compute the time derivative of $\hat{\mathbf{r}}$. This vector is a function of θ and φ, these angles both change with time as the particle moves. Using the chain rule it thus follows that:

$$\frac{d\hat{\mathbf{r}}}{dt} = \frac{d\hat{\mathbf{r}}(\theta,\varphi)}{dt} = \frac{d\theta}{dt}\frac{\partial\hat{\mathbf{r}}}{\partial\theta} + \frac{d\varphi}{dt}\frac{\partial\hat{\mathbf{r}}}{\partial\varphi} \ . \tag{3.22}$$

The derivatives $\partial\hat{\mathbf{r}}/\partial\theta$ and $\partial\hat{\mathbf{r}}/\partial\varphi$ can be eliminated with (3.12).

Problem a: Use the expressions (3.12) to eliminate the derivatives $\partial\hat{\mathbf{r}}/\partial\theta$ and $\partial\hat{\mathbf{r}}/\partial\varphi$ and carry out a similar analysis for the time-derivatives of the unit vectors $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ to show that:

$$\begin{aligned} \frac{d\hat{\mathbf{r}}}{dt} &= \dot{\theta}\ \hat{\boldsymbol{\theta}} + \sin\theta\ \dot{\varphi}\ \hat{\boldsymbol{\varphi}} \ , \\ \frac{d\hat{\boldsymbol{\theta}}}{dt} &= -\dot{\theta}\ \hat{\mathbf{r}} + \cos\theta\ \dot{\varphi}\ \hat{\boldsymbol{\varphi}} \ , \\ \frac{d\hat{\boldsymbol{\varphi}}}{dt} &= -\sin\theta\ \dot{\varphi}\ \hat{\mathbf{r}} - \cos\theta\ \dot{\varphi}\ \hat{\boldsymbol{\theta}} \ . \end{aligned} \tag{3.23}$$

In this expressions and other expressions in this section a dot is used to denote the time-derivative: $\dot{F} \equiv dF/dt$.

Problem b: Use the first line of (3.23) and the definition $\mathbf{v} = d\mathbf{r}/dt$ to show that in spherical coordinates:

$$\mathbf{v} = \dot{r}\hat{\mathbf{r}} + r\dot{\theta}\hat{\boldsymbol{\theta}} + r\sin\theta\ \dot{\varphi}\ \hat{\boldsymbol{\varphi}} \ . \tag{3.24}$$

In spherical coordinates the components of the velocity are thus given by:

$$\begin{aligned} v_r &= \dot{r} \\ v_\theta &= r\dot{\theta} \\ v_\varphi &= r\sin\theta\ \dot{\varphi} \end{aligned} \tag{3.25}$$

This result can be interpreted geometrically. As an example, let us consider the radial component of the velocity, see figure (3.2). To obtain the radial component of the velocity we keep the angles θ and φ fixed and let the radius $r(t)$ change to $r(t + \Delta t)$ over a time

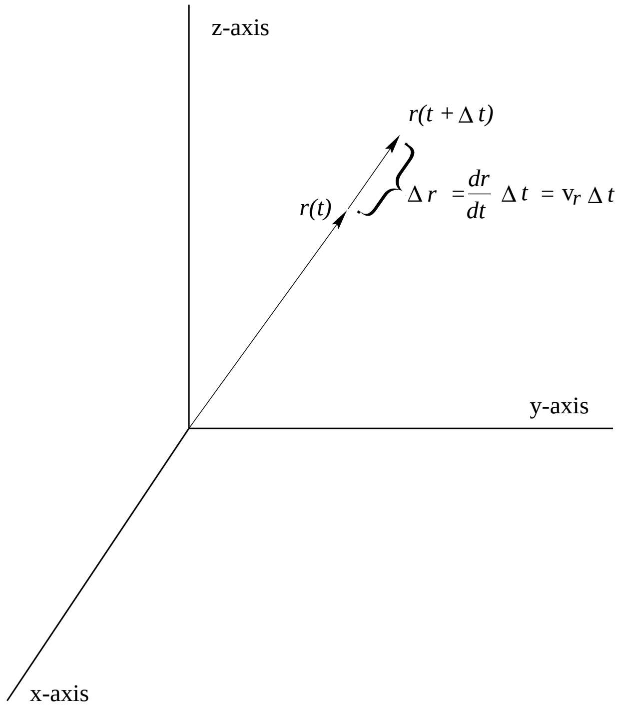


Figure 3.2: Definition of the geometric variables used to derive the radial component of the velocity.

Δt. The particle has moved a distance $r(t+\Delta t) - r(t) = dr/dt\ \Delta t$ in a time Δt, so that the radial component of the velocity is given by $v_r = dr/dt = \dot{r}$. This is the result given by the first line of (3.25).

Problem c: Use similar geometric arguments to explain the form of the velocity components v_θ and v_φ given in (3.25).

Problem d: We are now in the position to compute the acceleration is spherical coordinates. To do this differentiate (3.24) with respect to time and use expression (3.23) to eliminate the time-derivatives of the basis vectors. Use this to show that the acceleration **a** is given by:

$$\mathbf{a} = \left(\dot{v}_r - \dot{\theta} v_\theta - \sin\theta\ \dot{\varphi} v_\varphi\right)\hat{\mathbf{r}} + \left(\dot{v}_\theta + \dot{\theta} v_r - \cos\theta\ \dot{\varphi} v_\varphi\right)\hat{\boldsymbol{\theta}} + \left(\dot{v}_\varphi + \sin\theta\ \dot{\varphi} v_r + \cos\theta\ \dot{\varphi} v_\theta\right)\hat{\boldsymbol{\varphi}}\ . \tag{3.26}$$

Problem e: This expression is not quite satisfactory because it contains both the components of the velocity as well as the time-derivatives $\dot{\theta}$ and $\dot{\varphi}$ of the angles. Eliminate the time-derivatives with respect to the angles in favor of the components of the velocity using the expressions (3.25) to show that the components of the acceleration in spherical coordinates are given by:

$$a_r = \dot{v}_r - \frac{v_\theta^2 + v_\varphi^2}{r}$$

$$a_\theta = \dot{v}_\theta + \frac{v_r v_\theta}{r} - \frac{v_\varphi^2}{r \tan\theta} \tag{3.27}$$
$$a_\varphi = \dot{v}_\varphi + \frac{v_r v_\varphi}{r} + \frac{v_\theta v_\varphi}{r \tan\theta}$$

It thus follows that the components of the acceleration in a spherical coordinate system are not simply the time-derivative of the components of the velocity in that system. The reason for this is that the spherical coordinate system uses basis vectors that change when the particle moves. Expression (3.27) plays a crucial role in meteorology and oceanography where one describes the motion of the atmosphere or ocean [30]. Of course, in that application one should account for the Earth's rotation as well so that terms accounting for the Coriolis force and the centrifugal force need to be added, see section (10.4). It also should be added that the analysis of this section has been oversimplified when applied to the ocean or atmosphere because the advective terms $(\mathbf{v} \cdot \boldsymbol{\nabla})\,\mathbf{v}$ have not been taken into account. A complete treatment is given by *Holton*[30].

3.4 Volume integration in spherical coordinates

Carrying out a volume integral in Cartesian coordinates involves multiplying the function to be integrated by an infinitesimal volume element $dxdydz$ and integrating over all volume elements: $\iiint F dV = \iiint F(x,y,z) dxdydz$. Although this seems to be a simple procedure, it can be quite complex when the function F depends in a complex way on the coordinates (x,y,z) or when the limits of integration are not simple functions of x, y and z.

Problem a: Compute the volume of a sphere of radius R by taking $F = 1$ and integrating the volume integral in Cartesian coordinates over the volume of the sphere. Show first that in Cartesian coordinates the volume of the sphere can be written as

$$volume = \int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \int_{-\sqrt{R^2-x^2-y^2}}^{\sqrt{R^2-x^2-y^2}} dzdydx \ , \tag{3.28}$$

and carry out the integrations next.

After carrying out this exercise you probably have become convinced that using Cartesian coordinates is not the most efficient way to derive that the volume of a sphere with radius R is given by $4\pi R^3/3$. Using spherical coordinates appears to be the way to go, but for this one needs to be able to express an infinitesimal volume element dV in spherical coordinates. In doing this we will use that the volume spanned by three vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ is given by

$$volume = \det(\ \mathbf{a}\ ,\ \mathbf{b}\ ,\ \mathbf{c}\ ) = \begin{vmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ a_z & b_z & c_z \end{vmatrix} \ . \tag{3.29}$$

If we change the spherical coordinate θ with an increment $d\theta$, the position vector will change from $\mathbf{r}(r,\theta,\varphi)$ to $\mathbf{r}(r,\theta+d\theta,\varphi)$, this corresponds to a change $\mathbf{r}(r,\theta+d\theta,\varphi) - \mathbf{r}(r,\theta,\varphi) = \partial\mathbf{r}/\partial\theta\, d\theta$ in the position vector. Using the same reasoning for the variation of the position vector with r and φ it follows that the infinitesimal volume dV corresponding to changes increments dr, $d\theta$ and $d\varphi$ is given by

$$dV = \det\left(\frac{\partial \mathbf{r}}{\partial r} dr \ , \ \frac{\partial \mathbf{r}}{\partial \theta} d\theta \ , \ \frac{\partial \mathbf{r}}{\partial \varphi} d\varphi \ \right) . \tag{3.30}$$

Problem b: Show that this can be written as:

$$dV = \underbrace{\left| \begin{array}{ccc} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \varphi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \varphi} \end{array} \right|}_{J} dr d\theta d\varphi = J dr d\theta d\varphi \ . \tag{3.31}$$

The determinant J is called the *Jacobian*, the Jacobian is sometimes also written as:

$$J = \frac{\partial \left(x, y, z \right)}{\partial \left(r, \theta, \varphi \right)} \ , \tag{3.32}$$

but is should be kept in mind that this is nothing more than a new notation for the determinant in (3.31).

Problem c: Use the expressions (3.3) and (3.31) to show that

$$J = r^2 \sin \theta \ . \tag{3.33}$$

Note that the Jacobian J in (3.33) is the product of the scale factors defined in equation (3.8): $J = h_r h_\theta h_\varphi$. This is not a coincidence; in general the scale factors contain all the information needed to compute the Jacobian for a curvilinear coordinate system, see *Butkov*[14] for details.

Problem d: A volume element dV is in spherical coordinates thus given by $dV = r^2 \sin \theta \ dr d\theta d\varphi$. Consider the volume element dV in figure (3.3) that is defined by infinitesimal increments dr, $d\theta$ and $d\varphi$. Give an alternative derivation of this expression for dV that is based on geometric arguments only.

In some applications one wants to integrate over the surface of a sphere rather than integrating over a volume. For example, if one wants to compute the cooling of the Earth, one needs to integrate the heat flow over the Earth's surface. The treatment used for deriving the volume integral in spherical coordinates can also be used to derive the surface integral. A key element in the analysis is that the surface spanned by two vectors $\mathbf{a}$ and $\mathbf{b}$ is given by $|\mathbf{a} \times \mathbf{b}|$. Again, an increment $d\theta$ of the angle θ corresponds to a change $\partial \mathbf{r} / \partial \theta \, d\theta$ of the position vector. A similar result holds when the angle φ is changed.

Problem e: Use these results to show that the surface element dS corresponding to infinitesimal changes $d\theta$ and $d\varphi$ is given by

$$dS = \left| \frac{\partial \mathbf{r}}{\partial \theta} \times \frac{\partial \mathbf{r}}{\partial \varphi} \right| d\theta d\varphi \ . \tag{3.34}$$

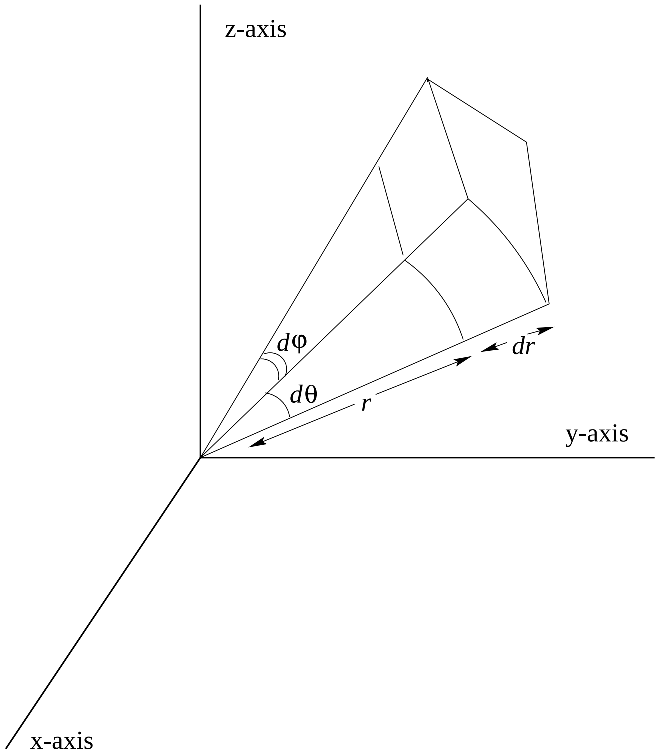


Figure 3.3: Definition of the geometric variables for an infinitesimal volume element dV.

Problem f: Use expression (3.3) to compute the vectors in the cross product and use this to derive that

$$dS = r^2 \sin\theta \; d\theta d\varphi \; . \tag{3.35}$$

Problem g: Using the geometric variables in figure (3.3) give an alternative derivation of this expression for a surface element that is based on geometric arguments only.

Problem h: Compute the volume of a sphere with radius R using spherical coordinates. Pay special attention to the range of integration for the angles θ and φ, see section (3.1).

3.5 Cylinder coordinates

Cylinder coordinates are useful in problems that exhibit cylinder symmetry rather than spherical symmetry. An example is the generation of water waves when a stone is thrown in a pond, or more importantly when an earthquake excites a tsunami in the ocean. In cylinder coordinates a point is specified by giving its distance $r = \sqrt{x^2 + y^2}$ to the z-axis, the angle φ and the z-coordinate, see figure (3.4) for the definition of variables. All the results we need could be derived using an analysis as shown in the previous sections. However, in such an approach we would do a large amount of unnecessary work. The key is to realize that at the equator of a spherical coordinate system (i.e. at the locations where $\theta = \pi/2$) the spherical coordinate system and the cylinder coordinate system are identical,

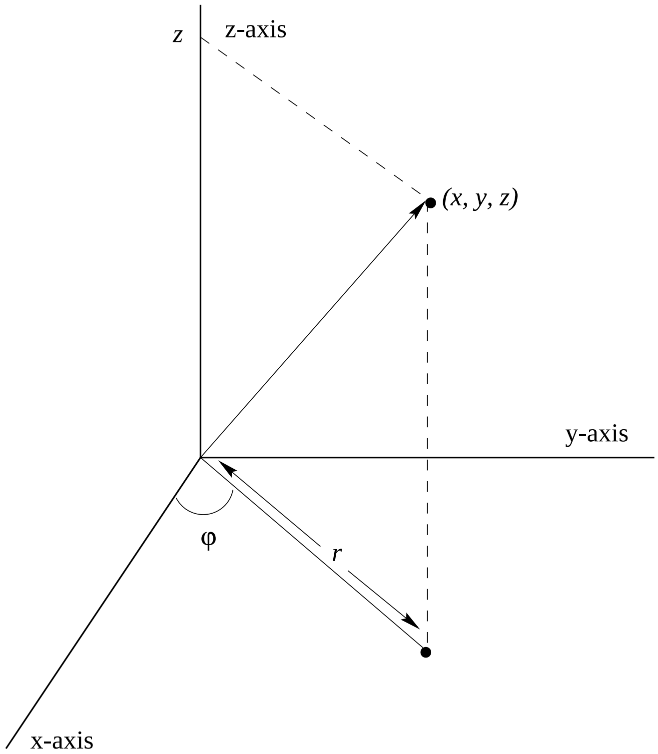


Figure 3.4: Definition of the geometric variables used in cylinder coordinates.

see figure (3.5). An inspection of this figure shows that all results obtained for spherical coordinates can be used for cylinder coordinates by making the following substitutions:

$$\begin{aligned} & r=\sqrt{x^2+y^2+z^2} \rightarrow \sqrt{x^2+y^2} \\ & \theta \rightarrow \pi/2 \\ & \hat{\boldsymbol{\theta}} \rightarrow -\hat{\mathbf{z}} \\ & r d\theta \rightarrow -dz \end{aligned} \tag{3.36}$$

Problem a: Convince yourself of this. To derive the third line consider the unit vectors pointing in the direction of increasing values of θ and z at the equator.

Problem b: Use the results of the previous sections and the substitutions (3.36) to show the following properties for a system of cylinder coordinates:

$$\begin{aligned} x &= r\cos\varphi \\ y &= r\sin\varphi \\ z &= z \end{aligned} \tag{3.37}$$

$$\hat{\mathbf{r}}=\begin{pmatrix}\cos\varphi\\ \sin\varphi\\ 0\end{pmatrix} \quad , \quad \hat{\boldsymbol{\varphi}}=\begin{pmatrix}-\sin\varphi\\ \cos\varphi\\ 0\end{pmatrix} \quad , \quad \hat{\mathbf{z}}=\begin{pmatrix}0\\ 0\\ 1\end{pmatrix} \quad , \tag{3.38}$$

$$dV = r dr d\varphi dz \; , \tag{3.39}$$

$$dS = r dz d\varphi \; . \tag{3.40}$$

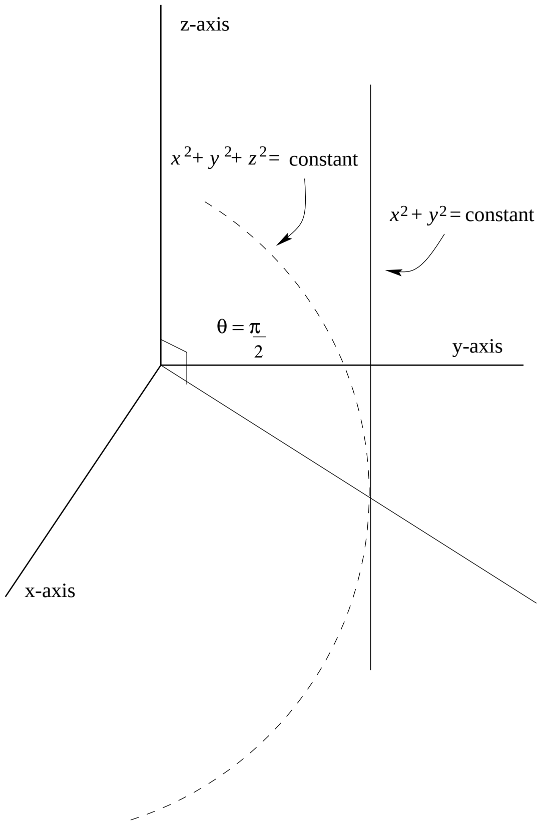


Figure 3.5: At the equator the spherical coordinate system has the same properties as a system of cylinder coordinates.

Problem c: Derive these properties directly using geometric arguments.

Chapter 4

The divergence of a vector field

The physical meaning of the divergence cannot be understood without understanding what the flux of a vector field is, and what the sources and sinks of a vector field are.

4.1 The flux of a vector field

To fix our mind, let us consider a vector field $\mathbf{v}(\mathbf{r})$ that represents the flow of a fluid that has a constant density. We define a surface S in this fluid. Of course the surface has an orientation in space, and the unit vector perpendicular to S is denoted by $\hat{\mathbf{n}}$. Infinitesimal elements of this surface are denoted with $d\mathbf{S} \equiv \hat{\mathbf{n}} dS$. Now suppose we are interested in the volume of fluid that flows per unit time through the surface S, this quantity is called Φ. When we want to know the flow through the surface, we only need to consider the component of $\mathbf{v}$ perpendicular to the surface, the flow along the surface is not relevant.

Problem a: Show that the component of the flow *across* the surface is given by $(\mathbf{v} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$ and that the flow *along* the surface is given by $\mathbf{v} - (\mathbf{v} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$. If you find this problem difficult you may want to look ahead in section (10.1).

Using this result the volume of the flow through the surface per unit time is given by:

$$\Phi = \iint (\mathbf{v} \cdot \hat{\mathbf{n}}) dS = \iint \mathbf{v} \cdot d\mathbf{S} \; , \tag{4.1}$$

this expression defines the flux Φ of the vector field $\mathbf{v}$ through the surface S. The definition of a flux is not restricted to the flow of fluids, a flux can be computed for any vector field. However, the analogy of fluid flow often is very useful to understand the meaning of the flux and divergence.

Problem b: The electric field generated by a point charge q in the origin is given by

$$\mathbf{E}(\mathbf{r}) = \frac{q\hat{\mathbf{r}}}{4\pi\varepsilon_0 r^2}, \tag{4.2}$$

in this expression $\hat{\mathbf{r}}$ is the unit vector in the radial direction and ε_0 is the permittivity. Compute the flux of the electric field through a spherical surface with radius R with the point charge in its center. Show explicitly that this flux is independent of the radius R and find its relation to the charge q and the permittivity ε_0. Choose the coordinate system you use for the integration carefully.

Problem c: To first order the magnetic field of the Earth is a dipole field. (This is the field generated by a magnetic north pole and magnetic south pole very close together.) The dipole vector $\mathbf{m}$ points from the south pole of the dipole to the north pole and its size is given by the strength of the dipole. The magnetic field $\mathbf{B}(\mathbf{r})$ is given by (ref. [31], p. 182):

$$\mathbf{B}(\mathbf{r}) = \frac{3\hat{\mathbf{r}}(\hat{\mathbf{r}} \cdot \mathbf{m}) - \mathbf{m}}{r^3}. \tag{4.3}$$

Compute the flux of the magnetic field through the surface of the Earth, take a sphere with radius R for this. Hint, when you select a coordinate system, think not only about the geometry of the coordinate system (i.e. Cartesian or spherical coordinates), but also choose the *direction* of the axes of your coordinate system with care.

4.2 Introduction of the divergence

In order to introduce the divergence, consider an infinitesimal rectangular volume with sides dx, dy and dz, see fig (4.1) for the definition of the geometric variables. The

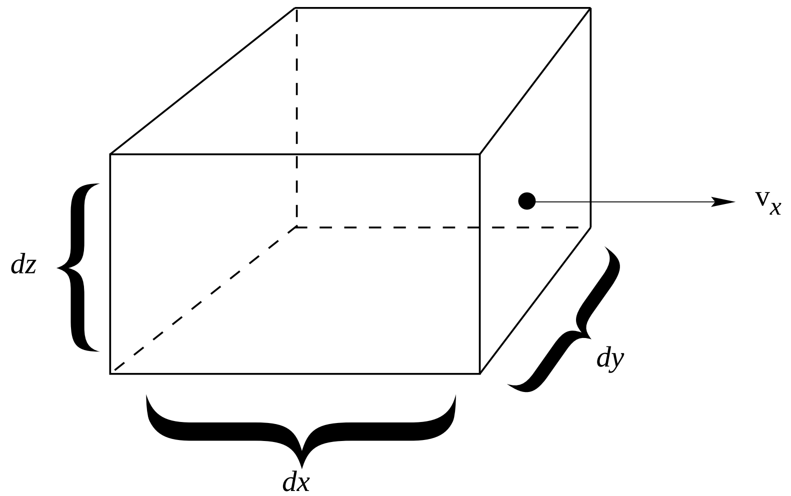


Figure 4.1: Definition of the geometric variables in the calculation of the flux of a vector field through an infinitesimal rectangular volume.

outward flux through the right surface perpendicular through the x-axis is given by $v_x(x + dx, y, z)dydz$, because $v_x(x + dx, y, z)$ is the component of the flow perpendicular to that surface and $dydz$ is the area of the surface. By the same token, the flux through the left surface perpendicular through the x-axis is given by $-v_x(x, y, z)dydz$, the $-$ sign is due to the fact the component of $\mathbf{v}$ in the direction *outward* of the cube is given by $-v_x$. (Alternatively one can say that for this surface the unit vector perpendicular to the surface and pointing outwards is given by $\hat{\mathbf{n}} = -\hat{\mathbf{x}}$.) This means that the total outward flux through the two surfaces is given by $v_x(x + dx, y, z)dydz - v_x(x, y, z)dydz = \frac{\partial v_x}{\partial x}dxdydz$. The same reasoning applies to the surfaces perpendicular to the y- and z-axes. This means that the total outward flux through the sides of the cubes is:

$$d\Phi = \left(\frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right) dV = (\nabla \cdot \mathbf{v})\, dV, \tag{4.4}$$

where dV is the volume $dxdydz$ of the cube and $(\nabla \cdot \mathbf{v})$ is the *divergence* of the vector field $\mathbf{v}$.

The above definition does not really tell us yet what the divergence really is. Dividing (4.4) by dV one obtains $(\nabla \cdot \mathbf{v}) = d\Phi/dV$. This allows us to state in words what the divergence is:

> *The divergence of a vector field is the outward flux of the vector field per unit volume.*

To fix our mind again let us consider a physical example where in two dimensions fluid is pumped into this two dimensional space at location $\mathbf{r} = 0$. For simplicity we assume that the fluid is incompressible, that means that the mass-density is constant. We do not know yet what the resulting flow field is, but we know two things. Away from the source at $\mathbf{r} = 0$ there are no sources or sinks of fluid flow. This means that the flux of the flow through any closed surface S must be zero. ("What goes in must come out.") This means that the divergence of the flow is zero, except possibly near the source at $\mathbf{r} = 0$:

$$(\nabla \cdot \mathbf{v}) = 0 \qquad for \qquad \mathbf{r} \neq 0. \tag{4.5}$$

In addition we know that due to the symmetry of the problem the flow is directed in the radial direction and depends on the radius r only:

$$\mathbf{v}(\mathbf{r}) = f(r)\mathbf{r}. \tag{4.6}$$

Problem a: Show this.

This is enough information to determine the flow field. Of course, it is a problem that we cannot immediately insert (4.6) in (4.5) because we have not yet derived an expression for the divergence in cylinder coordinates. However, there is another way to determine the flow from the expression above.

Problem b: Using that $r = \sqrt{x^2 + y^2}$ show that

$$\frac{\partial r}{\partial x} = \frac{x}{r}, \tag{4.7}$$

and derive the corresponding equation for y. Using expressions (4.6), (4.7) and the chain rule for differentiation show that

$$\nabla \cdot \mathbf{v} = 2f(r) + r\frac{df}{dr} \qquad (cilinder\ coordinates). \tag{4.8}$$

Problem c: Insert this result in (4.5) and show that the flow field is given by $\mathbf{v}(\mathbf{r}) = A\mathbf{r}/r^2$. Make a sketch of the flow field.

The constant A is yet to be determined. Let at the source $\mathbf{r} = 0$ a volume V per unit time be injected.

Problem d: Show that $V = \int \mathbf{v} \cdot d\mathbf{S}$ (where the integration is over an arbitrary surface around the source at $\mathbf{r} = 0$). By choosing a suitable surface derive that

$$\mathbf{v}(\mathbf{r}) = \frac{V}{2\pi} \frac{\hat{\mathbf{r}}}{r}. \tag{4.9}$$

From this simple example of a single source at $\mathbf{r} = 0$ more complex examples can be obtained. Suppose we have a source at $\mathbf{r}_+ = (L, 0)$ where a volume V is injected per unit time and a sink at $\mathbf{r}_- = (-L, 0)$ where a volume $-V$ is removed per unit time. The total flow field can be obtained by superposition of flow fields of the form (4.9) for the source and the sink.

Problem e: Show that the x- and y-components of the flow field in this case are given by:

$$v_x(x,y) = \frac{V}{2\pi} \left(\frac{x-L}{(x-L)^2 + y^2} - \frac{x+L}{(x+L)^2 + y^2} \right), \tag{4.10}$$

$$v_y(x,y) = \frac{V}{2\pi} \left(\frac{y}{(x-L)^2 + y^2} - \frac{y}{(x+L)^2 + y^2} \right), \tag{4.11}$$

and sketch the resulting flow field. This is most easily accomplished by determining from the expressions above the flow field at some selected lines such as the x- and y-axes.

One may also be interested in computing the *streamlines* of the flow. These are the lines along which material particles flow. The streamlines can be found by using the fact that the time derivative of the position of a material particle is the velocity: $d\mathbf{r}/dt = \mathbf{v}(\mathbf{r})$. Inserting expressions (4.10) and (4.11) leads to two coupled differential equations for $x(t)$ and $y(t)$ which are difficult to solve. Fortunately, there are more intelligent ways of retrieving the streamlines. We will return to this issue in section (12.3).

4.3 Sources and sinks

In the example of the fluid flow given above the fluid flow moves away from the source and converges on the sink of the fluid flow. The terms "source" and "sink" have a clear physical meaning since they are directly related to the "source" of water as from a tap, and a "sink" as the sink in a bathtub. The flow lines of the water flow diverge from the source while they convergence towards the sinks. This explains the term "divergence", because this quantity simply indicates to what extent flow lines originate (in case of a source) or end (in case of a sink).

This definition of sources and sinks is not restricted to fluid flow. For example, for the electric field the term "fluid flow" should be replaced by the term "field lines." Electrical field lines originate at positive charges and end at negative charges.

Problem a: To verify this, show that the divergence of the electrical field (4.2) for a point charge in three dimensions vanishes except near the point charge at $\mathbf{r} = 0$. Show also that the net flux through a small sphere surrounding the charge is positive (negative) when the charge q is positive (negative).

The result we have just discovered is that the electric charge is the source of the electric field. This is reflected in the Maxwell equation for the divergence of the electric field:

$$(\nabla \cdot \mathbf{E}) = \rho(\mathbf{r})/\varepsilon_0. \tag{4.12}$$

In this expression $\rho(\mathbf{r})$ is the charge density, this is simply the electric charge per unit volume just as the mass-density denotes the mass per unit volume. In addition, expression (4.12) contains the permittivity ε_0. This term serves as a *coupling constant* since it describes how "much" electrical field is generated by a given electrical charge density. It is obvious that a constant is needed here because the charge density and the electrical field have different physical dimensions, hence a proportionality factor must be present. However, the physical meaning of a coupling constant goes much deeper, because it prescribes how strong the field is that is generated by a given source. This constant describes how strong cause (the source) and effect (the field) are coupled.

Problem b: Show that the divergence of the magnetic field (4.3) for a dipole $\mathbf{m}$ at the origin is zero everywhere, *including* the location of the dipole.

By analogy with (4.12) one might expect that the divergence of the magnetic field is related to a magnetic charge density: $(\nabla \cdot \mathbf{B}) =$*coupling const.* $\rho_B(\mathbf{r})$, where ρ_B would be the "density of magnetic charge." However, particles with a magnetic charge (usually called "magnetic monopoles") have not been found in nature despite extensive searches. Therefore the Maxwell equation for the divergence of the magnetic field is:

$$(\nabla \cdot \mathbf{B}) = 0, \tag{4.13}$$

but we should remember that this divergence is zero because of the observational absence of magnetic monopoles rather than a vanishing coupling constant.

4.4 The divergence in cylinder coordinates

In the previous analysis we have only used the expression of the divergence is Cartesian coordinates: $\nabla \cdot \mathbf{v} = \partial_x v_x + \partial_y v_y + \partial_z v_z$. As you have (hopefully) discovered, the use of other coordinate systems such as cylinder coordinates or spherical coordinates can make life much simpler. Here we derive an expression for the divergence in cylinder coordinates. In this system, the distance $r = \sqrt{x^2 + y^2}$ of a point to the z-axis, the azimuth $\varphi (= \arctan(y/x)\ )$ and z are used as coordinates, see section (3.5). A vector $\mathbf{v}$ can be decomposed in components in this coordinate system:

$$\mathbf{v} = v_r \hat{\mathbf{r}} + v_\varphi \hat{\boldsymbol{\varphi}} + v_z \hat{\mathbf{z}} \tag{4.14}$$

where $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\varphi}}$ and $\hat{\mathbf{z}}$ are unit vectors in the direction of increasing values of r, φ and z respectively. As shown in section (4.2) the divergence is the flux per unit volume. Let

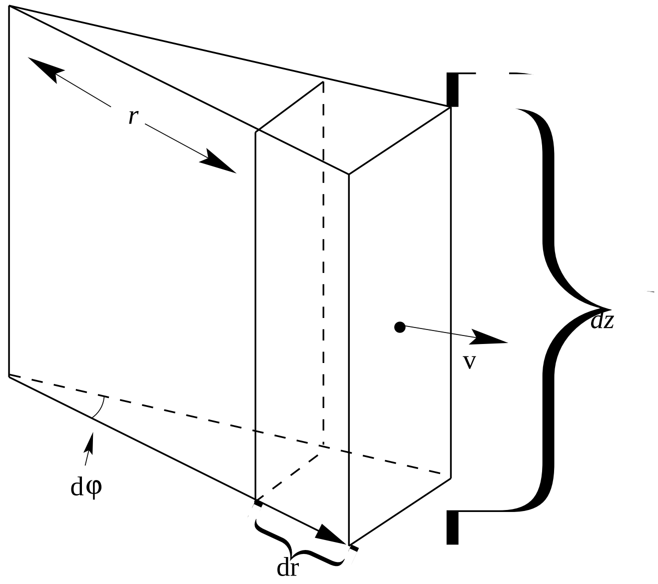


Figure 4.2: Definition of the geometric variables for the computation of the divergence in cylinder coordinates.

us consider the infinitesimal volume corresponding to increments dr, $d\varphi$ and dz shown in figure (4.2). Let us first consider the flux of $\mathbf{v}$ through the surface elements perpendicular to $\hat{\mathbf{r}}$. The size of this surface is $rd\varphi dz$ and $(r+dr)d\varphi dz$ respectively at r and $r+dr$. The normal components of $\mathbf{v}$ through these surfaces are $v_r(r,\varphi,z)$ and $v_r(r+dr,\varphi,z)$ respectively. Hence the total flux through these two surface is given by $v_r(r+dr,\varphi,z)(r+dr)d\varphi dz - v_r(r,\varphi,z)(r)d\varphi dz$.

Problem a: Show that to first order in dr this quantity is equal to $\frac{\partial}{\partial r}(rv_r)\,drd\varphi dz$. Hint, use a first order Taylor expansion for $v_r(r+dr,\varphi,z)$ in the quantity dr.

Problem b: Show that the flux through the surfaces perpendicular to $\hat{\boldsymbol{\varphi}}$ is to first order in $d\varphi$ given by $\frac{\partial v_\varphi}{\partial \varphi}drd\varphi dz$.

Problem c: Show that the flux through the surfaces perpendicular to $\hat{\mathbf{z}}$ is to first order in dz given by $\frac{\partial v_z}{\partial z}rdrd\varphi dz$.

The volume of the infinitesimal part of space shown in figure (4.2) is given by $rdrd\varphi dz$.

Problem d: Use the fact that the divergence is the flux per unit volume to show that in cylinder coordinates:

$$\nabla\cdot\mathbf{v} = \frac{1}{r}\frac{\partial}{\partial r}(rv_r) + \frac{1}{r}\frac{\partial v_\varphi}{\partial\varphi} + \frac{\partial v_z}{\partial z}\ . \tag{4.15}$$

Problem e: Use this result to re-derive equation (4.8) without using Cartesian coordinates as an intermediary.

In spherical coordinates a vector $\mathbf{v}$ can be expended in the components v_r, v_θ and v_φ in the directions of increasing values of r, θ and φ respectively. In this coordinate system r has a different meaning than in cylinder coordinates because in spherical coordinates $r = \sqrt{x^2 + y^2 + z^2}$.

Problem f: Show that in spherical coordinates

$$\nabla \cdot \mathbf{v} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 v_r\right) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\; v_\theta\right) + \frac{1}{r\sin\theta}\frac{\partial v_\varphi}{\partial\varphi} \tag{4.16}$$

4.5 Is life possible in a 5-dimensional world?

In this section we will investigate whether the motion of the earth around the sun is stable or not. This means that we ask ourselves the question that when the position of the earth a perturbed, for example by the gravitational attraction of the other planets or by a passing asteroid, whether the gravitational force brings the earth back to its original position (stability) or whether the earth spirals away from the sun (or towards the sun). It turns out that the stability properties depend on the spatial dimension! We know that we live in a world of three spatial dimensions, but it is interesting to investigate if the orbit of the earth would also be stable in a world with a different number of spatial dimensions.

In the Newtonian theory the gravitational field $\mathbf{g}(\mathbf{r})$ satisfies (see ref: [42]):

$$(\nabla \cdot \mathbf{g}) = -4\pi G\rho \ , \tag{4.17}$$

where $\rho(\mathbf{r})$ is the mass density and G is the gravitational constant which has a value of $6.67 \times 10^{-8}\ cm^3 g^{-1} s^{-2}$. The term G plays the role of a coupling constant, just as the 1/permittivity in (4.12). Note that the right hand side of the gravitational field equation (4.17) has an opposite sign as the right hand side of the electric field equation (4.12). This is due to the fact that two electric charges of equal sign repel each other, while two masses of equal sign (mass being positive) attract each other. If the sign of the right hand side of (4.17) would be positive, masses would repel each other and structures such as planets, the solar system and stellar systems would not exist.

Problem a: We argued in section (4.3) that electric field lines start at positive charges and end at negative charges. By analogy we expect that gravitational field lines end at the (positive) masses that generate the field. However, where do the gravitational field lines start?

Let us first determine the gravitational field of the sun in N dimensions. Outside the sun the mass-density vanishes, this means that $(\nabla \cdot \mathbf{g}) = 0$. We assume that the mass density in the sun is spherically symmetric, the gravitational field must be spherically symmetric too and is thus of the form:

$$\mathbf{g}(\mathbf{r}) = f(r)\mathbf{r} \ . \tag{4.18}$$

In order to make further progress we must derive the divergence of a spherically symmetric vector field in N dimensions. Generalizing expression (4.16) to an arbitrary number of dimensions is not trivial, but fortunately this is not needed. We will make use of the property that in N dimensions: $r = \sqrt{\sum_{i=1}^N x_i^2}$.

Problem b: Derive from this expression that

$$\partial r/\partial x_j = x_j/r \ . \tag{4.19}$$

Use this result to derive that for a vector field of the form (4.18):

$$(\nabla \cdot \mathbf{g}) = N f(r) + r\frac{\partial f}{\partial r} \ . \tag{4.20}$$

Outside the sun, where the mass-density vanishes and $(\nabla \cdot \mathbf{g}) = 0$ we can use this result to solve for the gravitational field.

Problem c: Derive that

$$\mathbf{g}(\mathbf{r}) = -\frac{A}{r^{N-1}}\hat{\mathbf{r}} \ , \tag{4.21}$$

and check this result for three spatial dimensions.

At this point the constant A is not determined, but this is not important for the coming arguments. The minus sign is added for convenience, the gravitational field points towards the sun hence $A > 0$.

Associated with the gravitational field is a gravitational force that attracts the earth towards the sun. If the mass of the earth is denoted by m, this force is given by

$$\mathbf{F}_{grav} = -\frac{Am}{r^{N-1}}\hat{\mathbf{r}} \ , \tag{4.22}$$

and is directed towards the sun. For simplicity we assume that the earth is in a circular orbit. This means that the attractive gravitational force is balanced by the repulsive centrifugal force which is given by

$$\mathbf{F}_{cent} = \frac{mv^2}{r}\hat{\mathbf{r}} \ . \tag{4.23}$$

In equilibrium these forces balance: $\mathbf{F}_{grav} + \mathbf{F}_{cent} = 0$.

Problem d: Derive the velocity v from this requirement.

We now assume that the distance to the sun is perturbed from its original distance r to a new distance $r + \delta r$, the perturbation in the position is therefore $\delta\mathbf{r} = \delta r \ \hat{\mathbf{r}}$. Because of this perturbation, the gravitational force and the centrifugal force are perturbed too, these quantities will be denoted by $\delta\mathbf{F}_{grav}$ and $\delta\mathbf{F}_{cent}$ respectively, see figure (4.3).

Problem e: Show that the earth moves back to its original position when:

$$(\delta\mathbf{F}_{grav} + \delta\mathbf{F}_{cent}) \cdot \delta\mathbf{r} < 0 \qquad (stability) \ . \tag{4.24}$$

Hint: consider the case where the radius is increased ($\delta r > 0$) and decreased ($\delta r < 0$) separately.

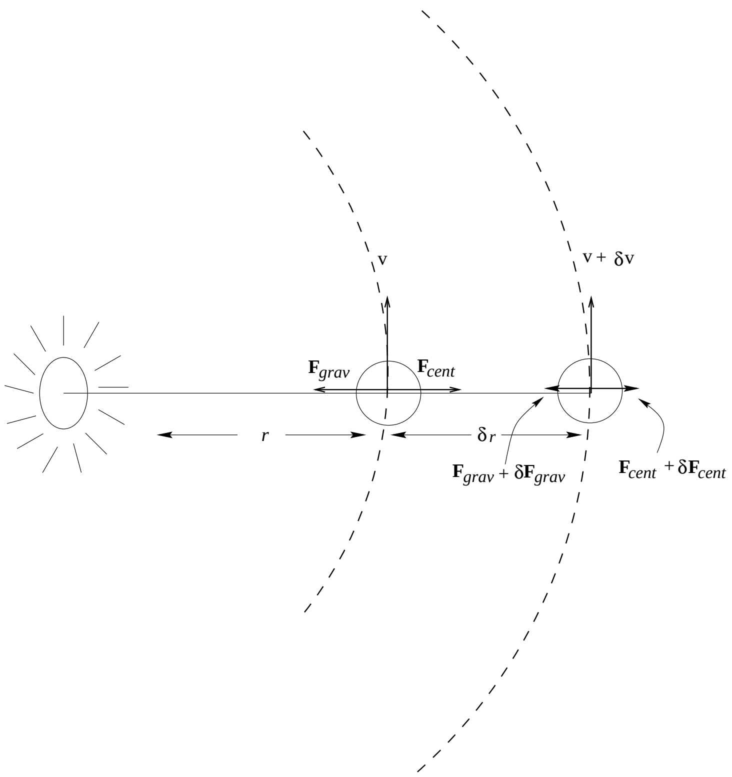


Figure 4.3: Definition of variables for the perturbed orbit of the earth.

Hence the orbital motion is stable for perturbations when the gravitational field satisfies the criterion (4.24). In order to compute the change in the centrifugal force we use that angular momentum is conserved, i.e. $mrv = m(r + \delta r)(v + \delta v)$. In what follows we will consider small perturbations and will retain only terms of first order in the perturbation. This means that we will ignore higher order terms such as the product $\delta r \delta v$.

Problem f: Determine δv and derive that

$$\delta \mathbf{F}_{cent} = -\frac{3mv^2}{r^2}\delta \mathbf{r} \ , \tag{4.25}$$

and use (4.22) to show that

$$\delta \mathbf{F}_{grav} = (N-1)\ \frac{Am}{r^N}\delta \mathbf{r} \tag{4.26}$$

Note that the perturbation of the centrifugal force does not depend on the number of spatial dimensions, but that the perturbation of the gravitational force does depend on N.

Problem g: Using the value of the velocity derived in **problem d** and expressions (4.25)-(4.26) show that according to the criterion (4.24) the orbital motion is stable in less than four spatial dimensions. Show also that the requirement for stability is independent of the original distance r.

This is a very interesting result. It implies that orbital motion is unstable in more than four spatial dimensions. This means that in a world with five spatial dimensions the solar system would not be stable. Life seems to be tied to planetary systems with a central star which supplies the energy to sustain life on the orbiting planet(s). This implies that life would be impossible in a five-dimensional world! Note also that the stability requirement is independent of r, i.e. the stability properties of orbital motion does not depend on the size of the orbit. This implies that the gravitational field does not have "stable regions" and "unstable regions", the stability property depends only on the number of spatial dimensions.

Chapter 5

The curl of a vector field

5.1 Introduction of the curl

We will introduce the *curl* of a vector field $\mathbf{v}$ by its formal definition in terms of Cartesian coordinates (x, y, z) and unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ in the x, y and z-direction respectively:

$$curl\ \mathbf{v} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \partial_x & \partial_y & \partial_z \\ v_x & v_y & v_z \end{vmatrix} = \begin{pmatrix} \partial_y v_z - \partial_z v_y \\ \partial_z v_x - \partial_x v_z \\ \partial_x v_y - \partial_y v_x \end{pmatrix} . \tag{5.1}$$

It can be seen that the *curl* is a *vector*, this is in contrast to the divergence which is a scalar. The notation with the determinant is of course incorrect because the entries in a determinant should be numbers rather than vectors such as $\hat{\mathbf{x}}$ or differentiation operators such as $\partial_y = \partial/\partial y$. However, the notation in terms of a determinant is a simple rule to remember the definition of the *curl* in Cartesian coordinates. We will write the *curl* of a vector field also as: $curl\ \mathbf{v} = \nabla \times \mathbf{v}$.

Problem a: Verify that this notation with the curl expressed as the outer product of the operator ∇ and the vector $\mathbf{v}$ is consistent with the definition (5.1).

In general the *curl* is a three-dimensional vector. To see the physical interpretation of the *curl*, we will make life easy for ourselves by choosing a Cartesian coordinate system where the z-axis is aligned with $curl\ \mathbf{v}$. In that coordinate system the *curl* is given by: $curl\ \mathbf{v} = (\partial_x v_y - \partial_y v_x)\hat{\mathbf{z}}$. Consider a little rectangular surface element oriented perpendicular to the z-axis with sides dx and dy respectively, see figure (5.1). We will consider the line integral $\oint_{dxdy} \mathbf{v} \cdot d\mathbf{r}$ along a closed loop defined by the sides of this surface element integrating in the counter-clockwise direction. This line integral can be written as the sum of the integral over the four sides of the surface element.

Problem b: Show that the line integral is given by $\oint_{dxdy} \mathbf{v} \cdot d\mathbf{r} = v_x(x, y)dx + v_y(x + dx, y)dy - v_x(x, y + dy)dx - v_y(x, y)dy$, and use a first order Taylor expansion to write this as

$$\oint_{dxdy} \mathbf{v} \cdot d\mathbf{r} = (\partial_x v_y - \partial_y v_x)dxdy . \tag{5.2}$$

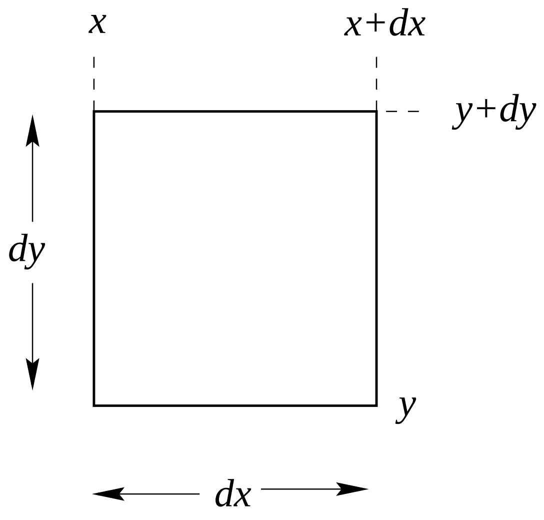


Figure 5.1: Definition of the geometric variables for the interpretation of the curl.

This expression can be rewritten as:

$$(curl\ \mathbf{v})_z = (\partial_x v_y - \partial_y v_x) = \frac{\oint_{dxdy} \mathbf{v} \cdot d\mathbf{r}}{dxdy} \ . \tag{5.3}$$

In this form we can express the meaning of the *curl* in words:

The curl of $\mathbf{v}$ *is the closed line integral of* $\mathbf{v}$ *per unit surface area.*

Note that this interpretation is similar to the interpretation of the divergence given in section (4.2). There is, however, one major difference. The *curl* is a vector while the divergence is a scalar. This is reflected in our interpretation of the curl because a surface has an orientation defined by its normal vector, hence the curl is a vector too.

5.2 What is the curl of the vector field?

In order to discover the meaning of the curl, we will consider again an incompressible fluid and will consider the *curl* of the velocity vector $\mathbf{v}$, because this will allow us to discover when the *curl* is nonzero. It is not only for a didactic purpose that we consider the *curl* of fluid flow. In fluid mechanics this quantity plays such a crucial role that it is given a special name, the *vorticity* $\boldsymbol{\omega}$:

$$\boldsymbol{\omega} \equiv \boldsymbol{\nabla} \times \mathbf{v} \ . \tag{5.4}$$

To simplify things further we assume that the fluid moves in the x, y-plane only (i.e. $v_z = 0$) and that the flow depends only on x and y: $\mathbf{v} = \mathbf{v}(x, y)$.

Problem a: Show that for such a flow

$$\boldsymbol{\omega} = \boldsymbol{\nabla} \times \mathbf{v} = (\partial_x v_y - \partial_y v_x)\hat{\mathbf{z}} \ . \tag{5.5}$$

We will first consider an axi-symmetric flow field. Such a flow field has rotation symmetry around an axis, we will take the z-axis for this. Because of the cylinder symmetry and the fact that it is assumed that the flow does not depend on z, the components v_r, v_φ and v_z depend neither on the azimuth φ $(= \arctan y/x)$ used in the cylinder coordinates nor on z but only on the distance $r = \sqrt{x^2+y^2}$ to the z-axis.

Problem b: Show that it follows from expression (4.15) for the divergence in cylinder coordinates that for an axisymmetric flow field for an incompressible fluid (where $(\nabla \cdot \mathbf{v})=0$ everywhere including the z-axis where $r = \sqrt{x^2+y^2} = 0$) that the radial component of the velocity must vanish: $v_r = 0$.

This result simply reflects that for an incompressible flow with cylinder symmetry there can be no net flow towards (or away from) the symmetry axis. The only nonzero component of the flow is therefore in the direction of $\hat{\boldsymbol{\varphi}}$. This implies that the velocity field must be of the form:

$$\mathbf{v} = \hat{\boldsymbol{\varphi}} v(r) \ , \tag{5.6}$$

see figure (5.2) for a sketch of this flow field. The problem we now face is that

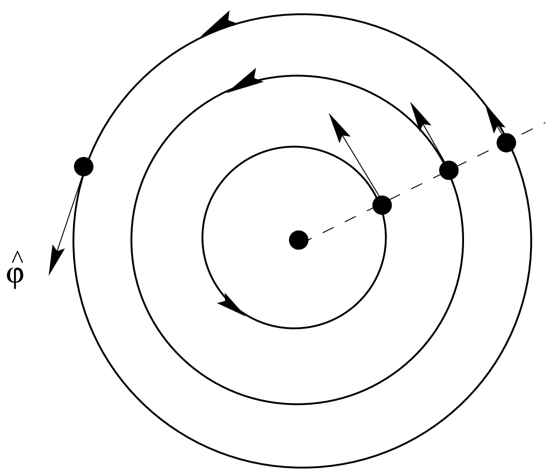


Figure 5.2: Sketch of an axi-symmetric source-free flow in the x,y-plane.

definition (5.1) is expressed in Cartesian coordinates while the velocity in equation (5.6) is expressed in cylinder coordinates. In section (5.6) an expression for the curl in cylinder coordinates will be derived. As an alternative, one can express the unit vector $\hat{\boldsymbol{\varphi}}$ in Cartesian coordinates.

Problem c: Verify that:

$$\hat{\boldsymbol{\varphi}} = \begin{pmatrix} -y/r \\ x/r \\ 0 \end{pmatrix} \ . \tag{5.7}$$

Hints, make a figure of this vector in the x, y-plane, verify that this vector is perpendicular to the position vector $\mathbf{r}$ and that it is of unit length. Alternatively you can use expression (3.36) of section (3.5).

Problem d: Use the expressions (5.5), (5.7) and the chain rule for differentiation to show that for the flow field (5.6):

$$(\nabla \times \mathbf{v})_z = \frac{\partial v}{\partial r} + \frac{v}{r} \ . \tag{5.8}$$

Hint, you have to use the derivatives $\partial r/\partial x$ and $\partial r/\partial y$ again. You have learned this in section (4.2).

5.3 The first source of vorticity; rigid rotation

In general, a nonzero curl of a vector field can have two origins, in this section we will treat the effect of rigid rotation. Because we will use fluid flow as an example we will speak about the vorticity, but keep in mind that the results of this section (and the next) apply to any vector field. We will consider a velocity field that describes a rigid rotation with the z-axis as rotation axis and angular velocity Ω.

Problem a: Show that the associated velocity field is of the form (5.6) with $v(r) = \Omega r$. Verify explicitly that every particle in the flow makes one revolution in a time $T = 2\pi/\Omega$ and that this time does not depend on the position of the particle.

Problem b: Show that for this velocity field: $\nabla \times \mathbf{v} = 2\Omega\hat{\mathbf{z}}$.

This means that the vorticity is twice the rotation vector $\Omega\hat{\mathbf{z}}$. This result is derived here for the special case that the z-axis is the axis of rotation. (This can always be achieved because one is free in the choice of the orientation of the coordinate system.) In section (6.11) of *Boas*[11] it is shown with a very different derivation that the vorticity for rigid rotation is given by $\boldsymbol{\omega} = \nabla \times \mathbf{v} = 2\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is the rotation vector. (Beware, the notation used by *Boas* is different from ours in a deceptive way!)

We see that rigid rotation leads to a vorticity that is twice the rotation rate. Imagine we place a paddle-wheel in the flow field that is associated with the rigid rotation, see figure (5.3). This paddle-wheel moves with the flow and makes one revolution along its axis in a time $2\pi/\Omega$. Note also that for the sense of rotation shown in figure (5.3) the paddle wheel moves in the counterclockwise direction and that the curl points along the positive z-axis. This implies that the rotation of the paddle-wheel not only denotes that the curl is nonzero, the rotation vector of the paddle is directed along the *curl*! This actually explains the origin of the word *vorticity*. In a vortex, the flow rotates around a rotation axis. The *curl* increases with the rotation rate, hence it increases with the strength of the vortex. This strength of the vortex has been dubbed *vorticity,* and this term therefore reflects the fact that the curl of velocity denotes the (local) intensity of rotation in the flow.

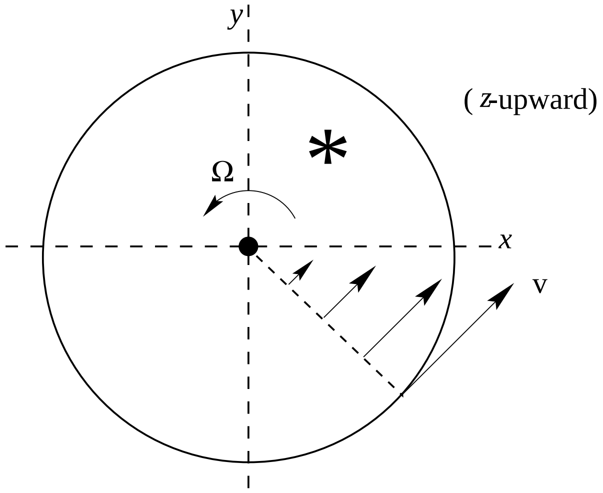


Figure 5.3: The vorticity for a rigid rotation.

5.4 The second source of vorticity; shear

In addition to rigid rotation, shear is another cause of vorticity. In order to see this we consider a fluid in which the flow is only in the x-direction and where the flow depends on the y-coordinate only: $v_y = v_z = 0$, $v_x = f(y)$.

Problem a: Show that this flow does not describe a rigid rotation. Hint: how long does it take before a fluid particle returns to its original position?

Problem b: Show that for this flow

$$\nabla \times \mathbf{v} = -\frac{\partial f}{\partial y}\hat{\mathbf{z}} \ . \tag{5.9}$$

As a special example consider the velocity given by:

$$v_x = f(y) = v_0 \exp\left(-y^2/L^2\right) \ . \tag{5.10}$$

This flow field is sketched in figure (5.4).

Problem c: Verify for yourself that paddle-wheels placed in the flow rotate in the sense indicated in figure (5.4)

Problem d: Compute $\nabla \times \mathbf{v}$ for this flow field and verify that both the *curl* and the rotation vector of the paddle wheels are aligned with the z-axis. Show that the vorticity is positive where the paddle-wheels rotate in the counterclockwise direction and that the vorticity is negative where the paddle-wheels rotate in the clockwise direction.

It follows from the example of this section and the example of section (5.3) that both rotation and shear cause a nonzero vorticity. Both phenomena lead to the rotation of imaginary paddle-wheels embedded in the vector field. Therefore, the curl of a

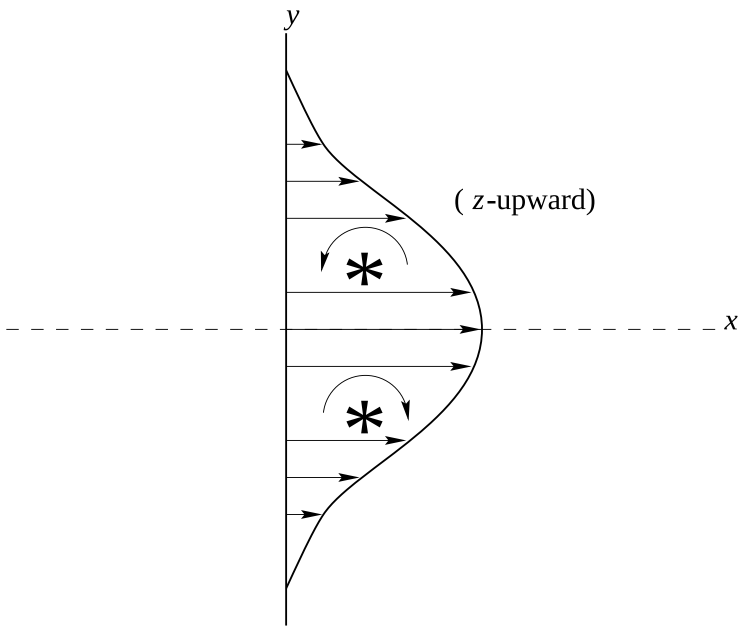


Figure 5.4: Sketch of the flow field for a shear flow.

vector field measures the local rotation of the vector field (in a literal sense). This explains why in some languages (i.e. Dutch) the notation *rot* $\mathbf{v}$ is used rather than *curl* $\mathbf{v}$. Note that this interpretation of the *curl* as a measure of (local) rotation is consistent with equation (5.3) where the *curl* is related to the value of the line integral along the small contour. If the flow (locally) rotates and if we integrate along the fluid flow, the line integral $\oint \mathbf{v} \cdot d\mathbf{r}$ will be relatively large, so that this line integral indeed measures the local rotation.

Rotation and shear each contribute to the *curl* of a vector field. Let us consider once again a vector field of the form (5.6) which is axially symmetric around the z-axis. In the following we don't require the rotation around the z-axis to be rigid, so that $v(r)$ in (5.6) is still arbitrary. We know that both the rotation around the z-axis and the shear are a source of vorticity.

Problem e: Show that for the flow

$$v(r) = \frac{A}{r} \tag{5.11}$$

the vorticity vanishes, with A a constant that is not yet determined. Make a sketch of this flow field.

The vorticity of this flow vanishes despite the fact that the flow rotates around the z-axis (but not in rigid rotation) and that the flow has a nonzero shear. The reason that the vorticity vanishes is that the contribution of the rotation around the z-axis to the vorticity is equal but of opposite sign from the contribution of the shear, so that the total vorticity vanishes. Note that this implies that a paddle-wheel does not change its orientation as it moves with this flow!

5.5 The magnetic field induced by a straight current

At this point you may have the impression that the flow field (5.11) is contrived in an artificial way. However, keep in mind that all the arguments of the previous section apply to any vector field and that fluid flow was used only as an example to fix our mind. As an example we consider the generation of the magnetic field **B** by an electrical current **J** that is independent of time. The Maxwell equation for the curl of the magnetic field in vacuum is for time-independent fields given by:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \ , \tag{5.12}$$

see equation (5.22) in ref. [31]. In this expression μ_0 is the magnetic permeability of vacuum. It plays the role of a coupling constant since it governs the strength of the magnetic field that is generated by a given current. It plays the same role as 1/permittivity in (4.12) or the gravitational constant G in (4.17). The vector **J** denotes the electric current per unit volume (properly called the electric current density).

For simplicity we will consider an electric current running through an infinite straight wire along the z-axis. Because of rotational symmetry around the z-axis and because of translational invariance along the z-axis the magnetic field depends neither on φ nor on z and must be of the form (5.6). Away from the wire the electrical current **J** vanishes.

Problem a: Show that

$$\mathbf{B} = \frac{A}{r}\hat{\boldsymbol{\varphi}} \ . \tag{5.13}$$

A comparison with equation (5.11) shows that for this magnetic field the contribution of the "rotation" around the z-axis to $\nabla \times \mathbf{B}$ is exactly balanced by the contribution of the"magnetic shear" to $\nabla \times \mathbf{B}$. It should be noted that the magnetic field derived in this section is of great importance because this field has been used to define the unit of electrical current, the Ampère. However, this can only be done when the constant A in expression (5.13) is known.

Problem b: Why does the treatment of this section not tell us what the relation is between the constant A and the current **J** in the wire?

We will return to this issue in section (7.3).

5.6 Spherical coordinates and cylinder coordinates

In section (4.4), expressions for the divergence in spherical coordinates and cylinder coordinates were derived. Here we will do the same for the *curl* because these expressions are frequently very useful. It is possible to derive the *curl* in curvilinear coordinates by systematically carrying out the effect of the coordinate transformation from Cartesian coordinates to curvilinear coordinates on all the elements of the involved vectors and on all the differentiations. As an alternative, we will use the physical interpretation of the *curl* given by expression (5.3) to derive the *curl* in spherical coordinates. This expression simply states that a certain component of the *curl* of a vector field $\mathbf{v}$ is the line integral $\oint \mathbf{v} \cdot d\mathbf{r}$ along a contour perpendicular to the component of the *curl* that we are considering, normalized by the surface area bounded by that contour. As an example we will derive for a system of spherical coordinates the φ-component of the *curl*, see figure (5.5) for the definition of the geometric variables.

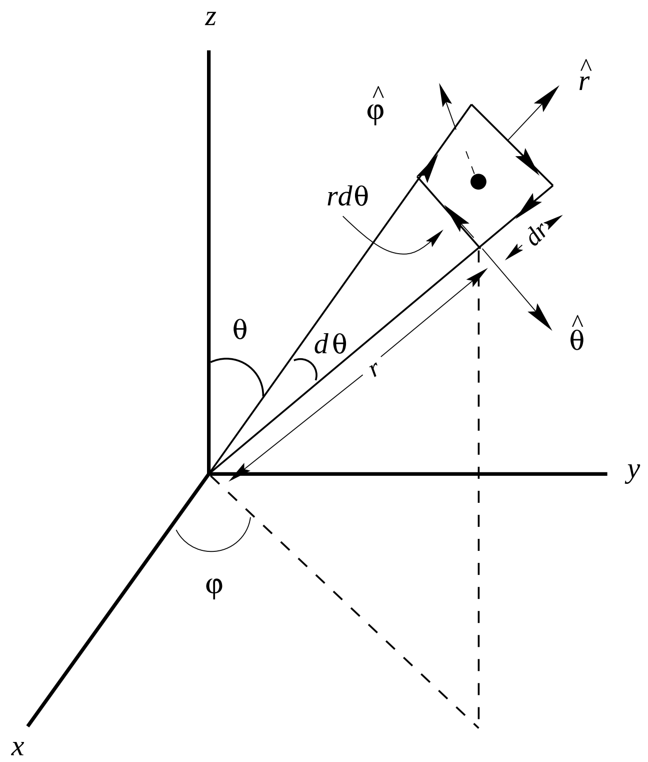


Figure 5.5: Definition of the geometric variables for the computation of the curl in spherical coordinates.

Consider in figure (5.5) the little surface. When we carry out the line integral along the surface we integrate in the direction shown in the figure. The reason for this is that the azimuth φ increases when we move into the figure, hence $\hat{\boldsymbol{\varphi}}$ point into the figure. Following the rules of a right-handed screw this corresponds with the indicated sense of integration. The area enclosed by the contour is given by $rd\theta dr$. By summing the contributions of the four sides of the contour we find using expression (5.3) that the φ-component of $\nabla \times \mathbf{v}$ is given by:

$$(\nabla \times \mathbf{v})_\varphi = \frac{1}{rd\theta dr} \left\{ v_\theta(r+dr,\theta)(r+dr)d\theta - v_r(r,\theta+d\theta)dr - v_\theta(r,\theta)rd\theta + v_r(r,\theta)dr \right\} . \tag{5.14}$$

In this expression v_r and v_θ denote the components of $\mathbf{v}$ in the radial direction and in the direction of $\hat{\boldsymbol{\theta}}$ respectively.

Problem a: Verify expression (5.14).

This result can be simplified by Taylor expanding the components of $\mathbf{v}$ in dr and $d\theta$ and linearizing the resulting expression in the infinitesimal increments dr and $d\theta$.

Problem b: Do this and show that the final result does not depend on dr and $d\theta$ and is given by:

$$(\nabla \times \mathbf{v})_{\varphi} = \frac{1}{r}\frac{\partial}{\partial r}(r v_{\theta}) - \frac{1}{r}\frac{\partial v_r}{\partial \theta} . \tag{5.15}$$

The same treatment can be applied to the other components of the curl. This leads to the following expression for the curl in spherical coordinates:

$$\nabla \times \mathbf{v} = \hat{\mathbf{r}}\tfrac{1}{r\sin\theta}\left\{\tfrac{\partial}{\partial\theta}(\sin\theta v_{\varphi}) - \tfrac{\partial v_{\theta}}{\partial\varphi}\right\} + \hat{\boldsymbol{\theta}}\tfrac{1}{r}\left\{\tfrac{1}{\sin\theta}\tfrac{\partial v_r}{\partial\varphi} - \tfrac{\partial}{\partial r}(r v_{\varphi})\right\} + \hat{\boldsymbol{\varphi}}\tfrac{1}{r}\left\{\tfrac{\partial}{\partial r}(r v_{\theta}) - \tfrac{\partial v_r}{\partial\theta}\right\} \tag{5.16}$$

Problem c: Show that in cylinder coordinates (r, φ, z) the curl is given by:

$$\nabla \times \mathbf{v} = \hat{\mathbf{r}}\left\{\tfrac{1}{r}\tfrac{\partial v_z}{\partial\varphi} - \tfrac{\partial v_{\varphi}}{\partial z}\right\} + \hat{\boldsymbol{\varphi}}\left\{\tfrac{\partial v_r}{\partial z} - \tfrac{\partial v_z}{\partial r}\right\} + \hat{\mathbf{z}}\tfrac{1}{r}\left\{\tfrac{\partial}{\partial r}(r v_{\varphi}) - \tfrac{\partial v_r}{\partial\varphi}\right\} , \tag{5.17}$$

with $r = \sqrt{x^2 + y^2}$.

Problem d: Use this result to re-derive (5.8) for vector fields of the form $\mathbf{v} = \mathbf{v}(r)\hat{\boldsymbol{\varphi}}$. Hint: use the same method as used in the derivation of (5.14) and treat the three components of the curl separately.

Chapter 6

The theorem of Gauss

In section (4.5) we have determined the gravitational field in N-dimensions using as only ingredient that in free space, where the mass density vanishes, the divergence of the gravitational field vanishes; $(\nabla \cdot \mathbf{g}) = 0$. This was sufficient to determine the gravitational field in expression (4.21). However, that expression is not quite satisfactory because it contains a constant A that is unknown. In fact, at this point we have no idea how this constant is related to the mass M that causes the gravitational field! The reason for this is simple, in order to derive the gravitational field in (4.21) we have only used the field equation (4.17) for free space (where $\rho = 0$). However, if we want to find the relation between the mass and the resulting gravitational field we must also use the field equation $(\nabla \cdot \mathbf{g}) = -4\pi G\rho$ at places where the mass is present. More specifically, we have to *integrate* the field equation in order to find the total effect of the mass. The theorem of Gauss gives us an expression for the volume integral of the divergence of a vector field.

6.1 Statement of Gauss' law

In section (4.2) it was shown that the divergence is the flux per unit volume. In fact, equation (4.4) gives us the outward flux $d\Phi$ through an infinitesimal volume dV; $d\Phi = (\nabla \cdot \mathbf{v}) dV$. We can immediately integrate this expression to find the total flux through the surface S which encloses the total volume V:

$$\oint_S \mathbf{v} \cdot d\mathbf{S} = \int_V (\nabla \cdot \mathbf{v}) dV \ . \tag{6.1}$$

In deriving this expression (4.4) has been used to express the total flux in the left hand side of (6.1). This expression is called the theorem of Gauss..

Note that in the derivation of (6.1) we did not use the dimensionality of the space, this relation holds in any number of dimensions. You may recognize the one-dimensional version of (6.1). In one dimension the vector $\mathbf{v}$ has only one component v_x, hence $(\nabla \cdot \mathbf{v}) = \partial_x v_x$. A "volume" in one dimension is simply a line, let this line run from $x = a$ to $x = b$. The "surface" of a one-dimensional volume consists of the endpoints of this line, so that the left hand side of (6.1) is the difference of the function v_x at its endpoints. This implies that the theorem of Gauss is in one-dimension:

$$v_x(b) - v_x(a) = \int_a^b \frac{\partial v_x}{\partial x} dx \ . \tag{6.2}$$

This expression will be familiar to you. We will use the 2-dimensional version of the theorem of Gauss in section (7.2) to derive the theorem of Stokes.

Problem a: Compute the flux of the vector field $\mathbf{v}(x, y, z) = (x + y + z)\hat{\mathbf{z}}$ through a sphere with radius R centered on the origin by explicitly computing the integral that defines the flux.

Problem b: Show that the total flux of the magnetic field of the earth through your skin is zero.

Problem c: Solve **problem a** without carrying out any integration explicitly.

6.2 The gravitational field of a spherically symmetric mass

In this section we will use Gauss's law (6.1) to show that the gravitational field of a body with a spherically symmetric mass density ρ depends only on the total mass but not on the distribution of the mass over that body. For a spherically symmetric body the mass density depends only on radius: $\rho = \rho(r)$. Because of the spherical symmetry of the mass, the gravitational field is spherically symmetric and points in the radial direction

$$\mathbf{g}(\mathbf{r}) = g(r)\hat{\mathbf{r}} \ . \tag{6.3}$$

Problem a: Use the field equation (4.17) for the gravitational field and Gauss's law (applied to a surface that completely encloses the mass) to show that

$$\oint_S \mathbf{g} \cdot d\mathbf{S} = -\, 4\pi G M \ , \tag{6.4}$$

where M is the total mass of the body.

Problem b: Use a sphere with radius r as the surface in (6.4) to show that the gravitational field is in three dimensions given by

$$\mathbf{g}(\mathbf{r}) = -\, \frac{GM}{r^2}\hat{\mathbf{r}} \ . \tag{6.5}$$

This is an intriguing result. What we have shown here is that the gravitational field depends only the total mass of the spherically symmetric body, but not on the distribution of the mass within that body. As an example consider two bodies with the same mass. One body has all the mass located in a small ball near the origin and the other body has all the mass distributed on a thin spherical shell with radius R, see figure (6.1). According to expression (6.5) these bodies generate exactly the same gravitational field outside the body. This implies that gravitational measurements taken outside the two bodies cannot be used to distinguish between them. The non-unique relation between the gravity field and the underlying mass-distribution is of importance for the interpretation of gravitational measurements taken in geophysical surveys.

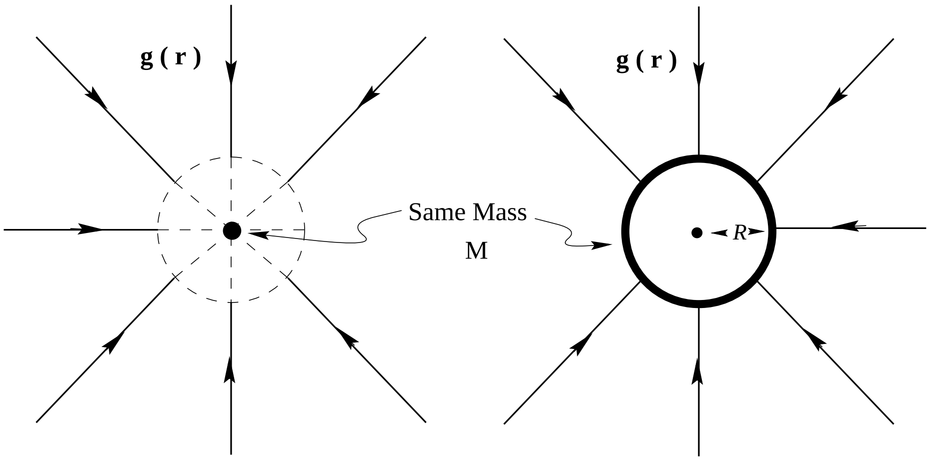


Figure 6.1: Two different bodies with a different mass distribution that generate the same gravitational field for distances larger than the radius of the body on the right.

Problem c: Let us assume that the mass is located within a sphere with radius R, and that the mass density within that sphere is constant. Integrate equation (4.17) over a sphere with radius $r < R$ to show that the gravitational field within the sphere is given by:

$$\mathbf{g}(\mathbf{r}) = -\frac{MGr}{R^3}\hat{\mathbf{r}} \ . \tag{6.6}$$

Plot the gravitational field as a function from r when the distance increases from zero to a distance larger than the radius R. Verify explicitly that the gravitational field is continuous at the radius R of the sphere.

Note that all conclusions hold identically for the electrical field when we replace the mass density by the charge density, because expression (4.12) for the divergence of the electric field has the same form as equation (4.17) for the gravitational field. As an example we will consider a hollow spherical shell with radius R. On the spherical shell electrical charge is distributed with a constant charge density: $\rho = const.$

Problem d: Use expression (4.12) for the electric field and Gauss's law to show that within the hollow sphere the electric field vanishes: $\mathbf{E}(\mathbf{r}) = 0$ for $r < R$.

This result implies that when a charge is placed within such a spherical shell the electrical field generated by the charge on the shell exerts no net force on this charge; the charge will not move. Since the electrical potential satisfies $\mathbf{E} = -\nabla V$, the result derived in **problem d** implies that the potential is constant within the sphere. This property has actually been used to determine experimentally whether the electric field indeed satisfies (4.12) (which implies that the field of point charge decays as $1/r^2$). Measurement of the potential differences within a hollow spherical shell as described in **problem d** can be carried out with very great sensitivity. Experiments based on this principle (usually in a more elaborated form) have been used to ascertain the decay of the electric field of a point charge with distance. Writing the field strength as $1/r^{2+\varepsilon}$ is has now be shown that $\varepsilon = (2.7 \pm 3.1) \times 10^{-16}$, see section I.2 of *Jackson*[31] for a discussion. The small value of ε is a remarkable experimental confirmation of equation (4.12) for the electric field.

6.3 A representation theorem for acoustic waves

Acoustic waves are waves that propagate through a gas or fluid. You can hear the voice of others because acoustic waves propagate from their vocal tract to your ear. Acoustic waves are frequently used to describe the propagation of waves through the earth. Since the earth is a solid body, this is strictly speaking not correct, but under certain conditions (small scattering angles) the errors can be acceptable. The pressure field $p(\mathbf{r})$ of acoustic waves satisfy in the frequency domain the following partial differential equation:

$$\nabla \cdot \left(\frac{1}{\rho} \nabla p \right) + \frac{\omega^2}{\kappa} p = f \ . \tag{6.7}$$

In this expression $\rho(\mathbf{r})$ is the mass density of the medium while $\kappa(\mathbf{r})$ is the compressibility (a factor that describes how strongly the medium resists changes in its volume). The right hand side $f(\mathbf{r})$ describes the source of the acoustic wave. This term accounts for example for the action of your voice.

We will now consider two pressure fields $p_1(\mathbf{r})$ and $p_2(\mathbf{r})$ that both satisfy (6.7) with sources $f_1(\mathbf{r})$ and $f_2(\mathbf{r})$ in the right hand side of the equation.

Problem a: Multiply equation (6.7) for p_1 with p_2, multiply equation (6.7) for p_2 with p_1 and subtract the resulting expressions. Integrate the result over a volume V to show that:

$$\int_V \left\{ p_2 \nabla \cdot \left(\frac{1}{\rho} \nabla p_1 \right) - p_1 \nabla \cdot \left(\frac{1}{\rho} \nabla p_2 \right) \right\} dV = \int_V \{ p_2 f_1 - p_1 f_2 \} \, dV \ . \tag{6.8}$$

Ultimately we want to relate the wavefield at the surface S that encloses the volume V to the wavefield within the volume. Obviously, Gauss's law is the tool for doing this. The problem we face is that Gauss's law holds for the volume integral of the divergence, whereas in expression (6.8) we have the *product* of a divergence (such as $\nabla \cdot \left(\frac{1}{\rho} \nabla p_1 \right)$) with another function (such as p_2).

Problem b: This means we have to "make" a divergence. Show that

$$p_2 \nabla \cdot \left(\frac{1}{\rho} \nabla p_1 \right) = \nabla \cdot \left(\frac{1}{\rho} p_2 \nabla p_1 \right) - \frac{1}{\rho} (\nabla p_1 \cdot \nabla p_2) \ . \tag{6.9}$$

What we are doing here is similar to the standard derivation of integration by parts. The easiest way to show that $\int_a^b f(\partial g/\partial x) dx = [f(x)g(x)]_a^b - \int_a^b (\partial f/\partial x) g dx$, is to integrate the identity $f(\partial g/\partial x) = \partial(fg)/dx - (\partial f/\partial x) g$ from $x = a$ to $x = b$. This last equation has exactly the same structure as expression (6.9).

Problem c: Use expressions (6.8), (6.9) and Gauss's law to derive that

$$\oint_S \frac{1}{\rho} \left(p_2 \nabla p_1 - p_1 \nabla p_2 \right) \cdot d\mathbf{S} = \int_V \{ p_2 f_1 - p_1 f_2 \} \, dV \ . \tag{6.10}$$

This expression forms the basis for the proof that reciprocity holds for acoustic waves. (Reciprocity means that the wavefield propagating from point A to point B is identical to the wavefield that propagates in the reverse direction from point B to point A.) To see the power of expression (6.10), consider the special case that the source f_2 of p_2 is of unit strength and that this source is localized in a very small volume around a point $\mathbf{r}_0$ within the volume. This means that f_2 in the right hand side of (6.10) is only nonzero at $\mathbf{r}_0$. The corresponding volume integral $\int_V p_1 f_2 dV$ is in that case given by $p_1(\mathbf{r}_0)$. The wavefield $p_2(\mathbf{r})$ generated by this point source is called the *Green's function,* this special solution is denoted by $G(\mathbf{r}, \mathbf{r}_0)$. (The concept Green's function is introduced in great detail in chapter 14.) The argument $\mathbf{r}_0$ is added to indicate that this is the wavefield at location $\mathbf{r}$ due to a unit source at location $\mathbf{r}_0$. We will now consider a solution p_1 that has no sources within the volume V (i.e. $f_1 = 0$). Let us simplify the notation further by dropping the subscript "1" in p_1.

Problem d: Show by making all these changes that equation (6.10) can be written as:

$$p(\mathbf{r}_0) = \oint_S \frac{1}{\rho} \left(p(\mathbf{r}) \nabla G(\mathbf{r}, \mathbf{r}_0) - G(\mathbf{r}, \mathbf{r}_0) \nabla p(\mathbf{r}) \right) \cdot d\mathbf{S} \ . \tag{6.11}$$

This result is called the "representation theorem" because it gives the wavefield inside the volume when the wavefield (an its gradient) are specified on the surface that bounds this volume. Expression (6.11) can be used to formally derive Huygens' principle which states that every point on a wavefront acts as a source for other waves and that interference of these waves determine the propagation of the wavefront. Equation (6.11) also forms the basis for imaging techniques for seismic data, see for example ref. [53]. In seismic exploration one records the wavefield at the earth's surface. This can be used by taking the earth's surface as the surface S over which the integration is carried out. If the Green's function $G(\mathbf{r}, \mathbf{r}_0)$ is known, one can use expression (6.11) to compute the wavefield in the interior in the earth. Once the wavefield in the interior of the earth is known, one can deduce some of the properties of the material in the earth. In this way, equation (6.11) (or its elastic generalization) forms the basis of seismic imaging techniques.

Problem e: This almost sounds too good to be true! Can you find the catch?

6.4 Flowing probability

In classical mechanics, the motion of a particle with mass m is governed by Newton's law: $m\ddot{\mathbf{r}} = \mathbf{F}$. When the force $\mathbf{F}$ is associated with a potential $V(\mathbf{r})$ the motion of the particle satisfies:

$$m \frac{d^2\mathbf{r}}{dt^2} = -\nabla V(\mathbf{r}) \ . \tag{6.12}$$

However, this law does not hold for particles that are very small. Microscopic particles such as electrons are not described accurately by (6.12). It is one of the outstanding features of quantum mechanics that microscopic particles are treated as waves rather than particles. The wave function $\psi(\mathbf{r},t)$ that describes a particle that moves under the influence of a potential $V(\mathbf{r})$ satisfies the Schrödinger equation[39]:

$$-\frac{\hbar}{i} \frac{\partial \psi(\mathbf{r}, t)}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi(\mathbf{r}, t) + V(\mathbf{r}) \psi(\mathbf{r}, t) \ . \tag{6.13}$$

In this expression, $\hbar$ is Planck's constant h divided by 2π.

Problem a: Check that Planck's constant has the dimension of angular momentum.

Planck's constant has the numerical value $h = 6.626 \times 10^{-34}\ kg\ m^2/s$. Suppose we are willing to accept that the motion of an electron is described by the Schrödinger equation, then the following question arises: What is the position of the electron as a function of time? According to the Copenhagen interpretation of quantum mechanics this is a meaningless question because the electron behaves like a wave and does not have a definite location. Instead, the wavefunction $\psi(\mathbf{r}, t)$ dictates how likely it is that the particle is at location $\mathbf{r}$ at time t. Specifically, the quantity $|\psi(\mathbf{r}, t)|^2$ is the probability density of finding the particle at location $\mathbf{r}$ at time t. This implies that the probability P_V that the particle is located within the volume V is given by $P_V = \int_V |\psi|^2 dV$. (Take care not to confuse the volume with the potential, because they are both indicated with the same symbol V.) This implies that the wavefunction is related to a probability. Instead of the motion of the electron, Schrödinger's equation dictates how the probability density of the particle moves through space as time progresses. One expects that a "probability current" is associated with this movement. In this section we will determine this current using the theorem of Gauss.

Problem b: In the following we need the time-derivative of $\psi^*(\mathbf{r}, t)$, where the asterisk denotes the complex conjugate. Derive the differential equation that $\psi^*(\mathbf{r}, t)$ obeys by taking the complex conjugate of Schrödinger's equation (6.13).

Problem c: Use this result to derive that for a volume V that is fixed in time:

$$\frac{\partial}{\partial t}\int_V |\psi|^2 dV = \frac{i\hbar}{2m}\int_V (\psi^*\nabla^2\psi - \psi\nabla^2\psi^*)dV \ . \tag{6.14}$$

Problem d: Use Gauss's law to rewrite this expression as:

$$\frac{\partial}{\partial t}\int_V |\psi|^2 dV = \frac{i\hbar}{2m}\oint (\psi^*\nabla\psi - \psi\nabla\psi^*) \cdot d\mathbf{S} \ . \tag{6.15}$$

Hint, spot the divergence in (6.14) first.

The left hand side of this expression gives the time-derivative of the probability that the particle is within the volume V. The only way the particle can enter or leave the volume is through the enclosing surface S. The right hand side therefore describes the "flow" of probability through the surface S. More accurately, one can formulate this as the flux of the probability density current.

Problem e: Show from (6.15) that the probability density current $\mathbf{J}$ is given by:

$$\mathbf{J} = \frac{i\hbar}{2m}(\psi\nabla\psi^* - \psi^*\nabla\psi) \tag{6.16}$$

Pay in particular attention to the sign of the terms in this expression.

As an example let us consider a plane wave:

$$\psi(\mathbf{r},t) = A \exp i(\mathbf{k}\cdot\mathbf{r} - \omega t) \ , \tag{6.17}$$

where $\mathbf{k}$ is the wavevector and A an unspecified constant.

Problem f: Show that the wavelength λ is related to the wavevector by the relation $\lambda = 2\pi/\left|\mathbf{k}\right|$. In which direction does the wave propagate?

Problem g: Show that the probability density current $\mathbf{J}$ for this wavefunction satisfies:

$$\mathbf{J} = \frac{\hbar\mathbf{k}}{m}|\psi|^2 \ . \tag{6.18}$$

This is a very interesting expression. The term $|\psi|^2$ gives the probability density of the particle, while the probability density current $\mathbf{J}$ physically describes the current of this probability density. Since the probability current moves with the velocity of the particle (why?), the remaining terms in the right hand side of (6.18) must denote the velocity of the particle:

$$\mathbf{v} = \frac{\hbar\mathbf{k}}{m} \ . \tag{6.19}$$

Since the momentum $\mathbf{p}$ is the mass times the velocity, equation (6.19) can also be written as $\mathbf{p} = \hbar\mathbf{k}$. This relation was proposed by de Broglie in 1924 using completely different arguments than we have used here[13]. Its discovery was a major step in the development of quantum mechanics.

Problem h: Use this expression and the result of **problem f** to compute your own wavelength while you are riding your bicycle. Are quantum-mechanical phenomena important when you ride you bicycle? Use your wavelength as an argument. Did you know you possessed a wavelength?

Chapter 7

The theorem of Stokes

In section 6 we have noted that in order to find the gravitational field of a mass we have to integrate the field equation (4.17) over the mass. Gauss's theorem can then be used to compute the integral of the divergence of the gravitational field. For the *curl* the situation is similar. In section (5.5) we computed the magnetic field generated by a current in a straight infinite wire. The field equation

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \qquad (5.12) \quad again$$

was used to compute the field away from the wire. However, the solution (5.13) contained an unknown constant A. The reason for this is that the field equation (5.12) was only used outside the wire, where $\mathbf{J} = 0$. The treatment of section (5.5) therefore did not provide us with the relation between the field $\mathbf{B}$ and its source $\mathbf{J}$. The only way to obtain this relation is to integrate the field equation. This implies we have to compute the integral of the *curl* of a vector field. The theorem of Stokes tells us how to do this.

7.1 Statement of Stokes' law

The theorem of Stokes is based on the principle that the *curl* of a vector field is the closed line integral of the vector field per unit surface area, see section (5.1). Mathematically this statement is expressed by equation (5.2) that we write in a slightly different form as:

$$\oint_{d\mathbf{S}} \mathbf{v} \cdot d\mathbf{r} = (\boldsymbol{\nabla} \times \mathbf{v}) \cdot \hat{\mathbf{n}}\ dS = (\boldsymbol{\nabla} \times \mathbf{v}) \cdot d\mathbf{S}\ . \tag{7.1}$$

The only difference with (5.2) is that in the expression above we have not aligned the z-axis with the vector $\boldsymbol{\nabla} \times \mathbf{v}$. The infinitesimal surface therefore is not necessarily confined to the x, y-plane and the z-component of the *curl* is replaced by the component of the *curl* normal to the surface, hence the occurrence of the terms $\hat{\mathbf{n}}\ dS$ in (7.1). Expression (7.1) holds for an infinitesimal surface area. However, this expression can immediately be integrated to give the surface integral of the *curl* over a finite surface S that is bounded by the curve C:

$$\oint_C \mathbf{v} \cdot d\mathbf{r} = \int_S (\nabla \times \mathbf{v}) \cdot d\mathbf{S}\ . \tag{7.2}$$

This result is known as the theorem of Stokes (or Stokes' law). The line integral in the left hand side is over the curve that bounds the surface S. A proper derivation of Stokes' law can be found in ref. [38].

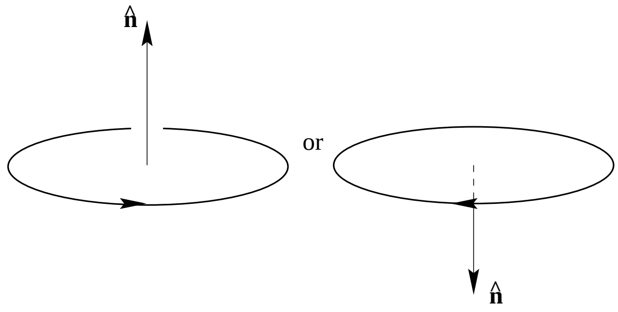


Figure 7.1: The relation between the sense of integration and the orientation of the surface.

Note that a line integration along a closed surface can be carried out in two directions. What is the direction of the line integral in the left hand side of Stokes' law (7.2)? To see this, we have to realize that Stokes' law was ultimately based on equation (5.2). The orientation of the line integration used in that expression is defined in figure (5.1), where it can be seen that the line integration is in the counterclockwise direction. In figure (5.1) the z-axis points out off the paper, this implies that the vector $d\mathbf{S}$ also points out of the paper. *This means that in Stokes' law the sense of the line integration and the direction of the surface vector* $d\mathbf{S}$ *are related through the rule for a right-handed screw.*

There is something strange about Stokes' law. If we define a curve C over which we carry out the line integral, we can define many different surfaces S that are bounded by the same curve C. Apparently, the surface integral in the right hand side of Stokes' law does not depend on the specific choice of the surface S as long as it is bounded by the curve C.

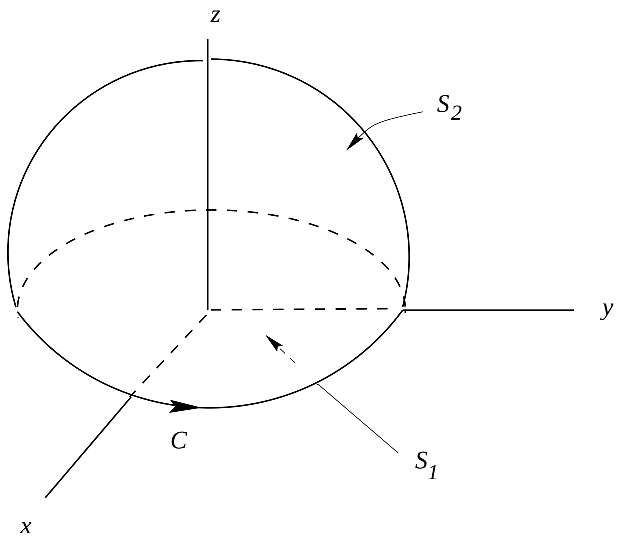


Figure 7.2: Definition of the geometric variables for problem a.

Problem a: Let us verify this property for an example. Consider the vector field $\mathbf{v} = r\hat{\boldsymbol{\varphi}}$. Let the curve C used for the line integral be a circle in the x, y-plane with radius R, see figure (**??**) for the geometry of the problem. (i) Compute the line integral in the left hand side of (7.2) by direct integration. Compute the surface integral in the right hand side of (7.2) by (ii) integrating over a circle of radius R in the x, y-plane (the surface S_1 in figure (**??**)) and by (iii) integrating over the upper half of a sphere with radius R (the surface S_2 in figure (**??**)). Verify that the three integrals are identical.

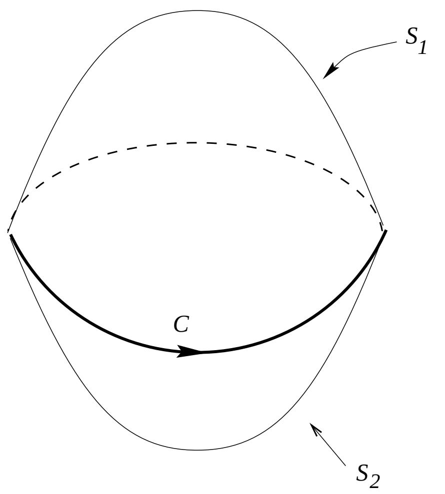


Figure 7.3: Two surfaces that are bouded by the same contour C.

It is actually not difficult to prove that the surface integral in Stokes' law is independent of the specific choice of the surface S as long as it is bounded by the same contour C. Consider figure (7.3) where the two surfaces S_1 and S_2 are bounded by the same contour C. We want to show that the surface integral of $\nabla \times \mathbf{v}$ is the same for the two surfaces, i.e. that:

$$\int_{S_1} (\nabla \times \mathbf{v}) \cdot d\mathbf{S} = \int_{S_2} (\nabla \times \mathbf{v}) \cdot d\mathbf{S} \ . \tag{7.3}$$

We can form a closed surface S by combining the surfaces S_1 and S_2.

Problem b: Show that equation (7.3) is equivalent to the condition

$$\oint_S (\nabla \times \mathbf{v}) \cdot d\mathbf{S} = 0 \ , \tag{7.4}$$

where the integration is over the closed surfaces defined by the combination of S_1 and S_2. Pay in particular attention to the sign of the different terms.

Problem c: Use Gauss' law to convert (7.4) to a volume integral and show that the integral is indeed identical to zero.

The result you obtained in **problem c** implies that the condition (7.3) is indeed satisfied and that in the application of Stokes' law you can choose *any* surface as long as it is bounded by the contour over which the line integration is carried out. This is a very useful result because often the surface integration can be simplified by choosing the surface carefully.

7.2 Stokes' theorem from the theorem of Gauss

Stokes' law is concerned with surface integrations. Since the *curl* is intrinsically a three-dimensional vector, Stokes's law is inherently related to three space dimensions. However, if we consider a vector field that depends only on the coordinates x and y $(\mathbf{v} = \mathbf{v}(x, y))$ and that has a vanishing component in the z-direction $(v_z = 0)$, then $\nabla \times \mathbf{v}$ points along the z-axis. If we consider a contour C that is confined to the x, y-plane, Stokes' law takes for such a vector field the form

$$\oint_C (v_x dx + v_y dy) = \int_S (\partial_x v_y - \partial_y v_x)\, dx dy \ . \tag{7.5}$$

Problem a: Verify this.

This result can be derived from the theorem of Gauss in two dimensions.

Problem b: Show that Gauss' law (6.1) for a vector field $\mathbf{u}$ in two dimensions can be written as

$$\oint_C (\mathbf{u} \cdot \hat{\mathbf{n}}) ds = \int_S (\partial_x u_x + \partial_y u_y) dx dy \ , \tag{7.6}$$

where the unit vector $\hat{\mathbf{n}}$ is perpendicular to the curve C (see figure (7.4)) and where ds denotes the integration over the arclength of the curve C.

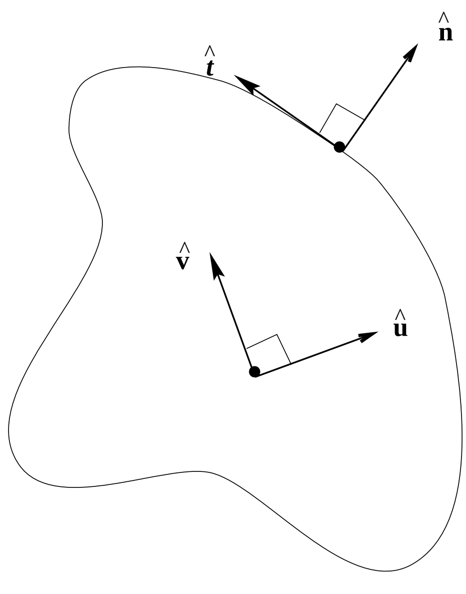


Figure 7.4: Definition of the geometric variables for the derivation of Stokes' law from the theorem of Gauss.

In order to derive the special form of Stokes' law (7.5) from Gauss' law (7.6) we have to define the relation between the vectors $\mathbf{u}$ and $\mathbf{v}$. Let the vector $\mathbf{u}$ follow from $\mathbf{v}$ by a clockwise rotation over 90 degrees, see figure (7.4).

Problem c: Show that:

$$v_x = -u_y \qquad and \qquad v_y = u_x \ . \tag{7.7}$$

We now define the unit vector $\hat{\mathbf{t}}$ to be directed along the curve C, see figure (7.4). Since a rotation is an orthonormal transformation the inner product of two vectors is invariant for a rotation over 90 degrees so that $(\mathbf{u} \cdot \hat{\mathbf{n}}) = (\mathbf{v} \cdot \hat{\mathbf{t}})$.

Problem d: Verify this by expressing the components of $\hat{\mathbf{t}}$ in the components of $\hat{\mathbf{n}}$ and by using (7.7).

Problem e: Use these results to show that (7.5) follows from (7.6).

What you have shown here is that Stokes' law for the special case considered in this section is identical to the theorem of Gauss for two spatial dimensions.

7.3 The magnetic field of a current in a straight wire

We now return to the problem of the generation of the magnetic field induced by a current in an infinite straight wire that was discussed in section (5.5). Because of the cylinder symmetry of the problem, we know that the magnetic field is in the direction of the unit vector $\hat{\boldsymbol{\varphi}}$ and that the field only depends on the distance $r = \sqrt{x^2 + y^2}$ to the wire:

$$\mathbf{B} = B(r)\hat{\boldsymbol{\varphi}} \ . \tag{7.8}$$

The field can be found by integrating the field equation $\nabla \times \mathbf{B} = \mu_0 \mathbf{J}$ over a disc with radius r perpendicular to the wire, see figure (7.5). When the disc is larger than the thickness of the wire the surface integral of $\mathbf{J}$ gives the electric current I through the wire: $I = \int \mathbf{J} \cdot d\mathbf{S}$.

Problem a: Use these results and Stokes' law to show that:

$$\mathbf{B} = \frac{\mu_0 I}{2\pi r}\hat{\boldsymbol{\varphi}} \ . \tag{7.9}$$

We now have a relation between the magnetic field and the current that generates the field, hence the constant A in expression (5.13) is now determined. Note that the magnetic field depends only on the total current through the wire, but that is does not depend on the distribution of the electric current density $\mathbf{J}$ within the wire as long as the electric current density exhibits cylinder symmetry. Compare this with the result you obtained in **problem b** of section (6.2)!

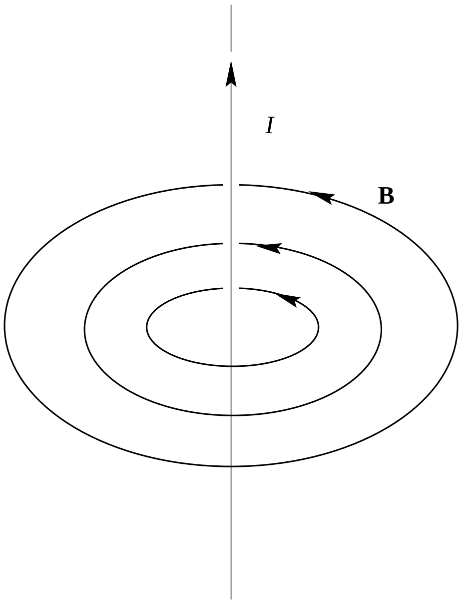


Figure 7.5: Geometry of the magnetic field induced by a current in a straight infinite wire.

7.4 Magnetic induction and Lenz's law

The theory of the previous section deals with the generation of a magnetic field by a current. A magnet placed in this field will experience a force exerted by the magnetic field. This force is essentially the driving force in electric motors; using an electrical current that changes with time a time-dependent magnetic field is generated that exerts a force on magnets attached to a rotation axis.

In this section we will study the reverse effect; what is the electrical field generated by a magnetic field that changes with time? In a dynamo, a moving part (e.g. your bicycle wheel) drives a magnet. This creates a time-dependent electric field. This process is called magnetic induction and is described by the following Maxwell equation (see ref. [31]):

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{7.10}$$

To fix our mind let us consider a wire with endpoints A en B, see figure (7.6). The direction of the magnetic field is indicated in this figure. In order to find the electric field induced in the wire, integrate equation (7.10) over the surface enclosed by the wire

$$\int_S (\nabla \times \mathbf{E}) \cdot d\mathbf{S} = -\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} \ . \tag{7.11}$$

Problem a: Show that the right hand side of (7.11) is given by $-\partial\Phi/\partial t$, where Φ is the magnetic flux through the wire. (See section (4.1) for the definition of the flux.)

We have discovered that a change in the magnetic flux is the source of an electric field. The resulting field can be characterized by the *electromotive force* F_{AB} which is a measure

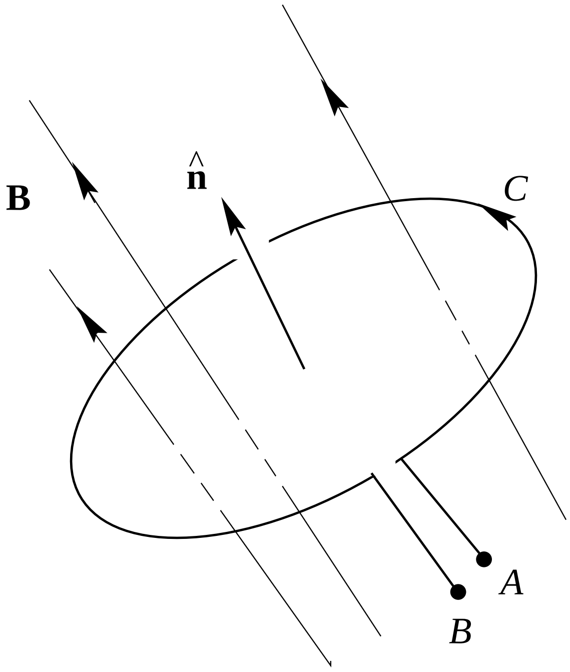


Figure 7.6: A wire-loop in a time-dependent magnetic field.

of the work done by the electric field on a unit charge when it moves from point A to point B, see figure (7.6):

$$F_{AB} \equiv \int_A^B \mathbf{E} \cdot d\mathbf{r} \ . \tag{7.12}$$

Problem b: Show that the electromotive force satisfies

$$F_{AB} = -\frac{\partial \Phi}{\partial t} \ . \tag{7.13}$$

Problem c: Because of the electromotive force an electric current will flow through the wire. Determine the direction of the electric current in the wire. Show that this current generates a magnetic field that opposes the change in the magnetic field that generates this current. You have learned in section (7.3) the direction of the magnetic field that is generated by an electric current in a wire.

What we have discovered in **problem c** is Lenz's law, which states that induction currents lead to a secondary magnetic field which opposes the change in the primary magnetic field that generates the electric current. This implies that coils in electrical systems exhibit a certain inertia in the sense that they resist changes in the magnetic field that passes through the coil. The amount of inertia is described by a quantity called the inductance L. This quantity plays a similar role as mass in classical mechanics because the mass of a body also describes how strongly a body resists changing its velocity when an external force is applied.

7.5 The Aharonov-Bohm effect

It was shown in section (4.3) that because of the absence of magnetic monopoles the magnetic field is source-free: $(\nabla \cdot \mathbf{B}) = 0$. In electromagnetism one often expresses the magnetic field as the curl of a vector field $\mathbf{A}$:

$$\mathbf{B} = \nabla \times \mathbf{A} \ . \tag{7.14}$$

The advantage of writing the magnetic field in this way is that for *any* field $\mathbf{A}$ the magnetic field satisfies $(\nabla \cdot \mathbf{B}) = 0$ because $\nabla \cdot (\nabla \times \mathbf{A}) = 0$.

Problem a: Give a proof of this last identity.

The vector field $\mathbf{A}$ is called the *vector potential*. The reason for this name is that it plays a similar role as the electric potential V. Both the electric and the magnetic field follows from V and $\mathbf{A}$ respectively by differentiation: $\mathbf{E} = -\nabla V$ and $\mathbf{B} = \nabla \times \mathbf{A}$. The vector potential has the strange property that it can be nonzero (and variable) in parts of space where the magnetic field vanishes. As an example, consider a magnetic field with cylinder symmetry along the z-axis that is constant for $r < R$ and which vanishes for $r > R$:

$$\mathbf{B} = \begin{cases} B_0 \hat{\mathbf{z}} & for \quad r < R \\ 0 & for \quad r > R \end{cases} \tag{7.15}$$

see figure (7.7) for a sketch of the magnetic field. Because of cylinder symmetry the vector potential is a function of the distance r to the z-axis only and does not depend on z or φ.

Problem b: Show that a vector potential of the form

$$\mathbf{A} = f(r) \hat{\boldsymbol{\varphi}} \tag{7.16}$$

gives a magnetic field in the required direction. Give a derivation that $f(r)$ satisfies the following differential equation:

$$\frac{1}{r} \frac{\partial}{\partial r} (r f(r)) = \begin{cases} B_0 & for \quad r < R \\ 0 & for \quad r > R \end{cases} \tag{7.17}$$

These differential equations can immediately be integrated. After integration two integration constants are present. These constants follow from the requirement that the vector potential is continuous at $r = R$ and from the requirement that $f(r = 0) = 0$. (This requirement is needed because the direction of the unit vector $\hat{\boldsymbol{\varphi}}$ is undefined on the z-axis where $r = 0$. The vector potential therefore only has a unique value at the z-axis when $f(r = 0) = 0$.)

Problem c: Integrate the differential equation (7.17) and use that with the requirements described above the vector potential is given by

$$\mathbf{A} = \begin{cases} \frac{1}{2} B_0 r \hat{\boldsymbol{\varphi}} & for \quad r < R \\ \frac{1}{2} B_0 \frac{R^2}{r} \hat{\boldsymbol{\varphi}} & for \quad r > R \end{cases} \tag{7.18}$$

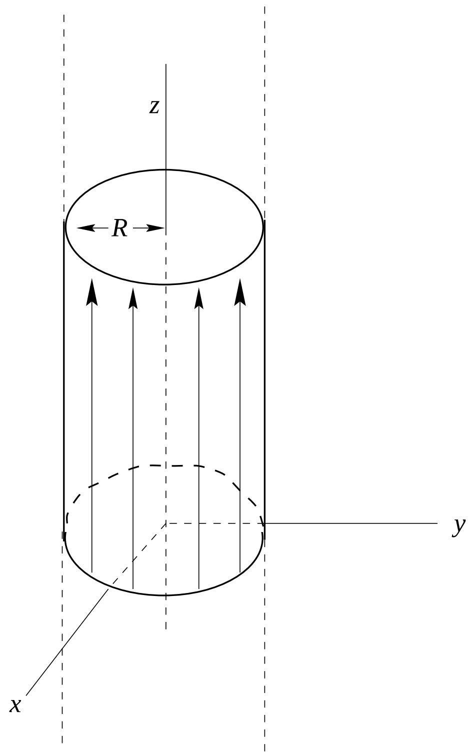


Figure 7.7: Geometry of the magnetic field.

The important point of this expression is that although the magnetic field is only nonzero for $r < R$, the vector potential (and its gradient) is nonzero everywhere in space! The vector potential is thus much more non-local than the magnetic field. This leads to a very interesting effect in quantum mechanics; the Aharonov Bohm effect.

Before introducing this effect we need to know more about quantum mechanics. As you have seen in section (6.4) microscopic "particles" such as electrons behave more like a wave than like a particle. Their wave properties are described by Schrödinger's equation (6.13). When different waves propagate in the same region of space, interference can occur. In some parts of space the waves may enhance each other (constructive interference) while in other parts the waves cancel each other (destructive interference). This is observed for "particle waves" when electrons are being send through two slits and where the electrons are detected on a screen behind these slits, see the left panel of figure (7.8). You might expect that the electrons propagate like bullets along straight lines and that they are only detected in two points after the two slits. However, this is not the case, in experiments one observes a pattern of fringes on the screen that are caused by the constructive and destructive interference of the electron waves. This interference pattern is sketched in figure (7.8) on the right side of the screens. This remarkable confirmation of the wave-property of particles is described clearly in ref. [21]. (The situation is even more remarkable, when one send the electrons through the slits "one-by-one" so that only one electron passes through the slits at a time, one sees a dot at the detector for each electron. However, after many particles have arrived at the detector this pattern of dots forms the interference pattern of the waves, see ref. [54].)

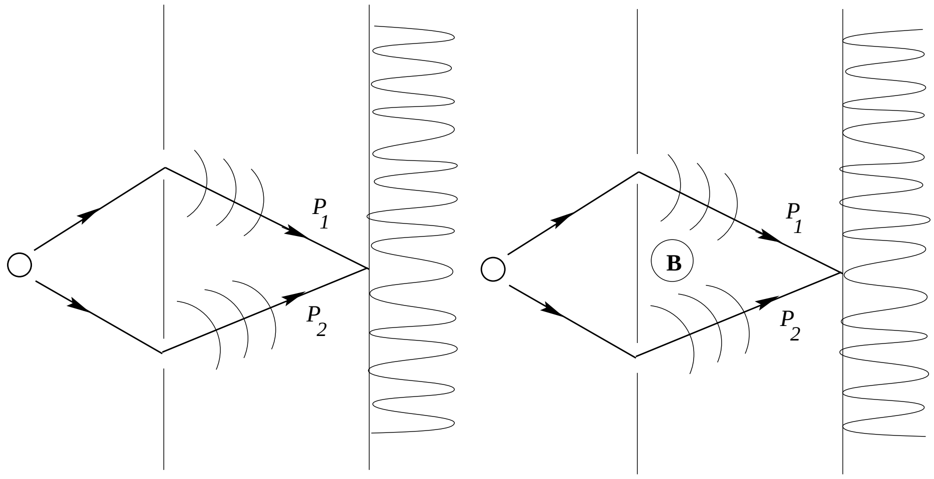


Figure 7.8: Experiment where electrons travel through two slits and are detected on a screen behind the slits. The resulting interference pattern is sketched. The experiment without magnetic field is shown on the left, the experiment with magnetic field is shown on the right. Note the shift in the maxima and minima of the interference pattern between the two experiments.

Let us now consider the same experiment, but with a magnetic field given by equation (7.15) placed between the two slits. Since the electrons do not pass through this field one expects that the electrons are not influenced by this field and that the magnetic field does not change the observed interference pattern at the detector. However, it is an observational fact that the magnetic field *does* change the interference pattern at the detector, see ref. [54] for examples. This surprising effect is called the Aharonov-Bohm effect.

In order to understand this effect, we should note that a magnetic field in quantum mechanics leads to a phase shift of the wavefunction. If the wavefunction in the absence is given by $\psi(\mathbf{r})$, the wavefunction in the presence of the magnetic field is given by $\psi(\mathbf{r})\times \exp\left(\frac{ie}{\hbar c}\int_P \mathbf{A}\cdot d\mathbf{r}\right)$, see ref. [52]. In this expression $\hbar$ is Planck's constant (divided by 2π), c is the speed of light and $\mathbf{A}$ is the vector potential associated with the magnetic field. The integration is over the path P from the source of the particles to the detector. Consider now the waves that interfere in the two-slit experiment in the right panel of figure (7.8). The wave that travels through the upper slit experiences a phase shift $\exp\left(\frac{ie}{\hbar c}\int_{P_1} \mathbf{A}\cdot d\mathbf{r}\right)$, where the integration is over the path P_1 through the upper slit. The wave that travels through the lower slit obtains a phase shift $\exp\left(\frac{ie}{\hbar c}\int_{P_2} \mathbf{A}\cdot d\mathbf{r}\right)$ where the path P_2 runs through the lower slit.

Problem d: Show that the phase difference $\delta\varphi$ between the two waves due to the presence of the magnetic field is given by

$$\delta\varphi = \frac{e}{\hbar c}\oint_P \mathbf{A}\cdot d\mathbf{r} \ , \tag{7.19}$$

where the path P is the closed path from the source through the upper slit to the detector and back through the lower slit to the source.

This phase difference affects the interference pattern because it is the *relative* phase between interfering waves that determines whether the interference is constructive or destructive.

Problem e: Show that the phase difference can be written as

$$\delta\varphi = \frac{e\Phi}{\hbar c} \,, \tag{7.20}$$

where Φ is the magnetic flux through the area enclosed by the path P.

This expression shows that the phase shift between the interfering waves is proportional to the magnetic field enclosed by the paths of the interfering waves, *despite the fact that the electrons never move through the magnetic field* **B**. Mathematically the reason for this surprising effect is that the vector potential is nonzero throughout space even when the magnetic field is confined to a small region of space, see expression (7.18) as an example. However, this explanation is purely mathematical and does not seem to agree with common sense. This has led to speculations that the vector potential is actually a more "fundamental" quantity than the magnetic field[54].

7.6 Wingtips vortices

Figure 7.9: Vortices trailing form the wingtips of a Boeing 727.

If you have been watching aircraft closely, you may have noticed that sometimes a little stream of condensation is left behind by the wingtips, see figure (7.9). This is a different condensation trail than the thick contrails created by the engines. The condensation trails that start at the wingtips is due to a vortex (a spinning motion of the air) that is generated at the wingtips. This vortex is called the wingtip-vortex. In this section we will use Stokes' law to see that this wingtip-vortex is closely related to the lift that is generated by the airflow along a wing.

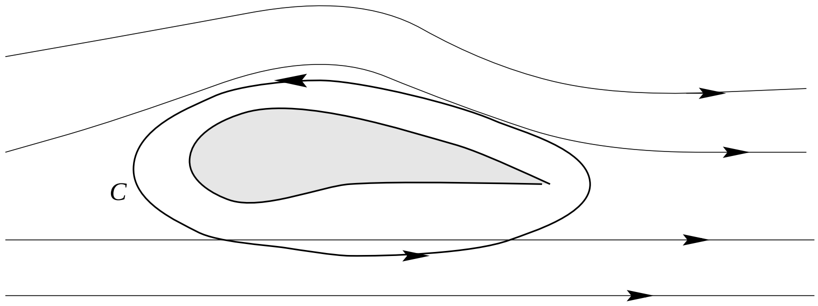


Figure 7.10: Sketch of the flow along an airfoil. The wing is shown in grey, the contour C is shown by the thick solid line.

Let us first consider the flow along a wing, see figure (7.10). A wing can only generate lift when it is curved. In figure (7.10) the air traverses a longer path along the upper part of the wing than along the lower part. The velocity of the airstream along the upper part of the wing is therefore larger than the velocity along the lower part. Because of Bernoulli's law this is the reason that a wing generates lift. (For details of Bernoulli's law and other aspects of the flow along wings see ref. [60].)

Problem a: The *circulation* is defined as the line integral $\oint_C \mathbf{v} \cdot d\mathbf{r}$ of the velocity along a curve. Is the circulation positive or negative for the curve C in figure (7.10) for the indicated sense of integration?

Problem b: Consider now the surface S shown in figure (7.11). Show that the circulation satisfies

$$\oint_C \mathbf{v} \cdot d\mathbf{r} = \int_S \boldsymbol{\omega} \cdot d\mathbf{S} \ , \tag{7.21}$$

where $\boldsymbol{\omega}$ is the vorticity. (See the sections (5.2)-(5.4).)

This expression implies that whenever lift is generated by the circulation along the contour C around the wing, the integral of the vorticity over a surface that envelopes the wingtip is nonzero. The vorticity depends on the derivative of the velocity. Since the flow is relatively smooth along the wing, the derivative of the velocity field is largest near the wingtips. Therefore, expression (7.21) implies that vorticity is generated at the wingtips. As shown in section (5.3) the vorticity is a measure of the local vortex strength. A wing can only produce lift when the circulation along the curve C is nonzero. The above reasoning implies that wingtip vortices are unavoidably associated with the lift produced by an airfoil.

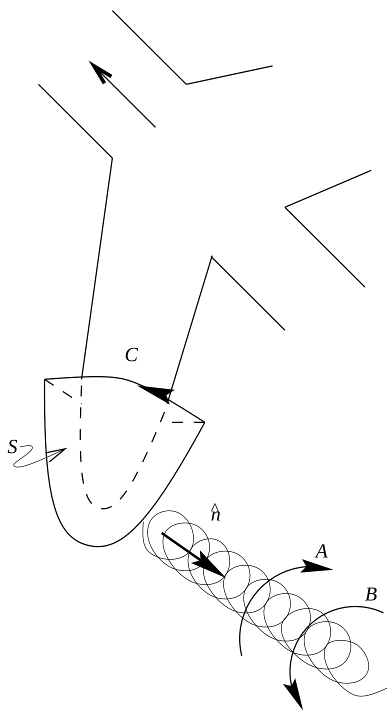


Figure 7.11: Geometry of the surface S and the wingtip vortex for an aircraft seen from above.

Problem c: Consider the wingtip vortex shown in figure (7.11). You have obtained the sign of the circulation $\oint_C \mathbf{v}\cdot d\mathbf{r}$ in **problem a**. Does this imply that the wingtip vortex rotates in the direction A of figure (7.11) or in the direction B? Use equation (7.21) in your argumentation. You may assume that the vorticity is mostly concentrated at the trailing edge of the wingtips, see figure (7.11).

Problem d: The wingtip-vortex obviously carries kinetic energy. As such it entails an undesirable loss of energy for a moving aircraft. Why do aircraft such as the Boeing 747-400 have wingtips that are turned upward? (These are called "winglets.")

Problem e: Just like aircraft, sailing boats suffer from energy loss due a vortex that is generated at the upper part of the sail, see the discussion of *Marchaj*[37]. (A sail can be considered to be a "vertical wing.") Consider the two boats shown in figure (7.12). Suppose that the sails have the same area. If you could choose one of these boats for a race, would you choose the one on the left or on the right? Use equation (7.21) to motivate your choice.

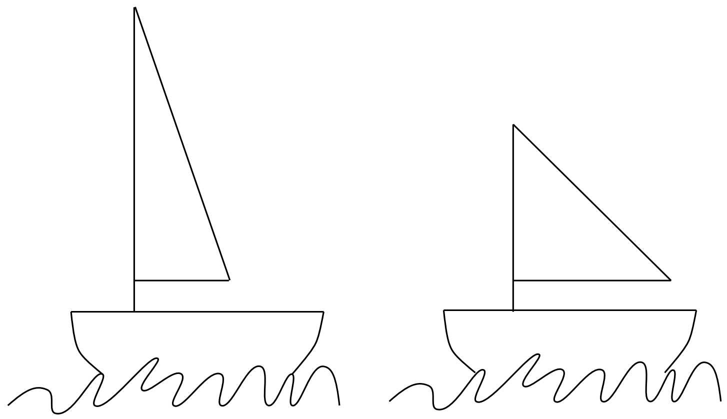


Figure 7.12: Two boats carrying sails with a very different aspect ratio.

Chapter 8

Conservation laws

In physics one frequently handles the change of a property with time by considering properties that do *not* change with time. For example, when two particles collide, the momentum and the energy of each particle may change. However, this change can be found from the consideration that the total momentum and energy of the system are conserved. Often in physics, such conservation laws are main ingredients for describing a system. In this section we deal with conservation laws for continuous systems. These are systems where the physical properties are a continuous function of the space coordinates. Examples are the motion in a fluid or solid, the temperature distribution in a body. The introduced conservation laws are not only of great importance in physics, they also provide worthwhile exercises of the vector calculus introduced in the previous sections.

8.1 The general form of conservation laws

In this section a general derivation of conservation laws is given. Suppose we consider a physical quantity Q. This quantity could denote the mass density of a fluid, the heat content within a solid or any other type of physical variable. In fact, there is no reason why Q should be a scalar, it could also be a vector (such as the momentum density) or a higher order tensor. Let us consider a volume V in space that does not change with time. This volume is bounded by a surface ∂V. The total amount of Q within this volume is given by the integral $\int_V QdV$. The rate of change of this quantity with time is given by $\frac{\partial}{\partial t}\int_V QdV$.

In general, there are two reason for the quantity $\int_V QdV$ to change with time. First, the field Q may have sources or sinks within the volume V, the net source of the field Q per unit volume is denoted with the symbol S. The total source of Q within the volume is simply the volume integral $\int_V SdV$ of the source density. Second, it may be that the quantity Q is transported in the medium. With this transport process, a current $\mathbf{J}$ is associated.

As an example one can think of Q being the mass density of a fluid. In that case $\int_V QdV$ is the total mass of the fluid in the volume. This total mass can change because there is a source of fluid within the volume (i.e. a tap or a bathroom sink), or the total mass may change because of the flow through the boundary of the volume.

The rate of change of $\int_V QdV$ by the current is given by the *inward* flux of the current

J through the surface ∂V. If we retain the convention that the surface element $d\mathbf{S}$ points out off the volume, the inward flux is given by $-\oint_{\partial V} \mathbf{J}\cdot d\mathbf{S}$. Together with the rate of change due to the source density S within the volume this implies that the rate of change of the total amount of Q within the volume satisfies:

$$\frac{\partial}{\partial t}\int_V Q dV = -\oint_{\partial V} \mathbf{J}\cdot d\mathbf{S} + \int_V S dV \ . \tag{8.1}$$

Using Gauss' law (6.1), the surface integral in the right hand side can be written as $-\int_V (\nabla \cdot \mathbf{J}) dV$, so that the expression above is equivalent with

$$\frac{\partial}{\partial t}\int_V Q dV + \int_V (\nabla \cdot \mathbf{J}) dV = \int_V S dV \ . \tag{8.2}$$

Since the volume V is assumed to be fixed with time, the time derivative of the volume integral is the volume integral of the time derivative: $\frac{\partial}{\partial t}\int_V Q dV = \int_V \frac{\partial Q}{\partial t} dV$. It should be noted that expression (8.2) holds for *any* volume V. If the volume is an infinitesimal volume, the volume integrals in (8.2) can be replaced by the integrand multiplied with the infinitesimal volume. Using these results, one finds that expression (8.2) is equivalent with:

$$\frac{\partial Q}{\partial t} + (\nabla \cdot \mathbf{J}) = S \ . \tag{8.3}$$

This is the general form of a conservation law in physics, it simply states that the rate of change of a quantity is due to the sources (or sinks) of that quantity and due to the divergence of the current of that quantity. Of course, the general conservation law (8.3) is not very meaningful as long as we don't provide expressions for the current **J** and the source S. In this section we will see examples where the current and the source follow from physical theory, but we will also encounter examples where they follow from an "educated" guess.

Equation (8.3) will not be completely new to you. In section (6.4) the probability density current for a quantum mechanical system was derived.

Problem a: Use the derivation of this section to show that expression (6.15) can be written as

$$\frac{\partial}{\partial t}|\psi|^2 + (\nabla \cdot \mathbf{J}) = 0 \ , \tag{8.4}$$

with **J** given by expression (6.16).

This equation constitutes a conservation law for the probability density of a particle. Note that equation (6.15) could be derived rigorously from the Schrödinger equation (6.13) so that the conservation law (8.4) and the expression for the current **J** follow from the basic equation of the system.

Problem b: Why is the source term on the right hand side of (8.4) equal to zero?

8.2 The continuity equation

In this section we will consider the conservation of mass in a continuous medium such as a fluid or a solid. In that case, the quantity Q is the mass-density ρ. If we assume that mass is not created or destroyed, the source term vanishes: $S = 0$. The vector $\mathbf{J}$ is the mass current, this quantity denotes the flow of mass per unit volume. Let us consider a small volume δV. The mass within this volume is equal to $\rho\ \delta V$. If the velocity of the medium is denoted with $\mathbf{v}$, the mass-flow is given by $\rho\ \delta V\ \mathbf{v}$. Dividing this by the volume δV one obtains the mass flow per unit volume, this quantity is called the mass density current:

$$\mathbf{J} = \rho\mathbf{v}\ . \tag{8.5}$$

Using these results, the principle of the conservation of mass can be expressed as

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0\ . \tag{8.6}$$

This expression plays a very important role in continuum mechanics and is called the *continuity equation.*

Up to this point the reasoning was based on a volume V that did not change with time. This means that our treatment was strictly Eulerian; we considered the change of physical properties at a fixed location. As an alternative, a Lagrangian description of the same process can be given. In such an approach one specifies how physical properties change as they are moved along with the flow. In that approach one seeks an expression for the total time derivative $\frac{d}{dt}$ of physical properties rather than expressions for the partial derivative $\frac{\partial}{\partial t}$. These two derivatives are related in the following way:

$$\frac{d}{dt} = \frac{\partial}{\partial t} + (\mathbf{v}\cdot\nabla)\ . \tag{8.7}$$

Problem a: Show that the total derivative of the mass density is given by:

$$\frac{d\rho}{dt} + \rho(\nabla\cdot\mathbf{v}) = 0\ . \tag{8.8}$$

Problem b: This Lagrangian expression gives the change of the density when one follows the flow. Let us consider a infinitesimal volume δV that is carried around with the flow. The mass of this volume is given by $\delta m = \rho\delta V$. The mass within that volume is conserved (why?), so that $\delta\dot{m} = 0$. (The dot denotes the time derivative.) Use this expression and equation (8.8) to show that $(\nabla\cdot\mathbf{v})$ is the rate of change of the volume normalized by size of the volume:

$$\frac{\delta\dot{V}}{\delta V} = (\nabla\cdot\mathbf{v})\ . \tag{8.9}$$

We have learned a new meaning of the divergence of the velocity field, it equals the relative change in volume per unit time.

8.3 Conservation of momentum and energy

In the description of a point mass in classical mechanics, the conservation of momentum and energy can be derived from Newton's third law. The same is true for a continuous medium such as a fluid or a solid. In order to formulate Newton's law for a continuous medium we start with a Lagrangian point of view and consider a volume δV that moves with the flow. The mass of this volume is given by $\delta m = \rho \delta V$. This mass is constant. Let the force per unit volume be denoted by $\mathbf{F}$, so that the total force acting on the volume is $\mathbf{F}\delta V$. The force $\mathbf{F}$ contains both forces generated by external agents (such as gravity) and internal agents such as the pressure force $-\nabla p$ or the effect of internal stresses $(\nabla \cdot \sigma)$ (with σ being the stress tensor). Newton's law applied to the volume δV takes the form:

$$\frac{d}{dt}\left(\rho \delta V \mathbf{v}\right) = \mathbf{F}\delta V \ . \tag{8.10}$$

Since the mass $\delta m = \rho \delta V$ is constant with time it can be taken outside the derivative in (8.10). Dividing the resulting expression by δV leads to the Lagrangian form of the equation of motion:

$$\rho \frac{d\mathbf{v}}{dt} = \mathbf{F} \ . \tag{8.11}$$

Note that the density appears *outside* the time derivative, despite the fact that the density may vary with time. Using the prescription (8.7) one obtains the Eulerian form of Newton's law for a continuous medium:

$$\rho \frac{\partial \mathbf{v}}{\partial t} + \rho \mathbf{v} \cdot \nabla \mathbf{v} = \mathbf{F} \ . \tag{8.12}$$

This equation is not yet in the general form (8.3) of conservation laws because in the first term on the left hand side we have the density *times* a time derivative, and because the second term on the left hand side is not the divergence of some current.

Problem a: Use expression (8.12) and the continuity equation (8.6) to show that:

$$\frac{\partial(\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v}\mathbf{v}) = \mathbf{F} \ . \tag{8.13}$$

This expression does take the form of a conservation law; it expresses that the momentum (density) $\rho \mathbf{v}$ is conserved. (For brevity we will often not include the affix "density" in the description of the different quantities, but remember that all quantities are given per unit volume). The source of momentum is given by the force $\mathbf{F}$, this reflects that forces are the cause of changes in momentum. In addition there is a momentum current $\mathbf{J} = \rho \mathbf{v}\mathbf{v}$ that describes the transport of momentum by the flow. This momentum current is not a simple vector, it is a dyad and hence is represented by a 3×3 matrix. This is not surprising since the momentum is a vector with three components and each component can be transported in three spatial directions.

You may find the inner products of vectors and the ∇-operator in expressions such (8.12)-confusing, and indeed a notation such as $\rho \mathbf{v} \cdot \nabla \mathbf{v}$ can be a source of error and confusion. Working with quantities like this is simpler by explicitly writing out the components of all vectors or tensors and by using the Einstein summation convention. (In this convention one sums over all indices that are repeated on one side of the equality sign.) This notation implies the following identities: $\mathbf{v} \cdot \nabla Q = \sum_{i=1}^{3} v_i \partial_i Q = v_i \partial_i Q$ (where ∂_i is an abbreviated equation for $\partial/\partial x_i$), $v^2 = \sum_{i=1}^{3} v_i v_i = v_i v_i$ and an equation such as (8.12) is written in this notation as:

$$\rho \frac{\partial v_i}{\partial t} + \rho v_j \partial_j v_i = F_i \ . \tag{8.14}$$

Problem b: Rewrite the continuity equation (8.6) in component form and redo the derivation of **problem a** with all equations in component form to arrive at the conservation law of momentum in component form:

$$\frac{\partial(\rho v_i)}{\partial t} + \partial_j (\rho v_j v_i) = F_i \ . \tag{8.15}$$

In order to derive the law of energy conservation we start by deriving the conservation law for the kinetic energy (density)

$$E_K = \frac{1}{2}\rho v^2 = \frac{1}{2}\rho v_i v_i \ . \tag{8.16}$$

Problem c: Express the partial time-derivative $\partial(\rho v^2)/\partial t$ in the time derivatives $\partial(\rho v_i)/\partial t$ and $\partial v_i/\partial t$, use the expressions (8.14) and (8.15) to eliminate these time derivatives and write the final results as:

$$\frac{\partial(\frac{1}{2}\rho v_i v_i)}{\partial t} = -\partial_j \left(\frac{1}{2}\rho v_i v_i v_j \right) + v_j F_j \ . \tag{8.17}$$

Problem d: Use definition (8.16) to rewrite the expression above as the conservation law of kinetic energy:

$$\frac{\partial E_K}{\partial t} + \nabla \cdot (\mathbf{v} E_K) = (\mathbf{v} \cdot \mathbf{F}) \ . \tag{8.18}$$

This equation states that the kinetic energy current is given by $\mathbf{J} = \mathbf{v} E_K$, this term describes how kinetic energy is transported by the flow. The term $(\mathbf{v} \cdot \mathbf{F})$ on the right hand side denotes the source of kinetic energy. This term is relatively easy to understand. Suppose the force $\mathbf{F}$ acts on the fluid over a distance $\delta\mathbf{r}$, the work carried out by the force is given by $(\delta\mathbf{r} \cdot \mathbf{F})$. If it takes the fluid a time δt to move over the distance $\delta\mathbf{r}$ the work per unit time is given by $(\delta\mathbf{r}/\delta t \cdot \mathbf{F})$. However, $\delta\mathbf{r}/\delta t$ is simply the velocity $\mathbf{v}$, and hence the term $(\mathbf{v} \cdot \mathbf{F})$ denotes the work performed by the force per unit time. Since work per unit time is called the *power*, equation (8.18) states that the power produced by the force $\mathbf{F}$ is the source of kinetic energy.

In order to invoke the potential energy as well we assume for the moment that the force $\mathbf{F}$ is the gravitational force. Suppose there is a gravitational potential $V(\mathbf{r})$, then the gravitational force is given by

$$\mathbf{F} = -\,\rho \nabla V \ , \tag{8.19}$$

and the potential energy E_P is given by

$$E_P = \rho V \ . \tag{8.20}$$

Problem e: Take the (partial) time derivative of (8.20), use the continuity equation (8.6) to eliminate $\partial\rho/\partial t$, use that the potential V does not depend explicitly on time and employ expressions (8.19) and (8.20) to derive the conservation law of potential energy:

$$\frac{\partial E_P}{\partial t} + \nabla \cdot (\mathbf{v} E_P) = -(\mathbf{v} \cdot \mathbf{F}) \ . \tag{8.21}$$

Note that this conservation law is very similar to the conservation law (8.18) for kinetic energy. The meaning of the second term on the left hand side will be clear to you by now, it denotes the divergence of the current $\mathbf{v}E_P$ of potential energy. Note that the right hand side of (8.20) has the opposite sign of the right hand side of (8.18). This reflects the fact that when the force $\mathbf{F}$ acts as a source of kinetic energy, it acts as a sink of potential energy; the opposite signs imply that kinetic and potential energy are converted into each other. However, the total energy $E = E_K + E_P$ should have no source or sink.

Problem f: Show that the total energy is source-free:

$$\frac{\partial E}{\partial t} + \nabla \cdot (\mathbf{v} E) = 0 \ . \tag{8.22}$$

8.4 The heat equation

In the previous section the momentum and energy current could be derived from Newton's law. Such a rigorous derivation is not always possible. In this section the transport of heat is treated, and we will see that the law for heat transport cannot be derived rigorously. Consider the general conservation equation (8.3) where T is the temperature. (Strictly speaking we should derive the heat equation using a conservation law for the heat content rather than the temperature. The heat content is given by CT, with C the heat capacity. When the specific heat is constant the distinction between heat and temperature implies multiplication with a constant, for simplicity this multiplication is left out.)

The source term in the conservation equation is simply the amount of heat (normalized by the heat capacity) supplied to the medium. For example, the decay of radioactive isotopes is a major source of the heat budget of the earth. The transport of heat is affected by the heat current $\mathbf{J}$. In the earth, heat can be transported by two mechanisms, heat conduction and heat advection. The first process is similar to the process of diffusion, it accounts for the fact that heat flows from warm regions to colder regions. The second process accounts for the heat that is transported by the flow field $\mathbf{v}$ is the medium. Therefore, the current $\mathbf{J}$ can be written as a sum of two components:

$$\mathbf{J} = \mathbf{J}^{conduction} + \mathbf{J}^{advection} \ . \tag{8.23}$$

The heat advection is simply given by

$$\mathbf{J}^{advection} = \mathbf{v} T \ , \tag{8.24}$$

which reflects that heat is simply carried around by the flow. This viewpoint of the process of heat transport is in fact too simplistic in many situation. *Fletcher* [24] describes how the human body during outdoor activities looses heat through four processes; conduction, advection, evaporation and radiation. He describes in detail the conditions under which each of these processes dominate, and how the associated heat loss can be reduced. In the physics of the atmosphere, energy transport by radiation and by evaporation (or condensation) also plays a crucial role.

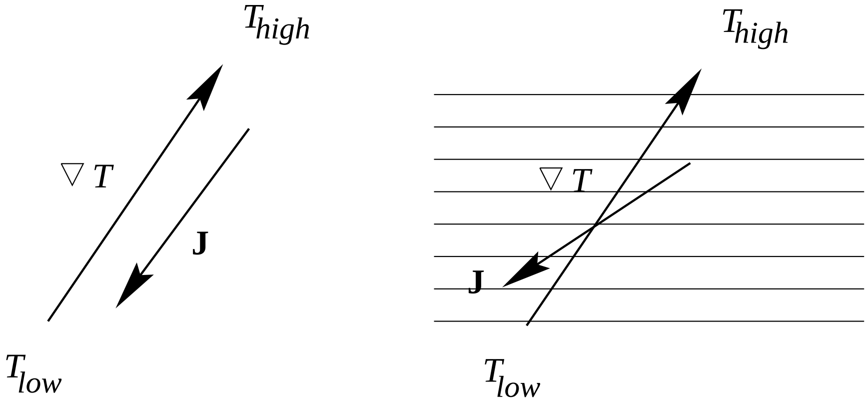


Figure 8.1: Heat flow and temperature gradient in an isotropic medium (left panel) and in a medium consisting of alternating layers of copper and styrofoam (right panel).

For the moment we will focus on the heat conduction.This quantity cannot be derived from first principles. In general, heat flows from warm regions to cold regions. The vector ∇T points from cold regions to warm regions. It therefore is logical that the heat conduction points in the opposite direction from the temperature gradient:

$$\mathbf{J}^{conduction} = -\kappa \nabla T \ , \tag{8.25}$$

see the left panel of figure (8.1). The constant κ is the heat conductivity. (For a given value of ∇T the heat conduction increases when κ increases, hence it measures indeed the conductivity.) However, the simple law (8.25) does not hold for every medium. Consider a medium consisting of alternating layers of a good heat conductor (such as copper) and a poor heat conductor (such as styrofoam). In such a medium the heat will be preferentially transported along the planes of the good heat conductor and the conductive heat flow $\mathbf{J}^{conduction}$ and the temperature gradient are not antiparallel, see the right panel in figure (8.1). In that case there is a matrix operator that relates $\mathbf{J}^{conduction}$ and ∇T: $J_i^{conduction} = -\kappa_{ij}\partial_j T$, with κ_{ij} the heat conductivity tensor. In this section we will restrict ourselves to the simple conduction law (8.25). Combining this law with the expressions (8.23), (8.24) and the conservation law (8.3) for heat gives:

$$\frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{v}T - \kappa \nabla T) = S \ . \tag{8.26}$$

As a first example we will consider a solid in which there is no flow ($\mathbf{v} = 0$). For a constant heat conductivity κ, expression (8.26) reduces to:

$$\frac{\partial T}{\partial t} = \kappa \nabla^2 T + S \ . \tag{8.27}$$

The expression is called the "heat equation", despite the fact that it holds only under special conditions. This expression is identical to Fick's law that accounts for diffusion processes. This is not surprising since heat is transported by a diffusive process in the absence of advection.

We now consider heat transport in a one-dimensional medium (such as a bar) when there is no source of heat. In that case the heat equation reduces to

$$\frac{\partial T}{\partial t} = \kappa \frac{\partial^2 T}{\partial x^2} \ . \tag{8.28}$$

If we know the temperature throughout the medium at some initial time (i.e. $T(x, t = 0)$ is known), then (8.28) can be used to compute the temperature at later times. As a special case we consider a Gaussian shaped temperature distribution at $t = 0$:

$$T(x, t = 0) = T_0 \exp\left(-\frac{x^2}{L^2}\right) \ . \tag{8.29}$$

Problem a: Sketch this temperature distribution and indicate the role of the constants T_0 and L.

We will assume that the temperature profile maintains a Gaussian shape at later times but that the peak value and the width may change, i.e. we will consider a solution of the following form:

$$T(x, t) = F(t) \exp\left(-H(t)x^2\right) \ . \tag{8.30}$$

At this point the function $F(t)$ and $H(t)$ are not yet known.

Problem b: Show that these functions satisfy the initial conditions:

$$F(0) = T_0 \qquad , \qquad H(0) = 1/L^2 \ . \tag{8.31}$$

Problem c: Show that for the special solution (8.30) the heat equation reduces to:

$$\frac{\partial F}{\partial t} - x^2 F \frac{\partial H}{\partial t} = \kappa \left(4FH^2x^2 - 2FH\right) \ . \tag{8.32}$$

It is possible to derive equations for the time evolution of F and H by recognizing that (8.32) can only be satisfied for *all values of* x when all terms proportional to x^2 balance and when the terms independent of x balance.

Problem d: Use this to show that $F(t)$ and $H(t)$ satisfy the following differential equations:

$$\frac{\partial F}{\partial t} = -2\kappa FH \ , \tag{8.33}$$

$$\frac{\partial H}{\partial t} = -4\kappa H^2 \ . \tag{8.34}$$

It is easiest to solve the last equation first because it contains only $H(t)$ whereas (8.33) contains both $F(t)$ and $H(t)$.

Problem e: Solve (8.34) with the initial condition (8.31) and show that:

$$H(t) = \frac{1}{4\kappa t + L^2} \; . \tag{8.35}$$

Problem f: Solve (8.33) with the initial condition (8.31) and show that:

$$F(t) = T_0 \frac{L}{\sqrt{4\kappa t + L^2}} \; . \tag{8.36}$$

Inserting these solutions in expression (8.30) gives the temperature field at all times $t \geq 0$:

$$T(x,t) = T_0 \frac{L}{\sqrt{4\kappa t + L^2}} \exp\left(- \frac{x^2}{4\kappa t + L^2} \right) \; . \tag{8.37}$$

Problem g: Sketch the temperature for several later times and describe using the solution (8.37) how the temperature profile changes as time progresses.

The total heat $Q^{total}(t)$ at time t is given by $Q^{total}(t) = C \int_{-\infty}^{\infty} T(x,t)dx$, where C is the heat capacity.

Problem h: Show that the total heat does not change with time for the solution (8.37). Hint: reduce any integral of the form $\int_{-\infty}^{\infty} \exp\left(-\alpha x^2\right) dx$ to the integral $\int_{-\infty}^{\infty} \exp\left(-u^2\right) du$ with a suitable change of variables. You don't even have to use that $\int_{-\infty}^{\infty} \exp\left(-u^2\right) du = \sqrt{\pi}$.

Problem i: Show that for *any* solution of the heat equation (8.28) where the heat flux vanishes at the endpoints ($\kappa \partial_x T(x = \pm\infty, t) = 0$) the total heat $Q^{total}(t)$ is constant in time.

Problem j: What happens to the special solution (8.37) when the temperature field evolves backward in time? Consider in particular times earlier than $t = -L^2/4\kappa$.

Problem k: The peak value of the temperature field (8.37) decays as $1/\sqrt{4\kappa t + L^2}$ with time. Do you expect that in more dimensions this decay is more rapid or more slowly with time? Don't do any calculations but use your common sense only!

Up to this point, we considered the conduction of heat in a medium without flow ($\mathbf{v} = 0$). In many applications the flow in the medium plays a crucial role in redistributing heat. This is particular the case when heat is the source of convective motions, as for example in the earth's mantle, the atmosphere and the central heating system in buildings. As an example of the role of advection we will consider the cooling model of the oceanic lithosphere proposed by Parsons and Sclater[45].

At the mid-oceanic ridges lithospheric material with thickness H is produced. At the ridge the temperature of this material is essentially the temperature T_m of mantle material. As shown in figure (8.2) this implies that at $x = 0$ and at depth $z = H$ the temperature is given by the mantle temperature: $T(x = 0, z) = T(x, z = H) = T_m$. We assume that the velocity with which the plate moves away from the ridge is constant:

$$\mathbf{v} = U\hat{\mathbf{x}} \; . \tag{8.38}$$

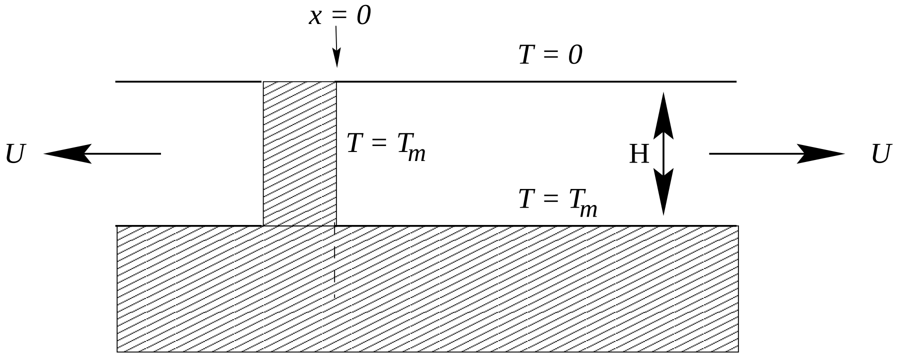


Figure 8.2: Sketch of the cooling model of the oceanic lithosphere.

We will consider the situation that the temperature is stationary. This does not imply that the flow vanishes; it means that the partial time-derivatives vanish: $\partial T/\partial t = 0$, $\partial \mathbf{v}/\partial t = 0$.

Problem l: Show that in the absence of heat sources ($S = 0$) the conservation equation (8.26) reduces to:

$$U\frac{\partial T}{\partial x} = \kappa\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial z^2}\right) \ . \tag{8.39}$$

In general the thickness of the oceanic lithosphere is less than 100 km, whereas the width of ocean basins is several thousand kilometers.

Problem m: Use this fact to explain that the following expression is a reasonable approximation to (8.39):

$$U\frac{\partial T}{\partial x} = \kappa\frac{\partial^2 T}{\partial z^2} \ . \tag{8.40}$$

Problem n: Show that with the replacement $\tau = x/U$ this expression is identical to the heat equation (8.28).

Note that τ is the time it has taken the oceanic plate to move from its point of creation ($x = 0$) to the point of consideration (x), hence the time τ simply is the *age* of the oceanic lithosphere. This implies that solutions of the one-dimensional heat equation can be used to describe the cooling of oceanic lithosphere with the age of the lithosphere taken as the time-variable. Accounting for cooling with such a model leads to a prediction of the depth of the ocean that increases as $\sqrt{t}$ with the age of the lithosphere. For ages less than about 100 $Myear$ this is in very good agreement with the observed ocean depth[45].

8.5 The explosion of a nuclear bomb

As an example of the use of conservation equations we will now study the condition under which a ball of Uranium of Plutonium can explode through a nuclear chain reaction. The starting point is once again the general conservation law (8.3), where Q is the concentration $N(t)$ of neutrons per unit volume. We will assume that the material is solid and assume there is no flow: $\mathbf{v} = 0$. The neutron concentration is affected by two processes. First, the neutrons experience normal diffusion. For simplicity we assume that the neutron current is given by expression (8.25): $\mathbf{J} = -\kappa \nabla N$. Second, neutrons are produced in the nuclear chain reaction. For example, when an atom of U_{235} absorbs one neutron, it may fission and emit three free neutrons. This effectively constitutes a source of neutrons. The intensity of this source depends on the neutrons that are around to produce fission of atoms. This implies that the source term is proportional to the neutron concentration: $S = \lambda N$, where λ is a positive constant that depends on the details of the nuclear reaction.

Problem a: Show that the neutron concentration satisfies:

$$\frac{\partial N}{\partial t} = \kappa \nabla^2 N + \lambda N \ . \tag{8.41}$$

This equation needs to be supplemented with boundary conditions. We will assume that the material that fissions is a sphere with radius R. At the edge of the sphere the neutron concentration vanishes while at the center of the sphere the neutron concentration must remain finite for finite times:

$$N(r = R, t) = 0 \qquad and \qquad N(r = 0, t) \ is\ finite \ . \tag{8.42}$$

We restrict our attention to solutions that are spherically symmetric: $N = N(r, t)$.

Problem b: Apply separation of variables by writing the neutron concentration as $N(r, t) = F(r)H(t)$ and show that $F(r)$ and $H(t)$ satisfy the following equations:

$$\frac{\partial H}{\partial t} = \mu H \ , \tag{8.43}$$

$$\nabla^2 F + \frac{(\lambda - \mu)}{\kappa} F = 0 \ , \tag{8.44}$$

where μ is a separation constant that is not yet known.

Problem c: Show that for positive μ there is an exponential growth of the neutron concentration with characteristic growth time $\tau = 1/\mu$.

Problem d: Use the expression of the Laplacian in spherical coordinates to rewrite (8.44). Make the substitution $F(r) = f(r)/r$ and show that $f(r)$ satisfies:

$$\frac{\partial^2 f}{\partial r^2} + \frac{(\lambda - \mu)}{\kappa} f = 0 \ . \tag{8.45}$$

Problem e: Derive the boundary conditions at $r = 0$ and $r = R$ for $f(r)$.

Problem f: Show that equation (8.45) with the boundary condition derived in **problem e** can only be satisfied when

$$\mu = \lambda - \left(\frac{n\pi}{R} \right)^2 \kappa \qquad for\ integer\ n \ . \tag{8.46}$$

Problem g: Show that for $n = 0$ the neutron concentration vanishes so that we only need to consider values $n \geq 1$.

Equation (8.46) gives the growth rate of the neutron concentration. It can be seen that the effects of unstable nuclear reactions and of neutron diffusion oppose each other. The λ-term accounts for the growth of the neutron concentration through fission reactions, this term makes the inverse growth rate μ more positive. Conversely, the κ-term accounts for diffusion, this term gives a negative contribution to μ.

Problem h: What value of n gives the largest growth rate? Show that exponential growth of the neutron concentration (i.e. a nuclear explosion) can only occur when

$$R > \pi\sqrt{\frac{\kappa}{\lambda}} \; . \tag{8.47}$$

This implies that a nuclear explosion can only occur when the ball of fissionable material is larger than a certain critical size. If the size is smaller than the critical size, more neutrons diffuse out of the ball than are created by fission, hence the nuclear reaction stops. In some of the earliest nuclear devices an explosion was created by bringing two halve spheres that each were a stable together to form one whole sphere that was unstable.

Problem g: Suppose you had a ball of fissionable material that is just unstable and that you shape this material in a cube rather than a ball. Do you expect this cube to be stable or unstable? Don't use any equations!

8.6 Viscosity and the Navier-Stokes equation

Many fluids exhibit a certain degree of viscosity. In this section it will be shown that viscosity can be seen as an ad-hoc description of the momentum current in a fluid by small-scale movements in the fluid. Starting point of the analysis is the equation of momentum conservation in a fluid:

$$\frac{\partial(\rho\mathbf{v})}{\partial t} + \nabla \cdot (\rho\mathbf{v}\mathbf{v}) = \mathbf{F} \qquad\qquad (8.13) \; again.$$

In a real fluid, motion takes places at a large range of length scales from microscopic eddies to organized motions with a size comparable to the size of the fluid body. Whenever we describe a fluid, it is impossible to account for the motions at the very small length scales. This not only so in analytical descriptions, but it is in particular the case in numerical simulations of fluid flow. For example, in current weather prediction schemes the motion of the air is computed on a grid with a distance of about 100 km between the gridpoints. When you look at the weather it is obvious that there is considerable motion at smaller length scales (e.g. cumulus clouds indicating convection, fronts, etc.). In general one cannot simply ignore the motion at these short length scales because these small-scale fluid motions transport significant amounts of momentum, heat and other quantities such as moisture.

One way to account for the effect of the small-scale motion is to express the small-scale motion in the large-scale motion. It is not obvious that this is consistent with reality, but it appears to be the only way to avoid a complete description of the small-scale motion of the fluid (which would be impossible).

In order to do this, we assume there is some length scale that separates the small-scale flow from the large scale flow, and we decompose the velocity in a long-wavelength component $\mathbf{v}^L$ and a short-wavelength component $\mathbf{v}^S$:

$$\mathbf{v} = \mathbf{v}^L + \mathbf{v}^S \ . \tag{8.48}$$

In addition, we will take spatial averages over a length scale that corresponds to the length scale that distinguishes the large-scale flow from the small-scale flow. This average is indicated by brackets: $\langle \cdots \rangle$. The average of the small-scale flow is zero ($\left\langle \mathbf{v}^S \right\rangle = 0$) while the average of the large-scale flow is equal to the large-scale flow ($\left\langle \mathbf{v}^L \right\rangle = \mathbf{v}^L$) because the large-scale flow by definition does not vary over the averaging length. For simplicity we will assume that the density does not vary.

Problem a: Show that the momentum equation for the large-scale flow is given by:

$$\frac{\partial(\rho\mathbf{v}^L)}{\partial t} + \nabla\cdot(\rho\mathbf{v}^L\mathbf{v}^L) + \nabla\cdot(\left\langle \rho\mathbf{v}^S\mathbf{v}^S \right\rangle) = \mathbf{F} \ . \tag{8.49}$$

Show in particular why this expression contains a contribution that is quadratic in the small-scale flow, but that the terms that are linear in $\mathbf{v}^S$ do not contribute.

All terms in (8.49) are familiar, except the last term in the left hand side. This term exemplifies the effect of the small-scale flow on the large-scale flow since it accounts for the transport of momentum by the small-scale flow. It looks that at this point further progress in impossible without knowing the small scale flow $\mathbf{v}^S$. One way to make further progress is to express the small-scale momentum current $\left\langle \rho\mathbf{v}^S\mathbf{v}^S \right\rangle$ in the large scale flow.

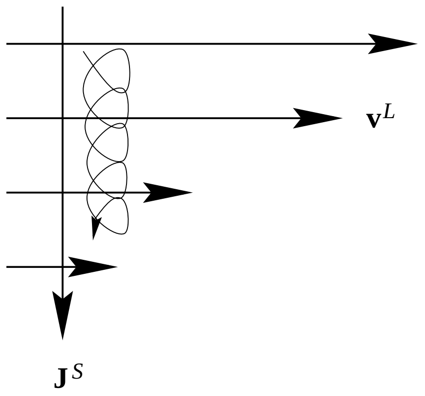


Figure 8.3: he direction of momentum transport within a large-scale flow by small-scale motions.

Consider the large-scale flow shown in figure (8.3). Whatever the small-scale motions are, in general they will have the character of mixing. In the example of the figure, the momentum is large at the top of the figure and the momentum is smaller at the bottom. As a first approximation one may assume that the small-scale motions transport momentum in the direction opposite to the momentum gradient of the large-scale flow. By analogy with (8.25) we can approximate the momentum transport by the small-scale flow by:

$$\mathbf{J}^S \equiv \left\langle \rho \mathbf{v}^S \mathbf{v}^S \right\rangle \approx -\mu \nabla \mathbf{v}^L \ , \tag{8.50}$$

where μ plays the role of a diffusion constant.

Problem b: Insert this relation in (8.49), drop the superscript L of $\mathbf{v}^L$ to show that large-scale flow satisfies:

$$\frac{\partial(\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v}\mathbf{v}) = \mu \nabla^2 \mathbf{v} + \mathbf{F} \ . \tag{8.51}$$

This equation is called the Navier-Stokes equation. The first term on the right hand side accounts for the momentum transport by small-scale motions. Effectively this leads to viscosity of the fluid.

Problem c: Viscosity tends to damp motions at smaller length-scales more than motion at larger length scales. Show that the term $\mu \nabla^2 \mathbf{v}$ indeed affects shorter length scales more than larger length scales.

Problem d: Do you think this treatment of the momentum flux due to small-scale motions is realistic? Can you think of an alternative?

Despite reservations that you may (or may not) have against the treatment of viscosity in this section, you should realize that the Navier-Stokes equation (8.51) is widely used in fluid mechanics.

8.7 Quantum mechanics = hydrodynamics

As we have seen in section (6.4) the behavior of microscopic particles is described by Schrödinger's equation

$$-\frac{\hbar}{i}\frac{\partial \psi(\mathbf{r},t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2 \psi(\mathbf{r},t) + V(\mathbf{r})\psi(\mathbf{r},t) \ , \qquad (6.13)\ \textit{again}$$

rather than Newton's law. In this section we reformulate the linear wave equation (6.13) as the laws of conservation of mass and momentum for a normal fluid. In order to do this write the wave function ψ as

$$\psi = \sqrt{\rho} \times \exp\left(\frac{i}{\hbar}\varphi\right) \ . \tag{8.52}$$

This equation is simply the decomposition of a complex function in its absolute value and its phase, hence ρ and φ are real functions. The factor $\hbar$ is added for notational convenience.

Problem a: Insert the decomposition (8.52) in Schrödinger's equation (6.13), divide by $\sqrt{\rho}\exp\left(\frac{i}{\hbar}\varphi\right)$ and separate the result in real and imaginary parts to show that ρ and φ satisfy the following differential equations:

$$\partial_t\rho + \nabla\cdot\left(\rho\frac{1}{m}\nabla\varphi\right) = 0 \ , \tag{8.53}$$

$$\partial_t\varphi + \frac{1}{2m}\left|\nabla\varphi\right|^2 + \frac{\hbar^2}{8m}\left(\frac{1}{\rho^2}\left|\nabla\rho\right|^2 - \frac{2}{\rho}\nabla^2\rho\right) = -V \ . \tag{8.54}$$

The problem is that at this point we do not have a velocity yet. Let us define the following velocity vector:

$$\mathbf{v} \equiv \frac{1}{m}\nabla\varphi \ . \tag{8.55}$$

Problem b: Show that this definition of the velocity is identical to the velocity obtained in equation (6.19) of section (6.4).

Problem c: Show that with this definition of the velocity, expression (8.53) is identical to the continuity equation:

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0 \ . \qquad (8.6)\ \textit{again}$$

Problem d: In order to reformulate (8.54) as an equation of conservation of momentum, differentiate (8.54) with respect to x_i. Do this, use the definition (8.55) and the relation between force and potential ($\mathbf{F} = -\nabla V$) to write the result as:

$$\partial_t v_i + \frac{1}{2}\partial_i(v_jv_j) + \frac{\hbar^2}{8m}\left(\partial_i\left(\frac{1}{\rho^2}\left|\nabla\rho\right|^2\right) - 2\partial_i\left(\frac{1}{\rho}\nabla^2\rho\right)\right) = \frac{1}{m}F_i \ . \tag{8.56}$$

The second term on the left hand side does not look very much to the term $\partial_j(\rho v_jv_i)$ in the left hand side of (8.13) To make progress we need to rewrite the term $\partial_i(v_jv_j)$ into a term of the form $\partial_j(v_jv_i)$. In general these terms are different.

Problem e: Show that for the special case that the velocity is the gradient of a scalar function (as in expression (8.55)) that:

$$\frac{1}{2}\partial_i(v_jv_j) = \partial_j(v_jv_i) \ . \tag{8.57}$$

With this step we can rewrite the second term on the left hand side of (8.56). Part of the third term in (8.56) we will designate as Q_i:

$$Q_i \equiv -\frac{1}{8}\left(\partial_i\left(\frac{1}{\rho^2}\left|\nabla\rho\right|^2\right) - 2\partial_i\left(\frac{1}{\rho}\nabla^2\rho\right)\right) \ . \tag{8.58}$$

Problem f: Using equations (8.6) and (8.56) through (8.58) derive that:

$$\partial_t(\rho\mathbf{v}) + \nabla\cdot(\rho\mathbf{v}\mathbf{v}) = \frac{\rho}{m}\left(\mathbf{F}+\hbar^2\mathbf{Q}\right) \ . \tag{8.59}$$

Note that this equation is identical with the momentum equation (8.13). This implies that the Schrödinger equation is equivalent with the continuity equation (8.6) and the momentum equation (8.13) for a classical fluid. In section (6.4) we have seen that microscopic particles behave as waves rather than point-like particles. In this section we discovered that particles also behave like a fluid. This has led to hydrodynamic formulations of quantum mechanics[28]. In general, quantum-mechanical phenomena depend critically on Planck's constant. Quantum mechanics reduces to classical mechanics in the limit $\hbar \to 0$. The only place where Planck's constant occurs in (8.59) is the additional force $\mathbf{Q}$ that multiplied with Planck's constant. This implies that the action of the force term $\mathbf{Q}$ is fundamentally quantum-mechanical, it has no analogue in classical mechanics.

Problem g: Suppose we consider a particle in one dimension that is represented by the following wave function:

$$\psi(x,t) = \exp\left(-\frac{x^2}{L^2}\right) \exp i\left(kx - \omega t\right) \ . \tag{8.60}$$

Sketch the corresponding probability density ρ and use (8.58) to deduce that the quantum force acts to broaden the wave function with time.

This example shows that (at least for this case) the quantum force $\mathbf{Q}$ makes the wave function "spread-out" with time. This reflects the fact that if a particle propagates with time, its position becomes more and more uncertain.

Chapter 9

Scale analysis

In most situations, the equations that we would like to solve in mathematical physics are too complicated to solve analytically. One of the reasons for this is often that an equation contains many different terms which make the problem simply too complex to be manageable. However, many of these terms may in practice be very small. Ignoring these small terms can simplify the problem to such an extent that it can be solved in closed form. Moreover, by deleting terms that are small one is able to focus on the terms that are significant and that contain the relevant physics. In this sense, ignoring small terms can actually give a better physical insight in the processes that really do matter.

Scale analysis is a technique where one estimates the different terms in an equation by considering the scale over which the relevant parameters vary. This is an extremely powerful too for simplifying problems. A comprehensive overview of this technique with many applications is given by *Kline*[33].

Many of the equations that are used in physics are differential equations. For this reason it is crucial in scale analysis to be able to estimate the order of magnitude of derivatives. The estimation of derivatives is therefore treated first. In subsequent sections this is then applied to a variety of different problems.

9.1 Three ways to estimate a derivative

In this section three different ways are derived to estimate the derivative of a function $f(x)$. The *first* way to estimate the derivative is to realize that the derivative is nothing but the slope of the function $f(x)$. Consider figure 9.1 in which the function $f(x)$ is assumed to be known in neighboring points x and $x+h$.

Problem a: Deduce from the geometry of this figure that the slope of the function at x is approximately given by $(f(x+h)-f(x))/h$.

Since the slope is the derivative this means that the derivative of the function is approximately given by

$$\frac{df}{dx} \approx \frac{f(x+h)-f(x)}{h} \tag{9.1}$$

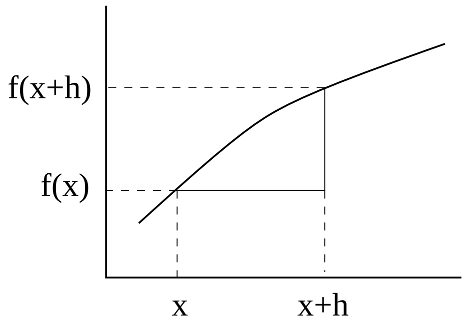


Figure 9.1: The slope of a function $f(x)$ that is known at positions x and $x+h$.

The *second* way to derive the same result is to realize that the derive is defined by the following limit:

$$\frac{df}{dx} \equiv \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \ . \tag{9.2}$$

If we consider the right hand side of this expression without taking the limit, we do not quite obtain the derivative, but as long as h is sufficiently small we obtain the approximation (9.1).

The problem with estimating the derivative of $f(x)$ in the previous ways is that we do obtain an estimate of the derivative, but we do not know how good these estimates are. We do know that if $f(x)$ would be a straight line, which has a constant slope, that the estimate (9.1) would be exact. Hence is it the deviation of $f(x)$ from a straight line that makes (9.1) only an approximation. This means that it is the curvature of $f(x)$ that accounts for the error in the approximation (9.1). The *third* way of estimating the derivative provides this error estimate as well.

Problem b: Consider the Taylor series (2.17) of section 2.1. Truncate this series after the second order term and solve the resulting expression for df/dx to derive that

$$\frac{df}{dx} = \frac{f(x+h) - f(x)}{h} - \frac{1}{2}\frac{d^2 f}{dx^2} h + \cdots \tag{9.3}$$

where the dots indicate terms of order h^2.

In the limit $h \to 0$ the last term vanishes and expression (9.2) is obtained. When one ignores the last term in (9.3) for finite h one obtains the approximation (9.1) once more.

Problem c: Use expression (9.3) to show that the error made in the approximation (9.1) depends indeed on the *curvature* of the function $f(x)$.

The approximation (9.1) has a variety of applications. The first is the numerical solution of differential equations. Suppose one has a differential equation that one cannot solve in closed form. to fix out mind consider the differential equation

$$\frac{df}{dx} = G(f(x), x) \ , \tag{9.4}$$

with initial value

$$f(0) = f_0 \; . \tag{9.5}$$

When this equation cannot be solved in closed form, one can solve it numerically by evaluating the function $f(x)$ not for every value of x, but only at a finite number of x-values that are separated by a distance h. These points x_n are given by $x_n = nh$, and the function $f(x)$ at location x_n is denoted by f_n:

$$f_n \equiv f(x_n) \; . \tag{9.6}$$

Problem d: Show that the derivative df/dx at location x_n can be approximated by:

$$\frac{df}{dx}(x_n) = \frac{1}{h}(f_{n+1} - f_n) \; . \tag{9.7}$$

Problem e: Insert this result in the differential equation (9.4) and solve the resulting expression for f_{n+1} to show that:

$$f_{n+1} = f_n + hG(f_n, x_n) \; . \tag{9.8}$$

This is all we need to solve the differential equation (9.4) with the boundary condition (9.5) numerically. Once f_n is known, (9.8) can be used to compute f_{n+1}. This means that the function can be computed at all values of the grid points x_n recursively. To start this process, one uses the boundary condition (9.5) that gives the value of the function at location $x_0 = 0$. This technique for estimating the derivative of a function can be extended to higher order derivatives as well so that second order differential equations can also be solved numerically. In practice, one has to pay serious attention to the *stability* of the numerical solution. The requirements of stability and numerical efficiency have led to many refinements of the numerical methods for solving differential equations. The interested reader can consult *Press et al.*[47] as an introduction and many practical algorithms.

The estimate (9.1) has a second important application because it allows us to estimate the order of magnitude of a derivative. Suppose a function $f(x)$ varies over a characteristic range of values F and that this variation takes place over a characteristic distance L. It follows from (9.1) that the derivative of $f(x)$ is of the order of the ratio of the variation of the function $f(x)$ divided by the length-scale over which the function varies. In other words:

$$\left|\frac{df}{dx}\right| \approx \frac{\text{variation of the function } f(x)}{\text{length scale of the variation}} \sim \frac{F}{L} \; . \tag{9.9}$$

In this expression the term $\sim F/L$ indicates that the derivative is of the order F/L. Note that this is in general not an accurate estimate of the precise value of the function $f(x)$, it only provides us with an estimate of the order of magnitude of a derivative. However, this is all we need to carry out scale analysis.

Problem f: Suppose $f(x)$ is a sinusoidal wave with amplitude A and wavelength λ:

$$f(x) = A \sin\left(\frac{2\pi x}{\lambda}\right) \; . \tag{9.10}$$

Show that (9.9) implies that the order of magnitude of the derivative of this function is given by $|df/dx| \sim O(A/\lambda)$. Compare this estimate of the order of magnitude with the true value of the derivative and pay attention both to the numerical value as well as to the spatial variation.

From the previous estimate we can learn two things. First, the estimate (9.9) is only a rough estimate that *locally* can be very poor. One should always be aware that the estimate (9.9) may break down at certain points and that this can cause errors in the subsequent scale analysis. Second, the estimate (9.9) differs by a factor 2π from the true derivative. However, $2\pi = 6.28\cdots$ which is not a small number. Therefore you must be aware that hidden numerical factors may enter scaling arguments.

9.2 The advective terms in the equation of motion

As a first example of scale analysis we consider the role of advective terms in the equation of motion. As shown in expression (8.12) of section 8.3 the equation of motion for a continuous medium is given by

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = \frac{1}{\rho} \mathbf{F} \ . \tag{9.11}$$

Note that we have divided by the density compared to the original expression (8.12). This equation can describe the propagation of acoustic waves when $\mathbf{F}$ is the pressure force, it accounts for elastic waves when $\mathbf{F}$ is given by the elastic forces in the medium. We will be interested in the situation where waves with a wavelength λ and a period T propagate through the medium.

The advective terms $\mathbf{v} \cdot \nabla \mathbf{v}$ often pose a problem in solving this equation. The reason is that the partial time derivative $\partial \mathbf{v}/\partial t$ is *linear* in the velocity $\mathbf{v}$ but that the advective terms $\mathbf{v} \cdot \nabla \mathbf{v}$ are *nonlinear* in the velocity $\mathbf{v}$. Since linear equations are in general much easier to solve than nonlinear equations it is very useful to know under which conditions the advective terms $\mathbf{v} \cdot \nabla \mathbf{v}$ can be ignored compared to the partial derivative $\partial \mathbf{v}/\partial t$.

Problem a: Let the velocity of the continuous medium have a characteristic value V. Show that $|\partial \mathbf{v}/\partial t| \sim V/T$ and that $|\mathbf{v} \cdot \nabla \mathbf{v}| \sim V^2/\lambda$.

Problem b: Show that this means that the ratio of the advective terms to the partial time derivative is given by

$$\frac{|\mathbf{v} \cdot \nabla \mathbf{v}|}{|\partial \mathbf{v}/\partial t|} \sim \frac{V}{c} \ , \tag{9.12}$$

where c is the velocity with which the waves propagate through the medium.

This result implies that the advective terms can be ignored when the velocity of the medium itself is much less than the velocity which the waves propagate through the medium:

$$V \ll c \, . \tag{9.13}$$

In other words, when the amplitude of the wave motion is so small that the velocity of the medium is much less than the wave velocity one can ignore the advective terms in the equation of motion.

Problem c: Suppose an earthquake causes at a large distance a ground displacement of 1 mm at a frequency of 1 Hz. The wave velocity of seismic P-waves is of the order of 5 km/s near the surface. Show that in that case $V/c \sim 10^{-9}$.

The small value of V/c implies that for the propagation of elastic waves due to earthquakes one can ignore advective terms in the equation of motion. Note, however, that this is not necessarily true near the earthquake where the motion is much more violent and where the associated velocity of the rocks is not necessarily much smaller than the wave velocity.

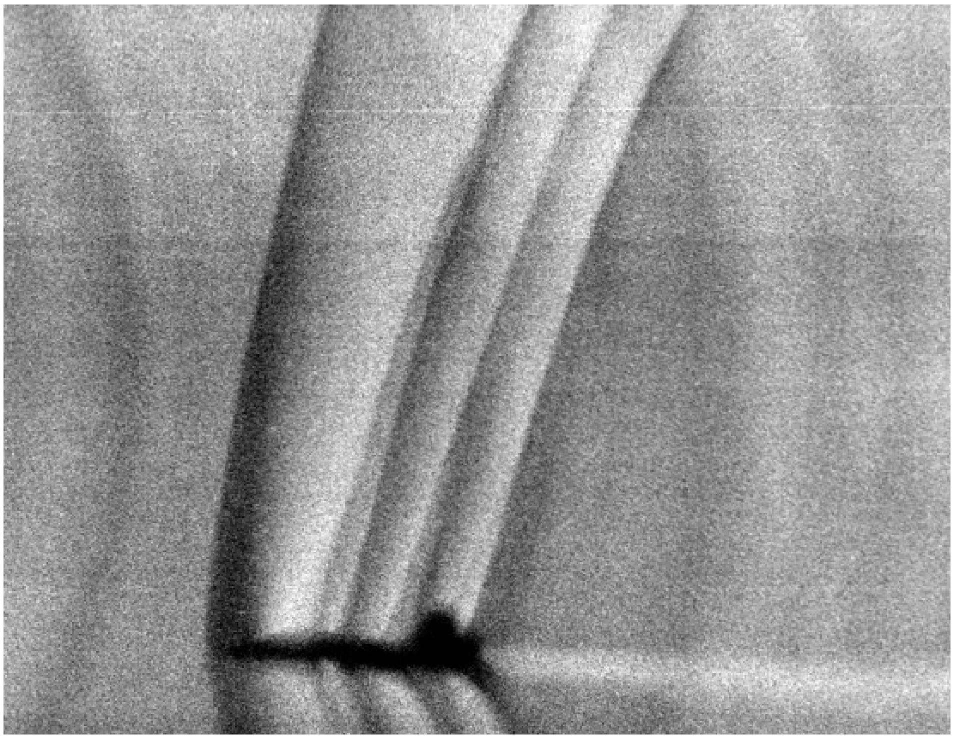


Figure 9.2: The shock waves generated by a T38 flying at Mach 1.1 (a speed of 1.1 times the speed of sound) as made visible as made visibible with the schlieren method.

There are a number of physical phenomena that are intimately related to the presence of the advective terms in the equation of motion. One important phenomenon is the occurrence of shock waves when the motion of the medium is comparable to the wave velocity. A prime example of shock waves is the sonic boom made by aircraft that move at a velocity equal to the speed of sound[32]. Since the air pushed around by the aircraft moves with the same velocity as the aircraft, shock waves are generated when the velocity of the aircraft is equal to the speed of sound. A spectacular example can be seen in figure 9.2 where the shock waves generated by an T38 flying at a speed of Mach 1.1 at an altitude of 13.700 ft can be seen. These shock waves are visualised using the schlieren method [36] which is an optical technique to convert phase differences of light waves in amplitude differences.

Another example of shock waves is the formation of the *hydraulic jump*. You may not known what a hydraulic jump is, but you have surely seen one! Consider water flowing down a channel such as a mountain stream as shown in figure 9.3. The flow velocity is

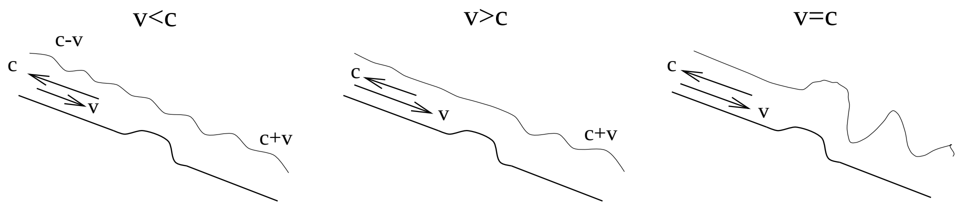


Figure 9.3: Waves on a water flowing over a rock when $v < c$ (left panel), $v > c$ (middle panel) and $v = c$ (right panel.

denoted by v. At the bottom of the channel a rock is disrupting the flow. This rock generates water-waves that propagate with a velocity c compared to the moving water. When the flow velocity is less than the wave velocity ($v < c$, see the left panel of figure 9.3) the waves propagate upstream with an absolute velocity $c - v$ and propagate downstream with an absolute velocity $c + v$. When the flow velocity is larger than the wave velocity ($v > c$, see the middle panel of figure 9.3) the waves move downstream only because the wave velocity is not sufficiently large to move the waves against the current. The most interesting case is when the flow velocity equals the wave velocity ($v = c$, see the right panel of figure 9.3). In that case the waves that move upstream have an absolute velocity given by $c - v = 0$. In other words, these waves do not move with respect to the rock that generates the waves. This wave is continuously excited by the rock, and through a process similar to an oscillator that is driven at its resonance frequency the wave grows and grows until it ultimately breaks and becomes turbulent. This is the reason why one can see strong turbulent waves over boulders and other irregularities in streams. For further details on channel flow and hydraulic jumps the reader can consult chapter 9 of *Whitaker*[67]. In general the advective terms play a crucial role steepening and breaking of waves and the formation of shock waves. This is described in much detail by *Whitham*[66].

9.3 Geometric ray theory

Geometric ray theory is an approximation that accounts for the propagation of waves along lines through space. The theory finds is conceptual roots in optics, where for a long time one has observed that a light beam propagates along a well-defined trajectory through lenses and many other optical devices. Mathematically, this behavior of waves is accounted for in geometric ray theory, or more briefly "ray theory."

Ray theory is derived here for the acoustic wave equation rather than for the propagation of light because pressure waves are described by a scalar equation rather than the vector equation that governs the propagation of electromagnetic waves. The starting point is the acoustic wave equation (6.7) of section 6.3:

$$\rho \nabla \cdot \left(\frac{1}{\rho} \nabla p \right) + \frac{\omega^2}{c^2} p = 0. \tag{9.14}$$

For simplicity the source term in the right hand side has been set to zero. In addition, the relation $c^2 = \kappa/\rho$ has been used to eliminate the bulk modulus κ in favor of the wave velocity c. Both the density and the wave velocity are arbitrary functions of space.

In general it is not possible to solve this differential equation in closed form. Instead we will seek an approximation by writing the pressure as:

$$p(\mathbf{r},\omega)=A(\mathbf{r},\omega)e^{i\psi(\mathbf{r},\omega)}\ , \tag{9.15}$$

with A and ψ real functions. Any function $p(\mathbf{r},\omega)$ can be written in this way.

Problem a: Insert the solution (9.15) in the acoustic wave equation (9.14), separate the real and imaginary parts of the resulting equation to deduce that (9.14) is equivalent to the following equations:

$$\underbrace{\nabla^2 A}_{(1)}-\underbrace{A\left|\nabla\psi\right|^2}_{(2)}-\underbrace{\frac{1}{\rho}\left(\nabla\rho\cdot\nabla A\right)}_{(3)}+\underbrace{\frac{\omega^2}{c^2}A}_{(4)}=0\ , \tag{9.16}$$

and

$$2\left(\nabla A\cdot\nabla\psi\right)+A\nabla^2\psi-\frac{1}{\rho}\left(\nabla\rho\cdot\nabla\psi\right)A=0\ . \tag{9.17}$$

The equations are even harder to solve than the acoustic wave equation because they are nonlinear in the unknown functions A and ψ whereas the acoustic wave equation is linear in the pressure p. However, the equations (9.16) and (9.17) form a good starting point for making the ray-geometric approximation. First we will analyze expression (9.16).

Assume that the density varies on a length scale L_ρ, that the amplitude A of the wave-field varies on a characteristic length scale L_A. Furthermore the wavelength of the waves is denoted by λ.

Problem b: Explain that the wavelength is the length-scale over which the phase ψ of the waves varies.

Problem c: Use the results of section 9.1 to obtain the following estimates of the order of magnitude of the terms (1) $-$ (4) in equation (9.16):

$$\left|\nabla^2 A\right|\sim\frac{A}{L_A^2}\qquad A\left|\nabla\psi\right|^2\sim\frac{A}{\lambda^2}\qquad\left|\frac{1}{\rho}\left(\nabla\rho\cdot\nabla A\right)\right|\sim\frac{A}{L_A L_\rho}\qquad\frac{\omega^2}{c^2}A\sim\frac{A}{\lambda^2} \tag{9.18}$$

To make further progress we assume that the length-scale of both the density variations and the amplitude variations are much longer than a wavelength: $\lambda\ll L_A$ and $\lambda\ll L_\rho$.

Problem d: Show that under this assumption the terms (1) and (3) in equation (9.16) are much smaller than the terms (2) and (4).

Problem e: Convince yourself that ignoring the terms (1) and (3) in (9.16) gives the following (approximate) expression:

$$\left|\nabla\psi\right|^2=\frac{\omega^2}{c^2}\ . \tag{9.19}$$

Problem f: The approximation (9.19) was obtained under the premise that $\left|\nabla\psi\right|\sim 1/\lambda$. Show that this assumption is satisfied by the function ψ in (9.19).

Whenever one makes approximations by deleting terms that scale-analysis predicts to be small one has to check that the final solution is consistent with the scale-analysis that is used to derive the approximation.

Note that the original equation (9.16) contains both the amplitude A and the phase ψ but that (9.19) contains the phase only. The approximation that we have made has thus decoupled the phase from the amplitude, this simplifies the problem considerably. The frequency enters the right hand side of this equation only through a simple multiplication with ω^2. The frequency dependence of ψ can be found by substituting

$$\psi(\mathbf{r}, \omega) = \omega\tau(\mathbf{r}) \ . \tag{9.20}$$

Problem g: Show that the equations (9.19) and (9.17) after this substitution are given by:

$$|\nabla\tau(\mathbf{r})|^2 = \frac{1}{c^2} \ , \tag{9.21}$$

and

$$2\left(\nabla A \cdot \nabla\tau\right) + A\nabla^2\tau - \frac{1}{\rho}\left(\nabla\rho \cdot \nabla\tau\right) A = 0 \ . \tag{9.22}$$

According to (9.21) the function $\tau(\mathbf{r})$ does not depend on frequency. Note that equation (9.22) for the amplitude does not contain any frequency dependence either. This means that the amplitude also does not depend on frequency: $A = A(\mathbf{r})$. This has important consequences for the shape of the wave-field in the ray-geometric approximation. Suppose that the wave-field is excited by a source-function $s(t)$ in the time domain that is represented in the frequency domain by a complex function $S(\omega)$. (The forward and backward Fourier-transform is defined by the equations (11.42) and (11.43) of section 11.5.) In the frequency domain the response is given by expression (9.15) multiplied with the source function $S(\omega)$. Using that A and τ do not depend on frequency the pressure in the time domain can be written as:

$$p(\mathbf{r}, t) = \int_{-\infty}^{\infty} A(\mathbf{r}) e^{i\omega\tau(\mathbf{r})} e^{-i\omega t} S(\omega) d\omega \ . \tag{9.23}$$

Problem h: Use this expression to show that the pressure in the time domain can be written as:

$$p(\mathbf{r}, t) = A(\mathbf{r}) s(t - \tau(\mathbf{r})) \ . \tag{9.24}$$

This is a very important result because it implies that the time-dependence of the wave-field is everywhere given by the same source-time function $s(t)$. In a ray-geometric approximation the shape of the waveforms is everywhere the same. The are no frequency-dependent effects in a ray geometric approximation.

Problem i: Explain why this implies that geometric ray theory can not be used to explain why the sky is blue.

The absence of any frequency-dependent wave propagation effects is both the strength and the weakness is ray theory. It is a strength because the wave-fields can be computed in a simple way once $\tau(\mathbf{r})$ and $A(\mathbf{r})$ are known. The theory also tells us that this is an adequate description of the wave-field as long as the frequency is sufficiently high that

$\lambda \ll L_A$ and $\lambda \ll L_\rho$. However, many wave propagation phenomena are in practice frequency-dependent, it is the weakness of ray theory that it cannot account for these phenomena.

According to expression (9.24) the function $\tau(\mathbf{r})$ accounts for the time-delay of the waves to travel to the point $\mathbf{r}$. Therefore, $\tau(\mathbf{r})$ is the *travel time* of the wave-field. The travel time is described by the differential equation (9.21), this equation is called the *eikonal equation.*

Problem j: Show that it follows from the eikonal equation that $\nabla\tau$ can be written as:

$$\nabla\tau = \hat{\mathbf{n}}/\mathbf{c}\,, \tag{9.25}$$

where $\hat{\mathbf{n}}$ is a unit vector. Show also that $\hat{\mathbf{n}}$ is perpendicular to the surface τ =constant.

The vector $\hat{\mathbf{n}}$ defines the direction of the *rays* along which the wave energy propagates through the medium. Taking suitable derivatives of expression (9.25) one can derive the *equation of kinematic ray-tracing.* This is a second-order differential equation for the position of the rays, details are given by *Virieux*[63] or *Aki and Richards*[2].

Once $\tau(\mathbf{r})$ is known, one can compute the amplitude $A(\mathbf{r})$ from equation (9.22). We have not yet applied any scale-analysis to this expression. We will not do this, because it can be solved exactly. Let us first simplify this differential equation by considering the dependence on the density ρ in more detail.

Problem k: Write $A=\rho^\alpha B$, where the constant α is not yet determined. Show that the transport equation results in the following differential equation for $B(\mathbf{r})$:

$$(2\alpha-1)\left(\nabla\rho\cdot\nabla\tau\right)B + 2\rho\left(\nabla B\cdot\nabla\tau\right) + \rho B\nabla^2\tau = 0\ . \tag{9.26}$$

Choose the constant α in such a way that the gradient of the density disappears from the equation and show that the remaining terms can be written as $\nabla\cdot\left(B^2\nabla\tau\right)=0$. Show finally using (9.25) that this implies the following differential equation for the amplitude:

$$\nabla\cdot\left(\frac{1}{\rho c}A^2\hat{\mathbf{n}}\right) = 0\ . \tag{9.27}$$

Equation (9.27) states that the divergence of the vector $\left(A^2/\rho c\right)\hat{\mathbf{n}}$ vanishes, hence the flux of this vector through any closed surface that does not contain the source of the wave-field vanishes, see section 6.1. This is not surprising, because the vector $\left(A^2/\rho c\right)\hat{\mathbf{n}}$ accounts for the energy flux of acoustic waves. Expression (9.27) implies that the net flux of this vector through any closed surface is equal to zero. This means that all the energy that flows in the surface must also flow out through the surface again. The transport equation in the form (9.27) is therefore a statement of energy conservation. *Virieux*[63] or *Aki and Richards*[2] show how one can compute this amplitude once the location of rays is known.

An interesting complication arises when the energy is focussed in a point or on a surface in space. Such an area of focussing is called a *caustic.* A familiar example of a caustic is the rainbow. One can show that at a caustic, the ray-geometric approximation leads to an infinte amplitude of the wave-field [63].

Problem l: Show that when the amplitude becomes infinite in a finite region of space the condition $\lambda \ll L_A$ must be violated.

This means that ray theory is not valid in or near a caustic. A clear account of the physics of caustics can be found in refs. [9] and [34]. The former reference contains many beautiful images of caustics.

9.4 Is there convection in the Earth's mantle?

The Earth is a body that continuously looses heat to outer space. This heat is a remnant of the heat that has been converted from the gravitational energy during the Earth's formation, but more importantly this heat is generated by the decay of unstable isotopes in the Earth. This heat is transported to the Earth's surface, and the question we aim to address here is: is the heat transported by conduction or by convection?

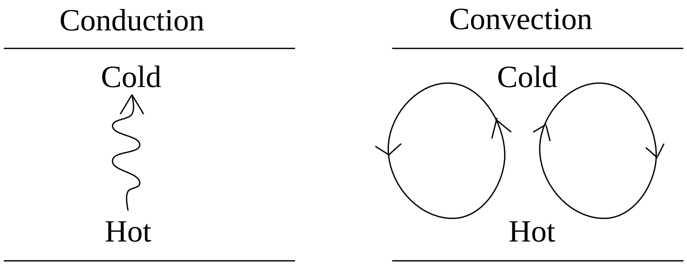


Figure 9.4: Two alternatives for the heat transport in the Earth. In the left panel the material does not move and heat is transported by conduction. In the right panel the material flows and heat is tranported by convection.

If the material in the Earth would not flow, heat could only be transported by conduction. This means that it is the average transfer of the molecular motion from warm regions to cold regions that is responsible for the transport of heat. On the other hand, if the material in the Earth would flow, heat could be carried by the flow. This process is called convection.

The starting point of the analysis is the heat equation (8.26) of section 8.4. In the absence of source terms this equation can for the special case of a constant heat conduction coefficient κ be written as:

$$\frac{\partial T}{\partial t} + \nabla \cdot (\mathbf{v}T) = \kappa \nabla^2 T \, . \tag{9.28}$$

The term $\nabla \cdot (\mathbf{v}T)$ describes the convective heat transport while the term $\kappa \nabla^2 T$ accounts for the conductive heat transport.

Problem a: Let the characteristic velocity be denoted by V, the characteristic length scale by L, and the characteristic temperature perturbation by T. Show that the ratio of the convective heat transport to the conductive heat transport is of the following order:

$$\frac{\text{convective heat transport}}{\text{conductive heat transport}} \sim \frac{VL}{\kappa} \tag{9.29}$$

This estimate gives the ratio of the two modes of heat transport, but it does not help us too much yet because we do not know the order of magnitude V of the flow velocity. This quantity can be obtained from the Navier-Stokes equation of section 8.6:

$$\frac{\partial(\rho\mathbf{v})}{\partial t}+\nabla\cdot(\rho\mathbf{v}\mathbf{v})=\mu\nabla^2\mathbf{v}+\mathbf{F} \qquad (8.51) \qquad \textit{again}$$

The force $\mathbf{F}$ in the right hand side is the buoyancy force that is associated with the flow while the term $\mu\nabla^2\mathbf{v}$ accounts for the viscosity of the flow with viscosity coefficient μ. The mantle of Earth's is extremely viscous and mantle convection (if it exists at all) is a very slow process. We will therefore assume that the inertia term $\partial(\rho\mathbf{v})/\partial t$ and the advection term $\nabla\cdot(\rho\mathbf{v}\mathbf{v})$ are small compared to the viscous term $\mu\nabla^2\mathbf{v}$. (This assumption would have to be supported by a proper scale analysis.) Under this assumption, the mantle flow is predominantly governed by a balance between the viscous force and the buoyancy force:

$$\mu\nabla^2\mathbf{v}=-\mathbf{F}\ . \tag{9.30}$$

The next step is to relate the buoyancy force in the temperature perturbation T. A temperature perturbation T from a reference temperature T_0 leads to a density perturbation ρ from the reference temperature ρ_0 given by:

$$\rho=-\alpha T\ . \tag{9.31}$$

In this expression α is the thermal expansion coefficient that accounts for the expansion or contraction of material due to temperature changes.

Problem b: Explain why for most materials $\alpha>0$. A notable exception is water at temperatures below $4^\circ C$.

Problem c: Write $\rho\left(T_0+T\right)=\rho_0+\rho$ and use the Taylor expansion (2.11) of section 2.1 truncated after the first order term to show that the expansion coefficient is given by $\alpha=-\partial\rho/\partial T$.

Problem d: The buoyancy forces is given by Archimedes' law which states that this force equals the weight of the displaced fluid. Use this result, (9.30) and (9.31) in a scale analysis to show that the velocity is of the following order:

$$V\sim\frac{g\alpha TL^2}{\mu}\ , \tag{9.32}$$

where g is the acceleration of gravity.

Problem e: Use this to derive that the ratio of the convective heat transport to the conductive heat transport is given by:

$$\frac{\text{convective heat transport}}{\text{conductive heat transport}}\sim\frac{g\alpha TL^2}{\mu\kappa} \tag{9.33}$$

The right hand side of this expression is dimensionless, this term is called the *Rayleigh number* which is denoted by Ra:

$$Ra \equiv \frac{g\alpha T L^2}{\mu\kappa} . \tag{9.34}$$

The Rayleigh number is an indicator for the mode of heat transport. When $Ra \gg 1$ heat is predominantly transported by convection. When the thermal expansion coefficient α is large and when the viscosity μ and the heat conduction coefficient κ are small the Rayleigh number is large and heat is transported by convection.

Problem f: Explain physically why a large value of α and small values of μ and κ lead to convective heat transport rather than conductive heat transport.

Dimensionless numbers play a crucial role in fluid mechanics. A discussion of the Rayleigh number and other dimesion-less diagnostics such as the Prandtl number and the Grashof number can be found in section 14.2 of *Tritton*[60]. The implications on the different values of the Rayleigh number on the character of convection in the Earth's mantle is discussed in refs. [43] and [62]. Of course, if one want to use a scale analysis one must know the values of the physical properties involved. For the Earth's mantle, the thermal expansion coefficient α is not very well known because of the complications involved in laboratory measurements of the thermal expansion under the extremely high ambient pressure of Earth's mantle[16].

9.5 Making an equation dimensionless

Usually the terms in the equations that one wants to analyze have a physical dimension such as temperature, velocity, etc. It can sometimes be useful to re-scale all the variables in the equation in such a way that the rescaled variables are dimensionless. This is convenient when setting up numerical solutions of the equations, but in general it also introduces dimensionless numbers that govern the physics of the problem in a natural way. As an example we will apply this technique here to the heat equation (9.28).

Any variable can be made dimensionless by dividing out a constant that has the dimension of the variable. As an example, let the characteristic temperature variation be denoted by T_0, the dimensional temperature perturbation can then be written as:

$$T = T_0 T' . \tag{9.35}$$

The quantity T' is dimesion-less. In this section, dimensionless variables are denoted with a prime. Of course we may not know all the suitable scale factors a-priori. For example, let the characteristic time used for scale the time-variable be denoted by τ:

$$t = \tau t' . \tag{9.36}$$

We can still leave τ open and later choose a value that simplifies the equations as much as possible. Of course when we want to express the heat equation (9.28) in the new time variable we need to specify how the dimensional time derivative $\partial/\partial t$ is related to the dimensionless time derivative $\partial/\partial t'$.

Problem a: Use the chain-rule for differentiation to show that

$$\frac{\partial}{\partial t} = \frac{1}{\tau}\frac{\partial}{\partial t'} \ . \tag{9.37}$$

Problem b: Let the velocity be scaled with the characteristic velocity (9.32):

$$\mathbf{v} = \frac{g\alpha T_0 L^2}{\mu}\mathbf{v}' \ , \tag{9.38}$$

and let the position vector be scaled with the characteristic length L of the system: $\mathbf{r} = L\mathbf{r}'$. Use a result similar to (9.37) to convert the spatial derivatives to the new space coordinate and re-scale all terms in the heat equation (9.28) to derive the following dimensionless form of this equation

$$\frac{L^2}{\tau\kappa}\frac{\partial T'}{\partial t'} + \frac{g\alpha T_0 L^3}{\mu\kappa}\nabla' \cdot \left(\mathbf{v}' T'\right) = \nabla'^2 T' \ , \tag{9.39}$$

where ∇' is the gradient operator with respect to the dimensionless coordinates $\mathbf{r}'$.

At this point we have not specified the time-scale τ for the scaling of the time variable yet. The equation (9.39) simplifies as much as possible when we choose τ in such a way that the constant that multiplies $\partial T'/\partial t'$ is equal to unity:

$$\tau = L^2/\kappa \ . \tag{9.40}$$

Problem c: Suppose heat would only be transported by conduction: $\partial T/\partial t = \kappa\nabla^2 T$. Use a scale analysis to show that τ given by (9.40) is the characteristic time-scale for heat conduction.

This means that the scaling τ of the time variable expresses the time in units of the characteristic diffusion time for heat.

Problem d: Show that with this choice of τ the dimensionless heat equation is given by:

$$\frac{\partial T'}{\partial t'} + Ra\nabla' \cdot \left(\mathbf{v}' T'\right) = \nabla'^2 T' \ , \tag{9.41}$$

where Ra is the Rayleigh number.

The advantage of this dimensionless equation over the original heat equation is that (9.41) contains only a single constant Ra whereas the dimensional heat equation (9.28) depends on a large number of constant. In addition, the scaling of the heat equation has led in a natural way to the key role of the Rayleigh number in the mode of heat transport in a fluid.

Problem e: Use (9.41) to show that convective heat transport dominates over conductive heat transport when $Ra \gg 1$.

Problem f: Suppose that this condition is satisfied and that heat conduction plays a negligible role. Show that the characteristic time-scale of the dimensionless time $t^{'}$ is much less than unity. Give a physical interpretation of this result.

Transforming dimensional equations to dimensionless equations is often used to derive the relevant dimensionless physical constants of the system as well as for setting up algorithms for solving systems numerically. The basic rationale behind this approach is that the physical units that are used are completely arbitrary. It is immaterial whether we express length in meters or in inches, but of course the numerical values of a given length changes when we change form meters to inches. Making the system dimensionless removes all physical units from the system because all the resulting terms in the equation are dimensionless.

Chapter 10

Linear algebra

In this chapter several elements of linear algebra are treated that have important applications in (geo)physics or that serve to illustrate methodologies used in other areas of mathematical physics

10.1 Projections and the completeness relation

In mathematical physics, projections play an extremely important role. This is not only in linear algebra, but also in the analysis of linear systems such as linear filters in data processing (see section 11.10) and the analysis of vibrating systems such as the normal modes of the earth. Let us consider a vector **v** that we want to project along a unit vector **n̂**, see figure (10.1). In the examples of this section we will work in a three-dimensional space, but the arguments presented here can be generalized to any number of dimensions.

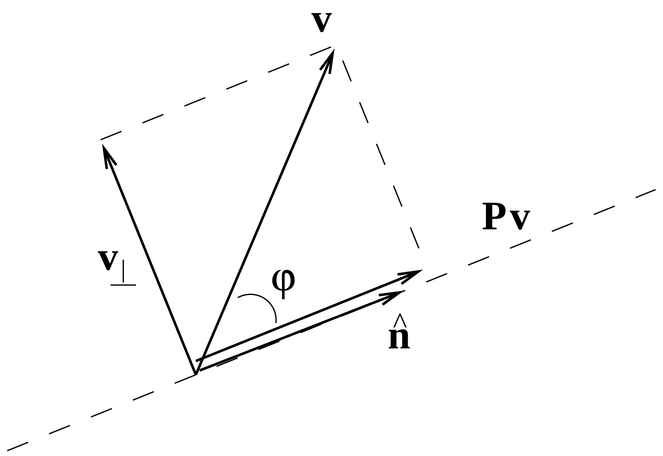


Figure 10.1: Definition of the geometric variables for the projection of a vector.

We will denote the projection of **v** along **n̂** as **Pv**, where **P** stands for the projection operator. In a three-dimensional space this operator can be represented by a 3×3 matrix. It is our goal to find the operator **P** in terms of the unit vector **n̂** as well as the matrix form of this operator. By definition the projection of $\mathbf{v}$ is directed along $\hat{\mathbf{n}}$, hence:

$$\mathbf{P}\mathbf{v} = C\hat{\mathbf{n}} \; . \tag{10.1}$$

This means that we know the projection operator once the constant C is known.

Problem a: Express the length of the vector $\mathbf{P}\mathbf{v}$ in the length of the vector $\mathbf{v}$ and the angle φ of figure (10.1) and express the angle φ in the inner product of the vectors $\mathbf{v}$ and $\hat{\mathbf{n}}$ to show that: $C = (\hat{\mathbf{n}} \cdot \mathbf{v})$.

Inserting this expression for the constant C in (10.1) leads to an expression for the projection $\mathbf{P}\mathbf{v}$:

$$\mathbf{P}\mathbf{v} = \hat{\mathbf{n}} \left(\hat{\mathbf{n}} \cdot \mathbf{v} \right) \; . \tag{10.2}$$

Problem b: Show that the component $\mathbf{v}_\perp$ perpendicular to $\hat{\mathbf{n}}$ as defined in figure (10.1) is given by:

$$\mathbf{v}_\perp = \mathbf{v} - \hat{\mathbf{n}} \left(\hat{\mathbf{n}} \cdot \mathbf{v} \right) \; . \tag{10.3}$$

Problem c: As an example, consider the projection along the unit vector along the x-axis: $\hat{\mathbf{n}} = \hat{\mathbf{x}}$. Show using the equations (10.2) and (10.3) that in that case:

$$\mathbf{P}\mathbf{v} = \begin{pmatrix} v_x \\ 0 \\ 0 \end{pmatrix} \qquad and \qquad \mathbf{v}_\perp = \begin{pmatrix} 0 \\ v_y \\ v_z \end{pmatrix} .$$

Problem d: When we project the projected vector $\mathbf{P}\mathbf{v}$ once more along the same unit vector $\hat{\mathbf{n}}$ the vector will not change. We therefore expect that $\mathbf{P}(\mathbf{P}\mathbf{v}) = \mathbf{P}\mathbf{v}$. Show using expression (10.2) that this is indeed the case. Since this property holds for any vector $\mathbf{v}$ we can also write it as:

$$\mathbf{P}^2 = \mathbf{P} \; . \tag{10.4}$$

Problem e: If $\mathbf{P}$ would be a scalar the expression above would imply that $\mathbf{P}$ is the identity operator $\mathbf{I}$. Can you explain why (10.4) does *not* imply that $\mathbf{P}$ is the identity operator?

In expression (10.2) we derived the action of the projection operator on a vector $\mathbf{v}$. Since this expression holds for any vector $\mathbf{v}$ it can be used to derive an explicit form of the projection operator:

$$\mathbf{P} = \hat{\mathbf{n}}\hat{\mathbf{n}}^T \; . \tag{10.5}$$

This expression should not be confused withe the inner product $(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}})$, instead it denotes the dyad of the vector $\hat{\mathbf{n}}$ and itself. The superscript T denotes the transpose of a vector or matrix. The transpose of a vector (or matrix) is found by interchanging rows and columns. For example, the transpose $\mathbf{A}^T$ of a matrix $\mathbf{A}$ is defined by:

$$A^T_{ij} = A_{ji} \; , \tag{10.6}$$

and the transpose of the vector $\mathbf{u}$ is defined by:

$$\mathbf{u}^T = (u_x, u_y, u_z) \qquad when \qquad \mathbf{u} = \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} , \tag{10.7}$$

i.e. taking the transpose converts a column vector into a row vector. The projection operator $\mathbf{P}$ is written in (10.5) as a dyad. In general the dyad $\mathbf{T}$ of two vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$\mathbf{T} = \mathbf{u}\mathbf{v}^T . \tag{10.8}$$

This is an abstract way to define a dyad, it simply means that the components T_{ij} of the dyad are defined by

$$T_{ij} = u_i v_j , \tag{10.9}$$

where u_i is the i-component of $\mathbf{u}$ and v_j is the j-component of $\mathbf{v}$.

In the literature you will find different notations for the inner-product of two vectors. The inner product of the vectors $\mathbf{u}$ and $\mathbf{v}$ is sometimes written as

$$(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u}^T \mathbf{v} . \tag{10.10}$$

Problem f: Considering the vector $\mathbf{v}$ as a 1×3 matrix and the vector $\mathbf{v}^T$ as a 3×1 matrix, show that the notation used in the right hand sides of (10.10) and (10.8) is consistent with the normal rules for matrix multiplication.

Equation (10.5) relates the projection operator $\mathbf{P}$ to the unit vector $\hat{\mathbf{n}}$. From this the representation of the projection operator as a 3×3-matrix can be found by computing the dyad $\hat{\mathbf{n}}\hat{\mathbf{n}}^T$.

Problem g: Show that the operator for the projection along the unit vector $\hat{\mathbf{n}} = \frac{1}{\sqrt{14}} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ is given by

$$\mathbf{P} = \frac{1}{14} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix} .$$

Verify explicitly that for this example $\mathbf{P}\hat{\mathbf{n}} = \hat{\mathbf{n}}$, and explain this result.

Up to this point we projected the vector $\mathbf{v}$ along a single unit vector $\hat{\mathbf{n}}$. Suppose we have a set of mutually orthogonal unit vectors $\hat{\mathbf{n}}_i$. The fact that these unit vectors are mutually orthogonal means that different unit vectors are perpendicular to each other: $(\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_j) = 0$ when $i \neq j$. We can project $\mathbf{v}$ on each of these unit vectors and add these projections. This gives us the projection of $\mathbf{v}$ on the subspace spanned by the unit vectors $\hat{\mathbf{n}}_i$:

$$\mathbf{P}\mathbf{v} = \sum_i \hat{\mathbf{n}}_i \left(\hat{\mathbf{n}}_i \cdot \mathbf{v} \right) . \tag{10.11}$$

When the unit vectors $\hat{\mathbf{n}}_i$ span the full space we work in, the projected vector is identical to the original vector. To see this, consider for example a three-dimensional space. Any vector can be decomposed in the components along the x, y and z-axis, this can be written as:

$$\mathbf{v} = v_x\hat{\mathbf{x}} + v_y\hat{\mathbf{y}} + v_z\hat{\mathbf{z}} = \hat{\mathbf{x}}\left(\hat{\mathbf{x}}\cdot\mathbf{v}\right) + \hat{\mathbf{y}}\left(\hat{\mathbf{y}}\cdot\mathbf{v}\right) + \hat{\mathbf{z}}\left(\hat{\mathbf{z}}\cdot\mathbf{v}\right) \ , \tag{10.12}$$

note that this expression has the same form as (10.11). This implies that when we sum in (10.11) over a set of unit vectors that completely spans the space we work in, the right hand side of (10.11) is identical to the original vector $\mathbf{v}$, i.e. $\sum_i \hat{\mathbf{n}}_i\left(\hat{\mathbf{n}}_i\cdot\mathbf{v}\right) = \mathbf{v}$. The operator of the left hand side of this equality is therefore identical to the identity operator $\mathbf{I}$:

$$\sum_{i=1}^{N} \hat{\mathbf{n}}_i\hat{\mathbf{n}}_i^T = \mathbf{I}\ . \tag{10.13}$$

Keep in mind that N is the dimension of the space we work in, if we sum over a smaller number of unit vectors we project on a subspace of the N-dimensional space. Expression (10.13) expresses that the vectors $\hat{\mathbf{n}}_i$ (with $i = 1, \cdots, N$) can be used to give a complete representation of any vector. Such a set of vectors is called a *complete set*, and expression (10.13) is called the *closure relation*.

Problem h: Verify explicitly that when the unit vectors $\hat{\mathbf{n}}_i$ are chosen to be the unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ along the x, y and z-axis that the right hand side of (10.13) is given by the 3×3 identity matrix.

There are of course many different ways of choosing a set of three orthogonal unit vectors in three dimensions. Expression (10.13) should hold for every choice of a complete set of unit vectors.

Problem i: Verify explicitly that when the unit vectors $\hat{\mathbf{n}}_i$ are chosen to be the unit vectors $\hat{\mathbf{r}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ defined in equations (3.6) for a system of spherical coordinates that the right hand side of (10.13) is given by the 3×3 identity matrix.

10.2 A projection on vectors that are not orthogonal

In the previous section we considered the projection on a set of orthogonal unit vectors. In this section we consider an example of a projection on a set of vectors that is not necessarily orthogonal. Consider two vectors $\mathbf{a}$ and $\mathbf{b}$ in a three-dimensional space. These two vectors span a two-dimensional plane. In this section we determine the projection of a vector $\mathbf{v}$ on the plane spanned by the vectors $\mathbf{a}$ and $\mathbf{b}$, see figure (10.2) for the geometry of the problem. The projection of $\mathbf{v}$ on the plane will be denoted by $\mathbf{v}_P$.

By definition the projected vector $\mathbf{v}_P$ lies in the plane spanned by $\mathbf{a}$ and $\mathbf{b}$, this vector can therefore be written as:

$$\mathbf{v}_P = \alpha\mathbf{a} + \beta\mathbf{b}\ . \tag{10.14}$$

The task of finding the projection can therefore be reduced to finding the two coefficients α and β. These constants follow from the requirement that the vector joining $\mathbf{v}$ with its projection $\mathbf{v}_P = \mathbf{P}\mathbf{v}$ is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$, see figure (10.2).

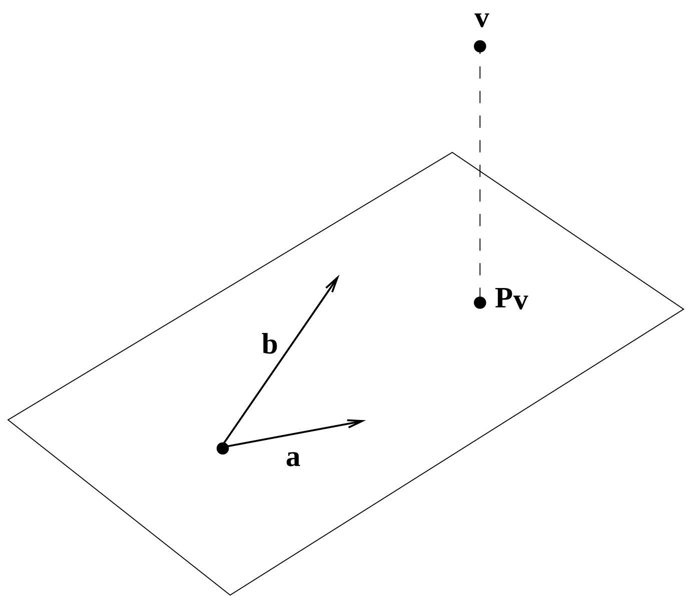


Figure 10.2: Definition of the geometric variables for the projection on a plane.

Problem a: Show that this requirement is equivalent with the following system of equations for α and β:

$$\begin{array}{c} \alpha\,(\mathbf{a}\cdot\mathbf{a})+\beta\,(\mathbf{a}\cdot\mathbf{b})=(\mathbf{a}\cdot\mathbf{v}) \\ \alpha\,(\mathbf{a}\cdot\mathbf{b})+\beta\,(\mathbf{b}\cdot\mathbf{b})=(\mathbf{b}\cdot\mathbf{v}) \end{array} \tag{10.15}$$

Problem b: Show that the solution of this system is given by

$$\alpha=\frac{\left(b^2\mathbf{a}-(\mathbf{a}\cdot\mathbf{b})\,\mathbf{b}\right)\cdot\mathbf{v}}{a^2b^2-(\mathbf{a}\cdot\mathbf{b})^2}\ ,$$

$$\beta=\frac{\left(a^2\mathbf{b}-(\mathbf{a}\cdot\mathbf{b})\,\mathbf{a}\right)\cdot\mathbf{v}}{a^2b^2-(\mathbf{a}\cdot\mathbf{b})^2}\ , \tag{10.16}$$

where a denotes the length of the vector $\mathbf{a}$: $a\equiv|\mathbf{a}|$, and a similar notation is used for the vector $\mathbf{b}$.

Problem c: Show using (10.14) and (10.16) that the projection operator for the projection on the plane ($\mathbf{Pv}=\mathbf{v}_P$) is given by

$$\mathbf{P}=\frac{1}{a^2b^2-(\mathbf{a}\cdot\mathbf{b})^2}\left(b^2\mathbf{a}\mathbf{a}^T+a^2\mathbf{b}\mathbf{b}^T-(\mathbf{a}\cdot\mathbf{b})\left(\mathbf{a}\mathbf{b}^T+\mathbf{b}\mathbf{a}^T\right)\right)\ . \tag{10.17}$$

This example shows that projection on a set of non-orthogonal basis vectors is much more complex than projecting on a set of orthonormal basis vectors. A different way of finding the projection operator of expression (10.17) is by first finding two orthogonal unit vectors in the plane spanned by $\mathbf{a}$ and $\mathbf{b}$ and then using expression (10.11). One unit vector can be found by dividing $\mathbf{a}$ by its length to give the unit vector $\hat{\mathbf{a}}=\mathbf{a}/|\mathbf{a}|$. The second unit vector can be found by considering the component $\mathbf{b}_\perp$ of $\mathbf{b}$ perpendicular to $\hat{\mathbf{a}}$ and by normalizing the resulting vector to form the unit vector $\hat{\mathbf{b}}_\perp$ that is perpendicular to $\hat{\mathbf{a}}$, see figure (10.3).

Problem d: Use expression (10.3) to find $\hat{\mathbf{b}}_\perp$ and show that the projection operator $\mathbf{P}$ of expression (10.17) can also be written as

$$\mathbf{P} = \hat{\mathbf{a}}\hat{\mathbf{a}}^T + \hat{\mathbf{b}}_\perp \hat{\mathbf{b}}_\perp^T \ . \tag{10.18}$$

Note that this expression is consistent with (10.11).

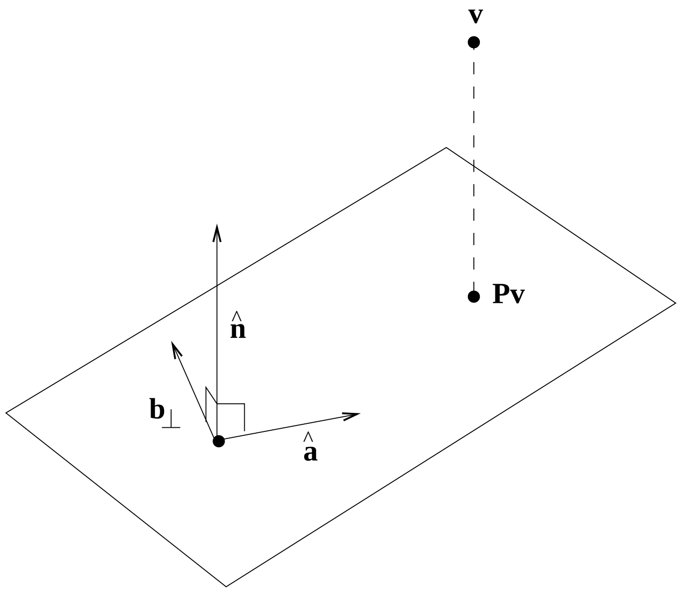


Figure 10.3: Definition of the normal vector to a plane.

Up to this point the plane was defined by the vectors $\mathbf{a}$ and $\mathbf{b}$ (or equivalently by the orthonormal unit vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}_\perp$). However, a plane can also be defined by the unit vector $\hat{\mathbf{n}}$ that is perpendicular to the plane, see figure (10.3). In fact, the unit vectors $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}_\perp$ and $\hat{\mathbf{n}}$ form a complete orthonormal basis of the three-dimensional space. According to equation (10.13) this implies that $\hat{\mathbf{a}}\hat{\mathbf{a}}^T + \hat{\mathbf{b}}_\perp \hat{\mathbf{b}}_\perp^T + \hat{\mathbf{n}}\hat{\mathbf{n}}^T = \mathbf{I}$. With (10.18) this implies that the projection operator $\mathbf{P}$ can also be written as

$$\mathbf{P} = \mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}^T \ . \tag{10.19}$$

Problem e: Give an alternative derivation of this result. Hint, let the operator in equation (10.19) act on an arbitrary vector $\mathbf{v}$.

10.3 The Householder transformation

Linear systems of equations can be solved in a systematic way by sweeping columns of the matrix that defines the system of equations. As an example consider the system

$$\begin{aligned} x + y + 3z &= 5 \\ -x + 2z &= 1 \\ 2x + y + 2z &= 5 \end{aligned} \tag{10.20}$$

This system of equations will be written here also as:

$$\left(\begin{array}{rrr|r} 1 & 1 & 3 & 5 \\ -1 & 0 & 2 & 1 \\ 2 & 1 & 2 & 5 \end{array} \right) \tag{10.21}$$

This is nothing but a compressed notation of the equations (10.20), the matrix shown in (10.21) is called the *augmented matrix* because the matrix defining the left hand side of (10.20) is augmented with the right hand side of (10.20). The linear equations can be solved by adding the first row to the second row and subtracting the first row twice from the third row, the resulting system of equations is then represented by the following augmented matrix:

$$\left(\begin{array}{rrr|r} 1 & 1 & 3 & 5 \\ 0 & 1 & 5 & 6 \\ 0 & -1 & -4 & -5 \end{array} \right) \tag{10.22}$$

Note that in the first column all elements below the first elements are equal to zero. By adding the second row to the third row we can also make all elements below the second element in the second column equal to zero:

$$\left(\begin{array}{rrr|r} 1 & 1 & 3 & 5 \\ 0 & 1 & 5 & 6 \\ 0 & 0 & 1 & 1 \end{array} \right) \tag{10.23}$$

The system is now in *upper-triangular form*, this is a different way of saying that all matrix elements below the diagonal vanish. This is convenient because the system can now be solved by *backsubstitution*. To see how this works note that the augmented matrix (10.23) is a shorthand notation for the following system of equations:

$$\begin{aligned} x + y + 3z &= 5 \\ y + 5z &= 6 \\ z &= 1 \end{aligned} \tag{10.24}$$

The value of z follows from the last equation, given this value of z the value of y follows from the middle equations, given y and z the value of x follows from the top equation.

Problem a: Show that the solution of the linear equations is given by $x = y = z = 1$.

For small systems of linear equations this process for solving linear equations can be carried out by hand. For large systems of equations this process must be carried out on a computer. This is only possible when one has a systematic and efficient way of carrying out this sweeping process. Suppose we have an $N \times N$ matrix $\mathbf{A}$:

$$\mathbf{A} = \left(\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{array} \right) . \tag{10.25}$$

We want to find an operator $\mathbf{Q}$ such that when $\mathbf{A}$ is multiplied with $\mathbf{Q}$ all elements in the first column are zero except the element above or on the diagonal, i.e. we want to find $\mathbf{Q}$ such that:

$$\mathbf{QA} = \begin{pmatrix} a'_{11} & a'_{12} & \cdots & a'_{1N} \\ 0 & a'_{22} & \cdots & a'_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a'_{N2} & \cdots & a'_{NN} \end{pmatrix} . \tag{10.26}$$

This problem can be formulated slightly differently, suppose we denote the first columns of $\mathbf{A}$ by the vector $\mathbf{u}$:

$$\mathbf{u} \equiv \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{N1} \end{pmatrix} . \tag{10.27}$$

The operator $\mathbf{Q}$ that we want to find maps this vector to a new vector which only has a nonzero component in the first element:

$$\mathbf{Qu} = \begin{pmatrix} a'_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = a'_{11}\hat{\mathbf{e}}_1 , \tag{10.28}$$

where $\hat{\mathbf{e}}_1$ is the unit vector in the x_1-direction:

$$\hat{\mathbf{e}}_1 \equiv \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{10.29}$$

The desired operator $\mathbf{Q}$ can be found with a Householder transformation. For a given unit vector $\hat{\mathbf{n}}$ the Householder transformation is defined by:

$$\mathbf{Q} \equiv \mathbf{I} - \mathbf{2}\hat{\mathbf{n}}\hat{\mathbf{n}}^T . \tag{10.30}$$

Problem b: Show that the Householder transformation can be written as $\mathbf{Q} = \mathbf{I} - \mathbf{2P}$, where $\mathbf{P}$ is the operator for projection along $\hat{\mathbf{n}}$.

Problem c: It follows from (10.3) that any vector $\mathbf{v}$ can be decomposed in a component along $\hat{\mathbf{n}}$ and a perpendicular component: $\mathbf{v} = \hat{\mathbf{n}}\left(\hat{\mathbf{n}} \cdot \mathbf{v}\right) + \mathbf{v}_\perp$. Show that after the Householder transformation the vector is given by:

$$\mathbf{Qv} = -\hat{\mathbf{n}}\left(\hat{\mathbf{n}} \cdot \mathbf{v}\right) + \mathbf{v}_\perp \tag{10.31}$$

Problem d: Convince yourself that the Householder transformation of $\mathbf{v}$ is correctly shown in figure (10.4).

Problem e: Use equation (10.31) to show that $\mathbf{Q}$ does not change the length of a vector. Use this result to show that a'_{11} in equation (10.28) is given by $a'_{11} = |\mathbf{u}|$.

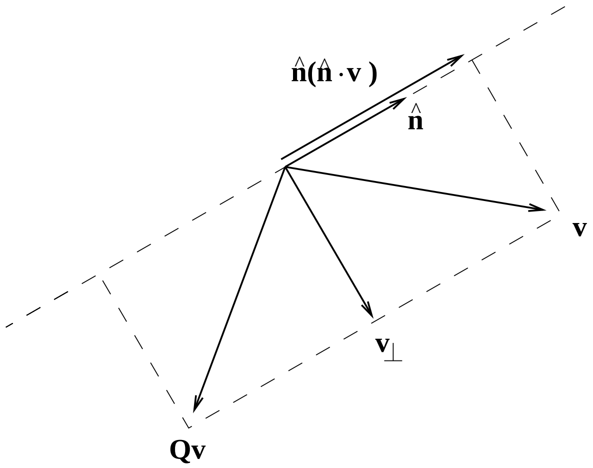


Figure 10.4: Geometrical interpretation of the Householder transformation.

With (10.28) this means that the Householder transformation should satisfy

$$\mathbf{Q}\mathbf{u} = |\mathbf{u}|\,\hat{\mathbf{e}}_1 \ . \tag{10.32}$$

Our goal is now to find a unit vector $\hat{\mathbf{n}}$ such that this expression is satisfied.

Problem f: Use (10.30) to show if $\mathbf{Q}$ satisfies the requirement (10.32) that $\hat{\mathbf{n}}$ must satisfy the following equation:

$$2\hat{\mathbf{n}}\,(\hat{\mathbf{n}}\cdot\hat{\mathbf{u}}) = \hat{\mathbf{u}} - \hat{\mathbf{e}}_1 \ , \tag{10.33}$$

in this expression $\hat{\mathbf{u}}$ is the unit vector in the direction $\mathbf{u}$.

Problem g: Equation (10.33) implies that $\hat{\mathbf{n}}$ is directed in the direction of the vector $\hat{\mathbf{u}} - \hat{\mathbf{e}}_1$, therefore $\hat{\mathbf{n}}$ can be written as $\hat{\mathbf{n}} = C\,(\hat{\mathbf{u}} - \hat{\mathbf{e}}_1)$, with C an undetermined constant. Show that (10.32) implies that $C = 1/\sqrt{2\,(1-(\hat{\mathbf{u}}\cdot\hat{\mathbf{e}}_1))}$. Also show that this value of C indeed leads to a vector $\hat{\mathbf{n}}$ that is of unit length.

This value of C implies that the unit vector $\hat{\mathbf{n}}$ to be used in the Householder transformation (10.30) is given by

$$\hat{\mathbf{n}} = \frac{\hat{\mathbf{u}} - \hat{\mathbf{e}}_1}{\sqrt{2\,(1-(\hat{\mathbf{u}}\cdot\hat{\mathbf{e}}_1))}} \ . \tag{10.34}$$

To see how the Householder transformation can be used to render the matrix elements below the diagonal equal to zero apply the transformation $\mathbf{Q}$ to the linear equation $\mathbf{A}\mathbf{x} = \mathbf{y}$.

Problem h: Show that this leads to a new system of equations given by

$$\begin{pmatrix} |\mathbf{u}| & a'_{12} & \cdots & a'_{1N} \\ 0 & a'_{22} & \cdots & a'_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a'_{N2} & \cdots & a'_{NN} \end{pmatrix} \mathbf{x} = \mathbf{Q}\mathbf{y} \ . \tag{10.35}$$

A second Householder transformation can now be applied to render all elements in the second column below the diagonal element a'_{22} equal to zero. In this way, all the columns of $\mathbf{A}$ can successively be swiped. Note that in order to apply the Householder transformation one only needs to compute the expressions (10.34) and (10.30) one needs to carry out a matrix multiplication. These operations can be carried out efficiently on computers.

10.4 The Coriolis force and Centrifugal force

As an example of working with the cross-product of vectors we consider the inertia forces that occur in the mechanics of rotating coordinate systems. This is of great importance in the earth sciences, because the rotation of the earth plays a crucial role in the motion of wind and currents in the atmosphere and in the ocean. In addition, the earth's rotation is essential for the generation of the magnetic field of the earth in the outer core.

In order to describe the motion of a particle in a rotating coordinate system we need to characterize the rotation somehow. This can be achieved by introducing a vector $\mathbf{\Omega}$ that is aligned with the rotation axis and whose length is given by *rate* of rotation expressed in radians per seconds.

Problem a: Compute the direction of $\mathbf{\Omega}$ and the length $\Omega = |\mathbf{\Omega}|$ for the earth's rotation.

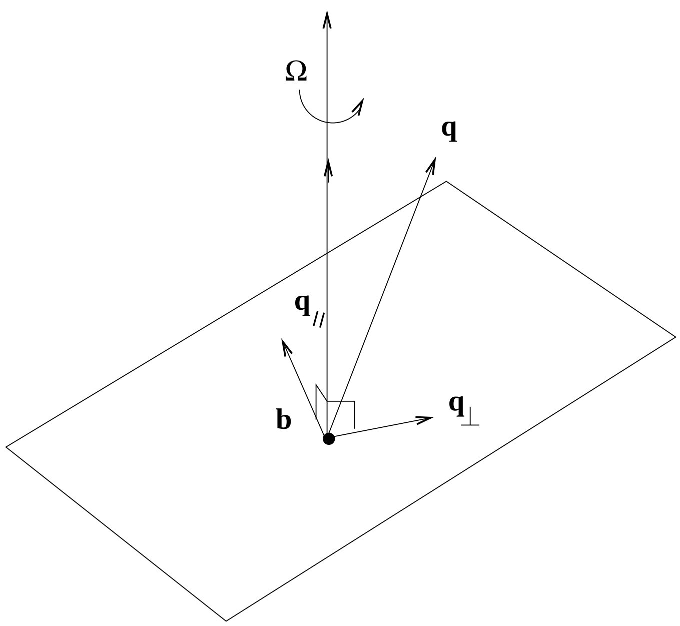


Figure 10.5: Decomposition of a vector in a rotating coordinate system.

Let us assume we are considering a vector $\mathbf{q}$ that is constant in the rotating coordinate system. In a non-rotating system this vector changes with time because it co-rotates with the rotating system. The vector $\mathbf{q}$ can be decomposed in a component $\mathbf{q}_{//}$ along the rotation vector and a component $\mathbf{q}_{\perp}$ to the rotation vector. In addition, a vector $\mathbf{b}$ is defined in figure (10.5) that is perpendicular to both $\mathbf{q}_{\perp}$ and $\mathbf{\Omega}$ in such a way that $\mathbf{\Omega}$, $\mathbf{q}_{\perp}$ and $\mathbf{b}$ form a right handed orthogonal system.

Problem b: Show that:

$$\begin{aligned} \mathbf{q}_{//} &= \hat{\mathbf{\Omega}} \left(\hat{\mathbf{\Omega}} \cdot \mathbf{q} \right) \ , \\ \mathbf{q}_{\perp} &= \mathbf{q} - \hat{\mathbf{\Omega}} \left(\hat{\mathbf{\Omega}} \cdot \mathbf{q} \right) \\ \mathbf{b} &= \hat{\mathbf{\Omega}} \times \mathbf{q} \end{aligned} \tag{10.36}$$

Problem c: In a fixed non-rotating coordinate system, the vector $\mathbf{q}$ rotates, hence it position is time dependent: $\mathbf{q} = \mathbf{q}(t)$. Let us consider how the vector changes over a time interval Δt. Since the component $\mathbf{q}_{//}$ is at all times directed along the rotation vector $\mathbf{\Omega}$, it is constant in time. Over a time interval Δt the coordinate system rotates over an angle $\Omega \Delta t$. Use this to show that the component of $\mathbf{q}$ perpendicular to the rotation vector satisfies:

$$\mathbf{q}_{\perp}(t + \Delta t) = \cos\left(\Omega \Delta t\right) \ \mathbf{q}_{\perp}(t) + \sin\left(\Omega \Delta t\right) \ \mathbf{b} \ , \tag{10.37}$$

and that time evolution of $\mathbf{q}$ is therefore given by

$$\mathbf{q}(t + \Delta t) = \mathbf{q}(t) + \left(\cos\left(\Omega \Delta t\right) - 1\right) \ \mathbf{q}_{\perp}(t) + \sin\left(\Omega \Delta t\right) \ \mathbf{b} \tag{10.38}$$

Problem d: The goal is to obtain the time-derivative of the vector $\mathbf{q}$. This quantity can be computed using the rule $d\mathbf{q}/dt = \lim_{\Delta t \to 0} (\mathbf{q}(t + \Delta t) - \mathbf{q}(t))/\Delta t$. Use this, and equation (10.38) to show that

$$\dot{\mathbf{q}} = \Omega \mathbf{b} \ , \tag{10.39}$$

where the dot denotes the time-derivative. Use (10.36) to show that the time derivative of the vector $\mathbf{q}$ is given by

$$\dot{\mathbf{q}} = \mathbf{\Omega} \times \mathbf{q} \ . \tag{10.40}$$

At this point the vector $\mathbf{q}$ can be any vector that co-rotates with the rotating coordinate system. In this rotating coordinate system, three Cartesian basis vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ can be used as a basis to decompose the position vector:

$$\mathbf{r}_{rot} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}} \ . \tag{10.41}$$

Since these basis vectors are constant in the rotating coordinate system, they satisfy (10.40) so that:

$$\begin{aligned} d\hat{\mathbf{x}}/dt &= \mathbf{\Omega} \times \hat{\mathbf{x}} \ , \\ d\hat{\mathbf{y}}/dt &= \mathbf{\Omega} \times \hat{\mathbf{y}} \ , \\ d\hat{\mathbf{z}}/dt &= \mathbf{\Omega} \times \hat{\mathbf{z}} \ . \end{aligned} \tag{10.42}$$

It should be noted that we have *not* assumed that the position vector $\mathbf{r}_{rot}$ in (10.41) rotates with the coordinate system, we only assumed that the unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ rotate with the coordinate system. Of course, this will leave an imprint on the velocity and the acceleration. In general the velocity and the acceleration follow by differentiating (10.41) with time. If the unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ would be fixed, they would not contribute to the time derivative. However, the unit vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ rotate with the coordinate system and the associated time derivative is given by (10.42).

Problem e: Differentiate the position vector in (10.41) with respect to time and show that the velocity vector $\mathbf{v}$ is given by:

$$\mathbf{v} = \dot{x}\hat{\mathbf{x}} + \dot{y}\hat{\mathbf{y}} + \dot{z}\hat{\mathbf{z}} + \mathbf{\Omega} \times \mathbf{r} \ . \tag{10.43}$$

The terms $\dot{x}\hat{\mathbf{x}} + \dot{y}\hat{\mathbf{y}} + \dot{z}\hat{\mathbf{z}}$ is the velocity as seen in the rotating coordinate system, this velocity is denoted by $\mathbf{v}_{rot}$. The velocity vector can therefore be written as:

$$\mathbf{v} = \mathbf{v}_{rot} + \mathbf{\Omega} \times \mathbf{r} \ . \tag{10.44}$$

Problem f: Give an interpretation of the last term in this expression.

Problem g: The acceleration follows by differentiation expression (10.43) for the velocity once more with respect to time. Show that the acceleration is given by

$$\mathbf{a} = \ddot{x}\hat{\mathbf{x}} + \ddot{y}\hat{\mathbf{y}} + \ddot{z}\hat{\mathbf{z}} + 2\mathbf{\Omega} \times \left(\dot{x}\hat{\mathbf{x}} + \dot{y}\hat{\mathbf{y}} + \dot{z}\hat{\mathbf{z}}\right) + \mathbf{\Omega} \times \left(\mathbf{\Omega} \times \mathbf{r}\right) \ . \tag{10.45}$$

The terms $\ddot{x}\hat{\mathbf{x}} + \ddot{y}\hat{\mathbf{y}} + \ddot{z}\hat{\mathbf{z}}$ in the right hand side denote the acceleration as seen in the rotating coordinate system, this quantity will be denoted by $\mathbf{a}_{rot}$. The terms $\dot{x}\hat{\mathbf{x}} + \dot{y}\hat{\mathbf{y}} + \dot{z}\hat{\mathbf{z}}$ again denote the velocity $\mathbf{v}_{rot}$ as seen in the rotating coordinate system. The left hand side is by Newton's law equal to $\mathbf{F}/m$, where $\mathbf{F}$ is the force acting on the particle.

Problem h: Use this to show that in the rotating coordinate system Newton's law is given by:

$$m\mathbf{a}_{rot} = \mathbf{F} - 2m\mathbf{\Omega} \times \mathbf{v}_{rot} - m\mathbf{\Omega} \times \left(\mathbf{\Omega} \times \mathbf{r}\right) \ . \tag{10.46}$$

The rotation manifests itself through two additional forces. The term $-2m\mathbf{\Omega} \times \mathbf{v}_{rot}$ describes the *Coriolis force* and the term $-m\mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{r})$ describes the centrifugal force.

Problem i: Show that the centrifugal force is perpendicular to the rotation axis and is directed from the rotation axis to the particle.

Problem j: Air flows from high pressure areas to low pressure areas. As air flows in the northern hemisphere from a high pressure area to a low-pressure area, is it deflected towards the right or towards the left when seen from above?

Problem k: Compute the magnitude of the centrifugal force and the Coriolis force you experience due to the earth's rotation when you ride your bicycle. Compare this with the force $m\mathbf{g}$ you experience due to the gravitational attraction of the earth. It suffices to compute orders of magnitude of the different terms. Does the Coriolis force deflect you on the northern hemisphere to the left or to the right? Did you ever notice a tilt while riding your bicycle due to the Coriolis force?

In meteorology and oceanography it is often convenient to describe the motion of air or water along the earth's surface using a Cartesian coordinate system that rotates with the earth with unit vectors pointing in the eastwards ($\hat{\mathbf{e}}_1$), northwards ($\hat{\mathbf{e}}_2$) and upwards ($\hat{\mathbf{e}}_3$), see figure (10.6). The unit vectors can be related to the unit vectors $\hat{\mathbf{r}}$, $\hat{\varphi}$ and $\hat{\theta}$ that are defined in equation (3.7) of section (3.1). Let the velocity in the eastward direction be denoted by u, the velocity in the northward direction by v and the vertical velocity by w.

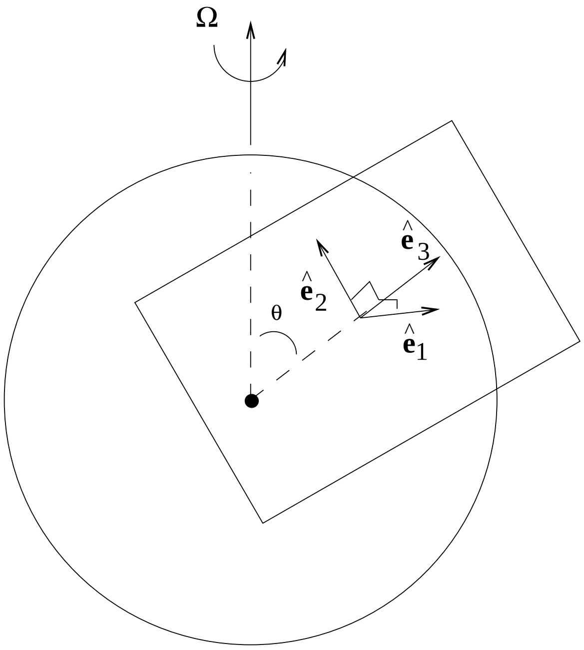


Figure 10.6: Definition of a local Cartesian coordinate system that is aligned with the earth's surface.

Problem l: Show that:

$$\hat{\mathbf{e}}_1 = \hat{\varphi} \quad , \quad \hat{\mathbf{e}}_2 = -\hat{\theta} \quad , \quad \hat{\mathbf{e}}_3 = \hat{\mathbf{r}} \ , \tag{10.47}$$

and that the velocity in this rotating coordinate system is given by

$$\mathbf{v} = u\hat{\mathbf{e}}_1 + v\hat{\mathbf{e}}_2 + w\hat{\mathbf{e}}_3 \ . \tag{10.48}$$

Problem m: We will assume that the axes of the spherical coordinate system are chosen in such a way that the direction $\theta = 0$ is aligned with the rotation axis. This is a different way of saying the rotation vector is parallel to the z-axis: $\mathbf{\Omega} = \Omega\hat{\mathbf{z}}$. Use the first two expressions of equation (3.13) of section (3.1) to show that the rotation vector has the following expansion in the unit vectors $\hat{\mathbf{r}}$ and $\hat{\theta}$:

$$\mathbf{\Omega} = \Omega\left(\cos\theta\ \hat{\mathbf{r}} - \sin\theta\ \hat{\theta}\right) \ . \tag{10.49}$$

Problem n: In the rotating coordinate system, the Coriolis force is given by $\mathbf{F}_{cor} = -2m\mathbf{\Omega}\times\mathbf{v}$. Use the expressions (10.47)-(10.49) and the relations (3.11) of section (3.1) for the cross product of the unit vectors to show that the Coriolis force is given by

$$\mathbf{F}_{cor} = 2m\Omega\sin\theta\ u\ \hat{\mathbf{r}} + 2m\Omega\cos\theta\ u\ \hat{\theta} + 2m\Omega\left(v\cos\theta\ - w\sin\theta\right)\hat{\varphi} \ . \tag{10.50}$$

Problem o: Both the ocean or atmosphere are shallow in the sense that the vertical length scale (a few kilometers for the ocean and around 10 kilometers for the atmosphere) is much less than the horizontal length scale. This causes the vertical velocity to be much smaller than the horizontal velocity. For this reason the vertical velocity w will be neglected in expression (10.50). Use this approximation and the definition (10.47) to show that the horizontal component $\mathbf{a}_{cor}^H$ of the Coriolis acceleration is in this approach given by:

$$\mathbf{a}_{cor}^H = -f\hat{\mathbf{e}}_3 \times \mathbf{v} , \tag{10.51}$$

with

$$f = 2\Omega\cos\theta \tag{10.52}$$

This result is widely used in meteorology and oceanography, because equation (10.51) states that in the Cartesian coordinate system aligned with the earth's surface, the Coriolis force generated by the rotation around the true earth's axis of rotation is identical to the Coriolis force generated by the rotation around a vertical axis with a rotation rate given by $\Omega\cos\theta$. This rotation rate is largest at the poles where $\cos\theta = \pm 1$, and this rotation rate vanishes at the equator where $\cos\theta = 0$. The parameter f in equation (10.51) acts as a coupling parameter, it is called the *Coriolis parameter.* (In the literature on geophysical fluid dynamics one often uses latitude rather than the co-latitude θ that is used here, for this reason one often sees a sin-term rather than a cos-term in the definition of the Coriolis parameter.) In many applications one disregards the dependence of f on the co-latitude θ; in that approach f is a constant and one speaks of the *f-plane approximation.* However, the dependence of the Coriolis parameter on θ is crucial in explaining a number of atmospheric and oceanographic phenomena such as the propagation of Rossby waves and the formation of the Gulfstream. In a further refinement one linearizes the dependence of the Coriolis parameter with co-latitude. This leads to the *β-plane approximation.* Details can be found in the books of *Holton* [30] and *Pedlosky* [46].

10.5 The eigenvalue decomposition of a square matrix

In this section we consider the way in which a square $N \times N$ matrix $\mathbf{A}$ operates on a vector. Since a matrix describes a linear transformation from a vector to a new vector, the action of the matrix $\mathbf{A}$ can be quite complex. However, suppose the matrix has a set of eigenvectors $\hat{\mathbf{v}}^{(n)}$. We assume these eigenvectors are normalized, hence a caret is used in the notation $\hat{\mathbf{v}}^{(n)}$. These eigenvectors are extremely useful because the action of $\mathbf{A}$ on an eigenvector $\hat{\mathbf{v}}^{(n)}$ is very simple:

$$\mathbf{A}\hat{\mathbf{v}}^{(n)} = \lambda_n \hat{\mathbf{v}}^{(n)} , \tag{10.53}$$

where λ_n is the eigenvalue of the eigenvector $\hat{\mathbf{v}}^{(n)}$. When $\mathbf{A}$ acts on an eigenvector, the resulting vector is parallel to the original vector, the only effect of $\mathbf{A}$ on this vector is to either elongate the vector (when $\lambda_n \geq 1$), compress the vector (when $0 \leq \lambda_n < 1$) or reverse the vector (when $\lambda_n < 0$). We will restrict ourselves to matrices that are real and symmetric.

Problem a: Show that for such a matrix the eigenvalues are real and the eigenvectors are orthogonal.

The fact that the eigenvectors $\hat{\mathbf{v}}^{(n)}$ are normalized and mutually orthogonal can be expressed as

$$\left(\hat{\mathbf{v}}^{(n)} \cdot \hat{\mathbf{v}}^{(m)}\right) = \delta_{nm} , \tag{10.54}$$

where δ_{nm} is the Kronecker delta, this quantity is equal to 1 when $n = m$ and is equal to zero when $n \neq m$. The eigenvectors $\hat{\mathbf{v}}^{(n)}$ can be used to define the columns of a matrix $\mathbf{V}$:

$$\mathbf{V} = \begin{pmatrix} \vdots & \vdots & & \vdots \\ \hat{\mathbf{v}}^{(1)} & \hat{\mathbf{v}}^{(2)} & \cdots & \hat{\mathbf{v}}^{(N)} \\ \vdots & \vdots & & \vdots \end{pmatrix} , \tag{10.55}$$

this definition implies that

$$V_{ij} \equiv v_i^{(j)} . \tag{10.56}$$

Problem b: Use the orthogonality of the eigenvectors $\hat{\mathbf{v}}^{(n)}$ (expression (10.54)) to show that the matrix $\mathbf{V}$ is unitary, i.e. to show that

$$\mathbf{V}^T\mathbf{V} = \mathbf{I} , \tag{10.57}$$

where $\mathbf{I}$ is the identity matrix with elements $I_{kl} = \delta_{kl}$. The superscript T denotes the transpose.

Since there are N eigenvectors that the orthonormal in an N-dimensional space, these eigenvectors form a complete set and analogously to (10.13) the completeness relation can be expressed as

$$\mathbf{I} = \sum_{n=1}^{N} \hat{\mathbf{v}}^{(n)}\hat{\mathbf{v}}^{(n)T} . \tag{10.58}$$

When the terms in this expression operate on a arbitrary vector $\mathbf{p}$, an expansion of $\mathbf{p}$ in the eigenvectors is obtained that is completely analogous to equation (10.11):

$$\mathbf{p} = \sum_{n=1}^{N} \hat{\mathbf{v}}^{(n)}\hat{\mathbf{v}}^{(n)T}\mathbf{p} = \sum_{n=1}^{N} \hat{\mathbf{v}}^{(n)} \left(\hat{\mathbf{v}}^{(n)} \cdot \mathbf{p}\right) . \tag{10.59}$$

This is a useful expression, because is can be used to simplify the effect of the matrix $\mathbf{A}$ on an arbitrary vector $\mathbf{p}$.

Problem c: Let $\mathbf{A}$ act on expression (10.59) and show that:

$$\mathbf{A}\mathbf{p} = \sum_{n=1}^{N} \lambda_n \hat{\mathbf{v}}^{(n)} \left(\hat{\mathbf{v}}^{(n)} \cdot \mathbf{p}\right) . \tag{10.60}$$

This expression has an interesting geometric interpretation. When $\mathbf{A}$ acts on $\mathbf{p}$, the vector $\mathbf{p}$ is projected on each of the eigenvectors, this is described by the term $\left(\hat{\mathbf{v}}^{(n)} \cdot \mathbf{p}\right)$. The corresponding eigenvector $\hat{\mathbf{v}}^{(n)}$ is multiplied with the eigenvalue $\hat{\mathbf{v}}^{(n)} \to \lambda_n \hat{\mathbf{v}}^{(n)}$ and the result is summed over all the eigenvectors. The action of $\mathbf{A}$ can thus be reduced to a projection on eigenvectors, a multiplication with the corresponding eigenvalue and a summation over all eigenvectors. The eigenvalue λ_n can be seen as the sensitivity of the eigenvector $\hat{\mathbf{v}}^{(n)}$ to the matrix $\mathbf{A}$.

Problem d: Expression (10.60) holds for every vector $\mathbf{p}$. Use this to show that $\mathbf{A}$ can be written as:

$$\mathbf{A} = \sum_{n=1}^{N} \lambda_n \hat{\mathbf{v}}^{(n)} \hat{\mathbf{v}}^{(n)T} . \tag{10.61}$$

Problem e: Show that with the definition (10.55) this result can also be written as:

$$\mathbf{A} = \mathbf{V} \mathbf{\Sigma} \mathbf{V}^T , \tag{10.62}$$

where $\mathbf{\Sigma}$ is a matrix that has the eigenvalues on the diagonal and whose other elements are equal to zero:

$$\mathbf{\Sigma} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \lambda_N \end{pmatrix} . \tag{10.63}$$

Hint: let (10.62) act on a arbitrary vector, use the definition (10.56) and see what happens.

10.6 Computing a function of a matrix

The expansion (10.61) (or equivalently (10.62)) is very useful because it provides a way to compute the inverse of a matrix and to complete complex functions of a matrix such as the exponential of a matrix. Let us first use (10.61) to compute the inverse $\mathbf{A}^{-1}$ of the matrix. In order to do this we must know the effect of $\mathbf{A}^{-1}$ on the eigenvectors $\hat{\mathbf{v}}^{(n)}$.

Problem a: Use the relation $\hat{\mathbf{v}}^{(n)} = \mathbf{I}\hat{\mathbf{v}}^{(\mathbf{n})} = \mathbf{A}^{-1}\mathbf{A}\hat{\mathbf{v}}^{(n)}$ to show that $\hat{\mathbf{v}}^{(n)}$ is also an eigenvector of the inverse $\mathbf{A}^{-1}$ with eigenvalue $1/\lambda_n$:

$$\mathbf{A}^{-1}\hat{\mathbf{v}}^{(n)} = \frac{1}{\lambda_n}\hat{\mathbf{v}}^{(n)} . \tag{10.64}$$

Problem b: Use this result and the eigenvector decomposition (10.59) to show that the effect of $\mathbf{A}^{-1}$ on a vector $\mathbf{p}$ can be written as

$$\mathbf{A}^{-1}\mathbf{p} = \sum_{n=1}^{N} \frac{1}{\lambda_n}\hat{\mathbf{v}}^{(n)} \left(\hat{\mathbf{v}}^{(n)} \cdot \mathbf{p}\right) . \tag{10.65}$$

Also show that this implies that $\mathbf{A}^{-1}$ can also be written as:

$$\mathbf{A}^{-1} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{V}^T , \tag{10.66}$$

with

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & 1/\lambda_N \end{pmatrix} . \tag{10.67}$$

This is an important result, it means that once we have computed the eigenvectors and eigenvalues of a matrix, we can compute the inverse matrix very efficiently. Note that this procedure gives problems when one of the eigenvalues vanishes because for such an eigenvalue $1/\lambda_n$ is not defined. However, this makes sense; when one (or more) of the eigenvalues vanishes the matrix is singular and the inverse does not exist. Also when one of the eigenvalues is nonzero but close to zero, the corresponding term $1/\lambda_n$ is very large, in practice this gives rise to numerical instabilities. In this situation the inverse of the matrix exist, but the result is very sensitive to computational (and other) errors. Such a matrix is called *poorly conditioned.*

In general, a function of a matrix, such as the exponent of a matrix, is not defined. However, suppose we have a function $f(z)$ that operates on a scalar z and that this function can be written as a power series:

$$f(z) = \sum_p a_p z^p \ . \tag{10.68}$$

For example, when $f(z) = \exp(z)$, then $f(z) = \sum_{p=0}^{\infty}(1/p!)z^p$. Replacing the scalar z by the matrix $\mathbf{A}$ the power series expansion can be used to *define* the effect of the function f when it operates on the matrix $\mathbf{A}$:

$$f(\mathbf{A}) \equiv \sum_p a_p \mathbf{A}^p \ . \tag{10.69}$$

Although this may seem to be a simple rule to compute $f(\mathbf{A})$, it is actually not so useful because in many applications the summation (10.69) consists of infinitely many terms and the computation of $\mathbf{A}^p$ can computationally be very demanding. Again, the eigenvalue decomposition (10.61) or (10.62) allows us to simplify the evaluation of $f(\mathbf{A})$.

Problem c: Show that $\hat{\mathbf{v}}^{(n)}$ is also an eigenvector of $\mathbf{A}^p$ with eigenvalue $(\lambda_n)^p$, i.e. show that

$$\mathbf{A}^p \hat{\mathbf{v}}^{(n)} = (\lambda_n)^p \, \hat{\mathbf{v}}^{(n)} \ . \tag{10.70}$$

Hint, first compute $\mathbf{A}^2\hat{\mathbf{v}}^{(n)} = \mathbf{A}\left(\mathbf{A}\hat{\mathbf{v}}^{(n)}\right)$, then $\mathbf{A}^3\hat{\mathbf{v}}^{(n)}$, etc.

Problem d: Use this result to show that (10.62) can be generalized to:

$$\mathbf{A}^p = \mathbf{V}\mathbf{\Sigma}^p\mathbf{V}^T \ , \tag{10.71}$$

with $\mathbf{\Sigma}^p$ given by

$$\mathbf{\Sigma}^p = \begin{pmatrix} \lambda_1^p & 0 & \cdots & 0 \\ 0 & \lambda_2^p & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \lambda_N^p \end{pmatrix} \ . \tag{10.72}$$

Problem e: Finally use (10.69) to show that $f(\mathbf{A})$ can be written as:

$$f(\mathbf{A}) = \mathbf{V} f\left(\mathbf{\Sigma}\right) \mathbf{V}^T \ , \tag{10.73}$$

with $f(\mathbf{\Sigma})$ given by

$$f(\mathbf{\Sigma}) = \begin{pmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & f(\lambda_N) \end{pmatrix} \tag{10.74}$$

Problem f: In order to revert to an explicit eigenvector expansion, show that (10.73) can be written as:

$$f(\mathbf{A}) = \sum_{n=1}^{N} f(\lambda_n)\, \hat{\mathbf{v}}^{(n)} \hat{\mathbf{v}}^{(n)T} \,. \tag{10.75}$$

With this expression (or the equivalent expression (10.73)) the evaluation of $f(\mathbf{A})$ is simple once the eigenvectors and eigenvalues of $\mathbf{A}$ are known, because in (10.75) the function f only acts on the eigenvalues, but not on the matrix. Since the function f normally acts on a scalar (such as the eigenvalues), the eigenvector decomposition has obviated the need for computing higher powers of the matrix $\mathbf{A}$. However, from a numerical point of view computing functions of matrices can be a tricky issue. For example, *Moler and van Loan*[40] give nineteen dubious ways to compute the exponential of a matrix.

10.7 The normal modes of a vibrating system

An eigenvector decomposition is not only useful for computing the inverse of a matrix or other functions of a matrix, it also provides a way for analyzing characteristics of dynamical systems. As an example, a simple model for the oscillations of a vibrating molecule is shown here. This system is the prototype of a vibrating system that has different modes of vibration. The natural modes of vibration are usually called the *normal modes* of that system. Consider the mechanical system shown in figure (10.7). Three particles with mass m are coupled by two springs with spring constants k. It is assumed that the three masses are constrained to move along a line. The displacement of the masses from their equilibrium positions are denoted with x_1, x_2 and x_3 respectively. This mechanical model can considered to be a grossly oversimplified model of a tri-atomic molecule such as CO_2 or H_2O.

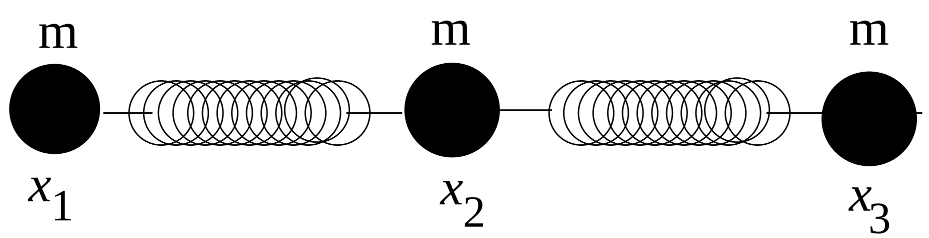


Figure 10.7: Definition of variables for a simple vibrating system.

Each of the masses can experience an external force F_i, where the subscript i denotes the mass under consideration. The equations of motion for the three masses is given by:

$$\begin{aligned} m\ddot{x}_1 &= k(x_2 - x_1) + F_1 \ , \\ m\ddot{x}_2 &= -k(x_2 - x_1) + k(x_3 - x_2) + F_2 \ , \\ m\ddot{x}_3 &= -k(x_3 - x_2) + F_3 \ , \end{aligned} \tag{10.76}$$

For the moment we will consider harmonic oscillations, i.e. we assume that the both the driving forces F_i and the displacements x_i vary with time as $\exp -i\omega t$. The displacements x_1, x_2 and x_3 can be used to form a vector $\mathbf{x}$, and summarily a vector $\mathbf{F}$ can be formed from the three forces F_1, F_2 and F_3 that act on the three masses.

Problem a: Show that for an harmonic motion with frequency ω the equations of motion can be written in vector form as:

$$\left(\mathbf{A} - \frac{m\omega^2}{k}\mathbf{I}\right)\mathbf{x} = \frac{1}{k}\mathbf{F} \ , \tag{10.77}$$

with the matrix $\mathbf{A}$ given by

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix} . \tag{10.78}$$

The normal modes of the system are given by the patterns of oscillations of the system when there is no driving force. For this reason, we set the driving force $\mathbf{F}$ in the right hand side of (10.77) momentarily to zero. Equation (10.77) then reduces to a homogeneous system of linear equations, such a system of equations can only have nonzero solutions when the determinant of the matrix vanishes. Since the matrix $\mathbf{A}$ has only three eigenvalues, the system can only oscillate freely at three discrete eigenfrequencies. The system can only oscillate at other frequencies when it is driven by the force $\mathbf{F}$ at such a frequency.

Problem b: Show that the eigenfrequencies ω_i of the vibrating system are given by

$$\omega_i = \sqrt{\frac{k\lambda_i}{m}} \ , \tag{10.79}$$

where λ_i are the eigenvalues of the matrix $\mathbf{A}$.

Problem c: Show that the eigenfrequencies of the system are given by:

$$\omega_1 = 0 \quad , \quad \omega_2 = \sqrt{\frac{k}{m}} \quad , \quad \omega_3 = \sqrt{\frac{3k}{m}} \ . \tag{10.80}$$

Problem d: The frequencies do not give the vibrations of each of the three particles respectively. Instead these frequencies give the eigenfrequencies of the three modes of oscillation of the system. The eigenvector that corresponds to each eigenvalue gives the displacement of each particle for that mode of oscillation. Show that these eigenvectors are given by:

$$\hat{\mathbf{v}}^{(1)} = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix} \quad , \quad \hat{\mathbf{v}}^{(2)} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix} \quad , \quad \hat{\mathbf{v}}^{(3)} = \frac{1}{\sqrt{6}}\begin{pmatrix}1\\-2\\1\end{pmatrix} \tag{10.81}$$

Remember that the eigenvectors can be multiplied with an arbitrary constant, this constant is chosen in such a way that each eigenvector has length 1.

Problem e: Show that these eigenvectors satisfy the requirement (10.54).

Problem f: Sketch the motion of the three masses of each normal mode. Explain physically why the third mode with frequency ω_3 has a higher eigenfrequency than the second mode ω_2.

Problem g: Explain physically why the second mode has an eigenfrequency $\omega_2 = \sqrt{k/m}$ that is identical to the frequency of a single mass m that is suspended by a spring with spring constant k.

Problem h: What type of motion does the first mode with eigenfrequency ω_1 describe? Explain physically why this frequency is independent of the spring constant k and the mass m.

Now we know the normal modes of the system, we consider the case where the system is driven by a force $\mathbf{F}$ that varies in time as $\exp -i\omega t$. For simplicity it is assumed that the frequency ω of the driving force differs from the eigenfrequencies of the system: $\omega \neq \omega_i$. The eigenvectors $\hat{\mathbf{v}}^{(n)}$ defined in (10.81) form a complete orthonormal set, hence both the driving force $\mathbf{F}$ and the displacement $\mathbf{x}$ can be expanded in this set. Using (10.59) the driving force can be expanded as

$$\mathbf{F} = \sum_{n=1}^{3} \hat{\mathbf{v}}^{(n)}(\hat{\mathbf{v}}^{(n)} \cdot \mathbf{F}) \; . \tag{10.82}$$

Problem i: Write the displacement vector as a superposition of the normal mode displacements: $\mathbf{x} = \sum_{n=1}^{3} c_n \hat{\mathbf{v}}^{(n)}$, use the expansion (10.82) for the driving force and insert these equations in the equation of motion (10.77) to solve for the unknown coefficients c_n. Eliminate the eigenvalues with (10.79) and show that the displacement is given by:

$$\mathbf{x} = \frac{1}{m}\sum_{n=1}^{3} \frac{\hat{\mathbf{v}}^{(n)}(\hat{\mathbf{v}}^{(n)} \cdot \mathbf{F})}{(\omega_n^2 - \omega^2)} \; . \tag{10.83}$$

This expression has a nice physical interpretation. Expression (10.83) states that the total response of the system can be written as a superposition of the different normal modes (the $\sum_{n=1}^{3} \hat{\mathbf{v}}^{(n)}$ terms). The effect that the force has on each normal mode is given by the inner product $(\hat{\mathbf{v}}^{(n)} \cdot \mathbf{F})$. This is nothing but the component of the force $\mathbf{F}$ along the eigenvector $\hat{\mathbf{v}}^{(n)}$, see equation (10.2). The term $1/\left(\omega_n^2 - \omega^2\right)$ gives the sensitivity of the system to a driving force with frequency ω, this term can be called a sensitivity term.

When the driving force is close to one of the eigenfrequencies, $1/\left(\omega_n^2-\omega^2\right)$ is very large. In that case the system is close to resonance and the resulting displacement will be very large. On the other hand, when the frequency of the driving force is very far from the eigenfrequencies of the system, $1/\left(\omega_n^2-\omega^2\right)$ will be small and the system will give a very small response. The total response can be seen as a combination of three basic operations: eigenvector expansion, projection and multiplication with a response function. Note that the same operations were used in the explanation of the action of a matrix $\mathbf{A}$ below equation (10.60).

10.8 Singular value decomposition

In section (10.5) the decomposition of a square matrix in terms of eigenvectors was treated. In many practical applications, such as inverse problems, one encounters a system of equations that is not square:

$$\underbrace{\mathbf{A}}_{\substack{M\times N\\ matrix}} \; \underbrace{\mathbf{x}}_{\substack{N\\ rows}} = \underbrace{\mathbf{y}}_{\substack{M\\ rows}} \tag{10.84}$$

Consider the example that the vector $\mathbf{x}$ has N components and that there are M equations. In that case the vector $\mathbf{y}$ has M components and the matrix $\mathbf{A}$ has M rows and N columns, i.e. it is an $M\times N$ matrix. A relation such as (10.53) which states that $\mathbf{A}\hat{\mathbf{v}}^{(n)}=\lambda_n\hat{\mathbf{v}}^{(n)}$ cannot possibly hold because when the matrix $\mathbf{A}$ acts on an N-vector it produces an M-vector whereas in (10.53) the vector in the right hand side has the same number of components as the vector in the left hand side. It will be clear that the theory of section (10.5) cannot be applied when the matrix is not square. However, it is possible to generalize the theory of section (10.5) when $\mathbf{A}$ is not square. For simplicity it is assumed that $\mathbf{A}$ is a real matrix.

In section (10.5) a single set of orthonormal eigenvectors $\hat{\mathbf{v}}^{(n)}$ was used to analyze the problem. Since the vectors $\mathbf{x}$ and $\mathbf{y}$ in (10.84) have different dimensions, it is necessary to expand the vector $\mathbf{x}$ in a set of N orthogonal vectors $\hat{\mathbf{v}}^{(n)}$ that each have N components and to expand $\mathbf{y}$ in a different set of M orthogonal vectors $\hat{\mathbf{u}}^{(m)}$ that each have M components. Suppose we have chosen a set $\hat{\mathbf{v}}^{(n)}$, let us define vectors $\hat{\mathbf{u}}^{(n)}$ by the following relation:

$$\mathbf{A}\hat{\mathbf{v}}^{(n)}=\lambda_n\hat{\mathbf{u}}^{(n)} \; . \tag{10.85}$$

The constant λ_n should not be confused with an eigenvalue, this constant follows from the requirement that $\hat{\mathbf{v}}^{(n)}$ and $\hat{\mathbf{u}}^{(n)}$ are both vectors of unit length. At this point, the choice of $\hat{\mathbf{v}}^{(n)}$ is still open. The vectors $\hat{\mathbf{v}}^{(n)}$ will now be constrained that they satisfy in addition to (10.85) the following requirement:

$$\mathbf{A}^T\hat{\mathbf{u}}^{(n)}=\mu_n\hat{\mathbf{v}}^{(n)} \; , \tag{10.86}$$

where $\mathbf{A}^T$ is the transpose of $\mathbf{A}$.

Problem a: In order to find the vectors $\hat{\mathbf{v}}^{(n)}$ and $\hat{\mathbf{u}}^{(n)}$ that satisfy both (10.85) and (10.86), multiply (10.85) with $\mathbf{A}^T$ and use (10.86) to eliminate $\hat{\mathbf{u}}^{(n)}$. Do this to show that $\hat{\mathbf{v}}^{(n)}$ satisfies:

$$\left(\mathbf{A}^T\mathbf{A}\right)\hat{\mathbf{v}}^{(n)} = \lambda_n\mu_n\hat{\mathbf{v}}^{(n)} \ . \tag{10.87}$$

Use similar steps to show that $\hat{\mathbf{u}}^{(n)}$ satisfies

$$\left(\mathbf{A}\mathbf{A}^T\right)\hat{\mathbf{u}}^{(n)} = \lambda_n\mu_n\hat{\mathbf{u}}^{(n)} \ . \tag{10.88}$$

These equations state that the $\hat{\mathbf{v}}^{(n)}$ are the eigenvectors of $\mathbf{A}^T\mathbf{A}$ and that the $\hat{\mathbf{u}}^{(n)}$ are the eigenvectors of $\mathbf{A}\mathbf{A}^T$.

Problem b: Show that both $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ are real symmetric matrices and show that this implies that the basis vectors $\hat{\mathbf{v}}^{(n)}$ $(n = 1, \cdots, N)$ and $\hat{\mathbf{u}}^{(m)}$ $(m = 1, \cdots, M)$ are both orthonormal:

$$\left(\hat{\mathbf{v}}^{(n)} \cdot \hat{\mathbf{v}}^{(m)}\right) = \left(\hat{\mathbf{u}}^{(n)} \cdot \hat{\mathbf{u}}^{(m)}\right) = \delta_{nm} \ . \tag{10.89}$$

Although (10.87) and (10.88) can be used to find the basis vectors $\hat{\mathbf{v}}^{(n)}$ and $\hat{\mathbf{u}}^{(n)}$, these expressions cannot be used to find the constants λ_n and μ_n, because these expressions state that the *product* $\lambda_n\mu_n$ is equal to the eigenvalues of $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$. This implies that only the product of λ_n and μ_n is defined.

Problem c: In order to find the relation between λ_n and μ_n, take the inner product of (10.85) with $\hat{\mathbf{u}}^{(n)}$ and use the orthogonality relation (10.89) to show that:

$$\lambda_n = \left(\hat{\mathbf{u}}^{(n)} \cdot \mathbf{A}\hat{\mathbf{v}}^{(n)}\right) \ . \tag{10.90}$$

Problem d: Show that for arbitrary vectors $\mathbf{p}$ and $\mathbf{q}$ that

$$(\mathbf{p} \cdot \mathbf{A}\mathbf{q}) = \left(\mathbf{A}^T\mathbf{p} \cdot \mathbf{q}\right) \ . \tag{10.91}$$

Problem e: Apply this relation to (10.90) and use (10.86) to show that

$$\lambda_n = \mu_n \ . \tag{10.92}$$

This is all the information we need to find both λ_n and μ_n. Since these quantities are equal, and since by virtue of (10.87) these eigenvectors are equal to the eigenvectors of $\mathbf{A}^T\mathbf{A}$, it follows that both λ_n and μ_n are given by the square-root of the eigenvalues of $\mathbf{A}^T\mathbf{A}$. Note that is follows from (10.88) that the product $\lambda_n\mu_n$ also equals the eigenvalues of $\mathbf{A}\mathbf{A}^T$. This can only be the case when $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ have the same eigenvalues. Before we proceed let us show that this is indeed the case. Let the eigenvalues of $\mathbf{A}^T\mathbf{A}$ be denoted by Λ_n and the eigenvalues of $\mathbf{A}\mathbf{A}^T$ by Υ_n, i.e. that

$$\mathbf{A}^T\mathbf{A}\hat{\mathbf{v}}^{(n)} = \Lambda_n\hat{\mathbf{v}}^{(n)} \ , \tag{10.93}$$

and

$$\mathbf{A}\mathbf{A}^T\hat{\mathbf{u}}^{(n)} = \Upsilon_n\hat{\mathbf{u}}^{(n)} \ . \tag{10.94}$$

Problem f: Take the inner product of (10.93) with $\hat{\mathbf{v}}^{(n)}$ to show that $\Lambda_n = \left(\hat{\mathbf{v}}^{(n)} \cdot \mathbf{A}^T\mathbf{A}\hat{\mathbf{v}}^{(n)}\right)$, use the properties (10.91) and $\mathbf{A}^{TT} = \mathbf{A}$ and (10.85) to show that $\lambda_n^2 = \Lambda_n$. Use similar steps to show that $\mu_n^2 = \Upsilon_n$. With (10.92) this implies that $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ have the same eigenvalues.

The proof that $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ have the same eigenvalues was not only given as a check of the consistency of the theory, the fact that $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ have the same eigenvalues has important implications. Since $\mathbf{A}\mathbf{A}^T$ is an $M \times M$ matrix, it has M eigenvalues, and since $\mathbf{A}^T\mathbf{A}$ is an $N \times N$ matrix it has N eigenvalues. The only way for these matrices to have the same eigenvalues, but to have a different number of eigenvalues is that the number of nonzero eigenvalues is given by the minimum of N and M. In practice, some of the eigenvalues of $\mathbf{A}\mathbf{A}^T$ may be zero, hence the number of nonzero eigenvalues of $\mathbf{A}\mathbf{A}^T$ can be less that M. By the same token, the number of nonzero eigenvalues of $\mathbf{A}^T\mathbf{A}$ can be less than N. The number of nonzero eigenvalues will be denoted by P. It is not know a-priori how many nonzero eigenvalues there are, but it follows from the arguments above that P is smaller or equal than M and N. This implies that

$$P \leq \min(N, M) \; , \tag{10.95}$$

where $\min(N, M)$ denotes the minimum of N and M. Therefore, whenever a summation over eigenvalues occurs, we need to take only P eigenvalues into account. Since the ordering of the eigenvalues is arbitrary, it is assumed in the following that the eigenvectors are ordered by decreasing size: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. In this ordering the eigenvalues for $n > P$ are equal to zero so that the summation over eigenvalues runs from 1 to P.

Problem g: The matrices $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ have the same eigenvalues. When you need the eigenvalues and eigenvectors, would it be from the point of view of computational efficiency be more efficient to compute the eigenvalues and eigenvectors of $\mathbf{A}^T\mathbf{A}$ or of $\mathbf{A}\mathbf{A}^T$? Consider the situations $M > N$ and $M < N$ separately.

Let us now return to the task of making an eigenvalue decomposition of the matrix $\mathbf{A}$. The vectors $\hat{\mathbf{v}}^{(n)}$ form a basis in N-dimensional space. Since the vector $\mathbf{x}$ is N-dimensional, every vector $\mathbf{x}$ can de decomposed according to equation (10.59): $\mathbf{x} = \sum_{n=1}^{N} \hat{\mathbf{v}}^{(n)} \left(\hat{\mathbf{v}}^{(n)} \cdot \mathbf{x}\right)$.

Problem h: Let the matrix $\mathbf{A}$ act on this expression and use (10.85) to show that:

$$\mathbf{A}\mathbf{x} = \sum_{n=1}^{P} \lambda_n \hat{\mathbf{u}}^{(n)} \left(\hat{\mathbf{v}}^{(n)} \cdot \mathbf{x}\right) \; . \tag{10.96}$$

Problem i: This expression must hold for any vector $\mathbf{x}$. Use this property to deduce that:

$$\mathbf{A} = \sum_{n=1}^{P} \lambda_n \hat{\mathbf{u}}^{(n)} \hat{\mathbf{v}}^{(n)T} \; . \tag{10.97}$$

Problem j: The eigenvectors $\hat{\mathbf{v}}^{(n)}$ can be arranged in an $N \times N$ matrix $\mathbf{V}$ defined in (10.55). Similarly the eigenvectors $\hat{\mathbf{u}}^{(n)}$ can be used to form the columns of an

$M \times M$ matrix $\mathbf{U}$:

$$\mathbf{U} = \begin{pmatrix} \vdots & \vdots & & \vdots \\ \hat{\mathbf{u}}^{(1)} & \hat{\mathbf{u}}^{(2)} & \cdots & \hat{\mathbf{u}}^{(M)} \\ \vdots & \vdots & & \vdots \end{pmatrix} . \tag{10.98}$$

Show that $\mathbf{A}$ can also be written as:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T , \tag{10.99}$$

with the diagonal matrix $\mathbf{\Sigma}$ defined in (10.63).

This decomposition of $\mathbf{A}$ in terms of eigenvectors is called the *Singular Value Decomposition* of the matrix. This is frequently abbreviated as SVD.

Problem k: You may have noticed the similarity between the expression (10.97) and the equation (10.62) for a square matrix and expression (10.99) and equation (10.61). Show that for the special case $M = N$ the theory of this section is identical to the eigenvalue decomposition for a square matrix presented in section (10.5). Hint: what are the vectors $\hat{\mathbf{u}}^{(n)}$ when $M = N$?

Let us now solve the original system of linear equations (10.84) for the unknown vector $\mathbf{x}$. In order to do this, expand the vector $\mathbf{y}$ in the vectors $\hat{\mathbf{u}}^{(n)}$ that span the M-dimensional space: $\mathbf{y} = \sum_{m=1}^{M} \hat{\mathbf{u}}^{(m)} \left(\hat{\mathbf{u}}^{(m)} \cdot \mathbf{y}\right)$, and expand the vector $\mathbf{x}$ in the vectors $\hat{\mathbf{v}}^{(n)}$ that span the N-dimensional space:

$$\mathbf{x} = \sum_{n=1}^{N} c_n \hat{\mathbf{v}}^{(n)} . \tag{10.100}$$

Problem l: At this point the coefficients c_n are unknown. Insert the expansions for $\mathbf{y}$ and $\mathbf{x}$ and the expansion (10.97) for the matrix $\mathbf{A}$ in the linear system (10.84) and use the orthogonality properties of the eigenvectors to show that $c_n = \left(\hat{\mathbf{u}}^{(n)} \cdot \mathbf{y}\right) / \lambda_n$ so that

$$\mathbf{x} = \sum_{n=1}^{P} \frac{1}{\lambda_n} \left(\hat{\mathbf{u}}^{(n)} \cdot \mathbf{y}\right) \hat{\mathbf{v}}^{(n)} . \tag{10.101}$$

Note that although in the original expansion (10.100) of $\mathbf{x}$ a summation is carried out over all N basisvectors, whereas in the solution (10.101) a summation is carried out over the first P basisvectors only. The reason for this is that the remaining eigenvectors have eigenvalues that are equal to zero so that they could be left out of the expansion (10.97) of the matrix $\mathbf{A}$. Indeed, these eigenvalues would give rise to problems because if they were retained they would lead to infinite contributions $1/\lambda \to \infty$ in the solution (10.101). In practice, some eigenvalues may be nonzero, but close to zero so that the term $1/\lambda$ gives rise to numerical instabilities. In practice, one therefore often leaves out nonzero but small eigenvalues as well in the summation (10.101).

This may appear to be a objective procedure for defining solutions for linear problems that are undetermined or for problems that are otherwise ill-conditioned, but there is a price once pays for leaving out basisvectors in the construction of the solution. The vector

$\mathbf{x}$ is N-dimensional, hence one needs N basisvectors to construct an arbitrary vector $\mathbf{x}$, see equation (10.100). The solution vector given in (10.101) is build by superposing only P basisvectors. This implies that the solution vector is constrained to be within the P-dimensional subspace spanned by the first P eigenvectors. Therefore, there it is not clear that the solution vector in (10.101) is identical to the true vector $\mathbf{x}$. However, the point of using the singular value decomposition is that the solution is only constrained by the linear system of equations (10.84) within the subspace spanned by the first P basisvectors $\hat{\mathbf{v}}^{(n)}$. The solution (10.101) ensures that only the components of $\mathbf{x}$ within that subspace are affected by the right-hand side vector $\mathbf{y}$. This technique is extremely important in the analysis of linear inverse problems.

Chapter 11

Fourier analysis

Fourier analysis is concerned with the decomposition of signals in sine and cosine waves. This technique is of obvious relevance for spectral analysis where one decomposes a time signal in its different frequency components. As an example, the spectrum of a low-C on

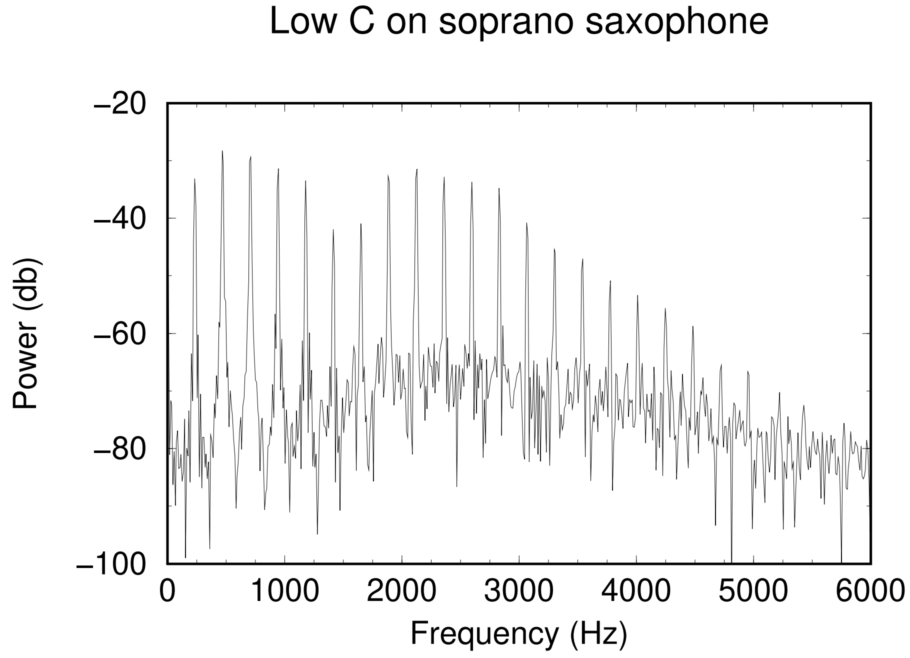


Figure 11.1: The energy of the sound made by the author playing a low C on his soprano saxophone as a function of frequency. The unit used for the horizontal axis is Hertz (number of oscillations per second), the unit on the vertical axis is decibels (a logarithmic measure of energy).

a soprano saxophone shown in figure 11.1. However, the use of Fourier analysis goes far beyond this application because Fourier analysis can also be used for finding solutions of differential equations and a large number of other applications. In this chapter the real Fourier transform on a finite interval is used as a starting point. From this the complex Fourier transform and the Fourier transform on an infinite interval are derived. In several stages of the analysis, the similarity of Fourier analysis and linear algebra will be made apparent.

11.1 The real Fourier series on a finite interval

Consider a function $f(x)$ that is defined on the interval $-L < x \leq L$. This interval is of length $2L$, and let us assume that $f(x)$ is periodic with period $2L$. This means that if one translates this function over a distance $2L$ the value does not change:

$$f(x + 2L) = f(x) \ . \tag{11.1}$$

We want to expand this function in a set of basis functions. Since $f(x)$ is periodic with period $2L$, these basis functions must be periodic with the same period.

Problem a: Show that the functions $\cos(n\pi x/L)$ and $\sin(n\pi x/L)$ with integer n are periodic with period $2L$, i.e. show that these functions satisfy (11.1).

The main statement of Fourier analysis is that one can write $f(x)$ as a superposition of these periodic sine and cosine waves:

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(n\pi x/L) + \sum_{n=1}^{\infty} b_n \sin(n\pi x/L) \ . \tag{11.2}$$

The factor 1/2 in the coefficient a_0 has no special significance, it is used to simplify subsequent expressions. To show that (11.2) is actually true is not trivial. Providing this proof essentially amounts to showing that the functions $\cos(n\pi x/L)$ and $\sin(n\pi x/L)$ actually contain enough "degrees of freedom" to describe $f(x)$. However, since $f(x)$ is a function of a continuous variable x this function has infinitely many degrees of freedom and since there are infinitely many coefficients a_n and b_n counting the number of degrees of freedom does not work. Mathematically one would say that one needs to show that the set of functions $\cos(n\pi x/L)$ and $\sin(n\pi x/L)$ is a "complete set." We will not concern us here with this proof, and simply start working with the Fourier series (11.2).

At this point it is not clear yet what the coefficients a_n and b_n are. In order to derive these coefficients one needs to use the following integrals:

$$\int_{-L}^{L} \cos^2(n\pi x/L)\, dx = \int_{-L}^{L} \sin^2(n\pi x/L)\, dx = L \qquad (n \geq 1) \ , \tag{11.3}$$

$$\int_{-L}^{L} \cos(n\pi x/L) \cos(m\pi x/L)\, dx = 0 \qquad if \qquad n \neq m \ , \tag{11.4}$$

$$\int_{-L}^{L} \sin(n\pi x/L) \sin(m\pi x/L)\, dx = 0 \qquad if \qquad n \neq m \ , \tag{11.5}$$

$$\int_{-L}^{L} \cos(n\pi x/L) \sin(m\pi x/L)\, dx = 0 \qquad all\ n, m \ . \tag{11.6}$$

Problem b: Derive these identities. In doing so you need to use trigonometric identities such as $\cos\alpha \cos\beta = (\cos(\alpha + \beta) + \cos(\alpha - \beta))/2$. If you have difficulties deriving these identities you may want to consult a textbook such as *Boas*[11].

Problem c: In order to find the coefficient b_m, multiply the Fourier expansion (11.2) with $\sin(m\pi x/L)$, integrate the result from $-L$ to L and use the relations (11.3)-(11.6) to evaluate the integrals. Show that this gives:

$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin(n\pi x/L)\,dx\,. \tag{11.7}$$

Problem d: Use a similar analysis to show that:

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos(n\pi x/L)\,dx\,. \tag{11.8}$$

In deriving this result treat the cases $n \neq 0$ and $n = 0$ separately. It is now clear why the factor $1/2$ is introduced in the a_0-term of (11.2); without this factor expression (11.8) would have an additional factor 2 for $n = 0$.

There is a close relation between the Fourier series (11.2) and the coefficients given in the expressions above and the projection of a vector on a number of basis vectors in linear algebra as shown in section (10.1). To see this we will restrict ourselves for simplicity to functions $f(x)$ that are odd functions of x: $f(-x) = -f(x)$, but this restriction is by no means essential. For these functions all coefficients a_n are equal to zero. As an analogue of a basis vector in linear algebra let us define the following basis function $u_n(x)$:

$$u_n(x) \equiv \frac{1}{\sqrt{L}}\sin(n\pi x/L)\ . \tag{11.9}$$

An essential ingredient in the projection operators of section (10.1) is the inner product between vectors. It is also possible to define an inner product for functions, and for the present example the inner product of two functions $f(x)$ and $g(x)$ is defined as:

$$(f\cdot g) \equiv \int_{-L}^{L} f(x)g(x)dx\ . \tag{11.10}$$

Problem e: The basis functions $u_n(x)$ defined in (11.9) are the analogue of a set of orthonormal basis vectors. To see this, use (11.3) and (11.5) to show that

$$(u_n\cdot u_m) = \delta_{nm}\ , \tag{11.11}$$

where δ_{nm} is the Kronecker delta.

This expression implies that the basis functions $u_n(x)$ are mutually orthogonal. If the norm of such a basis function is defined as $\|u_n\| \equiv \sqrt{(u_n\cdot u_n)}$, expression (11.11) implies that the basis functions are normalized (i.e. have norm 1). These functions are the generalization of orthogonal unit vectors to a function space. The (odd) function $f(x)$ can be written as a sum of the basis functions $u_n(x)$:

$$f(x) = \sum_{n=1}^{\infty} c_n u_n(x)\ . \tag{11.12}$$

Problem f: Take the inner product of (11.12) with $u_m(x)$ and show that $c_m = (u_m \cdot f)$. Use this to show that the Fourier expansion of $f(x)$ can be written as: $f(x) = \sum_{n=1}^{\infty} u_n(x)\,(u_n \cdot f)$. and that leaving out the explicit dependence on the variable x the result is given by

$$f = \sum_{n=1}^{\infty} u_n\,(u_n \cdot f) \; . \tag{11.13}$$

This equation bears a close resemblance to the expression derived in section (10.1) for the projection of vectors. The projection of a vector $\mathbf{v}$ along a unit vector $\hat{\mathbf{n}}$ was shown to be

$$\mathbf{P}\mathbf{v} = \hat{\mathbf{n}}\,(\hat{\mathbf{n}} \cdot \mathbf{v}) \qquad (10.2) \quad again \; .$$

A comparison with equation (11.13) shows that $u_n(x)\,(u_n \cdot f)$ can be interpreted as the projection of the function $f(x)$ on the function $u_n(x)$. To reconstruct the function, one must sum over the projections along all basis functions, hence the summation in (11.13). It is shown in equation (10.12) of section (10.1) that in order to find the projection of the vector $\mathbf{v}$ onto the subspace spanned by a finite number of orthonormal basis vectors one simply has to sum the projections of the vector $\mathbf{v}$ on all the basis vectors that span the subspace: $\mathbf{P}\mathbf{v} = \sum_i \hat{\mathbf{n}}_i\,(\hat{\mathbf{n}}_i \cdot \mathbf{v})$. In a similar way, one can sum the Fourier series (11.13) over only a limited number of basis functions to obtain the projection of $f(x)$ on a limited number of basis functions:

$$f_{filtered} = \sum_{n=n_1}^{n_2} u_n\,(u_n \cdot f) \; , \tag{11.14}$$

in this expression it was assumed that only values $n_1 \le n \le n_2$ have been used. The projected function is called $f_{filtered}$ because this projection really is a filtering operation.

Problem g: To see this, show that the functions $u_n(x)$ are sinusoidal waves with wavelength $\lambda = 2L/n$.

This means that restricting the n-values in the sum (11.14) amounts to using only wavelengths between $2L/n_2$ and $2L/n_1$ for the projected function. Since only certain wavelengths are used, this projection really acts as a filter that allows only certain wavelengths in the filtered function.

It is the filtering property that makes the Fourier transform so useful for filtering data sets for excluding wavelengths that are unwanted. In fact, the Fourier transform forms the basis of digital filtering techniques that have many applications in science and engineering, see for example the books of *Claerbout*[15] or *Robinson and Treitel*[51].

11.2 The complex Fourier series on a finite interval

In the theory of the preceding section there is no reason why the function $f(x)$ should be real. Although the basis functions $\cos(n\pi x/L)$ and $\sin(n\pi x/L)$ are real, the Fourier sum (11.2) can be complex because the coefficients a_n and b_n can be complex. The equation of de Moivre gives the relation between these basis functions and complex exponential functions:

$$e^{in\pi x/L} = \cos(n\pi x/L) + i\sin(n\pi x/L) \tag{11.15}$$

This expression can be used to rewrite the Fourier series (11.2) using the basis functions $\exp in\pi x/L$ rather than sine and cosines.

Problem a: Replace n by $-n$ in (11.15) to show that:

$$\begin{aligned} \cos(n\pi x/L) &= \tfrac{1}{2}\left(e^{in\pi x/L} + e^{-in\pi x/L}\right) , \\ \sin(n\pi x/L) &= \tfrac{1}{2i}\left(e^{in\pi x/L} - e^{-in\pi x/L}\right) . \end{aligned} \tag{11.16}$$

Problem b: Insert this relation in the Fourier series (11.2) to show that this Fourier series can also be written as:

$$f(x) = \sum_{n=-\infty}^{\infty} c_n \, e^{in\pi x/L} \;, \tag{11.17}$$

with the coefficients c_n given by:

$$\begin{aligned} c_n &= (a_n - ib_n)/2 & \quad for\ n > 0 \\ c_n &= (a_{|n|} + ib_{|n|})/2 & \quad for\ n < 0 \\ c_0 &= a_0/2 & \end{aligned} \tag{11.18}$$

Note that the absolute value $|n|$ is used for $n < 0$.

Problem c: Explain why the n-summation in (11.17) extends from $-\infty$ to ∞ rather than from 0 to ∞.

Problem d: The relations (11.7) and (11.8) can be used to express the coefficients c_n in the function $f(x)$. Treat the cases $n > 0$, $n < 0$ and $n = 0$ separately to show that for all values of n the coefficient c_n is given by:

$$c_n = \frac{1}{2L} \int_{-L}^{L} f(x) e^{-in\pi x/L} dx \,. \tag{11.19}$$

The sum (11.17) with expression (11.19) constitutes the complex Fourier transform over a finite interval. Again, there is a close analogy with the projections of vectors shown in section (10.1). Before we can explore this analogy, the inner product between two complex functions $f(x)$ and $g(x)$ needs to be defined. This inner product is not given by $(f \cdot g) = \int f(x)g(x)dx$. The reason for this is that the length of a vector is defined by $\|\mathbf{v}\|^2 = (\mathbf{v} \cdot \mathbf{v})$, a straightforward generalization of this expression to functions using the inner product given above would give for the norm of the function $f(x)$: $\|f\|^2 = (f \cdot f) = \int f^2(x)dx$. However, when $f(x)$ is purely imaginary this would lead to a negative norm. This can be avoided by defining the inner product of two complex functions by:

$$(f \cdot g) \equiv \int_{-L}^{L} f^*(x)g(x)dx \;, \tag{11.20}$$

where the asterisk denotes the complex conjugate.

Problem e: Show that with this definition the norm of $f(x)$ is given by $\|f\|^2 = (f \cdot f) = \int |f(x)|^2 \, dx$.

With this inner product the norm of the function is guaranteed to be positive. Now that we have an inner product the analogy with the projections in linear algebra can be explored. In order to do this, define the following basis functions:

$$u_n(x) \equiv \frac{1}{\sqrt{2L}} e^{in\pi x/L} . \tag{11.21}$$

Problem f: Show that these functions are orthonormal with respect to the inner product (11.20), i.e. show that:

$$(u_n \cdot u_m) = \delta_{nm} . \tag{11.22}$$

Pay special attention to the normalization of these functions; i.e. to the case $n = m$.

Problem g: Expand $f(x)$ in these basis functions, $f(x) = \sum_{n=-\infty}^{\infty} \gamma_n u_n(x)$ and show that $f(x)$ can be written as:

$$f = \sum_{n=-\infty}^{\infty} u_n \left(u_n \cdot f\right) . \tag{11.23}$$

Problem h: Make the comparison between this expression and the expressions for the projections of vectors in section (10.1).

11.3 The Fourier transform on an infinite interval

In several applications, one wants to compute the Fourier transform of a function that is defined on an infinite interval. This amounts to taking the limit $L \to \infty$. However, a simple inspection of (11.19) shows that one cannot simply take the limit $L \to \infty$ of the expressions of the previous section because in that limit $c_n = 0$. In order to define the Fourier transform for an infinite interval define the variable k by:

$$k \equiv \frac{n\pi}{L} . \tag{11.24}$$

An increment Δn corresponds to an increment Δk given by: $\Delta k = \pi \Delta n / L$. In the summation over n in the Fourier expansion (11.17), n is incremented by unity: $\Delta n = 1$. This corresponds to an increment $\Delta k = \pi / L$ of the variable k. In the limit $L \to \infty$ this increment goes to zero, this implies that the summation over n should be replaced by an integration over k:

$$\sum_{n=-\infty}^{\infty} (\cdots) \to \frac{\Delta n}{\Delta k} \int_{-\infty}^{\infty} (\cdots)\, dk = \frac{L}{\pi} \int_{-\infty}^{\infty} (\cdots)\, dk \qquad as\ L \to \infty . \tag{11.25}$$

Problem a: Explain the presence of the factor $\Delta n / \Delta k$ and show the last identity.

This is not enough to generalize the Fourier transform of the previous section to an infinite interval. As noted earlier, the coefficients c_n vanish in the limit $L \to \infty$. Also note that the integral in the right hand side of (11.25) is multiplied by L/π, this coefficient is infinite in the limit $L \to \infty$. Both problems can be solved by defining the following function:

$$F(k) \equiv \frac{L}{\pi} c_n , \tag{11.26}$$

where the relation between k and n is given by (11.24).

Problem b: Show that with the replacements (11.25) and (11.26) the limit $L \to \infty$ of the complex Fourier transform (11.17) and (11.19) can be taken and that the result can be written as:

$$f(x) = \int_{-\infty}^{\infty} F(k) e^{ikx} dk \ , \tag{11.27}$$

$$F(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-ikx} dx \ . \tag{11.28}$$

11.4 The Fourier transform and the delta function

In this section the Fourier transform of the delta function is treated. This is not only useful in a variety of applications, but it will also establish the relation between the Fourier transform and the closure relation introduced in section (10.1). Consider the delta function centered at $x = x_0$:

$$f(x) = \delta\left(x - x_0\right) \ . \tag{11.29}$$

Problem a: Show that the Fourier transform $F(k)$ of this function is given by:

$$F(k) = \frac{1}{2\pi} e^{-ikx_0} \ . \tag{11.30}$$

Problem b: Show that this implies that the Fourier transform of the delta function $\delta(x)$ centered at $x = 0$ is a constant. Determine this constant.

Problem c: Use expression (11.27) to show that

$$\delta\left(x - x_0\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ik(x-x_0)} dk \ . \tag{11.31}$$

Problem d: Use a similar analysis to derive that

$$\delta\left(k - k_0\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i(k-k_0)x} dx \ . \tag{11.32}$$

These expressions are very useful in a number of applications. Again, there is close analogy between this expression and the projection of vectors introduced in section (10.1). To establish this connection let use define the following basis functions:

$$u_k(x) \equiv \frac{1}{\sqrt{2\pi}} e^{ikx} \ , \tag{11.33}$$

and use the inner product defined in (11.20) with the integration limits extending from $-\infty$ to ∞.

Problem e: Show that expression (11.32) implies that

$$\left(u_k \cdot u_{k_0}\right) = \delta(k - k_0) \ . \tag{11.34}$$

Why does this imply that the functions $u_k(x)$ form an orthonormal set?

Problem f: Use (11.31) to derive that:

$$\int_{-\infty}^{\infty} u_k(x)u_k^*(x_0)\, dk = \delta\left(x - x_0\right) \ . \tag{11.35}$$

This expression is the counterpart of the closure relation (10.13) introduced in section (10.1) for finite-dimensional vector spaces. Note that the delta function $\delta(x - x_0)$ plays the role of the identity operator $\mathbf{I}$ with components $I_{ij} = \delta_{ij}$ in equation (10.13) and that the summation $\sum_{i=1}^{N}$ over the basis vectors is replaced by an integration $\int_{-\infty}^{\infty} dk$ over the basis functions. Both differences are due to the fact that we are dealing in this section with an infinitely-dimensional function space rather than a finite-dimensional vector space. Also note that in (11.35) the complex conjugate is taken of $u_k(x_0)$. The reason for this is that for complex unit vectors $\hat{\mathbf{n}}$ the transpose in the completeness relation (10.13) should be replaced by the Hermitian conjugate. This involves taking the complex conjugate as well as taking the transpose.

11.5 Changing the sign and scale factor

In the Fourier transform (11.27) from the wave number domain (k) to the position domain (x), the exponent has a plus sign $\exp(+ikx)$ and the coefficient multiplying the integral is given by 1. In other texts on Fourier transforms you may encounter a different sign of the exponent and different scale factors are sometimes used in the Fourier transform. For example, the exponent in the Fourier transform from the wave number domain to the position domain may have a minus sign $exp\,(-ikx)$ and there may be a scale factor such as $1/\sqrt{2\pi}$ that differs from 1. It turns out that there is a freedom in choosing the sign of the exponential of the Fourier transform as well as in the scaling of the Fourier transform. We will first study the effect of a scaling parameter on the Fourier transform.

Problem a: Let the function $F(k)$ defined in (11.28) be related to a new function $\tilde{F}(k)$ by a scaling with a scale factor C: $F(k) = C\tilde{F}(k)$. Use the expressions (11.27) and (11.28) to show that:

$$f(x) = C\int_{-\infty}^{\infty} \tilde{F}(k)e^{ikx}dk \ , \tag{11.36}$$

$$\tilde{F}(k) = \frac{1}{2\pi C}\int_{-\infty}^{\infty} f(x)e^{-ikx}dx \ . \tag{11.37}$$

These expressions are completely equivalent to the original Fourier transform pair (11.27) and (11.28). The constant C is completely arbitrary. This implies that one may take any multiplication constant for the Fourier transform; the only restriction is that the product of the coefficients for Fourier transform and the backward transform is equal to $1/2\pi$.

Problem b: Show this last statement.

In the literature, notably in quantum mechanics, one often encounters the Fourier transform pair using the value $C = 1/\sqrt{2\pi}$. This leads to the Fourier transform pair:

$$f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \tilde{F}(k)e^{ikx}dk \ , \tag{11.38}$$

$$\tilde{F}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-ikx} dx \; . \qquad (11.39)$$

This normalization not only has the advantage that the multiplication factors for the forward and backward are identical $(1/\sqrt{2\pi})$, but the constants are also identical to the constant used in (11.33) to create a set of orthonormal functions.

Next we will investigate a change in the sign of the exponent in the Fourier transform. To do this, we will use the function $\tilde{F}(k)$ defined by: $\tilde{F}(k) = F(-k)$.

Problem c: Change the integration variable k in (11.27) to $-k$ and show that the Fourier transform pair (11.27) and (11.28) is equivalent to:

$$f(x) = \int_{-\infty}^{\infty} \tilde{F}(k) e^{-ikx} dk \; , \qquad (11.40)$$

$$\tilde{F}(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{ikx} dx \; . \qquad (11.41)$$

Note that these expressions only differ from earlier expression by the sign of the exponent. This means that there is a freedom in the choice of this sign. It does not matter which sign convention you use. Any choice of the sign and the multiplication constant for the Fourier transform can be used as long as:

(i) The product of the constants for the forward and backward transform is equal to $1/2\pi$ and (ii) the sign of the exponent for the forward and the backward a transform is opposite.

In this book, the Fourier transform pair (11.27) and (11.28) will mostly be used for the Fourier transform from the space (x) domain to the wave number (k) domain.

Of course, the Fourier transform can also be used to transform a function in the time (t) domain to the frequency (ω) domain. Perhaps illogically the following convention will used in this book for this Fourier transform pair:

$$f(t) = \int_{-\infty}^{\infty} F(\omega) e^{-i\omega t} d\omega \; , \qquad (11.42)$$

$$F(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt \; . \qquad (11.43)$$

The reason for this choice is that the combined Fourier transform from the (x,t)-domain to the (k,ω)-domain that is obtained by combining (11.27) and (11.42) is given by:

$$f(x,t) = \iint_{-\infty}^{\infty} F(k,\omega) e^{i(kx-\omega t)} dk d\omega \; . \qquad (11.44)$$

The function $e^{i(kx-\omega t)}$ in this integral describes a wave that moves for positive values of k and ω in the direction of increasing values of x. To see this, let us assume we are at a crest of this wave and that we follow the motion of the crest over a time Δt and that we want to find the distance Δx that the crest has moved in that time interval. If we follow a wave crest, the phase of the wave is constant, and hence $kx - \omega t$ is constant.

Problem d: Show that this implies that $\Delta x = c\Delta t$, with c given by $c = \omega/k$. Why does this imply that the wave moves with velocity c?

The exponential in the double Fourier transform (11.44) therefore describes for positive values of ω and k a wave travelling in the positive direction with velocity $c = \omega/k$. However, note that this is no proof that we should use the Fourier transform (11.44) and not a transform with a different choice of the sign in the exponent. In fact, one should realize that in the Fourier transform (11.44) one needs to integrate over *all* values of ω and k so that negative values of ω and k contribute to the integral as well.

Problem e: Use (11.28) and (11.43) to derive the inverse of the double Fourier transform (11.44).

11.6 The convolution and correlation of two signals

There are different ways in which one can combine signals to create a new signal. In this section the convolution and correlation of two signals is treated. For the sake of argument the signals are taken to be functions of time, and the Fourier transform pair (11.42) and (11.43) is used for the forward and inverse Fourier transform. Suppose a function $f(t)$ has a Fourier transform $F(\omega)$ defined by (11.42) and another function $h(t)$ has a similar Fourier transform $H(\omega)$:

$$h(t) = \int_{-\infty}^{\infty} H(\omega) e^{-i\omega t} d\omega \ . \tag{11.45}$$

The two Fourier transforms $F(\omega)$ and $H(\omega)$ can be multiplied in the frequency domain, and we want to find out what the Fourier transform of the product $F(\omega)H(\omega)$ is in the time domain.

Problem a: Show that:

$$F(\omega)H(\omega) = \frac{1}{(2\pi)^2} \iint_{-\infty}^{\infty} f(t_1) h(t_2) e^{i\omega(t_1+t_2)} dt_1 dt_2 \ . \tag{11.46}$$

Problem b: Show that after a Fourier transform this function corresponds in the time domain to:

$$\int_{-\infty}^{\infty} F(\omega)H(\omega)e^{-i\omega t} d\omega = \frac{1}{(2\pi)^2} \iiint_{-\infty}^{\infty} f(t_1) h(t_2) e^{i\omega(t_1+t_2-t)} dt_1 dt_2 d\omega \ . \tag{11.47}$$

Problem c: Use the representation (11.31) of the delta function to carry out the integration over ω and show that this gives:

$$\int_{-\infty}^{\infty} F(\omega)H(\omega)e^{-i\omega t} d\omega = \frac{1}{2\pi} \iint_{-\infty}^{\infty} f(t_1) h(t_2) \delta(t_1 + t_2 - t) dt_1 dt_2 \ . \tag{11.48}$$

Problem d: The integration over t_1 can now be carried out. Do this, and show that after renaming the variable t_2 to τ the result can be written as:

$$\int_{-\infty}^{\infty} F(\omega)H(\omega)e^{-i\omega t}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t-\tau)h(\tau)d\tau = \frac{1}{2\pi}(f*h)(t) \ . \tag{11.49}$$

The $\tau-$integral in the middle term is called the *convolution* of the functions f and h, this operation is denoted by the symbol $(f*h)$. Equation (11.49) states that a multiplication of the spectra of two functions in the frequency domain corresponds to the convolution of these functions in the time domain. For this reason, equation (11.49) is called the *convolution theorem*. This theorem is schematically indicated in the following diagram:

$$\begin{array}{rcl} f(t) & \longleftrightarrow & F(\omega) \\ h(t) & \longleftrightarrow & H(\omega) \\ \frac{1}{2\pi}(f*h) & \longleftrightarrow & F(\omega)H(\omega) \end{array}$$

Note that in the convolution theorem, a scale factor $1/2\pi$ is present in the left hand side. This scale factor depends on the choice of the scale factors that one uses in the Fourier transform, see section (11.5).

Problem e: Use a change of the integration variable to show that the convolution of f and h can be written in the following two ways:

$$(f*h)(t) = \int_{-\infty}^{\infty} f(t-\tau)h(\tau)d\tau = \int_{-\infty}^{\infty} f(\tau)h(t-\tau)d\tau \ . \tag{11.50}$$

Problem f: In order to see what the convolution theorem looks like when a different scale factor is used in the Fourier transform define $F(\omega) = C\tilde{F}(\omega)$, and a similar scaling for $H(\omega)$. Show that with this choice of the scale factors, the Fourier transform of $\tilde{F}(\omega)\tilde{H}(\omega)$ is in the time domain given by $(1/2\pi C)(f*h)(t)$. Hint: first determine the scale factor that one needs to use in the transformation from the frequency domain to the time domain.

The convolution of two time series plays a very important role in exploration geophysics. Suppose one carries out a seismic experiment where one uses a source such as dynamite to generate waves that propagate through the earth. Let the source signal in the frequency domain be given by $S(\omega)$. The waves reflect at layers in the earth and are recorded by geophones. In the ideal case, the source signal would have the shape of a simple spike, and the waves reflected by all the reflectors would show up as a sequence of individual spikes. In that case the recorded data would indicate the true reflectors in the earth. Let the signal $r(t)$ recorded in this ideal case have a Fourier transform $R(\omega)$ in the frequency domain. The problem that one faces is that a realistic seismic source is often not very impulsive. If the recorded data $d(t)$ have a Fourier transform $D(\omega)$ in the frequency domain, then this Fourier transform is given by

$$D(\omega) = R(\omega)S(\omega) \ . \tag{11.51}$$

One is only interested in $R(\omega)$ which is the earth response in the frequency domain, but in practice one records the product $R(\omega)S(\omega)$. In the time domain this is equivalent to saying that one has recorded the convolution $\int_{-\infty}^{\infty} r(\tau)s(t-\tau)d\tau$ of the earth response with the source signal, but that one is only interested in the earth response $r(t)$. One would like to "undo" this convolution, this process is called *deconvolution.* Carrying out the deconvolution seems trivial in the frequency domain. According to (11.51) one only needs to divide the data in the frequency domain by the source spectrum $S(\omega)$ to obtain $R(\omega)$. The problem is that in practice one often does not know the source spectrum $S(\omega)$. This makes seismic deconvolution a difficult process, see the collection of articles compiled by *Webster*[65]. It has been strongly argued by *Ziolkowski*[69] that the seismic industry should make a larger effort to record the source signal accurately.

The convolution of two signal was obtained in this section by taking the product $F(\omega)H(\omega)$ and carrying out a Fourier transform back to the time domain. The same steps can be taken by multiplying $F(\omega)$ with the complex conjugate $H^*(\omega)$ and by applying a Fourier transform to go the time domain.

Problem g: Take the similar steps as in the derivation of the convolution to show that

$$\int_{-\infty}^{\infty} F(\omega)H^*(\omega)e^{-i\omega t}d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(t+\tau)h^*(\tau)d\tau \; . \tag{11.52}$$

The right hand side of this expression is called the *correlation* of the functions $f(t)$ and $h^*(t)$. Note that this expression is very similar to the convolution theorem (11.50). This result implies that the Fourier transform of the product of a function and the complex conjugate in the frequency domain corresponds with the correlation in the time domain. Note again the constant $1/2\pi$ in the right hand side. This constant again depends on the scale factors used in the Fourier transform.

Problem h: Set $t = 0$ in expression (11.52) and let the function $h(t)$ be equal to $f(t)$. Show that this gives:

$$\int_{-\infty}^{\infty} |F(\omega)|^2 \, d\omega = \frac{1}{2\pi}\int_{-\infty}^{\infty} |f(t)|^2 \, dt \; . \tag{11.53}$$

This equality is known as *Parseval's theorem.* To see its significance, note that $\int_{-\infty}^{\infty} |f(t)|^2 \, dt = (f \cdot f)$, with the inner product of equation (11.20) with t as integration variable and with the integration extending from $-\infty$ to ∞. Since $\sqrt{(f \cdot f)}$ is the norm of f measured in the time domain, and since $\int_{-\infty}^{\infty} |F(\omega)|^2 \, d\omega$ is square of the norm of F measured in the frequency domain, Parseval's theorem states that with this definition of the norm, the norm of a function is equal in the time domain and in the frequency domain (up to the scale factor $1/2\pi$).

11.7 Linear filters and the convolution theorem

Let us consider a linear system that has an output signal $o(t)$ when it is given an input signal $i(t)$, see figure (11.2). There are numerous examples of this kind of systems. As an example, consider a damped harmonic oscillator that is driven by a force, this system is described by the differential equation $\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = F/m$, where the dot denotes a time derivative. The force $F(t)$ can be seen as the input signal, and the response $x(t)$ of the oscillator can be seen as the output signal. The relation between the input signal and the output signal is governed by the characteristics of the system under consideration, in this example it is the physics of the damped harmonic oscillator that determines the relation between the input signal $F(t)$ and the output signal $x(t)$.

Note that we have not defined yet what a *linear* filter is. A filter is linear when an input $c_1 i_1(t)+c_2 i_2(t)$ leads to an output $c_1 o_1(t)+c_2 o_2(t)$, when $o_1(t)$ is the output corresponding to the input $i_1(t)$ and $o_2(t)$ is the output the input $i_2(t)$.

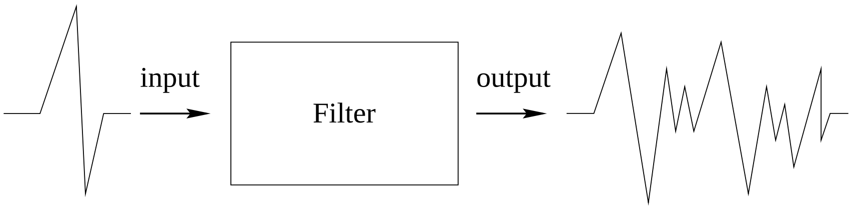


Figure 11.2: Schematic representation of a linear filter.

Problem a: Can you think of another example of a linear filter?

Problem b: Can you think of a system that has one input signal and one output signal, where these signals are related through a nonlinear relation? This would be an example of a nonlinear filter, the theory of this section would not apply to such a filter.

It is possible to determine the output $o(t)$ for any input $i(t)$ if the output to a delta function input is known. Consider the special input signal $\delta(t-\tau)$ that consists of a delta function centered at $t=\tau$. Since a delta function has "zero-width" (if it has a width at all) such an input function is very impulsive. Let the output for this particular input be denoted by $g(t,\tau)$. Since this function is the response at time t to an impulsive input at time τ this function is called the *impulse response*:

> *The impulse response function $g(t,\tau)$ is the output of the system at time t due to an impulsive input at time τ.*

How can the impulse response be used to find the response to an arbitrary input function? Any input function can be written as:

$$i(t)=\int_{-\infty}^{\infty}\delta(t-\tau)i(\tau)d\tau \ . \tag{11.54}$$

This identity follows from the definition of the delta function. However, we can also look at this expression from a different point of view. The integral in the right hand side of (11.54) can be seen as a superposition of infinitely many delta functions $\delta(t-\tau)$. Each delta function when considered as a function of t is centered at time τ. Since we integrate over τ these different delta functions are superposed to construct the input signal $i(t)$. Each of the delta functions in the integral (11.54) is multiplied with $i(\tau)$. This term plays the role of a coefficients that gives a weight to the delta function $\delta(t-\tau)$.

At this point it is crucial to use that the filter is linear. Since the response to the input $\delta(t-\tau)$ is the impulse response $g(t,\tau)$, and since the input can be written as the superposition (11.54) of delta function input signals $\delta(t-\tau)$, the output can be written as the same superposition of impulse response signals $g(t,\tau)$:

$$o(t) = \int_{-\infty}^{\infty} g(t,\tau) i(\tau) d\tau \ . \tag{11.55}$$

Problem c: Carefully compare the expressions (11.54) and (11.55). Note the similarity and make sure you understand the reasoning that has led to the previous expression.

You may find this "derivation" of (11.55) rather vague. The notion of the impulse response will be treated in much greater detail in chapter (14) because it plays a crucial role in mathematical physics.

At this point we will make another assumption about the system. Apart from the linearity we will also assume it is *invariant for translations in time*. This is a complex way of saying that we assume that the properties of the filter do not change with time. This is the case for the damped harmonic oscillator used in the beginning of this section. However, this oscillator would not be invariant for translations in time if the damping parameter would be a function of time as well: $\beta = \beta(t)$. In that case, the system would give a different response when the same input is used at different times.

When the properties of filter do not depend on time, the impulse response $g(t,\tau)$ depends only on the *difference* $t-\tau$. To see this, consider the damped harmonic oscillator again. The response at a certain time depends only the time that has lapsed between the excitation at time τ and the time of observation t. Therefore, for a time-invariant filter:

$$g(t,\tau) = g(t-\tau) \ . \tag{11.56}$$

Inserting this in (11.55) shows that for a linear time-invariant filter the output is given by the convolution of the input with the impulse response:

$$o(t) = \int_{-\infty}^{\infty} g(t-\tau) i(\tau) d\tau = (g * i)(t) \ . \tag{11.57}$$

Problem d: Let the Fourier transform of $i(t)$ be given by $I(\omega)$, the Fourier transform of $o(t)$ by $O(\omega)$ and the Fourier transform of $g(t)$ by $G(\omega)$. Use (11.57) to show that these Fourier transforms are related by:

$$O(\omega) = 2\pi G(\omega) I(\omega) \ . \tag{11.58}$$

Expressions (11.57) and (11.58) are key results in the theory in linear time-invariant filters. The first expression states that one only needs to know the response $g(t)$ to a single impulse to compute the output of the filter to *any* input signal $i(t)$. Equation (11.58) has two important consequences. First, if one knows the Fourier transform $G(\omega)$ of the impulse response, one can compute the Fourier transform $O(\omega)$ of the output. An inverse Fourier transform then gives the output $o(t)$ in the time domain.

Problem e: Show that $G(\omega)e^{-i\omega t}$ is the response of the system to the input signal $e^{-i\omega t}$.

This means that if one knows the response of the filter to the harmonic signal $e^{-i\omega t}$ at any frequency, one knows $G(\omega)$ and the response to any input signal can be determined.

The second important consequence of (11.58) is that the output at frequency ω does depend only at the input and impulse response at the same frequency ω, but not on other frequencies. This last property does not hold for nonlinear systems, because in that case different frequency components of the input signal are mixed by the non-linearity of the system. An example of this phenomenon is given by *Snieder*[55] who shows that observed variations in the earth's climate contain frequency components that cannot be explained by periodic variations in the orbital parameters in the earth, but which are due to the nonlinear character of the climate response to the amount of energy received by the sun.

The fact that a filter can either be used by specifying its Fourier transform $G(\omega)$ (or equivalently the response to an harmonic input $\exp -i\omega t)$ or by prescribing the impulse response $g(t)$ implies that a filter can be designed either in the frequency domain or in the time domain. In section (11.8) the action of a filter is designed in the time domain. A Fourier transform then leads to a compact description of the filter response in the frequency domain. In section (11.9) the converse route is taken; the filter is designed in the frequency domain, and a Fourier transform is used to derive an expression for the filter in the time domain.

As a last reminder it should be mentioned that although the theory of linear filters is introduced here for filters that act in the time domain, the theory is of course equally valid for filters in the spatial domain. In the case the wave number k plays the role that the angular frequency played in this section. Since there may be more than one spatial dimension, the theory must in that case be generalized to include higher-dimensional spatial Fourier transforms. However, this does not change the principles involved.

11.8 The dereverberation filter

As an example of a filter that is derived in the time domain we consider here the description of reverberations on marine seismics. Suppose a seismic survey is carried out at sea. In such an experiment a ship tows a string of hydrophones that record the pressure variations in the water just below the surface of the water, see figure (11.3). Since the pressure at the surface of the water vanishes, the surface of the water totally reflects pressure waves and the reflection coeficient for reflection at the water surface is equal to -1. Let the reflection coeficient for waves reflecting upwards from the water bottom be denoted by r. Since the constrast between the water and the solid earth below is not small, this reflection coefficient can be considerable.

Problem a: Give a physical argument why this reflection coefficients must be smaller or equal than unity: $r \leq 1$.

Since the reflection coefficient of the water bottom is not small, waves can bounce back and forth repeatedly between the water surface and the water bottom. These reverberations are an unwanted artifact in seismic experiments. The reason for this is that a wave that has bounced back and forth in the water layer can be misinterpreted on a seismic section as a reflector in the earth. For this reason one wants to eliminate these reverberations from seismic data.

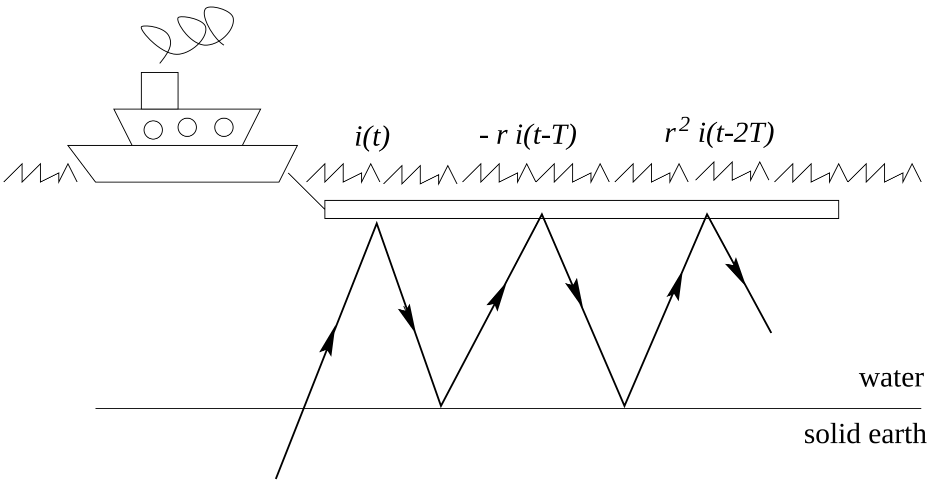


Figure 11.3: The generation of reverberations in a marine seismic experiment.

Suppose the the wavefield recorded by the hydrophones in the absense of reverberations is denoted by $i(t)$. Let the time it takes for wave to travel from the water surface to the water bottom and back be denoted by T.

Problem b: Show that the wave that has bounced back and forth once is given by $-r\ i(t-T)$. Hint; determine the amplitude of this wave from the reflection coefficients it encounters on its path and account for the time delay due to the bouncing up and down once in the water layer.

Problem c: Generalize this result to the wave that bounces back and forth n-times in the water layer and show that the signal $o(t)$ recorded by the hydrophones is given by:

$$o(t) = i(t) - r\ i(t-T) + r^2\ i(t-2T) + \cdots$$

or

$$o(t) = \sum_{n=0}^{\infty} (-r)^n\, i(t-nT) \tag{11.59}$$

see figure (11.3).

The notation $i(t)$ and $o(t)$ that was used in the previous section is deliberately used here. The action of the reverberation in the water layer is seen as a linear filter. The input of the filter $i(t)$ is the wavefield that would have been recorded if the waves would not bounce back and forth in the water layer. The output is the wavefield that results from the reverberations in the water layer. In a marine seismic experiment one records the wavefield $o(t)$ while one would like to know the signal $i(t)$ that contains just the reflections from below the water bottom. The process of removing the reverberations from the signal is called "dereverberation." The aim of this section is to derive a dereverberation filter that allows us to extract the input $i(t)$ from the recorded output $o(t)$.

Problem d: Can you see a way to determine $i(t)$ from (11.59) when $o(t)$ is given?

Problem e: It may not be obvious that expression (11.59) describes a linear filter of the form (11.57) that maps the input $i(t)$ onto the output $o(t)$. Show that expression (11.59) can be written in the form (11.57) with the impulse response $g(t)$ given by:

$$g(t) = \sum_{n=0}^{\infty} (-r)^n \, \delta(t - nT) \ , \tag{11.60}$$

with $\delta(t)$ the Dirac delta function.

Problem f: Show that $g(t)$ is indeed the impulse response, in other words: show that if a delta function is incident as a primary arrival at the water surface, that the reverbarations within the water layer lead to the signal (11.60).

You probably discovered it is not simple to solve **problem d**. However, the problem becomes much simpler by carrying out the analysis in the frequency domain. Let the Fourier transforms of $i(t)$ and $o(t)$ as defined by the transform (11.43) be denoted by $I(\omega)$ and $O(\omega)$ respectively. It follows from expression (11.59) that one needs to find the Fourier transform of $i(t - nT)$.

Problem g: According to the definition (11.43) the Fourier transform of $i(t-\tau)$ is given by $1/2\pi \int_{-\infty}^{\infty} i(t - \tau) \exp i\omega t \; dt$. Use a change of the integration variable to show that the Fourier transform of $i(t - \tau)$ is given by $I(\omega) \exp i\omega\tau$.

What you have derived here is the *shift property* of the Fourier transform, a translation of a function over a time τ corresponds in the frequency domain to a multiplication with $\exp i\omega\tau$:

$$\begin{array}{ccc} i(t) & \longleftrightarrow & I(\omega) \\ i(t-\tau) & \longleftrightarrow & I(\omega) \exp i\omega\tau \end{array} \tag{11.61}$$

Problem h: Apply a Fourier transform to expression (11.59) for the output, use the shift property (11.61) for each term and show that the output in the frequency domain is related to the Fourier transform of the input by the following expression:

$$O(\omega) = \sum_{n=0}^{\infty} (-r)^n \, e^{i\omega nT} I(\omega) \ . \tag{11.62}$$

Problem i: Use the theory of section (11.7) to show that the filter that describes the generation of reverberations is in the frequency domain given by:

$$G(\omega) = \frac{1}{2\pi} \sum_{n=0}^{\infty} (-r)^n \, e^{i\omega nT} \ . \tag{11.63}$$

Problem j: Since we know that the reflection coefficient r is less or equal to 1 (see **problem a**), this series is guaranteed to converge. Sum this series to show that

$$G(\omega) = \frac{1}{2\pi} \frac{1}{1 + re^{i\omega T}} \ . \tag{11.64}$$

This is a very useful result because it implies that the output and the input are in the frequency domain related by

$$O(\omega) = \frac{1}{1 + re^{i\omega T}} \, I(\omega) \; . \tag{11.65}$$

Note that the action of the reverberation leads in the frequency domain to a simple division by $(1 + r \exp i\omega T)$. Note that this expression (11.65) has a similar form as equation (2.32) of section (2.3) that accounts for the reverberation of waves between two stacks of reflectors. This resemblance is no coincidence because the physics of waves bouncing back and forth between two reflectors is similar.

Problem k: The goal of this section was to derive the dereverberation filter that produces $i(t)$ when $o(t)$ is given. Use expression (11.65) to derive the dereverberation filter in the frequency domain.

The dereverberation filter you have just derived is very simple in the frequency domain, it only involves a multiplication of every frequency component $O(\omega)$ with a scalar. Since multiplication is a simple and efficient procedure it is attractive to carry out dereverberation in the frequency domain. The dereverberation filter you have just derived was developed originally by *Backus*[4].

The simplicity of the dereverberation filter hides a nasty complication. If the reflection coefficient r and the two-way travel time T are exactly known and if the water bottom is exactly horizontal there is no problem with the dereverberation filter. However, in practice one only has *estimates* of these quantities, let these estimates be denoted by r' and T' respectively. The reverbarations lead in the frequency domain to a division by $1 + r \exp i\omega T$ while the dereverberation filter based on the estimated parameters leads to a multiplication with $1 + r' \exp i\omega T'$. The net effect of the generation of the reverberations and the subsequent dereverberation thus is in the frequency domain given by a multiplication with

$$\frac{1 + r' \exp i\omega T'}{1 + r \exp i\omega T}$$

Problem l: Show that when the reflection coefficients are close to unity and when the estimate of the travel time is not accurate ($T' \neq T$) the term given above differs appreciably from unity. Explain that this implies that the dereverberation does not work well.

In practice one does not only face the problem that the estimates of the reflection coefficients and the two-way travel time may be inaccurate. In addition the water bottom may not be exactly flat and there may be variations in the reflections coefficient along the water bottom. In that case the action of the dereverberation filter can be significantly degraded.

11.9 Design of frequency filters

In this section we consider the problem that a time series $i(t)$ is recorded and that this time series is contaminated with high-frequency noise. The aim of this section is to derive a filter in the time domain that removes the frequency components with a frequency greater than a cut-off frequency ω_0 from the time series. Such a filter is called a low-pass filter because only frequencies components lower than the threshold ω_0 pass the filter.

Problem a: Show that this filter is in the frequency domain given by:

$$G(\omega) = \left\{ \begin{array}{lll} 1 & if & |\omega| \leq \omega_0 \\ 0 & if & |\omega| > \omega_0 \end{array} \right. \tag{11.66}$$

Problem b: Explain why the absolute value of the frequency should be used in this expression.

Problem c: Show that this filter is in the time domain given by

$$g(t) = \int_{-\omega_0}^{\omega_0} e^{-i\omega t} \, d\omega \ . \tag{11.67}$$

Problem d: Carry out the integration over frequency to derive that the filter is explicitly given by

$$g(t) = 2\omega_0 \, \mathrm{sinc}\left(\omega_0 t\right) \ , \tag{11.68}$$

where the sinc-function is defined by

$$\mathrm{sinc}\left(x\right) \equiv \frac{\sin\left(x\right)}{x} \ . \tag{11.69}$$

Problem e: Sketch the impulse response (11.68) of the low-pass filter as a function of time. Determine the behaviour of the filter for $t = 0$ and show that the first zero crossing of the filter is at time $t = \pm\pi/\omega_0$.

The zero crossing of the filter is of fundamental importance. It implies that the width of the impulse response in the time domain is given by $2\pi/\omega_0$.

Problem f: Show that the width of the filter in the frequency domain is given by $2\omega_0$.

This means that when the the cut-off frequency ω_0 is increased, the width of the filter in the frequency domain increases but the width of the filter in the time domain decreases. A large width of the filter in the frequency domain corresponds to a small width of the filter in the time domain and vice versa.

Problem g: Show that the product of the width of the filter in the time domain and the width of the same filter in the frequency domain is given by 4π.

The significance of this result is that this product is independent of frequency. This implies that the filter cannot be arbitrary peaked both in the time domain and the frequency domain. This effect has pronounced consequences since it is the essence of the *uncertainty relation of Heisenberg* which states that the position and momentum of a particle can never be known exactly, more details can be found in the book of *Mertzbacher*[**?**].

The filter (11.68) does actually not have very desirable properties, it has two basic problems. The first problem is that the filter decays only slowly with time. This means that the filter is very long in the time domain, and hence the convolution of a time series with the filter is numerically a rather inefficient process. This can be solved by making the cutoff of the filter in the frequency domain more gradual than the frequency cut-off defined in expression (11.66), for example by using the filter $G(\omega) = \left(1 + \frac{|\omega|}{\omega_0}\right)^{-n}$ with n a positive integer.

Problem h: Does this filter have the steepest cutoff for low values of n or for high values of n? Hint: make a plot of $G(\omega)$ as a function of ω.

The second problem is that the filter is not *causal*. This means that when a function is convolved with the filter (11.68), the output of the filter depends on the value of the input at later times, i.e. the filter output depends on the input in the future.

Problem i: Show that this is the case, and that the output depends on the the input on earlier times only when $g(t) = 0$ for $t < 0$.

A causal filter can be designed by using the theory of analytic functions shown in chapter (12). The design of filters is quite an art, details can be found for example in the books of *Robinson and Treitel*[51] or *Claerbout*[15].

11.10 Linear filters and linear algebra

There is a close analogy between the theory of linear filters of section (11.7) and the eigenvector decomposition of a matrix in linear algebra as treated in section (10.5). To see this we will use the same notation as in section (11.7) and use the Fourier transform (11.45) to write the output of the filter in the time domain as:

$$o(t) = \int_{-\infty}^{\infty} O(\omega) e^{-i\omega t} d\omega \ . \tag{11.70}$$

Problem a: Use expression (11.58) to show that this can be written as

$$o(t) = 2\pi \int_{-\infty}^{\infty} G(\omega) I(\omega) e^{-i\omega t} d\omega \ , \tag{11.71}$$

and carry out an inverse Fourier transform of $I(\omega)$ to find the following expression

$$o(t) = \iint_{-\infty}^{\infty} G(\omega) e^{-i\omega t} e^{i\omega\tau} i(\tau) d\omega d\tau \ . \tag{11.72}$$

In order to establish the connection with linear algebra we introduce by analogy with (11.33) the following basis functions:

$$u_\omega(t) \equiv \frac{1}{\sqrt{2\pi}} e^{-i\omega t} \ , \tag{11.73}$$

and the inner product

$$(f \cdot g) \equiv \int_{-\infty}^{\infty} f^*(t) g(t) dt \ . \tag{11.74}$$

Problem b: Show that these basisfunctions are orthonormal for this inner product in the sense that

$$(u_\omega \cdot u_{\omega'}) = \delta(\omega - \omega') \ . \tag{11.75}$$

Problem c: These functions play the same role as the eigenvectors in section (10.5). To which expression in section (10.5) does the above expression correspond?

Problem d: Show that equation (11.72) can be written as

$$o(t) = \int_{-\infty}^{\infty} G(\omega) u_\omega(t) \left(u_\omega \cdot i\right) d\omega \ . \tag{11.76}$$

This expression should be compared with equation (10.60) of section (10.5)

$$\mathbf{Ap} = \sum_{n=1}^{N} \lambda_n \hat{\mathbf{v}}^{(n)} \left(\hat{\mathbf{v}}^{(n)} \cdot \mathbf{p}\right) \qquad (10.60) \quad \textit{again} \ .$$

The integration over frequency plays the same role as the summation over eigenvectors in equation (10.60). Expression (11.76) can be seen as a description for the operator $g(t)$ in the time domain that maps the input function $i(t)$ onto the output $o(t)$.

Problem e: Use the equations (11.57), (11.74) and (11.76) to show that:

$$g(t-\tau) = 2\pi \int_{-\infty}^{\infty} G(\omega) u_\omega(t) u_\omega^*(\tau) d\omega \ . \tag{11.77}$$

There is a close analogy between this expression and the dyadic decompostion of a matrix in its eigenvectors and eigenvalues derived in section (10.5).

Problem f: To see this connection show that equation (10.61) can be written in component form as:

$$A_{ij} = \sum_{n=1}^{N} \lambda_n \hat{v}_i^{(n)} \hat{v}_j^{(n)T} \ . \tag{11.78}$$

The sum over eigenvalues in (11.78) corresponds with the integration over frequency in (11.77). In section (10.5) linear algebra in a finite-dimensional vector space was treated, in such a space there is a finite number of eigenvalues. In this section, a function space with infinitely many degrees of freedom is analyzed; it will be no surprise that for this reason the sum over a finite number of eigenvules should be replaced by an integration over the continuous variable ω. The index i in (11.78) corresponds with the variable t in (11.77) while the index j corresponds with the variable τ.

Problem g: Establish the connection between *all* variables in the equations (11.77) and (11.78). Show specifically that $G(\omega)$ plays the role of eigenvalue and u_ω plays the role of eigenvector. Which operation in (11.77) corresponds to the transpose that is taken of the second eigenvector in (11.78)?

You may wonder why the function $u_\omega(t) = \exp(-i\omega t)/\sqrt{2\pi}$ defined in (11.73) and not some other function plays the role of eigenvector of the impulse respons operator $g(t-\tau)$. To see this we have to understand what a linear filter actually does. Let us first consider the example of the reverberation filter of section (11.8). According to (11.59) the reverberation filter is given by:

$$o(t) = i(t) - r\, i(t-T) + r^2\, i(t-2T) + \cdots \qquad (11.59)\ again$$

It follows from this expression that what the filter really does is to take the input $i(t)$, translate it over a time nT to a new function $i(t-nT)$, multiply each term with $(-r)^n$ and sum over all values of n. This means that the filter is a combination of three operations, (i) translation in time, (ii) multiplication and (iii) summation over n. The same conclusion holds for any general time-invariant linear filter.

Problem h: Use a change of the integration variable to show that the action of a time-invariant linear filter as given in (11.57) can be written as

$$o(t) = \int_{-\infty}^{\infty} g(\tau) i(t-\tau) d\tau \; . \qquad (11.79)$$

The function $i(t-\tau)$ is the function $i(t)$ translated over a time τ. This translated function is multiplied with $g(\tau)$ and an integration over all values of τ is carried out. This means that in general the action of a linear filter can be seen as a combination of translation in time, multiplication and integration over all translations τ. How can this be used to explain that the correct eigenfunctions to be used are $u_\omega(t) = \exp(-i\omega t)/\sqrt{2\pi}$? The answer does not lie in the multiplication because *any* function is eigenfunction of the operator that carries out multplication with a constant, i.e. $af(t) = \lambda f(t)$ for every function $f(t)$.

Problem i: What is the eigenvalue λ?

This implies that the translation operator is the reason that the eigenfunctions are $u_\omega(t) = \exp(-i\omega t)/\sqrt{2\pi}$. Let the operator that carries out a translation over a time τ be denoted by T_τ:

$$T_\tau f(t) \equiv f(t-\tau) \; . \qquad (11.80)$$

Problem j: Show that the functions $u_\omega(t)$ defined in (11.73) are the eigenfunctions of the translation operator T_τ, i.e. show that $T_\tau u_\omega(t) = \lambda u_\omega(t)$. Express the eigenvalue λ of the translation operator in the translation time τ.

Problem k: Compare this result with the shift property of the Fourier transform that was derived in (11.61).

This means that the functions $u_\omega(t)$ are the eigenfunctions to be used for the eigenfunction decomposition of a linear time-invariant filter, because these functions are eigenfunctions of the translation operator.

Problem l: You identified in **problem e** the eigenvalues of the filter with $G(\omega)$. Show that this interpretation is correct, in other words show that when the filter g acts on the function $u_\omega(t)$ the result can be written as $G(\omega)u_\omega(t)$. Hint: go back to **problem e** of section (11.7).

This analysis shows that the Fourier transform, which uses the functions $\exp(-i\omega t)$ is so useful because these functions are the eigenfunctions of the translation operator. However, this also points to a limitation of the Fourier transform. Consider a linear filter that is not time-invariant, that is a filter where the output does not depend only on the *difference* between the input time τ and the output time t. Such a filter satisfies the general equation (11.55) rather than the convolution integral (11.57). The action of a filter that is not time-invariant can in general not be written as a combination of the following operations: multiplication, translation and integration. This means that for such a filter the functions $\exp(-i\omega t)$ that form the basis of the Fourier transform are not the appropriate eigenfunctions. The upshot of this is that in practice the Fourier transform is only useful for systems that are time-dependent, or in general that are translationally invariant in the coordinate that is used.

Chapter 12

Analytic functions

In this section we will consider complex functions in the complex plane. The reason for doing this is that the requirement that the function "behaves well" (this is defined later) imposes remarkable constraints on such complex functions. Since these constraints coincide with some of the laws of physics, the theory of complex functions has a number of important applications in mathematical physics. In this chapter complex functions $h(z)$ are treated that are decomposed in a real and imaginary parts:

$$h(z) = f(z) + ig(z) \ , \tag{12.1}$$

hence the functions $f(z)$ and $g(z)$ are assumed to be real. The complex number z will frequently be written as $z = x + iy$, so that $x = \Re(z)$ and $y = \Im(z)$ where $\Re$ and $\Im$ denote the real and imaginary part respectively.

12.1 The theorem of Cauchy-Riemann

Let us first consider a real function $F(x)$ of a real variable x. The derivative of such a function is defined by the rule

$$\frac{dF}{dx} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} \ . \tag{12.2}$$

In general there are two ways in which Δx can approach zero; from above and from below. For a function that is differentiable it does not matter whether Δx approaches zero from above or from below. If the limits $\Delta x \downarrow 0$ and $\Delta x \uparrow 0$ do give a different result it is a sign that the function does not behave well, it has a kink and the derivative is not unambiguously defines, see figure (12.1).

For complex functions the derivative is defined in the same way as in equation (12.1) for real functions:

$$\frac{dh}{dz} = \lim_{\Delta z \to 0} \frac{h(z + \Delta z) - h(z)}{\Delta z} \ . \tag{12.3}$$

For real functions, Δx could approach zero in two ways, from below and from above. However, the limit $\Delta z \to 0$ in (12.3) can be taken in infinitely many ways. As an example see figure (12.2) where several paths are sketched that one can use to let Δz approach zero. This does not always give the same result.

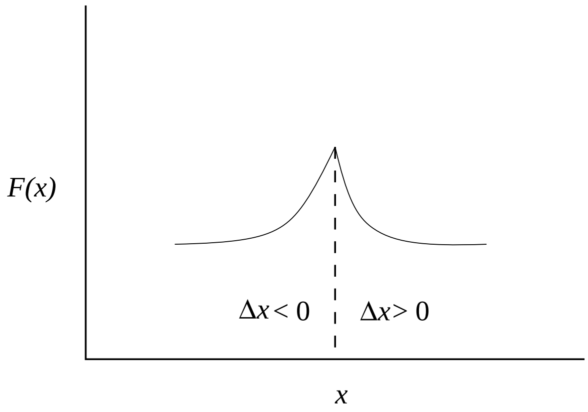


Figure 12.1: A function F(x) that is not differentiable.

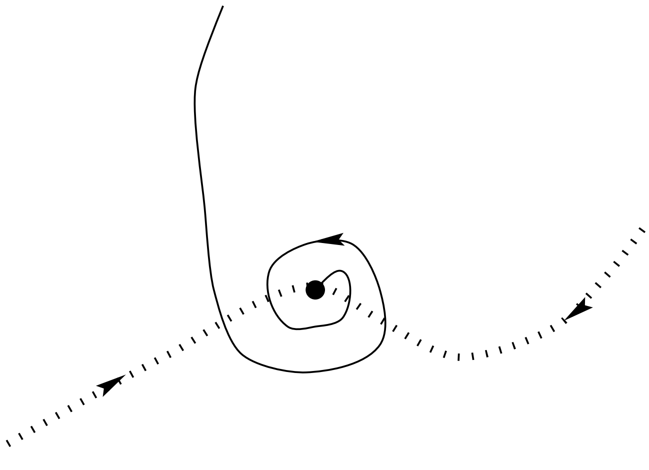


Figure 12.2: Examples of paths along which the limit can be taken.

Problem a: Consider the function $h(z) = \exp(1/z)$. Using the definition (12.3) compute dh/dz at the point $z = 0$ when Δz approaches zero (i) from the positive real axis, (ii) from the negative real axis, (iii) from the positive imaginary axis and (iv) from the negative imaginary axis.

You have discovered that for some functions the result of the limit Δz depends critically on the path that one uses in the limit process. The derivative of such a function is not defined unambiguously. However, for many functions the value of the derivative does *not* depend on the way that Δz approaches zero. When these functions and their derivative are also finite, they are called *analytic functions.* The requirement that the derivative does not depend on the way in which Δz approaches zero imposes a strong constraint on the real and imaginary part of the complex function. To see this we will let Δz approach zero along the real axis and along the imaginary axis.

Problem b: Consider a complex function of the form (12.1) and compute the derivative dh/dz by setting $\Delta z = \Delta x$ with Δx a real number. (Hence Δz approaches zero along the real axis). Show that the derivative is given by $dh/dz = \partial f/\partial x + i\partial g/\partial x$.

Problem c: Compute the derivative dh/dz also by setting $\Delta z = i\Delta y$ with Δy a real number. (Hence Δz approaches zero along the imaginary axis.) Show that the derivative is given by $dh/dz = \partial g/\partial y - i\partial f/\partial y$.

Problem d: When $h(z)$ is analytic these two expressions for the derivative are by definition equal. Show that this implies that:

$$\frac{\partial f}{\partial x} = \frac{\partial g}{\partial y} , \tag{12.4}$$

$$\frac{\partial g}{\partial x} = -\frac{\partial f}{\partial y} . \tag{12.5}$$

These are puzzling expressions since the conditions (12.4) and (12.5) imply that the real and imaginary part of an analytic complex functions are not independent of each other, they are coupled by the constraints imposed by the equations above. The expressions (12.4) and (12.5) are called the Cauchy-Riemann relations.

Problem e: Use these relations to show that both $f(x, y)$ and $g(x, y)$ are harmonic functions. These are functions for which the Laplacian vanishes:

$$\nabla^2 f = \nabla^2 g = 0 . \tag{12.6}$$

Hence we have found not only that f and g are coupled to each other; in addition the functions f and g must be harmonic functions. This is exactly the reason why this theory is so useful in mathematical physics because harmonic functions arise in several applications, see the examples of the coming sections. However, we have not found all the properties of harmonic functions yet.

Problem f: Show that:

$$(\nabla f \cdot \nabla g) = 0 . \tag{12.7}$$

Since the gradient of a function is perpendicular to the lines where the function is constant this implies that the curves where f is constant and where g is constant intersect each other at a fixed angle.

Problem g: Determine this angle.

Problem h: Verify the properties (12.4) through (12.7) explicitly for the function $h(z) = z^2$. Also sketch the lines in the complex plane where $f = \Re(h)$ and $g = \Im(h)$ are constant.

Still we have not fully explored all the properties of analytic functions. Let us consider a line integral $\oint_C h(z)dz$ along a closed contour C in the complex plane.

Problem i: Use the property $dz = dx + idy$ to deduce that:

$$\oint_C h(z)dz = \oint_C \mathbf{v}\cdot d\mathbf{r} + i\oint_C \mathbf{w}\cdot d\mathbf{r} \ , \tag{12.8}$$

where $d\mathbf{r} = \begin{pmatrix} dx \\ dy \end{pmatrix}$ and with the vectors $\mathbf{v}$ and $\mathbf{w}$ defined by:

$$\mathbf{v} = \begin{pmatrix} f \\ -g \end{pmatrix} \quad , \quad \mathbf{w} = \begin{pmatrix} g \\ f \end{pmatrix} . \tag{12.9}$$

Note that we now use x and y both as the real and imaginary part of a complex number, but also as the Cartesian coordinates in a plane. In the following problem we will (perhaps confusingly) use the notation z both for a complex number in the complex plane, as well as for the familiar z-coordinate in a three-dimensional Cartesian coordinate system.

Problem j: Show that the Cauchy-Riemann relations (12.4)-(12.5) imply that the z-component of the *curl* of $\mathbf{v}$ and $\mathbf{w}$ vanishes: $(\nabla\times\mathbf{v})_z = (\nabla\times\mathbf{w})_z = 0$, and use (12.8) and the theorem of Stokes (7.2) to show that when $h(z)$ is analytic everywhere within the contour C that:

$$\oint_C h(z)dz = 0 \qquad , h(z) \ \textit{analytic within } C \ . \tag{12.10}$$

This means that the line integral of a complex functions along *any* contour that encloses a region of the complex plane where that function is analytic is equal to zero. We will make extensive use of this property in section (13) where we deal with integration in the complex plane.

12.2 The electric potential

Analytic functions are often useful in the determination of the electric field and the potential for two-dimensional problems. The electric field satisfies the field equation (6.12): $(\nabla\cdot\mathbf{E}) = \rho(\mathbf{r})/\varepsilon_0$. In free space the charge density vanishes, hence $(\nabla\cdot\mathbf{E}) = 0$. The electric field is related to the potential V through the relation

$$\mathbf{E} = -\nabla V \ . \tag{12.11}$$

Problem a: Show that in free space the potential is a harmonic function:

$$\nabla^2 V(x, y) = 0 \ . \tag{12.12}$$

We can exploit the theory of analytic functions by noting that the real and imaginary parts of analytic functions both satisfy (12.12). This implies that if we take $V(x, y)$ to be the real part of a complex analytic function $h(x+iy)$, the condition (12.12) is automatically satisfied.

Problem b: If follows from (12.11) that the electric field is perpendicular to the lines where V is constant. Show that this implies that the electric field lines are also perpendicular to the lines $V = const.$ Use the theory of the previous section to argue that the field lines are the lines where the imaginary part of $h(x+iy)$ is constant.

This means that we receive a bonus by expressing the potential V as the real part of a complex analytic function, because the field lines simply follow from the requirement that $\Im(h) = const.$

Suppose we want to know the potential in the half space $y \geq 0$ when we have specified the potential on the x-axis. (Mathematically this means that we want to solve the equation $\nabla^2 V = 0$ for $y \geq 0$ when $V(x, y = 0)$ is given.) If we can find an analytic function $h(x+iy)$ such that on the x-axis (where $y = 0$) the real part of h is equal to the potential, we have solved our problem because the real part of h satisfies by definition the required boundary condition and it satisfies the field equation (12.12).

Problem c: Consider a potential that is given on the x-axis by

$$V(x, y = 0) = V_0 \exp\left(-x^2/a^2\right) \ . \tag{12.13}$$

Show that on the x-axis this function can be written as $V = \Re(h)$ with

$$h(z) = V_0 \exp\left(-z^2/a^2\right) \ . \tag{12.14}$$

Problem d: This means that we can determine the potential and the field lines throughout the half-plane $y \geq 0$. Use the theory of this section to show that the potential is given by

$$V(x, y) = V_0 \exp\left(\frac{y^2 - x^2}{a^2}\right) \cos\left(\frac{2xy}{a^2}\right) \ . \tag{12.15}$$

Problem e: Verify explicitly that this solution satisfies the boundary condition at the x-axis and that is satisfies the field equation (12.12).

Problem f: Show that the field lines are given by the relation

$$\exp\left(\frac{y^2 - x^2}{a^2}\right) \sin\left(\frac{2xy}{a^2}\right) = const. \tag{12.16}$$

Problem g: Sketch the field lines and the lines where the potential is constant in the half space $y \geq 0$.

In this derivation we have extended the solution $V(x, y)$ into the upper half plane by identifying it with an analytic function in the half plane that has on the x-axis a real part that equals the potential on the x-axis. Note that we found the solution to this problem without explicitly solving the partial differential equation (12.12) that governs the potential. The approach we have taken is called *analytic continuation* since we continue an analytic function from one region (the x-axis) into the upper half plane. Analytic continuation turns out to be a very unstable process. This can be verified explicitly for this example.

Problem h: Sketch the potential $V(x, y)$ as a function of x for the values $y = 0$, $y = a$ and $y = 10a$. What is the wavelength of the oscillations in the x-direction of the potential $V(x, y)$ for these values of y? Show that when we slightly perturb the constant a that the perturbation of the potential increases for increasing values of y. This implies that when we slightly perturb the boundary condition, that the solution is more perturbed as we move away from that boundary!

12.3 Fluid flow and analytic functions

As a second application of the theory of analytic functions we consider fluid flow. At the end of section (4.2) we have seen that the computation of the streamlines by solving the differential equation $d\mathbf{r}/dt = \mathbf{v}(\mathbf{r})$ for the velocity field (4.10)-(4.11) is extremely complex. Here the theory of analytic functions is used to solve this problem in a simple way. We consider once again a fluid that is incompressible ($(\nabla \cdot \mathbf{v}) = 0$) and will specialize to the special case that the vorticity of the flow vanishes:

$$\nabla \times \mathbf{v} = 0 \; . \tag{12.17}$$

Such a flow is called *irrotational* because it does not rotate (see the sections (5.3) and (5.4)). The requirement (12.17) is automatically satisfied when the velocity is the gradient of a scalar function f:

$$\mathbf{v} = \nabla f \; . \tag{12.18}$$

Problem a: Show this.

The function f plays for the velocity field $\mathbf{v}$ the same role as the electric potential V for the electric field $\mathbf{E}$. For this reason, flow with a vorticity equal to zero is called *potential flow.*

Problem b: Show that the requirement that the flow is incompressible implies that

$$\nabla^2 f = 0 \; . \tag{12.19}$$

We now specialize to the special case of incompressible and irrotational flow in two dimensions. In that case we can use the theory of analytic functions again to describe the flow field. Once we can identify the potential $f(x, y)$ with the real part of an analytic function $h(x + iy)$ we know that (12.19) must be satisfied.

Problem c: Consider the velocity field (4.9) due to a point source at $\mathbf{r} = 0$. Show that this flow field follows from the potential

$$f(x,y) = \frac{V}{2\pi} \ln r \; , \tag{12.20}$$

where $r = \sqrt{x^2 + y^2}$.

Problem d: Verify explicitly that this potential satisfies (12.19) except for $r = 0$. What is the physical reason that (12.19) is not satisfied at $r = 0$?

We now want to identify the potential $f(x,y)$ with the real part of an analytic function $h(x+iy)$. We know already that the flow follows from the requirement that the velocity is the gradient of the potential, hence it follows by taking the gradient of the real part of h. The curves $f = const.$ are perpendicular to the flow because $\mathbf{v} = \nabla f$ is perpendicular to these curves. However, it was shown in section (12.1) that the curves $g = \Im(h) = const.$ are perpendicular to the curves $f = \Re(h) = const.$ This means that the flow lines are given by the curves $g = \Im(h) = const.$ In order to use this we first have to find an analytic function with a real part given by (12.20). The simplest guess is to replace r in (12.20) by the complex variable z:

$$h(z) = \frac{V}{2\pi} \ln z \; . \tag{12.21}$$

Problem e: Verify that the real part of this function indeed satisfies (12.20) and that the imaginary part g of this function is given by:

$$g = \frac{V}{2\pi} \varphi \; , \tag{12.22}$$

where $\varphi = \arctan(y/x)$ is the argument of the complex number. Hint: use the representation $z = r \exp i\varphi$ for the complex numbers.

Problem f: Sketch the lines $g(x,y) = const.$ and verify that these lines indeed represent the flow lines in this flow.

Now we will consider the more complicated problem of section (4.2) where the flow has a source at $\mathbf{r}_+ = (L, 0)$ and a sink at $\mathbf{r}_- = (-L, 0)$. The velocity field is given by the equations (4.10) and (4.11) and our goal is to determine the slow lines without solving the differential equation $d\mathbf{r}/dt = \mathbf{v}(\mathbf{r})$. The points $\mathbf{r}_+$ and $\mathbf{r}_-$ can be represented in the complex plane by the complex numbers $z_+ = L + i0$ and $z_- = -L + i0$ respectively. For the source at $\mathbf{r}_+$ flow is represented by the potential $\frac{V}{2\pi} \ln|z - z_+|$, this follows from the solution (12.21) for a single source by moving this source to $z = z_+$

Problem g: Using a similar reasoning determine the contribution to the potential f by the sink at $\mathbf{r}_-$. Show that the potential of the total flow is given by:

$$f(z) = \frac{V}{2\pi} \ln \left(\frac{|z - z_+|}{|z - z_-|} \right) \; . \tag{12.23}$$

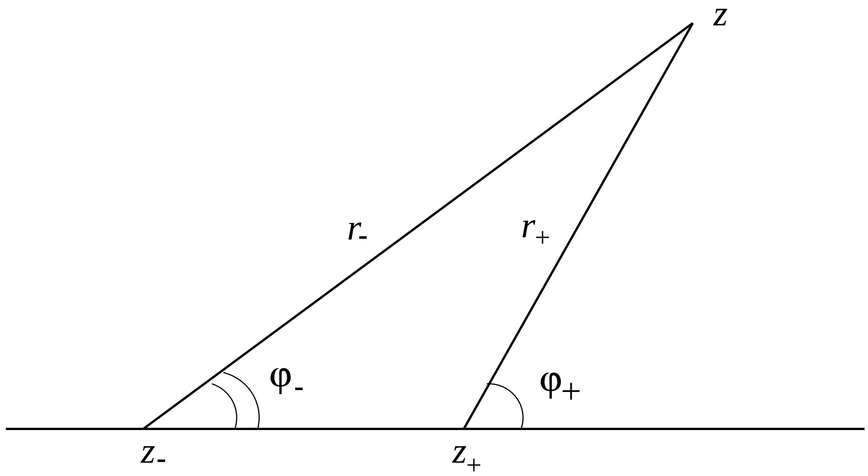


Figure 12.3: Definition of the geometric variables for the fluid flow with a source and a sink.

Problem h: Express this potential in x and y, compute the gradient and verify that this potential indeed gives the flow of the equations (4.10)-(4.11). You may find figure (12.3) helpful.

We have found that the potential f is the real part of the complex function

$$h(z) = \frac{V}{2\pi} \ln \left(\frac{z - z_+}{z - z_-} \right) \ . \tag{12.24}$$

Problem i: Write $z - z_\pm = r_\pm \exp i\varphi_\pm$ (with $r_\pm$ and $\varphi_\pm$ defined in figure (12.3)), take the imaginary part of (12.24), and show that $g = \Im(h)$ is given by:

$$g = \frac{V}{2\pi} \left(\varphi_+ - \varphi_- \right) \ . \tag{12.25}$$

Problem j: Show that the streamlines of the flow are given by the relation

$$\arctan \left(\frac{y}{x - L} \right) - \arctan \left(\frac{y}{x + L} \right) = const. \tag{12.26}$$

A figure of the streamlines can thus be found by making a contour plot of the function in the left hand side of (12.26). This treatment is definitely much simpler than solving the differential equation $d\mathbf{r}/dt = \mathbf{v}(\mathbf{r})$.

Chapter 13

Complex integration

In chapter (12) the properties of analytic functions in the complex plane were treated. One of the key-results is that the contour integral of a complex function is equal to zero when the function is analytic everywhere in the area of the complex plane enclosed by that contour, see expression (12.10). From this it follows that the integral of a complex function along a closed contour is only nonzero when the function is not analytic in the area enclosed by the contour. Functions that are not analytic come in different types. In this section complex functions are considered that are not analytic only at isolated points. These points where the function is not analytic are called the *poles* of the function.

13.1 Non-analytic functions

When a complex function is analytic at a point z_0, it can be expanded in a Taylor series around that point. This implies that within a certain region around z_0 the function can be written as:

$$h(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \ . \tag{13.1}$$

Note that in this sum only positive powers of $(z - z_0)$ appear.

Problem a: Show that the function $h(z) = \frac{\sin z}{z}$ can be written as a Taylor series around the point $z_0 = 0$ of the form (13.1) and determine the coefficients a_n.

Not all functions can be represented in a series of the form (13.1). As an example consider the function $h(z) = \exp(1/z)$. The function is not analytic at the point $z = 0$ (why?). Expanding the exponential leads to the expansion

$$h(z) = \exp(1/z) = 1 + \frac{1}{z} + \frac{1}{2!}\frac{1}{z^2} + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}\frac{1}{z^n} \ . \tag{13.2}$$

In this case an expansion in *negative* powers of z is needed to represent this function. Of course, each of the terms $1/z^n$ is for $n \geq 1$ singular at $z = 0$, this reflects the fact that the function $\exp(1/z)$ has a pole at $z = 0$. In this section we consider complex functions that can be expanded around a point z_0 as an infinite sum of *integer* powers of $(z - z_0)$:

$$h(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n \ . \tag{13.3}$$

However, it should be noted that not every function can be represented as such a sum.

Problem b: Can you think of one?

13.2 The residue theorem

It was argued at the beginning of this section that the integral of a complex function around a closed contour in the complex plane is only nonzero when the function is not analytic at some point in the area enclosed by the contour. In this section we will derive what the value is of the contour integral. Let us integrate a complex function $h(z)$ along a contour C in the complex plane that encloses a pole of the function at the point z_0, see the left panel of figure (13.1). Note that the integration is carried out in the counter-clockwise direction. It is assumed that around the point z_0 the function $h(z)$ can be expanded in a power series of the form (13.3). It is our goal to determine the value of the contour integral $\oint_C h(z)dz$.

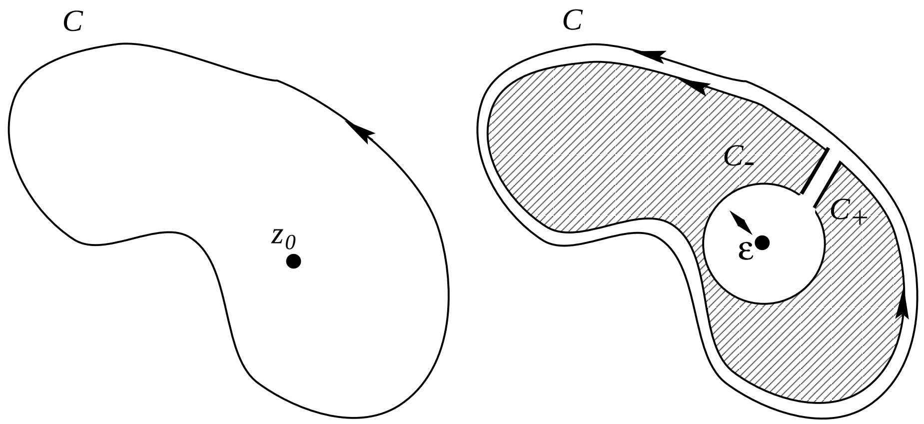


Figure 13.1: Definition of the contours for the contour integration.

The first step in the determination of the contour integral is to recognize that within the shaded area in the right panel of figure (13.1) the function $h(z)$ is analytic because we assumed that $h(z)$ is only non-analytic at the point z_0. By virtue of the identity (12.10) this implies that

$$\oint_{C^*} h(z)dz = 0 \ , \tag{13.4}$$

where the path C^* consists of the contour C, a small circle with radius ε around z_0 and the paths C^+ and C^- in the right panel of figure (13.1).

Problem a: Show that the integrals along C^+ and C^- do not give a net contribution to the total integral:

$$\int_{C^+} h(z)dz + \int_{C^-} h(z)dz = 0 \ . \tag{13.5}$$

Hint: note the opposite directions of integration along the paths C^+ and C^-.

Problem b: Use this result and expression (13.4) to show that the integral along the original contour C is identical to the integral along the small circle C_ε around the point where $h(z)$ is not analytic:

$$\oint_C h(z)dz = \oint_{C_\varepsilon} h(z)dz \ . \tag{13.6}$$

Problem c: The integration along C is in the counter-clockwise direction. Is the integration along C_ε in the clockwise or in the counter-clockwise direction?

Expression (13.6) is very useful because the integral along the small circle can be evaluated by using that close to z_0 the function $h(z)$ can be written as the series (13.3). When one does this the integration path C_ε needs to be parameterized. This can be achieved by writing the points on the path C_ε as

$$z = z_0 + \varepsilon \exp i\varphi \ , \tag{13.7}$$

with φ running from 0 to 2π since C_ε is a complete circle.

Problem d: Use the expressions (13.3), (13.6) and (13.7) to derive that

$$\oint_C h(z)dz = \sum_{n=-\infty}^{\infty} i a_n \varepsilon^{(n+1)} \int_0^{2\pi} \exp\left(i(n+1)\varphi\right) d\varphi \ . \tag{13.8}$$

This expression is very useful because it expresses the contour integral in the coefficients a_n of the expansion (13.3). It turns out that only the coefficient a_{-1} gives a nonzero contribution.

Problem e: Show by direct integration that:

$$\int_0^{2\pi} \exp\left(im\varphi\right) d\varphi = \left\{ \begin{array}{ll} 0 & for\ m \neq 0 \\ 2\pi & for\ m = 0 \end{array} \right. \tag{13.9}$$

Problem f: Use this result to derive that only the term $n = -1$ contributes to the sum in the right-hand side of (13.8) and that

$$\oint_C h(z)dz = 2\pi i a_{-1} \tag{13.10}$$

It may seem surprising that only the term $n = -1$ contributes to the sum in the right hand side of equation (13.8). However, we could have anticipated this result because we had already discovered that the contour integral does not depend on the precise choice of the integration path. It can be seen that in the sum (13.8) each term is proportional to $\varepsilon^{(n+1)}$. Since we know that the integral does not depend on the choice of the integration path, and hence on the size of the circle C_ε, one would expect that only terms that do not depend on the radius ε contribute. This is only the case when $n + 1 = 0$, hence only for the term $n = -1$ is the contribution independent of the size of the circle. It is indeed only this term that gives a nonzero contribution to the contour integral.

The coefficient a_{-1} is usually called the *residue* and is denoted by the symbol $Res\ h(z_0)$ rather than a_{-1}. However, remember that there is nothing mysterious about the residue, it is simply defined by the definition

$$Res\ h(z_0) \equiv a_{-1}\ . \tag{13.11}$$

With this definition the result (13.10) can trivially be written as

$$\oint_C h(z)dz = 2\pi i Res\ h(z_0) \qquad (counter-clockwise\ direction)\ . \tag{13.12}$$

This may appear to be a rather uninformative rewrite of expression (13.10) but it is the form (13.12) that you will find in the literature. The identity (13.12) is called the *residue theorem*.

Of course the residue theorem is only useful when one can determine the coefficient a_{-1} in the expansion (13.3). You can find in section (2.12) of the book of *Butkov*[14] an overview of methods for computing the residue. Here we will present the two most widely used methods. The first method is to determine the power series expansion (13.3) of the function explicitly.

Problem g: Determine the power series expansion of the function $h(z) = \frac{\sin z}{z^4}$ around $z = 0$ and use this expansion to determine the residue.

Unfortunately, this method does not always work. For some special functions other tricks can be used. Here we will consider functions with a simple pole, these are functions where in the expansion (13.3) the terms for $n < -1$ do not contribute:

$$h(z) = \sum_{n=-1}^{\infty} a_n (z - z_0)^n\ . \tag{13.13}$$

An example of such a function is $h(z) = \frac{\cos z}{z}$. The residue at the point z_0 follows by "extracting" the coefficient a_{-1} from the series (13.13).

Problem h: Multiply (13.13) with $(z - z_0)$, take the limit $z \to z_0$ to show that:

$$Res\ h(z_0) = \lim_{z \to z_0} (z - z_0) h(z_0) \qquad (simple\ pole)\ . \tag{13.14}$$

However, remember that this only works for functions with a simple pole, this recipe gives the wrong answer (infinity) when applied to a function that has nonzero coefficients a_n for $n < -1$ in the expansion (13.3).

In the treatment of this section we considered an integration in the counter-clockwise direction around the pole z_0.

Problem i: Redo the derivation of this section for a contour integration in the *clockwise* direction around the pole z_0 and show that in that case

$$\oint_C h(z)dz = -2\pi i Res\ h(z_0) \qquad (clockwise\ direction)\ . \tag{13.15}$$

Find out in which step of the derivation the minus sign is picked up!

Problem j: It may happen that a contour encloses not a single pole but a number of poles at points z_j. Find for this case a contour similar to the contour C^* in the right panel of figure (13.1) to show that the contour integral is equal to the *sum* of the contour integrals around the individual poles z_j. Use this to show that for this situation:

$$\oint_C h(z)dz = 2\pi i \sum_j Res\ h(z_j) \qquad (counter-clockwise\ direction)\ . \qquad (13.16)$$

13.3 Application to some integrals

The residue theorem has some applications to integrals that do not contain a complex variable at all! As an example consider the integral

$$I = \int_{-\infty}^{\infty} \frac{1}{1+x^2} dx\ . \qquad (13.17)$$

If you know that $1/(1+x^2)$ is the derivative of $\arctan x$ it is not difficult to solve this integral:

$$I = [\arctan x]_{-\infty}^{\infty} = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi\ . \qquad (13.18)$$

Now suppose you did not know that $\arctan x$ is the primitive function of $1/(1+x^2)$. In that case you would be unable to see that the integral (13.17) is equal to π. Complex integration offers a way to obtain the value of this integral in a systematic fashion.

First note that the path of integration in the integral can be viewed as the real axis in the complex plane. Nothing prevents us from viewing the real function $1/(1+x^2)$ as a complex function $1/(1+z^2)$ because on the real axis z is equal to x. This means that

$$I = \int_{C_{real}} \frac{1}{1+z^2} dz\ , \qquad (13.19)$$

where C_{real} denotes the real axis as integration path. At this point we cannot apply the residue theorem yet because the integration is not over a closed contour in the complex plane, see figure (13.2). Let us close the integration path using the circular path C_R in the upper half plane with radius R, see figure (13.2). In the end of the calculation we will let R go to infinity so that the integration over the semicircle moves to infinity.

Problem a: Show that

$$I = \int_{C_{real}} \frac{1}{1+z^2} dz = \oint_C \frac{1}{1+z^2} dz - \int_{C_R} \frac{1}{1+z^2} dz\ . \qquad (13.20)$$

The circular integral is over the closed contour in figure (13.2). What we have done is that we have closed the contour and that we subtract the integral over the semicircle that we added to obtain a closed integration path. This is the general approach in the application of the residue theorem; one adds segments to the integration path to obtain an integral over a closed contour in the complex plane, and corrects for the segments that one has added to the path. Obviously, this is only useful when the integral over the segment that

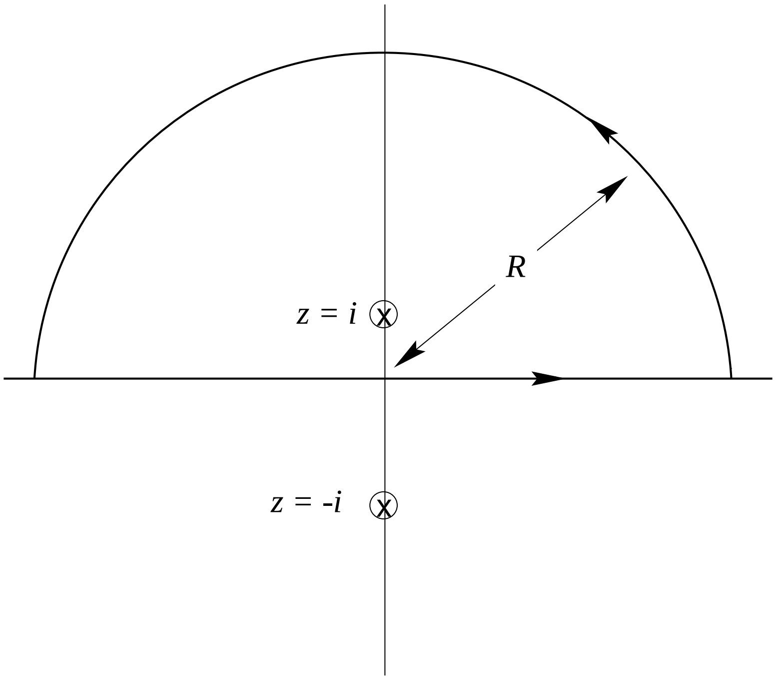


Figure 13.2: Definition of the integration paths in the complex plane.

one has added can be computed easily or when it vanishes. In this example the integral over the semicircle vanishes as $R \to \infty$. This can be seen with the following estimations:

$$\left| \int_{C_R} \frac{1}{1+z^2} dz \right| \leq \int_{C_R} \left| \frac{1}{1+z^2} \right| |dz| \leq \int_{C_R} \frac{1}{|z|^2 - 1} |dz| = \frac{\pi R}{R^2 - 1} \to 0 \qquad as\ R \to \infty \ . \tag{13.21}$$

Problem b: Justify each of these steps.

The estimate (13.21) implies that the last integral in (13.20) vanishes in the limit $R \to \infty$. This means that

$$I = \oint_C \frac{1}{1+z^2} dz \ . \tag{13.22}$$

Now we are in the position to apply the residue theorem because we have reduced the problem to the evaluation of an integral along a closed contour in the complex plane. We know from section (13.2) that this integral is determined by the poles of the function that is integrated within the contour.

Problem c: Show that the function $1/(1+z^2)$ has poles for $z = +i$ and $z = -i$.

Only the pole at $z = +i$ is within the contour C, see figure (13.2). Since $1/(1+z^2) = 1/\{(z-i)(z+i)\}$ this pole is a simple (why?).

Problem d: Use equation (13.14) to show that for the pole at $z = i$ the residue is given by: $Res = 1/2i$.

Problem e: Use the residue theorem (13.12) to deduce that

$$I = \int_{-\infty}^{\infty} \frac{1}{1+x^2} dx = \pi \ . \tag{13.23}$$

This value is identical to the value obtained at the beginning of this section by using that the primitive of $1/(1+x^2)$ is equal to $\arctan x$. Note that the analysis is very systematic and that we did not need to "know" the primitive function.

In the treatment of this problem there is no reason why the contour should be closed in the upper half plane. The estimate (13.21) hold equally well for a semicircle in the lower half-plane.

Problem f: Redo the analysis of this section when you close the contour in the lower half plane. Use that now the pole at $z = -i$ contributes and take into account that the sense of integration now is in the clockwise direction rather than the anti-clockwise direction. Show that this leads to the same result (13.23) that was obtained by closing the contour in the upper half-plane.

In the evaluation of the integral (13.17) there was a freedom whether to close the contour in the upper half-plane or in the lower half-plane. This is not always the case. To see this consider the integral

$$J = \int_{-\infty}^{\infty} \frac{\cos{(x-\pi/4)}}{1+x^2} dx \ . \tag{13.24}$$

Since $\exp ix = \cos x + i \sin x$ this integral can be written as

$$J = \Re\left(\int_{-\infty}^{\infty} \frac{e^{i(x-\pi/4)}}{1+x^2} dx\right) \ , \tag{13.25}$$

where $\Re(\cdots)$ again denotes the real part. We want to evaluate this integral by closing this integration path with a semicircle either in the upper half-plane or in the lower half-plane. Due to the term $\exp{(ix)}$ in the integral we now have no real choice in this issue. The decision whether to close the integral in the upper half-plane or in the lower half-plane is dictated by the requirement that the integral over the semicircle vanishes as $R \to \infty$. This can only happen when the integral vanishes (faster than $1/R$) as $R \to \infty$. Let z be a point in the complex plane on the semicircle C_R that we use for closing the contour. On the semicircle z can be written as $z = R \exp i\varphi$. In the upper half-plane $0 \leq \varphi < \pi$ and for the lower half-plane $\pi \leq \varphi < 2\pi$.

Problem g: Use this representation of z to show that

$$\left| e^{iz} \right| = e^{-R\sin\varphi} \ . \tag{13.26}$$

Problem h: Show that in the limit $R \to \infty$ this term only goes to zero when z is in the upper half-plane.

This means that the integral over the semicircle only vanishes when we close the contour in the upper half-plane. Using steps similar as in (13.21) one can show that the integral over a semicircle in the upper half-plane vanishes as $R \to \infty$.

Problem i: Take exactly the same steps as in the derivation of (13.23) and show that

$$J = \int_{-\infty}^{\infty} \frac{\cos(x - \pi/4)}{1 + x^2} dx = \frac{\pi}{\sqrt{2}\, e} \ . \tag{13.27}$$

Problem j: Determine the integral $\int_{-\infty}^{\infty} \frac{\sin(x-\pi/4)}{1+x^2} dx$ without doing any additional calculations. Hint; look carefully at (13.25) and spot $\sin(x - \pi/4)$.

13.4 Response of a particle in syrup

Up to this point, contour integration has been applied to mathematical problems. However, this technique does have important application in physical problems. In this section we consider a particle with mass m on which a force $f(t)$ is acting. The particle is suspended in syrup, this damps the velocity of the particle and it is assumed that this damping force is proportional to the velocity $v(t)$ of the particle. The equation of motion of the particle is given by

$$m\frac{dv}{dt} + \beta v = f(t) \ . \tag{13.28}$$

where β is a parameter that determines the strength of the damping of the motion by the fluid. The question we want to solve is: what is the velocity $v(t)$ for a given force $f(t)$?

We will solve this problem by using a Fourier transform technique. The Fourier transform of $v(t)$ is denoted by $V(\omega)$. The velocity in the frequency domain is related to the velocity in the time domain by the relation:

$$v(t) = \int_{-\infty}^{\infty} V(\omega) e^{-i\omega t} d\omega \ , \tag{13.29}$$

and its inverse

$$V(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} v(t) e^{+i\omega t} dt \ . \tag{13.30}$$

You may be used to seeing a different sign in the exponents and by seeing the factor 2π in a different place, but as long as the exponents in the forward and backward Fourier transform have opposite sign and the product of the scale factors in the forward and backward Fourier transform is equal to $1/2\pi$ the different conventions are equally valid and lead to the same final result. This issue is treated in detail in section (11.5). The force $f(t)$ is Fourier transformed using the same expressions; in the frequency domain it is denoted by $F(\omega)$.

Problem a: Use the definitions of the Fourier transform to show that the equation of motion (13.28) is in the frequency domain given by

$$-i\omega m V(\omega) + \beta V(\omega) = F(\omega) \ . \tag{13.31}$$

Comparing this with the original equation (13.28) we can see immediately why the Fourier transform is so useful. The original expression (13.28) is a differential equation while expression (13.31) is an algebraic equation. Since algebraic equations are much easier to solve than differential equations we have made considerable progress.

Problem b: Solve the algebraic equation (13.31) for $V(\omega)$ and use the Fourier transform (13.29) to derive that

$$v(t) = \frac{i}{m} \int_{-\infty}^{\infty} \frac{F(\omega)e^{-i\omega t}}{\left(\omega + \frac{i\beta}{m}\right)} \, d\omega \; . \tag{13.32}$$

Now we have an explicit relation between the velocity $v(t)$ in the time-domain and the force $F(\omega)$ in the frequency domain. This is not quite what we want since we want to find the relation of the velocity with the force $f(t)$ in the time domain.

Problem c: Use the inverse Fourier transform (13.30) (but for the force) to show that

$$v(t) = \frac{i}{2\pi m} \int_{-\infty}^{\infty} f(t') \int_{-\infty}^{\infty} \frac{e^{-i\omega(t-t')}}{\left(\omega + \frac{i\beta}{m}\right)} \, d\omega \; dt' \; . \tag{13.33}$$

This equation looks messy, but we will simplify it by writing it as

$$v(t) = \frac{i}{m} \int_{-\infty}^{\infty} f(t') I(t-t') \; dt' \; , \tag{13.34}$$

with

$$I(t-t') = \int_{-\infty}^{\infty} \frac{e^{-i\omega(t-t')}}{\left(\omega + \frac{i\beta}{m}\right)} \, d\omega \; . \tag{13.35}$$

The problem we now face is to evaluate this integral. For this we will use complex integration. The integration variable is now called ω rather than z but this does not change the principles. We will close the contour by adding a semicircle either in the upper half-plane or in the lower half-plane to the integral (13.35) along the real axis. On a semicircle with radius R the complex number ω can be written as $\omega = R \exp i\varphi$.

Problem d: Show that $\left| e^{-i\omega(t-t')} \right| = \exp\left(-R \sin\varphi \; \times (t'-t)\right)$.

Problem e: The integral along the semicircle should vanish in the limit $R \to \infty$. Use the result of **problem d** to show that for $t < t'$ the contour should be closed in the upper half plane and for $t > t'$ in the lower half-plane, see figure (13.3).

Problem f: Show that the integrand in (13.35) has one pole at the negative imaginary axis at $\omega = -i\beta/m$ and that the residue at this pole is given by

$$Res = \exp\left(-\frac{\beta}{m}(t-t')\right) \; . \tag{13.36}$$

Problem g: Use these results and the theorems derived in section (13.2) that:

$$I(t-t') = \begin{cases} 0 & for \; t < t' \\ -i \exp\left(-\frac{\beta}{m}(t-t')\right) & for \; t > t' \end{cases} \tag{13.37}$$

Hint; treat the cases $t < t'$ and $t > t'$ separately.

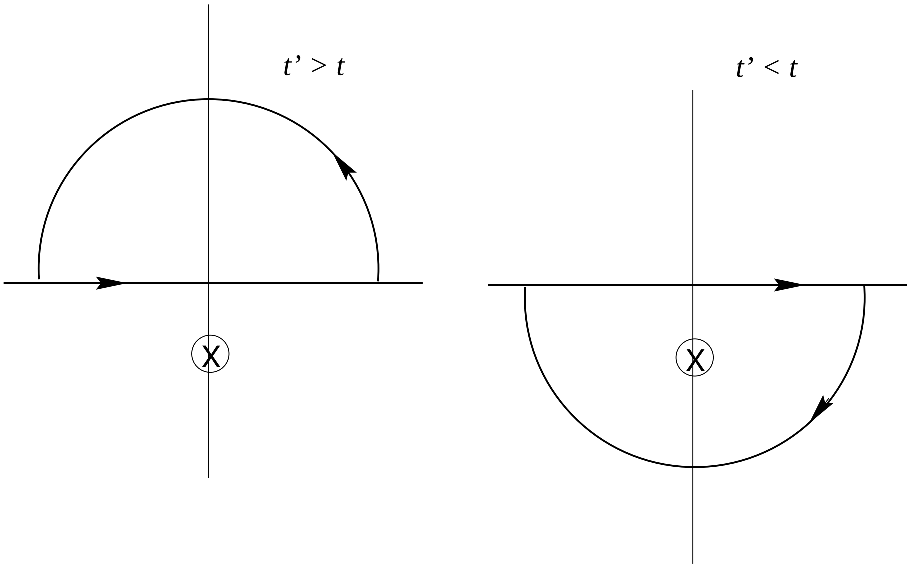


Figure 13.3: The poles in the complex plane and the closure of contours for t' larger than t (left) and t' smaller than t (right).

Let us first consider (13.37). One can see from (13.33) that $I(t - t')$ is a function that describes the effect of a force acting at time t' on the velocity at time t. Expression (13.37) tells us that this effect is zero when $t < t'$. In other words, this expression tells us that the force $f(t')$ has no effect on the velocity $v(t)$ when $t < t'$. This is equivalent to saying that the force only affects the velocity at *later* times. In this sense, equation (13.37) can be seen as an expression of *causality*; the cause $f(t')$ only influences the effect $v(t)$ for later times.

Problem h: Insert the equations (13.37) and (13.35) in (13.33) to show that

$$v(t) = \frac{1}{m} \int_{-\infty}^{t} \exp\left(- \frac{\beta}{m}(t - t') \right) f(t')dt' \ . \tag{13.38}$$

Pay in particular attention to the limits of integration.

Problem i: Convince yourself that this expression states that the force (the "cause") only has an influence on the velocity (the "effect") for *later* times.

You may be very happy that for this problem we have managed to give a proof of the causality principle. However, there is a problem hidden in the analysis. Suppose we switch off the damping parameter β, i.e. we remove the syrup from the problem. One can easily see that setting $\beta = 0$ in the final result (13.38) poses no problem. However, suppose that we would have set $\beta = 0$ at the start of the problem.

Problem j: Show that in that case the pole in figure (13.3) is located on the *real* axis rather than the negative imaginary axis.

This implies that it is not clear how this pole affects the response. In particular, it is not clear whether this pole gives a nonzero contribution for $t < t'$ (as it would when we consider it to lie in the upper half-plane) or for $t > t'$(as it would when we consider it to lie in the lower half-plane). This is a disconcerting result since it implies that causality only follows from the analysis when the problem contains some dissipation. This is not an artifact of the employed analysis using complex integration. What we encounter here is a manifestation of the problem that in the absence of dissipation the laws of physics are symmetric for time-reversal, whereas the world around us seems to move in one time direction only. This is the poorly resolved issue of the "arrow of time."

Chapter 14

Green's functions, principles

Green's function play a very important role in mathematical physics. The Green's function plays a similar role as the impulse response for linear filters that was treated in section (11.7). The general idea is that if one knows the response of a system to a delta-function input, the response of the system to any input can be reconstructed by superposing the response to the delta function input in an appropriate manner. However, the use of Green's functions suffers from the same limitation as the use of the impulse response for linear filters; since the superposition principle underlies the use of Green's functions they are only useful for systems that are linear. Excellent treatments of Green's functions can be found in the book of *Barton*[5] that is completely devoted to Green's functions and in *Butkov*[14]. Since the principles of Green's functions are so fundamental, the first section of this chapter is not presented as a set of problems.

14.1 The girl on a swing

In order to become familiar with Green's functions let us consider the example of a girl on a swing that is pushed by her mother, see figure (14.1). When the amplitude of the swing is not too large, the motion is of the swing is described by the equation of a harmonic oscillator that is driven by an external force $F(t)$ that is a general function of time:

$$\ddot{x} + \omega_0^2 x = F(t)/m \; . \tag{14.1}$$

The eigen-frequency of the oscillator is denoted by ω_0. It is not easy to solve this equation for a general driving force. For simplicity, we will solve equation (14.1) for the special case that the mother gives a single push to her daughter. The push is given at time $t = 0$ and has duration Δ and the magnitude of the force is denoted by F_0, this means that:

$$F(t) = \begin{cases} 0 & for \quad t < 0 \\ F_0 & for \quad 0 \le t < \Delta \\ 0 & for \quad \Delta \le t \end{cases} \tag{14.2}$$

We will look here for a *causal* solution. This is another way of saying that we are looking for a solution where the cause (the driving force) precedes the effect (the motion of the oscillator). This means that we require that the oscillator does not move before for $t < 0$. For $t \ge \Delta$ the driving force vanishes and the solution is given by a linear combination of

Figure 14.1: The girl on a swing.

$\cos(\omega_0 t)$ and $\sin(\omega_0 t)$. For $0 \le t < \Delta$ the solution can be found by making the substitution $x = x' + F_0/m\omega_o^2$.

When $t < 0$ and $t \ge \Delta$ the function $x(t)$ satisfies the differential equation $\ddot{x} + \omega_0^2 x = 0$. This differential equation has general solutions of the form $x(t) = A\cos(\omega_0 t) + B\sin(\omega_0 t)$. For $t < 0$ the displacement vanishes, hence the constants A and B vanish. This determines the solution for $t < 0$ and $t \ge \Delta$. For $0 \le t < \Delta$ the displacement satisfies the differential equation $\ddot{x} + \omega_0^2 x = F_0/m$. The solution can be found by writing $x(t) = F_0/m\omega_0^2 + y(t)$. The function $y(t)$ then satisfies the equation $\ddot{y} + \omega_0^2 y = 0$, which has the general solution $C\cos(\omega_0 t) + D\sin(\omega_0 t)$. This means that the general solution of the oscillator is given by

$$x(t) = \begin{cases} 0 & for \quad t < 0 \\ \frac{F_0}{m\omega_0^2} + C\cos(\omega_0 t) + D\sin(\omega_0 t) & for \quad 0 \le t < \Delta \\ A\cos(\omega_0 t) + B\sin(\omega_0 t) & for \quad \Delta \le t \end{cases}, \qquad (14.3)$$

where A, B, C and D are integration constants that are not yet known.

These integration constants follow from the requirement that the motion $x(t)$ of the oscillator is at all time continuous and that the velocity $\dot{x}(t)$ of the oscillator is at all time continuous. The last condition follows from the consideration that when the force is finite, the acceleration is finite and the velocity is therefore continuous. The requirement that the both $x(t)$ and $\dot{x}(t)$ are continuous at $t = 0$ and at $t = \Delta$ lead to the following equations:

$$\begin{array}{c} \frac{F_0}{m\omega_0^2} + C = 0 \\ \omega_0 D = 0 \\ \frac{F_0}{m\omega_0^2} + C\cos(\omega_0\Delta) + D\sin(\omega_0\Delta) = A\cos(\omega_0\Delta) + B\sin(\omega_0\Delta) \\ -C\sin(\omega_0\Delta) + D\cos(\omega_0\Delta) = -A\sin(\omega_0\Delta) + B\cos(\omega_0\Delta) \end{array} \tag{14.4}$$

These equations are four linear equations for the four unknown integration constants A, B, C and D. The two upper equations can be solved directly for the constants C and D to give the values $C = -\left(F_0/m\omega_0^2\right)$ and $D = 0$. These values for C and D can be inserted in the lower two equations. Solving these equations then for the constant A and B gives the values $A = -\left(F_0/m\omega_0^2\right)(1 - \cos(\omega_0\Delta))$ and $B = \left(F_0/m\omega_0^2\right)\sin(\omega_0\Delta)$. Inserting these values of the constants in (14.3) shows that the motion of the oscillator is given by:

$$x(t) = \begin{cases} 0 & for \quad t < 0 \\ \frac{F_0}{m\omega_0^2}\{1 - \cos(\omega_0 t)\} & for \quad 0 \le t < \Delta \\ \frac{F_0}{m\omega_0^2}\{\cos(\omega_0(t-\Delta)) - \cos(\omega_0 t)\} & for \quad \Delta \le t \end{cases} \tag{14.5}$$

This is the solution for a push with duration Δ delivered at time $t = 0$. Suppose now that the push is very short. When the duration of the push is much shorter than the period of the oscillator $\omega_0\Delta \ll 1$. In that case one can use a Taylor expansion in $\omega_0\Delta$ for the term $\cos(\omega_0(t-\Delta))$ in (14.5). This can be achieved by using that $\cos(\omega_0(t-\Delta)) = \cos(\omega_0 t)\cos(\omega_0\Delta) - \sin(\omega_0 t)\sin(\omega_0\Delta)$ and by using the Taylor expansions $\sin(x) = x - x^3/6 + O(x^5)$ and $\cos(x) = 1 - x^2/2 + O(x^4)$ for $\sin(\omega_0\Delta)$ and $\cos(\omega_0\Delta)$. Retaining term of order $(\omega_0\Delta)$ and ignoring terms of higher order in $(\omega_0\Delta)$ shows that for an impulsive push $(\omega_0\Delta \ll 1)$ the solution is given by:

$$x(t) = \begin{cases} 0 & for \quad t < 0 \\ \frac{F_0}{m\omega_0^2}(\omega_0\Delta)\sin(\omega_0 t) & for \quad t > \Delta \end{cases} \tag{14.6}$$

We will not bother anymore with the solution between $0 \le t < \Delta$ because in the limit $\Delta \to 0$ this interval is of vanishing duration.

At this point we have all the ingredients needed to determine the response of the oscillator for a general driving force $F(t)$. Suppose we divide the time-axis in intervals of duration Δ. In the i-th interval, the force is given by $F_i = F(t_i)$ where t_i is the time of the i-th interval. We know from expression (14.6) the response to a force of duration Δ at time $t = 0$. The response to a force F_i at time t_i follows by replacing F_0 by F_i and by replacing t by $t - t_i$. Making these replacements it this follows that the respond to a force F_i delivered over a time interval Δ at time t_i is given by:

$$x(t) = \begin{cases} 0 & for \quad t < t_i \\ \frac{1}{m\omega_0}\sin(\omega_0(t - t_i))\ F(t_i)\Delta & for \quad t > t_i \end{cases} \tag{14.7}$$

This is the response due to the force acting at time t_i only. To obtain the response to the full force $F(t)$ one should sum over the forces delivered at all the times t_i. In the language of the girl on the swing one would say that equation (14.6) gives the motion of the swing for a single impulsive push, and that expression (14.7) gives the response of the swing to a sequence of pushes given by the mother. Since the differential equation (14.1) is linear we can use the superposition principle that states that the response to the superposition of two pushes is the sum of the response to the individual pushes. (In the language of section 11.7 we would say that the swing is a linear system.) This means that when the swing receives a number of pushes at different times t_i the response can be written as the sum of the response to every individual push. With (14.7) this gives:

$$x(t) = \sum_{t_i<t} \frac{1}{m\omega_0} \sin\left(\omega_0\left(t-t_i\right)\right) \; F(t_i)\Delta \; . \tag{14.8}$$

Note that in (14.7) the response to a push at time t before the push at time t_i vanishes. For this reason one only needs to sum in (14.8) over the pushes at *earlier* times because the pushes at later times give a vanishing contribution. For this reason the summation extended to times $t \geq t_i$.
Suppose now that the swing is not given a finite number of impulse pushes but that the driving force is a continuous function. This case can be handled by taking the limit $\Delta \rightarrow 0$. The summation in (14.8) then needs to be replaced by an integration. This can naturally be achieved because the duration Δ is equal to the infinitesimal interval dt used in the integration. What we really are doing here is replacing the continuous function $F(t)$ by a function that is constant within every interval Δ at times t_i , see figure (14.2), and then taking the limit where the width of the intervals goes to zero $\Delta \rightarrow 0$. A similar treatment may be familiar to you from the theory of integration. When the limit $\Delta \rightarrow 0$ is taken

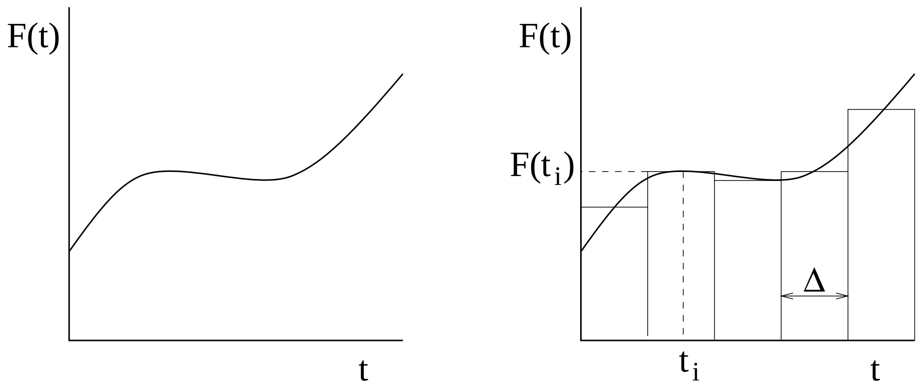


Figure 14.2: A continous function (left) and an approximation to this function that is constant within finite intervals (right).

the summation over t_i can be replaced by an integration: $\sum_{t_i} (\cdots)\Delta \rightarrow \int (\cdots)\, d\tau$. The integration variable τ plays the role of the summation variable t_i and the time-interval Δ is replaced $d\tau$. The response of the oscillator to a continuous force $F(t)$ is then given by

$$x(t) = \int_{-\infty}^{t} \frac{1}{m\omega_0} \sin\left(\omega_0\left(t-\tau\right)\right) F(\tau) d\tau \; . \tag{14.9}$$

The integration is only carried out over times $\tau < t$ because in the summation (14.8) extends only over the times $t_i < t$.

With a slight change in notation this result can be written as:

$$x(t) = \int_{-\infty}^{\infty} G(t,\tau)F(\tau)d\tau \ , \tag{14.10}$$

with

$$G(t,\tau) = \begin{cases} 0 & for \quad t < \tau \\ \frac{1}{m\omega_0}\sin\left(\omega_0\left(t-\tau\right)\right) & for \quad t > \tau \end{cases} \tag{14.11}$$

The function $G(t,\tau)$ in expression (14.11) is called the *Green's function* of the harmonic oscillator. Note that equation (14.10) is very similar to the response of a linear filter that was derived in equation (11.55) of section (11.7). This is not surprising, in both examples the response of a linear system to impulsive input was determined, it will be no surprise that the results are identical. In fact, the Green's function is defined as the response of a linear system to a delta-function input. Although Green's functions often are presented in a rather abstract way, one should remember that:

> *The Green's function of a system is nothing but the impulse response of a system, i.e. it is the response of the system to a delta-function excitation.*

14.2 You have seen Green's functions before!

Although the concept of a Green's function may appear to be new to you, you have already seen ample examples of Green's function, although the term "Green's function" might not have been used in that context. One example is the electric field generated by a point charge q in the origin that was used in section (4.1):

$$\mathbf{E}(\mathbf{r}) = \frac{q\hat{\mathbf{r}}}{4\pi\varepsilon_0 r^2} \qquad \text{(4.2) } again$$

Since this is the electric field generated by a delta-function charge in the origin, this field is very closely related to the Green's function for this problem. The field equation (4.12) of the electric field is invariant for translations in space. This is a complex way of saying that the electric field depends only on the *relative* position between the point charge and the point of observation.

Problem a: Show that this implies that the electric field at location $\mathbf{r}$ due to a point charge at location $\mathbf{r}'$ is given by:

$$\mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\varepsilon_0}\frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3} \tag{14.12}$$

Now suppose that we don't have a single point charge, but a system of point charges q_i at locations $\mathbf{r}_i$. Since the field equation is linear, the electric field generated by a sum of point charges is the sum of the fields generated by each point charge:

$$\mathbf{E}(\mathbf{r}) = \sum_i \frac{q_i}{4\pi\varepsilon_0}\frac{(\mathbf{r}-\mathbf{r}_i)}{|\mathbf{r}-\mathbf{r}_i|^3} \tag{14.13}$$

Problem b: To which expression of the previous section does this equation correspond?

Just as in the previous sections we now make the transition from a finite number of discrete inputs (either pushes of the swing or point charges) to an input function that is a continuous function (either the applied force to the oscillator as a function of time or a continuous electric charge). Let the electric charge per unit volume be denoted by $\rho(\mathbf{r})$, this means that the electric charge in a volume dV is given by $\rho(\mathbf{r})dV$.

Problem c: Replacing the sum in (14.13) by an integration over volume and using that the appropriate charge for each volume element dV show that the electric field for a continuous charge distribution is given by:

$$\mathbf{E}(\mathbf{r}) = \iiint \frac{\rho(\mathbf{r}')}{4\pi\varepsilon_0} \frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3} dV' \ , \tag{14.14}$$

where the volume integration is over $\mathbf{r}'$.

Problem d: Show that this implies that the electric field can be written as

$$\mathbf{E}(\mathbf{r}) = \iiint \mathbf{G}(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')dV' \ , \tag{14.15}$$

with the Green's function given by

$$\mathbf{G}(\mathbf{r},\mathbf{r}') = \frac{1}{4\pi\varepsilon_0} \frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3} \tag{14.16}$$

Note that this Green's function has the same form as the electric field for a point charge shown in (14.13).

Problem e: Show that the Green's function is only a function of the relative distance $\mathbf{r}-\mathbf{r}'$. Explain why the integral (14.15) can be seen as a three-dimensional convolution: $\mathbf{E}(\mathbf{r}) = \iiint \mathbf{G}(\mathbf{r}-\mathbf{r}')\rho(\mathbf{r}')dV'$.

The main purpose of this section was not to show you that you had seen an example of a Green's function before. Instead, it provides an example that the Green's function is not necessarily a function of time and that the Green's function is not necessarily a scalar function; the Green's function (14.16) depends only on the position and not on time and it describes a vector field rather than a scalar. The most important thing to remember is that the Green's function is the impulse response of a linear system.

Problem f: You have seen another Green's function before if you have worked through section (13.4) where the response of a particle in syrup was treated. Find the Green's function in that section and spot the equivalent expressions of the equations (14.10) and (14.11) of this section

14.3 The Green's function as impulse response

You may have found the derivation of the Green's function in section (14.1) rather complex. The reason for this is that in expression (14.3) the motion of the swing was determined before the push ($t < 0$), during the push ($0 < t < \Delta$) and after the push ($t > \Delta$). The requirement that the displacement x and the velocity $\dot{x}$ were continuous then led to the system of equations (14.4) with four unknowns. However, in the end we took the limit $\Delta \rightarrow 0$ and never used the solution for time $0 < t < \Delta$. This suggests that this method of solution is unnecessarily complicated. This is indeed the case. In this section an alternative derivation of the Green's function (14.11) is given that is based directly on the idea that the Green's function $G(t,\tau)$ describes the motion of the oscillator due to a delta-function force at time τ:

$$\ddot{G}(t,\tau) + \omega_0^2 G(t,\tau) = \frac{1}{m}\delta(t-\tau) \; . \tag{14.17}$$

Problem a: For $t \neq \tau$ the delta function vanishes and the right hand side of this expression is equal to zero. We are looking for the *causal* Green's function, this is the solution where the cause (the force) precedes the effect (the motion of the oscillator). Show that these conditions imply that for $t \neq \tau$ the Green's function is given by:

$$G(t,\tau) = \begin{cases} 0 & for \quad t < \tau \\ A\cos\left(\omega_0\left(t-\tau\right)\right) \; + B\sin\left(\omega_0\left(t-\tau\right)\right) & for \quad t > \tau \end{cases} \tag{14.18}$$

where A and B are unknown integration constants.

The integration constants follow from conditions at $t = \tau$. Since we have two unknown parameters we need to impose two conditions. The first conditions is that the motion of the oscillator is continuous at $t = \tau$. If this would not be the case the velocity of the oscillator would be infinite at that moment.

Problem b: Show that the requirement of continuity of the Green's function at $t = \tau$ implies that $A = 0$.

The second condition requires more care. We will derive the second condition first mathematically and then explore the physical meaning. The second condition follows by integrating expression (14.17) over t from $\tau - \varepsilon$ to $\tau + \varepsilon$ and by taking the limit $\varepsilon \downarrow 0$. Integrating (14.17) in this way gives:

$$\int_{\tau-\varepsilon}^{\tau+\varepsilon} \ddot{G}(t,\tau)dt + \omega_0^2 \int_{\tau-\varepsilon}^{\tau+\varepsilon} G(t,\tau)dt = \frac{1}{m}\int_{\tau-\varepsilon}^{\tau+\varepsilon} \delta(t-\tau)dt \; . \tag{14.19}$$

Problem c: Show that the right-hand side is equal to $1/m$, regardless of the value of ε.

Problem d: Show that the absolute value of the middle term is smaller than $2\varepsilon\omega_0^2 \max(G)$, where $\max(G)$ is the maximum of G over the integration interval. Since the Green's function is finite this means that the middle term vanishes in the limit $\varepsilon \downarrow 0$.

Problem e: Show that the left term in (14.19) is equal to $\dot{G}(t = \tau+\varepsilon, \tau) - \dot{G}(t = \tau-\varepsilon, \tau)$. This quantity will be denoted by $\left[\dot{G}(t,\tau)\right]_{t=\tau-\varepsilon}^{t=\tau+\varepsilon}$.

Problem f: Show that in the limit $\varepsilon \downarrow 0$ equation (14.19) gives:

$$\left[\dot{G}(t,\tau)\right]_{t=\tau-\varepsilon}^{t=\tau+\varepsilon} = \frac{1}{m} \, . \tag{14.20}$$

Problem g: Show that this condition together with the continuity of G implies that the integration constants in expression (14.18) have the values $A = 0$ and $B = 1/m\omega_0$, i.e. that the Green's function is given by:

$$G(t,\tau) = \begin{cases} 0 & for \quad t < \tau \\ \frac{1}{m\omega_0} \sin\left(\omega_0 \left(t - \tau\right)\right) & for \quad t > \tau \end{cases} \tag{14.21}$$

A comparison with (14.11) shows that the Green's function derived in this section is identical to the Green's function derived in section (14.1). Note that the solution was obtained here without invoking the motion of the oscillator during the moment of excitation. This also would have been very difficult because the duration of the excitation (a delta function) is equal to zero, if it can be defined at all.

There is however something strange about the derivation in this section. In section (14.1) the solution was found by requiring that the displacement x and its first derivative $\dot{x}$ were continuous at all times. As used in **problem b** the first condition is also met by the solution (14.21). However, the derivative $\dot{G}$ is not continuous at $t = \tau$.

Problem h: Which of the equations that you derived above states that the first derivative is not continuous?

Problem i: $G(t,\tau)$ denotes the displacement of the oscillator. Show that expression (14.20) states that the velocity of the oscillator is changes discontinuously at $t = \tau$.

Problem j: Give a physical reason why the velocity of the oscillator was continuous in the first part of section (14.1) and why the velocity is discontinuous for the Green's function derived in this section. Hint: how large is the force needed produce a finite jump in the velocity of a particle when the force is applied over a time interval of length zero (the width of the delta-function excitation).

How can we reconcile this result with the solution obtained in section (14.1)?

Problem k: Show that the change in the velocity in the solution $x(t)$ in equation (14.7) is proportional to $F(t_i)\Delta$, i.e. that

$$[\dot{x}]_{t_i-\varepsilon}^{t_i+\varepsilon} = \frac{1}{m} F(t_i)\Delta \tag{14.22}$$

This means that the change in the velocity depends on the strength of the force times the duration of the force. The physical reason for this is that the change in the velocity depends on the integral of the force over time divided by the mass of the particle.

Problem l: Derive this last statement also directly from Newton's law ($F = ma$).

When the force is finite and when $\Delta \to 0$, the jump in the velocity is zero and the velocity is continuous. However, when the force is infinite (as is the case for a delta function), the jump in the velocity is nonzero and the velocity is discontinuous.

In many applications the Green's function is the solution of a differential equation with a delta function as excitation. This implies that some derivative, or combination of derivatives, of the Green's function are equal to a delta function at the point (or time) of excitation. This usually has the effect that the Green's function or its derivative are not continuous functions. The delta function in the differential equation usually leads to a singularity in the Green's function or its derivative.

14.4 The Green's function for a general problem

In this section, the theory of Green's functions are treated in a more abstract fashion. Every linear differential equation for a function u with a source term F can symbolically be written as:

$$Lu = F \; . \tag{14.23}$$

For example in equation (14.1) for the girl on the swing, u is the displacement $x(t)$ while L is a differential operator given by

$$L = m\frac{d^2}{dt^2} + m\omega_0^2 \; , \tag{14.24}$$

where is understood that a differential operator acts term by term on the function to the right of the operator.

Problem a: Find the differential operator L and the source term F for the electrical field treated in section (14.2) from the field equation (4.12).

In the notation used in this section, the Green's function depends on the position vector $\mathbf{r}$, but the results derived here are equally valid for a Green's function that depends only on time or on position and time. In general, the differential equation (14.23) must be supplemented with boundary conditions to give an unique solution. In this section the position of the boundary is denoted by $\mathbf{r}_B$ and it is assumed that the function u has the value u_B at the boundary:

$$u(\mathbf{r}_B) = u_B \; . \tag{14.25}$$

Let us first find a single solution to the differential equation without bothering about boundary conditions. We will follow the same treatment as in section (11.7) where in equation (11.54) the input of a linear function was written as a superposition of delta functions. In the same way, the source function can be written as:

$$F(\mathbf{r}) = \int \delta(\mathbf{r} - \mathbf{r}')F(\mathbf{r}')dV' \; . \tag{14.26}$$

This expression follows from the properties of the delta function. One can interpret this expression as an expansion of the function $F(\mathbf{r})$ in delta functions because the integral (14.26) describes a superposition of delta functions $\delta(\mathbf{r} - \mathbf{r}')$ centered at $\mathbf{r} = \mathbf{r}'$; each of these delta functions is given a weight $F(\mathbf{r}')$. We want to use a Green's function to construct a solution. The Green's function $G(\mathbf{r},\mathbf{r}')$ is the response at location $\mathbf{r}$ due to a delta function source at location $\mathbf{r}'$, i.e. the Green's function satisfies:

$$LG(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}') \ . \tag{14.27}$$

The response to the input $\delta(\mathbf{r}-\mathbf{r}')$ is given by $G(\mathbf{r},\mathbf{r}')$, and the source functions can be written as a superposition of these delta functions with weight $F(\mathbf{r}')$. This suggest that a solution of the problem (14.23) is given by a superposition of Green's functions $G(\mathbf{r},\mathbf{r}')$ where each Green's function has the same weight factor as the delta function $\delta(\mathbf{r}-\mathbf{r}')$ in the expansion (14.26) of $F(\mathbf{r})$ in delta functions. This means that the solution of (14.23) is given by:

$$u_P(\mathbf{r}) = \int G(\mathbf{r},\mathbf{r}')F(\mathbf{r}')dV' \ . \tag{14.28}$$

Problem b: In case you worked through section (11.7) discuss the relation between this expression and equation (11.55) for the output of a linear function.

It is crucial to understand at this point that we have used three steps to arrive at (14.28): (i) The source function is written as a superposition of delta functions, (ii) the response of the system to each delta function input is defined and (iii) the solution is written as the same superposition of Green's function as was used in the expansion of the source function in delta functions:

$$\begin{array}{ccc} \delta(\mathbf{r}-\mathbf{r}') & \leftrightarrow & F(\mathbf{r}) = \int \delta(\mathbf{r}-\mathbf{r}')F(\mathbf{r}')dV' \\ \Downarrow & & \Downarrow \\ G(\mathbf{r},\mathbf{r}') & \leftrightarrow & u_P(\mathbf{r}) = \int G(\mathbf{r},\mathbf{r}')F(\mathbf{r}')dV' \end{array} \tag{14.29}$$

Problem c: Although this reasoning may sound plausible, we have not proven that $u_P(\mathbf{r})$ in equation (14.28) actually is a solution of the differential equation (14.23). Give a proof that this is indeed the case by letting the operator L act on (14.28) and by using equation (14.27) for the Green's function. Hint: the operator L acts on $\mathbf{r}$ while the integration is over $\mathbf{r}'$, the operator can thus be taken inside the integral.

It should be noted that we have not solved our problem yet, because u_P does not necessarily satisfy the boundary conditions. In fact, the solution (14.28) is just one of the many possible solutions to the problem (14.23). It is a particular solution of the inhomogeneous equation (14.23), and this is the reason why the subscript P is used. Equation (14.23) is called an *inhomogeneous equation* because the right-hand-side is nonzero. If the right-hand-side is zero one speaks of the *homogeneous equation*. This implies that a solution u_0 of the homogenous equation satisfies

$$Lu_0 = 0 \ . \tag{14.30}$$

Problem d: In general one can add a solution of the homogeneous equation (14.30) to a particular solution, and the result still satisfies the inhomogeneous equation (14.23). Give a proof of this statement by showing that the function $u = u_P + u_0$ is a solution of (14.23). In other words show that the general solution of (14.23) is given by:

$$u(\mathbf{r}) = u_0(\mathbf{r}) + \int G(\mathbf{r},\mathbf{r}')F(\mathbf{r}')dV' \ . \tag{14.31}$$

Problem e: The problem is that we still need to enforce the boundary conditions (14.25). This can be achieved by requiring that the solution u_0 satisfies specific boundary conditions at $\mathbf{r}_B$. Insert (14.31) in the boundary conditions (14.25) and show that the required solution u_0 of the homogeneous equation must satisfy the following boundary conditions:

$$u_0(\mathbf{r}_B) = u_B(\mathbf{r}_B) - \int G(\mathbf{r}_B, \mathbf{r}')F(\mathbf{r}')dV' \ . \tag{14.32}$$

This is all we need to solve the problem. What we have shown is that:

the total solution (14.31) is given by the sum of the particular solution (14.28) plus a solution of the homogeneous equation (14.30) that satisfies the boundary condition (14.32).

This construction may appear to be very complex to you. However, you should realize that the main complexity is the treatment of the boundary condition. In many problems, the boundary condition dictates that the function vanishes at the boundary ($u_B = 0$) and the Green's function also vanishes at the boundary. It follows from (14.31) that in that case the boundary condition for the homogeneous solution is $u_0(\mathbf{r}_B) = 0$. This boundary condition is satisfied by the solution $u_0(\mathbf{r}) = 0$ which implies that one can dispense with the addition of u_0 to the particular solution $u_P(\mathbf{r})$.

Problem f: Suppose that the boundary conditions do not prescribe the value of the solution at the boundary but that instead of (14.25) the normal derivative of the solution is prescribed by the boundary conditions:

$$\frac{\partial u}{\partial n}(\mathbf{r}_B) = \hat{\mathbf{n}} \cdot \boldsymbol{\nabla} u(\mathbf{r}_B) = w_B \ , \tag{14.33}$$

where $\hat{\mathbf{n}}$ is the unit vector perpendicular to the boundary. How should the theory of this section be modified to accommodate this boundary condition?

The theory of this section is rather abstract. In order to make the issues at stake more explicit the theory is applied in the next section to the calculation of the temperature in the earth.

14.5 Radiogenic heating and the earth's temperature

As an application of the use of Green's function we consider in this section the calculation of the temperature in the earth and specifically the effect of the decay of radioactive elements in the crust on the temperature in the earth. Several radioactive elements such as U_{235} do not fit well in the lattice of mantle rocks. For this reason, these elements are expelled from material in the earth's mantle and they accumulate in the crust. Radioactive decay of these elements then leads to a production of heat at the place where these elements accumulate.

As a simplified example of this problem we assume that the temperature T and the radiogenic heating Q depend only on depth and that we can ignore the sphericity of the earth. In addition, we assume that the radiogenic heating does not depend on time and that we consider only the equilibrium temperature.

Problem a: Show that these assumptions imply that the temperature is only a function of the z-coordinate: $T = T(z)$.

The temperature field satisfies the heat equation derived in section (8.4):

$$\frac{\partial T}{\partial t} = \kappa \nabla^2 T + Q \qquad \text{(8.27) } again$$

Problem b: Use this expression to show that for the problem of this section the temperature field satisfies

$$\frac{d^2 T}{dz^2} = -\frac{Q(z)}{\kappa} . \tag{14.34}$$

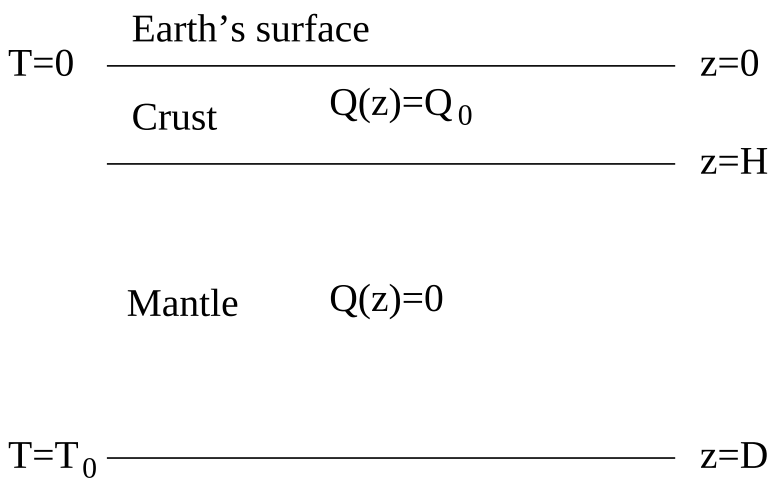


Figure 14.3: Definition of the geometric variables and boundary conditions for the temperature in the earth.

This equation can be solved when the boundary conditions are specified. The thickness of the crust is denoted by H, see figure (14.3). The temperature is assumed to vanish at the earth's surface. In addition, it is assumed that at a fixed depth D the temperature has a fixed value T_0. This implies that the boundary conditions are:

$$T(z = 0) = 0 \qquad , \qquad T(Z = D) = T_0 . \tag{14.35}$$

In this section we will solve the differential equation (14.34) with the boundary conditions (14.35) using the Green's function technique described in the previous section. Analogously to expression (14.28) we will first determine a particular solution T_P of the differential equation (14.34) and worry about the boundary conditions later. The Green's function $G(z, z')$ to be used is the temperature at depth z due to delta function heating at depth z':

$$\frac{d^2 G(z, z')}{dz^2} = \delta(z - z') . \tag{14.36}$$

Problem c: Use the theory of the previous section that the following function satisfies the heat equation (14.34):

$$T_P(z) = -\frac{1}{\kappa} \int_0^D G(z, z') Q(z') dz' . \tag{14.37}$$

Before further progress can be made it is necessary to find the Green's function, i.e. to solve the differential equation (14.36). In order to do this the boundary conditions for the Green's function need to be specified. In this example we will use a Green's function that vanishes at the endpoints of the depth interval:

$$G(z=0,z') = G(Z=D,z') = 0 \; . \tag{14.38}$$

Problem d: Use (14.36) to show that for $z \neq z'$ the Green's function satisfies the differential equation $d^2G(z,z')/dz^2 = 0$ and use this to show that the Green's function that satisfies the boundary conditions (14.38) must be of the form

$$G(z,z') = \begin{cases} \beta z & for\ z < z' \\ \gamma (z-D) & for\ z > z' \end{cases} \tag{14.39}$$

with β and γ constants that need to be determined.

Problem e: Since there are two unknown constants, two conditions are needed. The first condition is that the Green's function is continuous for $z = z'$. Use the theory of section (14.3) and the differential equation (14.36) to show that the second requirement is:

$$\lim_{\varepsilon \downarrow 0} \left[\frac{dG(z,z')}{dz} \right]_{z=z'-\varepsilon}^{z=z'+\varepsilon} = 1 \; , \tag{14.40}$$

i.e. that the first derivative makes a unit jump at the point of excitation.

Problem f: Apply these two conditions to the solution (14.39) to determine the constants β and γ and show that the Green's function is given by:

$$G(z,z') = \begin{cases} -\left(\frac{D-z'}{D}\right) z & for\ z < z' \\ -\frac{z'}{D}(D-z) & for\ z > z' \end{cases} \tag{14.41}$$

In this notation the two regions $z < z'$ and $z > z'$ are separated. Note, however, that the solution in the two regions has a highly symmetric form. In the literature you will find that a solution such as (14.41) is often rewritten by defining $z_>$ to be the maximum of z and z' and $z_<$ to be the minimum of z and z':

$$\begin{aligned} z_> &\equiv \max(z,z') \\ z_< &\equiv \min(z,z') \end{aligned} \tag{14.42}$$

Problem g: Show that in this notation the Green's function (14.41) can be written as:

$$G(z,z') = -\frac{D-z_>}{D} z_< \; . \tag{14.43}$$

As a particular heating function we will assume that the heating Q is only nonzero in the crust. This is a first-order description of the radiogenic heating in the shallow layers in the earth. The reason for this is that many of the radiogenic elements such as U_{235} fit much better in the crystal lattice of crustal material than in mantle material. For simplicity we will assume that the radiogenic heating is constant in the crust:

$$Q(z) = \begin{cases} Q_0 & for\ 0 < z < H \\ 0 & for\ H < z < D \end{cases} \tag{14.44}$$

Problem h: Show that the particular solution (14.37) for this heating function is given by

$$T_P(z) = \begin{cases} \frac{Q_0 H^2}{2\kappa}\left\{-\left(\frac{z}{H}\right)^2 + \frac{z}{H}\left(2 - \frac{H}{D}\right)\right\} & for\ 0 < z < H \\ \frac{Q_0 H^2}{2\kappa}\left(1 - \frac{z}{D}\right) & for\ H < z < D \end{cases} \tag{14.45}$$

Problem i: Show that this particular solution satisfies the boundary conditions

$$T_P(z=0) = T_P(z=D) = 0\ . \tag{14.46}$$

Problem j: This means that this solution does not satisfy the boundary conditions (14.35) of our problem. Use the theory of section (14.4) to derive that to this particular solution we must add a solution T_0 of the homogenous equation $d^2T_0/dz^2 = 0$ that satisfies the boundary conditions $T_0(z=0) = 0$ and $T_0(z=D) = T_0$.

Problem k: Show that the solution to this equation is given by $T_0(z) = T_0 z/D$ and that the final solution is given by

$$T(z) = \begin{cases} T_0\frac{z}{D} + \frac{Q_0 H^2}{2\kappa}\left\{-\left(\frac{z}{H}\right)^2 + \frac{z}{H}\left(2 - \frac{H}{D}\right)\right\} & for\ 0 < z < H \\ T_0\frac{z}{D} + \frac{Q_0 H^2}{2\kappa}\left(1 - \frac{z}{D}\right) & for\ H < z < D \end{cases} \tag{14.47}$$

Problem l: Verify explicitly that this solution satisfies the differential equation (14.34) with the boundary conditions (14.35).

As shown in expression (8.25) of section (8.4) the conductive heat-flow is given by $\mathbf{J} = -\kappa\nabla\mathrm{T}$. Since the problem is one-dimensional the heat flow is given by

$$J = -\kappa\frac{dT}{dz}\ . \tag{14.48}$$

Problem m: Compute the heat-flow at the top of the layer ($z = 0$) and at the bottom ($z = D$). Assuming that T_0 and Q_0 are both positive, does the heat-flow at these locations increase or decrease because of the radiogenic heating Q_0? Give a physical interpretation of this result. Use this result also to explain why people who often feel cold like to use an electric blanket while sleeping.

The derivation of this section used a Green's function that satisfied the boundary conditions (14.38) rather than the boundary conditions (14.35) of the temperature field. However, there is no particular reason why one should use these boundary conditions. To wit, one might think one could avoid the step of adding a solution $T_0(z)$ of the homogeneous equation by using a Green's function $\tilde{G}$ that satisfies the differential equation (14.39) and the same boundary conditions as the temperature field:

$$\tilde{G}(z=0,z') = 0 \quad , \quad \tilde{G}(z=D,z') = T_0\ . \tag{14.49}$$

Problem n: Go through the same steps as you did earlier in this section by constructing the Green's function $\tilde{G}(z,z')$, computing the corresponding particular solution $\tilde{T}_P(z)$, verifying whether the boundary conditions (14.35) are satisfied by this particular solution and if necessary adding a solution of the homogeneous equation in order to satisfy the boundary conditions. Show that this again leads to the solution (14.47).

Problem o: If you carried out the previous problem you will have discovered that the trick to use a Green's function that satisfied the boundary condition at $z = D$ did not lead to a particular solution that satisfied the same boundary condition at that point. Why did that trick not work?

The lesson from the last problems is that usually one needs to add to solution of the homogeneous equation to a particular solution in order to satisfy the boundary conditions. However, suppose that the boundary conditions of the temperature field would be homogeneous as well $(T = (z = 0) = T(z = D) = 0)$. In that case the particular solution (14.45) that was constructed using a Green's function that satisfies the same homogeneous boundary conditions (14.38) satisfies the boundary conditions of the full problem. This implies that it only pays off to use a Green's function that satisfies the boundary conditions of the full problem when these boundary conditions are *homogeneous*, i.e. when the function itself vanishes $(T = 0)$ or when the normal gradient of the function vanishes $(\partial T/\partial n = 0)$ or when a linear combination of these quantities vanishes $(aT + b\partial T/\partial n = 0)$. In all other cases one cannot avoid adding a solution of the homogeneous equation in order to satisfy the boundary conditions and the most efficient procedure is usually to use the Green's function that can most easily be computed.

14.6 Nonlinear systems and Green's functions

Up to this point, Green's function were applied to linear systems. The definition of a linear system was introduced in section (11.7). Suppose that a forcing F_1 leads to a response x_1 and that a forcing F_2 leads to a response x_2. A system is linear when the response to the linear combination $c_1F_1 + c_2F_2$ (with c_1 and c_2 constants) leads to the response $c_1x_1 + c_2x_2$.

Problem a: Show that this definition implies that the response to the input times a constant is given by the response that is multiplied by the same constant. In other words show that for a linear system an input that is twice as large leads to a response that is twice as large.

Problem b: Show that the definition of linearity given above implies that the response to the sum of two forcing functions is the sum of the responses to the individual forcing functions.

This last property reflects that a linear system satisfies the *superposition principle* which states that for a linear system one can superpose the response to a sum of forcing functions.

Not every system is linear, and we will exploit here to what extent Green's functions are useful for nonlinear systems. As an example we will consider the Verhulst equation:

$$\dot{x} = x - x^2 + F(t) \; . \tag{14.50}$$

This equation has been used in mathematical biology to describe the growth of a population. Suppose that only the term x was present in the right hand side. In that case the solution would be given by $x(t) = C \exp(t)$. This means that the first term on the right hand side accounts for the exponential population growth that is due to the fact that the number of offspring is proportional to the size of the population. However, a population cannot grow indefinitely, when a population is too large limited resources restrict the growth, this is accounted for by the $-x^2$ term in the right hand side. The term $F(t)$ accounts for external influences on the population. For example, a mass-extinction could be described by a strongly negative forcing function $F(t)$. We will consider first the solution for the case that $F(t) = 0$. Since the population size is positive we consider only positive solutions $x(t)$.

Problem c: Show that for the case $F(t) = 0$ the change of variable $y = 1/x$ leads to the linear equation $\dot{y} = 1 - y$. Solve this equation and show that the general solution of (14.50) (with $F(t) = 0$) is given by:

$$x(t) = \frac{1}{Ae^{-t} + 1} , \tag{14.51}$$

with A an integration constant.

Problem d: Use this solution to show that any solution of the unforced equation goes to 1 for infinite times:

$$\lim_{t \to \infty} x(t) = 1 . \tag{14.52}$$

In other words, the population of the unforced Verhulst equation always converges to the same population size. Note that when the forcing vanishes after a finite time, the solution after that time must satisfy (14.51) which implies that the long-time limit is then also given by (14.52).

Now, consider the response to a delta function excitation at time t_0 with strength F_0. The associated response $g(t, t_0)$ thus satisfies

$$\dot{g} - g + g^2 = F_0 \ \delta(t - t_0) . \tag{14.53}$$

Since this function is the impulse response of the system the notation g is used in order to bring out the resemblance with the Green's functions used earlier. We will consider only causal solution, i.e. we require that $g(t, t_0)$ vanishes for $t < t_0$: $g(t, t_0) = 0$ for $t < t_0$. For $t > t_0$ the solution satisfies the Verhulst equation without forcing, hence the general form is given by (14.51). The only remaining task is to find the integration constant A. This constant follows by a treatment similar to the analysis of section (14.3).

Problem e: Integrate (14.53) over t from $t_0 - \varepsilon$ to $t_0 + \varepsilon$, take the limit $\varepsilon \downarrow 0$ and show that this leads to the following requirement for the discontinuity in g:

$$\lim_{\varepsilon \downarrow 0} \left[g(t, t_0) \right]_{t_0 - \varepsilon}^{t_0 + \varepsilon} = F_0 . \tag{14.54}$$

Problem f: Use this condition to show that the constant A in the solution (14.51) is given by $A = \left(\frac{1}{F_0} - 1 \right) \exp t_0$ and that the solution is given by:

$$g(t, t_0) = \begin{cases} 0 & for \quad t < t_0 \\ \frac{F_0}{(1-F_0)e^{-(t-t_0)} + F_0} & for \quad t > t_0 \end{cases} \tag{14.55}$$

At this point you should be suspicious for interpreting $g(t, t_0)$ as a Green's function. An important property of linear systems is that the response is proportional to the forcing. However, the solution $g(t, t_0)$ in (14.55) is not proportional to the strength F_0 of the forcing.

Let us now check if we can use the superposition principle. Suppose the forcing function is the superposition of a delta-function forcing F_1 at $t = t_1$ and a delta-function forcing F_2 at $t = t_2$:

$$F(t) = F_1 \delta(t - t_1) + F_2 \delta(t - t_2) \ . \tag{14.56}$$

By analogy with expression (14.10) you might think that a Green's function-type solution is given by:

$$x^{Green}(t) = \frac{F_1}{(1 - F_1)e^{-(t-t_1)} + F_1} + \frac{F_2}{(1 - F_2)e^{-(t-t_2)} + F_2} \ , \tag{14.57}$$

for times larger than both t_1 and t_2. You could verify by direct substitution that this function is not a solution of the differential equation (14.50). However, this process is rather tedious and there is a simpler way to see that the function $x^{Green}(t)$ violates the differential equation (14.50).

Problem g: To see this, show that the solution $x^{Green}(t)$ has the following long-time behavior:

$$\lim_{t \to \infty} x^{Green}(t) = 2 \ . \tag{14.58}$$

This limit is at odds with the limit (14.52) that every solution of the differential equation (14.50) should satisfy when the forcing vanishes after a certain finite time. This proves that $x^{Green}(t)$ is not a solution of the Verhulst equation.

This implies that the Green's function technique introduced in the previous sections cannot be used for a nonlinear equation such as the forced Verhulst equation. The reason for this is that Green's function are based on the superposition principle; by knowing the response to a delta-function forcing and by writing a general forcing as a superposition of delta functions one can construct a solution by making the corresponding superposition of Green's functions, see (14.29). However, solutions of a nonlinear equation such as the Verhulst equation do not satisfy the principle of superposition. This implies that Green's function cannot be used effectively to construct behavior of nonlinear systems. It is for this reason that Green's function are in practice only used for constructing the response of linear systems.

Chapter 15

Green's functions, examples

In the previous section the basic theory of Green's function was introduced. In this chapter a number of examples of Green's functions are introduced that are often used in mathematical physics.

15.1 The heat equation in N-dimensions

In this section we consider once again the heat equation as introduced in section (8.4):

$$\frac{\partial T}{\partial t} = \kappa \nabla^2 T + Q \qquad (8.27)\ \textit{again}$$

In this section we will construct a Green's function for this equation in N space dimensions. The reason for this is that the analysis for N dimensions is just as easy (or difficult) as the analysis for only one spatial dimension.

The heat equation is invariant for translations in both space and time. For this reason the Green's function $G(\mathbf{r},t;\mathbf{r}_0,t_0)$ that gives the temperature at location $\mathbf{r}$ and time t to a delta-function heat source at location $\mathbf{r}_0$ and time t_0 depends only on the relative distance $\mathbf{r}-\mathbf{r}_0$ and the relative time $t-t_0$.

Problem a: Show that this implies that $G(\mathbf{r},t;\mathbf{r}_0,t_0) = G(\mathbf{r}-\mathbf{r}_0,t-t_0)$.

Since the Green's function depends only on $\mathbf{r}-\mathbf{r}_0$ and $t-t_0$ it suffices to construct the simplest solution by considering the special case of a source at $\mathbf{r}_0 = 0$ at time $t_0 = 0$. This means that we will construct the Green's function $G(\mathbf{r},t)$ that satisfies:

$$\frac{\partial G(\mathbf{r},t)}{\partial t} - \kappa \nabla^2 G(\mathbf{r},t) = \delta(r)\delta(t)\ . \tag{15.1}$$

This Green's function can most easily be constructed by carrying out a spatial Fourier transform. Using the Fourier transform (11.27) for each of the N spatial dimensions one finds that the Green's function has the following Fourier expansion:

$$G(\mathbf{r},t) = \frac{1}{(2\pi)^N}\int g(\mathbf{k},t)e^{i\mathbf{k}\cdot\mathbf{r}}d^N k\ . \tag{15.2}$$

Note that the Fourier transform is only carried out over the spatial dimensions but not over time. This implies that $g(\mathbf{k},t)$ is a function of time as well. The differential equation that g satisfies can be obtained by inserting the Fourier representation (15.2) in the differential equation (15.1). In doing this we also need the Fourier representation of $\nabla^2 G(\mathbf{r},t)$.

Problem b: Show by applying the Laplacian to the Fourier integral (15.1) that:

$$\nabla^2 G(\mathbf{r},t) = \frac{-1}{(2\pi)^N} \int k^2 g(\mathbf{k},t) e^{i\mathbf{k}\cdot\mathbf{r}} d^N k \; . \tag{15.3}$$

Problem c: As a last ingredient we need the Fourier representation of the delta function in the right hand side of (15.1). This multidimensional delta function is a shorthand notation for $\delta(\mathbf{r}) = \delta(x_1)\delta(x_2)\cdots\delta(x_N)$. Use the Fourier representation (11.31) of the delta function to show that:

$$\delta(\mathbf{r}) = \frac{1}{(2\pi)^N} \int e^{i\mathbf{k}\cdot\mathbf{r}} d^N k \; . \tag{15.4}$$

Problem d: Insert these results in the differential equation (15.1) of the Green's function to show that $g(\mathbf{k},t)$ satisfies the differential equation

$$\frac{\partial g(\mathbf{k},t)}{\partial t} + \kappa k^2 g(\mathbf{k},t) = \delta(t) \; . \tag{15.5}$$

We have made considerable progress. The original equation (15.1) was a partial differential equation, whereas equation (15.5) is an ordinary differential equation for g because only a time-derivative is taken. In fact, you have seen this equation before when you have read section (13.4) that dealt with the response of a particle in syrup. Equation (15.5) is equivalent to the equation of motion (13.28) for a particle in syrup when the forcing forcing is a delta function.

Problem e: Use the theory of section (14.3) to show that the causal solution of (15.5) is given by:

$$g(\mathbf{k},t) = \exp\left(-\kappa k^2 t\right) \; . \tag{15.6}$$

This solution can be inserted in the Fourier representation (15.2) of the Green's function, this gives:

$$G(\mathbf{r},t) = \frac{1}{(2\pi)^N} \int e^{-\kappa k^2 t + i\mathbf{k}\cdot\mathbf{r}} d^N k \; . \tag{15.7}$$

The Green's function can be found by solving this Fourier integral. Before we do this, let us pause and consider the solution (15.6) for the Green's function in the wave-number-time domain. The function $g(\mathbf{k},t)$ gives the coefficient of the plane wave component $\exp(i\mathbf{k}\cdot\mathbf{r})$ as a function of time. According to (15.6) each Fourier component decays exponentially in time with a characteristic decay time $1/\left(\kappa k^2\right)$.

Problem f: Show that this implies that in the Fourier expansion (15.2) plane waves with a smaller wavelength decay faster with time than plane waves with larger wavelength. Explain this result physically.

In order to find the Green's function, we need to solve the Fourier integral (15.7). The integrations over the different components k_i of the wave-number integration all have the same form.

Problem g: Show this by giving a proof that the Green's function can be written as:

$$G(\mathbf{r},t) = \frac{1}{(2\pi)^N}\left(\int e^{-\kappa k_1^2 t + ik_1 x_1} dk_1\right)\left(\int e^{-\kappa k_2^2 t + ik_2 x_2} dk_2\right)\cdots\left(\int e^{-\kappa k_N^2 t + ik_N x_N} dk_N\right) \tag{15.8}$$

You will notice that the each of the integrals is of the same form, hence the Green's function can be written as $G(x_1, x_2, \cdots, x_N, t) = I(x_1,t)I(x_2,t)\cdots I(x_N,t)$ with $I(x,t)$ given by

$$I(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\kappa k^2 t + ikx} dk \ . \tag{15.9}$$

This means that our problem is solved when the one-dimensional Fourier integral (15.9) is solved. In order to solve this integral it is important to realize that the exponent in the integral is a quadratic function of the integration variable k. If the integral would be of the form $\int_{-\infty}^{\infty} e^{-\alpha k^2} dk$ the problem would not be difficult because it known that this integral has the value $\sqrt{\pi/\alpha}$. The problem can be solved by rewriting the integral (15.9) in the form of the integral $\int_{-\infty}^{\infty} e^{-\alpha k^2} dk$.

Problem h: Complete the square of the exponent in (15.9), i.e. show that

$$-\kappa k^2 t + ikx = -\kappa t\left(k - \frac{ix}{2\kappa t}\right)^2 - \frac{x^2}{4\kappa t} \ , \tag{15.10}$$

and use this result to show that $I(x,t)$ can be written as:

$$I(x,t) = \frac{1}{2\pi} e^{-x^2/4\kappa t}\int_{-\infty - ix/2\kappa t}^{\infty - ix/2\kappa t} e^{-\kappa k'^2 t} dk' \ . \tag{15.11}$$

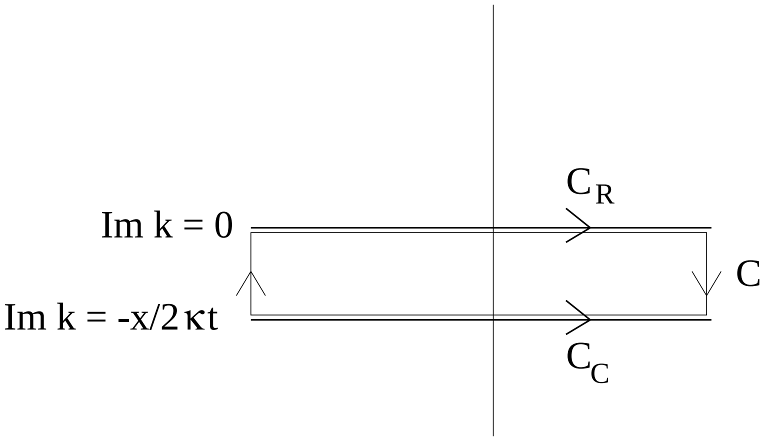


Figure 15.1: The contours C_R, C_C and C in the complex k-plane.

With these steps we have achieved our goal of having an integrand of the form $\exp\left(-\alpha k^2\right)$, but have paid a price. In the integral (15.9) the integration was along the real axis C_R, see figure (15.1). In the transformed integral the integration now takes place along the integration path C_C in the complex plane that lies below the real axis, see figure (15.1). However, one can show that when the integration path C_C is replaced by an integration along the real axis the integral has the same value:

$$I(x,t) = \frac{1}{2\pi} e^{-x^2/4\kappa t} \int_{-\infty}^{\infty} e^{-\kappa k^2 t} dk \; . \tag{15.12}$$

Problem i: When you have studied section (13.2) you have seen all the material to give a proof that (15.12) is indeed identical to (15.11). Show that this is indeed the case by using that the closed integral along the closed contour C in figure (15.1) vanishes.

Problem j: Carry out the integration in (15.12) and derive that

$$I(x,t) = \frac{e^{-x^2/4\kappa t}}{\sqrt{4\pi\kappa t}} \; , \tag{15.13}$$

and show that this implies that the Green's function is given by

$$G(\mathbf{r},t) = \frac{1}{(4\pi\kappa t)^{N/2}} \exp\left(-r^2/4\kappa t\right) \; . \tag{15.14}$$

Problem k: This result implies that the Green's function has in any dimension the form of the Gaussian. Show that this Gaussian changes shape with time. Is the Gaussian broadest at early times or at late times? What is the shape of the Green's function in the limit $t \downarrow 0$, i.e. at the time just after the heat forcing has been applied.

Problem l: Sketch the time-behavior of the Green's function for a fixed distance r. Does the Green's function decay more rapidly as a function of time in three dimensions than in one dimension? Give a physical interpretation of this result.

It is a remarkable property of the derivation in this section that the Green's function could be derived with a single derivation for every number of dimension. It should be noted that this is not the generic case. In many problems, the behavior of the system depends critically of the number of spatial dimensions. We will see in section 15.4 that wave propagation in two dimensions is fundamentally different from wave propagation in one or three dimensions. Another example is chaotic behavior of dynamical systems where the occurrence of chaos is intricately linked to the number of dimensions, see the discussion given by *Tabor*[59].

15.2 The Schrödinger equation with an impulsive source

In this section we will study the Green's function for the Schrödinger equation that was introduced in section (6.4):

$$-\frac{\hbar}{i}\frac{\partial\psi(\mathbf{r},t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r},t) + V(\mathbf{r})\psi(\mathbf{r},t) \qquad (6.13) \quad \textit{again}$$

Solving this equation for a general potential $V(\mathbf{r})$ is a formidable problem, and solutions are known for only very few examples such as the free particle, the harmonic oscillator and the Coulomb potential. We will restrict ourselves to the simplest case of a free particle, this is the case where the potential vanishes ($V(\mathbf{r}) = 0$). The corresponding Green's function satisfies the following partial differential equation:

$$\frac{\hbar}{i}\frac{\partial G(\mathbf{r},t)}{\partial t} - \frac{\hbar^2}{2m}\nabla^2 G(\mathbf{r},t) = \delta(\mathbf{r})\delta(t) \ . \tag{15.15}$$

Before we compute the Green's function for this problem, let us pause to find the meaning of this Green's function. First, the Green's function is for $\mathbf{r} \neq 0$ and $t \neq 0$ a solution of Schrödinger's equation. This means that $|G|^2$ gives the probability density of a particle (see also section 6.4). However, the right hand side of (15.15) contains a delta function forcing at time $t = 0$ at location $\mathbf{r} = 0$. This is a source term of G and hence this is a source of probability for the presence of the particle. One can say that this source term creates probability for having a particle in the origin at $t = 0$. Of course, this particle will not necessarily remain in the origin, it will move according to the laws of quantum mechanics. This motion is described by equation (15.15). This means that this equation describes the time evolution of matter waves when matter is injected at $t = 0$ at location $\mathbf{r} = 0$.

Problem a: The Green's function $G(\mathbf{r},t;\mathbf{r}',t')$ gives the wave-function at location $\mathbf{r}$ and time t for a source of particles at location $\mathbf{r}'$ at time t'. Express the Green's function $G(\mathbf{r},t;\mathbf{r}',t')$ in the solution $G(\mathbf{r},t)$ of (15.15), and show how you obtain this result. Is this result also valid for the Green's function for the quantum-mechanical harmonic oscillator (where the potential $V(\mathbf{r})$ depends on position)?

In the previous section the Green's function gave the evolution of the temperature field due to a delta function injection of heat in the origin at time $t = 0$. Similarly, the Green's function of this section describes the time-evolution of probability for a delta function injection of matter waves in the origin at time $t = 0$. These two Green's functions are not only conceptually very similar. The differential equations (15.1) for the temperature field and (15.15) for the Schrödinger equation are first order differential equations in time and second order differential equations in the space coordinate that have a delta-function excitation in the right hand side. In this section we will exploit this similarity and derive the Green's function for the Schrödinger's equation from the Green's function for the heat equation derived in the previous section rather than constructing the solution from first principles. This approach is admittedly not very rigorous, but it shows that analogies are useful for making shortcuts.

The principle difference between (15.1) and (15.15) is that the time-derivative for Schrödinger's equation is multiplied with $i = \sqrt{-1}$ whereas the heat equation is purely real. We will relate the two equation by introducing the new time variable τ for the Schrödinger equation that is proportional to the original time: $\tau = \gamma t$.

Problem b: How should the proportionality constant γ be chosen so that (15.15) transform to:

$$\frac{\partial G(\mathbf{r},\tau)}{\partial \tau} - \frac{\hbar^2}{2m}\nabla^2 G(\mathbf{r},\tau) = C\delta(\mathbf{r})\delta(\tau) \ . \tag{15.16}$$

The constant C in the right hand side cannot easily be determined from the change of variables $\tau = \gamma t$ because γ is not necessarily real and it is not clear how a delta function with a complex argument should be interpreted. For this reason we will bother to specify C.

The key point to note is that this equation is of exactly the same form as the heat equation (15.1), where $\hbar^2/2m$ plays the role of the heat conductivity κ. The only difference is the constant C in the right hand side of (15.16). However, since the equation is linear, this term only leads to an overall multiplication with C.

Problem c: Show that the Green's function for the Green's function can be obtained from the Green's function (15.14) for the heat equation by making the following substitutions:

$$\begin{array}{ccc} t & \longrightarrow & it/\hbar \\ \kappa & \longrightarrow & \hbar^2/2m \\ G & \longrightarrow & C\,G \end{array} \tag{15.17}$$

It is interesting to note that the "diffusion constant" κ that governs the spreading of the waves with time is proportional to the square of Planck's constant. Classical mechanics follows from quantum mechanics by letting Planck's constant go to zero: $\hbar \rightarrow 0$. It follows from (15.17) that in that limit the diffusion constant of the matter waves goes to zero. This reflects the fact that in classical mechanics the probability of the presence for a particle does not spread-out with time.

Problem d: Use the substitutions (15.17) to show that the Green's function for the Schrödinger equation in N-dimensions is given by:

$$G(\mathbf{r},t) = C\frac{1}{(2\pi i\hbar t/m)^{N/2}}\exp\left(imr^2/2\hbar t\right) . \tag{15.18}$$

This Green's function plays a crucial role in the formulation of the *Feynman path integrals* that have been a breakthrough both within quantum mechanics as well as in other fields. A very clear description of the Feynman path integrals is given by *Feynman and Hibbs*[22].

Problem e: Sketch the real part of the exponential $\exp\left(imr^2/2\hbar t\right)$ in the Green's function for a fixed time as a function of radius r. Does the wavelength of the Green's function increase or decrease with distance?

The Green's function (15.18) actually has an interesting physical meaning which is based on the fact that it describes the propagation of matter waves injected at $t = 0$ in the origin. The Green's function can be written as $G = C\left(2\pi i\hbar t/m\right)^{-N/2}\exp\left(i\Phi\right)$, where the phase of the Green's function is given by

$$\Phi = \frac{mr^2}{2\hbar t} . \tag{15.19}$$

As you noted in **problem e** the wave-number of the waves depends on position. For a plane wave $\exp(i\mathbf{k}\cdot\mathbf{r})$ the phase is given by $\Phi = (\mathbf{k}\cdot\mathbf{r})$ and the wave-number follows by taking the gradient of this function.

Problem f: Show that for a plane wave that

$$\mathbf{k} = \boldsymbol{\nabla}\Phi \ . \tag{15.20}$$

The relation (15.20) has a wider applicability than plane waves. It is shown by *Whitham*[66] that for a general phase function $\Phi(\mathbf{r})$ that varies smoothly with $\mathbf{r}$ the local wave-number $\mathbf{k}(\mathbf{r})$ is defined by (15.20).

Problem g: Use this to show that for the Green's function of the Schrödinger equation the local wave-number is given by

$$\mathbf{k} = \frac{m\mathbf{r}}{\hbar t} \ . \tag{15.21}$$

Problem h: Show that this expression is equivalent to expression (6.19) of section (6.4):

$$\mathbf{v} = \frac{\hbar \mathbf{k}}{m} \qquad (6.19) \quad \textit{again}$$

In **problem e** you discovered that for a fixed time, the wavelength of the waves decreases when the distance r to the source is increased. This is consistent with expression (6.19); when a particle has moved further away form the source in a fixed time, its velocity is larger. This corresponds according to (6.19) with a larger wave-number and hence with a smaller wavelength. This is indeed the behavior that is exhibited by the full wave function (15.18).

The analysis of this chapter was not very rigorous because the substitution $t \to (i/\hbar)\, t$ implies that the independent parameter is purely imaginary rather than real. This means that all the arguments used in the previous section for the complex integration should be carefully re-examined. However, a more rigorous analysis shows that (15.18) is indeed the correct Green's function for the Schrödinger equation[**?**]. However, the approach taken in this section shows that an educated guess can be very useful in deriving new results. One can in fact argue that many innovations in mathematical physics have been obtained using intuition or analogies rather than formal derivations. Of course, a formal derivation should ultimately substantiate the results obtained from a more intuitive approach.

15.3 The Helmholtz equation in 1,2,3 dimensions

The Helmholtz equation plays an important role in mathematical physics because it is closely related to the wave equation. A very complete analysis of the Green's function for the wave equation and the Helmholtz equation in different dimensions is given by *DeSanto*[**?**]. The Green's function for the wave equation for a medium with constant velocity c satisfies:

$$\nabla^2 G(\mathbf{r}, t; \mathbf{r}_0, t_0) - \frac{1}{c^2}\frac{\partial^2 G(\mathbf{r}, t; \mathbf{r}_0, t_0)}{\partial t^2} = \delta(\mathbf{r} - \mathbf{r}_0)\delta(t - t_0) \ . \tag{15.22}$$

As shown in section (15.1) the Green's function depends only on the relative location $\mathbf{r} - \mathbf{r}_0$ and the relative time $t - t_0$ so that without loss of generality we can take the source at the origin ($\mathbf{r}_0 = 0$) and let the source act at time $t_0 = 0$. In addition it follows from symmetry considerations that the Green's function depends only on the relative distance $|\mathbf{r} - \mathbf{r}_0|$ but not on the orientation of the vector $\mathbf{r} - \mathbf{r}_0$. This means that the Green's function then satisfies $G(\mathbf{r}, t; \mathbf{r}_0, t_0) = G(|\mathbf{r} - \mathbf{r}_0|, t - t_0)$ and we need to solve the following equation:

$$\nabla^2 G(r,t) - \frac{1}{c^2}\frac{\partial^2 G(r,t)}{\partial t^2} = \delta(\mathbf{r})\delta(t) \ . \tag{15.23}$$

Problem a: Under which conditions is this approach justified?

Problem b: Use a similar treatment as in section (15.1) to show that when the Fourier transform (11.43) is used the Green's function satisfies in the frequency domain the following equation:

$$\nabla^2 G(r,\omega) + k^2 G(r,\omega) = \delta(\mathbf{r}) \ , \tag{15.24}$$

where the wave number k satisfies $k = \omega/c$.

This equation is called the *Helmholtz equation*, it is the reformulation of the wave equation in the frequency domain. In the following we will suppress the factor ω *in*the Green's function but it should be remembered that the Green's function depends on frequency.

We will solve (15.24) for 1,2 and 3 space dimensions. To do this we will consider the case of N-dimensions and derive the Laplacian of a function $F(r)$ that depends only on the distance $r = \sqrt{\sum_{j=1}^{N} x_j^2}$. According to expression (4.19) $\partial r/\partial x_j = x_j/r$. This means that the derivative $\partial F/\partial x_j$ can be written as $\partial F/\partial x_j = (\partial r/\partial x_j)\,\partial F/\partial r = (x_j/r)\,\partial F/\partial r$.

Problem c: Use these results to show that

$$\frac{\partial^2 F(r)}{\partial x_j^2} = \frac{x_j^2}{r^2}\frac{\partial^2 F}{\partial r^2} + \left(\frac{1}{r} - \frac{x_j^2}{r^3}\right)\frac{\partial F}{\partial r} \tag{15.25}$$

and that the Laplacian $\sum_{j=1}^{N} \partial^2 F/\partial x_j^2$ is given by

$$\nabla^2 F(r) = \frac{\partial^2 F}{\partial r^2} + \frac{N-1}{r}\frac{\partial F}{\partial r} = \frac{1}{r^{N-1}}\frac{\partial}{\partial r}\left(r^{N-1}\frac{\partial F}{\partial r}\right) \ . \tag{15.26}$$

Using this expression the differential equation for the Green's function in N-dimension is given by

$$\frac{1}{r^{N-1}}\frac{\partial}{\partial r}\left(r^{N-1}\frac{\partial G}{\partial r}\right) + k^2 G(r,\omega) = \delta(\mathbf{r}) \ . \tag{15.27}$$

This differential equation is not difficult to solve for 1, 2 or 3 space dimensions for locations away from the source ($r \neq 0$). However, we need to consider carefully how the source $\delta(\mathbf{r})$ should be coupled to the solution of the differential equation. For the case of one dimension this can be achieved using the theory of section (14.3). The derivation of that section needs to be generalized to more space dimensions.

This can be achieved by integrating (15.24) over a sphere with radius R centered at the source and letting the radius go to zero.

Problem d: Integrate (15.24) over this volume, use Gauss' law and let the radius R go to zero to show that the Green's function satisfies

$$\oint_{S_R} \frac{\partial G}{\partial r} dS = 1 \ , \tag{15.28}$$

where the surface integral is over a sphere S_R with radius R in the limit $R \downarrow 0$. Show that this can also be written as

$$\lim_{r \downarrow 0} S_r \frac{\partial G}{\partial r} = 1 \ , \tag{15.29}$$

where S_r is the surface of a sphere in N dimensions with radius r.

Note that the surface of the sphere in general goes to zero as $r \downarrow 0$ (except in one dimension), this implies that $\partial G / \partial r$ must be infinite in the limit $r \downarrow 0$ in more than one space dimension.

The differential equation (15.27) is a second order differential equation. Such an equation must be supplemented with two boundary conditions. The first boundary condition is given by (15.29), this condition specifies how the solution is coupled to the source at $\mathbf{r} = 0$. The second boundary condition that we will use reflects the fact that the waves generated by the source will move away from the source. The solutions that we will find will behave for large distance as $\exp(\pm ikr)$, but it is not clear whether we should use the upper sign $(+)$ or the lower sign $(-)$.

Problem e: Use the Fourier transform (11.42) and the relation $k = \omega / c$ to show that the integrand in the Fourier transform to the time domain is proportional to $\exp\left(-i\omega(t \mp \frac{r}{c}\right)$. Show that the waves only move away from the source for the upper sign. This means that the second boundary condition dictates that the solution behave in the limit $r \to \infty$ as $\exp(+ikr)$.

The derivative of function $\exp(+ikr)$ is given by $ik \exp(+ikr)$, i.e. the derivative is ik times the original function. When the Green's function behaves for large r as $\exp(+ikr)$, then the derivative of the Green's must satisfy the same relation as the derivative of $\exp(+ikr)$. This means that the Green's function satisfies for large distance r:

$$\frac{\partial G}{\partial r} = ikG \ . \tag{15.30}$$

This relation specifies that the energy radiates *away* from the source. For this reason expression (15.30) is called the *radiation boundary condition.*

Now we are at the point where we can actually construct the solution for each dimension. Let us first determine the solution in one space dimension.

Problem f: Show that for one dimension ($N = 1$) the differential equation (15.27) has away from the source the general form $G = C \exp(\pm ikr)$, where r is the distance to the origin: $r = |x|$. Use the result of **problem e** to show that the plus sign should be used in the exponent and equation (15.28) to derive that the constant C is given by $C = -i/2k$. (Hint, what is the surface of a one-dimensional "volume"?) Show that this implies that the Green's function in one dimension is given by

$$G^{1D}(x) = \frac{-i}{2k} e^{ik|x|} \ . \tag{15.31}$$

Before we go to two dimensions we will first solve the Green's function in three dimensions.

Problem g: Make for three dimensions ($N = 3$) the substitution $G(r) = f(r)/r$ and show that (15.27) implies that away from the source the function $f(r)$ satisfies

$$\frac{\partial^2 f}{\partial r^2} + k^2 f = 0 \; . \tag{15.32}$$

This equation has the solution $C\exp(\pm ikr)$. According to **problem e** the upper sign should be used and the Green's function is given by $G(r) = Ce^{ikr}/r$. Show that the condition (15.29) dictates that $C = -1/4\pi$, so that in three dimensions the Green's function is given by:

$$G^{3D}(r) = \frac{-1}{4\pi}\frac{e^{ikr}}{r} \; . \tag{15.33}$$

The problem is actually most difficult in two dimensions because in that case the Green's function cannot be expressed in the simplest elementary functions.

Problem h: Show that in two dimensions ($N = 2$) the differential equation of the Green's function is away from the source given by

$$\frac{\partial^2 G}{\partial r^2} + \frac{1}{r}\frac{\partial G}{\partial r} + k^2 G(r) = 0 \qquad , \; r \neq 0 \; . \tag{15.34}$$

Problem i: This equation cannot be solved in terms of elementary functions. However there is a close relation between equation (15.34) and the Bessel equation

$$\frac{d^2 F}{dx^2} + \frac{1}{x}\frac{dF}{dx} + (1 - \frac{m^2}{x^2})F = 0 \; . \tag{15.35}$$

Show that the $G(kr)$ satisfies the Bessel equation for order $m = 0$.

This implies that the Green's function is given by the solution of the zeroth-order Bessel equation with argument kr. The Bessel equation is a second order differential equation, there are therefore two independent solutions. The solution that is finite everywhere is denoted by $J_m(x)$, it is the called the regular Bessel function. The second solution is singular at the point $x = 0$ and is called the Neumann function denoted by $N_m(x)$. The Green's function obviously is a linear combination of $J_0(kr)$ and $N_0(kr)$. In order to determine how this linear combination is constructed it is crucial to consider the behavior of these functions at the source (i.e. for $x = 0$) and at infinity (i.e. for $x \gg 1$). The required asymptotic behavior can be found in textbooks such as *Butkov*[14] and *Arfken*[3] and is summarized in table (15.1).

Problem j: Show that neither $J_0(kr)$ nor $N_0(kr)$ behave for large values of r as $\exp(+ikr)$. Show that the linear combination $J_0(kr) + iN_0(kr)$ does behave as $\exp(+ikr)$.

The Green's function thus is a linear combination of the regular Bessel function and the Neumann function. This particular combination is called the *first Hankel function of*

	$J_0(x)$	$N_0(x)$
$x \to 0$	$1 - \frac{1}{4}x^2 + O(x^4)$	$\frac{2}{\pi}\ln(x) + O(1)$
$x \gg 1$	$\sqrt{\frac{2}{\pi x}}\cos\left(x - \frac{\pi}{4}\right) + O(x^{-3/2})$	$\sqrt{\frac{2}{\pi x}}\sin\left(x - \frac{\pi}{4}\right) + O(x^{-3/2})$

Table 15.1: Leading asymptotic behaviour of the Bessel function and Neumann function of order zero.

degree zero and is denoted by $H_0^{(1)}(kr)$. In general the Hankel functions are simply linear combinations of the Bessel function and the Neumann function:

$$\begin{aligned} H_m^{(1)}(x) &\equiv J_m(x) + iN_m(x) \\ H_m^{(2)}(x) &\equiv J_m(x) - iN_m(x) \end{aligned} \tag{15.36}$$

Problem k: Show that $H_0^{(1)}(kr)$ behaves for large values of r as $\exp(+i\,(kr) - i\pi/4)/\sqrt{\frac{\pi}{2}kr}$ and that in this limit $H_0^{(2)}(kr)$ behaves as $\exp(-i\,(kr) - i\pi/4)/\sqrt{\frac{\pi}{2}kr}$. Use this to argue that the Green's function is given by

$$G(r) = CH_0^{(1)}(kr)\ , \tag{15.37}$$

where the constant C still needs to be determined.

Problem l: This constant follows from the requirement (15.29) at the source. Use (15.36) and the asymptotic value of the Bessel function and the Neumann function given in table (15.1) to derive the asymptotic behavior of the Green's function near the source and use this to show that $C = -i/4$.

This result implies that in two dimensions the Green's function of the Helmholtz equation is given by

$$G^{2D}(r) = \frac{-i}{4}H_0^{(1)}(kr)\ . \tag{15.38}$$

Summarizing these results and reverting to the more general case of a source at location $\mathbf{r}_0$ it follows that the Green's functions of the Helmholtz equation is in one, two and three dimensions given by:

$$\begin{aligned} G^{1D}(x, x_0) &= \tfrac{-i}{2k}e^{ik|x-x_0|} \\ G^{2D}(\mathbf{r}, \mathbf{r}_0) &= \tfrac{-i}{4}H_0^{(1)}(k\,|\mathbf{r}-\mathbf{r}_0|) \\ G^{3D}(\mathbf{r}, \mathbf{r}_0) &= \tfrac{-1}{4\pi}\tfrac{e^{ik|\mathbf{r}-\mathbf{r}_0|}}{|\mathbf{r}-\mathbf{r}_0|} \end{aligned} \tag{15.39}$$

Note that in two and three dimensions the Green's function is singular at the source $\mathbf{r}_0$.

Problem m: Show that these singularity is *integrable*, i.e. show that when the Green's function is integrated over a sphere with finite radius around the source the result is finite.

There is a physical reason why the Green's function in two and three dimensions has an integrable singularity. Suppose one has a source that is not a point source but that the source is constant within a sphere with radius R centered around the origin. The response p to this source is given by $p(\mathbf{r}) = \int_{r'<R} G(\mathbf{r}, \mathbf{r}')dV'$ where the integration over the variable $\mathbf{r}'$ is over a sphere with radius R. It follows from this expression that the response in the origin is given by

$$p(\mathbf{r}=0) = \int_{r'<R} G(\mathbf{r}=0, \mathbf{r}')dV' \ . \tag{15.40}$$

Since the excitation of this field is finite everywhere, the response $p(\mathbf{r}=0)$ should be finite as well. This implies that the integral (15.40) should be finite as well, which is a different way of stating that the singularity of the Green's function must be integrable.

15.4 The wave equation in 1,2,3 dimensions

In this section we will consider the Green's function for the wave equation in 1,2 and 3 dimensions. This means that we consider solutions to the wave equation with an impulsive source at location $\mathbf{r}_0$ at time t_0:

$$\nabla^2 G(\mathbf{r}, t; \mathbf{r}_0, t_0) - \frac{1}{c^2}\frac{\partial^2 G(\mathbf{r}, t; \mathbf{r}_0, t_0)}{\partial t^2} = \delta(\mathbf{r}-\mathbf{r}_0)\delta(t-t_0) \qquad (15.22) \quad \textit{again.}$$

It was shown in the previous section that this Green's function depends only on the relative distance $|\mathbf{r}-\mathbf{r}_0|$ and the relative time $t-t_0$. For the case of a source in the origin ($\mathbf{r}_0 = 0$) acting at time zero ($t_0 = 0$) the time domain solution follows by applying a Fourier transform to the solution $G(\mathbf{r}, \omega)$ of the previous section. This Fourier transform is simplest in three dimensions, hence we will start with this case.

Problem a: Apply the Fourier transform (11.42) to the 3D Green's function (15.33) and use the relation $k = \omega/c$ and the properties of the delta function to show that the Green's function is in the time given by

$$G^{3D}(\mathbf{r}, t) = -\frac{1}{4\pi r}\delta\left(t - \frac{r}{c}\right) \ . \tag{15.41}$$

Problem b: Now consider the wave equation with a general source term $S(\mathbf{r},t)$:

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2}\frac{\partial^2 p(\mathbf{r}, t_0)}{\partial t^2} = S(\mathbf{r}, t) \ . \tag{15.42}$$

Use the Green's function (15.41) to show that a solution of this equation is given by

$$p(\mathbf{r}, t) = -\frac{1}{4\pi}\int \frac{S\left(t - \frac{|\mathbf{r}-\mathbf{r}'|}{c}\right)}{|\mathbf{r}-\mathbf{r}'|}dV' \ . \tag{15.43}$$

Note that since $|\mathbf{r}-\mathbf{r}'|$ is always positive, the response $p(\mathbf{r},t)$ depends only on the source function at *earlier* times. The solution therefore has a causal behavior and the Green's function (15.41) is called the *retarded Green's function.* However, in several applications one does not want to use a Green's function that depends on excitation on earlier times. An example is reflection seismology. In that case one records the wave field at the surface, and from these observations one wants to reconstruct the wave field at *earlier* times while it was being reflected off layers inside the earth. (See the treatment in section 6.3 and paper of *Schneider*[53].) A Green's function with waves that propagate towards the source and are then annihilated by the source can be obtained by replacing the radiation condition (15.30) by $\partial G/\partial r = -ikG$. The only difference is the minus sign in the right hand side, this is equivalent to replacing k by $-k$.

Problem c: Apply the Fourier transform (11.42) to the 3D Green's function (15.33) with k replaced by $-k$ and show that the resulting Green's function is in the time given by

$$G^{3D,\,advanced}(\mathbf{r},t) = -\frac{1}{4\pi r}\delta\left(t+\frac{r}{c}\right) \ , \tag{15.44}$$

and that the following function is a solution of the wave equation (15.42):

$$p(\mathbf{r},t) = -\frac{1}{4\pi}\int \frac{S\left(t+\frac{|\mathbf{r}-\mathbf{r}'|}{c}\right)}{|\mathbf{r}-\mathbf{r}'|}dV' \ . \tag{15.45}$$

Note that in this representation the wave field is expressed in the source function at *later* times. For this reason the Green's function (15.44) is called the *advanced Green's function.* The fact that the wave equation has both a retarded and an advanced solution is that the wave equation (15.42) is invariant for time-reversal, i.e. when one replaces t by $-t$ the equation does not change. In practice one works most often with the retarded Green's function, but keep in mind that in some application such as exploration seismology the advanced Green's functions are crucial. In the remaining part of this section we will focus exclusively of the retarded Green's functions that represent causal solutions.

In order to obtain the Green's function for two dimensions in the time domain one should apply a Fourier transform to the solution (15.38). This involves taking the Fourier transform of a Hankel function, and it is not obvious how this Fourier integral should be solved (although it can be solved). Here we will follow an alternative route by recognizing that the Green's function in two dimensions is identical to the solution of the wave equation in three dimensions when the source is not a point source but a cylinder-source. In other words, we obtain the 2D Green's function by considering the wave field in three dimensions that is generated by a source that is distributed homogeneously along the z-axis. In order to separate the distance to the origin from the distance to the z-axis the variables r and ρ are used, see figure (15.2).

Problem d: Show that

$$G^{2D}(\rho,t) = \int_{-\infty}^{\infty} G^{3D}(r,t)dz \ . \tag{15.46}$$

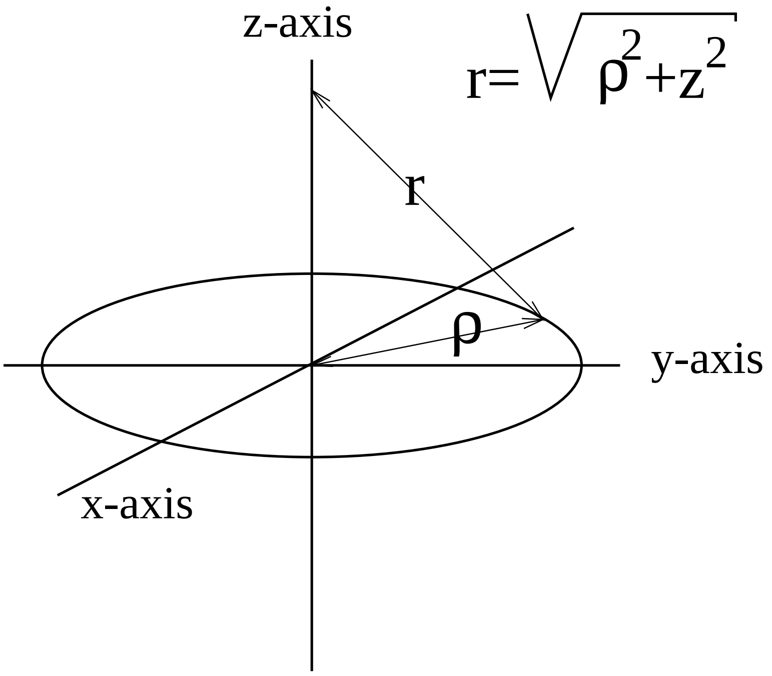


Figure 15.2: Definition of the variables r and ρ.

Problem e: Use the Green's function (15.41) and the relation $r = \sqrt{\rho^2 + z^2}$ to show that

$$G^{2D}(\rho, t) = -\frac{1}{2\pi} \int_0^\infty \frac{\delta\left(t - \frac{\sqrt{\rho^2+z^2}}{c}\right)}{\sqrt{\rho^2 + z^2}} dz \ . \tag{15.47}$$

Note that the integration interval has been changed from $(-\infty, \infty)$ to $(0, \infty)$, show how this can be achieved.

The distance r in three dimensions does not appear in this expression anymore. Without loss of generality we variable ρ can therefore be replaced by r.

Problem f: The integral (15.47) (with ρ replaced by r) can be solved by introducing the new integration variable $u \equiv \sqrt{r^2 + z^2}$ instead of the old integration variable z. Show that the integral (15.47) can with this new variable be written as

$$G^{2D}(r, t) = -\frac{1}{2\pi} \int_r^\infty \frac{\delta\left(t - \frac{u}{c}\right)}{\sqrt{u^2 - r^2}} du \ , \tag{15.48}$$

pay attention to the limits of integration!

Problem g: Use the property $\delta(ax) = \delta(x)/|a|$ to rewrite this integral and evaluate the resulting integral separately for $t < r/c$ and $t > r/c$ to show that:

$$G^{2D}(r, t) = \begin{cases} 0 & for \quad t < r/c \\ -\frac{1}{2\pi} \frac{1}{\sqrt{t^2 - r^2/c^2}} & for \quad t > r/c \end{cases} \tag{15.49}$$

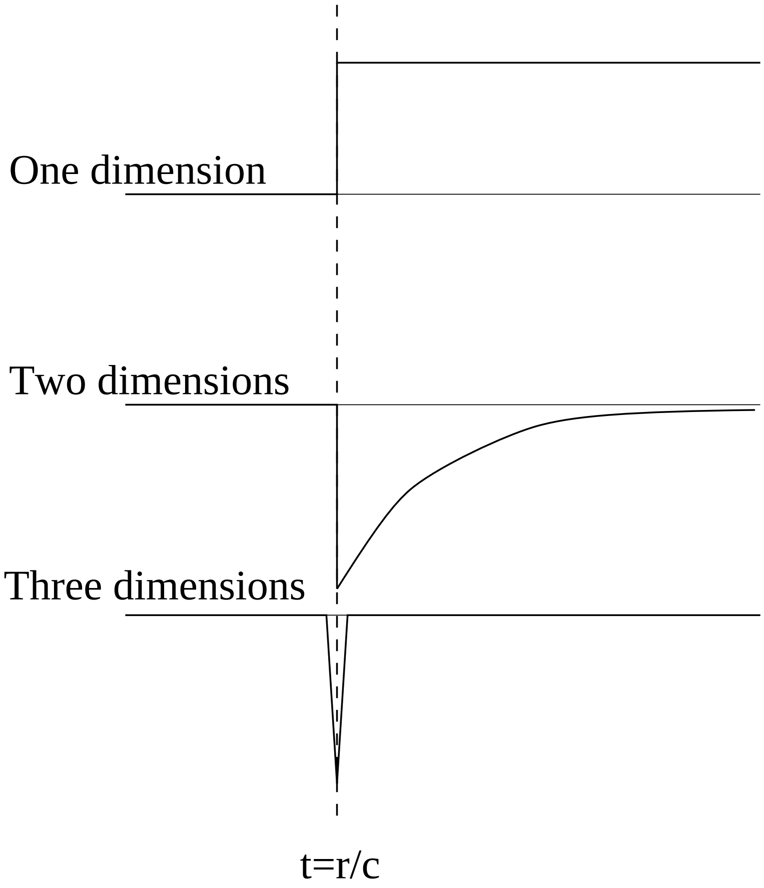


Figure 15.3: The Green's function of the wave equation in 1, 2 and 3 dimensios as a function of time.

This Green's function as well as the Green's function for the three-dimensional case is shown in figure (15.3). There is a fundamental difference between the Green's function for two dimensions and the Green's function (15.41) for three dimensions. In three dimensions the Green's function is a delta function $\delta(t-r/c)$ modulated by the geometrical spreading $-1/4\pi r$. This means that the response to a delta function source has the same shape as the input function $\delta(t)$ that excites the wave field. An impulsive input leads to an impulsive output with a time delay given by r/c and the solution is only nonzero at the wave front $t = r/c$. However, expressions (15.49) shows that an impulsive input in two dimensions leads to a response that is not impulsive. The response has an infinite duration and decays with time as $1/\sqrt{t^2 - r^2/c^2}$, the solution is not only nonzero *at* the wave front $t = r/c$, but it is nonzero everywhere *within* this wave front.. This means that in two dimensions an impulsive input leads to a sound response that is of infinite duration. One can therefore say that:

Any word spoken in two dimensions will reverberate forever (albeit weakly).

The approach we have taken is to compute the Green's function in two dimension is interesting in that we solved the problem first in a higher dimension and retrieved the solution by integrating over one space dimension. Note that for this trick it is not necessary that this higher dimensional space indeed exists! (Although in this case it does.) Remember that we took this approach because we did not want to evaluate the Fourier transform of a Hankel function. We can also turn this around; the Green's function (15.49) can be used to determine the Fourier transform of the Hankel function.

Problem h: Show that the Fourier transform of the Hankel function is given by:

$$\int_{-\infty}^{\infty} H_0^{(1)}(x)e^{-iqx}dx = \begin{cases} 0 & for \quad q<1 \\ \frac{2}{i\pi}\frac{1}{\sqrt{q^2-1}} & for \quad q>1 \end{cases} \tag{15.50}$$

Let us continue with the Green's function of the wave equation in one dimension in the time domain.

Problem i: Use the Green's function for one dimension of the last section to show that in the time domain

$$G^{1D}(x,t) = -\frac{ic}{4\pi}\int_{-\infty}^{\infty}\frac{1}{\omega}e^{-i\omega(t-|x|/c)}d\omega\ . \tag{15.51}$$

This integral resembles the integral used for the calculation of the Green's function in three dimensions. The only difference is the term $1/\omega$ in the integrand, because of this term we cannot immediately evaluate the integral. However, the $1/\omega$ term can be removed by differentiating expression (15.51) with respect to time, and the remaining integral can be evaluated analytically.

Problem j: Show that

$$\frac{\partial G^{1D}(x,t)}{\partial t} = \frac{c}{2}\delta\left(t-\frac{|x|}{c}\right)\ . \tag{15.52}$$

Problem k: This expression can be integrated but one condition is needed to specify the integration constant that appears. We will use here that at $t=-\infty$ the Green's function vanishes. Show that with this condition the Green's function is given by:

$$G^{1D}(x,t) = \begin{cases} 0 & for \quad t<|x|/c \\ c/2 & for \quad t>|x|/c \end{cases} \tag{15.53}$$

Just as in two dimensions the solution is nonzero everywhere *within* the expanding wave front and not only on the wave front $|x|=ct$ such as in three dimensions. However, there is an important difference; in two dimensions the solution changes for all times with time whereas in one dimension the solution is constant except for $t=|x|/c$. Humans cannot detect a static change in pressure (did you ever hear something when you drove in the mountains?), therefore a one-dimensional human will only hear a sound at $t=|x|/c$ but not at later times.

In order to appreciate the difference in the sound propagation in 1, 2 and 3 space dimensions the Green's functions for the different dimensions is shown in figure (15.3). Note the dramatic change in the response for different numbers of dimensions. This change in the properties of the Green's function with change in dimension has been used somewhat jokingly by *Morley*[41] to give "a simple proof that the world is three dimensional." When you have worked through the sections (15.1) and (15.2) you have learned that both for the heat equation and the Schrödinger equation the solution does not depend fundamentally on the number of dimensions. This is in stark contrast with the solutions of the wave equation that depend critically on the number of dimensions.

Chapter 16

Normal modes

Many physical systems have the property that they can carry out oscillations only at certain specific frequencies. As a child (and hopefully also as an adult) you will have discovered that a swing on a playground will move only with a very specific natural period, and that the force that pushes the swing is only effective when the period of the force matches the period of the swing. The patterns of motion at which a system oscillates are called the *normal modes* of the system. A swing may has one normal mode, but you have seen in section 10.7 that a simple model of a tri-atomic molecule has three normal modes. An example of a normal mode of a system is shown in figure 16.1. Shown is the

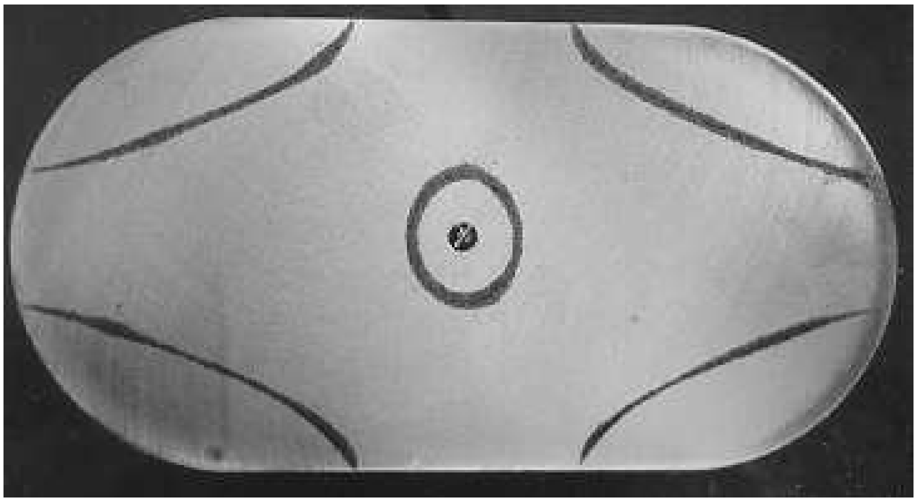


Figure 16.1: Sand on a metal plate that is driven by an oscillator at a frequency that corresponds to one of the eigen-frequencies of the plate. This figure was prepared by John Scales at the Colorado School of Mines.

pattern of oscillation of a metal plate that is driven by an oscillator at a fixed frequency. The screw in the middle of the plate shows the point at which the force on the plate is applied. Sand is sprinkled on the plate. When the frequency of the external force is equal to the frequency of a normal mode of the plate, the motion of the plate is given by the motion that corresponds to that specific normal mode. Such a pattern of oscillation has nodal lines where the motion vanishes. These nodal lines are visible because the sand on the plate collects at the these lines.

In this chapter, the normal modes of a variety of systems are analyzed. Normal modes play an important role in a variety of applications because the eigen-frequencies of normal modes provide important information of physical systems. Examples are the normal modes the Earth that provide information about the internal structure of our planet, or the spectral lines of light emitted by atoms that have led to the advent of quantum mechanics and its description of the internal structure of atoms. In addition, normal modes are used in this chapter to introduce some properties of special functions such as Bessel functions and Legendre functions. This is achieved by analyzing the normal modes of a system in 1, 2 and 3 dimensions in the sections 16.1 through 16.3.

16.1 The normal modes of a string

In this section and the following two sections we assume that the motion of the system is governed by the Helmholtz equation

$$\nabla^2 u + k^2 u = 0 \, . \tag{16.1}$$

In this expression the wave-number k is related to the angular frequency ω by the relation

$$k = \frac{\omega}{c} \, . \tag{16.2}$$

For simplicity we assume the system to be homogeneous, this means that the velocity c is constant. This in turn implies that the wave-number k is constant. In the sections 16.1 through 16.3 we consider a body with size R. Since a circle or sphere with radius R has a diameter $2R$ we will consider here a string with length $2R$ in order to be able to make meaningful comparisons. It is assumed that the endpoints of the string are fixed so that the boundary conditions are:

$$u(0) = u(2R) = 0 \, . \tag{16.3}$$

Problem a: Show that the solutions of (16.1) that satisfy the boundary conditions (16.3) are given by $\sin k_n r$ with the wave-number k_n given by

$$k_n = \frac{n\pi}{2R} \, , \tag{16.4}$$

where n is an integer.

For a number of purposes is it useful to normalize the modes, this means that we require that the modes $u_n(x)$ satisfy the condition $\int_0^{2R} u_n^2(x) dx = 1$.

Problem b: Show that the normalized modes are given by

$$u_n(x) = \frac{1}{\sqrt{R}} \sin k_n r \, . \tag{16.5}$$

Problem c: Sketch the modes for several values of n as a function of the distance x.

Problem d: The modes $u_n(x)$ are orthogonal, which means that the inner product $\int_0^{2R} u_n(x)u_m(x)dx$ vanishes when $n \neq m$. Give a proof of this property to derive that

$$\int_0^{2R} u_n(x)u_m(x)dx = \delta_{nm} . \tag{16.6}$$

We conclude from this section that the modes of a string are oscillatory functions with a wave-number that can only have discrete well-defined values k_n. According to expression (16.2) this means that the string can only vibrate at discrete frequencies that are given by

$$\omega_n = \frac{n\pi c}{2R} . \tag{16.7}$$

This property will be familiar to you because you probably know that a guitar string vibrates only at a very specific frequencies that determines the pitch of the sound that you hear. The results of this section imply that each string does not only oscillate at one particular frequency, but at many discrete frequencies. The oscillation with the lowest frequency is given by (16.7) with $n = 1$, this is called the *fundamental mode* or *ground-tone*. This is what the ear perceives as the pitch of the tone. The oscillations corresponding to larger values of n are called the *higher modes* or *overtones*. The particular mix of overtones determines the timbre of the signal. If the higher modes are strongly excited the ear perceives this sound as metallic, whereas the fundamental mode only is perceived as a smooth sound. The reader who is interested in the theory of musical instruments can consult Ref.[49].

The discrete modes are not a peculiarity of the string. Most systems that support waves and that are of a finite extend support modes. For example, in figure 11.1 of chapter 11 the spectrum of the sound of a soprano saxophone is shown. This spectrum is characterized by well-defined peaks that corresponds to the modes of the air-waves in the instrument. Mechanical systems in general have discrete modes, these modes can be destructive when they are excited at their resonance frequency. The matter waves in atoms are organized in modes as well, this is ultimately the reason why atoms in an excited state emit only light are very specific frequencies, called spectral lines.

16.2 The normal modes of drum

In the previous section we looked at the modes of a one-dimensional system. Here we will derive the modes of a two-dimensional system which is a model of a drum. We consider a two-dimensional membrane that satisfies the Helmholtz equation (16.1). The membrane is circular and has a radius R. At the edge, the membrane cannot move, this means that in cylinder coordinates the boundary condition for the waves $u(r, \varphi)$ is given by:

$$u(R, \varphi) = 0 . \tag{16.8}$$

In order to find the modes of the drum we will use *separation of variables*, this means that we seek solutions that can be written as a product of the function that depends only on r and a function that depends only φ:

$$u(r, \varphi) = F(r)G(\varphi) \tag{16.9}$$

Problem a: Insert this solution in the Helmholtz equation, use the expression of the Laplacian in cylinder coordinates, and show that the resulting equation can be written as

$$\underbrace{\left(\frac{1}{F(r)}r\frac{\partial}{\partial r}\left(r\frac{\partial F}{\partial r}\right)+k^2r^2\right)}_{\text{(A)}} = - \underbrace{\frac{1}{G(\varphi)}\frac{\partial^2 G}{\partial \varphi^2}}_{\text{(B)}} \tag{16.10}$$

Problem b: The terms labelled (A) depend on the variable r only whereas the terms labelled (B) depend only on the variable φ. These terms can only be equal for all values of r and φ when they depend neither on r nor on φ, i.e. when they are a constant. Use this to show that $F(r)$ and $G(\varphi)$ satisfy the following differential equations:

$$\frac{d^2F}{dr^2}+\frac{1}{r}\frac{dF}{dr}+\left(k^2-\frac{\mu}{r^2}\right)F=0\ , \tag{16.11}$$

$$\frac{d^2G}{d\varphi^2}+\mu G=0\ , \tag{16.12}$$

where μ is a constant that is not yet known.

These differential equations need to be supplemented with boundary conditions. The boundary conditions for $F(r)$ follow from the requirement that this function is finite everywhere and that the displacement vanishes at the edge of the drum:

$$F(r) \text{ is finite everywhere} \quad , \quad F(R)=0. \tag{16.13}$$

The boundary condition for $G(\varphi)$ follows from the requirement that if we rotate the drum over 360°, every point on the drum returns to its original position. This means that the modes satisfy the requirement that $u(r,\varphi)=u(r,\varphi+2\pi)$. This implies that $G(\varphi)$ satisfies the *periodic boundary condition*:

$$G(\varphi)=G(\varphi+2\pi)\ . \tag{16.14}$$

Problem c: The general solution of (16.12) is given by $G(\varphi)=\exp\left(\pm i\sqrt{\mu}\varphi\right)$. Show that the boundary condition (16.14) implies that $\mu=m^2$, with m an integer.

This means that the dependence of the modes on the angle φ is given by:

$$G(\varphi)=e^{im\varphi}\ . \tag{16.15}$$

The value $\mu=m^2$ can be inserted in (16.11). The resulting equation then bears a close resemblance to the Bessel equation:

$$\frac{d^2J_m}{dx^2}+\frac{1}{x}\frac{dJ_m}{dx}+\left(1-\frac{m^2}{x^2}\right)J_m=0\ . \tag{16.16}$$

This equation has two independent solutions; the Bessel function $J_m(x)$ that is finite everywhere and the Neumann function $N_m(x)$ that is singular at $x=0$.

Problem d: Show that the general solution of (16.11) can be written as:

$$F(r) = AJ_m(kr) + BN_m(kr) \ , \tag{16.17}$$

with A and B integration constants.

Problem e: Use the boundary conditions of $F(r)$ show that $B = 0$ and that the wave-number k must take a value such that $J_m(kR) = 0$.

This last condition for the wave-number is analogous to the condition (16.4) for the one-dimensional string. For both the string and the drum the wave-number can only take discrete values, these values are dictated by the condition that the displacement vanishes at the outer boundary of the string or the drum. It follows from (16.4) that for the string there are infinitely many wave-numbers k_n. Similarly, for the drum there are for every value of the the angular degree m infinitely many wave-numbers that satisfy the requirement $J_m(kR) = 0$. These wave-numbers are labelled with a subscript n, but since these wave-numbers are different for each value of the angular order m, the allowed wave-numbers carry two indices and are denoted by $k_n^{(m)}$. They satisfy the condition

$$J_m(k_n^{(m)}R) = 0 \ . \tag{16.18}$$

The zeroes of the Bessel function $J_m(x)$ are not known in closed form. However, tables exists of the zero crossings of Bessel functions, see for example, table 9.4 of *Abramowitz and Stegun*[1]. Take a look at this reference which contains a bewildering collection of formulas, graphs and tables of mathematical functions. The lowest order zeroes of the Bessel functions $J_0(x)$, $J_1(x), \ldots, J_5(x)$ are shown in table 16.1.

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$
$n=1$	2.40482	3.83171	5.13562	6.38016	7.58834	8.77148
$n=2$	5.52007	7.01559	8.41724	9.76102	11.06471	12.33860
$n=3$	8.65372	10.17347	11.61984	13.01520	14.37254	15.70017
$n=4$	11.79153	13.32369	14.79595	16.22347	17.61597	18.98013
$n=5$	14.93091	16.47063	17.95982	19.4092	20.82693	22.21780
$n=6$	18.07106	19.61586	21.11700	22.58273	24.01902	25.43034
$n=7$	21.21163	22.76008	24.27011	25.74817	27.19909	28.62662

Table 16.1: The zeroes of the Bessel function $J_m(x)$.

Problem f: Find the eigen-frequencies of the four modes of the drum with the lowest frequencies and make a sketch of the associated wave standing wave of the drum.

Problem g: Compute the separation between the different zero crossing for a fixed value of m. To which number does this separation converge for the zero crossings at large values of x?

Using the results of this section it follows that the modes of the drum are given by

$$u_{nm}(r, \varphi) = J_m(k_n^{(m)}r)e^{im\varphi} \ . \tag{16.19}$$

Problem h: Let us first consider the φ-dependence of these modes. Show that when one follows the mode $u_{nm}(r, \varphi)$ along a complete circle around the origin that one encounters exactly m oscillations of that mode.

The shape of the Bessel function is more difficult to see than the properties of the functions $e^{im\varphi}$. As shown in section 9.7 of *Butkov*[14] these functions satisfy a large number of properties that include recursion relations and series expansions. However, at this point the following facts are most important:

- The Bessel functions $J_m(x)$ are oscillatory functions that decay with distance, in a sense they behave as decaying standing waves We will return to this issue in section 16.5.

- The Bessel functions satisfy an orthogonality relation similar to the orthogonality relation (16.6) for the modes of the string. This orthogonality relation is treated in more detail in section 16.4.

16.3 The normal modes of a sphere

In this section we consider the normal modes of a spherical surface with radius R. We only consider the modes that are associated with the waves that propagate along the surface, hence we do not consider wave motion in the *interior* of the sphere. The modes satisfy the wave equation (16.1). Since the waves propagate on the spherical surface, they are only a function of the angles θ and φ that are used in spherical coordinates: $u = u(\theta, \varphi)$. Using the expression of the Laplacian in spherical coordinates the wave equation (16.1) is then given by

$$\frac{1}{R^2}\left\{\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial u}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2 u}{\partial\varphi^2}\right\}+k^2u=0\ . \tag{16.20}$$

Again, we will seek a solution by applying separation of variables by writing the solution in a form similar to (16.9):

$$u(\theta,\varphi)=F(\theta)G(\varphi)\ . \tag{16.21}$$

Problem a: Insert this in (16.20) and apply separation of variables to show that $F(\theta)$ satisfies the following differential equation:

$$\sin\theta\frac{d}{d\theta}\left(\sin\theta\frac{dF}{d\theta}\right)+\left(k^2R^2\sin^2\theta-\mu\right)F=0\ , \tag{16.22}$$

and that $G(\varphi)$ satisfies (16.12), where the unknown constant μ does not depend on θ or φ.

To make further progress we have to apply boundary conditions. Just as with the drum of section 16.2 the system is invariant when a rotation over 2π is applied: $u(\theta, \varphi) = u(\theta, \varphi + 2\pi)$. This means that $G(\varphi)$ satisfies the same differential equation (16.12) as for the case of the drum and satisfies the same periodic boundary condition (16.14). The solution is therefore given by $G(\varphi) = e^{im\varphi}$ and the separation constant satisfies $\mu = -m^2$, with m an integer. Using this, the differential equation for $F(\theta)$ can be written as:

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dF}{d\theta}\right) + \left(k^2R^2 - \frac{m^2}{\sin^2\theta}\right)F = 0 \ . \tag{16.23}$$

Before we continue let us compare this equation with expression (16.11) for the modes of the drum that we can rewrite as

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dF}{dr}\right) + \left(k^2 - \frac{m^2}{r^2}\right)F = 0 \qquad \text{(16.11) again}$$

Note that these equations are identical when we compare r in (16.11) with $\sin\theta$ in (16.23). There is a good reason for this. Suppose the we have a source in the middle of the drum. In that case the variable r measures the distance of a point on the drum to the source. This can be compared with the case of waves on a spherical surface that are excited by a source at the north-pole. In that case, $\sin\theta$ is a measure of the distance of a point to the source point. The only difference is that $\sin\theta$ enters the equation rather than the true angular distance θ. This is a consequence of the fact that the surface is curved, this curvature leaves an imprint on the differential equation that the modes satisfy.

Problem b: The differential equation (16.11) was reduced in section 16.2 to the Bessel equation by changing to a new variable $x = kr$. Define a new variable

$$x \equiv \cos\theta \tag{16.24}$$

and show that the differential equation for F is given by

$$\frac{d}{dx}\left(\left(1-x^2\right)\frac{dF}{dx}\right) + \left(k^2R^2 - \frac{m^2}{1-x^2}\right)F = 0 \ . \tag{16.25}$$

The solution of this differential equation is given by the associated Legendre functions $P_l^m(x)$. These functions are described in great detail in section 9.8 of *Butkov*[14]. In fact, just as the Bessel equation, the differential equation (16.25) has a solution that is regular as well a solution $Q_l^m(x)$ that is singular at the point $x = 1$ where $\theta = 0$. However, since the modes are finite everywhere, they are given by the regular solution $P_l^m(x)$ only.

The wave-number k is related to frequency by the relation $k = \omega/c$. At this point it is not clear what k is, hence the eigen-frequencies of the spherical surface are not yet known. It is shown in section 9.8 of *Butkov*[14] that:

- The associated Legendre functions are only finite when the wave-number satisfies

$$k^2R^2 = l(l+1) \ , \tag{16.26}$$

where l is a positive integer. Using this in (16.25) implies that the associated Legendre functions satisfy the following differential equation:

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dP_l^m(\cos\theta)}{d\theta}\right) + \left(l\,(l+1) - \frac{m^2}{\sin^2\theta}\right)P_l^m(\cos\theta) = 0 \ . \tag{16.27}$$

Seen as a function of x $(= \cos\theta)$ this is equivalent to the following differential equation

$$\frac{d}{dx}\left(\left(1-x^2\right)\frac{dP_l^m(x)}{dx}\right)+\left(l\,(l+1)-\frac{m^2}{1-x^2}\right)P_l^m(x)=0\;. \tag{16.28}$$

- The integer l must be larger or equal than the absolute value of the angular order m.

Problem c: Show that the last condition can also be written as:

$$-l \leq m \leq l\;. \tag{16.29}$$

Problem d: Derive that the eigen-frequencies of the modes are given by

$$\omega_l = \frac{\sqrt{l(l+1)}R}{c}\;. \tag{16.30}$$

It is interesting to compare this result with the eigen-frequencies (16.7) of the string. The eigen-frequencies of the string all have the same spacing in frequency, but the eigen-frequencies of the spherical surface are not spaced at the same interval. In musical jargon one would say that the overtones of a string are harmonious, this means that the eigen-frequencies of the overtones are multiples of the eigen-frequency of the ground tone. In contrast, the overtones of a spherical surface are not harmonious.

Problem e: Show that for large values of l the eigen-frequencies of the spherical surface have an almost equal spacing.

Problem f: The eigen-frequency ω_l only depends on the order l but not on the degree m. For each value of l, the angular degree m can according to (16.29) take the values $-l, -l+1, \cdots, l-1, l$. Show that this implies that for every value of l, there are $(2l+1)$ modes with the same eigen-frequency.

When different modes have the same eigen-frequency one speaks of *degenerate modes*.

The results we obtained imply that the modes on a spherical surface are given by $P_l^m(\cos\theta)e^{im\varphi}$. We used here that the variable x is related to the angle θ through the relation (16.24). The modes of the spherical surface are called *spherical harmonics*. These eigen-functions are for $m \geq 0$ given by:

$$Y_{lm}(\theta,\varphi) = (-1)^m\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)e^{im\varphi} \qquad m \geq 0\;. \tag{16.31}$$

For $m < 0$ the spherical harmonics are defined by the relation

$$Y_{lm}(\theta,\varphi) = (-1)^m\,Y_{l,-m}(\theta,\varphi) \tag{16.32}$$

You may wonder where the square-root in front of the associated Legendre function comes from. One can show that with this numerical factor the spherical harmonics are normalized when integrated over the sphere:

$$\iint |Y_{lm}|^2\,d\Omega = 1\;, \tag{16.33}$$

where $\iint \cdots d\Omega$ denotes an integration over the unit sphere. You have to be aware of the fact that different authors use different definitions of the spherical harmonics. For example, one could define the spherical harmonics also as $\tilde{Y}_{lm}(\theta,\varphi) = P_l^m(\cos\theta)e^{im\varphi}$ because the functions also account for the normal modes of a spherical surface.

Problem g: Show that the modes defined in this way satisfy $\iint \left|\tilde{Y}_{lm}\right|^2 d\Omega = 4\pi/\left(2l+1\right)\times (l+m)!/(l-m)!$.

This means that the modes defined in this way are not normalized when integrated over the sphere. There is no reason why one cannot work with this convention, as long as one accounts for the fact that in this definition the modes are not normalized. Throughout this book we will use the definition (16.31) for the spherical harmonics. In doing so we follow the normalization that is used by *Edmonds*[20].

Just as with the Bessel functions, the associated Legendre functions satisfy recursion relations and a large number of other properties that are described in detail in section 9.8 of *Butkov*[14]. The most important properties of the spherical harmonics $Y_{lm}(\theta,\varphi)$ are:

- These functions display m oscillations when the angle φ increases with 2π. In other words, there are m oscillations along one circle of constant latitude.
- The associated Legendre functions $P_l^m(\cos\theta)$ behave like Bessel functions that they behave like standing waves with an amplitude that decays from the pole. We return to this issue in section 16.6.
- There are $l-m$ oscillations between the north pole of the sphere and the south pole of the sphere.
- The spherical harmonics are orthogonal for a suitably chosen inner product, this orthogonality relation is derived in section 16.4.

A last and very important property is that the spherical harmonics are the eigen-functions of the Laplacian in the sphere.

Problem h: Give a proof of this last property by showing that

$$\nabla_1^2 Y_{lm}(\theta,\varphi) = -l\left(l+1\right)Y_{lm}(\theta,\varphi) \; , \tag{16.34}$$

where the Laplacian on the unit sphere is given by

$$\nabla_1^2 = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2} \; . \tag{16.35}$$

This property is in many applications extremely useful, because the action of the Laplacian on a sphere can be replaced by the much simpler multiplication with the constant $-l\left(l+1\right)$ when spherical harmonics are concerned.

16.4 Normal modes and orthogonality relations

The normal modes of a physical system often satisfy orthogonality relations when a suitably chosen inner product for the eigen-functions is used. In this section this is illustrated by studying once again the normal modes of the Helmholtz equation (16.1) for different geometries. In this section we derive first the general orthogonality relation for these normal modes. This is then applied to the normal modes of the previous sections to derive the orthogonality relations for Bessel functions and associated Legendre functions.

Let us consider two normal modes of the Helmholtz equation (16.1), and let these modes be called u_p and u_q. At this point we leave it open whether the modes are defined on a line, on a surface of arbitrary shape or a volume. The integration over the region of space in which the modes are defined is denoted as $\int \cdots d^N x$, where N is the dimension of this space. The wave-number of these modes that acts as an eigenvalue in the Helmholtz equation is defined by k_p and k_q respectively. In other words, the modes satisfy the equations:

$$\nabla^2 u_p + k_p^2 u = 0 \; , \tag{16.36}$$

$$\nabla^2 u_q + k_q^2 u_q = 0 \; . \tag{16.37}$$

The subscript p may stand for a single mode index such as in the index n for the wave-number k_n for the modes of a string, or it may stand for a number of indices such as the indices nm that label the eigen-functions (16.19) of a circular drum.

Problem a: Multiply (16.36) with u_q^*, take the complex conjugate of (16.37) and multiply the result with u_p. Subtract the resulting equations and integrate this over the region of space for which the modes are defined to show that

$$\int \left(u_q^* \nabla^2 u_p - u_p \nabla^2 u_q^* \right) d^N x + \left(k_p^2 - k_q^{*2} \right) \int u_q^* u_p d^N x = 0 \; . \tag{16.38}$$

Problem b: Use the theorem of Gauss to derive that

$$\int u_q^* \nabla^2 u_p d^N x = \oint u_q^* \nabla u_p \cdot d\mathbf{S} - \int \left(\nabla u_q^* \cdot \nabla u_p \right) d^N x \; , \tag{16.39}$$

where the integral $\oint \cdots \cdot d\mathbf{S}$ is over the surface that bounds the body. If you have trouble deriving this, you can consult expression (6.9) of section 6.3 where a similar result was used for the derivation of the representation theorem for acoustic waves.

Problem c: Use the last result to show that

$$\oint \left(u_q^* \nabla u_p - u_p^* \nabla u_q \right) \cdot d\mathbf{S} + \left(k_p^2 - k_q^{*2} \right) \int u_q^* u_p d^N x = 0 \; . \tag{16.40}$$

Problem d: The result is now expressed in the first term as an integral over the boundary of the body. Let us assume that the modes satisfy on this boundary one of the three boundary conditions: (i) $u = 0$, (ii) $\hat{\mathbf{n}} \cdot \boldsymbol{\nabla} u = 0$ (where $\hat{\mathbf{n}}$ is the unit vector perpendicular to the surface) or (iii) $\hat{\mathbf{n}} \cdot \boldsymbol{\nabla} u = \alpha u$ (where α is a constant). Show that for all of these boundary conditions the surface integral in (16.40) vanishes.

The last result implies that when the modes satisfy one of these boundary conditions that

$$\left(k_p^2 - k_q^{*2}\right) \int u_q^* u_p d^N x = 0 \; . \tag{16.41}$$

Let us first consider the case that the modes are equal, i.e. that $p = q$. In that case the integral reduces to $\int |u_p|^2 \, d^N x$ which is guaranteed to be positive. Equation (16.41) then implies that $k_p^2 = k_p^{*2}$, so that the wave-numbers k_p must be real: $k_p = k_p^*$. For this reason the complex conjugate of the wave-numbers can be dropped and (16.41) can be written as:

$$\left(k_p^2 - k_q^2\right) \int u_q^* u_p d^N x = 0 \; . \tag{16.42}$$

Now consider the case of two different modes for which the wave-numbers k_p and k_q are different. In that case the term $\left(k_p^2 - k_q^2\right)$ is nonzero, hence in order to satisfy (16.42) the modes must satisfy

$$\int u_q^* u_p d^N x = 0 \qquad \text{for} \qquad k_p \neq k_q \; . \tag{16.43}$$

This finally gives the *orthogonality relation* of the modes in the sense that it states that the modes are orthogonal for the following inner product: $\langle f \cdot g \rangle \equiv \int f^* g \; d^N x$. Note that the inner product for which the modes are orthogonal follows from the Helmholtz equation (16.1) that defines the modes.

Let us now consider this orthogonality relation for the modes of the string, the drum and the spherical surface of the previous sections. For the string the orthogonality relation was derived in **problem d** of section 16.1 and you can see that equation (16.9) is identical to the general orthogonality relation (16.43). For the circular drum the modes are given by equation (16.19).

Problem e: Use expression (16.19) for the modes of the circular drum to show that the orthogonality relation (16.43) for this case can be written as:

$$\int_0^R \int_0^{2\pi} J_{m_1}(k_{n_1}^{(m_1)} r) J_{m_2}(k_{n_2}^{(m_2)} r) e^{i(m_1 - m_2)\varphi} d\varphi \; r dr = 0 \qquad \text{for} \quad k_{n_1}^{(m_1)} \neq k_{n_2}^{(m_2)} \tag{16.44}$$

Explain where the factor r comes from in the integration.

Problem f: This integral can be separated in an integral over φ and an integral over r. The φ-integral is given by $\int_0^{2\pi} e^{i(m_1 - m_2)\varphi} d\varphi$. Show that this integral vanishes when $m_1 \neq m_2$:

$$\int_0^{2\pi} e^{i(m_1 - m_2)\varphi} d\varphi = 0 \qquad \text{for} \qquad m_1 \neq m_2 \; . \tag{16.45}$$

Note that you have derived this relation earlier in expression (13.9) of section 13.2 in the derivation of the residue theorem.

Expression (16.45) implies that the modes $u_{n_1 m_1}(r, \varphi)$ and $u_{n_2 m_2}(r, \varphi)$ are orthogonal when $m_1 \neq m_2$ because the φ-integral in (16.44) vanishes when $m_1 \neq m_2$. Let us now consider why the different modes of the drum are orthogonal when m_1 and m_2 are equal to the same integer m. In that case (16.44) implies that

$$\int_0^R J_m(k_{n_1}^{(m)} r) J_m(k_{n_2}^{(m)} r) \; r \; dr = 0 \qquad \text{for} \quad n_1 \neq n_2 \; . \tag{16.46}$$

Note that we have used here that $k_{n_1}^{(m)} \neq k_{n_2}^{(m)}$ when $n_1 \neq n_2$. This integral defines an orthogonality relation for Bessel functions. Note that both Bessel functions in this relation are of the same degree m but that the wave-numbers in the argument of the Bessel functions differ. Note the resemblance between this expression and the orthogonality relation of the modes of the string that can be written as

$$\int_0^{2R} \sin k_n x \ \sin k_m x \ dx = 0 \qquad \text{for} \quad n \neq m \ . \tag{16.47}$$

The presence of the term r in the integral (16.46) comes from the fact that the modes of the drum are orthogonal for the integration over the total area of the drum. In cylinder coordinates this leads to a factor r in the integration.

Problem g: Take you favorite book on mathematical physics and find an alternative derivation of the orthogonality relation (16.46) of the Bessel functions of the same degree m.

Note finally that the modes $u_{n_1 m_1}(r,\varphi)$ and $u_{n_2 m_2}(r,\varphi)$ are orthogonal when $m_1 \neq m_2$ because the φ-integral satisfies (16.45) whereas the modes are orthogonal when $n_1 \neq n_2$ but the same order m because the r-integral (16.46) vanishes in that case. This implies that the eigen-functions of the drum defined in (16.19) satisfy the following orthogonality relation:

$$\int_0^R \int_0^{2\pi} u_{n_1 m_1}^*(r,\varphi) u_{n_2 m_2}(r,\varphi) \ d\varphi \ r dr = C \delta_{n_1 n_2} \delta_{m_1 m_2} \ , \tag{16.48}$$

where δ_{ij} is the Kronecker delta and C is a constant that depends on n_1 and m_1.

A similar analysis can be applied to the spherical harmonics $Y_{lm}(\theta,\varphi)$ that are the eigen-functions of the Helmholtz equation on a spherical surface. You may wonder in that case what the boundary conditions of these eigen-functions are because in the step from equation (16.40) to (16.41) the boundary conditions of the modes have been used. A closed surface has, however, no boundary. This means that the surface integral in (16.40) vanishes. This means that the orthogonality relation (16.43) holds despite the fact that the spherical harmonics do not satisfy one of the boundary conditions that has been used in **problem d**. Let us now consider the inner product of two spherical harmonics on the sphere: $\iint Y_{l_1 m_1}^*(\theta,\varphi) Y_{l_2 m_2}(\theta,\varphi) d\Omega$.

Problem h: Show that the φ-integral in the integration over the sphere is of the form $\int_0^{2\pi} \exp i\,(m_2 - m_1)\, d\varphi$ and that this integral is equal to $2\pi\delta_{m_1 m_2}$.

This implies that the spherical harmonics are orthogonal when $m_1 \neq m_2$ because of the φ-integration. We will now continue with the case that $m_1 = m_2$, and denote this common value with the single index m.

Problem i: Use the general orthogonality relation (16.43) to derive that the associated Legendre functions satisfy the following orthogonality relation:

$$\int_0^{\pi} P_{l_1}^m(\cos\theta) P_{l_2}^m(\cos\theta) \sin\theta d\theta = 0 \qquad \text{when } l_1 \neq l_2 \ . \tag{16.49}$$

Note the common value of the degree m in the two associated Legendre functions. Show also explicitly that the condition $k_{l_1} \neq k_{l_2}$ is equivalent to the condition $l_1 \neq l_2$.

Problem j: Use a substitution of variables to show that this orthogonality relation can also be written as

$$\int_{-1}^{1} P_{l_1}^m(x)P_{l_2}^m(x)dx = 0 \qquad \text{when } l_1 \neq l_2 \ . \tag{16.50}$$

Problem k: Find an alternative derivation of this orthogonality relation in the literature.

The result you obtained in **problem h** implies that the spherical harmonics are orthogonal when $m_1 \neq m_2$ because of the φ-integration, whereas **problem i** implies that the spherical harmonics are orthogonal when $l_1 \neq l_2$ because of the θ-integration. This means that the spherical harmonics satisfy the following orthogonality relation:

$$\iint Y_{l_1 m_1}^*(\theta,\varphi)Y_{l_2 m_2}(\theta,\varphi)d\Omega = \delta_{l_1 l_2}\delta_{m_1 m_2} \ . \tag{16.51}$$

The numerical constant multiplying the delta functions is equal to 1, this is a consequence of the square-root term in (16.31) that pre-multiplies the associated Legendre functions. One should be aware of the fact that when a different convention is used for the normalization of the spherical harmonics a normalization factor appears in the right hand side of the orthogonality relation (16.51) of the spherical harmonics.

16.5 Bessel functions are decaying cosines

As we have seen in section 16.2 the modes of the circular drum are given by $J_m(kr)e^{im\varphi}$ where the Bessel function satisfies the differential equation (16.16) and where k is a wavenumber chosen in such a way that the displacement at the edge of the drum vanishes. We will show in this section that the waves that propagate through the drum have approximately a constant wavelength, but that their amplitude decays with the distance to the center of the drum. The starting point of the analysis is the Bessel equation

$$\frac{d^2 J_m}{dx^2} + \frac{1}{x}\frac{dJ_m}{dx} + \left(1 - \frac{m^2}{x^2}\right) J_m = 0 \qquad \text{(16.16) again.}$$

If the terms $\frac{1}{x}\frac{dJ_m}{dx}$ and $\frac{m^2}{x^2}$ would be absent in (16.16) the Bessel equation would reduce to the differential equation $d^2F/dx^2 + F = 0$ whose solutions are given by a superposition of $\cos x$ and $\sin x$. We therefore can expect the Bessel functions to display an oscillatory behavior when x is large.

It follows directly from (16.16) that the term m^2/x^2 is relatively small for large values of x, specifically when $x \gg m$. However, it is not obvious under which conditions the term $\frac{1}{x}\frac{dJ_m}{dx}$ is relatively small. Fortunately this term can be transformed away.

Problem a: Write $J_m(x) = x^\alpha g_m(x)$, insert this in the Bessel equation (16.16), show that the term with the first derivative vanishes when $\alpha = -1/2$ and that the resulting differential equation for $g_m(x)$ is given by

$$\frac{d^2 g_m}{dx^2} + \left(1 - \frac{m^2 - 1/4}{x^2}\right) g_m = 0 \ . \tag{16.52}$$

Up to this point we have made no approximation. Although we have transformed the first derivative term out of the Bessel equation, we still cannot solve (16.52). However, when $x \gg m$ the term proportional to $1/x^2$ in this expression is relatively small. This means that for large values of x the function $g_m(x)$ satisfies the following approximate differential equation $d^2 g_m/dx^2 + g_m \approx 0$.

Problem b: Show that the solution of this equation is given by $g_m(x) \approx A\cos(x+\varphi)$, where A and φ are constants. Also show that this implies that the Bessel function is approximately given by:

$$J_m(x) \approx A\frac{\cos(x+\varphi)}{\sqrt{x}} \; . \tag{16.53}$$

This approximation is obtained from a local analysis of the Bessel equation. Since all values of the constants A and φ lead to a solution that approximately satisfies the differential equation (16.52), it is not possible to retrieve the precise values of these constant from the analysis of this section. An analysis based on the asymptotic evaluation of the integral representation of the Bessel function [7] shows that:

$$J_m(x) = \sqrt{\frac{2}{\pi x}}\cos\left(x - (2m+1)\,\frac{\pi}{4}\right) + O(x^{-3/2}) \; . \tag{16.54}$$

Problem c: As a check on the accuracy of this asymptotic expression let us compare the zeroes of this approximation with the zeroes of the Bessel functions as given in table 16.1 of section 16.2. In **problem g** of section 16.2 you found that the separation of the zero crossings tends to π for large values of x. Explain this using the approximate expression (16.54) How large must x be for the different values of the order m so that the error in the spacing of the zero crossing is less than 0.01?

Physically, expression (16.54) states that Bessel functions behave like standing waves with a constant wavelength and which decay with distance as $1/\sqrt{kr}$. (Here it is used that the modes are given by the Bessel functions with argument $x = kr$.) How can we explain this decay of the amplitude with distance? First let us note that (16.54) expresses the Bessel function in a cosine, hence this is a representation of the Bessel function as a standing wave. However, using the relation $\cos x = \left(e^{ix} + e^{-ix}\right)/2$ the Bessel function can be written as two travelling waves that depend on the distance as $(\exp \pm ikr)/\sqrt{kr}$ and that interfere to give the standing wave pattern of the Bessel function. Now let us consider a propagating wave $A(r)\exp(ikr)$ in two dimensions, in this expression $A(r)$ is an amplitude that is at this point unknown. The energy varies with the square of the wave-field, and thus depends on $|A(r)|^2$. The energy current therefore also varies as $|A(r)|^2$. Consider an outgoing wave as shown in figure 16.2. The total energy flux through a ring of radius r is given by the energy current times the circumference of the ring, this means that the flux is equal to $2\pi r\,|A(r)|^2$. Since energy is conserved, this total energy flux is the same for all values of r, which means that $2\pi r\,|A(r)|^2 = \text{constant}$.

Problem d: Show that this implies that $A(r) \sim 1/\sqrt{r}$.

This is the same dependence on distance as the $1/\sqrt{x}$ decay of the approximation (16.54) of the Bessel function. This means that the decay of the Bessel function with distance is dictated by the requirement of energy conservation.

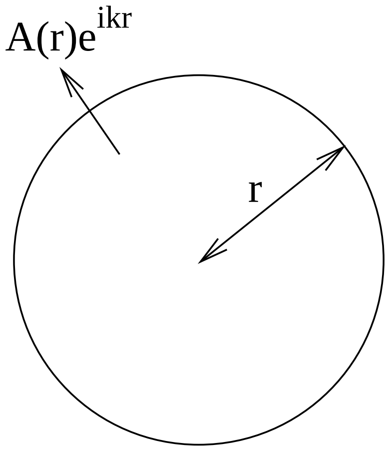


Figure 16.2: An expanding wavefront with radius r.

16.6 Legendre functions are decaying cosines

The technique used in the previous section for the approximation of the Bessel function can also be applied to spherical harmonics. We will show in this section that the spherical harmonics behave asymptotically as standing waves on a sphere with an amplitude decay that is determined by the condition that energy is conserved. The spherical harmonics are proportional to the associated Legendre functions with argument $\cos\theta$, the starting point of our analysis therefore is the differential equation for $P_l^m(\cos\theta)$ that was derived in section 16.3:

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dP_l^m(\cos\theta)}{d\theta}\right)+\left(l\,(l+1)-\frac{m^2}{\sin^2\theta}\right)P_l^m(\cos\theta)=0 \qquad (16.27)\text{ again}$$

Let us assume we have a source at the north pole, where $\theta = 0$. Far away from the source, the term $m^2/\sin^2\theta$ in the last term is much smaller than the constant $l\,(l+1)$.

Problem a: Show that the words "far away from the source" stand for the requirement

$$\sin\theta \gg \frac{m}{\sqrt{l\,(l+1)}}\ , \qquad (16.55)$$

and show that this implies that the approximation that we will derive will break down near the north pole as well as near the south pole of the employed system of spherical coordinates. In addition, the asymptotic expression that we will derive will be most accurate for large values of the angular order l.

Problem b: Just as in the previous section we will transform the first derivative in the differential equation (16.27) away, here this can be achieved by writing $P_l^m(\cos\theta) = (\sin\theta)^\alpha\, g_l^m(\theta)$. Insert this substitution in the differential equation (16.27), show that the first derivative $dg_l^m/d\theta$ disappears when $\alpha = -1/2$, and that the resulting differential equation for $g_l^m(\theta)$ is given by:

$$\frac{d^2 g_l^m}{d\theta^2}+\left\{\left(l+\frac{1}{2}\right)^2-\frac{m^2}{\sin^2\theta}-\frac{1}{4}\frac{\cos^2\theta}{\sin^2\theta}\right\}g_l^m(\theta)=0\ . \qquad (16.56)$$

Problem c: If the terms $m^2/\sin^2\theta$ and $\cos^2\theta/4\sin^2\theta$ would be absent this equation would be simple to solve. Show that these terms are small compared to the constant $\left(l+\frac{1}{2}\right)^2$ when the requirement (16.55) is satisfied.

Problem d: Show that under this condition the associated Legendre functions satisfy the following approximation

$$P_l^m(\cos\theta) \approx A\frac{\cos\left(\left(l+\frac{1}{2}\right)\theta+\gamma\right)}{\sqrt{\sin\theta}}\ , \tag{16.57}$$

where A and γ are constants.

Just as in the previous section the constants A and γ cannot be obtained from this analysis because (16.57) satisfies the approximate differential equation for *any* values of these constant. As shown in expression (2.5.58) of Ref. [20] the asymptotic relation of the associated Legendre functions is given by:

$$P_l^m(\cos\theta) \approx (-l)^m\sqrt{\frac{2}{\pi l\sin\theta}}\cos\left(\left(l+\frac{1}{2}\right)\theta-(2m+1)\frac{\pi}{4}\right)+O(l^{-3/2}) \tag{16.58}$$

This means that the spherical harmonics also have the same approximate dependence on the angle θ. Just like the Bessel functions the spherical harmonics behave like standing wave given by a cosine that is multiplied by a factor $1/\sqrt{\sin\theta}$ that modulates the amplitude.

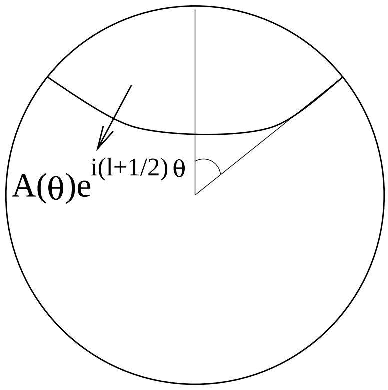


Figure 16.3: An expanding wavefront on a spherical surface at a distance θ from the source.

Problem e: Use a reasoning as you used in **problem d** of section 16.5 to explain that this amplitude decrease follows from the requirement of energy conservation. In doing so you may find figure 16.3 helpful.

Problem f: Deduce from (16.58) that the wavelength of the associated Legendre functions measured in radians is given by $2\pi/\left(l+\frac{1}{2}\right)$.

This last result can be used to find the number of oscillations in the spherical harmonics when one moves around the globe once. For simplicity we consider here the case of a spherical harmonic $Y_l^0(\theta, \varphi)$ for degree $m = 0$. When one goes from the north pole to the south pole, the angle θ increases from 0 to π. The number of oscillations that fit in this interval is given by $\pi/wavelength$, according to **problem f** this number is equal to $\pi/\left(2\pi/\left(l+\frac{1}{2}\right)\right) = \left(l+\frac{1}{2}\right)/2$. This is the number of wavelenths that fit on half the globe. When one returns from the south pole to the north pole one encounters another $\left(l+\frac{1}{2}\right)$ oscillations. This means that the total number of waves that fit around the globe is given by $\left(l+\frac{1}{2}\right)$. It may surprise you that the number of oscillations that one encounters making one loop around the globe is not an integer. One would expect that the requirement of constructive interference dictates that an integer number of wavelengths should "fit" in this interval. The reason that the total number of oscillations is $\left(l+\frac{1}{2}\right)$ rather than the integer l is that near the north pole and near the south pole the asymptotic approximation (16.58) breaks down, this follows from the requirement (16.55).

The fact that $\left(l+\frac{1}{2}\right)$ rather than l oscillations fit on the globe has a profound effect on quantum mechanics. In the first attempts to explain the line spectra of light emitted by atoms, Bohr postulated that an integer number of waves has to fit on a sphere, this can be expressed as $\oint k ds = 2\pi n$, where k is the local wave-number. This condition could not explain the observed spectra of light emitted by atoms. However, the arguments of this section imply that the number of wavelengths that fit on a sphere should be given by the requirement

$$\oint k ds = 2\pi \left(n + \frac{1}{2}\right) . \tag{16.59}$$

This is the *Bohr-Sommerfeld quantization rule*, which was the earliest result in quantum mechanics that provided an explanation of the line-spectra of light emitted by atoms.[1] More details on this issue and the cause of the factor $\frac{1}{2}$ in the quantization rule can be found in the Refs. [59] and [12].

The asymptotic expression (16.58) can give a useful insight in the relation between modes and travelling waves on a sphere. Let us first return to the modes on the string, which according to (16.5) are given by $\sin k_n x$. For simplicity, we will leave out normalization constants in the arguments. The wave motion associated with this mode is given by the real part of $\sin k_n x \ \exp(-i\omega_n t)$, with $\omega_n = k_n/c$. These modes therefore denote a standing wave. However, using the decomposition $\sin k_n x = \left(e^{ik_n x} - e^{-ik_n x}\right)/2i$, the mode can in the time domain also be seen as a superposition of two waves $e^{i(k_n x - \omega_n t)}$ and

[1]The fact that $(l+1/2)$ oscillations of the spherical harmonics fit on the sphere appears to be in contrast with the statement made in section 16.3 that the spherical harmonic Y_l^m has exactly $l-m$ oscillations. The reason for this discrepancy is that the sperical harmonics are only oscillatory for an angle θ that satisfies the inequality (16.55). This means for angular degree m that is nonzero the spherical harmonics only oscillate with wavelength $2\pi/(l+1/2)$ on only part of the sphere. This leads to a reduction of the number of oscillations of the spherical harmonics between the poles with increasing degree m. However, the quantization condition (16.59) holds for every degree m because a proper treatment of this quatization condition stipulates that the integral is taken over a region where the modes are oscillatory. This means that one should not integrate from pole-to-pole, but that the integration must be taken over a closed path that is not aligned with the north-south direction on the sphere. One can show that one encounters exactly $(l+1/2)$ oscillations while making a closed loop along such a path. This issue is explained in a clear and pictorial way by *Dahlen and Henson* [18].

$e^{-i(k_n x+\omega_n t)}$. These are two travelling waves that move in opposite directions.

Problem g: Suppose we excite a string at the left side at $x = 0$. We know we can account for the motion of the string as a superposition of standing waves $\sin k_n x$. However, we can consider these modes to exist also as a superposition of waves $e^{\pm ik_n x}$ that move in opposite directions. The wave $e^{ik_n x}$ moves away from the source at $x = 0$. However the wave $e^{-ik_n x}$ moves *towards* the source at $x = 0$. Give a physical explanation why in the string travelling waves also move *towards* the source.

On the sphere the situation is completely analogous. The modes can be written according to (16.58) as standing waves $\cos\left(\left(l+\frac{1}{2}\right)\theta-(2m+1)\frac{\pi}{4}\right)/\sqrt{\sin\theta}$ on the sphere. However, using the relation $\cos x = \left(e^{ix}+e^{-ix}\right)/2$ the modes can also be seen as a superposition of travelling waves $e^{i\left(l+\frac{1}{2}\right)\theta}/\sqrt{\sin\theta}$ and $e^{-i\left(l+\frac{1}{2}\right)\theta}/\sqrt{\sin\theta}$ on the sphere.

Problem h: Explain why the first wave travels away from the north pole while the second wave travels towards the north pole.

Problem i: Suppose the waves are excited by a source at the north pole. According to the last problem the motion of the sphere can alternatively be seen as a superposition of standing waves or of travelling waves. The travelling wave $e^{i\left(l+\frac{1}{2}\right)\theta}/\sqrt{\sin\theta}$ moves away from the source. Explain physically why there is also a travelling wave $e^{-i\left(l+\frac{1}{2}\right)\theta}/\sqrt{\sin\theta}$ moving *towards* the source.

These results imply that the motion of the Earth can either be seen as a superposition of normal modes, or of a superposition of waves that travel along the Earth's surface in opposite directions. The waves that travel along the Earth's surface are called *surface waves.* The relation between normal modes and surface waves is treated in more detail by *Dahlen*[17] and by *Snieder and Nolet*[57].

16.7 Normal modes and the Green's function

In section (10.7) we analyzed the normal modes of a system of three coupled masses. This system had three normal modes, and each mode could be characterized with a vector $\mathbf{x}$ with the displacement of the three masses. The response of the system to a force $\mathbf{F}$ acting on the three masses with time dependence $\exp(-i\omega t)$ was derived to be:

$$\mathbf{x} = \frac{1}{m}\sum_{n=1}^{3}\frac{\hat{\mathbf{v}}^{(n)}(\hat{\mathbf{v}}^{(n)}\cdot\mathbf{F})}{(\omega_n^2-\omega^2)} \qquad (10.83) \quad \text{again} .$$

This means that the Green's function of this system is given by the following dyad:

$$\mathbf{G} = \frac{1}{2\pi m}\sum_{n=1}^{3}\frac{\hat{\mathbf{v}}^{(n)}\hat{\mathbf{v}}^{(n)T}}{(\omega_n^2-\omega^2)} . \tag{16.60}$$

The factor $1/2\pi$ is due to the fact that a delta function $f(t) = \delta(t)$ force in the time-domain corresponds with the Fourier transform (11.43) to $F(\omega) = 1/2\pi$ in the frequency-domain.

In this section we derive the Green's function for a very general oscillating system that can be continuous. An important example is the Earth, which a body that has well-defined normal modes and where the displacement is a continuous functions of the space coordinates.

We consider a system that satisfies the following equation of motion:

$$\rho\ddot{u} + Hu = F \ . \tag{16.61}$$

The field u can either be a scalar field or a vector field. The operator H is at this point very general, the only requirement that we impose is that this operator is Hermitian, this means that we require that

$$(f \cdot Hg) = (Hf \cdot g) \ , \tag{16.62}$$

where the inner product of is defined as $(f \cdot h) \equiv \int f^* g \, dV$. In the frequency domain, the equation of motion is given by

$$-\rho\omega^2 u + Hu = F(\omega) \ . \tag{16.63}$$

Let the normal modes of the system be denoted by $u^{(n)}$, the normal modes describe the oscillations of the system in the absence of any external force. The normal modes therefore satisfy the following expression

$$Hu^{(n)} = \rho\omega_n^2 u^{(n)} \ , \tag{16.64}$$

where ω_n is the eigen-frequency of this mode.

Problem a: Take the inner product of this expression with a mode $u^{(m)}$, use that H is Hermitian to derive that

$$\left(\omega_n^2 - \omega_m^{*2}\right)\left(u^{(m)} \cdot \rho u^{(n)}\right) = 0 \ . \tag{16.65}$$

Note the resemblance of this expression with (16.41) for the modes of a system that obeys the Helmholtz equation.

Problem b: Just as in section 16.4 one can show that the eigen-frequencies are real by setting $m = n$, and one can derive that different modes are orthogonal with respect to the following inner product:

$$\left(u^{(m)} \cdot \rho u^{(n)}\right) = \delta_{nm} \qquad \text{for} \quad \omega_m \neq \omega_n \ . \tag{16.66}$$

Give a proof of this orthogonality relation.

Note the presence of the density term ρ in this inner product.

Let us now return to the inhomogeneous problem (16.63) where an external force $F(\omega)$ is present. Assuming that the normal modes form a complete set, the response to this force can be written as a sum of normal modes:

$$u = \sum_n c_n u^{(n)} \ , \tag{16.67}$$

where the c_n are unknown coefficients.

Problem c: Find these coefficients by inserting (16.67) in the equation of motion (16.63) and by taking the inner product of the result with a mode $u^{(m)}$ to derive that

$$c_m = \frac{\left(u^{(m)} \cdot F\right)}{\omega_m^2 - \omega^2} \; . \tag{16.68}$$

This means that the response of the system can be written as:

$$u = \sum_n \frac{u^{(n)} \left(u^{(n)} \cdot F\right)}{\omega_n^2 - \omega^2} \; . \tag{16.69}$$

Note the resemblance of this expression with equation (10.83) for a system of three masses. The main difference is that the derivation of this section is valid as well for continuous vibrating systems such as the Earth.

It is instructive to re-write this expression taking the dependence of the space coordinates explicitly into account:

$$u(\mathbf{r}) = \sum_n \frac{u^{(n)}(\mathbf{r}) \int u^{*(n)}(\mathbf{r}')F(\mathbf{r}')dV'}{\omega_n^2 - \omega^2} \; . \tag{16.70}$$

It follows from this expression that the Green's function is given by

$$G(\mathbf{r}, \mathbf{r}', \omega) = \frac{1}{2\pi} \sum_n \frac{u^{(n)}(\mathbf{r}) u^{*(n)}(\mathbf{r}')}{\omega_n^2 - \omega^2} \; . \tag{16.71}$$

When the mode is a vector, one should take the transpose of the mode $u^{*(n)}(\mathbf{r}')$. Note the similarity between this expression for the Green's function of a continuous medium with the Green's function (16.60) for a discrete system. In this sense, the Earth behaves in the same way as a tri-atomic molecule. For both systems, the dyadic representation of the Green's function provides a very compact way for accounting for the response of the system to external forces.

Note that the response is strongest when the frequency ω of the external force is close to one of the eigen-frequencies ω_n of the system. This implies for example for the Earth that the modes with a frequency close to the frequency of the external forcing are most strongly excited. If we jump up and down with a frequency of $1Hz$, we excite the gravest normal mode of the Earth with a period of about 1 hour only very weakly. In addition, a mode is most effectively excited when the inner product of the forcing $F(\mathbf{r}')$ in (16.70) is maximal. This means that a mode is most strongly excited when the spatial distribution of the force equals the displacement $u^{(n)}(\mathbf{r}')$ of the mode.

Problem d: Show that a mode is not excited when the force acts only at one of the nodal lines of that mode.

As a next step we consider the Green's function in the time domain. This function follows by applying the Fourier transform (11.42) to the Green's function (16.71).

Problem e: Show that this gives:

$$G(\mathbf{r}, \mathbf{r}', t) = \frac{1}{2\pi} \sum_n u^{(n)}(\mathbf{r}) u^{*(n)}(\mathbf{r}') \int_{-\infty}^{\infty} \frac{e^{-i\omega t}}{\omega_n^2 - \omega^2} d\omega \ . \tag{16.72}$$

The integrand is singular at the frequencies $\omega = \pm\omega_n$ of the normal modes. These singularities are located on the integration path, as shown in the left panel of figure 16.4. At the singularity at $\omega = \omega_n$ the integrand behaves as $1/\left(2\omega_n\left(\omega - \omega_n\right)\right)$. The contribution of these singularities is poorly defined because the integral $\int 1/\left(\omega - \omega_n\right)\, d\omega$ is not defined.

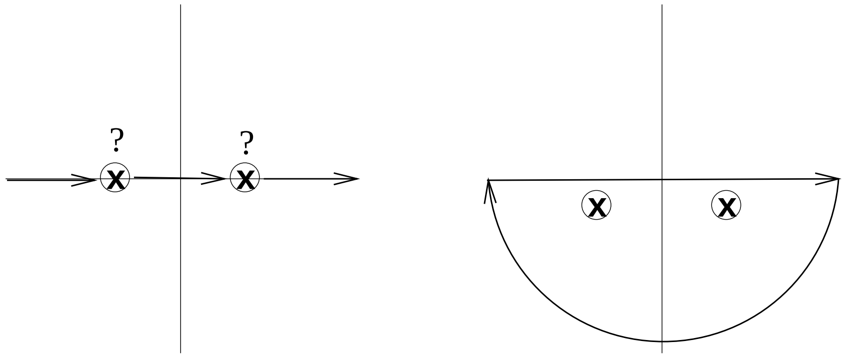


Figure 16.4: The location of the poles and the integration path in the complex ω-plane. The poles are indicared with a cross. Left panel, the original situation where the poles are located on the integration path at location $\pm\omega_n$. Right panel, the location of the poles when a slight anelastic damping is present.

This situation is comparable to the treatment in section 13.4 of the response of a particle in syrup to an external forcing. When this particle was subjected to a damping β, the integrand in the Fourier transform to the time domain had a singularity in the lower half plane. This gave a causal response; as shown in equation (13.35) the response was only different from zero at times *later* than the time at which the forcing was applied. This suggests that we can obtain a well-defined causal response of the Green's function (16.72) when we introduce a slight damping. This damping breaks the invariance of the problem for time-reversal, and is responsible for a causal response. At the end of the calculation we can let the damping parameter go to zero. Damping can be introduced by giving the eigen-frequencies of the normal modes a small negative imaginary component: $\pm\omega_n \to \pm\omega_n - i\eta$, where η is a small positive number.

Problem f: The time-dependence of the oscillation of a normal mode is given by $e^{-i\omega_n t}$. Show that with this replacement the modes decay with a decay time that is given by $\tau = 1/\eta$.

This last property means that when we ultimately set $\eta = 0$ that the decay time because infinite, in other words, the modes are not attenuated in that limit.

With the replacement $\pm\omega_n \to \pm\omega_n - i\eta$ the poles that are associated with the normal modes are located in the lower ω-plane, this situation is shown in figure 16.4. Now that the singularities are moved from the integration path the theory of complex integration can be used to evaluate the resulting integral.

Problem g: Use the theory of contour integration as treated in chapter 13 to derive that the Green's function is in the time domain given by:

$$G(\mathbf{r},\mathbf{r}',t)=\begin{cases} 0 & for\ t<0 \\ \sum_n \dfrac{u^{(n)}(\mathbf{r})u^{*(n)}(\mathbf{r}')}{\omega_n}\sin\omega_n t & for\ t>0 \end{cases} \tag{16.73}$$

Hint: use the same steps as in the derivation of the function (13.35) of section 13.4 and let the damping parameter η go to zero at the end of the integration.

This result gives a causal response because the Green's function is only nonzero at times $t>0$, which is later than the time $t=0$ when the delta function forcing is nonzero. The total response is given as a sum over all the modes. Each mode leads to a time signal $\sin\omega_n t$ in the modal sum, this is a periodic oscillation with the frequency ω_n of the mode. The singularities in the integrand of the Green's function (16.72) at the pole positions $\omega=\pm\omega_n$ are thus associated in the time domain with a harmonic oscillation with angular frequency ω_n. Note that the Green's function is continuous at the time $t=0$ of excitation.

Problem h: Use the Green's function (16.73) to derive that the response of the system to a force $F(\mathbf{r},t)$ is given by:

$$u(\mathbf{r},t)=\sum_n\frac{1}{\omega_n}u^{(n)}(\mathbf{r})\int\int_{-\infty}^{t}u^{*(n)}(\mathbf{r}')\sin\omega_n\left(t-t'\right)F(\mathbf{r}',t')\,dt'dV'. \tag{16.74}$$

Justify the integration limit in the t'-integration.

The results of this section imply that the total Green's function of a system is known once the normal modes are known. The total response can then be obtained by summing the contribution of each normal mode to the total response. This technique is called *normal-mode summation*, it is often use to obtain the low-frequency response of the Earth to an excitation [19]. However, in the seismological literature one usually treats a source signal that is given by a step function at $t=0$ rather than a delta function because this is a more accurate description of the slip on a fault during an earthquake[2]. This leads to a time-dependence $(1-\cos\omega_n t)$ rather than the time-dependence $\sin\omega_n t$ in the response (16.73) to an delta function excitation.

16.8 Guided waves in a low velocity channel

In this section we will consider a system that strictly speaking does not have normal modes, but that can support solutions that behave like travelling waves in one direction and as modes in another direction. The waves in such a system propagate as *guided waves*. Consider a system in two dimensions (x and z) where the velocity depends only on the z-coordinate. We assume that the wave-field satisfies in the frequency domain the Helmholtz equation (16.1):

$$\nabla^2 u+\frac{\omega^2}{c^2(z)}u=0\ . \tag{16.75}$$

In this section we consider a simple model of a layer with thickness H that extends from $z = 0$ to $z = H$ where the velocity is given by c_1. This layer is embedded in a medium with a constant velocity c_0. The geometry of the problem is shown in figure 16.5. Since the system is invariant in the x-direction, the problem can be simplified by a Fourier transform over the x-coordinate:

$$u(x,z) = \int_{-\infty}^{\infty} U(k,z) e^{ikx} dk \ . \tag{16.76}$$

Problem a: Show that $U(k,z)$ satisfies the following ordinary differential equation:

$$\frac{d^2 U}{dz^2} + \left(\frac{\omega^2}{c(z)^2} - k^2 \right) U = 0 \ . \tag{16.77}$$

It is important to note at this point that the frequency ω is a fixed constant, and that according to (16.76) the variable k is an integration variable that can be anything. For this reason one should at this point not use a relation $k = \omega / c(z)$ because k can still be anything.

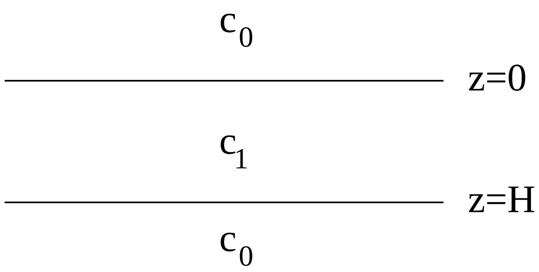


Figure 16.5: Geometry of the model of a single layer sandwiched between two homogeneous half-spaces.

Now consider the special case of the model shown in figure 16.5. We require that the waves outside the layer move away from the layer.

Problem b: Show that this implies that the solution for $z < 0$ is given by $A \exp(-ik_0 z)$ and the solution for $z > H$ is given by $B \exp(+ik_0 z)$ where A and B are unknown integration constants and where k_0 is given by

$$k_0 = \sqrt{\frac{\omega^2}{c_0^2} - k^2} \ . \tag{16.78}$$

Problem c: Show that within the layer the wave-field is given by $C \cos k_1 z + D \sin k_1 z$ with C and D integration constants and k_1 is given by

$$k_1 = \sqrt{\frac{\omega^2}{c_1^2} - k^2} \ . \tag{16.79}$$

The solution in the three regions of space therefore takes the following form:

$$U(k,z) = \begin{cases} A \exp(-ik_0 z) & for \quad z < 0 \\ C \cos k_1 z + D \sin k_1 z & for \quad 0 < z < H \\ B \exp(+ik_0 z) & for \quad z > H \end{cases} \tag{16.80}$$

We now have the general form of the solution within the layer and the two half-spaces on both sides of the layer. Boundary conditions are needed to find the integration constants A, B, C and D. For this system both U and dU/dz are continuous at $z = 0$ and $z = H$.

Problem d: Use the results of **problem b** and **problem c** to show that these requirement impose the following constraints on the integration constants:

$$\begin{aligned} A - C &= 0 \, , \\ ik_0 A + k_1 D &= 0 \, , \\ -Be^{ik_0 H} + C\cos k_1 H + D \sin k_1 H &= 0 \, , \\ ik_0 B e^{ik_0 H} + k_1 C \cos k_1 H - k_1 D \sin k_1 H &= 0 \, . \end{aligned} \tag{16.81}$$

This is a linear system of four equations for the four unknowns A, B, C and D. Note that this is a homogeneous system of equations, because the right hand side vanishes. Such a homogeneous system of equations only has nonzero solutions when the determinant of the system of equations vanishes.

Problem e: Show that this requirement leads to the following condition:

$$\tan k_1 H = \frac{-2ik_0 k_1}{k_1^2 + k_o^2} \tag{16.82}$$

This equation is implicitly an equation for the wave-number k, because according to (16.78) and (16.79) both k_0 and k_1 are a function of the wave-number k. Equation (16.82) implies that the system can only support waves when the wave-number k is such that expression (16.82) is satisfied. The system does strictly speaking not have normal modes, because the waves propagate in the x-direction. However, in the z-direction the waves only "fit" in the layer for very specific values of the wave-number k. These waves are called "guided waves" because they propagate along the layer with a well-defined phase velocity that follows from the relation $c(\omega) = \omega/k$. Be careful not to confuse this phase velocity $c(\omega)$ with the velocities c_1 and c_0 in the layer and the half-spaces outside the layer. At this point we do not know yet what the phase velocities of the guided waves are.

The phase velocity follows from expression (16.82) because this expression is implicitly an equation for the wave-number k. At this point we consider the case of a low-velocity layer, i.e. we assume that $c_1 < c_0$. In this case $1/c_0 < 1/c_1$. We will look for guided waves with a wave-number in the following interval: $\omega/c_0 < k < \omega/c_1$.

Problem f: Show that in that case k_1 is real and that k_0 is purely imaginary. Write $k_0 = i\kappa_0$ and show that

$$\kappa_0 = \sqrt{k^2 - \frac{\omega^2}{c_0^2}} \, . \tag{16.83}$$

Problem g: Show that the solution decays exponentially away from the low-velocity channel both in the half-space $z < 0$ and the half-space $z > H$.

The fact that the waves decay exponentially with the distance to the plate means that the guided waves are trapped near the low-velocity layer. Waves that decay exponentially are called *evanescent waves.*

Problem h: Use (16.82) to show that the wave-number of the guided waves satisfies the following relation

$$\tan\sqrt{\frac{\omega^2}{c_1^2}-k^2}H=\frac{2\sqrt{k^2-\frac{\omega^2}{c_0^2}}\sqrt{\frac{\omega^2}{c_1^2}-k^2}}{\omega^2\left(\frac{1}{c_1^2}-\frac{1}{c_0^2}\right)} \tag{16.84}$$

For a fixed value of ω this expression constitutes a constraint on the wave-number k of the guided waves. Unfortunately, it is not possible to solve this equation for k in closed form. Such an equation is called a *transcendental equation.*

Problem j: Make a sketch of both the left hand side and the right hand side of expression (16.84) as a function of k. Show that the two curves have a finite number of intersection points.

These intersection points correspond to the k-values of the guided waves. The corresponding phase velocity $c = \omega/k$ in general depends on the frequency ω. This means that these guided waves are *dispersive,* which means that the different frequency components travel with a different phase velocity. It is for this reason that expression (16.82) is called the *dispersion relation.*

Dispersive waves occur in many different situations. When electromagnetic waves propagate between plates or in a layered structure, guided waves result [31]. The atmosphere, and most importantly the ionosphere is an excellent waveguide for electromagnetic waves[27]. This leads to a large variety of electromagnetic guided waves in the upper atmosphere with exotic names such as "pearls", "whistlers", "tweaks", "hydromagnetic howling" and "serpentine emissions"; colorful names associated with the sounds these phenomena would make if they were audible, or with the patterns they generate in frequency-time diagrams. Guided waves play a crucial role in telecommunication, because light propagates through optical fibers as guided waves[36]. The fact that these waves are guided prohibits the light to propagate out of the fiber, this allows for the transmission of light signals over extremely large distances. In the Earth the wave velocity increases rapidly with depth. Elastic waves can be guided near the Earth's surface and the different modes are called "Rayleigh waves" and "Love waves"[2]. These surface waves in the Earth are a prime tool for mapping the shear-velocity within the Earth[56].

Since the surface waves in the Earth are trapped near the Earth's surface, they propagate effectively in two dimensions rather than in three dimensions. The surface waves therefore suffer less from geometrical spreading than the body waves that propagate through the interior of the Earth. For this reason, it is the surface waves that do most damage after an earthquake. This is illustrated in figure 16.6 which shows the vertical displacement at a seismic station in Naroch (Belarus) after an earthquake at Jan-Mayen island. Around $t = 300s$ and $t = 520s$ impulsive waves arrive, these are the body waves that travel through the interior of the Earth. The wave with the largest amplitude that arrives between $t = 650s$ and $t = 900s$ is the surface wave that is guided along the Earth's surface. Note that the waves that arrive around $t = 700s$ have a lower frequency content than the waves that arrive later around $t = 850s$. This is due to the fact that the

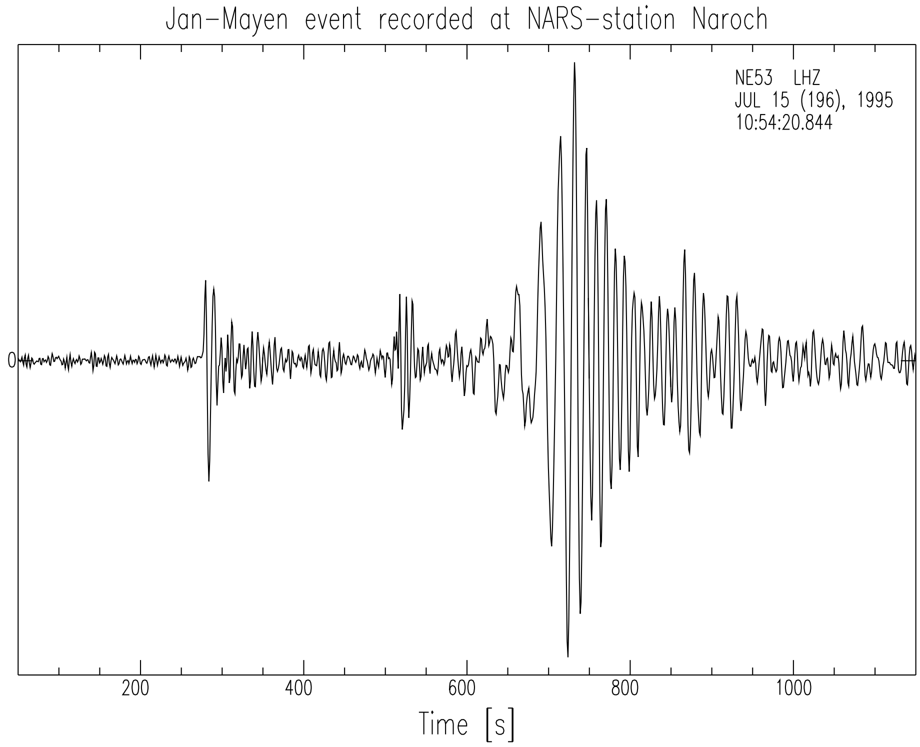


Figure 16.6: Vertical component of the ground motion at a seismic station in Naroch (Belarus) after an earthquake at Jan-Mayen island. This station is part of the Network of Autonomously Recording Seismographs (NARS) which is operated by Utrecht University.

group velocity of the low-frequency components of the surface wave is higher than the group-velocity of the high-frequency components. Hence it is ultimately the dispersion of the Rayleigh waves that causes the change in the apparent frequency of the surface wave arrival.

16.9 Leaky modes

The guided waves in the previous section decay exponentially with the distance to the low-velocity layer. Intuitively, the fact that the waves are confined to a region near a low-velocity layer can be understood as follows. Waves are refracted from regions of a high velocity to a region of low velocity. This means that the waves that stray out of the low-velocity channel are refracted back in the channel. Effectively this traps the waves near the vicinity of the channel. This explanation suggest that for a high-velocity channel the waves are refracted *away* from the channel. The resulting wave pattern will then correspond to waves that preferentially move away from the high velocity layer. For this reason we consider in this section the waves that propagate through the system shown in figure 16.5 but we will consider the case of a high-velocity layer where $c_1 > c_0$.

In this case, $1/c_1 < 1/c_0$, and we will first consider waves with a wave-number that is confined to the following interval: $\omega/c_1 < k < \omega/c_0$.

Problem a: Show that in this case the wave-number k_1 is imaginary and that it can be written as $k_1 = i\kappa_1$, with

$$\kappa_1 = \sqrt{k^2 - \frac{\omega^2}{c_1^2}} \ , \tag{16.85}$$

and show that the dispersion relation (16.82) is given by:

$$\tan i\kappa_1 H = \frac{-2k_0\kappa_1}{\kappa_1^2 - k_o^2} \ . \tag{16.86}$$

Problem b: Use the relation $\cos x = \left(e^{ix} + e^{-ix}\right)/2$ and the related expression for $\sin x$ to rewrite the dispersion relation (16.86) in the following form:

$$i \tanh \kappa_1 H = \frac{-2k_0\kappa_1}{\kappa_1^2 - k_o^2} \ . \tag{16.87}$$

In this expression all quantities are real when k is real. The factor i in the left hand side implies that this equation cannot be satisfied for real values of k. The only way in which the dispersion relation (16.87) can be satisfied is that k is complex. What does it mean that the wave-number is complex? Suppose that the dispersion relation is satisfied for a complex wave-number $k = k_r + ik_i$, with k_r and k_i the real and imaginary part. In the time domain a solution behaves for a fixed frequency as $U(k, z) \exp i\left(kx - \omega t\right)$. This means that for complex values of the wave-number the solution behaves as $U(k, z)e^{-k_i x} \exp i\left(k_r x - \omega t\right)$. This is a wave that propagates in the x-direction with phase velocity $c = \omega/k_r$ and that decays exponentially with the propagation distance x.

The exponential decay of the wave with the propagation distance x is due to the fact that the wave energy refracts out of the high-velocity layer. A different way of understanding this exponential decay is to consider the character of the wave-field outside the layer.

Problem c: Show that in the two half-spaces outside the high-velocity layer the waves propagate away from the layer. Hint: analyze the wave-number k_0 in the half-spaces and consider the corresponding solution in these regions.

This means that wave energy is continuously radiated away from the high-velocity layer. The exponential decay of the mode with propagation distance x is thus due to the fact that wave energy continuously leaks out of the layer. For this reason one speaks of *leaky modes*[64]. In the Earth a well-observed leaky mode is the S-PL wave. This is a mode where a transverse propagating wave in the mantle is coupled to a wave that is trapped in the Earth's crust.

In general there is no simple way to find the complex wave-number k for which the dispersion relation (16.87) is satisfied. However, the presence of leaky modes can be seen in figure 16.7 where the following function is shown in the complex plane:

$$F(k) \equiv 1/\left(i \tanh \kappa_1 H + \frac{2k_0\kappa_1}{\kappa_1^2 - k_o^2}\right) \ . \tag{16.88}$$

Problem d: Show that this function is infinite for the k-values that correspond to a leaky mode.

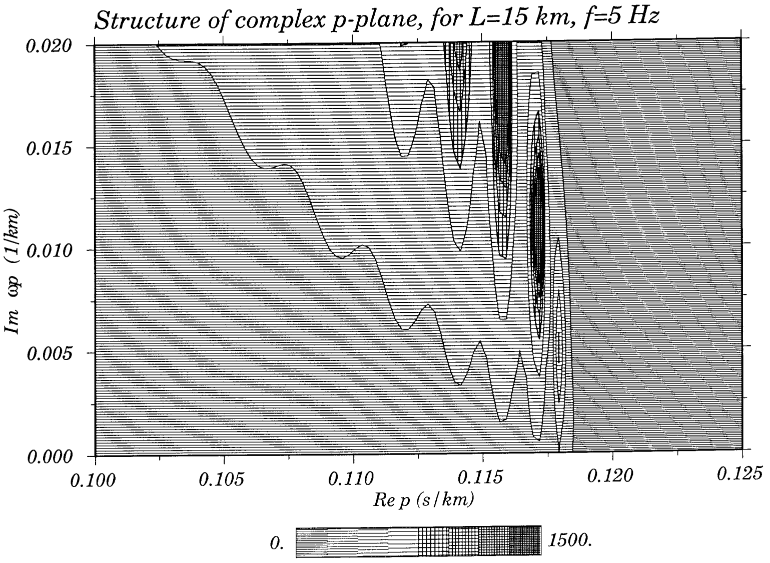


Figure 16.7: Contour diagram of the function $F(k)$ for a high-velocity layer with velocity $c_1 = 8.4km/s$ and a thickness $H = 15km$ that is embedded between two halfspaces with velocity $c_0 = 8km/s$, for waves with a frequency of $5Hz$. The horizontal axis is given by k_r/ω and the vertical axis by k_i.

The function $F(k)$ in figure 16.7 is computed for a high-velocity layer with a thickness of $15km$ and a velocity of $8.4km/s$ that is embedded between two half-spaces with a velocity of $8km/s$. The frequency of the wave is $5Hz$. In this figure, the horizontal axis is given by $\Re e\,(p) = k_r/\omega$ while the vertical axis is given by $\Im m\,(\omega p) = k_i$. The quantity p is called the slowness because $\Re e\,(p) = k_r/\omega = 1/c(\omega)$. The leaky modes show up in figure 16.7 as a number of localized singularities of the function $F(k)$.

Problem e: What is the propagation distance over which the amplitude of the mode with the lowest phase velocity decays with a factor $1/e$?

Leaky modes have been used by *Gubbins and Snieder*[26] to analyze waves that have propagated along a subduction zone. (A subduction zone is a plate in the Earth that slides downward in the mantle.) By a fortuitous geometry, compressive wave that are excited by earthquakes in the Tonga-Kermadec region that travel to a seismic station in Wellington (New Zealand) propagate for a large distance through the Tonga-Kermadec subduction zone. At the station in Wellington, a high-frequency wave arrives before the main compressive wave. This can be seen in figure 16.8 where such a seismogram is shown band-pass filtered at different frequencies. It can clearly be seen that the waves with a frequency around $6Hz$ arrive before the waves with a frequency around $1Hz$. This observation can be explained by the propagation of a leaky mode through the subduction zone. The physical reason that the high-frequency components arrive before the lower frequency components is that the high-frequency waves "fit" in the high-velocity layer in the subducting plate, whereas the lower frequency components do not fit in the high-velocity layer and are more influenced by the slower material outside the high-velocity layer. Of course, the energy leaks out of the high-velocity layer so that this arrival is very

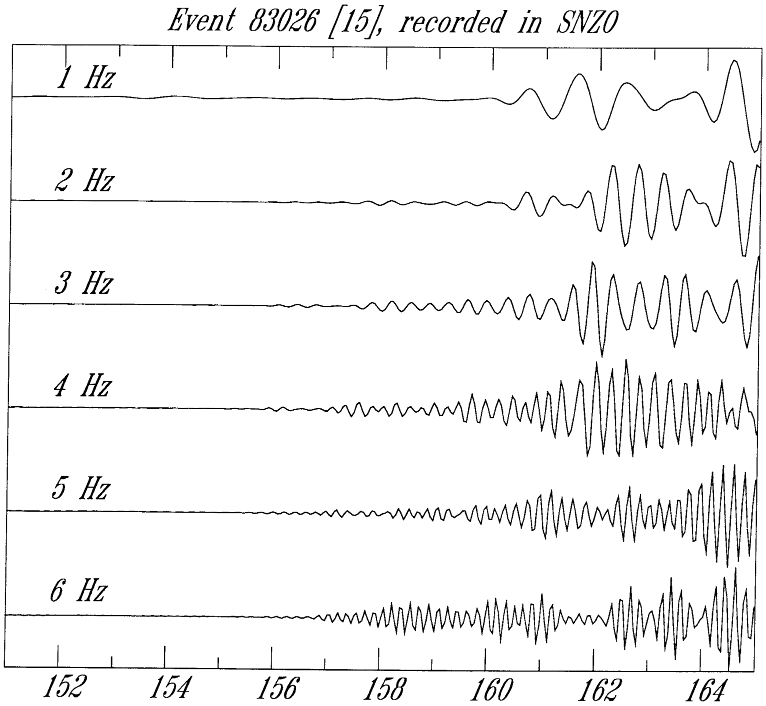


Figure 16.8: Seismic waves recorded in Wellington after an earthquake in the Tonga-Kermadec subduction zone, from Ref. [26]. The different traces correspond to the waves band-pass filtered with a center frequency indicated at each trace. The horizontal axis gives the time since the earthquake in seconds.

weak. From the data it could be inferred that in the subduction zone a high-velocity layer is present with a thickness between 6 and 10 km[26].

16.10 Radiation damping

Up to this point we have considered systems that are freely oscillating. When such systems are of finite extent, such a system displays undamped free oscillations. In the previous section leaky modes were introduced. In such a system, energy is radiated away, which leads to an exponentially of waves that propagate through the system. In a similar way, a system that has normal modes when it is isolated from its surroundings, can display damped oscillations when it is coupled to the external world.

As a simple prototype of such a system, consider a mass m that can move in the z-direction which is coupled to a spring with spring constant κ. The mass is attached to a string that is under a tension T and which has a mass ρ per unit length. The system is shown in figure 16.9. The total force acting on the mass is the sum of the force $-\kappa z$ exerted by the spring and the force F_s that is generated by the string:

$$m\ddot{z} + \kappa z = F_s \ , \tag{16.89}$$

where z denotes the vertical displacement of the mass. The motion of the waves that

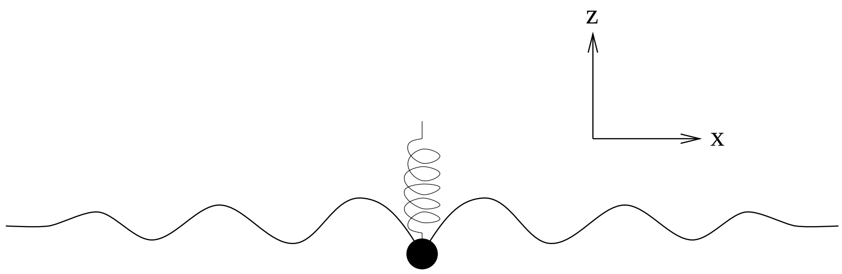


Figure 16.9: Geometry of an oscillating mass that is coupled to a spring.

propagate in the string is given by the wave equation:

$$u_{xx} - \frac{1}{c^2} u_{tt} = 0 \ , \tag{16.90}$$

where u is the displacement of the string in the vertical direction and c is given by

$$c = \sqrt{\frac{T}{\rho}} \ . \tag{16.91}$$

Let us first consider the case that no external force is present and that the mass is not coupled to the spring.

Problem a: Show that in that case the equation of motion is given by

$$\ddot{z} + \omega_0^2 z = 0 \ , \tag{16.92}$$

with ω_0 given by

$$\omega_0 = \sqrt{\frac{\kappa}{m}} \ . \tag{16.93}$$

One can say that the mass that is not coupled to the string has one free oscillation with angular frequency ω_0. The fact that the system has one free oscillation is a consequence from the fact that this mass can move only in the vertical direction, hence it has only one degree of freedom.

Before we couple the mass to the string let us first analyze the wave-motion in the string in the absence of the mass.

Problem b: Show that for any function $f(t - \frac{x}{c})$ satisfies the wave equation (16.90). Show that this function describes a wave that move in the positive x-direction with velocity c.

Problem c: Show that any function $g(t + \frac{x}{c})$ satisfies the wave equation (16.90). Show that this function describes a wave that move in the negative x-direction with velocity c.

The general solution is a superposition of the rightward and leftward moving waves:

$$u(x,t) = f(t - \frac{x}{c}) + g(t + \frac{x}{c}) \ . \tag{16.94}$$

This general solution is called the d'Alembert solution.

Now we want to describe the motion of the coupled system. Let us first assume that the mass oscillates with a prescribed displacement $z(t)$ and find the waves that this displacement generates in the string. We will consider here the situation that there are no waves moving towards the mass. This means that to the right of the mass the waves can only move rightward and to the left of the mass there are only leftward moving waves.

Problem d: Show that this radiation condition implies that the waves in the string are given by:

$$u(x,t) = \left\{ \begin{array}{ll} f(t-\frac{x}{c}) & for \quad x>0 \\ g(t+\frac{x}{c}) & for \quad x<0 \end{array} \right. \tag{16.95}$$

Problem e: At $x=0$ the displacement of the mass is the same as the displacement of the string. Show that this implies that $f(t)=g(t)=z(t)$, so that

$$u(x,t) = \left\{ \begin{array}{ll} z(t-\frac{x}{c}) & for \quad x>0 \\ z(t+\frac{x}{c}) & for \quad x<0 \end{array} \right. \tag{16.96}$$

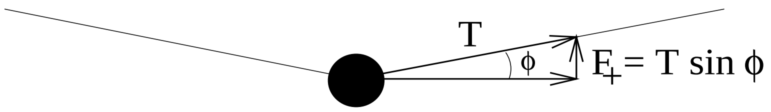


Figure 16.10: Sketch of the force exerted by the spring on the mass.

Now we have solved the problem of finding the wave motion in the string given the motion of the mass. To complete our description of the system we also need to specify how the motion of the string affects the mass. In other words, we need to find the force F_s in (16.89) given the motion of the string. This force can be derived from figure 16.10. The vertical component F_+ of the force acting on the mass from the right side of the string is given by $F_+ = T\sin\varphi$, where T is the tension in the string. When the motion in the spring is sufficiently weak we can approximate: $F_+ = T\sin\varphi \approx T\varphi \approx T\tan\varphi \approx Tu_x(x=0^+,t)$. In the last identity we used that the derivative $u_x(x=0^+,t)$ gives the slope of the string on the right of the point $x=0$.

Problem f: Use a similar reasoning to determine the force acting in the mass from the left part of the spring and show that the net force acting on the spring is given by

$$F_s(t) = T\left(u_x(x=0^+,t) - u_x(x=0^-,t)\right) \ , \tag{16.97}$$

where $u_x(x=0^-,t)$ is the x-derivative of the displacement in the string just to the left of the mass.

Problem g: Show that this expression implies that the net force that acts on the mass is equal to the *kink* in the spring at the location of the mass.

You may not feel comfortable with the fact the we used the approximation of a small angle φ in the derivation of (16.97). However, keep in mind that the wave equation (16.90) is derived using the same approximation and that this wave equation therefore is only valid for small displacements of the string.

At this point we have assembled all the ingredients for solving the coupled problem.

Problem h: Use (16.96) and (16.97) to derive that the force exerted by the spring on the mass is given by

$$F_s(t) = -\frac{2T}{c}\,\dot{z}\,, \tag{16.98}$$

and that the motion of the mass is therefore given by:

$$\ddot{z} + \frac{2T}{mc}\dot{z} + \omega_0^2 z = 0\,. \tag{16.99}$$

It is interesting to compare this expression for the motion of the mass that is coupled to the string with the equation of motion (16.92) for the mass that is not coupled to the string. The string leads to a term $\frac{2T}{mc}\dot{z}$ in the equation of motion that damps the motion of the mass. How can we explain this damping physically? When the mass moves, the string moves with it at location $x = 0$. Any motion in the string at that point will start propagating in the string. This means that the string radiates wave energy away from the mass whenever the mass moves. Since energy is conserved, the energy that is radiated in the string must be withdrawn from the energy of the moving mass. This means that the mass looses energy whenever it moves; this effect is described by the damping term in equation (16.99). This damping process is called *radiation damping*, because it is the radiation of waves that damps the motion of the mass.

The system described in this section is extremely simple. However, is does contain the essential physics of radiation damping. Many systems in physics that display normal modes are not quite isolated from their surroundings. Interactions of the system with their surrounding often lead to the radiation of energy, and hence to a damping of the oscillation of the system.

One example of such a system is an atom in an excited state. In the absence of external influences such an atom will not emit any light and decay. However, when such an atom can interact with electromagnetic fields, it can emit a photon and subsequently decay.

A second example is a charged particle that moves in a synchrotron. In the absence of external fields, such a particle will orbit forever in a circular orbit without any change in its speed. In reality, a charged particle is coupled to electromagnetic fields. This has the effect that a charged particle that is accelerated emits electromagnetic radiation, called synchrotron radiation[31]. The radiated energy means an energy loss of the particle, so that the particle slows down. This is actually the reason why accelerators such as used at CERN or Fermilab are so large. The acceleration of a particle in a circular orbit with radius r at the given velocity v is given by v^2/r. This means that for a fixed velocity v the larger the radius of the orbit is, the smaller the acceleration is, and the weaker the energy loss due to the emission of synchrotron radiation is. This is why one needs huge machines to accelerate tiny particles to an extreme energy.

Problem i: The modes in the plate in figure 16.1 are also damped because of radiation damping. What form of radiation is emitted by this oscillating plate?

Chapter 17

Potential theory

Potential fields play an important role in physics and geophysics because they describe the behavior of gravitational and electric fields as well as a number of other fields. Conversely, measurements of potential fields provide important information about the internal structure of bodies. For example, measurements of the electric potential at the Earth's surface when a current is send in the Earth gives information about the electrical conductivity while measurements of the Earth's gravity field or geoid provide information about the mass distribution within the Earth.

An example of this can be seen in figure 17.1 where the gravity anomaly over the northern part of the Yucatan peninsula in Mexico is shown[29]. The coast is visible as a thin white line. Note the clear ring-structure that is visible in the gravity signal. These rings have led to the discovery of the Chicxulub crater that was caused by the massive impact of a meteorite. Note that the diameter of the impact crater is about 150 km! This crater is presently hidden by thick layers of sediments, at the surface the only visible imprint of this crater is the presence of underground water-filled caves called "cenotes" at the outer edge of the crater. It is the measurement of the gravity field that made it possibly to find this massive impact crater.

The equation that the gravitational or electrical potential satisfies depends critically on the Laplacian of the potential. As shown in section 4.5 the gravitational field has the mass density as its source:

$$(\nabla \cdot \mathbf{g}) = -4\pi G\rho \qquad (4.17) \quad \textit{again}$$

The gravity field $\mathbf{g}$ is (minus) the gradient of the gravitational potential: $\mathbf{g} = -\nabla V$. This means that the gravitational potential satisfies the following partial differential equation:

$$\nabla^2 V(\mathbf{r}) = 4\pi G\rho \; . \tag{17.1}$$

This equation is called the Laplace equation, it is the prototype of the equations that occur in potential field theory. Note that the mathematical structure of the equations of the gravitational field and the electrical field is identical (compare the equations (4.12) and (4.17)), therefore the results derived in this chapter for the gravitational field can be used directly for the electrical field as well by replacing the mass density by the charge density and by making the following replacement:

$$\underbrace{4\pi G}_{\text{Gravity}} \Leftrightarrow \underbrace{-1/\varepsilon_0}_{\text{Electrostatics}} \quad . \tag{17.2}$$

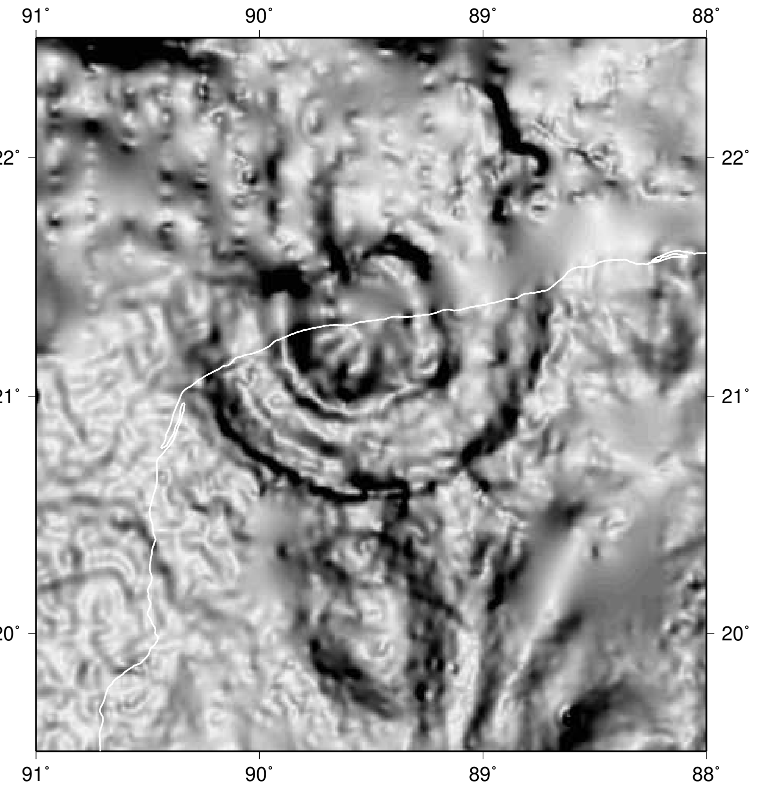


Figure 17.1: Gravity field over the Chicxulub impact crater on the northern coast of Yucatan (Mexico). The coastline is shown by a white line. The numbers along the vertical and horozontal axes refer to the latitutde and longitude respectively. The magnitude of the horizontal gradient of the Bouguer gravity anomaly is shown, details can be found in Ref. [29]. Courtesy of M. Pilkington and A.R. Hildebrand.

The theory of potentials fields is treated in great detail by *Blakeley*[10].

17.1 The Green's function of the gravitational potential

The Laplace equation (17.1) can be solved using a Green's function technique. In essence the derivation of the Green's function yields the well-known result that the gravitational potential for a point-mass m is given by $-Gm/r$. The use of Green's functions is introduced in great detail in chapter 14. The Green's function $G(\mathbf{r},\mathbf{r}')$ be the potential at location $\mathbf{r}$ generated by a point mass at location $\mathbf{r}'$ satisfies the following differential equation:

$$\nabla^2 G(\mathbf{r},\mathbf{r}') = \delta\left(\mathbf{r}-\mathbf{r}'\right) \ . \tag{17.3}$$

Take care not to confuse the Green's function $G(\mathbf{r},\mathbf{r}')$ with the gravitational constant G.

Problem a: Show that the solution of (17.1) is given by:

$$V(\mathbf{r}) = 4\pi G \int G(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}')dV' \ . \tag{17.4}$$

Problem b: The differential equation (17.3) has translational invariance, and is invariant for rotations. Show that this implies that $G(\mathbf{r}, \mathbf{r}') = G(|\mathbf{r} - \mathbf{r}'|)$. Show by placing the point mass in the origin by setting $\mathbf{r}' = 0$ that $G(r)$ satisfies

$$\nabla^2 G(r) = \delta(\mathbf{r}) \ . \tag{17.5}$$

Problem c: Express the Laplacian in spherical coordinates and show that for $r > 0$ equation (17.5) is given by

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial G(r)}{\partial r}\right) = 0 \ . \tag{17.6}$$

Problem d: Integrate this equation with respect to r to derive that the solution is given by $G(r) = A/r + Br$, where A and B are integration constants.

The constant B must be zero because the potential must remain finite as $r \to \infty$. The potential is therefore given by

$$G(r) = -\frac{A}{r} \ . \tag{17.7}$$

Problem e: The constant A can be found by integrating (17.5) over a sphere with radius R centered around the origin. Show that Gauss' theorem implies that $\int \nabla^2 G(r) dV = \oint \nabla G \cdot d\mathbf{S}$, use (**??**) in the right hand side of this expression and show that this gives $A = 1/4\pi$. Note that this result is independent of the radius R that you have used.

Problem f: Show that the Green's function is given by:

$$G(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi}\frac{1}{|\mathbf{r} - \mathbf{r}'|} \ . \tag{17.8}$$

With (17.4) this implies that the gravitational potential is given by:

$$V(\mathbf{r}) = -G \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' \ . \tag{17.9}$$

This general expression is useful for a variety of different purposes, we will make extensive use of it. By taking the gradient of this expression one obtains the gravitational acceleration $\mathbf{g}$. This acceleration was also derived in equation (6.5) of section 6.2 for the special case of a spherically symmetric mass distribution. Surprisingly it is a non-trivial calculation to derive (6.5) by taking the gradient of (17.9).

17.2 Upward continuation in a flat geometry

Suppose one has body with variable mass in two dimensions and that the mass density is only nonzero in the half-space $z < 0$. In this section we will determine the gravitational potential V above the half-space when the potential is specified at the plane $z = 0$ that forms the upper boundary of this body. The geometry of this problem is sketched in figure 17.2. This problem is of relevance for the interpretation of gravity measurements

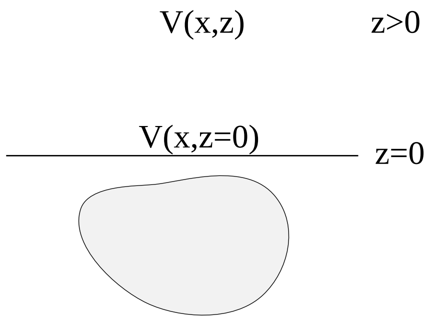


Figure 17.2: Geometry of the upward continuation problem. A mass anomaly (shaded) leaves an imprint on the potential at $z = 0$. The upward continuation problem states how the potential at the surface $z = 0$ is related to the potential $V(x, z)$ at greater height.

taken above the Earth's surface using aircraft or satellites orbits because the first step in this interpretation is to relate the values of the potential at the Earth's surface to the measurements taken above the surface. This process is called *upward continuation.*

Mathematically the problem can be stated this way. Suppose one is given the function $V(x, z = 0)$, what is the function $V(x, z)$? When we know that there is no mass above the surface it follows from (17.1) that the potential satisfies:

$$\nabla^2 V(\mathbf{r}) = 0 \qquad for \qquad z > 0 \; . \tag{17.10}$$

It is instructive to solve this problem by making a Fourier expansion of the potential in the variable x:

$$V(x, z) = \int_{-\infty}^{\infty} v(k, z) e^{ikx} dk \; . \tag{17.11}$$

Problem a: Show that for $z = 0$ the Fourier coefficients can be expressed in the known value of the potential at the edge of the half-space:

$$v(k, z = 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} V(x, z = 0) e^{-ikx} dx \; . \tag{17.12}$$

Problem b: Use Poisson's equation (17.10) and the Fourier expansion (17.11) to derive that the Fourier components of the potential satisfy for $z > 0$ the following differential equation:

$$\frac{\partial^2 v(k, z)}{\partial z^2} - k^2 v(k, z) = 0 \; . \tag{17.13}$$

Problem c: Show that the general solution of this differential equation can be written as $v(k, z) = A(k) \exp(+ |k| z) + B(k) \exp(- |k| z)$ and explain why the absolute values of the wave number k in the exponent can be taken .

Since the potential must remain finite at great height ($z \to \infty$) the coefficient $A(k)$ must be equal to zero. Setting $z = 0$ shows that $B(k) = v(k, z = 0)$, so that the potential is given by:

$$V(x, z) = \int_{-\infty}^{\infty} v(k, z = 0) e^{ikx} e^{-|k|z} dk \; . \tag{17.14}$$

This expression is interesting because it states that the different Fourier components of the potential decay as $\exp(- |k| z)$ with height.

Problem d: Explain that this implies that the short-wavelength components in the potential field decay must faster with height than the long-wavelength components.

The decrease of the Fourier components with the distance z to the surface is bad news when one wants to infer the mass-density in the body from measurements of the potential or from gravity at great height above the surface because the influence of mass perturbations on the gravitational field decays rapidly with height, the measurement of the short-wavelength component of the potential at great height therefore carries virtually no information about the density distribution within the Earth. This is the reason why gravity measurements from space are preferably carried out using satellites in low orbits rather than in high orbits. Similarly, for gravity surveys at sea a gravity meter has been developed that is towed far below the sea surface[70]. The idea is that by towing the gravity meter closer to the sea-bottom, the gravity signal generated at the sub-surface for short wavelength suffers less from the exponential decay due to upward continuation.

Problem e: Take the gradient of (17.14) to find the vertical component of the gravity field. Use the resulting expression to show that the gravity field $\mathbf{g}$ is less sensitive to the exponential decay due to upward continuation than the potential V.

This last result is the reasons why satellites in low orbits are used for measuring the Earth's gravitational potential and why for satellites in high orbits one measures gravity. In fact, presently a space-born gradiometer[50] is presently being developed. This instrument measures the *gradient* of the gravity vector by monitoring the differential motion between two masses in the satellite. Taking the gradient of the gravity leads to another factor of k in the Fourier expansion so that the effects of upward continuation are further reduced.

We will now explicitly express the potential at height z to the potential at the surface $z = 0$.

Problem f: Insert expression (17.12) in (17.14) to show that the upward continuation of the potential is given by:

$$V(x,z) = \int_{-\infty}^{\infty} H(x - x', z) V(x', z = 0) dx' \ , \tag{17.15}$$

with

$$H(x,z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-|k|z} e^{-ikx} dk \ . \tag{17.16}$$

Note that expression (17.15) has exactly the same structure as equation (11.57) of section 11.7 for a time-independent linear filter. The only difference is that the variable x now plays the role the role of the variable t in expression (11.57). This means that we can consider upward continuation as a linear filtering operation. The convolutional filter $H(x, z)$ maps the potential from the surface $z = 0$ onto the potential at height z.

Problem g: Show that this filter is given by:

$$H(x,z) = \frac{1}{\pi} \frac{z}{z^2 + x^2} \ . \tag{17.17}$$

Problem h: Sketch this filter as a function of x for a large value of z and a small value of z.

Problem i: Equation (17.15) implies that $H(x,z) = \delta(x)$, with $\delta(x)$ the Dirac delta function. Convince yourself of this by showing that $H(x,z)$ becomes more and more peaked round $x = 0$ when $z \to 0$ and by proving that for all values of z the filter function satisfies $\int_{-\infty}^{\infty} H(x,z)dx = 1$.

17.3 Upward continuation in a flat geometry in 3D

The analysis of the previous section is valid for a flat geometry in two dimensions. However,the theory can readily be extended to three dimensions by including another horizontal coordinate y in the derivation.

Problem a: Show that the theory of the previous section up to equation (17.16) can be generalized by carrying out a Fourier transform over both x and y. Show in particular that in three dimensions:

$$V(x,y,z) = \iint_{-\infty}^{\infty} H^{3D}(x-x', y-y', z)V(x', y', z=0)dx'dy' \; , \tag{17.18}$$

with

$$H^{3D}(x,y,z) = \frac{1}{(2\pi)^2} \iint_{-\infty}^{\infty} e^{-\sqrt{k_x^2+k_y^2}z} e^{-i(k_x x + k_y y)} dk_x dk_y \; . \tag{17.19}$$

The only difference with the case in two dimensions is that the integral (17.19) leads to a different upward continuation function than the integral (17.16) for the two-dimensional case. The integral can be solved by switching the k-integral to cylinder coordinates. The product $k_x x + k_y y$ can be written as $kr\cos\varphi$ where k and r are the length of the $\mathbf{k}$-vector and the position vector in the horizontal plane.

Problem b: Use this to show that H^{3D} can be written as:

$$H^{3D}(x,y,z) = \frac{1}{(2\pi)^2} \int_0^{\infty} \int_0^{2\pi} ke^{-kz} e^{-ikr\cos\varphi} d\varphi dk \; . \tag{17.20}$$

Problem c: The Bessel function has the following integral representation:

$$J_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{ix\sin\theta} d\theta \; . \tag{17.21}$$

Use this result to write the upward continuation filter as:

$$H^{3D}(x,y,z) = \frac{1}{2\pi} \int_0^{\infty} e^{-kz} J_0(kr) k dk \; . \tag{17.22}$$

It appears that we have only made the problem more complex, because the integral of the Bessel function is not trivial. Fortunately, books and tables exist with a bewildering collection of integrals. For example, in equation (6.621.1) of *Gradshteyn and Ryzhik*[25] you can find an expression for the following integral: $\int_0^{\infty} e^{-\alpha x} J_\nu(\beta x) x^{\mu-1} dx$.

Problem d: What are the values of α, ν, β and μ if we would like to use this integral to solve the integration in (17.22)?

Take a look at the form of integral in *Gradshteyn and Ryzhik*[25]. You will probably be discouraged by what you find because the result is expressed in hypergeometric functions, which means that you now have the new problem to find out what these functions are. There is however a way out, because in expression (6.611.1) of [25] gives the following integral:

$$\int_0^\infty e^{-\alpha x} J_\nu(\beta x) dx = \frac{\beta^{-\nu}\left\{\sqrt{\alpha^2+\beta^2}-\alpha\right\}^\nu}{\sqrt{\alpha^2+\beta^2}} \ . \tag{17.23}$$

This is not quite the integral that we want, because it does not contain a term that corresponds to the factor k in the integral (17.22). However, we can introduce such a factor by differentiating expression (17.23) with respect to α.

Problem e: Do this to show that the upward continuation operator is given by

$$H^{3D}(x,y,z) = \frac{1}{2\pi}\frac{z}{(x^2+y^2+z^2)^{3/2}} \ . \tag{17.24}$$

You can make the problem simpler by first inserting the appropriate value of ν in (17.23).

Problem f: Compare the upward continuation operator (17.24) for three-dimensions with the corresponding operator for two dimensions in equation (17.17). Which of these operators decays more rapidly as a function of the horizontal distance? Can you explain this difference physically?

Problem g: In section 17.2 you showed that the integral of the upward continuation operator over the horizontal distance is equal to one. Show that the same holds in three dimensions, i.e. show that $\iint_{-\infty}^{\infty} H^{3D}(x,y,z)dxdy = 1$. The integration simplifies by using cylinder coordinates.

Comparing the upward continuation operators in different dimensions one finds that these operators are different functions of the horizontal distance in the space domain. However, a comparison for (17.16) with (17.19) shows that in the wave-number domain the upward continuation operators in two and three dimensions have the same dependence on wave-number.. The same is actually true for the Green's function of the wave equation in 1,2 or 3 dimensions. As you can see in section 15.4 these Green's functions are very different in the space domain, but one can show that in the wave-number domain they are in each dimension given by the same expression.

17.4 The gravity field of the Earth

In this section we obtain an expression for the gravitational potential outside the Earth for an arbitrary distribution of the mass density $\rho(\mathbf{r})$ within the Earth. This could be done be using the Green's function that is appropriate for the Laplace equation (17.1). As an alternative we will solve the problem here by expanding both the mass density and the potential in spherical harmonics and by using a Green's function technique for every component in the spherical harmonics expansion separately.

When using spherical harmonics, the natural coordinate system is a system of spherical coordinates. For every value of the radius r both the density and the potential can be expanded in spherical harmonics:

$$\rho(r,\theta,\varphi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\rho_{lm}(r)Y_{lm}(\theta,\varphi)\ , \tag{17.25}$$

and

$$V(r,\theta,\varphi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}V_{lm}(r)Y_{lm}(\theta,\varphi)\ . \tag{17.26}$$

Problem a: Show that the expansion coefficients for the density are given by:

$$\rho_{lm}(r)=\int Y_{lm}^{*}(\theta,\varphi)\rho(r,\theta,\varphi)d\Omega\ , \tag{17.27}$$

where $\int(\cdots)\,d\Omega$ denotes an integration over the unit sphere.

Equation (17.1) for the gravitational potential contains the Laplacian. The Laplacian in spherical coordinates can be decomposed as

$$\nabla^2=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\nabla_1^2\ , \tag{17.28}$$

with ∇_1^2 the Laplacian on the unit sphere:

$$\nabla_1^2=\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}\ . \tag{17.29}$$

The reason why an expansion in spherical harmonics is used for the density and the potential is that the spherical harmonics are the eigenfunctions of the operator ∇_1^2 (see p. 379 of *Butkov*[14]):

$$\nabla_1^2Y_{lm}=l\,(l+1)\,Y_{lm}\ . \tag{17.30}$$

Problem b: Insert the expansions (17.25) and (17.26) in the Laplace equation, and use (17.28) and (17.29) for the Laplacian of the spherical harmonics to show that the expansion coefficients $V_{lm}(r)$ of the potential satisfy the following differential equation:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V_{lm}(r)}{\partial r}\right)-\frac{l\,(l+1)}{r^2}V_{lm}(r)=4\pi G\rho_{lm}(r)\ . \tag{17.31}$$

What we have gained by making the expansion in spherical harmonics is that (17.31) is an ordinary differential equation in the variable r whereas the original equation (17.1) is a partial differential equation in the variables r, θ and φ. The differential equation (17.31) can be solved using the Green's function technique that is described in section 14.3. Let us first consider a mass $\delta(r-r')$ located at a radius r' only. The response to this mass is the Green's function G_l that satisfies the following differential equation:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial G_l(r,r')}{\partial r}\right)-\frac{l\,(l+1)}{r^2}G_l(r,r')=\delta\left(r-r'\right)\ . \tag{17.32}$$

Note that this equation depends on the angular order l but not on the angular degree m. For this reason the Green's function $G_l(r,r')$ depends on l but not on m.

Problem c: The Green's function can be found by first solving the differential equation (17.32) for $r \neq r'$. Show that when $r \neq r'$ the general solution of the differential equations (17.32) can be written as $G_l = Ar^l + Br^{-(l+1)}$, where the constants A and B do not depend on r.

Problem d: In the regions $r < r'$ and $r > r'$ the constants A and B will in general have different values. Show that the requirement that the potential is everywhere finite implies that $B = 0$ for $r < r'$ and that $A = 0$ for $r > r'$. The solution can therefore be written as:

$$G_l(r,r')=\begin{cases} Ar^l & \text{for} \quad r<r' \\ Br^{-(l+1)} & \text{for} \quad r>r' \end{cases} \tag{17.33}$$

The integration constants follow in the same way as in the analysis of section 14.3. One constraint on the integration constants follows from the requirement than the Green's function is continuous in the point $r = r'$. The other constraint follows by multiplying (17.32) by r^2 and integrating the resulting equation over r from $r' - \varepsilon$ to $r' + \varepsilon$.

Problem e: Show that by taking the limit $\varepsilon \to 0$ this leads to the requirement

$$\left[r^2\frac{\partial G_l(r,r')}{\partial r}\right]_{r=r'-\varepsilon}^{r=r'+\varepsilon}=r'^2\ . \tag{17.34}$$

Problem f: Use this condition with the continuity of G_l to find the coefficients A and B and show that the Green's function is given by:

$$G_l(r,r')=\begin{cases} -\dfrac{1}{(2l+1)}\dfrac{r^l}{r'^{(l-1)}} & \text{for} \quad r<r' \\ -\dfrac{1}{(2l+1)}\dfrac{r'^{(l+2)}}{r^{(l+1)}} & \text{for} \quad r>r' \end{cases} \tag{17.35}$$

Problem g: Use this result to derive that the solution of equation (17.31) is given by:

$$\begin{aligned} V_{lm}(r) = \; & -\frac{4\pi G}{(2l+1)}\frac{1}{r^{l+1}}\int_0^r \rho_{lm}(r')r'^{(l+2)}dr' \\ & -\frac{4\pi G}{(2l+1)}r^l\int_r^\infty \rho_{lm}(r')\frac{1}{r'^{(l-2)}}dr'\ . \end{aligned} \tag{17.36}$$

Hint: split the integration over r' up the interval $0 < r' < r$ and the interval $r' > r$.

Problem h: Let us now consider the potential outside the Earth. Let us denote the radius of the Earth with the symbol a. Use the above expression to show that the potential outside the Earth is given by

$$V(r,\theta,\varphi) = -\sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{4\pi G}{(2l+1)} \frac{1}{r^{l+1}} \int_0^a \rho_{lm}(r') r'^{(l+2)} dr' \; Y_{lm}(\theta,\varphi) \; . \tag{17.37}$$

Problem i: Eliminate ρ_{lm} and show that the potential is finally given by:

$$V(r,\theta,\varphi) = -\sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{4\pi G}{(2l+1)} \frac{1}{r^{l+1}} \int_0^a \rho(r',\theta',\varphi') r'^{l} Y^*_{lm}(\theta',\varphi') dV' \; Y_{lm}(\theta,\varphi) \; . \tag{17.38}$$

Note that the integration in this expression is over the volume of the Earth rather than over the distance r' to the Earth's center.

Let us reflect on the relation between this result and the derivation of upward continuation in a Cartesian geometry of the section 17.2. Equation (17.38) can be compared with equation (17.14). In (17.14) the potential is written as an integration over wave-number and the potential is expanded in basis functions $\exp ikx$, whereas in (17.38) the potential is written as a summation over the degree l and order m and the potential is expanded in basis functions Y_{lm}. In both expressions the potential is written as a sum over basis functions with increasingly shorter wavelength as the summation index l or the integration variable k increases. This means that the decay of the potential with height is in both geometries faster for a potential with rapid horizontal variations than for a potential with smooth horizontal variations. In both cases the potential decreases when the height z (or the radius r) increases. In a flat geometry the potential decreases as $\exp(-|k|\,z)$ whereas in a spherical geometry the potential decreases as $r^{-(l+1)}$. This difference in the reduction of the potential with distance is due to the difference in the geometry in the two problems.

The expressions (17.9) and (17.38) both express the gravitational potential due a density distribution $\rho(\mathbf{r})$, therefore these expressions must be identical.

Problem j: Use the equivalence of these expressions to derive that for $r' < r$ the following identity holds:

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} \frac{4\pi}{(2l+1)} Y_{lm}(\theta,\varphi) Y^*_{lm}(\theta',\varphi') \frac{r'^{l}}{r^{l+1}} \; . \tag{17.39}$$

The derivation of this section could also have been made using (17.39) as a starting point because this expression can be derived by using the *generating function* Legendre polynomials and by using the addition theorem for obtaining the m-summation. However, these concepts are not needed in the treatment of this section that is based only on the expansion of functions in spherical harmonics and on the use of Green's functions.

As a last exercise let us consider the special case of a spherically symmetric mass distribution: $\rho = \rho(r)$. For such a mass distribution the potential is given by

$$V(\mathbf{r}) = -\frac{GM}{r} \; , \tag{17.40}$$

where M is the total mass of the body. The gradient of this potential is indeed equal to the gravitational acceleration given in expression (6.5) for a spherically symmetric mass M.

Problem i: Derive the potential (17.40) from (17.38) by considering the special case that the mass-density depends only on the radius.

17.5 Dipoles, quadrupoles and general relativity

We have seen in the section 6.2 that a spherically symmetric mass leads to a gravitational field $\mathbf{g}(\mathbf{r}) = -GM\hat{\mathbf{r}}/r^2$, which corresponds to a gravitational potential $V(\mathbf{r}) = -GM/r$. Similarly, the electric potential due to a spherically symmetric charge distribution is given by $V(\mathbf{r}) = q/4\pi\varepsilon_0 r$, where q is the total charge. In this section we investigate what happens if we place a positive charge and a negative charge close together. (Since there is no negative mass, we treat for the moment the electrical potential, but we will see in section 17.6 that the results also have a bearing on the gravitational potential.)

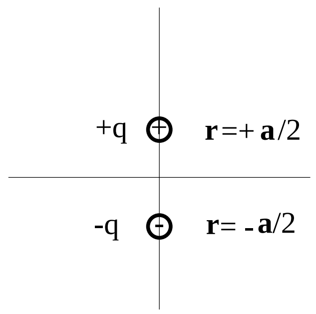


Figure 17.3: Two opposite charges that constitute an electric dipole.

Consider the case that a positive charge $+q$ is placed at position $\mathbf{a}/2$ and a negative charge $-q$ is placed at position $-\mathbf{a}/2$.

Problem a: The total charge of this system is zero. What would you expect the electric potential to be at positions that are very far from the charges compared to potential for a single point charge?

Problem b: The potential follows by adding the potentials for the two point charges. Show that the electric potential generated by these two charges is given by

$$4\pi\varepsilon_0 V(\mathbf{r}) = \frac{q}{|\mathbf{r}-\mathbf{a}/2|} - \frac{q}{|\mathbf{r}+\mathbf{a}/2|} \ . \tag{17.41}$$

Problem c: Ultimately we will place the charges very close to the origin by taking the limit $a \to 0$. We can therefore restrict our attention to the special case that $a \ll r$. Use a first order Taylor expansion to show that up to order a:

$$\frac{1}{|\mathbf{r}-\mathbf{a}/2|} = \frac{1}{r} - \frac{1}{2r^3}(\mathbf{r}\cdot\mathbf{a}) \ . \tag{17.42}$$

Problem d: Insert this in (17.41) and derive that the electric potential is given by:

$$4\pi\varepsilon_0 V(\mathbf{r}) = -\frac{q(\mathbf{r}\cdot\mathbf{a})}{r^3} \tag{17.43}$$

Now suppose we bring the charges in figure 17.3 closer and closer together, and suppose we let the charge q increase so that the product $\mathbf{p} = q\mathbf{a}$ is constant, then the electric potential is given by:

$$4\pi\varepsilon_0 V(\mathbf{r}) = -\frac{(\hat{\mathbf{r}}\cdot\mathbf{p})}{r^2}\ , \tag{17.44}$$

where we used that $\mathbf{r} = r\hat{\mathbf{r}}$. The vector $\mathbf{p}$ is called the *dipole vector*. We will see in the next section how the dipole vector can be defined for arbitrary charge or mass distributions.

In **problem a** you might have guessed that the electric potential would go to zero at great distance. Of course the potential due to the combined charges goes to zero much faster than the potential due to a single charge only, but note that the electric potential vanishes as $1/r^2$ compared to the $1/r$ decay of the potential for a single charge. Many physical systems, such as neutral atoms, consists of neutral combinations of positive and negative charges. The lesson we learn from equation (17.43) is that such a neutral combination of charges does generate a nonzero electric field and that such a system will in general interact with other electromagnetic systems. For example, atoms interact to leading order with the radiation (light) field through their *dipole moment*[52]. In chemistry, the dipole moment of molecules plays a crucial role in the distinction between polar and apolar substances. Water would not have its many wonderful properties if it would not have a dipole moment.

Let us now consider the electric field generated by an electric dipole.

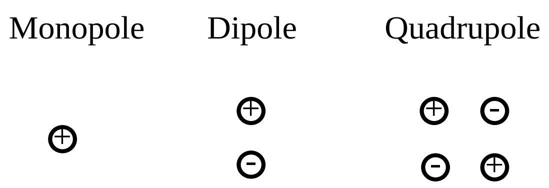
Problem e: Take the gradient of (17.44) to show that this field is given by

$$\mathbf{E}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0 r^3}\left(\mathbf{p} - 3\hat{\mathbf{r}}\left(\hat{\mathbf{r}}\cdot\mathbf{p}\right)\right)\ . \tag{17.45}$$

Hint, either use the expression of the gradient in spherical coordinates or take the gradient in Cartesian coordinates and use (4.7).

The electric field generated by an electric dipole has the same form as the magnetic field generated by a magnetic dipole as shown in expression (4.3) of section (4.1). The mathematical reason for this is that the magnetic field satisfies equation (4.13) which states that $(\nabla\cdot\mathbf{B}) = 0$ while the electric field in free space satisfies according to equation (4.12) the same field equation: $(\nabla\cdot\mathbf{E}) = 0$. However, there is an important difference, the electric field is generated by electric charges. In general, this field satisfies the equation $(\nabla\cdot\mathbf{E}) = \rho(\mathbf{r})/\varepsilon_0$. In the example of this section we created a dipole field by taking two opposite charges and putting them closer and closer together. However, the magnetic field satisfies $(\nabla\cdot\mathbf{B}) = 0$ *everywhere*. The reason for this is that the magnetic equivalent of electric charge, the *magnetic monopole,* has not been discovered in nature.

Problem f: The fact that magnetic monopoles have not been observed in nature seems puzzling, because we have seen that the magnetic dipole field has the same form as the electric dipole field that was constructed by putting two opposite electric charges close together. Can you think of a physical system or device that does generate the magnetic dipole field but that does not consist of two magnetic monopoles placed closely together?



.

Figure 17.4: The definition of the monopole, dipole and quadrupole in terms of electric charges

We have seen now the electric field of a single charge, this is called the monopole, see figure 17.4. The field of this charge decays as $1/r^2$. If we put two opposite charges close together we have seen we can create a dipole, see figure 17.4, as derived in expression (17.45) this field decays as $1/r^3$. We can also put two opposite dipoles together as shown in figure 17.4. The resulting charge distribution is called a quadrupole. To leading order the electric fields of the dipoles that constitute the quadrupole cancel, and we will see in the next section that the electric potential for quadrupole decays as $1/r^3$ so that the field decays with distance as $1/r^4$.

You may wonder whether the concept of a dipole or quadrupole can be used as well for the gravity field because these concept are for the electric field based on the presence of both positive and negative charges whereas we know that only positive mass occurs in nature. However, there is nothing that should keep us from computing the gravitational field for a negative mass, and this is actually quite useful. As an example, let us consider a double star that consists of two heavy stars that rotate around their joint center of gravity. The first order field is the monopole field that is generated by the joint mass of the stars. However, as shown in figure 17.5 the mass of the two stars can approximately

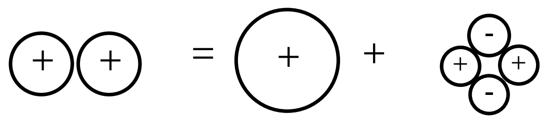


Figure 17.5: The decomposition of a double star in a gravitational monopole and a gravitational quadrupole.

be seen as the sum of a monopole and a quadrupole consisting of two positive and two negative charges. Since the stars rotate, the gravitational quadrupole rotates as well, and this is the reason why rotating double stars are seen as a prime source for the generation of *gravitational waves*[42]. However, gravitational waves that spread in space with time cannot be described by the classic expression (17.1) for the gravitational potential.

Problem g: Can you explain why (17.1) cannot account for propagating gravitational waves?

A proper description of gravitational waves depends on the general theory of relativity [42]. This is the reason why huge detectors for gravitational waves are being developed, because these waves can be used to investigate the theory of general relativity.

17.6 The multipole expansion

Now that we have learned that the concepts of monopole, dipole and quadrupole are relevant for both the electric field and the gravity field we will continue the analysis with the gravitational field. In this section we will derive the *multipole expansion* where the total field is written as a superposition of a monopole field, a dipole field, a quadrupole field, an octupole field, etc.

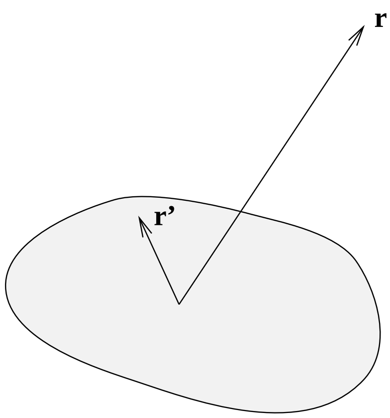


Figure 17.6: Definition of the integration variable $\mathbf{r}'$ within the mass and the observation point $\mathbf{r}$ outside the mass.

Consider the situation shown in figure 17.6 where a finite body has a mass density $\rho(\mathbf{r}')$. The gravitational potential generated by this mass is given by:

$$V(\mathbf{r}) = -G \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} dV' \qquad (17.9) \quad \text{again.}$$

We will consider the potential at a distance that is much larger than the size of the body. Since the integration variable $\mathbf{r}'$ is limited by the size of the body a "large distance" means in this context that $r \gg r'$. We will therefore make a Taylor expansion of the term $1/|\mathbf{r}-\mathbf{r}'|$ in the small parameter (r'/r) which is much smaller than unity.

Problem a: Show that

$$|\mathbf{r}-\mathbf{r}'| = \sqrt{r^2 - 2(\mathbf{r}\cdot\mathbf{r}') + r'^2} \qquad (17.46)$$

Problem b: Use a Taylor expansion in the small parameter r'/r to show that:

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \frac{1}{r}\left\{1 + \frac{1}{r}(\hat{\mathbf{r}}\cdot\mathbf{r}') + \frac{1}{2r^2}\left(3(\hat{\mathbf{r}}\cdot\mathbf{r}')^2 - r'^2\right) + O\left(\frac{r'}{r}\right)^3 .\right\} \qquad (17.47)$$

Be careful that you properly account for all the terms of order r' correctly. Also be aware of the distinction between the difference between the position vector $\mathbf{r}$ and the unit vector $\hat{\mathbf{r}}$.

From this point on we will ignore the terms of order $(r'/r)^3$.

Problem c: Insert the expansion (17.47) in (17.9) and show that the gravitational potential can be written as a sum of different contributions:

$$V(\mathbf{r}) = V_{mon}(\mathbf{r}) + V_{dip}(\mathbf{r}) + V_{qua}(\mathbf{r}) + \cdots \ , \tag{17.48}$$

with

$$V_{mon}(\mathbf{r}) = -\frac{G}{r}\int \rho(\mathbf{r}')dV' \ , \tag{17.49}$$

$$V_{dip}(\mathbf{r}) = -\frac{G}{r^2}\int \rho(\mathbf{r}')\left(\hat{\mathbf{r}}\cdot\mathbf{r}'\right)dV' \ , \tag{17.50}$$

$$V_{qua}(\mathbf{r}) = -\frac{G}{2r^3}\int \rho(\mathbf{r}')\left(3\left(\hat{\mathbf{r}}\cdot\mathbf{r}'\right)^2 - r'^2\right)dV' \ . \tag{17.51}$$

It follows that the gravitational potential can be written as sum of terms that decay with increasing powers of r^{-n}. Let us analyze these terms in turn. The term $V_{mon}(\mathbf{r})$ in (17.49) is the simplest since the volume integral of the mass density is simply the total mass of the body: $\int \rho(\mathbf{r}')dV' = M$. This means that this term is given by

$$V_{mon}(\mathbf{r}) = -\frac{GM}{r} \ . \tag{17.52}$$

This is the potential generated by a point mass M. To leading order, the gravitational field is the same as if all the mass of the body would be concentrated in the origin. The mass distribution within the body does not affect this part of the gravitational field at all. Because the resulting field is the same as for a point mass, this field is called the *monopole field.*

For the analysis of the term $V_{dip}(\mathbf{r})$ in (17.50) is us useful to define the center of gravity $\mathbf{r}_g$ of the body:

$$\mathbf{r}_g \equiv \frac{\int \rho(\mathbf{r}')\mathbf{r}'dV'}{\int \rho(\mathbf{r}')dV'} \tag{17.53}$$

This is simply a weighted average of the position vector with the mass density as weight functions. Note that the word "weight" here has a double meaning!

Problem d: Show that $V_{dip}(\mathbf{r})$ is given by:

$$V_{dip}(\mathbf{r}) = -\frac{GM}{r^2}\left(\hat{\mathbf{r}}\cdot\mathbf{r}_g\right) \ . \tag{17.54}$$

Note that this potential has exactly the same form as the potential (17.44) for an electric dipole. For this reason $V_{dip}(\mathbf{r})$ is called the dipole term, we will refer to this term also as the "dipole term."

Problem e: Compared to the monopole term, the dipole term decays as $1/r^2$ rather than $1/r$. The monopole term does not depend on $\hat{\mathbf{r}}$, the direction of observation. Show that the dipole term varies with the direction of observation as $\cos\theta$ and show how the angle θ must be defined.

Problem f: You may be puzzled by the fact that the gravitational potential contains a dipole term, despite the fact that there is no negative mass. Draw a figure similar to figure 17.4 to show that a displaced mass can be written as an undisplaced mass plus a mass-dipole.

Of course, one is completely free in the choice of the origin of the coordinate system. If one chooses the origin in the center of mass of the body, then $\mathbf{r}_g = 0$ and the dipole term vanishes.

We will now analyze the term $V_{qua}(\mathbf{r})$ in (17.51). It can be seen from this expression that this term decays with distance as $1/r^3$. For this reason, this term will be called the quadrupole term. The dependence of the quadrupole term on the direction is more complex than for the monopole term and the dipole term. In doing so it will be useful to use the double contraction between two tensors. This operation is defined as:

$$(\mathbf{A} : \mathbf{B}) \equiv \sum_{i,j} A_{ij} B_{ij} \ . \tag{17.55}$$

The double contraction generalizes the concept of the inner product of two vectors $(\mathbf{a} \cdot \mathbf{b}) = \sum_i a_i b_i$ to matrices or tensors of rank two. A double contraction occurs for example in the following identity: $1 = (\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}) = (\hat{\mathbf{r}} \cdot \mathbf{I}\hat{\mathbf{r}}) = (\hat{\mathbf{r}}\hat{\mathbf{r}} : \mathbf{I})$, where $\mathbf{I}$ is the identity operator. Note that the term $\hat{\mathbf{r}}\hat{\mathbf{r}}$ is a dyad. If you are unfamiliar with the concept of a dyad you may first want to look at section 10.1.

Problem g: Use these results to show that $V_{qua}(\mathbf{r})$ can be written as:

$$V_{qua}(\mathbf{r}) = -\frac{G}{2r^3} \left(\hat{\mathbf{r}}\hat{\mathbf{r}} : \mathbf{T} \right) \ , \tag{17.56}$$

where $\mathbf{T}$ is the *quadrupole moment tensor* defined as

$$\mathbf{T} = \int \rho(\mathbf{r}) \left(3\mathbf{r}\mathbf{r} - \mathbf{I}r^2 \right) dV \ . \tag{17.57}$$

Note that we renamed the integration variable $\mathbf{r}'$ in the quadrupole moment tensor as $\mathbf{r}$.

Problem h: Show that $\mathbf{T}$ is in explicit matrix notation given by:

$$\mathbf{T} = \int \rho(\mathbf{r}) \begin{pmatrix} 2x^2 - y^2 - z^2 & 3xy & 3xz \\ 3xy & 2y^2 - x^2 - z^2 & 3yz \\ 3xz & 3yz & 2z^2 - x^2 - y^2 \end{pmatrix} dV \ . \tag{17.58}$$

Note the resemblance between (17.56) and (17.54). For the dipole term the directional dependence is described by the single contraction $(\hat{\mathbf{r}} \cdot \mathbf{r}_g)$ whereas for the quadrupole term directional dependence is now given by the double contraction $\hat{\mathbf{r}}\hat{\mathbf{r}} : \mathbf{T}$. This double contraction leads to a more rapid angular dependence of the quadrupole term than for the monopole term and the dipole term.

To find the angular dependence, we will use that the inertia tensor $\mathbf{T}$ is a real symmetric 3×3 matrix. This matrix has therefore three orthogonal eigenvectors $\hat{\mathbf{v}}^{(i)}$ with corresponding eigenvalues λ_i. Using expression (10.61) of section (10.5) this implies that the quadrupole moment tensor can be written as:

$$\mathbf{T} = \sum_{i=1}^{3} \lambda_i \hat{\mathbf{v}}^{(i)} \hat{\mathbf{v}}^{(i)} . \tag{17.59}$$

Problem i: Use this result to show that the quadrupole term can be written as

$$V_{qua}(\mathbf{r}) = -\frac{G}{2r^3} \sum_{i=1}^{3} \lambda_i \cos^2 \Psi_i , \tag{17.60}$$

where the Ψ_i denote the angle between the eigenvector $\hat{\mathbf{v}}^{(i)}$ and the observation direction $\hat{\mathbf{r}}$, see figure 17.7 for the definition of these angles.

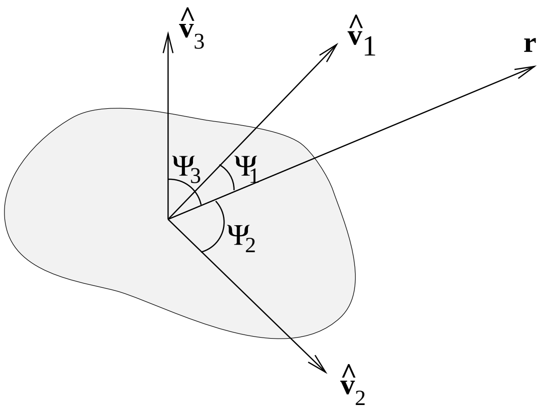


Figure 17.7: Definition of the angles Ψ_i.

Since the dependence of these angles goes as $\cos^2\Psi_i = (\cos 2\Psi_i + 1)/2$ this implies that the quadrupole varies through two periods when Ψ_i goes from 0 to 2π. This contrast with the monopole term, which does not depend on the direction, as well as with the dipole term that varies according to **problem e** as $\cos\theta$. There is actually a close connection between the different terms in the multipole expansion and spherical harmonics. This can be seen by comparing the multipole terms (17.49)-(17.51) with expression (17.38) for the gravitational potential. In the latter expression, the different terms decay with distance as $r^{-(l+1)}$ and have an angular dependence $Y_{lm}(\theta, \varphi)$. Similarly, the multipole terms decay as r^{-1}, r^{-2} and r^{-3} respectively and depend on the direction as $\cos 0$, $\cos\theta$ and $\cos 2\Psi$ respectively.

17.7 The quadrupole field of the Earth

Let us now investigate what the multipole expansion implies for the gravity field of the Earth. The monopole term is by far the dominant term. It explains why an apple falls from a tree, why the moon orbits the Earth and most other manifestations of gravity that we observe in daily life. The dipole term has in this context no physical meaning whatsoever. This can be seen from equation (17.54) which states that the dipole term only depends on the distance from the Earth's center of gravity to the origin of the coordinate system. Since we are completely free in choosing the origin, the dipole term can be made to vanish by choosing the origin of the coordinate system in the Earth's center of gravity. It is through the quadrupole term that some of the subtleties of the Earth's gravity field becomes manifest.

Problem a: The quadrupole term vanishes when the mass distribution in the Earth is spherically symmetric Earth. Show this by computing the inertia tensor $\mathbf{T}$ when $\rho = \rho(r)$.

The dominant departure of the Earth's figure from aspherical shape is the flattening of the Earth due to the rotation of the Earth. If that is the case, then by symmetry one eigenvector of $\mathbf{T}$ must be aligned with the Earth's axis of rotation, and the two other eigenvectors are perpendicular to the axis of rotation. By symmetry these other eigenvectors must correspond to equal eigenvalues. When we choose a coordinate system with the z-axis along the Earth's axis of rotation the eigenvectors are therefore given by the unit-vectors $\hat{\mathbf{z}}$, $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ with eigenvalues λ_z, λ_x and λ_y respectively. These last eigenvalues identical because of the rotational symmetry around the Earth's axis of rotation, hence $\lambda_y = \lambda_x$.

Let us first determine the eigenvalues. Once the eigenvalues are known, the quadrupole moment tensor follows from (17.60). The eigenvalues could be found in standard way by solving the equation $\det(\mathbf{T}-\lambda\mathbf{I})$, but this is unnecessarily difficult. Once we know the eigenvectors the eigenvalues can easily be found from expression (17.59).

Problem b: Take twice the inner product of (17.59) with the eigenvector $\hat{\mathbf{v}}^{(j)}$ to show that

$$\lambda_j = \hat{\mathbf{v}}^{(j)} \cdot \mathbf{T} \cdot \hat{\mathbf{v}}^{(j)} \ . \tag{17.61}$$

Problem c: Use this with expression (17.58) to show that the eigenvalues are given by:

$$\begin{aligned} \lambda_x &= \int \rho(\mathbf{r}) \left(2x^2 - y^2 - z^2\right) dV \ , \\ \lambda_y &= \int \rho(\mathbf{r}) \left(2y^2 - x^2 - z^2\right) dV \ , \\ \lambda_z &= \int \rho(\mathbf{r}) \left(2z^2 - x^2 - y^2\right) dV \ . \end{aligned} \tag{17.62}$$

It will be useful to relate these eigenvalues to the Earth's *moments of inertia*. The moment of inertia of the Earth around the z-axis is defined as:

$$C \equiv \int \rho(\mathbf{r}) \left(x^2 + y^2\right) dV \ , \tag{17.63}$$

whereas the moment of inertia around the x-axis is defined as

$$A \equiv \int \rho(\mathbf{r}) \left(y^2 + z^2\right) dV \ . \tag{17.64}$$

By symmetry the moment of inertia around the y-axis is given by the same moment A. These moments describe the rotational inertia around the coordinate axes as shown

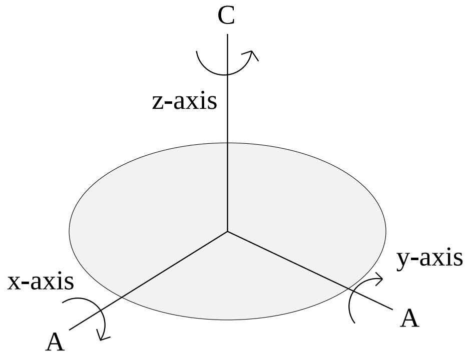


Figure 17.8: Definition of the moments of inertia A and C for an Earth with cylinder symmetry around the rotation axis.

in figure 17.8. The eigenvalues in (17.62) can be related to these moments of inertia. Because of the assumed axi-symmetric density distribution, the integral of y^2 in (17.62) is equal to the integral of x^2. The eigenvalue λ_x in (17.62) is therefore given by: $\lambda_x = \int \rho(\mathbf{r}) \left(x^2 - z^2\right) dV = \int \rho(\mathbf{r}) \left(x^2 + y^2 - y^2 - z^2\right) dV = C - A$.

Problem d: Apply a similar treatment to the other eigenvalues to show that:

$$\lambda_x = \lambda_y = C - A \quad , \quad \lambda_z = -2(C - A) \tag{17.65}$$

Problem e: Use these eigenvalues in (17.59) and use (17.56) and expression (3.7) for the unit vector $\hat{\mathbf{r}}$ to show that the quadrupole term of the potential is given by:

$$V_{qua}(\mathbf{r}) = \frac{G}{2r^3}\left(C - A\right)\left(3\cos^2\theta - 1\right) . \tag{17.66}$$

The Legendre polynomial of order 2 is given by: $P_2^0(x) = \frac{1}{2}\left(3x^2 - 1\right)$. The quadrupole term can therefore also be written as:

$$V_{qua}(\mathbf{r}) = \frac{G}{r^3}\left(C - A\right) P_2^0(\cos\theta) . \tag{17.67}$$

The term $C - A$ denotes the difference of the moments of inertia of the Earth around the rotation axis and around an axis through the equator, see figure 17.8. If the Earth would be a perfect sphere, these moments of inertia would be identical and the quadrupole term would vanish. However, the rotation of the Earth causes the Earth to bulge at the equator. This departure from spherical symmetry is responsible for the quadrupole term in the Earth's potential.

If the Earth would be spherical, the motion of satellites orbiting the Earth would satisfy Kepler's laws. The quadrupole term in the potential affects a measurable deviation of the trajectories of satellites from the orbits predicted by Kepler's laws. For example, if the potential is spherically symmetric, a satellite will orbit in a fixed plane. The quadrupole term of the potential causes the plane in which the satellite orbits to precess slightly. Observations of the orbits of satellites can therefore be used to deduce the departure of the Earth's shape from spherical symmetry[35]. Using these techniques it has been found that the difference $(C-A)$ in the moments of inertia has the numerical value[58]:

$$J_2 = \frac{(C-A)}{Ma^2} = 1.082626 \times 10^{-3} \,. \tag{17.68}$$

In this expression a is the radius of the Earth, and the term Ma^2 is a measure of the average moment of inertia of the Earth. Expression (17.68) states therefore that the relative departure of the mass distribution of the Earth from spherical symmetry is of the order 10^{-3}. This effect is small, but this number carries important information about the dynamics of our planet. In fact, the time-derivative $\dot{J}_2$ of this quantity has been measured[68] as well! This quantity is of importance because the rotation rate of the Earth slowly decreases because of the braking effect of the tidal forces. The Earth adjusts its shape to this deceleration.. The measurement of $\dot{J}_2$ therefore provides important information about the response of the Earth to a time-dependent loading.

17.8 Epilogue, the fifth force

Gravity is the force in nature that was understood first by mankind through the discovery by Newton of the law of gravitational attraction. The reason the gravitational force was understood first is that this force manifests itself in the macroscopic world in the motion of the sun, moon and planets. Later the electromagnetic force, and the strong and weak interactions were discovered. This means that presently, four forces are operative in nature.

In the 1980's, geophysical measurements of gravity suggested that the gravitational field behaves in a different way over geophysical length scales (between meters and kilometers) than over astronomical length scales ($>10^4$ km). This has led to the speculation that this discrepancy was due to a fifth force in nature. This speculation and the observations that fuelled this idea are clearly described by *Fishbach and Talmadge*[23]. The central idea is that in Newton's theory of gravity the gravitational potential generated by a point mass M is given by (17.40):

$$V_N(\mathbf{r}) = -\frac{GM}{r} \,. \tag{17.69}$$

The hypothesis of the fifth force presumes that a new potential should be added to this Newtonian potential that is given by

$$V_5(\mathbf{r}) = -\alpha \frac{GM}{r} e^{-r/\lambda} \,. \tag{17.70}$$

Note that this potential has almost the same form as the Newtonian potential $V_N(\mathbf{r})$, the main difference is that the fifth force decays exponentially with distance over a length λ and that it is weaker than Newtonian gravity by a factor α. This idea was prompted by measurements of gravity in mines, in the ice-cap of Greenland, in a 600m high telecommunication tower and a number of other experiments that seemed to disagree with the gravitational force that follows from the Newtonian potential $V_N(\mathbf{r})$.

Problem a: Effectively, the fifth force leads to a change of change of the gravitational constant G with distance. Compute the gravitational acceleration $\mathbf{g}(r)$ for the combined potential $V_N + V_5$ by taking the gradient and write the result as $-G(r)M\hat{\mathbf{r}}/r^2$

to show that the effective gravitational constant is given by:

$$G(r) = G\left(1 + \alpha\left(1 + \frac{r}{\lambda}\right)e^{-r/\lambda}\right) \ . \tag{17.71}$$

The fifth force thus effectively leads to a change of the gravitational constant over a characteristic distance λ. This effect is very small, in 1991 the value of α was estimated to be less than 10^{-3} for all estimates of λ longer than 1cm [23].

In doing geophysical measurements of gravity, one has to correct for perturbing effects such as the topography of the Earth's surface and density variations within the Earth's crust. It has been shown later that the uncertainties in these corrections are much larger than the observed discrepancy between the gravity measurements and Newtonian gravity[44]. This means that the issue of the fifth force seems be closed for the moment, and that the physical world appears to be governed again by only four fundamental forces.

Bibliography

[1] Abramowitz, M. and I.A. Stegun, 1965, Handbook of Mathematical Functions, Dover Publications, New York.

[2] Aki, K. and P.G. Richards, 1980, Quantitative seismology, volume 1, Freeman and Company, San Francsisco.

[3] Arfken, G.B., 1995 Mathematical methods for physicists, Academic Press, San Diego.

[4] Backus, M.M., 1959, Water reverberations - their nature and elimination, *Geophysics, 24,* 233-261.

[5] Barton, G., 1989, Elements of Green's functions and propagation, potentials, diffusion and waves, Oxford Scientific Publications, Oxford.

[6] Bellman, R. and R. Kalaba, 1960, Invariant embedding and mathematical physics I. Particle processes, *J. Math. Phys., 1,* 280-308.

[7] Bender, C.M. and S.A. Orszag, 1978, Advanced mathematical methds for scientists and engineers, McGraw-Hill, New York.

[8] Berry, M.V. and S. Klein, 1997, Transparent mirrors: rays, waves and localization, *Eur. J. Phys., 18,* 222-228.

[9] Berry, M.V. and C. Upstill, 1980, Catastrophe optics: Morphologies of caustics and their diffrection patterns, *Prog. Optics, 18*, 257-346.

[10] Blakeley, R.J., 1995, Potential theory in gravity and magnetics, Cambridge Univ. Press, Cambridge.

[11] Boas, M.L., 1983, Mathematical methods in the physical sciences, 2nd edition, Wiley, New York.

[12] Brack, M. and R.K. Bhaduri, 1997, Semiclassical physics, Addison-Wesley, Reading MA.

[13] Broglie, L. de, 1952, La théorie des particules de spin 1/2, Gauthier-Villars, Paris.

[14] Butkov, E., 1968, Mathematical physics, Addison Wesley, Reading MA.

[15] Claerbout, J.F., 1976, Fundamentals of geophysical data processing, McGraw-Hill, New york.

[16] Chopelas, A., 1996, Thermal expansivity of lower mantle phases MgO and $MgSiO_3$ perovskite at high pressure derived from vibrational spectroscopy, *Phys. Earth. Plan. Int., 98*, 3-15.

[17] Dahlen, F.A., 1979, The spectra of unresolved split normal mode multiples, *Geophys. J.R. Astron. Soc., 58*, 1-33.

[18] Dahlen, F.A. and I.H. Henson, 1985, Asymptotic normal modes of a laterally heoterogeneous Earth, *J. Geophys. Res., 90*, 12653-12681.

[19] Dziewonski, A.M., and J.H. Woodhouse, 1983, Studies of the seismic source using normal-mode theory, in *Earthquakes: Observations, theory and intepretation*, edited by H. Kanamori and E. Boschi, North Holland, Amsterdam, 45-137.

[20] Edmonds, A.R., 1974, Angular momentum in quantum mechanics, 3rd edition, Princeton Univ. Press, Princeton.

[21] Feynman, R.P., 1975, The character of physical law, MIT Press, Cambridge (MA).

[22] Feynman, R.P. and A.R. Hibbs, 1965, Quantum mechanics and path integrals, McGraw-Hill, New York.

[23] Fishbach, E. and C. Talmadge, 1992, Six years of the fifth force, *Nature, 356*, 207-214.

[24] Fletcher, C., 1996, The complete walker III, Alfred A. Knopf, New York.

[25] Gradshteyn, I.S. and I.M. Rhyzik, 1965, Tables of integrals, series and products, Academic Press, New York.

[26] Gubbins, D. and R. Snieder, 1991, Dispersion of P waves in subducted lithosphere: Evidence for an eclogite layer, *J. Geophys. Res., 96*, 6321-6333.

[27] Guglielmi, A.V. and O.A. Pokhotelov, 1996, Geoelectromagnetic waves, Inst. of Physics Publ., Bristol.

[28] Halbwachs, F., 1960, Théorie relativiste des fluides a spin, Gauthier-Villars, Paris.

[29] Hildebrand, A.R., M. Pilkington, M. Conners, C. Ortiz-Aleman and R.E. Chavez, 1995, Size and structure of the Chicxulub crater revealed by horizontal gravity and cenotes, *Nature, 376,* 415-417.

[30] Holton, J.R., 1992, An inroduction to dynamic meteorology, Academic Press, San Diego.

[31] Jackson, J.D., 1975, Classical electrodynamics, Wiley, New York.

[32] Kermode, A.C., 1972, Mechanics of flight, Longmanm Singapore.

[33] Kline, S.J., 1965, Similitude and approximation theory, McGraw-Hill, New York.

[34] Kravtsov, Yu.A., 1988, Ray and caustics as physical objects, *Prog. Optics, 26,* 228-348.

[35] Lambeck, K., 1988, Geophysical geodesy, Oxford University Press, Oxford.

[36] Lauterborn, W., T. Kurz and M. Wiesenfeldt, 1993, Coherent optics, Springer Verlag, Berlin.

[37] Marchaj, C.A., 1993, Aerohydrodynamics of sailing, 2nd edition, Adlard Coles Nautical, London.

[38] Marsden, J.E. and A.J. Tromba, 1988, Vector calculus, Freeman and Company, New York.

[39] Merzbacher, E., 1970, Quantum mechanics, Wiley, New York.

[40] Moler, C. and C. van Loan, 1978, Nineteen dubious ways to comopute the exponential of a matrix, *SIAM Review, 20*, 801-836.

[41] Morley, T., 1985, A simple proof that the world is three dimensional, *SIAM Review, 27*, 69-71.

[42] Ohanian, H.C. and R. Ruffini, 1994, Gravitation and Spacetime, Norton, New York.

[43] Olson, P., 1989, Mantle convection and plumes, in *The encyclopedia of solid earth geophysics*, Ed. D.E. James, Van Nostrand Reinholt, New York.

[44] Parker, R.L. and M.A. Zumberge, 1989, An analysis of geophysical experiments to test Newton's law of gravity, *Nature, 342*, 29-31.

[45] Parsons, B. and J.G. Sclater, 1977, An analysis of the variation of the ocean floor bathymetry and heat flow with age, *J. Geophys. Res., 32*, 803-827.

[46] Pedlosky, J., 1979, Geophysical Fluid Dynamics, Springer Verlag, Berlin.

[47] Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, 1986, Numerical Recipes, Cambridge Univ. Press, Cambridge.

[48] Rayleigh, Lord, 1917, On the reflection of light from a regularly stratified medium, *Proc. Roy. Soc. Lon., A93*, 565-577.

[49] Rossing, T.D., 1990, The science of sound, Addison Wesley, Reading MA.

[50] Rummel, R., 1986, Satellite gradiometry, in *Mathematical techniques for high-resolution mapping of the gravitational field, Lecture notes in the Earth Sciences, 7*, Ed. H. Suenkel, Springer Verlag, Berlin.

[51] Robinson, E.A. and S. Treitel, 1980, Geophysical singnal analysis, Prentive Hall, Englewood Cliffs NJ.

[52] Sakurai, J.J., 1978, Advanced quantum mechanics, Addison Wesley, Reading (MA).

[53] Schneider, W.A., 1978, Integral formulation for migration in two and three dimensions, *Geophysics, 43,* 49-76.

[54] Silverman, M.P., 1993, And yet it moves; strange systems and subtle questions in physics, Cambridge University Press, Cambridge (UK).

[55] Snieder, R.K., 1985, The origin of the 100,000 year cycle in a simple ice age model, *J. Geophys. Res., 90,* 5661-5664.

[56] Snieder, R., 1996, Surface wave inversions on a regional scale, in *Seismic modelling of Earth structure,* Eds. E. Boschi, G. Ekstrom and A. Morelli, Editrice Compositori, Bologna, 149-181.

[57] Snieder, R. and G. Nolet, 1987, Linearized scattering of surface wave on a spherical Earth, *J. Geophys., 61,* 55-63.

[58] Stacey, F.D., 1992, Physics of the Earth, 3ed., Brookfield Press, Brisbane.

[59] Tabor, M., 1989, Chaos and integrability in nonlinear dynamics, John Wiley, New York.

[60] Tritton, D.J., 1982, Physical fluid dynamics, Van Nostrand Reinhold, Wokingham (UK).

[61] Tromp, J. and R. Snieder, 1989. The reflection and transmission of plane P- and S-waves by a continuously stratified band: a new approach using invariant embedding, *Geophys. J., 96,* 447-456.

[62] Turcotte, D.L. and G. Schubert, 1982, Geodynamics, Wiley, New York.

[63] Virieux, J., 1996, Seismic ray tracing, in *Seismic modelling of Earth structure,* Eds. E. Boschi, G. Ekstrom and A. Morelli, Editrice Compositori, Bologna, 221-304.

[64] Watson, T.H., 1982, A real frequency, wave-number analysis of leaking modes, *Bull. Seismol. Soc. Am., 62,* 369-394.

[65] Webster, G.M. (Ed.), 1981, Deconvolution, *Geophysics reprint series, 1,* Society of Exploration Geophysicists, Tulsa OK.

[66] Whitham, G.B., 1974, Linear and nonlinear waves, Wiley, New York.

[67] Whitaker, S., 1968, Introduction to fluid mechanics, Prentice-Hall, Englewood Cliffs.

[68] Yoder, C.F., J.G. Williams, J.O. Dickey, B.E. Schutz, R.J. Eanes and B.D. Tapley, Secular variation of the Earth's gravitational harmonic J_2 coefficient from LAGEOS and nontidal acceleration of the Earth rotation, *Nature, 303*, 757-762, 1983.

[69] Ziolkowski, A., 1991, Why don't we measure seismic signatures? *Geophysics, 56*, 190-201.

[70] Zumberge, M.A., J.R. Ridgway and J.A. Hildebrand, 1997, A towed marine gravity meter for near-bottom surveys, *Geophysics, 62,* 1386-1393.